

Acta Crystallographica Section D

**Biological
Crystallography**

ISSN 0907-4449

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

Copyright © International Union of Crystallography

Author(s) of this paper may load this reprint on their own web site provided that this cover page is retained. Republication of this article or its storage in electronic databases or the like is not permitted without prior permission in writing from the IUCr.

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

Terence J. O'Loughlin,^{a†} Qian Xu,^{a†} Rebecca B. Kucera,^b Lydia F. Dorner,^b Suzanne Sweeney,^b Ira Schildkraut^b and Hwai-Chen Guo^{a*}

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

† These two authors contributed equally to this work.

Correspondence e-mail: hcguo@bu.edu

The *MspI* restriction endonuclease is a type II restriction enzyme. Unlike all other restriction enzymes with known structures, *MspI* recognizes the palindromic tetranucleotide sequence 5'-C/CGG and cleaves it as indicated by the '/' to produce DNA products with 5' two-base overhangs. Owing to the nature of its cleavage pattern, it is likely that *MspI* would represent a new structural class of restriction endonucleases. Crystals of the dimeric *MspI* restriction enzyme bound to a duplex DNA molecule containing the specific recognition sequence have been obtained by vapor-diffusion techniques in the presence of polyethylene glycol as precipitant. The crystals belong to the monoclinic space group $P2_1$, with unit-cell parameters $a = 50.2$, $b = 131.6$, $c = 59.3$ Å, $\beta = 109.7^\circ$. The crystals contain one dimeric complex in the asymmetric unit. A complete native data set has been collected to a resolution of 2.05 Å by cryo-crystallographic methods, with an R_{merge} of 4.0%.

Received 19 June 2000

Accepted 13 September 2000

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^{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 10^6 .

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), *Cfr10I* (Bozic *et al.*, 1996), *BglI* (Newman *et al.*, 1998), *MunI* (Deibert *et al.*, 1999) and *BglII* (Lukacs *et al.*, 2000). All eight of these enzymes recognize palindromic hexanucleotide sequences, although *Cfr10I* has a degenerated recognition sequence at the outer base pairs and *BglI* has an interrupted recognition sequence. Four other structures of type II enzymes and homing endonucleases, *FokI* (Wah *et al.*, 1997),

I-CreI (Heath *et al.*, 1997; Jurica *et al.*, 1998), *PI-SceI* (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*, *Cfr10I*, *MunI* and *BglII*) 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. *BglI* falls into the third group that recognizes an interrupted DNA sequence and produces 3' overhanging ends. Sequence recognition by *BglI* is mainly in the major groove, although it also contacts the minor groove. Interestingly, the core of *BglI* 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

Table 1
Data-collection statistics of native *MspI*-DNA co-crystal.

Resolution (Å)	R_{merge}	$\langle I \rangle / \sigma$	No. of observations	No. of unique reflections	Fraction complete	Data redundancy
5.56	0.020	52.3	6799	2183	0.925	3.11
4.42	0.022	41.4	6721	2244	0.952	3.00
3.86	0.023	40.4	6434	2211	0.963	2.91
3.51	0.027	38.1	6464	2212	0.953	2.92
3.25	0.027	42.7	6514	2223	0.957	2.93
3.06	0.034	34.2	6467	2200	0.954	2.94
2.91	0.041	26.2	6473	2207	0.964	2.93
2.78	0.047	23.4	6569	2250	0.969	2.92
2.68	0.057	19.8	6452	2206	0.965	2.92
2.58	0.065	16.8	6632	2248	0.962	2.95
2.50	0.071	14.8	6564	2235	0.976	2.94
2.43	0.077	13.7	6664	2254	0.972	2.96
2.37	0.092	11.2	6507	2216	0.971	2.94
2.31	0.100	10.0	6695	2287	0.988	2.93
2.26	0.131	9.3	6645	2268	0.977	2.93
2.21	0.145	7.7	6537	2265	0.994	2.87
2.16	0.136	7.1	6677	2269	0.981	2.94
2.12	0.156	6.2	6627	2285	0.990	2.90
2.09	0.193	5.6	6640	2275	0.988	2.92
2.05	0.229	4.8	6501	2243	0.990	2.90
100.0-2.05	0.040	25.7	131582	44781	0.969	2.94

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; Dorner *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'-AGATCT and

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 *BglII*-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 *BglII*-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 3' 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* methylase) and the pCAD39 (*MspI* restriction endonuclease under *ptac* promoter) plasmids were grown in LB containing 100 $\mu\text{g ml}^{-1}$ ampicillin by inoculation of a 100 l fermentation 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 $\mu\text{g ml}^{-1}$ 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 oligodeoxynucleotides 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 Centrprep and Centricon concentrators (Amicon). A ten-base self-complementary oligodeoxynucleotide 5'-CCCCCGGGG-3' with a dyad symmetry and carrying the specific

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'-TCCCCCGGGGA-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. Crystallization 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). Crystallization dishes (Linbro plates) were kept at 293 K in the dark to avoid any possible photosensitivity. 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).

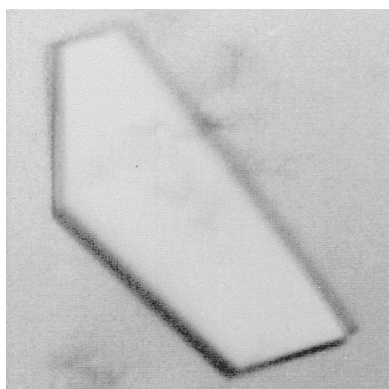


Figure 1
Co-crystals of restriction endonuclease *MspI*-cognate DNA. The dimensions of the crystal shown are about 500 × 150 × 20 μm.

