

# An Asymmetric Complex of Restriction Endonuclease MspI on Its Palindromic DNA Recognition Site

Qian Steven Xu,<sup>1,4</sup> Rebecca B. Kucera,<sup>2</sup>  
Richard J. Roberts,<sup>2</sup> and Hwai-Chen Guo<sup>1,\*</sup>

<sup>1</sup>Department of Physiology and Biophysics  
Boston University School of Medicine  
715 Albany Street  
Boston, Massachusetts 02118

<sup>2</sup>New England Biolabs  
32 Tozer Road  
Beverly, Massachusetts 01915

## Summary

Most well-known restriction endonucleases recognize palindromic DNA sequences and are classified as Type IIP. Due to the recognition and cleavage symmetry, Type IIP enzymes are usually found to act as homodimers in forming 2-fold symmetric enzyme-DNA complexes. Here we report an asymmetric complex of the Type IIP restriction enzyme MspI in complex with its cognate recognition sequence. Unlike any other Type IIP enzyme reported to date, an MspI monomer and not a dimer binds to a palindromic DNA sequence. The enzyme makes specific contacts with all 4 base pairs in the recognition sequence, by six direct and five water-mediated hydrogen bonds and numerous van der Waal contacts. This MspI-DNA structure represents the first example of asymmetric recognition of a palindromic DNA sequence by two different structural motifs in one polypeptide. A few possible pathways are discussed for MspI to cut both strands of DNA, either as a monomer or dimer.

## Introduction

Restriction endonucleases (REases) are components of restriction-modification systems that protect bacteria from invading foreign DNA (Wilson and Murray, 1991; Bickle and Kruger, 1993). REases are, nonetheless, best known for their roles in recombinant DNA technology and genetic manipulation. More than 3570 REases with 240 unique specificities have been biochemically characterized, many of them recognizing the same DNA sequence (Roberts et al., 2003b). However, of the 200 or so restriction enzyme genes that have been sequenced, little sequence similarity has been observed, suggesting diverse strategies for DNA recognition, even for the same DNA sequence. REases have been categorized into four groups, called Type I, II, III, and IV (Smith and Nathans, 1973; Roberts et al., 2003a). Of those characterized, over 98% of them belong to the Type II enzymes that recognize specific double-strand DNA sequences and cleave within or close to their recognition sites, and do not require ATP for their nucleolytic activity (Roberts and Halford, 1993; Pingoud and Jeltsch, 2001). Of the

588 commercially available REases, a great majority recognize double-strand DNA sequences of 4 to 8 base pairs with a dyad axis of symmetry, termed palindromes, and in the presence of Mg<sup>2+</sup> cleave both strands of the DNA at fixed symmetrical locations to generate “sticky” or blunt ends suitable for molecular cloning. These symmetrically cleaving enzymes are grouped into a subtype, called Type IIP (Roberts et al., 2003a). Biochemical and structural evidence for several Type IIP prototype enzymes indicate that they form symmetric dimers or tetramers to recognize their palindromic recognition sequences (Aggarwal, 1995; Siksnys et al., 1999; Hsieh et al., 2000; Deibert et al., 2000).

Crystal structures of 13 Type IIP REases have been reported before, and 11 of them are with structures of enzyme-DNA complexes (Roberts et al., 2003b). All these enzymes form dimers or tetramers in crystals either as free enzyme or in complex with specific or non-specific DNA. Comparisons of these structures indicate that domain organizations (and perhaps coupling of recognition to catalytic sites of enzyme) and intermonomer interactions (dimer structure) correlate well with their cleavage patterns on DNA (Guo, 2003). Thus, enzymes in the existing pool with different cleavage patterns are likely to reflect different control mechanisms that couple DNA recognition to DNA cleavage and/or dimer structures. To examine this hypothesis, our laboratory has embarked on a systematic study of REases that cleave DNA with unique patterns, starting with the MspI REase.

MspI, a Type IIP REase from *Moraxella* species (Nwan-kwo and Wilson, 1988; Lin et al., 1989), produces a different cleavage pattern from those of known structures. It recognizes the palindromic tetranucleotide sequence 5'-CCGG and cleaves between the first and second nucleotides, leaving 2 base 5' overhangs. A molecule of MspI has a molecular mass of 29 kDa and consists of 262 amino acid residues. MspI displays no significant sequence homology to other REases, with the one exception of its isoschizomer BsuFI (Kapfer et al., 1991) that carries the same DNA specificity and cleavage pattern as MspI. The crystal structure of an MspI-DNA complex has now been determined at 1.95 Å resolution. Unexpectedly, MspI interacts with the tetranucleotide sequence as a monomer. This is the first example illustrating that binding of the palindromic sequence by REase can be achieved by two different structural motifs in one polypeptide.

## Results and Discussion

### Structure of MspI-DNA Complex

The structure of MspI is the first crystal structure of a restriction enzyme that recognizes a tetranucleotide sequence. MspI molecules fold into an  $\alpha/\beta$  architecture, similar to that of BglI and EcoRV (both recognize hexanucleotide sequence) (Figure 1). It contains a five-stranded mixed  $\beta$  sheet sandwiched on both sides by  $\alpha$  helices, with approximate dimensions of 60 × 55 ×

\*Correspondence: hcguo@bu.edu

<sup>3</sup>Present address: Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720.

