The ApI I Restriction-Modification System in an Edible Cyanobacterium, Arthrospira (Spirulina) platensis NIES-39, Recognizes the Nucleotide Sequence 5’-CTGCAG-3’

Hideaki Shiraishi¹ and Yosuke Tabuse*¹

Graduate School of Biostudies, Kyoto University, Yoshida-Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan

Received November 30, 2012; Accepted January 18, 2013; Online Publication, April 7, 2013

[doi:10.1271/bbb.120919]

The degradation of foreign DNAs by restriction enzymes in an edible cyanobacterium, Arthrospira platensis, is a potential barrier for gene-transfer experiments in this economically valuable organism. We overproduced in Escherichia coli the proteins involved in a putative restriction-modification system of A. platensis NIES-39. The protein produced from the putative type II restriction enzyme gene NIES39 K04640 exhibited an endonuclease activity that cleaved DNA within the sequence 5’-CTGCAG-3’ between the A at the fifth position and the G at the sixth position. We designated this enzyme ApI. The protein from the adjacent gene NIES39 K04650, which encodes a putative DNA (cytosine-5')-methyltransferase, rendered DNA molecules resistant to ApI by modifying the C at the fourth position (but not the C at the first position) in the recognition sequence. This modification enzyme, M.ApI, should be useful for converting DNA molecules into ApI-resistant forms for use in gene-transfer experiments. A summary of restriction enzymes in various Arthrospira strains is also presented in this paper.

Key words: Arthrospira platensis; Spirulina; restriction enzyme; DNA methyltransferase; edible cyanobacteria

Arthrospira platensis, formerly known as Spirulina platensis, is an edible cyanobacterium that has traditionally been consumed as food by people of Chad and Niger along the shores of Lake Chad, which is characterized by high levels of carbonate/bicarbonate and high pH (for reviews of this organism, see refs. 1 and 2). Modern commercial cultivation of this filamentous cyanobacterium began in the 1970s, because it was revealed to contain high levels of proteins, vitamins, and minerals that make it suitable for human consumption. Thereafter, it has been cultivated and used worldwide as a source of food and food additives as well as animal and fish feed. As an edible photosynthetic microorganism that can be propagated outdoors as a monoalgal culture, it has the potential to be used in an economically advantageous manner for the production of vast amounts of nutrients and/or useful compounds. These efforts would be aided by genetic engineering of this organism.

One obstacle that limits the use of this cyanobacterium in genetic engineering is that all strains analyzed thus far contain multiple restriction enzymes that degrade foreign DNA molecules.¹⁻⁶ In accord with these observations, recent determinations of the nucleotide sequences of the whole genomes of A. platensis NIES-39, A. platensis C1 (PCC9438), and Arthrospira sp. PCC 8005 revealed that these strains also contain multiple genes encoding putative restriction endonucleases.⁷⁻¹³ Therefore, in order to apply bioengineering techniques to this species, it is important to characterize its restriction and modification (R/M) systems. The resulting knowledge will help provide means for evading the degradation of DNA molecules introduced into cells during gene-transfer experiments; specifically, it will allow design of DNA molecules devoid of susceptible sequences, as well as modification of DNA by cognate methylation enzymes to prevent recognition and cleavage at restriction sites.

Biochemical investigation of restriction enzymes in the 1980s and 1990s showed that Arthrospira strains contain different sets of restriction enzymes depending on the isolate.⁴⁻⁶ The restriction enzymes found in Arthrospira strains in that period were isoschizomers of nine restriction enzymes: BsiWI, Tth111I, HaeIII, PvuII, PvuII, HindIII, SnaBI, HgiCI, and HgiDI (Table 1). The activity of one to four members of these nine isoschizomers were detected in each strain. The results of recent whole-genome sequencing of A. platensis NIES-39, A. platensis C1 and Arthrospira sp. PCC 8005 indicate that the genomes of these strains contain many more putative restriction enzyme genes than expected from previous biochemical studies.⁷⁻¹³ As shown in Table 1, the genome of A. platensis NIES-39 contains genes encoding putative proteins having significant homology to BsiWI, SnaBI, HgiCI, and HgiDI, which correspond to four of the nine isoschizomers that had been detected in Arthrospira strains by biochemical studies. In the case of A. plantensis C1 and Arthrospira sp. PCC 8005, these contain putative genes for seven and six of these nine isoschizomers, respectively. In addition to these genes, genes for putative restriction enzymes that had not been found in previous biochemical studies were also found in the genome sequences. For example, in the case of A. platensis NIES-39, six
Table 1. Type II Restriction Enzymes in Arthrospira (Spirulina) Strains

