Cloning and Nucleotide Sequence of ApaLI Restriction-modification System from Acetobacter pasteurianus IFO 13753

Toshihiko Suzuki, Emi Sugimoto,* Yasutaka Tahara,* and Yuzo Yamada*††

United Graduate School of Agricultural Science, Gifu University (Shizuoka University), Gifu 501-11, Japan
*Laboratory of Applied Microbiology, Department of Applied Biological Chemistry, Faculty of Agriculture, Shizuoka University, 836 Ohyo, Shizuoka 422, Japan

Received September 21, 1995

The ApaLI restriction-modification system from Acetobacter pasteurianus IFO 13753 recognizes the nucleotide sequence GTGCAC. The gene coding for the ApaLI methylase (M.ApaLI) was cloned into Escherichia coli DH5zMCR, and the nucleotide sequence of the gene was analyzed. The M.ApaLI gene coded for a protein of 429 amino acid residues (molecular mass, 46,554 daltons). The ApaLI restriction endonuclease (R.ApaLI) gene was analyzed by inverse polymerase chain reaction. The R.ApaLI gene coded for a protein of 375 amino acid residues (molecular mass, 42,143 daltons). The two genes had the same orientation separated by two base pairs. The deduced amino acid sequence of M.ApaLI shows significant similarities to the family of cytosine-5 methylases. However, the deduced amino acid sequence of R.ApaLI did not have as much relatedness in the nucleotide sequence, when compared with those of the other restriction endonucleases already reported.

Key words: DNA methylase; restriction endonuclease; ApaLI; restriction-modification system; Acetobacter pasteurianus

Acetobacter pasteurianus IFO 13753 produces a new restriction endonuclease, designated ApaLI (=R.ApaLI), which recognizes the palindromic hexanucleotide sequence GTGCAC, cuts between G and T, and forms a 5′-cohesive tetranucleotide extension. We have attempted cloning of the genes responsible for the ApaLI restriction-modification system.

We have reported the cloning and the nucleotide sequences of the ApaLI methylase (M.ApaLI) and the ApaLI restriction endonuclease (R.ApaLI) genes.

Materials and Methods

Bacterial strains and plasmids. Acetobacter pasteurianus IFO 13753 was obtained from Institute for Fermentation, Osaka, Japan (IFO). Escherichia coli DH5zMCR, and pUC18 and pUC19 were used as a host and plasmid vectors, respectively, for constructing libraries and sequencing DNA.

Cloning of ApaLI methylase gene. Acetobacter pasteurianus IFO 13753 was grown as described previously. Total DNA from the strain was prepared by the procedure of Murray and Thompson and digested partially with SalI. The DNA obtained was ligated to pUC18, which had been cleaved with BamHI and dephosphorylated by treatment with bacterial alkaline phosphatase.

The ligated DNA was used to transform E. coli DH5zMCR by the standard methods.

The plasmid DNAs isolated from about 10,000 Ampicillin (Ap)-resistant transformants were digested completely with an excess of R.ApaLI, and the resulting digests were used to transform E. coli DH5zMCR. The plasmids carrying the M.ApaLI gene were isolated from the transformants which survive against the R.ApaLI digestion.

Assay for ApaLI methylase activity. The cells from individual colonies were inoculated in a small volume of culture broth (1.0 ml) and incubated overnight at 37°C. The plasmid and chromosomal DNAs were prepared from colonies and tested for their sensitivity to R.ApaLI. The DNAs (0.5 μg) were digested in reaction mixtures containing 10 mM MgCl2, 1 mM dithiothreitol, R.ApaLI (5 units), and 10 mM Tris-HCl (pH 7.5). The reaction mixtures were then analyzed by agarose gel electrophoresis, and the DNA modification was detected by a lack of digestion.

Nucleotide sequencing. The nucleotides were sequenced by the dideoxy chain termination method using a Dye deoxy terminator cycle sequencing kit (Applied Biosystems Inc., Foster, California, U.S.A.) and a model 373A DNA sequencer (Applied Biosystems Inc., Foster, California, U.S.A.). The deletion derivatives were obtained using a Kilo-sequence deletion kit (Takara Shuzo Co., Kyoto, Japan).