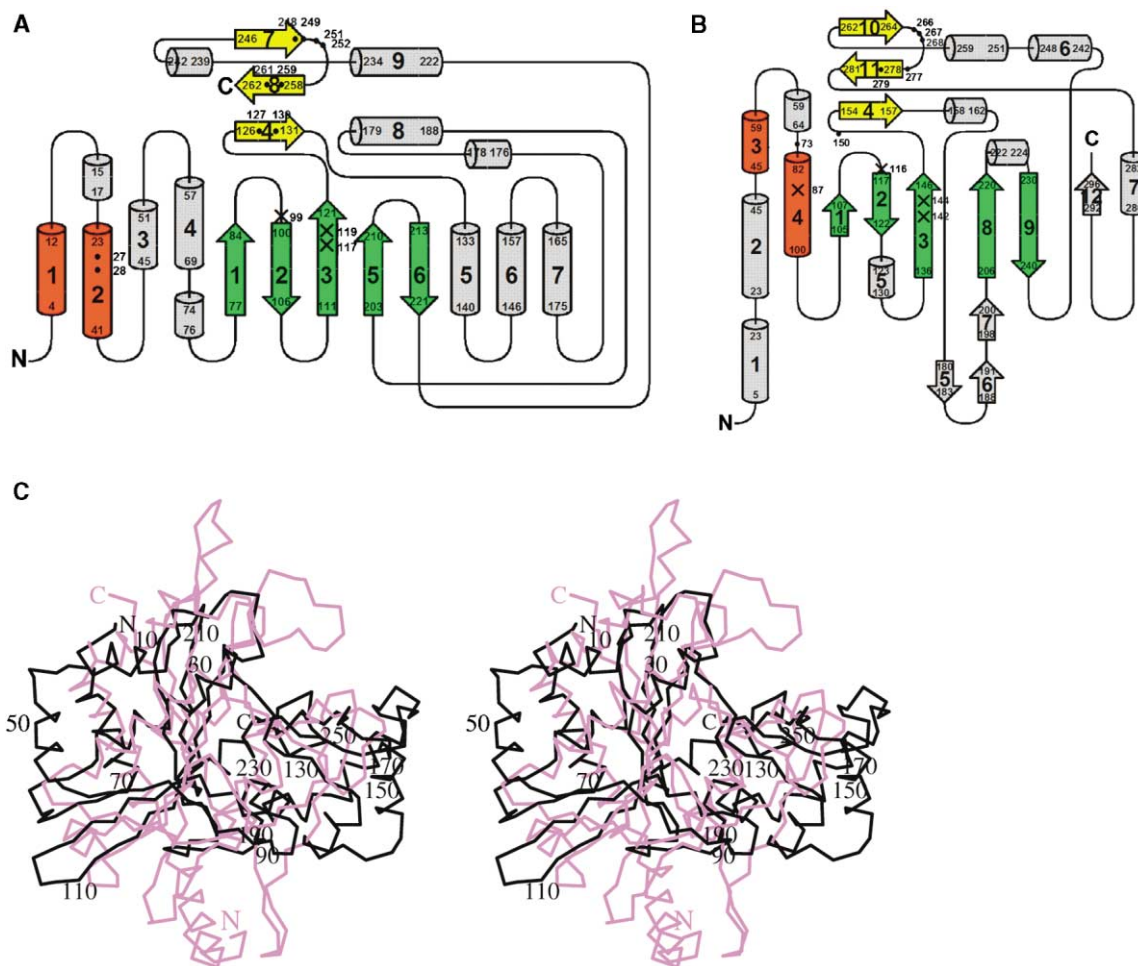


Figure 1. Structural Comparison of the  $\alpha/\beta$  Core Motif in MspI and BglI

(A and B) Schematic diagrams show the secondary structure elements of MspI (A) and BglI (B). The arrows mark  $\beta$  strands, numbered cylinders  $\alpha$  helices, and unnumbered cylinders  $3_{10}$  helices. The extent of each secondary structure element is indicated by residue numbers. Structurally equivalent regions are colored in red for N-terminal helices, green for the five-stranded central  $\beta$  sheet, and yellow for the  $\beta$  sheet involved in DNA recognition. The locations of the amino acids participating in DNA recognition and cleavage are marked with black dots and crosses, respectively.

(C) Stereo diagram of C $\alpha$  backbones of MspI (black) and BglI (magenta), superimposed by the common  $\alpha/\beta$  core motifs (colored in red, green, and yellow in [A] and [B]). Residue numbers of MspI are marked at every twentieth C $\alpha$ . The PDB code of the BglI structure used for comparison is 1DMU.

40 Å (Figure 2A). Within the sheet the strands  $\beta 1$ ,  $\beta 2$ , and  $\beta 3$  run antiparallel forming a  $\beta$ -meander that contains the potential catalytic residues Asp99, Asn117, and Lys119. Two N-terminal  $\alpha$  helices,  $\alpha 1$  and  $\alpha 2$ , connected by a  $3_{10}$  helix, are oriented antiparallel, packing against the concave surface of the central  $\beta$  sheet and protruding about 15 Å out of the molecule, while the following helices  $\alpha 3$  and  $\alpha 4$  are placed in a clamp-like fashion and linked to the first  $\beta$  strand of the central sheet through a second  $3_{10}$  helix. On the convex side of the sheet, three short  $\beta$  strands ( $\beta 4$ ,  $\beta 8$ , and  $\beta 7$ ) form a smaller antiparallel sheet that makes specific contacts with the DNA bases in the major groove. The loop connecting strands  $\beta 7$  and  $\beta 8$  sticks out of the subunit by nearly 10 Å. Between helices  $\alpha 7$  and  $\alpha 8$  is a proline (Pro177) that results in these two helices being kinked by approximately 80°. Together with the other three helices,  $\alpha 5$ ,

