Nucleotide Sequence of the Gene Coding for the BanIII DNA Methyltransferase in Bacillus aneurinolyticus

Bunsei Kawakami,* Akiko Sasaki, Masanori Oka and Yoshihiko Maekawa
Tsuruga Enzyme Plant, Toyobo Co., Ltd., Tsuruga, Fukui 914, Japan
Received June 18, 1990

The gene coding for the ATCGAT specific BanIII DNA methyltransferase (M-BanIII) of Bacillus aneurinolyticus was cloned and its nucleotides sequenced. The coding region was assigned on the nucleotide sequence on the basis of the N-terminal amino acid sequence and molecular weight of the enzyme. The M-BanIII gene coded for a protein of 580 amino acid residues (MW 66,344). Comparison with other methylases indicated that the M-BanIII sequence contained a segment of tetra-amino acids, NPPY, characteristic of N6-adenine methylases. In addition, some homologous regions were found in the sequences of type II adenine methylases PaeRII(CTCGAG), TaqI(TCGA) and PstI(CTGCAG), containing TCGA within the recognition sequences.

The type II restriction-modification system consists of two genes, one coding for a deoxyribonuclease (R) and the other for a methyltransferase (M). Many of these genes have been cloned in E. coli and their nucleotides sequenced.1-3) Comparison of the protein sequences encoded in these genes has shown that little sequence similarity exists between methylase and endonuclease in the same system, but they recognize the same target sequence.4) Nor was their sequence similarity among the endonucleases in different systems. It was noted that methylases that methylate cytosine residues at C5 and those that methylate adenine residues at N6, respectively, contain some common sequences.5-8) We have cloned the BanI and BanIII restriction-modification genes in E. coli,9) and sequenced the nucleotides of the BanI methylase gene. The BanI methylase was found to contain a sequence common to C5-cytosine methylases.1) In this report we sequenced the nucleotides of the BanIII methylase gene and compared its amino acid sequence with those of other methylases.

Materials and Methods

Strain and media. B. aneurinolyticus IAM1077 was obtained from the Institute of Applied Microbiology, the University of Tokyo. The E. coli strains used were HB10110) and JM10911). The plasmid vectors were pBR322 and Bluescript II.12) Plasmid pBanIIIRM12 (Fig. 1) was constructed as follows. Chromosomal DNA from B. aneurinolyticus IAM 1077 was partially digested with Sau3AI. Using T4 ligase, the DNA was ligated with

* Corresponding author.
pBR322, which had been digested with BamHI and treated with alkaline phosphatase. The ligation mixture was used to transform *E. coli* HB101. Ampicillin-resistant transformants (approximately 10,000) were grown to early stationary phase in 500 ml of LB containing 50 μg/ml ampicillin and the cells were used for preparation of plasmid DNA. These plasmids were digested with *BanIII* to remove the non-methylated DNA, and re-introduced into *E. coli* HB101. The cells were plated on LB agar plates containing ampicillin and the resistant transformants were selected. Among them, pBanIIIRM12 was obtained. *B. aneurinolyticus* IAM1077 and *E. coli* cells were grown in LB medium, at 30° and 37°C, respectively.

**Enzymes and chemicals.** All the restriction endonucleases, T4 DNA ligase, and bacterial alkaline phosphatase used were products of Toyobo Co., Ltd., Osaka. Other chemicals were reagent grade commercial products.

**Manipulation of DNA.** Preparation of plasmids, transformation of *E. coli* cells, restriction mapping, and agarose gel electrophoresis were done by the standard procedures.13

**Assay of M-BanIII activity.** The M-BanIII activity was assayed as follows. The enzyme solution (5 μl) was incubated for 1 hr at 37°C in a reaction mixture (50 μl) containing 1 μg of λ-DNA, 10 mM Tris-HCl, pH 7.5, 7 mM 2-mercaptoethanol, 10 mM MgCl2, and 80 μM S-adenosyl-L-methionine. The reaction was stopped by heating for 15 min at 65°C. Following addition of 10 units of R-BanIII, the solution was incubated for 1 hr at 37°C, and the products were analyzed by gel electrophoresis. One unit of enzyme activity was defined as the amount of enzyme required for complete protection from the R-BanIII digestion. Protein was measured by the method of Bradford et al.14

**DNA sequencing.** DNA fragments were cloned in Bluescript II and were sequenced by the dideoxy chain termination method using the Model 370A sequencer of Applied Biosystems. Inserts were shortened from the end by the combined action of exonuclease III and mung bean nuclease and sequenced with the standard M13 primer.

