Crystallization and preliminary X-ray diffraction analysis of *MspI* restriction endonuclease in complex with its cognate DNA

Terence J. O’Loughlin, Qian Xu, Rebecca B. Kucera, Lydia F. Dorner, Suzanne Sweeney, Ira Schildkraut and Hwai-Chen Guo
Crystallization and preliminary X-ray diffraction analysis of MspI restriction endonuclease in complex with its cognate DNA

Terence J. O’Loughlin,*† Qian Xu,a² Rebecca B. Kucera,b Lydia F. Dorner,b Suzanne Sweeney,b Ira Schildkrautb and Hwai-Chen Guoa*

*Department of Physiology and Structural Biology, Boston University School of Medicine, 715 Albany Street, Boston, MA 02118-2526, USA, and New England Biolabs, 32 Tozer Road, Beverly, MA 01915-5599, USA

a² These two authors contributed equally to this work.

Correspondence e-mail: hcguo@bu.edu

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

1. Introduction

Diverse recognition strategies and high specificity make restriction enzymes an ideal system for structural studies of protein–DNA interactions and DNA cleavage by proteins. The majority of restriction enzymes belong to type II endonucleases that form homodimers and depend only on Mg\textsuperscript{2+} for activity. They recognize DNA sequences of 4–8 base pairs with a dyad axis of symmetry, termed palindromes, and cleave within that sequence. There are currently over 3000 known restriction enzymes that recognize approximately 200 different DNA sequences (Roberts & Macelis, 2000). Of the 200 or so restriction enzymes that have been sequenced, little homology has been observed, suggesting diverse strategies for recognition of the same DNA sequence. Recognition is also highly specific, since a change of a single base pair in the cognate DNA sequence can reduce catalytic efficiency of the enzyme by as much as a factor of 10\textsuperscript{6}.

Eight structures of type II endonucleases have been reported: EcoRI (Kim et al., 1990), EcoRV (Winkler et al., 1993), PvuII (Athanasiadis et al., 1994; Cheng et al., 1994), BamHI (Newman et al., 1995), Cfr101I (Bozic et al., 1996), BglII (Newman et al., 1998), MunI (Deibert et al., 1999) and BgII (Lukacs et al., 2000). All eight of these enzymes recognize palindromic hexanucleotide sequences, although Cfr101I has a degenerated recognition sequence at the outer base pairs and BgII has an interrupted recognition sequence. Four other structures of type IIs enzymes and homing endonucleases, FokI (Wah et al., 1997), I-CreI (Heath et al., 1997; Jurica et al., 1998), PI-Scel (Duan et al., 1997) and I-DmoI (Silva et al., 1999) have also been reported (reviewed by Aggarwal & Wah, 1998). In contrast to the common type II enzymes, these type IIs and homing enzymes recognize asymmetric sequences and cleave at a short distance from that sequence. The eight type II endonuclease structures can be subdivided into three groups, correlating well with the type of cleavage pattern produced by these enzymes. The first group of five enzymes (EcoRI, BamHI, Cfr101I, MunI and BgII) bind their recognition sequences with the major groove of DNA facing the protein and cleave between the outer two bases of the recognition sequence, leaving a 5’ four-base overhang. They share a core architecture of five β-strands and two α-helices. The second group of two enzymes (EcoRV and PvuII) share a different core architecture of seven β-strands and three α-helices. This group of enzymes bind their recognition sequences with the minor groove of DNA facing the protein and cleave between the central two base pairs of the recognition sequence, leaving blunt ends. BglII falls into the third group that recognizes an interrupted DNA sequence and produces 3’ overhanging ends. Sequence recognition by BgII is mainly in the major groove, although it also contacts the minor groove. Interestingly, the core of BgII displays extensive similarities to the second group of EcoRV-like enzymes, but the dimer structure is dramatically different from other restriction-enzyme structures. Therefore, it appears that the two common core architectures may exist to properly place the active
sites by the scissile phosphates at the appropriate position of the DNA helix. Thus, cleavage pattern rather than DNA-recognition sequence might be the primary force organizing the core architecture and common elements of restriction enzymes.