Subcloning of ApaLI methylase gene. The inserted DNA fragment of pPAPALM3.2 was digested with restriction endonucleases, EcoRI, HindIII, and BamHI, and ligated into the multi-cloning site of pUC18 or pUC19. The ligated mixtures were used to transform E. coli DH5zMCR. The plasmids obtained were examined for their resistance to the digestion by R.ApaLI.

Purification and analysis of N-terminal amino acid sequence of R.ApaLI. R.ApaLI was purified from cells of A. pasteurianus IFO 13753 as described in a previous paper. The N-terminal amino acids were sequenced with a model PPSQ-10 amino acid sequencer (Shimadzu Co., Kyoto, Japan).

Polymerase chain reaction amplification and analysis of ApaLI restriction endonuclease gene. An inverse polymerase chain reaction (PCR) technique was used for analysis of the complete R.ApaLI gene. To obtain a suitable restriction fragment for inverse PCR, Southern hybridization was carried out. Southern hybridization was done using an ECL direct nucleic acid labelling and detection system according to the instructions of the supplier (Amersham, Buckinghamshire, England, U.K.).

The inverse PCR was done by 30 cycles using Pfu DNA polymerase according to the instructions of the supplier (Stratagene, La Jolla, California, U.S.A.). The cycling conditions were 94°C (1 min) for denaturation, 60°C (1 min) for annealing, and 72°C (2 min) for elongation. The PCR products were purified by agarose gel electrophoresis. The sequencing of the PCR product was done directly by primer walking using the Dye deoxy terminator cycle sequencing kit.

†† To whom reprint requests should be addressed (23 21 Seinamachi, Fujieda, Shizuoka 426, Japan). Retired on March 31, 1995 from Shizuoka University.

18 Studies on Restriction Endonucleases of Acetic Acid Bacteria and Allied Organisms. Part XIV. For Part XIII, see ref. 18.
Comparison of amino acid sequence. The deduced amino acid sequences of the ApaLI restriction-modification system were compared with those of the protein databases, Swiss-Prot and PIR, by the method of Smith and Waterman. The amino acid sequences of methylases were aligned by the program Clustal W.

Nucleotide sequence accession number. The nucleotide sequence data reported here will appear in the DDBJ, EMBL, and GenBank nucleotide sequence databases under the accession number D78276.

Enzymes and chemicals. All the restriction endonucleases used in this experiment, e.g., Sau3AI, ApaLI, BamHI, HindIII, EcoRI, and Clal, T4 DNA ligase, and bacterial alkaline phosphatase were obtained from Nippon Gene Co., Tokyo, Japan.

Results
Cloning of ApaLI methylase gene
The plasmid clones carrying the M.ApaLI gene were isolated by selecting the recombinants that protectively modified themselves against digestion by R.ApaLI. One plasmid clone was selected by this method. The recombinant plasmid, designated pAPALM3.2, contained an insert of approximately 3.2 kb.

Nucleotide sequence of 3.2-kb insert in pAPALM3.2
The nucleotide sequence of the 3.2-kb fragment was analyzed. In the 3.2-kb fragment, three open reading frames (ORF) were found in the same direction (Fig. 1A). Of the three, a termination codon was not observed in a third open reading frame. To assign the methylase gene, several deletion mutants of pAPALM3.2 were constructed. The deletion analysis of pAPALM3.2 showed that the plasmid containing the 1.8-kb fragment had resistance against R.ApaLI digestion. This fragment contained ORF2 but not ORF1. Thus, it was found that a second ORF encodes M.ApaLI. The sequence of the M.ApaLI gene is shown in Fig. 2. In the ORF assigned to the M.ApaLI gene, two ATG codons appeared close to each other at positions 1568 and 1580. Of the two ATG codons, the latter was preceded by a Shine–Dalgarno (SD)-like sequence (GGAT) five bp upstream, but the former was not. Therefore, we thought that the ATG at position 1580 is a real translational start codon in the ApaLI methylase gene. The M.ApaLI gene was 1287 base pairs (bp) long (coding for positions 1580 to 2866), corresponding to a protein of 429 amino acids with a calculated molecular mass of 46,554 daltons. Putative promoter-like structures were observed at −10 (GATGTT, positions 1467 to 1502) and −35 (GTGATC, positions 1474 to 1479) regions. The third ORF (300 bp long) was not terminated, however, this ORF was expected to encode R. ApaLI. Although the first ORF was about 600 bp long, the genetic role of the ORF was unclarified.