**Analysis of N-terminal amino acids sequences.** The sequential degradation of protein was done with the model 470A sequencer of applied Biosystems, equipped with the online HPLC apparatus Model 120A.

**Sequence comparison.** The software package GENETYX (SDC, Tokyo) was used for analysis of DNA and amino acid sequences. The sequences examined were the following DNA methylases: BsuRI, BspI, CviBIII, DdeI, DpnII, EcoRI, EcoRII, EcoRV, HhaI, HpaII, NgoPII, PvuII, PaeR7I, PstI, SstI, and TaqI.

**Results and Discussion**

**Purification of M-BanIII**

*E. coli* HB101 (pBanIIIRM12, Fig. 1) cells were cultured for 12 hr at 30°C in a 10-l jar fermentor. The cells were harvested by

---

**Table I. Purification of M-BanIII from 50 g of Cells of E. coli (pBanIIIRM12)**

<table>
<thead>
<tr>
<th>Fraction</th>
<th>Total protein (mg)</th>
<th>Total activity (kU)</th>
<th>Specific activity (kU/mg)</th>
<th>Yield (%)</th>
<th>Purification (fold)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crude extract</td>
<td>5,400</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>PEI treatment</td>
<td>2,863</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Ammonium sulfate</td>
<td>2,506</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Phosphocellulose</td>
<td>12.9</td>
<td>168</td>
<td>13.1</td>
<td>100</td>
<td>1</td>
</tr>
<tr>
<td>DE 52 cellulose</td>
<td>4.3</td>
<td>126</td>
<td>29.4</td>
<td>75</td>
<td>2.2</td>
</tr>
<tr>
<td>Heparin Sepharose</td>
<td>0.51</td>
<td>92</td>
<td>182</td>
<td>54.8</td>
<td>13.9</td>
</tr>
</tbody>
</table>

---

**Fig. 2.** Estimation of the Molecular Weight of M-BanIII by SDS-Polyacrylamide Gel Electrophoresis. Molecular weight markers [lane M] and the purified M-BanIII preparation were electrophoresed on gradient SDS-polyacrylamide gels (10-20%) and protein bands were made visible with silver stains. Numbers by the side of lane M are molecular masses of marker proteins (kilo daltons).
Fig. 3. The Entire Nucleotide Sequence of the \textit{Sall} (2)--HindIII Fragment.

The potential ribosome-binding sequence is dotted and a pair of arrows shows the palindromic sequence characteristic of the termination signal. The nucleotide sequence was numbered from the \textit{Sall} (2) site, and the predicted amino acid sequence for M-BanIII was indicated under the nucleotide sequence.

centrifugation and stored at \(-20^\circ\)C until use. The yield of cells was about 5g (wet weight) per liter. The cells (50g) were suspended in 250ml of 10mM Tris-HCl, pH 7.5, 10mM MgCl\(_2\), and 7mM 2-mercaptoethanol and disrupted by sonication. The cell debris was removed by centrifugation at \(8,000 \times g\) for 20min. To the supernatant was added NaCl to 0.1M and then a 10% stock solution of polyethyleneimine to 0.1%. After removal of the precipitates by centrifugation, solid ammonium sulfate was added to 60% saturation. The resulting precipitates were collected by centrifugation, dissolved in 10mM potassium phosphate buffer (pH 7.5) containing 7mM 2-mercaptoethanol and 5% glycerol (buffer A),
Fig. 4. The Amino Acid Sequence of M-BanlII (Upper Lane) Was Aligned with the Amino Acid Sequence of M-PaeR1 (Lower Lane) Using the GENETYX Program HOMOGAPP. The number of amino acids is indicated on the left margin. Star denotes that the amino acid sequences are identical. The NPPY sequence is indicated by a rectangle.