<table>
<thead>
<tr>
<th>Straina</th>
<th>Origin</th>
<th>Type II restriction enzymesb</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>S. platensis siamense</td>
<td>Ethiopia</td>
<td>+ + +</td>
<td>4</td>
</tr>
<tr>
<td>S. platensis UTEX2340c</td>
<td>Chad</td>
<td>+ + +</td>
<td>5</td>
</tr>
<tr>
<td>S. maxima UTEX2342c</td>
<td>Chad</td>
<td>+ + +</td>
<td>5</td>
</tr>
<tr>
<td>Spirulina sp. Kc</td>
<td>Kenya</td>
<td>+ + +</td>
<td>6</td>
</tr>
<tr>
<td>S. platensis pacifica</td>
<td>San Diego</td>
<td>+ + +</td>
<td>5</td>
</tr>
<tr>
<td>Spirulina sp. M1c</td>
<td>Mexico</td>
<td>+ + +</td>
<td>7</td>
</tr>
<tr>
<td>Spirulina sp. M2c</td>
<td>Mexico</td>
<td>+ + +</td>
<td>8</td>
</tr>
<tr>
<td>A. platensis NIES-39d</td>
<td>Chad</td>
<td>+ + + + + +</td>
<td>6a</td>
</tr>
<tr>
<td>A. platensis C1d</td>
<td></td>
<td>+ + + + + +</td>
<td>8</td>
</tr>
<tr>
<td>Arthrospira sp. PCC 8005d</td>
<td>Indiaa</td>
<td>+ + + +</td>
<td>9</td>
</tr>
</tbody>
</table>

aNames of strains are shown as in the original articles.
bNames of isoschizomers are shown. Blanks in this column indicate that there is no evidence of the corresponding enzymes; they do not necessarily mean an absence of the enzymes.
cFor these strains, enzyme activities detected in cellular extracts are indicated by +.dAdditional genes for putative type II restriction enzymes were reported (Table 1). Thus, in developing bioengineering systems utilizing these strains, it is important to characterize these newly identified putative R/M systems as well as those identified by previous studies.

Here we report the characterization of one of the putative R/M systems in A. platensis NIES-39, consisting of a protein encoded by the NIES39_K04640 gene, which has homology to type II restriction endonucleases of the BsuI/PstI family, and a protein encoded by an adjacent gene, NIES39_K04650, which has homology to DNA (cytosine-5-)-methyltransferases. Although the restriction enzyme of this putative R/M system shows similarity to restriction enzymes in the BsuI/PstI family, the cognate modification enzyme is strikingly different from the enzymes of well-characterized R/M systems in this family. Whereas BsuI and PstI are accompanied by N-6 adenine-specific methyletransferases, this putative restriction enzyme is accompanied by a C-5 cytosine-specific methyltransferase. In addition, this methyltransferase is most similar, among modification enzymes whose recognition sequence is known, to a modification enzyme M.HpII/C of a typeIIS R/M system HpII (E-value: 1e-25; identity: 34%), which recognizes non-palindromic nucleotide sequence 5'-GGTGA-3'. This is completely different from the recognition sequence of known proteins in the BsuI/PstI family, 5'-CTGCAAG-3'. Also, in addition to the difference in the cognate modification enzyme, the amino acid sequence of the putative restriction enzyme encoded by the NIES39_K04640 gene is less conserved than the sequences of BsuI and PstI. Thus, this R/M system as a whole appears to be evolutionarily distant from well-characterized R/M systems in the BsuI/PstI family. It is thus of interest to determine whether this R/M system shows the same sequence specificity as other members of the BsuI/PstI family. We show here that the restriction enzyme of this R/M system is isoschizomeric to BsuI and PstI, and that the cognate modification enzyme recognizes a target sequence in common with M.BsuI and M.PstI, even though it belongs to a different class of DNA methyltransferases.

