Acta Crystallographica Section F
**Structural Biology and Crystallization Communications**
ISSN 1744-3091

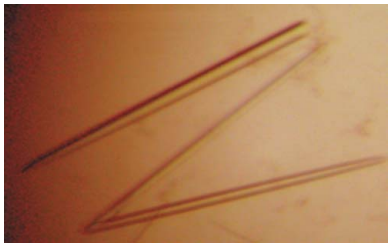
# Purification, crystallization and preliminary X-ray analysis of the *Bse*CI DNA methyltransferase from *Bacillus stearothermophilus* in complex with its cognate DNA

**Evangelia G. Kapetaniou,[a] Dina Kotsifaki,[a] Mary Providaki,[a] Maria Rina,[a] Vassilis Bouriotis[a,b] and Michael Kokkinidis[a,b]***

[a]Institute of Molecular Biology and Biotechnology (IMBB), PO Box 1527, GR-71110, Heraklion, Crete, Greece, and [b]University of Crete, Department of Biology, PO Box 2208, GR-71110, Heraklion, Crete, Greece

Correspondence e-mail: kokkinid@imbb.forth.gr

Received 5 October 2006
Accepted 28 November 2006





The DNA methyltransferase M.*Bse*CI from *Bacillus stearothermophilus* (EC 2.1.1.72), a 579-amino-acid enzyme, methylates the N6 atom of the 3′ adenine in the sequence 5′-ATCGAT-3′. M.*Bse*CI was crystallized in complex with its cognate DNA. The crystals were found to belong to the hexagonal space group *P*6, with unit-cell parameters $a = b = 87.0$, $c = 156.1$ Å, $\beta = 120.0^\circ$ and one molecule in the asymmetric unit. Two complete data sets were collected at wavelengths of 1.1 and 2.0 Å to 2.5 and 2.8 Å resolution, respectively, using synchrotron radiation at 100 K.

## 1. Introduction

DNA methyltransferases (DNA MTases) occur ubiquitously in all groups of prokaryotic and eukaryotic organisms. In higher eukaryotes they are implicated in the regulation of gene expression, while the prokaryotic enzymes are involved in the phage DNA-recognition mechanism (Hattman, 2005). Bacterial DNA MTases belong to restriction-modification (RM) systems, which comprise an endonuclease (ENase) and an MTase activity. There are three types of RM systems in bacteria, which are classified according to their subunit composition, cofactor requirements and mode of action (Smith & Nathans, 1973). Type II systems are the simplest and the most abundant, with the MTase and ENase activities exerted by two distinct enzymes encoded by gene pairs (Bujnicki, 2001).

DNA MTases from type II systems recognize sequence-specific duplex sites in DNA and catalyze irreversible methyl-group transfer from *S*-adenosyl-L-methionine (AdoMet), a methyl-group donor, to their target DNA base to form methylated DNA and *S*-adenosyl-L-homocysteine (AdoHcy; Adams, 1990). They are classified into three groups according to the targeted DNA base and the characteristics of the catalytic reaction. They either transfer a methyl group to the C5 atom of cytosine (C-MTases), forming $C^5$-methylcytosine ($m^5C$), or to the exocyclic amino group of cytosine or adenine (N-MTases), forming $N^4$-methylcytosine ($m^4C$) or $N^6$-methyladenine ($m^6A$), respectively (McClelland *et al.*, 1994). Comparative structural analysis of AdoMet-dependent MTases revealed a high degree of structural similarity, which is not reflected at the level of amino-acid sequence conservation (Egloff *et al.*, 2002). All MTases are two-domain proteins, with the DNA-binding cleft being located at the domain interface. The N-terminal domain forms the binding site for AdoMet and the catalytic centre of the enzyme and consists of a set of ten motifs, which differ between C-MTases and N-MTases (Malone *et al.*, 1995). The N-terminal domains of all MTase structures share a structurally conserved core, referred to as an AdoMet-dependent MTase fold, which consist of a seven-stranded $\beta$-sheet. The C-terminal domain is responsible for specific recognition and binding of the target DNA sequences and exhibits a considerable degree of structural and sequential variation. In contrast to C-MTases (Kumar *et al.*, 1994), N-MTases are a more heterogeneous class of enzymes and their catalytic mechanisms have been less well studied (Labahn *et al.*, 1994). Determination of the crystal structure of N6-M.*Taq*I in complex with DNA showed that the target adenine is rotated out of the DNA helix. This 'base-flipping' mechanism has also

been observed in the structures of the C-MTases *Hha*I (Klimasauskas *et al.*, 1994) and *Hae*II (Reinisch *et al.*, 1995). In addition, the M.*Taq*I structure revealed a previously unrecognized stabilization of the extrahelical target nucleotide *via* a local compression of DNA induced by protein–phosphate contacts (Goedecke *et al.*, 2001).