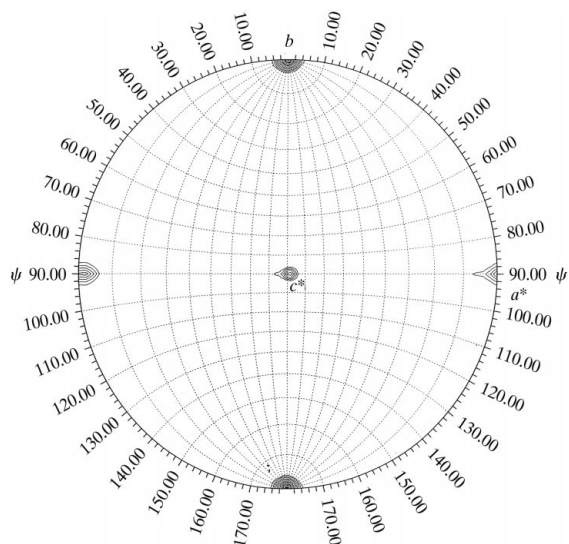


Figure 2
Representation of the Patterson self-rotation function of the set of structure factors in space group $P2_1$ for the *MspI*-DNA complex (using a spherical integration radius of 25 Å and diffraction data between 10 and 2.5 Å resolution) was produced using *GLRF* (Tong & Rossmann, 1990). Contours 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 ($\psi = 0, \psi = 180^\circ$), as expected for the crystallographic 2_1 screw axis. Additionally, there are non-crystallographic twofold peaks occurring in the ac plane ($\psi = 90, \psi = 270^\circ$, running nearly parallel to the a and c^* axes ($\psi = 179$ and 89° , respectively)).

(overall $(I)/\sigma = 25.7$), 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-duplex complexes 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 self-rotation 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.

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 $\lambda = 0.9207$ 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 (Otwinowski & Minor, 1997). The space group is monoclinic $P2_1$, with unit-cell parameters $a = 50.2, b = 131.6, c = 59.3$ Å, $\beta = 109.7^\circ$. 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

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

- Aggarwal, A. K. & Wah, D. A. (1998). *Curr. Opin. Struct. Biol.* **8**, 19–25.
- Alves, J., Ruter, T., Geiger, R., Fliess, A., Maass, G. & Pingoud, A. (1989). *Biochemistry*, **28**, 2678–2684.
- Athanasiadis, A., Vlassi, M., Kotsifaki, D., Tucker, P. A., Wilson, K. S. & Kokkinidis, M. (1994). *Nature Struct. Biol.* **1**, 469–475.
- Bozic, D., Grazulis, S., Siksnys, V. & Huber, R. (1996). *J. Mol. Biol.* **255**, 176–186.
- Cheng, X., Balendiran, K., Schildkraut, I. & Anderson, J. E. (1994). *EMBO J.* **13**, 3927–3935.
- Deibert, M., Grazulis, S., Janulaitis, A., Siksnys, V. & Huber, R. (1999). *EMBO J.* **18**, 5805–5816.
- Dorner, L. F., Bitinaite, J., Whitaker, R. D. & Schildkraut, I. (1999). *J. Mol. Biol.* **285**, 1515–1523.
- Duan, X., Gimble, F. S. & Quijcho, F. A. (1997). *Cell*, **89**, 555–564.
- Heath, P. J., Stephens, K. M., Monnat, R. J. Jr & Stoddard, B. L. (1997). *Nature Struct. Biol.* **4**, 468–476.
- Jurica, M. S., Monnat, R. J. Jr & Stoddard, B. L. (1998). *Mol. Cell*, **2**, 469–476.
- Kim, Y. C., Grable, J. C., Love, R., Greene, P. J. & Rosenberg, J. M. (1990). *Science*, **249**, 1307–1309.
- Lukacs, C. M., Kucera, R., Schildkraut, I. & Aggarwal, A. K. (2000). *Nature Struct. Biol.* **7**, 134–140.

- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Nastri, H. G., Evans, P. D., Walker, I. H. & Riggs, P. D. (1997). *J. Biol. Chem.* **272**, 25761–25767.
- Newman, M., Lunnen, K., Wilson, G., Greci, J., Schildkraut, I. & Phillips, S. E. V. (1998). *EMBO J.* **17**, 5466–5476.
- Newman, M., Strzelecka, T., Dorner, L. F., Schildkraut, I. & Aggarwal, A. K. (1995). *Science*, **269**, 656–663.
- Nwankwo, D. O. & Wilson, G. G. (1988). *Gene*, **64**, 1–8.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Roberts, R. J. & Macelis, D. (2000). *Nucleic Acids Res.* **28**, 306–307.
- Silva, G. H., Dalgaard, J. Z., Belfort, M. & Van Roey, P. (1999). *J. Mol. Biol.* **286**, 1123–1136.
- Smith, J., Berg, J. M. & Chandrasegaran, S. (1999). *Nucleic Acids Res.* **27**, 674–681.
- Teng, T. Y. (1990). *J. Appl. Cryst.* **23**, 387–391.
- Tong, L. & Rossmann, M. G. (1990). *Acta Cryst. A* **46**, 783–792.
- Wah, D. A., Hirsch, J. A., Dorner, L. F., Schildkraut, I. & Aggarwal, A. K. (1997). *Nature (London)*, **388**, 97–100.
- Winkler, F. K., Banner, D. W., Oefner, C., Tsernoglou, D., Brown, R. S., Heathman, S. P., Bryan, R. K., Martin, P. D., Petratos, K. & Wilson, K. S. (1993). *EMBO J.* **12**, 1781–1795.