With more than 3000 restriction endonucleases identified, the number of known restriction-endonuclease structures is still very small. This has posed severe limitations on our ability to rationally design new recognition specificities. In general, attempts to alter the specificity of restriction endonucleases by either site-directed mutagenesis or domain fusion have been unsuccessful. One possible approach is to mutate amino acids that make specific DNA contacts in the hope of creating new specificities. The results of these substitutions have failed to create new specificities but have only revealed acceptable modifications for the cognate DNA sequence (Alves et al., 1989; Nastri et al., 1997; Dornet et al., 1999). An alternative approach is to make a fusion protein between the zinc finger domain and the FokI cleavage domain (Smith et al., 1999). This chimeric enzyme produced some background smears of degraded DNA because of the uncontrolled nuclease activity of the endonuclease domain. Nonetheless, nature has created significant divergence within the family of restriction endonucleases. A striking divergence in DNA recognition was revealed by comparing the BglII structure with BamHI, two enzymes that cleave at closely related sites in the DNA sequences (Lukacs et al., 2000). The recognition sequences for BglII and BamHI are 5'-GGATCC (common base pairs in bold), respectively. These two enzymes bind DNA differently, leading to different protein–DNA contacts even for the common central four base pairs. In addition, the BgIII–DNA structure also reveals an active site that is different from other type II endonucleases, but is similar to an unrelated homing endonuclease I-CreI (Jurica et al., 1998). This unexpected diversity in DNA recognition, DNA structure and active site all indicate that there is still much for us to learn about the specificity and catalytic mechanisms of restriction endonucleases.

MspI, a type II restriction endonuclease from Moraxella sp. (Nwankwo & Wilson, 1988), produces a different cleavage pattern from those of known structures. Thus, MspI is likely to represent a novel structural class of endonucleases. It recognizes the palindromic tetranucleotide sequence 5'-CCGG and cleaves between the first and second nucleotides, leaving a 5' two-base overhang (Nwankwo & Wilson, 1988). A molecule of MspI has a molecular mass of 29 kDa and consists of 262 amino-acid residues. It displays no significant sequence homology to any other restriction endonuclease. Here, we report crystals of MspI in complex with a duplex DNA molecule containing the specific recognition site. The structure of the BgII–DNA complex shows that dimeric restriction endonucleases can use a conserved EcoRV-like core combined with alternative modes of dimerization to generate cleavage patterns with blunt or 5' overhanging ends. It remains to be seen whether restriction endonucleases that generate 5' overhangs of various lengths can achieve these cleavage patterns by alternative dimerization of a conserved BamHI-like core or whether they use a novel core structure. The crystal structure of the MspI–DNA complex should uncover information to address this question.

2. Purification

The MspI restriction/modification system was cloned from Moraxella sp. into Escherichia coli (Nwankwo & Wilson, 1988) and overexpressed (L. Greenough and W. E. Jack, personal communication). ER2502 cells containing the pCEW1 (MspI methylease) and the pCAD39 (MspI restriction endonuclease under ptac promoter) plasmids were grown in LB containing 100 μg ml⁻¹ ampicillin by inoculation of a 100 l fermenter with a 0.5% overnight inoculant. The cells were induced at a Klett value of 100 with 0.3 mM IPTG and were harvested after 3 h. After centrifugation, the cells were frozen as a wet cell pellet at 203 K until needed. Breakage was achieved by a single pass through a Menton–Gaulin press, cellular debris was removed by centrifugation and 25 μg ml⁻¹ PMSF was added to the crude supernatant. The enzyme was then purified by a series of low-pressure chromatographic steps consisting of phosphocellulose (Whatman P11), Affi-gel Blue (Bio-Rad), hydroxyapatite (Bio-Rad) and heparin-Sepharose (Pharmacia). The final eluted MspI enzyme was dialyzed against a storage buffer of 100 mM KCl, 10 mM HEPES pH 7.3, 1 mM DTT, 0.1 mM EDTA and 10% glycerol. Throughout the purification the MspI activity levels were monitored by incubation of the fractions containing MspI with λ DNA in NEBuffer #2 at 310 K; the final purified enzyme was determined to be free of contaminating exonucleases by incubations with ³H-labelled E. coli DNA and free of non-specific endonucleases by incubations with λ DNA. The oligodeoxyribonucleotides used to form MspI–DNA complexes were synthesized by standard methods, purified by HPLC and lyophilized for storage.

3. Crystallization

Before crystallization, MspI enzyme stored at 253 K was thawed on ice and concentrated to 20 mg ml⁻¹ using Centriprep and Centricron concentrators (Amicon). A ten-base self-complementary oligodeoxyribonucleotide 5'-CCCCCGGAGG-3' with a dyad symmetry and carrying the specific

---

**Table 1**

Data-collection statistics of native MspI–DNA co-crystal.