Results and Discussion

Expression in E. coli of the proteins for a putative R/M system from A. platensis NIES-39

In order to characterize the biochemical properties of the proteins encoded by the NIES39_K04640 and NIES39_K04650 genes (Fig. 1A), they were produced as recombinant proteins with a hexahistidine tag in Escherichia coli. They were then purified by means of a nickel-chelating column that binds the tag. The molecular masses of the recovered proteins were 40 kDa and 41 kDa respectively as estimated by mobility on an SDS-polyacrylamide gel (Fig. 1). These values were consistent with the expected molecular weights calculated from the amino acid sequences of the recombinant
proteins (39.3 kDa and 40.1 kDa for the recombinant proteins for NIES39 K04640 and NIES39 K04650 respectively).

The recombinant protein from the NIES39 K04640 gene had endonuclease activity

The recombinant protein from the NIES39 K04640 gene (K04640 protein) was assayed for endonuclease activity using the DNA of the λ phage as substrate. Since the optimal reaction conditions for this protein were unknown, enzymatic activity was first examined using several different buffers that had been used for the reactions of other restriction enzymes. Endonuclease activity of the K04640 protein was detected in many different buffers, but the most efficient cleavage was observed in a buffer containing 20 mM Tris–HCl (pH 8.5), 10 mM MgCl$_2$, 1 mM dithiothreitol, and 100 mM KCl (buffer K supplied by Takara Bio, Otsu, Japan). This buffer had a relatively basic pH compared with the others, whose pH values were 7.5–7.9. To determine whether the weakly alkaline pH of the buffer promoted efficient cleavage, we examined the activity of the K04640 protein under various pH conditions. Although the activity of the K04640 protein was detectable at pH 7.5, it was more active under basic conditions with optimal pH at about 9.5 (Fig. 2A and B). This optimal pH roughly corresponds to the conditions of the natural habitat of A. platensis. It is not conceivable, however, that the optimal pH conditions of this enzyme are directly related to normal conditions in healthy cells, since such alkaline conditions would impinge on many other enzymatic reactions in the cell. One possibility is that the high pH optimum of this enzyme was due to an uncommon event in which selection pressure was applied to unhealthy cells the intracellular pH conditions of which were temporarily affected by environmental conditions.

As Fig. 2A indicates, complete digestion of λ phage DNA by the K04640 protein resulted in the formation of DNA fragments of discrete lengths. This indicates that the K04640 protein cleaved DNA at specific positions.

We also examined the effects of monovalent cations (Fig. 2C). Endonuclease activity was enhanced in the presence of moderate concentrations (about 50 mM) of K$^+$ and of Na$^+$. In contrast, NH$_4^+$ appeared to have a negative effect under these conditions. In the presence of a high concentration of NH$_4$Cl, the reaction buffer was slightly acidified (pH 8.04 at 200 mM NH$_4$Cl), but the reduction in enzyme activity was greater than expected in view of the change of pH.

K04640 protein, or ApII, cleaved DNA at the 5'-CTGCAG-3' sequences

An examination of the DNA fragments generated by complete digestion of the λ phage DNA (Fig. 2A) suggested that their lengths were the same as those generated by PsI digestion. This was further confirmed by performing electrophoresis of PsI and K04640 protein digests of λ phage DNA on adjacent lanes on an agarose gel. As shown in Fig. 3A, PsI and the K04640 protein produced the same set of digestion products, indicating that the K04640 protein had the same sequence specificity as PsI, whose recognition sequence was 5'-CTGCAG-3'. In order to determine the exact cleavage site in the recognition sequence, a $^{32}$P-labeled double-stranded DNA having the recognition sequence was synthesized and cleaved with the K04640 protein, and the cleavage site was determined.
by electrophoresis on a sequencing gel. As shown in Fig. 3B, the K04640 protein cut the DNA at the phosphodiester bond between A and G in the recognition sequence $5'$-CTGCAG-3' to cleave it into $5'$-CTGCA and G-3'. Thus the cleavage site was also the same as for PstI.\(^\text{12}\) We designated this PstI isoschizomer of A. platensis NIES39 ApI, following the nomenclature for restriction enzymes.\(^\text{13}\)

ApI belongs to the BsuBI/PstI restriction endonuclease family (Pfam accession no. PF06616),\(^\text{14}\) sharing 37 and 30% identical amino acids with BsuBI and PstI respectively. The numbers of the identical amino acids are, however, lower than that between BsuBI and PstI, which share 46% of their amino acids.\(^\text{10}\) Nonetheless, our study indicates that ApI had the same recognition sequence and cleavage site as BsuBI and PstI.

