Cloning and structure of the *BepI* modification methylase

Dagmar Kupper*, Jian-Guang Zhou⁺, Pal Venetianer and Antal Kiss

Institute of Biochemistry, Biological Research Center, Hungarian Academy of Sciences, PO Box 521, 6701 Szeged, Hungary

Received October 11, 1988; Revised and Accepted January 12, 1989

EMBL accession no. X13555

ABSTRACT

The gene coding for a CGCG specific DNA methylase has been cloned in *E. coli* from *Brevibacterium epidermidis*. The enzyme, named *BepI* methylase, is probably the cognate methylase of the *FnuDII* isoschizomer *BepI* endonuclease isolated from this strain. The expression of *BepI* methylase in *E. coli* is dependent on the orientation of the cloned fragment suggesting that the gene is transcribed from a promoter on the plasmid vector. No *BepI* endonuclease could be detected in the clones producing *BepI* methylase. The nucleotide sequence of the *BepI* methylase gene has been determined, it predicts a protein of 403 amino acids (M_r : 45,447). Analysis of the amino acid sequence deduced from the nucleotide sequence revealed similarities between the *BepI* methylase and other cytosine methylases. *M.BepI* methylates the external cytosine in its recognition sequence.

INTRODUCTION

With the aim to analyze the molecular mechanism of specific interaction between the DNA and sequence specific DNA methylases, genes of several modification methylases were cloned and sequenced (1-22). Comparison of the derived amino acid sequences showed similarities between cytosine methylases suggesting that these enzymes have common evolutionary origin and similar structure (8,9,12,14,15,16,17,18,21,22). Less extensive but significant amino acid sequence similarities were found between adenine methylases (11,19,20,23,25). Only very limited sequence similarity was shown to exist between cytosine methylases and adenine methylases (25) and no significant similarity was found between modification methylases and their cognate endonucleases (1,2,6,12,13,15,21,24).

We initiated a systematic study of a family of DNA methylases which recognize GGCC or related sequences. It is hoped that

comparison of enzymes with related sequence specificities may help elucidate the structural basis of sequence specificity. Earlier we found very strong similarity between the protein sequences of two GGCC specific methylases (BspRI and BsuRI) and less pronounced similarity between these enzymes and another methylase (SPR, 12) which, in addition to GGCC, also recognizes CCGG and CC^A/_TGG.

Here we report the cloning and sequencing of the gene coding for a CGCG specific DNA methylase. This methylase was cloned from a strain of Brevibacterium epidermidis which produces the FnuDII isoschizomer BepI endonuclease (26). The methylase is probably the cognate methylase of the BepI restriction endonuclease.

MATERIALS AND METHODS

Strains and media

The Brevibacterium epidermidis strain producing BepI endonuclease and methylase was isolated in this laboratory (26). The BepI methylase gene was cloned in E. coli ER1398 (27). M13 phage was grown on JM107MA2 (mcrB⁻, 28). B. epidermidis was grown at 30 °C and E. coli at 37 °C in LB medium. (29)

Enzymes and chemicals

Restriction endonucleases were either prepared in this institute or were purchased from New England Biolabs or from BRL. DNA polymerase large fragment was from New England Biolabs and ³⁵S-α-dATP from Amersham. All other chemicals were reagent grade commercial products.

Cloning methods

Preparation of plasmids, transformation of E. coli, restriction mapping, agarose gel electrophoresis and subcloning of DNA fragments were done by standard procedures (29).

Cloning of the BepI methylase gene

Total cellular DNA of Brevibacterium epidermidis was completely digested with EcoRI or partially digested with MboI then it was ligated to EcoRI- or BamHI-cleaved and dephosphorylated pBR322 (30). The ligated DNA was transformed in E. coli ER1398. Amp^R transformants (approximately 10,000 recombinants) were grown to saturation in 50 ml LB medium containing 100 µg/ml ampicillin.

Plasmid DNA isolated from these cultures was digested to completion with BepI endonuclease. The digested DNA was used to transform E. coli ER1398. Amp^R transformants were selected.