<table>
<thead>
<tr>
<th>Resolution (Å)</th>
<th>Rmerge</th>
<th>(I/Io)</th>
<th>No. of observations</th>
<th>No. of unique reflections</th>
<th>Fraction complete</th>
<th>Data redundancy</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.56</td>
<td>0.020</td>
<td>52.3</td>
<td>6799</td>
<td>2183</td>
<td>0.925</td>
<td>3.11</td>
</tr>
<tr>
<td>4.42</td>
<td>0.022</td>
<td>41.4</td>
<td>6721</td>
<td>2244</td>
<td>0.952</td>
<td>3.00</td>
</tr>
<tr>
<td>3.86</td>
<td>0.023</td>
<td>40.4</td>
<td>6434</td>
<td>2211</td>
<td>0.963</td>
<td>2.91</td>
</tr>
<tr>
<td>3.51</td>
<td>0.027</td>
<td>38.1</td>
<td>6464</td>
<td>2212</td>
<td>0.953</td>
<td>2.92</td>
</tr>
<tr>
<td>3.25</td>
<td>0.027</td>
<td>42.7</td>
<td>6514</td>
<td>2223</td>
<td>0.957</td>
<td>2.93</td>
</tr>
<tr>
<td>3.06</td>
<td>0.034</td>
<td>34.2</td>
<td>6467</td>
<td>2200</td>
<td>0.954</td>
<td>2.94</td>
</tr>
<tr>
<td>2.91</td>
<td>0.041</td>
<td>26.2</td>
<td>6473</td>
<td>2207</td>
<td>0.964</td>
<td>2.93</td>
</tr>
<tr>
<td>2.78</td>
<td>0.047</td>
<td>23.4</td>
<td>6569</td>
<td>2250</td>
<td>0.969</td>
<td>2.92</td>
</tr>
<tr>
<td>2.68</td>
<td>0.057</td>
<td>19.8</td>
<td>6452</td>
<td>2206</td>
<td>0.965</td>
<td>2.92</td>
</tr>
<tr>
<td>2.58</td>
<td>0.065</td>
<td>16.8</td>
<td>6632</td>
<td>2248</td>
<td>0.963</td>
<td>2.95</td>
</tr>
<tr>
<td>2.50</td>
<td>0.071</td>
<td>14.8</td>
<td>6564</td>
<td>2235</td>
<td>0.976</td>
<td>2.94</td>
</tr>
<tr>
<td>2.43</td>
<td>0.077</td>
<td>13.7</td>
<td>6664</td>
<td>2254</td>
<td>0.972</td>
<td>2.96</td>
</tr>
<tr>
<td>2.37</td>
<td>0.092</td>
<td>11.2</td>
<td>6507</td>
<td>2216</td>
<td>0.971</td>
<td>2.94</td>
</tr>
<tr>
<td>2.31</td>
<td>0.100</td>
<td>10.0</td>
<td>6695</td>
<td>2287</td>
<td>0.988</td>
<td>2.93</td>
</tr>
<tr>
<td>2.26</td>
<td>0.131</td>
<td>9.3</td>
<td>6645</td>
<td>2268</td>
<td>0.977</td>
<td>2.93</td>
</tr>
<tr>
<td>2.21</td>
<td>0.145</td>
<td>7.7</td>
<td>6537</td>
<td>2265</td>
<td>0.994</td>
<td>2.87</td>
</tr>
<tr>
<td>2.16</td>
<td>0.136</td>
<td>7.1</td>
<td>6677</td>
<td>2269</td>
<td>0.981</td>
<td>2.94</td>
</tr>
<tr>
<td>2.12</td>
<td>0.156</td>
<td>6.2</td>
<td>6627</td>
<td>2285</td>
<td>0.990</td>
<td>2.90</td>
</tr>
<tr>
<td>2.09</td>
<td>0.193</td>
<td>5.6</td>
<td>6460</td>
<td>2275</td>
<td>0.988</td>
<td>2.92</td>
</tr>
<tr>
<td>2.05</td>
<td>0.229</td>
<td>4.8</td>
<td>6501</td>
<td>2243</td>
<td>0.990</td>
<td>2.90</td>
</tr>
<tr>
<td>100.0–2.05</td>
<td>0.040</td>
<td>25.7</td>
<td>131582</td>
<td>44781</td>
<td>0.969</td>
<td>2.94</td>
</tr>
</tbody>
</table>