$\alpha 6$ , and  $\alpha 9$ , they are assembled as a five-helix cluster, in which four helices, except  $\alpha 9$ , act as the crossover between strands  $\beta 3$  and  $\beta 5$ . The cluster is covered by the three-stranded sheet at the top like a lid, forming a big hydrophobic core consisting of a number of aromatic residues. The DNA binding cleft is formed by the N-terminal region of the  $\alpha 2$  helix and the C-terminal region of the  $\beta 3$  strand on one side, and the three-stranded  $\beta$  sheet ( $\beta 4$ ,  $\beta 8$  and  $\beta 7$ ) and the connecting loops on the other side. It exhibits a U-shape appearance when viewed down the DNA helix, caused by the protruding portions of the  $\alpha 2$  helix and the loop connecting strands  $\beta 7$  and  $\beta 8$ . The DNA is bound with the major groove facing the three-stranded  $\beta$  sheet and the minor groove primarily exposed to the solvent. The DNA duplexes are primarily in B-form with no major bends or kinks.

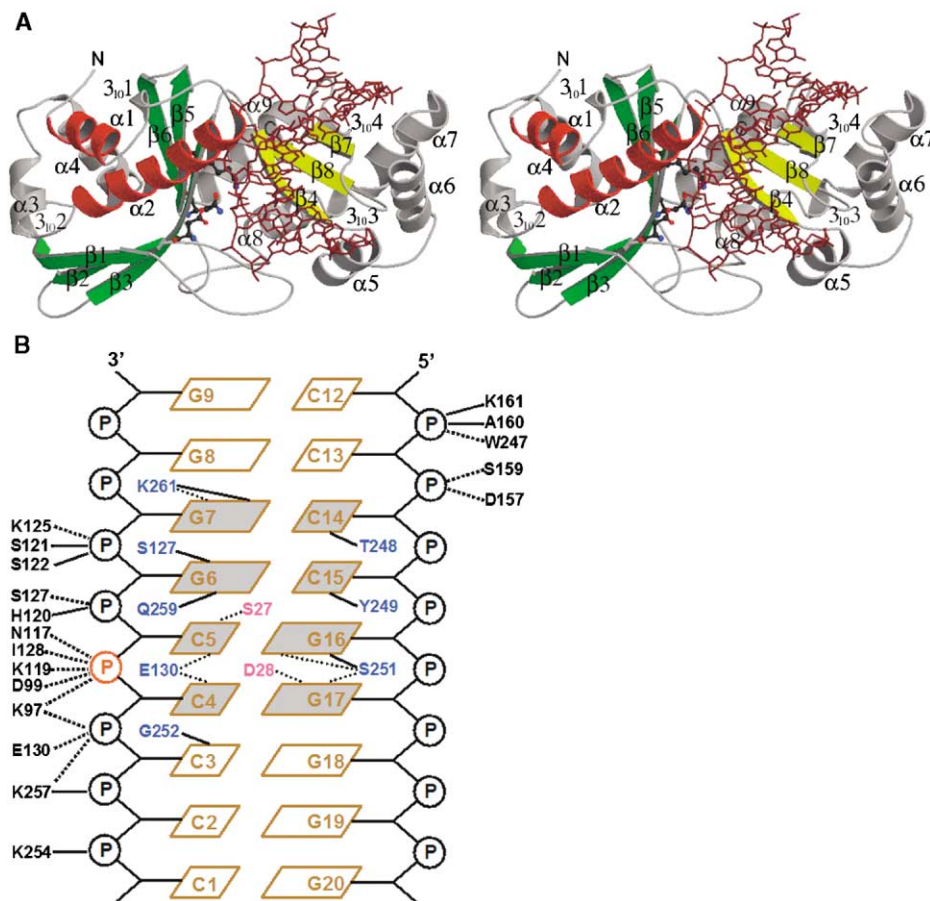


Figure 2. MspI-DNA Interactions

(A) Stereoview of the tertiary structure of the MspI-DNA complex. The enzyme is represented as ribbons and the DNA as a brown stick model. Secondary structure elements of MspI are labeled and the conserved structural core are colored as in Figure 1. Side chains of the catalytic site residues (Asp99, Asn117, and Lys119) are shown as ball-and-stick representations, with carbons in black, nitrogens in blue, and oxygens in red.

(B) Schematic diagram of hydrogen bonding between MspI and DNA. The DNA recognition sequence is shaded in gray, with the scissile phosphate (C4-C5) circled in red. One DNA base pair (G10:C11) is omitted in the final model (see Experimental Procedures). Blue and pink represent amino acids that bind to DNA bases in the major and minor groove, respectively. Amino acids that bind to the phosphate backbone of DNA are colored in black. Solid lines represent direct hydrogen bonds: Ser127-G6 (OG-N7), Thr248-C14 (OG1-N4), Tyr249-C15 (O-N4), Ser251-G16 (OG-N7), Gln259-G6 (NE2-O6), and Lys261-G7 (NZ-O6) in the recognition sequence; and Gly252-C3 (O-N4) outside the recognition sequence. Dotted lines represent water-mediated hydrogen bonds: Glu130-C4 (OE1-N4) through a water that also contacts Gly252 (N), Glu130-C5 (OE1-N4) through another water that also contacts Gln259 (NE2), Ser251-G16 (N-O6), Ser251-G17 (OG-O6), and Lys261-G7 (NZ-N7) in the major groove; and Ser27-C5 (OG-O2) and Asp28-G17 (OD1-N2) in the minor groove.