Determination of the nucleotide sequence

Specific fragments were cloned in mp18 and mp19 phage vectors (31,32) and were sequenced by the chain termination method (33,32) using ³⁵S- α -dATP (34, Amersham) and a 17mer oligonucleotide as universal sequencing primer (12). Both DNA strands were sequenced. The sequence has been deposited with the EMBL Data Library (accession number: X13555).

Detection of 5-methyl-cytosine

The methylation status of cytosines within the CGCG target sequence was determined by the chemical sequencing method (35). pBR322 and pBep1 plasmid DNAs were cleaved with SalI, dephosphorylated with bacterial alkaline phosphatase and labeled with ³²P using polynucleotide kinase. After digestion with EcoRI, the 3.7 kb SalI-EcoRI fragments were purified from gel, subjected to C + T and C specific chemical sequencing reactions (36) and run in a 6% polyacrylamide gel containing 8M urea (36).

RESULTS AND DISCUSSION

Cloning of the BepI methylase gene

The BepI methylase gene was cloned in E. coli ER1398 from two gene libraries as described in Materials and Methods. Plasmid DNA isolated from several clones showed complete or partial resistance to BepI endonuclease indicating that they contained the functional BepI methylase gene (pBR322 has 23 BepI sites). One such clone (pBep1) from the EcoRI library and one (pBep2) from the MboI library was chosen for further analysis. pBep1 was completely resistant to BepI endonuclease whereas pBep2 (and several other clones isolated from the MboI library) showed only partial resistance (Fig. 1.).

To test the specificity of the cloned methylase, pBep1 and pBep2 were digested with restriction enzymes whose target sequence is related to CGCG. pBep1 and pBep2 were completely sensitive to BspRI (GGCC), HpaII (CCGG), MspI (CCGG) and HhaI (GCGC), thus the enzyme we cloned is probably the modification

Fig. 1. Methylation test of plasmids pBep1 and pBep2.

Electrophoresis in 0.8 % agarose gel.

1. pBep1
2. pBep1 digested with BepI
3. pBep1 + pBR322 digested with BepI
4. pBep2
5. pBep2 digested with BepI
6. pBep2 + pBR322 digested with BepI
7. pBR322 digested with BepI

pBR322 DNA was added to digestion mixtures 3 and 6 to exclude the possibility of enzyme inhibition.

methylase of the BepI restriction-modification system, and not a nonspecific C-methylase.

Restriction mapping revealed that pBep1 contained a 5.8 kb EcoRI fragment which was partially overlapping with the 3.3 kb insert in pBep2 (Fig.2.).

Since in many cases the genes coding for the restriction endonuclease and the modification methylase were found closely linked we tested whether the clones carrying pBep1 or pBep2 produce BepI endonuclease. No BepI activity could be detected in cell-free extracts purified on a BioGel A-0.5m column.

Methylation specificity of M.BepI

To determine which C within the recognition sequence CGCG is methylated by M.BepI, pyrimidine specific chemical sequencing reactions were done with a DNA fragment isolated from both

Fig. 2. Schematic map of the DNA region coding for the BepI methylase.

- A) Circular map of plasmids carrying the BepI methylase gene. Symbols m⁺ and m⁻ indicate the modification phenotype of the clones.
- B) The sequenced region is shown by the thick horizontal bar and the coding sequence of the BepI methylase by the thick arrow. Selected restriction sites including those used for subcloning fragments for sequencing are shown.

unmodified (pBR322) and modified (pBep1) DNA. The missing band in the modified DNA (Fig.3.) identifies the external cytosine as the methylated base. M.BepI is similar in methylation specificity to the BsuE methyltransferase (37).