**Fig. 2.** Effects of pH and Monovalent Cation on the Activity of the K04640 Protein.

A. Effect of pH on the reaction by the K04640 protein. DNA of λ phage was incubated at 37°C for 2h with the K04640 protein under various pH conditions and was electrophoresed on a 1% agarose gel. Tris–HCl and Glycine-KOH buffers with the indicated pH values were used in the reactions. Also electrophoresed were λ phage DNA that was not treated with enzymes (No enzyme), λ phage DNA digested with HindIII (HindIII), and λ phage DNA that was completely digested with the K04640 protein by incubation for 16h (Complete digestion). B. Relative activity of the K04640 protein under various pH conditions. After performing the experiments described in A, the intensity of the bands shorter than 7 kb was quantitated and divided by the intensity of the whole bands in the same lane. The values obtained gave a rough estimate of enzyme activity under the conditions imposed. Relative values relative to a maximal value were then calculated and plotted as relative activity. Reactions in the Tris–HCl and Glycine-NaOH buffers are shown by open circles and closed circles respectively.

**Fig. 3.** Recognition Sequence and Cleavage Site of the K04640 Protein.

A. Digestion of λ phage DNA by PstI and the K04640 protein. The λ phage DNA was completely digested with PstI or the K04640 protein and electrophoresed on a 1% agarose gel. Intact λ phage DNA was also electrophoresed (No enzyme). B. Cleavage site by the K04640 protein. A double-stranded DNA that was radiolabeled with $^{32}$P at the 5' end of one strand was cleaved with the K04640 protein. The digestion products was electrophoresed on an 20% polyacrylamide gel containing 8M urea, along with a sequencing ladder (lanes G, A, T, and C) prepared by end-labeling a series of synthetic oligonucleotides with $^{32}$P.

The recombinant protein from the NIES39_K04650 gene had modification activity that rendered DNA resistant to ApI.

The protein encoded by the NIES39_K04650 gene, which resides adjacent to the ApI gene (NIES39_K04640), showed high sequence similarity to DNA (cytosine-5-)methyltransferases, including the signature sequence of the active site of C-5 cytosine-specific DNA methylases (PROSITE accession no. PS00094) at amino acid positions 85-97.\(^\text{15}\) To determine whether this protein renders DNA molecules resistant to ApI, the modification activity of the recombinant protein (K04650) from the NIES39_K04650 gene was assayed using λ phage DNA as substrate. As shown in Fig. 4A, the DNA pretreated with the K04650 protein in the presence of a methyl donor, S-adenosyl methionine, became resistant to both ApI and its isoschizomer, PstI. The modified DNA was susceptible to HindIII and EcoRI, indicating that the modification was specific to the restriction sites for ApI/PstI.

**Determination of the modification site by the K04650 protein, M.ApI**

The recognition sequence of ApI, $5'$-CTGCAG-3', contains two cytosine residues, at the first and fourth positions in the sequence. Since the K04650 protein belongs to the family of DNA (cytosine-5-)methyltransferases, it is most likely that the C5 position of one or both of these cytosines is the target of this protein. In order to examine the modification sites by the K04650 protein, a DNA molecule that contained a set of restriction sites, shown in Fig. 4B, was prepared. It contained a PvuII site that partly overlapped an ApI site. The cognate modification enzyme of PvuII is a...
This DNA molecule also contains a restriction site for the central four-nucleotide sequence. This positively indicates that the DNA has modifications. If the modified DNA is resistant to this enzyme, the position of C at the fourth position in (cytosine-N4-)-methyl transferase that modifies the N4-methyl modification at that position as well as to the N4-methylmodified DNA molecule modified by the K04650 protein, or M.AplI, modified only the C residue at the fourth position in the recognition sequence of M.AplI and works as a control. As shown in the bottom panel of Fig. 4B, the DNA molecule modified by the K04650 protein was insensitive to digestion by HindIII. Hence the K04560 protein modified the nucleotides in the central region of the ApII site, but not the C residue at the first position in the ApII site. This result indicates that the K04650 protein, or M.AplI, modified only the C residue at the fourth position in the recognition sequence of ApII.