### Asymmetric DNA Recognition

In the MspI-DNA complex, the enzyme approaches and recognizes DNA mainly from the major groove. Since MspI recognizes a palindromic sequence and cleaves the DNA at both strands symmetrically, there was an expectation that, like all other structures of Type IIP REases reported to date, MspI would bind to its cognate sequence as a homodimer. In contrast to this expectation, there is only one MspI molecule on each DNA palindrome (Figure 2A). Although there are two monomers of MspI in the asymmetric unit of the crystals, they bind to two separate DNA palindromes. As a result, two symmetric half-sites of DNA are recognized asymmetrically by the MspI monomer (Figure 2B). Within the 5'-CCGG palindromic recognition sequence, one outer base pair (G7:C14) is directly recognized by Thr248 and Lys261.

In addition, Lys261 also makes a water-mediated hydrogen bond to the N7 atom of the guanine base. However, the other dyad-related outer base pair (G17:C4) has no direct contact with the protein, but instead makes water-mediated hydrogen bonds to Ser251 and Glu130. Similarly, one inner base pair of the recognition sequence (G6:C15) forms direct hydrogen bonds with Ser127, Tyr249, and Gln259, whereas the dyad-related base pair (G16:C5) is involved in a direct hydrogen bonding with Ser251 and water-mediated interactions with Glu130 and Ser251. Together, there are six direct and five water-mediated hydrogen bonds from the MspI monomer to the 4 base pair recognition sequence, which nearly saturates the hydrogen-bonding potential in the major groove of the recognition site. With the exception of the N7 atom of one outer guanine base (G17), all atoms with

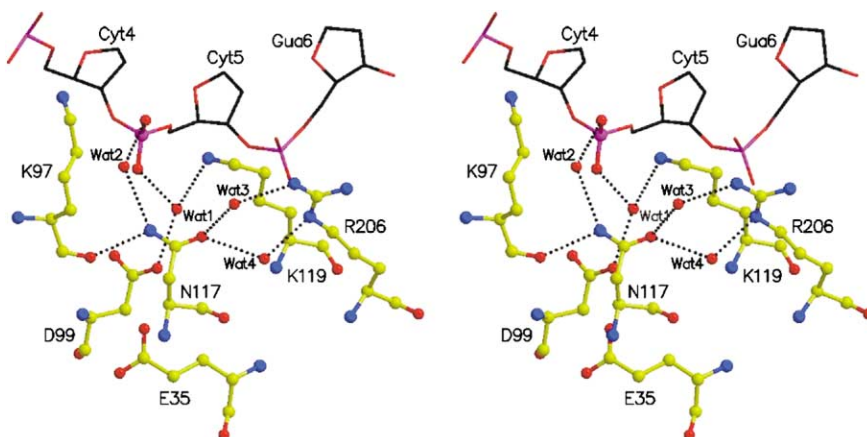


Figure 3. Stereoview of the MspI Catalytic Site on the Scissile Phosphodiester Bond

Active site residues and the DNA phosphate with the scissile bond are shown as ball-and-stick models. Other parts of a small segment of DNA juxtaposing the scissile bond are shown as a stick model, with bases omitted for clarity. Atom colors are as follows: nitrogen in blue, oxygen in red, phosphorus in purple, and carbon in yellow and black for MspI and DNA, respectively. Dashed lines indicate hydrogen bonds.

hydrogen bonding potential on the major groove side of the recognition sequence have direct or indirect contacts with the enzyme. In addition, the enzyme makes specific interactions with the DNA bases through many van der Waals contacts (Xu, 2003). Similar to the hydrogen bond patterns, the two dyad-related half-sites on DNA make asymmetric van der Waals contacts with different motifs in one MspI polypeptide. All the amino acids involved in specific contacts with the DNA within the major groove are located on, or near to, the small three-stranded recognition  $\beta$  sheet, comprising strands  $\beta 4$ ,  $\beta 7$ , and  $\beta 8$  (Figure 1). From the minor groove, the enzyme also makes two water-mediated hydrogen bonds to one half-site (C5 and G17) of the recognition sequence, using Ser27 and Asp28, respectively. Together, the bases within the recognition sequence are recognized predominantly in the major groove by the MspI molecule, with two indirect (water-mediated) hydrogen bonds from the minor groove of DNA. Outside the recognition sequence, MspI forms one hydrogen bond with the DNA, between the main chain carbonyl oxygen of Gly252 and the exocyclic amino group of the C3 base. Furthermore, MspI makes extensive contacts with the sugar-phosphate backbone of the DNA (Figure 2B), mainly within the recognition sequence and around the enzyme active site. Thus, these protein-phosphate interactions provide additional binding energy to position the active site at the scissile phosphate bond. Although homodimers had been previously reported to bind pseudopalindromic DNA sequences asymmetrically (Chen et al., 1998; Jin et al., 2003), the MspI-DNA complex presented here represents the first structure of a palindromic DNA sequence bound by an asymmetric protein monomer.