DNA sequence

The M.BepI gene was localized by subcloning it on shorter fragments. Deleting a PvuI fragment of pBep1 did not influence

Fig. 3. Detection of 5-methyl-cytosine in the DNA methylated by *M.BepI*. Autoradiographic picture of a 6% sequencing gel. C+T (lanes 1 and 4) and C specific (lanes 2 and 3) reactions with modified (lanes 1 and 2) and unmodified (lanes 3 and 4) pBR322 DNA. Part of the sequence (around the CGCG site at 702 in the pBR322 sequence) is shown. The arrow indicates the missing band in the modified DNA.

the methylation phenotype (pBep4, Fig. 2.). Deletion of an *EcoRI* fragment from pBep2 (Fig. 2.) yielded a plasmid (pBep5) which was completely sensitive to *BepI*. These data suggested that the *M.BepI* coding sequence (or most of it) lies between the *EcoRI* site and an *MboI* site as shown on the circular map of pBep2 (Fig. 2.). A detailed restriction map of this region was established, then specific fragments were cloned in M13 vectors and sequenced. The sequence of 1721 base pairs was determined: it

```

1  AAT ATT ATT TAT ATA AAT TAT ATT ATT TAT ATA AAT AAT ATT
43 ATC AAG AGT TTA CTT GTG ACA CAC GAT GAA ATT TTT AAA TGC
85 ATT GTC GAT GCT GTA AAT AAT GCT CCT AGA AAT AGA CAA ACT
127 GCT GAA TTA CAT ATT CAA ATG ATT AAA TAC GCT GAT GAG TTG
169 AAA TCC ATC ACA TCT AAA GAA TTC TGT GAA GAA ACA AAG TTT
211 CCG ATT GGA TAC GGT TCC GAA TTT AGC AAA ATG CGG AAT ATA
253 ACT GAA CGT TTA AAA GCA GCT GGA TTG GAT GTA TCC AAA CTT
295 TAA GCC ATA TCA TTA TTT TTA GAA GAG TTC ATT TTG AAA GTT
      M K V
337 TTA TCT TTA TTT TCA GGT TGT GGT GGT ATG GAT CTT GGT TTA
      L S L F S G C G G M D L G L
379 GAA GGC GGC TTC TTA GCA CAT AGA TCG TCT ATC AAT AGT GAT
      E G G F L A H R S S I N S D
421 TTA TAT GCT TCT TAT ATT TCA GAC CAT GAT GAA AAT TAT GTT
      L Y A S Y I S D H D E N Y V
463 TAT TTA AAA AAA ACA GGA TTT GAA ACT GTT TTT GCT AAT GAT
      Y L K K T G F E T V F A N D
505 ATT TTG CCT TTT GCA AAA TTG GCT TGG TGT AAT TTT TTC AAA
      I L P F A K L A W C N F K
547 AAT AGA GTA AAT CAA CCT GAA AAT ATC TTT CAT TTA GAA AGT
      N R V N Q P E N I F H L E S
589 ATC GTT GAT GTT GTC AAT AAT ATT GAG AAT AAG CAA TTT TCC
      I V D V V N N I E N K Q F S
631 TTT CCC AAT GAT ATT GAT GTA GTG ACT GGT GGA TTT CCA TGT
      F P N D I D V V T G G F P C
673 CAA GAC TTT AGC TTT GCA GGT AAA AGA GGG TTT GAT TCT
      Q D F S F A G K R K G F D S
715 CAT AAA GAT CAT AAT GGC ATC ATT TAT AAT GAA CCA ACT GAA
      H K D H N G I I Y N E P T E
757 GCT ACG CGA GGT CAA CTA TAT TTA TGG CTG AAA AAA GTA GTT
      A T R G Q L Y L W L K K V V
799 GAA ATC ACA AAA CCT AAA GTA TTC ATT GCA GAA AAT GTA AAG
      E I T K P K V F I A E N V K
841 GGG TTA GTA ACT TTA GGC GAT GTA AAA GAC ATT ATT CAA AAA
      G L V T L G D V K D I I Q K
883 GAT TTT AGA AAT ATT GAT GGG TAT GTT GTT CTT GAT GCT
      D F R N I D D G Y V V L D A
925 CAG GTT CTT AAT GCT AAA AAT TAT GGA GTT GCT CAA AAT AGA
      Q V L N A K N Y G V A Q N R
967 GAA AGA GTT ATT TTC ATT GGT ATA TCA AAA CGT TAT GCT AAC
      E R V I F I G I S K R Y A N
1009 AAA AAA ATC TTA GAT GAA CTA ATT TCT TTG CAA GAA AAA TCT
      K K I L D E L I S L Q E K S
1051 GAA GTC TAT CCA TAT CCA CCA TAT ACT GAT GGT ACT GAT CCA
      E V Y P Y P P Y T H G T D P
1093 GAG CTA AAA CCA TAT GCG ACT TTA AAT CAA ATT CTA GCT CAT
      E L K P Y A T L N Q I L A H
1135 CTT CCA GAG CCT GAG TTG GCA AGC ACA GAT AAG AGC CAA CAA
      L P E P E L A S T D K S Q Q
1177 TCT TAT TCT AAG GCT AAA CTT TTC AAA AAA ACT CAG GGC AAT
      S Y S K A K L F K K T Q G N
1219 ATT GAA GTT AAC ATG AAT GGC CAA GCT CCA ACT ATT AGA GCA
      I E V N M N G Q A P T I R A
1261 GAA CAT CAT GGA AAT ATT GAG TTT AGA AGA TTA TCC AAA GAA
      E H H G N I E F R R L S K E
1303 AAT GGC GGG ACT AAT CTA TCT GAA TTA CAT CTA CCA CAA AGA
      N G G T N L S E L H L P Q R
1345 CGT CTC ACA GTC GCG GAG TGC GCT TTA ATT CAA AGT TTT CCA
      R L T V R E C A L I Q S F P
1387 CCA GAT TAT GAG TTT GTA TTT AAC TAC GGA AAA GCC AAT TCA
      P D Y E F V F N Y G K A N S
1429 GTA TCA GCA TCG GCC GCA TAT AAA ATT ATC GGT AAT GCC GTT
      V S A S A A Y K I I G N A V
1471 CCT CCA TTA TTA GGA TTT GCA ATA GGC AGG CAT CTT AGT CAA
      P P L L G F A I G R H L S Q
1513 ATA TGG GAC AAA TTA TTT AAA ACA TGA ATT TAC TAT TCC ATA
      I W D K L F K T
1555 TTT TGA ATA TTA TTT AGA ATC ATT TTC ATA TCT GAA TCA TAT
1597 GAT TCA ACG ATA CAT ACA AAG TCT TTT AGA TGA CTA TAT TTT
1639 TTT AAA TTT CTA ATC CAA TAA CCT ATG GTT GTA CTA GAA CAA
1681 CCA TAG TAT TTA GCA ATA TCT TTG TAA TAG ACG TGA TAG TT