Analysis of the N-terminal amino acid sequence of R.ApaLI
The ApaLI restriction endonuclease was purified and estimated to be approximately 42,000 daltons by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–

![Fig. 1. Organization of ApaLI Restriction and Modification Genes.](image)

The open reading frames, ORF2, ORF3, and ORF1 of pAPALM3.2 correspond to the methylase (M.ApaLI) gene, a part of the endonuclease (R.ApaLI) gene, and an unidentified gene, respectively (A). The HindIII BamHI fragment of pAPAL1.0 was used as a probe for Southern hybridization. For the inverse PCR (B) of the ApaLI endonuclease gene (ORF3), the Clal restriction fragment (2.7 kb) was obtained from total DNA of A. pasteurianus 1F0 13753 by complete digestion. The two primers used in this experiment were designed at the positions indicated by the arrows. The upstream primer was composed of 24 bases, and the downstream primer was of 23 bases. The PCR product with 2.0 kb is indicated by a thick line. The constructions of derivative plasmids, the assay of methylase activities, and the PCR for the analysis of the complete R.ApaLI gene are described in the text. The box arrows indicate the direction of the three genes. plac, lac promoter.
The molecular mass of \( R. \) apoL1 was reported to be 26,000 daltons by gel filtration.\(^{21}\) In the estimation of molecular masses of proteins, SDS-PAGE is assumed to be more accurate than gel filtration.

Nineteen amino acids were detected for the N-terminal region, and the amino acid sequence was X-Thr-Arg-Gln-
Arg-Leu-Ser-Ala-Glu-Arg-Ser-Gln-Gln-Leu-Thr-Arg-Leu-Leu-Thr-Ile (X, unidentified). The N-terminal amino acid sequence detected was consistent with that deduced from the DNA sequence of the third ORF, which was supposed to be a part of the R.ApaLI gene.

**Inverse PCR amplification and nucleotide sequence of ApaLI restriction endonuclease gene**

To analyze the complete ApaLI restriction endonuclease gene, an inverse PCR amplification technique was used, which is rapid and simple to amplify the region adjacent to a known sequence. In addition, a combination of the inverse PCR and the direct sequencing enables to analyze without cloning. These procedures were effective in analyses of toxic genes, especially including endonucleases.

In advance of the inverse PCR, Southern hybridization was done to obtain a suitable restriction fragment as a template. In this experiment, the BamHI–HindIII fragment of pAPAL1.0 was used as a probe (Fig. 1A).

When the total DNA of A. pasteurianus IFO 13753 was treated with ClaI, the resulting 2.7-kb ClaI fragment was hybridized with the above-mentioned probe. The 2.7-kb fragment was suitable as a template for the inverse PCR, because the R.ApaLI gene corresponded to 1.2 kb

---

**Fig. 3.** Comparison of ApaLI Methylase with Cytosine-5 Methytransferase. The amino acid sequences of the methylases were aligned by the program Clustal W. The conserved regions (Regions I through X) are indicated by broken lines on the amino acids sequences. The asterisks under the sequences indicate identical amino acids.

---

NII-Electronic Library Service
calculated from the molecular mass (42,000 daltons) of \textit{R. Apa1}.

An upstream PCR primer was designed as 5'-ATGACT-ACACGGCAACGACTCAGC-3' based on the DNA sequence (positions 2869 to 2892) at the \textit{Nt} terminator of \textit{R. Apa1}. A downstream PCR primer was 5'-CTATCGAA-TGAATACAATCTCCTC-3', complementary to the DNA sequence (positions 2117 to 2139). The DNA sequences and loci of both primers are shown in Fig. 1B.

The total DNA of \textit{A. pasteurianus IFO 13753} was digested completely with \textit{ClaI}. The 2.7-kb fragment obtained was recovered by agarose gel electrophoresis and circularized by ligation. The inverse PCR was done using the circular 2.7-kb fragment as a template. As shown in Fig. 1B, a 2.0-kb PCR product was obtained by the inverse PCR amplification. This size agreed well with subtracting the distance (0.7 kb) between the two primers from the total (2.7 kb) of the circular fragment.