#### Unique Catalytic Site

This MspI-DNA complex also reveals unique features at the catalytic site. The location and essential residues of the active site have been identified in several REases. They make use of the same or similar type of active

center, represented by the PDX<sub>10-30</sub>(D/E)XK signature sequence (Anderson, 1993). Nonetheless, such a signature sequence was not found in MspI, nor was a potential metal binding site identified in this enzyme by the Fenton chemistry technique (Hlavaty et al., 2000). The structural equivalent of this motif in MspI is T<sup>96</sup>DX<sub>17</sub>NXXK, with Asn117 in the place of an Asp or Glu found in other enzymes (Figure 3). Although uncommon, asparagine has been found to bind the catalytic magnesium in other endonucleases (Flick et al., 1998; Miller et al., 1999). Therefore, Asn117 may play a similar role as the second acidic residue in the signature sequences of other REases, to coordinate magnesium ion. Also similar to a crystal structure of EcoRV-DNA-Ca<sup>2+</sup> complex (Horton et al., 1998), a bound water molecule is a possible candidate for the attacking nucleophile. This water molecule (Wat1) forms hydrogen bonds to the carboxylate of Asp99 and the pro-S<sub>P</sub>-oxygen of the scissile phosphate and is further stabilized and fixed by the side chain amine group of Lys119. It is poised at a distance of 3.7 Å from the scissile phosphorus and makes the water-P-O3' angle approximately 146°. When the active site of MspI was compared with that of EcoRV, it appears that the Asp99 of MspI (the first D in the motif) probably moves closer to the scissile phosphate. In fact, in MspI, the Asp99 contacts directly with the potential nucleophilic water molecule and could potentially serve as the general base, although its pK<sub>a</sub> appears to be not optimal. Alternatively, a magnesium-bound water can serve as the general base.

#### Comparison with Other Restriction Enzymes

Structural comparisons of the MspI monomer with other Type II REases indicate that MspI closely matches the BglI structure (Newman et al., 1998), with a rms deviation of 2.9 Å for 97 C $\alpha$  atoms that encompass both the DNA recognition and catalytic elements (Figure 1C). The common structural elements include a five-stranded  $\beta$  sheet (colored green in Figures 1 and 2A), two  $\alpha$  helices (colored red in Figures 1 and 2A), and three  $\beta$  strands (col-

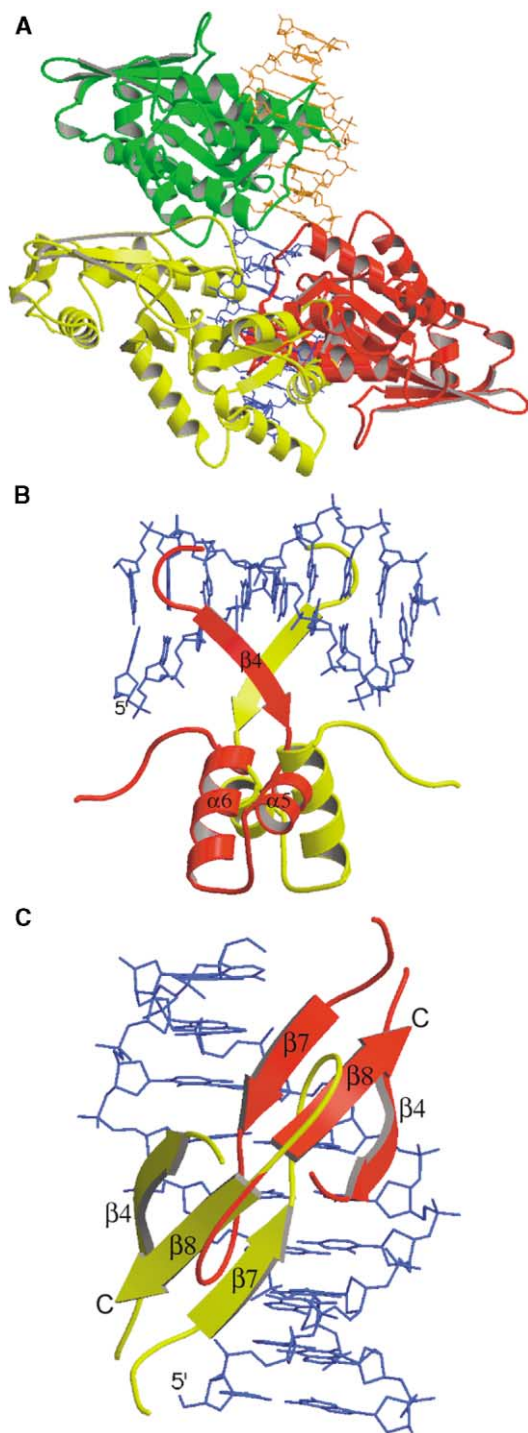


Figure 4. Crystallographic and Putative Dimer Structures of MspI-DNA Complex

(A) The relationship between a putative MspI dimer and the crystallographic dimer in an asymmetric unit of the MspI-DNA cocrystal. The enzymes are represented as ribbons and the DNAs as stick models. In a crystal, two MspI monomers (green and red) bind to two separate recognition sequences (brown and blue). To generate a putative dimer (gold and red), the red monomer was rotated 180° along the dyad axis (perpendicular to the paper plane) of the palindromic DNA (blue).