```

Fig. 4. Nucleotide sequence of the *BepI* methylase gene and the deduced amino acid sequence of the methylase protein. Dots indicate the putative ribosome binding site.

Fig. 5. Regions of cytosine methylases where amino acid sequence similarities are most apparent. Amino acids conserved in all 11 methylases are boxed. Numbers refer to the sequence-position of the first amino acid.

starts with an SspI site and ends with an HpaI site (Fig.2 and 4.).

The sequence contains only one reading frame longer than 150 amino acids. This reading frame has two potential start codons: TTG at 328 and ATG at 364. TTG₃₂₈ is more likely to be the real start codon of the methylase gene because the protein sequence defined by the nucleotide sequence between positions 328 and 364

BepI	244	Y	P	P	P	Y	T	H	G	T	D	P	E	L	K	P	Y	A	T	L	
BspRI	253	Y	E	Y	P	E	I	T	H	G	N	E	E	G	L	K	P	Y	V	T	L
BsuRI	254	Y	K	Y	P	S	P	T	H	G	E	E	T	G	L	K	P	F	K	T	L

Fig. 6. Region of partial homology in the variable segment of the BepI, BspRI and BsuRI methylase proteins. Amino acids conserved in all 3 methylases are boxed.

shows similarity with other cytosine methylases (Fig. 5.). TTG as initiation codon seems to be not unusual in cytosine methylases (8,9,17,18). Since there is an in-frame stop codon at position 295, it is unlikely that the translation would start further upstream. The reading frame between TTG₃₂₈ and stop codon TGA₁₅₃₇ defines a protein of 403 amino acids (M_r : 45,447). There are no typical E. coli transcriptional initiation and termination signals in the sequence.