In contrast to the similarity of restriction enzyme ApII to BstBI/PstI, the cognate modification enzyme M.AplI shares similarity with neither M.BsuBI nor M.PsrI, because M.AplI is an enzyme of the DNA (cytosine-5-)methyltransferase family, whereas M.BsuBI and M.PsrI are DNA (adenine-N6-)methyltransferases as mentioned above in the introduction. This striking difference indicates that M.AplI evolved independently of the M.BsuBI and M.PsrI to recognize the same recognition sequence as these modification enzymes.

We found in this study that the proteins produced by the NIES39_K04640 and NIES39_K04650 genes of A. platensis NIES-39 had restriction and modification activities respectively that recognize nucleotide sequence 5'–CTGCAG–3'. In the case of another filamentous cyanobacterium, Anabaena sp. PCC 7120, characterization of its R/M systems helped establish an efficient gene-transfer procedure. It has three type II restriction enzymes that are isoschizomeric to AvaI, AvaII, and AvaIII, and methylation of DNA by all three cognate modification enzymes resulted in a drastic improvement in gene-transfer efficiency. Although further study of other R/M systems is apparently needed for A. platensis NIES-39, the information on the recognition sequence of ApII and production of the recombinant M.AplI presented here should be helpful in establishing an efficient gene-transfer procedure for this strain in the future, as is the case for Anabaena sp. PCC 7120.

**Experimental**

**Bacterial strains.** A. platensis NIES-39 was obtained from the Microbial Culture Collection at the National Institute for Environmental Studies, Japan (NIES Collection). E. coli KRX, harboring a T7 RNA polymerase gene that is controlled by the rbad promoter, was purchased from Promega KK (Tokyo).

**Cell culture.** A. platensis NIES-39 was cultured at 30°C under a photoperiod of 12 h light-12 h dark in a medium slightly modified from the medium described previously. In preparing 1 L of the medium, 16.8 g of NaHCO3, 0.5 g of K2HPO4, 2.5 g of NaNO3, and 1 g of K2SO4 were dissolved in 970 mL of deionized water, and this was sterilized by autoclaving. After cooling, 20 mL of 50× macroelements and 1 mL of 1000× microelements that had been individually prepared and autoclaved were added. The 50× macroelements contained, in 100 mL 5 g of NaCl, 0.4 g of Na2EDTA, 50 mg of FeSO4–7H2O, 1 g of MgSO4–7H2O, 0.2 g of CuCl2–2H2O, and 14.3 mg of H3BO3, and the 1000× microelements contained, in 100 mL 218 mg of MnSO4–H2O, 22.2 mg of ZnSO4–7H2O, 7.9 mg of CuSO4–5H2O, and 2.1 mg of Na2MoO4–2H2O.

**Preparation of genomic DNA.** Genomic DNA of A. platensis NIES-39 was prepared by combining the methods for isolating nucleic acids from cyanobacteria and for enriching DNA from a mixture of DNA and RNA by differential precipitation with
isopropanol. In brief, exponential-phase cells of *A. platensis* NIES-39 (wet weight, 1 g) were washed with 5 mL of a buffer containing 50 mM Tris–HCl (pH 8.0), 50 mM NaCl, and 5 mM EDTA. The cells were suspended in 2 mL of a saturated solution of NaI and incubated for 20 min at 37 °C. After washing twice with distilled water, cells were suspended in 5 mL of 50 mM Tris–Cl (pH 8.5), 50 mM NaCl, 5 mM EDTA, and then incubated with lysozyme (1 mg/mL) for 45 min at 37 °C. Sodium dodecyl sulfate and proteinase K were added to make final concentrations of 1% (w/v) and 50 μg/mL, respectively. After incubation for 1 h at 37 °C, the solution was repeatedly extracted with phenol until no interface was visible. To 1 volume of purified nucleic acid solution, 2/3 volume of 5 M NaCl and 1 volume of isopropanol were added and mixed in. The precipitated DNA was recovered by centrifugation and dissolved in TE (10 mM Tris–HCl, pH 8.0). Precipitation was repeated once more. After precipitation, a clump of DNA that appeared in the supernatant was recovered by entangling it around the plastic tip of a micropipette and transferred into a 70% ethanol solution. The DNA, recovered from the 70% ethanol solution, was dissolved in TE (10 mM Tris–HCl, pH 8.0) and stored at −25 °C.