M.*Bse*CI DNA MTase from *Bacillus stearothermophilus* (M.*Bse*CI; MW = 66 700 Da) is a type II $m^6A$ MTase which methylates N6 of the 3′ adenine in the sequence 5′-ATCGAT-3′ with an optimum temperature for activity of 323–328 K and an optimum pH of 7.4 (Rina & Bouriotis, 1993). Here, we report its high-level expression, purification, cocrystallization with its cognate DNA and preliminary crystallographic analysis. Structure determination will pursued using experimental phasing methods. Since our knowledge of the catalytic mechanism of this class of enzymes is still limited, the structure will provide new insights into the structural basis of $N^6$-adenine methylation.

## 2. Materials and methods

### 2.1. Expression and purification

The *Bse*CI M8 gene encoding M.*Bse*CI (accession code P43423) was cloned as previously described (Rina & Bouriotis, 1993), inserted into a pQE60 vector (Qiagen) containing a C-terminal 6×His tag and transformed into *Escherichia coli* M15 cells. A sufficient amount of soluble protein for structural studies was obtained after expression using the following conditions. Cells were grown in 2 l LB medium containing 100 µg ml$^{-1}$ ampicillin and 50 µg ml$^{-1}$ kanamycin at 310 K until $OD_{600}$ reached 0.8. The culture was induced with 1 m*M* isopropyl *β*-D-thiogalactopyranoside (IPTG) for 4 h at 310 K and harvested by centrifugation at 5000*g* for 15 min at 277 K. The precipitated cells were resuspended in 50 ml lysis buffer containing 50 m*M* $Na_2HPO_4$ pH 8, 300 m*M* NaCl, 5 m*M* imidazole, 1 m*M* PMSF, 20 µg ml$^{-1}$ leupeptin and 150 µg ml$^{-1}$ benzamidine and homogenized by sonication. The precipitate was removed by centrifugation at 20 000*g* for 1 h at 277 K. Purification was effected *via* the His tag by affinity chromatography at 277 K on a 10 ml Ni–NTA chelating column (Qiagen) pre-equilibrated in lysis buffer. The column was washed with ten column volumes of 10 m*M* imidazole and five column volumes of 20 m*M* imidazole, followed by a gradient from 30 to 300 m*M* imidazole. M.*Bse*CI started to elute at 30 m*M* imidazole. Fractions containing more than 80% homogeneous M.*Bse*CI, as shown by SDS–PAGE (MW = 66 700 Da), were pooled and dialyzed extensively against 20 m*M* Tris–HCl pH 8.0 containing 200 m*M* NaCl. The protein solution was concentrated using Amicon Centriprep (YM-30) filters. Sephacryl S-200 size-exclusion chromatography was performed with 20 m*M* Tris–HCl pH 8 containing 200 m*M* NaCl (XK26/60 column, Amersham Pharmacia Biotech). M.*Bse*CI elutes after 170–200 ml of this buffer. Typical yields are 10 mg of homogeneous M.*Bse*CI from approximately 5.5 g cell paste. Prior to crystallization, M.*Bse*CI was concentrated to 10 mg ml$^{-1}$ in the gel-filtration buffer using Amicon Centriprep (YM-30) filters.

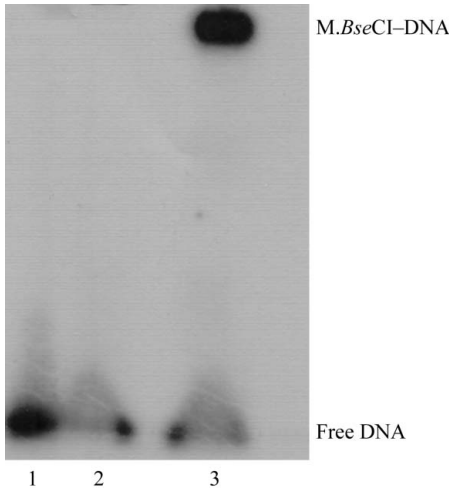



**Figure 1**
Polyacrylamide-gel mobility-shift assay of the M.*Bse*CI–DNA complex: lane 1, free dsDNA; lane 3, M.*Bse*CI–DNA complex.