(B) A four-helix bundle forms the putative MspI dimer on the palindromic recognition sequence. The secondary structure elements

ored yellow in Figures 1 and 2A) of the smaller  $\beta$  sheet for DNA recognition. It is noteworthy that MspI cleaves DNA giving a 2 base 5' overhang, whereas BglI cleaves DNA giving a 3 base 3' overhang. However, MspI and BglI share the same distance between the DNA recognition half-sites and the corresponding cleavage points (1 bp), which is distinct from any other Type II REases with known structures (Guo, 2003). Therefore, the new MspI monomer structure appears to support our hypothesis that the structural organization and possibly the coupling between DNA recognition and catalysis are determined by the distance between the DNA recognition half-sites and the corresponding cleavage points (Guo, 2003). However, as pointed out above, these two enzymes cleave DNA with different patterns of stagger of their cleavage points on the two DNA strands. This difference in stagger could be achieved by some structural differences between them, presumably related to dimerization (see below). In MspI, helix  $\alpha 2$  and the first strand in the central  $\beta$  sheet ( $\beta 1$ ) are connected by two  $\alpha$  helices  $\alpha 3$  and  $\alpha 4$  and a  $3_{10}$  helix, whereas in BglI the structurally equivalent  $\alpha$  helix and  $\beta$  strand are connected directly by a small loop. Likewise, a four-helix bundle consisting of  $\alpha 5$ ,  $\alpha 6$ ,  $\alpha 7$ , and  $\alpha 8$  acts as the crossover between strands  $\beta 3$  and  $\beta 5$  of MspI, to which no similar structure has been previously observed in any other Type II REases. Moreover, helix  $\alpha 3$  of MspI displays almost perfect spatial overlap with helix  $\alpha 2$  of BglI, but they run in the opposite orientation.

#### Possible Mechanisms for DNA Cleavage

The observation of a monomeric MspI-DNA complex in crystals (where MspI is in 18 mM concentration) suggests that such a unique DNA recognition by MspI is likely to occur in an optimal reaction solution where MspI concentrations are much lower (in submicromolar range). Interestingly, in a sedimentation velocity experiment with a 17  $\mu$ M MspI solution, the enzyme appears to exist as a homogeneous monomer (Xu, 2003). How then can MspI manage to cut both strands of DNA symmetrically? Since there is only one catalytic site in an MspI monomer, some kind of enzyme dimerization or flipping is necessary to cleave both strands of DNA at the recognition site. In a second preliminary study of sedimentation velocity experiment with a 10 base pair DNA containing the MspI recognition site, the enzyme appears to bind to DNA with a monomer-dimer equilibrium. It is thus plausible that MspI binds to its recognition sequence through the monomer-DNA intermediate captured in the crystals and then proceeds to a final homodimer-DNA complex responsible for cleavage of both strands of DNA. To generate such a putative MspI homodimer bound on a single palindromic recognition site, the crystal structure of MspI monomer was rotated 180° along the dyad axis of the palindromic DNA for the second monomer (Figure 4A). In the putative dimer, two

from two MspI monomers are colored as in (A) (gold and red). Other parts of the enzyme are omitted for clarity.

(C) Steric clashes between the two MspI monomers in the putative dimer structure. These clashes, however, can be avoided by some conformational adjustments (see text).

Table 1. Data Collection and Refinement Statistics

|   |                |
|---|----------------|
| Resolution (Å)  | 42.6–1.95      |
| R <sub>sym</sub> (%) <sup>a,b</sup>                   | 4.3 (27.5)     |
| Total observations                                    | 153,014        |
| Unique reflections                                    | 52,241         |
| Avg. redundancy                                       | 2.9            |
| <I/σ> <sup>a</sup>                                    | 24.5 (3.6)     |
| Completeness (%) <sup>a</sup>                         | 97.4 (99.9)    |
| <b>Refinement</b>                                     |                |
| Reflections (working/test)                            | 46,088/5,172   |
| R <sub>cryst</sub> /R <sub>free</sub> <sup>c</sup>    | 0.223/0.252    |
| Protein residues/DNA base pairs                       | 524/18         |
| Nonhydrogen atoms built in the asymmetric unit        |                |
| Protein   | 4,188          |
| DNA   | 726            |
| Water molecules                                       | 331            |
| Ions  | 2              |
| Rms deviation   |                |
| Bonds (Å)   | 0.005          |
| Angles (°)  | 1.16           |
| NCS related (Å) <sup>d</sup> (protein/DNA/complex)    | 0.54/0.16/0.52 |
| Average B value (Å <sup>2</sup> ) (protein/DNA/water) | 27.9/27.2/29.2 |
| Ramachandran analysis                                 |                |
| Most favored  | 429 (90.5%)    |
| Allowed   | 45 (9.5%)      |

<sup>a</sup>Numbers in parentheses are for highest resolution shell of 2.02–1.95 Å.

<sup>b</sup> $R_{sym} = (\sum_h \sum_i |I_{hi} - \langle I_h \rangle|) / \sum_h \sum_i I_{hi}$ , where  $\langle I_h \rangle$  is the average intensity of  $n$  independent observations  $I_{hi}$  of a given reflection  $h$ .

<sup>c</sup> $R_{cryst} = \sum_h |F_{obs} - F_{calc}| / \sum_h |F_{obs}|$ , where  $F_{obs}$  and  $F_{calc}$  are the observed and calculated structure factor amplitudes, respectively. The  $R_{free}$  was calculated using a randomly selected 10% reflections that were omitted from all stages of the refinement.

<sup>d</sup>The rms deviations between the two NCS-related MspI-DNA complexes in the asymmetric unit were calculated using all nonhydrogen atoms of protein subunits, DNA molecules, or entire protein-DNA complexes.

helices from each monomer ( $\alpha 5$  and  $\alpha 6$ ) pack against each other to form a four-helix bundle (Figure 4B). There are however some steric clashes (Figure 4C), mainly between the recognition  $\beta$  sheet (strands  $\beta 4$ ,  $\beta 7$ , and  $\beta 8$  in Figure 1). Nonetheless, these steric clashes could be avoided by some conformational adjustments on protein and/or DNA bending, as exemplified in other REases (Horton and Perona, 2000). Thus it is plausible that, like other Type IIP enzymes, MspI could bind to its recognition sequence as a homodimer to achieve double-strand DNA cleavage. Based on this putative MspI dimer, it appears to be a novel dimerization mode reflecting its unique pattern of stagger (2 base 5'-overhangs). Alternatively, MspI may achieve double-strand DNA cleavage by bringing together two monomers bound on two separate recognition sites, similar to the model proposed for FokI (Wah et al., 1998; Bitinaite et al., 1998; Vanamee et al., 2001). A third possibility is that it undergoes some conformational rearrangements to use a single active site of MspI to sequentially cut both DNA strands, similar to that proposed for BfiI (Sasnauskas et al., 2003). More studies are needed to elucidate the detailed mechanism of MspI.