Overproduction of recombinant proteins in *E. coli*. The coding regions for the NIES39*K04640* and NIES39*K04650* genes were individually amplified from the genomic DNA of *A. platensis* NIES-39 by PCR with pETtool high-fidelity DNA polymerase (Bio-Rad, Hercules, CA) following the manufacturer’s instructions. The primers used in PCR were designed to introduce *NcoI* and *Xhol* sites to the N-terminal side and the C-terminal side respectively of the protein-coding sequences retrieved from the CyanoBase. After digestion of the amplified DNA fragments with *NcoI* and *Xhol*, they were inserted between the *NcoI* and *Xhol* sites of a T7 promoter-based expression vector, pET22b(+) (Merck Millipore, Billerica, MA). The nucleotide sequences of the constructed plasmids were determined to confirm that the cloned DNA fragments had the same nucleotide sequences as the published ones. The plasmids, pET-K04640 and pET-K04650, thus constructed were introduced individually into *E. coli* KRKX. *E. coli* KRX cells harboring the expression plasmids were grown at 37 °C in LB medium supplemented with 0.1% (w/v) glucose and kanamycin (25 μg/mL). Production of recombinant proteins was then induced by adding rhamnose and isopropl β-d-1-thiogalactopyranoside to final concentrations of 0.1% (w/v) and 1 mM respectively when the optical density of the cultures was 0.4 at 600 nm. The cultures were harvested 2 h after induction.

Preparation of recombinant proteins. Unless otherwise stated, all of the following steps were carried out at 0–4 °C. *E. coli* cells containing recombinant proteins were disrupted with a sonicator (Ohkate Works, Tokyo, Model 5202PZT) in 20 mM sodium phosphate (pH 7.8), 0.5 mM NaCl. After centrifugation for 10 min at 10,000 g, supernatants were applied onto Ni-NTA agarose columns (Quagen, Venlo, Netherlands). After washing of the columns with the same buffer, recombinant proteins were eluted by stepwise washing with 20 mM sodium phosphate (pH 6.8), 0.5 mM NaCl containing 0, 10, 50, 100, and 150 mM imidazole. Enzymatically active proteins for NIES39*K04640* and NIES39*K04650* eluted from the columns with buffers containing 50 mM and 50 mM imidazole respectively. The protein fractions were dialyzed against 10 mM Tris–HCl (pH 7.5), 50 mM KCl, 0.5 mM dithiothreitol for 16 h and then against the same buffer containing 50% (v/v) glycerol for 8 h. The protein solutions were stored at −25 °C. Protein concentrations were determined by the Bradford method, with bovine γ-globulin as standard.

Enzymatic reaction with the K04640 and K04650 proteins. Unless otherwise stated, reactions by the K04640 protein were carried out at 37 °C for 2 h in 10 μL of a reaction mixture containing the K04640 protein (1 μg), 20 mM Tris–HCl (pH 8.5), 10 mM MgCl₂, 50 mM KCl, and 1 mM DTT, using the DNA (0.5 μg) of bacteriophage λ18S57am7 (Takara Bio) as substrate. In the experiments shown in Fig. 2, buffer or salt was changed from this solution as indicated. For complete digestion of DNA with the K04640 protein, incubation was prolonged to 16 h. The reaction was terminated by adding an equal volume of 20 mM Tris–HCl (pH 7.5), 12 mM EDTA, 0.1% SDS, 5% glycerol, and 0.005% bromphenol blue. The reaction products were electrophoresed on 1% agarose gels, followed by staining with ethidium bromide. Stained DNA was visualized and analyzed with a LAS-3000/Univ mini-lumino-image analyzer (FujiFilm, Tokyo).

Reactions by the K04650 protein were performed at 37 °C for 2 h in a mixture containing the protein (80 ng/μL), DNA substrate (50–70 ng/μL), 50 mM Tris–HCl (pH 8.0), 50 mM NaCl, 10 mM EDTA, and 133 μM 5'-adenosylmethionine. After recovery of DNA by phenol extraction and ethanol precipitation, it was digested with the K04640 protein or other restriction enzymes to analyze modifications in it. The digestion products were analyzed by electrophoresis on either 1% gels of UltraPure Agarose (Invitrogen) or 4% gels of NuSieve 3:1 Agarose (Takara Bio) depending on the size of the substrate.