---
MspI recognition sequence (in bold) was used for crystallization. The length of oligodeoxynucleotide is critical for crystallization, since no crystal was obtained when a 12-mer oligodeoxynucleotide 5' -TCCCCGGGGGA-3' was used in the MspI–DNA mixture. The lyophilized oligodeoxynucleotide was resuspended in water and reannealed by heating and slow cooling to room temperature. MspI–DNA mixtures were prepared to give final protein concentration of 6 mg ml⁻¹ and a twofold molar ratio of DNA with 5 mM EDTA. Crystalization of MspI–DNA complexes was performed using the hanging-drop vapor-diffusion technique with equal volumes of MspI–DNA mixture and crystallization solution (100 mM MES pH 6.5, 100 mM ammonium sulfate, 6 mM calcium chloride, 14% PEG 8000 and 12% glycerol). Crystalization dishes (Linbro plates) were kept at 293 K in the dark to avoid any possible photosensitiveness. Seeding techniques were routinely used to improve and ensure high-quality singular crystals. Thin-plate post-seeded crystals grew within a week to dimensions of about 400 × 100 × 30 μm (Fig. 1). These crystals were confirmed as enzyme–DNA co-crystals by staining with methylene blue and methyl violet for DNA and protein, respectively (data not shown).

4. X-ray analysis

Before data collection, crystals of the MspI–DNA complex were equilibrated with the crystallization buffer supplemented with 20% (v/v) glycerol as cryoprotectant. The crystals were then mounted in a thin film of crystallization buffer plus cryoprotectant, supported by a loop made of dental floss (Teng, 1990) and flash-cooled, either directly in the cold nitrogen stream with an MSC cryosystem or by plunging into liquid nitrogen for storage and transfer. X-ray diffraction data were collected at 100 K at the National Synchrotron Light Source (Brookhaven National Laboratory) on beamline X12C, which is equipped with a Brandeis 2×2 CCD detector. A native data set was collected at λ = 0.9207 Å from 1° oscillation photographs, sweeping a 140° wedge of reciprocal space at a crystal-to-detector distance of 85 mm. The exposure time per image was 45 s. Data reduction was performed and the space group and unit-cell parameters were determined using the HKL suite (Otwinski & Minor, 1997). The space group is monoclinic P2₁, with unit-cell parameters a = 50.2 Å, b = 131.6 Å, c = 59.3 Å, β = 109.7°. Data quality and completeness as a function of resolution are shown in Table 1. This native data set is at 2.05 Å resolution, with strong signal over noise (overall Rmerge = 25.7%), 96.9% completeness and a redundancy of 2.94. V_M calculations (Matthews, 1968) suggest that there is most likely to be one MspI dimer bound to the DNA duplex in the asymmetric unit, although the presence of one and a half MspI-dimer–DNA duplexes per asymmetric unit would also be within the acceptable range of solvent content. With an MspI dimer bound to the DNA duplex, the V_M is 2.86 Å³ Da⁻¹, corresponding to a solvent content of about 57%. The selfrotation function also showed non-crystallographic twofold peaks in the ac plane (Fig. 2). Various phasing approaches, including MIR and MAD methods with iodination or bromination of the DNA as well as conventional heavy-metal soaks, are being employed.

We thank Didem Vardar and Benjamin Frank for critical reading of the manuscript. Diffraction data for this study were collected at Brookhaven National Laboratory at beamline X12C in the National Synchrotron Light Source. This facility is supported by the United States Department of Energy under prime contract DE-AC02-98CH10886, by the National Science Foundation and by National Institutes of Health Grant 1P41 RR12408-01A1.

References


Figure 1
Co-crystals of restriction endonuclease MspI–cognate DNA. The dimensions of the crystal shown are about 500 × 150 × 20 μm.

Figure 2
Representation of the Patterson self-rotation function of the set of structure factors in space group P2₁ for the MspI–DNA complex (using a spherical integration radius of 25 Å and diffraction data between 10 and 2.5 Å resolution) was produced using GLRF (Tong & Rossmann, 1990). Contours start at 3 standard deviations (sd) with steps of 1 sd. This diagram indicates the presence of a strong twofold symmetry along the crystallographic b axis (φ = 0, ϑ = 0, α = 180°), as expected for the crystallographic 2₁ screw axis. Additionally, there are non-crystallographic twofold peaks occurring in the ac plane (φ = 90°, α = 180°), running nearly parallel to the a and c* axes (φ = 179° and 89°, respectively).