#### Experimental Procedures

##### Crystal Preparation and Data Collection

Crystals of MspI in complex with a DNA decamer containing the MspI recognition sequence were grown as previously described

(O'Loughlin et al., 2000). In the crystals, DNA cleavage was prevented by removing the required metal ion ( $Mg^{2+}$ ) with EDTA. The crystals of MspI-DNA complexes belong to the monoclinic space group  $P2_1$ , with unit cell constants  $a = 50.2$  Å,  $b = 131.6$  Å,  $c = 59.3$  Å, and  $\beta = 109.7^\circ$ . A high-resolution X-ray data set has now been collected up to 1.95 Å at 100 K at the National Synchrotron Light Source (NSLS) beamline X12C, Brookhaven National Laboratory. Data were processed using the HKL suite (Otwinowski and Minor, 1997), with  $R_{sym} = 4.3\%$  (highest resolution shell, 2.02–1.95 Å, = 27.5%) and overall completeness of 97.4% (highest resolution shell, 99.9%), and were converted to structure factors using TRUNCATE from the CCP4 suite (CCP4, 1994).

#### Structure Determination and Refinement

The complex structure is determined by a combination of MIRAS (multiple isomorphous replacement with anomalous scattering, with a figure of merit of 0.45 at 2.7 Å resolution using Hg and Sm derivatives) and MAD (multiwavelength anomalous diffraction) methods with an overall figure of merit of 0.79 after density modification at 2.7 Å resolution (Xu, 2003). An initial  $C\alpha$  backbone trace was built into the first 2.7 Å map using the program O (Jones et al., 1991). The model was subsequently completed and refined via rigid body, positional, group, and restrained individual temperature factor refinement, as well as simulated annealing using a slow-cooling protocol in CNS (Brunger et al., 1998). Program SIGMAA (CCP4, 1994) was used in the early cycles of refinement and manual building to combine model phases with experimental phases. Initially, strict NCS constraints or tight NCS restraints were applied, and in later stages of refinement, NCS restraints were loosened. After a few rounds of model rebuilding, stepwise resolution extension, and automated refinement, the final structure at 1.95 Å resolution contains all protein residues, 18 base pairs within the pseudocontinuous 19 bp DNA duplex, 331 solvent molecules, and 2 sodium ions. The side chains of Lys32, Gln169, Lys254 in subunit A and Lys193 and Lys254 in subunit B were partially disordered. The base pair at the joint region between two 10 bp DNA molecules was not built, as the density around there is actually the average of two different base pairs. Criteria for the inclusion of water molecules were the appearance of peaks at  $4.0 \sigma$  in ( $F_o - F_c$ ) maps,  $1.0 \sigma$  in ( $2F_o - F_c$ ) maps, and at least one hydrogen-bonding interaction with protein or DNA or first-layer water molecules. Any water with a temperature factor greater than  $50 \text{ \AA}^2$  was deleted. The statistics of the final structure are shown in Table 1.

#### Acknowledgments

We thank Dr. Ira Schildkraut for collaboration at the early stage of this research. We also thank Dr. Thomas M. Laue for providing assistance with analytical ultracentrifugation experiments; Dr. C. James McKnight for helpful discussions and comments on the manuscript; and C. Ogata, A. Saxena, Y. Wang, and X. Qian for assistance with data collection at the National Synchrotron Light Source, Brookhaven National Laboratory. This research was supported in part by a grant from the National Institutes of Health (DK053893).

Received: May 10, 2004

Revised: June 28, 2004

Accepted: July 6, 2004

Published: September 7, 2004

#### References

- Aggarwal, A.K. (1995). Structure and function of restriction endonucleases. *Curr. Opin. Struct. Biol.* 5, 11–19.
- Anderson, J.E. (1993). Restriction endonucleases and modification methylases. *Curr. Opin. Struct. Biol.* 3, 24–30.
- Bickle, T.A., and Kruger, D.H. (1993). Biology of DNA restriction. *Microbiol. Rev.* 57, 434–450.
- Bitinaite, J., Wah, D.A., Aggarwal, A.K., and Schildkraut, I. (1998). FokI dimerization is required for DNA cleavage. *Proc. Natl. Acad. Sci. USA* 95, 10570–10575.
- Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P.,

- Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54**, 905–921.
- CCP4 (Collaborative Computational Project, Number 4). (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763.
- Chen, Y.Q., Ghosh, S., and Ghosh, G. (1998). A novel DNA recognition mode by the NF-kappa B p65 homodimer. *Nat. Struct. Biol.* **5**, 67–73.
- Deibert, M., Grazulis, S., Sasnauskas, G., Siksnys, V., and Huber, R. (2000). Structure of the tetrameric restriction endonuclease NgoMIV in complex with cleaved DNA. *Nat. Struct. Biol.* **7**, 792–799.
- Flick, K.E., Jurica, M.S., Monnat, R.J., Jr., and Stoddard, B.L. (1998). DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-PpoI. *Nature* **394**, 96–101.
- Guo, H.-C. (2003). Type II restriction endonucleases: structures and applications. In *Recent Research Developments in Macromolecules*, S.G. Pandalai, ed. (Trivandrum, India: Research Signpost), pp. 225–245.
- Hlavaty, J.J., Benner, J.S., Hornstra, L.J., and Schildkraut, I. (2000). Identification of the metal-binding sites of restriction endonucleases by Fe<sup>2+</sup>-mediated oxidative cleavage. *Biochemistry* **39**, 3097–3105.
- Horton, N.C., and Perona, J.J. (2000). Crystallographic snapshots along a protein-induced DNA-bending pathway. *Proc. Natl. Acad. Sci. USA* **97**, 5729–5734.
- Horton, N.C., Newberry, K.J., and Perona, J.J. (1998). Metal ion-mediated substrate-assisted catalysis in type II restriction endonucleases. *Proc. Natl. Acad. Sci. USA* **95**, 13489–13494.
- Hsieh, P.C., Xiao, J., O'Loane, D., and Xu, S.Y. (2000). Cloning, expression, and purification of a thermostable nonhomodimeric restriction enzyme, BslI. *J. Bacteriol.* **182**, 949–955.
- Jin, L., Sliz, P., Chen, L., Macian, F., Rao, A., Hogan, P.G., and Harrison, S.C. (2003). An asymmetric NFAT1 dimer on a pseudopalindromic kappaB-like DNA site. *Nat. Struct. Biol.* **10**, 807–811.
- Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A* **47**, 110–119.
- Kapfer, W., Walter, J., and Trautner, T.A. (1991). Cloning, characterization and evolution of the BsuFI restriction endonuclease gene of *Bacillus subtilis* and purification of the enzyme. *Nucleic Acids Res.* **19**, 6457–6463.
- Lin, P.M., Lee, C.H., and Roberts, R.J. (1989). Cloning and characterization of the genes encoding the MspI restriction modification system. *Nucleic Acids Res.* **17**, 3001–3011.
- Miller, M.D., Cai, J., and Krause, K.L. (1999). The active site of *Serratia* endonuclease contains a conserved magnesium-water cluster. *J. Mol. Biol.* **288**, 975–987.
- Newman, M., Lunnen, K., Wilson, G., Greci, J., Schildkraut, I., and Phillips, S.E. (1998). Crystal structure of restriction endonuclease BglI bound to its interrupted DNA recognition sequence. *EMBO J.* **17**, 5466–5476.
- Nwankwo, D.O., and Wilson, G.G. (1988). Cloning and expression of the MspI restriction and modification genes. *Gene* **64**, 1–8.
- O'Loughlin, T.J., Xu, Q., Kucera, R.B., Dorner, L.F., Sweeney, S., Schildkraut, I., and Guo, H.-C. (2000). Crystallization and preliminary X-ray diffraction analysis of MspI restriction endonuclease in complex with its cognate DNA. *Acta Crystallogr. D Biol. Crystallogr.* **56**, 1652–1655.
- Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
- Pingoud, A., and Jeltsch, A. (2001). Structure and function of type II restriction endonucleases. *Nucleic Acids Res.* **29**, 3705–3727.
- Roberts, R.J., and Halford, S.E. (1993). Type II restriction endonucleases. In *Nucleases*, S.M. Linn, R.S. Lloyd, and R.J. Roberts, eds. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press), pp. 35–88.
- Roberts, R.J., Belfort, M., Bestor, T., Bhagwat, A.S., Bickle, T.A., Bitinaite, J., Blumenthal, R.M., Degtyarev, S.K.H., Dryden, D.T., Dybvig, K., et al. (2003a). A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res.* **31**, 1805–1812.
- Roberts, R.J., Vincze, T., Posfai, J., and Macelis, D. (2003b). REBASE: restriction enzymes and methyltransferases. *Nucleic Acids Res.* **31**, 418–420.
- Sasnauskas, G., Halford, S.E., and Siksnys, V. (2003). How the BfiI restriction enzyme uses one active site to cut two DNA strands. *Proc. Natl. Acad. Sci. USA* **100**, 6410–6415.
- Siksnys, V., Skirgaila, R., Sasnauskas, G., Urbanke, C., Cherny, D., Grazulis, S., and Huber, R. (1999). The Cfr10I restriction enzyme is functional as a tetramer. *J. Mol. Biol.* **291**, 1105–1118.
- Smith, H.O., and Nathans, D. (1973). Letter: a suggested nomenclature for bacterial host modification and restriction systems and their enzymes. *J. Mol. Biol.* **81**, 419–423.
- Vanamee, E.S., Santagata, S., and Aggarwal, A.K. (2001). FokI requires two specific DNA sites for cleavage. *J. Mol. Biol.* **309**, 69–78.
- Wah, D.A., Bitinaite, J., Schildkraut, I., and Aggarwal, A.K. (1998). Structure of FokI has implications for DNA cleavage. *Proc. Natl. Acad. Sci. USA* **95**, 10564–10569.
- Wilson, G.G., and Murray, N.E. (1991). Restriction and modification systems. *Annu. Rev. Genet.* **25**, 585–627.
- Xu, Q.S. (2003). Crystal structures of restriction endonuclease MspI-DNA complex and glycosylasparaginase. PhD thesis, Boston University, Boston, Massachusetts.

#### Accession Numbers

Atomic coordinates and structure-factor amplitudes have been deposited in the Protein Data Bank under accession code 1SA3.