Determination of cleavage site by the K04640 protein. Five picomoles of a synthetic oligonucleotide, 5'-AGCTGATGCGCATGCA-3'3' containing the recognition sequence of the K04640 protein as underlined in the nucleotide sequence, was labeled with γ-32P]ATP (PerkinElmer, Waltham, MA) and T4 polynucleotide kinase (Takara Bio). After the labeling reaction, the enzyme was inactivated by heating for 15 min at 70 °C, and the labeled oligonucleotide was mixed with 10 pmol of a complementary oligonucleotide. The resulting double-stranded DNA substrate was digested with the K04640 protein and electrophoresed on a 20% polyacrylamide gel (1/20 bisacrylamide) gel containing 8.3 M urea, along with a sequencing ladder prepared by mixing 5'-32P]-labeled synthetic oligonucleotides that had the same nucleotide sequence as the end region overhangs compatible with the substrate, but had successive deletions in the 3'-end region over the recognition sequence of the K04640 protein. Labeled DNA molecules on the gel were visualized and analyzed with image analyzer FLA-7000 (FujiFilm).

Preparation of a DNA fragment to determine the modification site by the K04650 protein. A double-stranded oligonucleotide with overlaps compatible with the *HindIII* site and the *PstI* site was prepared by annealing two oligonucleotides having sequences 5'-AGCTGATGCGCATGCA-3' and 5'-GCTGCCGATCACA-3'. This molecule was inserted between the *HindIII* and the *PstI* site of pUC118 to construct pUC118-Pv. This plasmid had a newly introduced *PvuII* site that partly overlapped a *PstI* site (see Fig. 4B for the nucleotide sequence of the relevant region). A DNA fragment including the multiple cloning site sequence region of pUC118-Pv was prepared by PCR with two primers that hybridized to the pUC118-Pv in the regions flanking the multiple cloning site sequence. The nucleotide sequences of the PCR primers used were 5'-GGCTCG-TATGTGTGTTGGAAAT-3' and 5'-GACAAGCGGATTAGTTGG-AAA-3', which were designed to hybridize to pUC plasmids in the reverse and universal orientations respectively. The PCR product was purified by electrophoresis on a 1.5% gel of low melting point agarose (SeaPlaque GTG agarose, Takara Bio).

BLAST searches. BLASTP searches for sequences similar to the amino acid sequences of R/M system proteins were carried out by CyanoBase Similarity Search (http://genome.microbedb.jp/blast/blast_search/cyanobasegenes) for *A. platensis* NIES-39 and NCBI/BLAST (http://blast.ncbi.nlm.nih.gov) for *A. platensis* C1 (taxid, 455949) and *A. platensis* sp. PPC 8005 (taxid, 376219). The amino acid sequences of the following proteins were used as query sequences in the BLASTP search (shown in parentheses are UniProt accession numbers): BoWI (E5Q8V3), Zh1111 (G33XCT7), HaIII (O68584), AqJ22M2 (Q7K621X), ProV (P23657), HindIII (P43870), SmdA (Q6SA25), HgCI (P25258), HgDI (P25459), Pstl (P00060), Eca2903J (Q46944), BsuBfI (P33355), NplQ (Q2ZHJP2), Ekorn (P42402), BsuW1 (Q6UQ65), Aqel (Q9K9HV6), BpUI10 α subunit (O52851), BpUI10 β subunit (O52852), BbCI subunit 1 (Q5D6Y5), BbCI subunit 2 (Q5D6Y4), and MjAfz22M2 (Q762K2). Sequences whose E-values were <1×10⁻13 were judged to be homologs. When the E-value of the most similar sequence was >0.01, it was judged that the homolog was absent from the genome. BLASTP searches against the UniProtKB/SwissProt database were also carried out by NCBI/BLAST using the amino acid sequences of the putative proteins encoded by the NIES39*K04640* gene and the NIES39*K04650* gene as query sequences to detect similar proteins.
Acknowledgment

We thank Dr. Tan Inoue for critical reading of the manuscript. This work was supported financially in part by the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References