1. MELTIEEMLIKQETGAHYTPTDLDI KALKRNLKKEGSSGT TKIRGLDPSCDGE LL
   MVDFILDLAGYTEDQPLHEKLLEPSFGGDFL

2. LS-LNRIAKFNNIDNIELIGEDKEAIKEADFRNLEMNGINDAKLATGDFDLMDLEGNL

3. 4" LPIIQRLSAWARAARPNGTEVDGDAOVELLHDTFRSTYAAVALLKREGLSANAAT

4. SLFDDLSK----IEPVD----LIIANPPYVRTQVLGADRAQKALKFN-LKGVRDMLYHA

5. ALADRLQSGDFLLAPLEQDFPVQVNPYYVRPEIPLAPLAEYRSRQTMYDRADITYIP

6. FLVAMLQLEKPGGLIGVITSKYNLAXSSGESIQRILFAENYDIIEIMDLGDTKLFSAVLQ

7. FIERSLTALSAGNLGFICADRMKNNYGGPLRSLVAERFLKVVYDMVDTAFHDV--

8. A1FFGRKVKNGIRQTAPANYKLYETEDPSKTEVISIKFETLFLGLESSNVTGFNVDKEF

9. -IAYPAITI-----ISREGGGA----TRIARHPSIRATLTLTLAGL----SAKLPDKAG--

10. YSVSCGKLVPDFSFKEPWMATDEYNYWNTININNSYCTIQDLCDVYGIKTDAKVF1K

11. -PVRELARYTNGAERGCWSLTT--RWRLFAWRARSHCSKRGLARGFIVGATGADKAFI--

12. STWEILPDEKPEVEVLKLLISTDHASKWPRLEIPNQILYTHENVNGKKAIIDFTQYP

13. GDFESLDVEPRKLPLTVTKDITGEGVW----RGQGVY----NPFAESGLVGLGEYP

14. HALAYLEHTRELEGKRKYIAKARNWNEIWPQNPWHALPKIVFPDISPEPKFYDEEG

15. RLRRLYERARDDVIAGRHCASSAPWNYRITIDRTAPLAARPKLIPDIFKGEHIVFEGGE

16. CCIDNCYIIPPKEENNNDLFLIGISNTKYMTNYHDIAPNKNLYPGRRRYLTQYVSKY

17. LYPFHLHLYVTSDDWLRLAGAAVLLASVRLFVATY-----STKMRGGFRFQAQLLYRRI

18. PLPNEANQSEIQIDLVRLEVFQNPNDKIEIENIENLTALAFGVERLX

19. RIPRWDVPEPLRRELAAEAIKRDVQACNRAFRLYGLSHEERSALGGGEX

Fig. 4. The Amino Acid Sequence of M-BanlII (Upper Lane) Was Aligned with the Amino Acid Sequence of M-PaeR1 (Lower Lane) Using the GENETYX Program HOMOGAPP. The number of amino acids is indicated on the left margin. Star denotes that the amino acid sequences are identical. The NPPY sequence is indicated by a rectangle.

and dialyzed against the same buffer. The solution was put onto a phosphocellulose column (Whatman P11 4 x 10 cm) and chromatographed with a linear gradient in buffer A (0-1 M KCl, 400 ml). M-BanlII activities were eluted at 0.1-0.2 M. The active fractions were pooled, dialyzed against buffer A, and put on a DE 52 cellulose column (Whatmann 2.5 x 10 cm), and eluted with a linear NaCl gradient in buffer A (0-0.5 M, 200 ml). The activity was eluted at the buffer of 0.4-0.5 M NaCl. The active fractions were pooled, dialyzed against buffer A, and put on a Heparin Sepharose column (Pharmacia 1 x 10 cm), and eluted with a linear KCl gradient in buffer A (0-1.0 M, 100 ml). The M-BanlII activity was eluted at 0.3-0.4 M. The enzyme fraction was concentrated by glycerol, and stored at -20°C. A summary of the purification is presented in Table I.