### 2.2. DNA preparation and gel-shift assay

Two lyophilized single-stranded ten-base oligodeoxynucleotides, 5′-CG**ATCGAT**GC-3′ and 5′-GC**ATCGAT**CG-3′ (METABION, Germany), carrying the specific M.*Bse*CI recognition sequence (in bold) were annealed in 0.3 *M* NaCl in equal concentrations by heating and slow cooling to room temperature to produce DNA duplex, which was used for crystallization. M.*Bse*CI–DNA mixtures (in 20 m*M* Tris buffer pH 8.0 with 200 m*M* NaCl, 10 m*M* $MgCl_2$ and 3 m*M* AdoHcy) were prepared to give final concentrations of 150 µ*M* protein and 200 µ*M* duplex DNA. After incubation at 277 K overnight, this protein mixture was used for crystallization trials.

The protein–DNA complex formation was controlled by gel electrophoresis in TBE buffer (25 m*M* Tris–HCl pH 8.0, 20 m*M* boric acid, 0.25 m*M* EDTA) on nondenaturating 15% polyacrylamide gels. M.*Bse*CI was mixed with $^{32}$P-labelled DNA heteroduplex in a stoichiometry of 1.3:1 and a clearly retarded band was detected on X-ray film after electrophoresis (Fig. 1).

### 2.3. Crystallization

Initial crystallization conditions for M.*Bse*CI–DNA complexes were determined using the hanging-drop vapour-diffusion method with the Natrix Screen (Hampton Research) and Linbro 24-well cell-culture plates. The drops consisted of 2.5 µl protein mixture and 2.5 µl reservoir solution. The drops were equilibrated against 1000 µl reservoir solution at 289 K. These conditions were refined and crystals of reasonable size for crystallographic studies were obtained with 28–30%(*v*/*v*) PEG 400, 50 m*M* Tris–HCl pH 8.5, 10 m*M* $MgCl_2$ and 0.1 *M* KCl (Fig. 2).

### 2.4. Data collection and processing

X-ray diffraction data were collected using synchrotron radiation. A native data set was collected from a single crystal at the EMBL X12 beamline at the DORIS storage ring, DESY, Hamburg. Crystals were

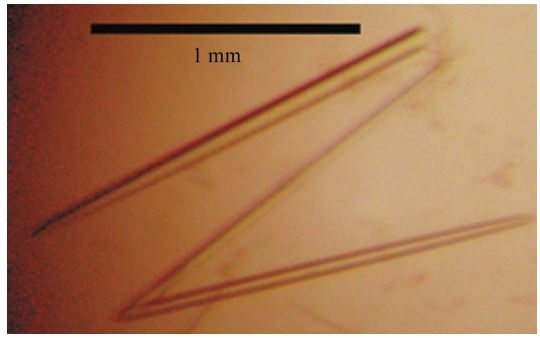



**Figure 2**
Crystals of the M.*Bse*CI–DNA complex.

**Table 1**
Data-collection and processing statistics.

Values in parentheses are for the highest resolution shell.

| Data set | 1.1 Å | 2.0 Å |
|---|---|---|
| Space group | *P*6 | *P*6 |
| Unit-cell parameters (Å, °) | $a = b = 87.03$, $c = 156.06$, $\gamma = 120$ | $a = b = 87.03$, $c = 156.06$, $\gamma = 120$ |
| Matthews coefficient (Å$^3$ Da$^{-1}$) | 2.5 | 2.5 |
| Resolution (Å) | 99–2.5 (2.64–2.5) | 99–2.8 (2.95–2.8) |
| Total observations | 169542 (24683) | 327115 (46638) |
| Unique reflections | 23041 (3358) | 16319 (2343) |
| Data completeness (%) | 99.5 (99.2) | 98.7 (97.7) |
| $R_{\rm sym}$† (%) | 5.9 (48.9) | 8.5 (72.5) |
| Average $I/\sigma(I)$ | 22.8 (3.9) | 34.4 (4.2) |
| Mosaicity (°) | 0.5 | 1.0 |

† $R_{\rm sym} = \sum_{\mathbf{h}} \sum_l |I_{\mathbf{h}l} - \langle I_{\mathbf{h}} \rangle| / \sum_{\mathbf{h}} \sum_l \langle I_{\mathbf{h}} \rangle$, where $I_l$ is the $l$th observation of reflection **h** and $\langle I_{\mathbf{h}} \rangle$ is the weighted average intensity for all observations $l$ of reflection **h**.