The difference in the methylation phenotypes of plasmids pBep1, pBep2, pBep4 and pBep5 and the lack of typical E. coli promoter structure before the coding sequence suggest that in E. coli cells containing pBep2, transcription of the BepI methylase gene starts at a sequence deleted in pBep5. Stronger expression of BepI methylase from pBep1 is probably due to transcription starting from the P1 promoter of pBR322 (38). Further evidence to support this comes from recloning the EcoRI fragment of pBep1 in pACYC184 (39). The BepI methylase was expressed only if the orientation of the EcoRI fragment was such that the M.BepI gene could be transcribed from the chloramphenicol acetyltransferase promoter (data not shown).

Protein structure

M.BepI is the first CGCG specific DNA methylase of which the coding sequence has been determined. The size of the predicted protein (403 amino acids, M_r =45,447) is similar to other cytosine methylases. Comparison of the amino acid sequences detected similarities between M.BepI and other cytosine methylases (Fig.5.). Most of these similarities fall in six blocks (A-F, as defined by Som et al., 14). The order of these blocks in M.BepI is the same as in other cytosine methylases. This reinforces the notion that cytosine methylases have similar structure.

It has been postulated that the conserved regions are responsible for the general steps of the methylation reaction (e.g. SAM binding) and the long variable region between homologous blocks D and E is responsible for target recognition (17,18). There is experimental evidence that in multispecific phage-coded methylases (SPR, ϕ 3T, ρ 11_g) the target recognizing domain is indeed located between homologous blocks D and E (40,41). This has not been shown for the monospecific methylases but the similarity detected between the two types of enzymes suggest a common building plan. In this respect it is perhaps interesting that in addition to the six conserved regions common in all C-methylases M.BepI shares another conserved region with two GGCC specific methylases (Fig.6.).

Construction of hybrids of different multispecific DNA methylases was very useful in identifying the target recognizing domains of these enzymes (40,41). Similar experiments with M.BepI and other monospecific methylases are underway in our laboratory.

ACKNOWLEDGEMENTS

We thank E. Csorba for her excellent technical assistance, A. Simoncsits for the sequencing primer and J. Posfai for help with part of the computing.

Present addresses: *Biotechnology Centre, Martin Luther University, Universitätsring 5, Halle 4020, GDR and +Molecular Genetics Centre, PO Box 130 (8), Beijing 100850, China

REFERENCES

1. Greene, P.J., Gupta, M., Boyer, H.W., Brown, W.E. and Rosenberg, J.M. (1981) *J. Biol. Chem.* **256**, 2143-2153.
2. Newman, A.K., Rubin, R.A., Kim, S. and Modrich, P. (1981) *J. Biol. Chem.* **256**, 2131-2139.
3. Schoner, B., Kelly, S. and Smith, H.O. (1983) *Gene* **24**, 227-236.
4. Brooks, J.E., Blumenthal, R.M., and Gingeras, T.R. (1983) *Nucl. Acids Res.* **11**, 837-851.
5. Posfai, G., Kiss, A., Erdei, S., Posfai, J. and Venetianer, P. (1983) *J. Mol. Biol.* **170**, 597-610.
6. Walder, R.Y., Walder, J.A. and Donelson, J.E. (1984) *J. Biol. Chem.* **259**, 8015-8026.
7. Bougueleret, L., Schwarzstein, M., Tsugita, A. and Zabeau, M. (1984) *Nucl. Acids Res.* **12**, 3659-3676.