Molecular weight and N-terminal amino acids sequence of M-BanlII

When the M-BanlII preparation purified
Nucleotide Sequence of the BanIII Methylase Gene

Fig. 5. The Amino Acid Sequence of M-BanIII (Upper Lane) Was Aligned with the Amino Acid Sequence of M-TaqI (Lower Lane) Using the GENETYX Program HOMOGAPP. The number of amino acids is indicated on the left margin. Star denotes that amino acid sequences are identical. The NPPY sequence is indicated by a rectangle.

above was analyzed by SDS-polyacrylamide gel electrophoresis (Fig. 2), a single band was obtained. Its molecular weight was estimated to be 66,000 from the mobility relative to markers. The N-terminal amino acid sequence of M-BanIII was analyzed by Edman degradation, and the first 6 amino acids from the N-terminal were deduced to be Met-Glu-Leu-Thr-Ile-Glu.

DNA sequence

According to deletion analysis done before, the R-BanIII and M-BanIII genes are located as indicated in Fig. 1. In agreement with this assignment, a derivative in which the SalI(1)–SalI(2) region was deleted produced a considerable level of the M-BanIII activity in E. coli cells. Thus we established the entire nucleotide sequence of the SalI(2)–HindIII region. Analysis of the reading frames indicated that the sequence contained one large open frame, 1,740 nucleotides in length. In this open frame, the ATG codon appeared at nucleotide position 102 and the termination codon at nucleotide position 1,842. We assigned this ATG codon to the initiation codon of M-BanIII, because the predicted molecular weight (66,344) agreed well with the molecular weight of M-BanIII estimated by SDS-polyacrylamide gel electrophoresis (Fig. 2). The start site of the BanIII gene was further confirmed by N-terminal amino acid analysis. The first 6 amino acid sequence from the N-terminus exactly corresponded to that
predicted from the nucleotide sequence at nucleotide position 102. In addition, an appropriate ribosome binding sequence, AGG, appeared 6 base pairs upstream of the ATG codon (see dots in Fig. 3). Immediately downstream from the M-BanII gene, a palindromic sequence followed by a stretch of thymine residues is seen (see arrows in Fig. 3). It is likely that transcription ends at this site.

**Comparisons of protein sequences**

The amino acid sequence of M-BanII was compared with published sequences of other methylases, and it was found that M-BanII contained a segment of tetra-amino acids that has been identified as a characteristic of N6-adenine methylases.\(^2,17-19\) The consensus sequence of the segment is Asp/Asn-Pro-Pro-Tyr(NPPY),\(^7,20\) and the sequence in BanII was Asn-Pro-Pro-Tyr, as indicated by a rectangle in Fig. 3. When the region of similarity was searched using the HOMOGAPP program from GENETYX, M-BanII was found to contain some homologous
regions surrounding the NPPY sequence with M-PaeR71\textsuperscript{17} (Fig. 4), M-TaqI\textsuperscript{18} (Fig. 5), and M-PstI\textsuperscript{19} (Fig. 6). With the comparison of amino acid sequences surrounding the NPPY sequence, Narva et al. grouped N6-adenine methylases into four classes.\textsuperscript{8} According to their classification, M-TaqI, M-CviBIII, M-PaeR7I, and M-PstI are included in class III, so M-BanIII may also be grouped in this class. It is assumed that the common sequence, NPPY, is located in some important domains for recognition and methylation. An interesting feature was that the degree of similarity between M-BanIII and M-PaeR7I was higher than that between M-BanIII and M-TaqI or M-PstI. Recently, Kaszubska et al. reported that M-EcoR1 and M-RsI show very little similarity, although both enzymes recognize the same sequence.\textsuperscript{4} An interpretation for the detected similarity could be that M-BanIII and M-PaeR7I evolved from a common ancestor.

Acknowledgment. We thank Dr. M. Takanami, Institute for Chemical Research of Kyoto University for his help during preparation of the manuscript and for encouragement.

References