flash-cooled to 100 K in a nitrogen-gas cold stream using an Oxford Cryosystem device. 240 images with 1° rotation each were collected to a resolution of 2.5 Å at a wavelength of 1.1 Å. Additionally, a complete data set was collected from a single crystal at a wavelength of 2.0 Å, aiming to maximize the inherent anomalous signal from the 12 S and the 20 P atoms present in the protein–DNA complex. In both cases, diffraction data were recorded on a MAR CCD detector with a diameter of 225 mm. Data were processed and scaled with *MOSFLM* (Leslie *et al.*, 1986) and *SCALA* (Evans, 1997) from the *CCP*4 program suite (Collaborative Computational Project, Number 4, 1994). The Matthews coefficient was determined using the program *MATTHEWS_COEFF* from the *CCP*4 program suite (Collaborative Computational Project, Number 4, 1994). The program *SHELXC* (Usón & Sheldrick, 1999) was used for detection of the anomalous signal.

## 3. Results and discussion

M.*Bse*CI was successfully crystallized in complex with its cognate DNA and a native data set was collected to 2.5 Å resolution using synchrotron radiation. From the diffraction data processing, the crystals were found to belong to the hexagonal space group *P*6, as systematic absences were not observed, with unit-cell parameters $a = 87.03$, $b = 87.03$, $c = 156.06$ Å, $\gamma = 120°$. Complete data-collection and processing statistics are given in Table 1. The merged data set was 99.2% complete to 2.5 Å resolution. Assuming the presence of one molecule in the asymmetric unit, the Matthews coefficient (Matthews, 1968) is 2.5 Å$^3$ Da$^{-1}$ and the solvent content is 41.4%. A highly redundant 2.8 Å data set was collected at a wavelength of 2 Å and the data-collection and processing statistics are given in Table 1. Although restricted to a single rotation axis, 600° of data were collected, providing an overall redundancy of 20. The anomalous signal-to-noise ratio, as calculated by *SHELXC* (Usón & Sheldrick, 1999), is 1.9 in the low-resolution shell (99–8 Å) and 0.7 at 2.8 Å, reflecting the small anomalous signal. Attempts to determine the structure are currently under way. Additionally, an SeMet derivative has been crystallized and MAD experiments will be carried out.

We thank the European Molecular Biology Laboratory, Hamburg Outstation and the European Union for support through the the EU-I3 access grant from the EU Research Infrastructure Action under the FP6 'Structuring the European Research Area Programme', contract No. RII3/CT/2004/5060008. Expert assistance by Dr Manfred Weiss is gratefully acknowledged. We thank Anastassia Gazi and Chrysa Meramveliotaki for assistance with data collection.

## References

Adams, R. L. P. (1990). *Biochem. J.* **265**, 309–320.
Bujnicki, J. M. (2001). *Acta Biochim. Pol.* **48**, 935–967.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Egloff, M.-P., Benarroch, D., Selisko, B., Romette, J.-L. & Canard, B. (2002). *EMBO J.* **21**, 2757–2768.
Evans, P. R. (1997). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **33**, 22–24.
Goedecke, K., Pignot, M., Goody, R. G., Scheidig, A. J. & Weinhold, E. (2001). *Nature Struct. Biol.* **8**, 121–125.
Hattman, S. (2005). *Biochemistry Mosc.* **70**, 550–558.
Klimasauskas, S., Kumar, S., Roberts, R. J. & Cheng, X. (1994). *Cell*, **76**, 357–369.
Kumar, S., Cheng, X., Klimasauskas, S., Mi, S., Posfai, J., Roberts, R. J. & Wilson, G. G. (1994). *Nucleic Acids Res.* **22**, 1–10.
Labahn, J., Granzin, J., Schluckebier, G., Robinson, D. P., Jack, W. E., Schildkraut, I. & Saenger, W. (1994). *Proc. Natl Acad. Sci. USA*, **91**, 10957–10961.
Leslie, A. G. W., Brick, P. & Wonacott, A. J. (1986). *CCP4 Newsl.* **18**, 33–39.
McClelland, M., Nelson, M. & Raschke, E. (1994). *Nucleic Acids Res.* **22**, 3640–3659.
Malone, T., Blumenthal, R. M. & Cheng, X. (1995). *J. Mol. Biol.* **253**, 618–632.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Reinisch, K. M., Chen, L., Verdine, G. L. & Lipscomb, W. N. (1995). *Cell*, **82**, 143–153.
Rina, M. & Bouriotis, V. (1993). *Gene*, **133**, 91–94
Smith, H. O. & Nathans, D. (1973). *J. Mol. Biol.* **81**, 419–423.
Usón, I. & Sheldrick, G. M. (1999). *Curr. Opin. Struct. Biol.* **9**, 643–648.