The sequencing of the PCR product showed that the \textit{R. Apa1} gene was 1125 bp long (coding for positions 2869 to 3993), corresponding to a protein of 375 amino acids with a calculated molecular mass of 42,143 daltons (Fig. 2). The predicted molecular mass agreed well with the value for \textit{R. Apa1} estimated by SDS-PAGE. A potential SD-like sequence (GGAT) was found ten bp upstream from the putative start ATG codon at position 2869. Putative promoter-like structures were observed at -10 (AATGAG, positions 2800 to 2805) and -35 (TTGTTC, positions 2778 to 2783) regions.

From the experiments mentioned above, the two genes encoding \textit{M. Apa1} and \textit{R. Apa1}, i.e., the \textit{Apa1} restriction-modification system, were in the same orientation, neighboring one another, and separated by only two bp, indicating that the stop (TAA) and the translational start (ATG) codons share A.

**Comparison of \textit{Apa1} restriction-modification system**

The amino acid sequences of the \textit{Apa1} proteins, viz., \textit{M. Apa1} and \textit{R. Apa1}, were compared with other methy- ylases and endonucleases. As expected, no similarity was detected between the \textit{R. Apa1} and the \textit{M. Apa1} sequences. No sequence similarity could be likewise observed between \textit{R. Apa1} and other endonucleases.

Comparison of the \textit{M. Apa1} sequence with those of cytosome-5 methylases showed significant similarities. The alignments of the deduced amino acid sequences of \textit{M. Apa1} and other methylases showed the ten conserved regions (Regions I through X) that are characteristic of cytosome-5 methylases\(^9,10\) (Fig. 3). In addition, the variable region of \textit{M. Apa1} (between Regions VIII and IX) had a moderate similarity to those of \textit{M. p111}\(^{11}\) and \textit{M. H21}\(^{12}\).

**Discussion**

Since our findings of a new restriction endonuclease, \textit{Apa1}, which recognizes GTGCAC and cuts between \textit{G} and \textit{T}, in \textit{A. pasteurianus IFO 13753},\(^{1,2}\) eleven isoschizomers, \textit{viz.}, \textit{Aqgl}, \textit{Ahi441}, \textit{VneI}, \textit{SnoI}, and so on, have been reported.\(^{1,3}\) However, there are no reports concerning cloning of their restriction-modification systems.

The two genes coding for endonuclease and methylase in the type II restriction-modification system are linked closely, but their arrangements are varied.\(^{14}\) The two genes in the \textit{Apa1} restriction-modification system are in the same orientation with the methylase gene preceding, and separated by two bp. This arrangement is similar to those of \textit{M. NgoMl}\(^{15}\) and \textit{PaeR71}\(^{16}\) systems but separated by three bp. The \textit{AcrI} system was separated by two bp but in a reverse order on two genes.\(^{17}\)

\textit{M. Apa1} showed the ten conserved regions (Regions I through X) which are characteristic of cytosome-5 methylases.\(^9,10\) The results obtained here suggested that the \textit{Apa1} methylase is classified in the cytosome-5 methylase family. The variable region of \textit{M. Apa1} (between regions VIII and IX) had a moderate similarity to those of \textit{M. p111} and \textit{M. H21}. The \textit{M. p111} and \textit{M. H21} are multispecific methylases of \textit{Bacillus subtilis} phase \textit{p11} and \textit{Bacillus amyloliquefaciens} phase \textit{H2}, respectively. The recognition sequence of \textit{M. Apa1} (GTGCAC) was included in the multispecific recognition sequence (GDGCHC; \textit{D}, not \textit{C}; \textit{H}, not \textit{G}) of the two phage methylases. It seems likely that the conserved amino acids in the variable region are responsible for the sequence specificities in the DNA recognition. In this study, the only methylase gene, viz., the \textit{M. Apa1} gene was cloned at a single step. However, the \textit{R. Apa1} gene was not cloned. In this case, the inverse PCR technique was very effective for identifying the nucleotide sequence which had not been cloned as yet.

Acknowledgements. We express our thanks to Dr. Kimio Sugiyama of this Department and Dr. Takamitsu Yorifuji, Faculty of Agriculture, Shimshu University, Minamininowa, Nagano, Japan, for their kind guidances and encouragements.

**References**