8. Buhk, H.J., Behrens, B., Tailor, R., Wilke, K., Prada, J.J., Gunthert, U., Noyer-Weidner, M., Jentsch, S. and Trautner, T.A. (1984) *Gene* 29, 51-61.
9. Posfai, G., Baldauf, F., Erdei, S., Posfai, J., Venetianer, P. and Kiss, A. (1984) *Nucl. Acids Res.* 12, 9039-9049.
10. Macdonald, P.M. and Mosig, G. (1984) *EMBO J.* 3, 2863-2871.
11. Mannarelli, B.M., Balganes, T.S., Greenberg, B., Springhorn, S.S., Lacks, S.A. (1985) *Proc. Natl. Acad. Sci. USA* 82, 4468-4472.
12. Kiss, A., Posfai, G., Keller, C.C., Venetianer, P. and Roberts, R.J. (1985) *Nucl. Acids Res.* 13, 6403-6421.
13. Theriault, G., Roy, P.H., Howard, K.A., Benner, J.S., Brooks, J.E., Waters, A.F. and Gingeras, T.R. (1985) *Nucl. Acids res.* 13, 8441-8461.
14. Som, S., Bhagwat, A.S. and Friedman, S. (1987) *Nucl. Acids Res.* 15, 313-331.
15. Sznyter, L.A., Slatko, B., Moran, L., O'Donnel, K.H. and Brooks, J.E. (1987) *Nucl. Acids Res.* 15, 8249-8266.
16. Caserta, M., Zacharias, W., Nwankwo, D., Wilson, G.G. and Wells, R.D. (1987) *J. Biol. Chem.* 262, 4770-4777.
17. Tran-Betcke, A., Behrens, B., Noyer-Weidner, M. and Trautner, T.A. (1986) *Gene* 42, 89-96.
18. Behrens, B., Noyer-Weidner, M., Pawlek, B., Lauster, R., Balganes, T.S. and Trautner, T.A. (1987) *EMBO J.* 6, 1137-1142.
19. Slatko, B.E., Benner, J.S., Jager-Quinton, T., Moran, L.S., Simcox, T.G., Van Cott, E.M., Wilson, G.G. (1987) *Nucl. Acids Res.* 15, 9781-9796.
20. Narva, K.E., Wendell, D.L., Skrdla, M.P. and Van Etten, J.L. (1987) *Nucl. Acids Res.* 15, 9807-9823.
21. Karreman, C. and de Waard, A. (1988) *J. Bact.* 170, 2527-2532.
22. Sullivan, K.M. and Saunders, J.R. (1988) *Nucl. Acids Res.* 16, 4369-4387.
23. Hattman, S., Wilkinson, J., Swinton, D., Schlagman, S., Macdonald, P.M. and Mosig, G. (1985) *J. Bact.* 164, 932-937.
24. Lacks, S.A., Mannarelli, B.M., Springhorn, S.S. and Greenberg, B. (1986) *Cell* 46, 993-1000.
25. Lauster, R., Kriebardis, A. and Guschlbauer, W. (1987) *FEBS Lett.* 220, 167-176.
26. Venetianer, P. and Orosz, A. (1988) *Nucl. Acids Res.* 16, 350.
27. Raleigh, E.A. and Wilson, G. (1986) *Proc. Natl. Acad. Sci. USA* 83, 9070-9074.
28. Blumenthal, R.M., Gregory, S.A. and Cooperider, J.S. (1985) *J. Bact.* 164, 501-509.
29. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
30. Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heynecker, H.L., Boyer, H.W., Cross, J.H. and Falkow, S. (1977) *Gene* 2, 95-113.
31. Norrander, J., Kempe, T. and Messing, J. (1983) *Gene* 26, 101-106.
32. Messing, J. (1983) *Methods Enzymol.* 101, 20-79.
33. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463-5467.

34. Biggin, M.D., Gibson, T.J. and Hong, G.F. (1983) Proc. Natl. Acad. Sci. USA 80, 3963-3965.
35. Ohmori, H., Tomizawa, J. and Maxam, A.M. (1978) Nucl. Acids Res. 5, 1479-1485.
36. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
37. Gaido, M.L., Prostko, C.R. and Strobl, J.S. (1988) J. Biol. Chem. 263, 4832-4836.
38. Stuber, D. and Bujard, H. (1981) Proc. Natl. Acad. Sci. USA 78, 167-171.
39. Chang, A.C.Y and Cohen, S. (1978) J. Bacteriol. 134, 1141-1156.
40. Balganesch, T.S., Reiners, L., Lauster, R., Noyer-Weidner, M., Wilke, K. and Trautner, T.A. (1987) EMBO J., 6, 3543-3549.
41. Trautner, T.A., Balganesch, T.S. and Pawlek, B. (1988) Nucl. Acids Res. 16, 6649-6658.