Cloning, nucleotide sequence, and expression of the *HincII* restriction-modification system

Hiroyuki Ito, Atsuko Sadaoka, Hirokazu Kotani, Nobutsugu Hiraoka and Teruya Nakamura
Bioproducts Development Center, Takara Shuzo Co., Ltd., Seta 3-4-1, Otsu, Shiga 520-21, Japan

Received March 9, 1990; Revised and Accepted May 30, 1990

EMBL accession no. X52124

ABSTRACT

Two genes, coding for the *HincII* from *Haemophilus influenzae* Rc restriction-modification system, were cloned and expressed in *Escherichia coli* RR1. Their DNA sequences were determined. The *HincII* methylase (*M.HincII*) gene was 1,506 base pairs (bp) long, corresponding to a protein of 502 amino acid residues ($M_r = 55,330$). The *HincII* endonuclease (*R.HincII*) gene was 774 bp long, corresponding to a protein of 258 amino acid residues ($M_r = 28,490$). The amino acid residues predicted from the *R.HincII* and the N-terminal amino acid sequence of the enzyme found by analysis were identical. These methylase and endonuclease genes overlapped by 1 bp on the *H. influenzae* Rc chromosomal DNA. The clone, named *E. coli* RR1-*Hinc*, overproduced *R.HincII*. The *R.HincII* activity of this clone was 1,000-fold that from *H. influenzae* Rc. The amino acid sequence of *M.HincII* was compared with the sequences of four other adenine-specific type II methylases. Important homology was found between the *M.HincII* and these other methylases.

INTRODUCTION

Various type II restriction-modification systems have been cloned, and their organization and nucleotide sequences have been analyzed (1–14). One problem in the cloning of restriction-modification systems is the choice of the bacterial host strain. *Escherichia coli* has several mechanisms for identifying foreign DNA and destroying it. Of these mechanisms, there are four restriction systems, called the modified cytosine restriction (*Mcr*) systems A, B, and C (15–19) and the methylated adenine recognition and restriction (*Mrr*) system (20). The *HincII* system recognizes double-stranded DNA at the palindromic sequence 5'-GTPyPuAC-3'. The *Mrr* system restricts the consensus sequence (G^mAC or C^mAG). Therefore, the *HincII* system can be cloned only into bacterial host strains that lack *mrr*. Two such strains in common use are *E. coli* HB101 and RR1.

Although consensus sequence has been not found among many restriction endonucleases, serious homology were recently found between the restriction endonuclease *EcoRI* and its isoschizomer *RsrI* (21). In contrast consensus sequences have already been found among many methylases (22–27). By the identification of consensus sequences between isoschizomers, it may be possible to pinpoint active sites and the like.

In this paper, we report on the cloning, characterization, and

nucleotide sequences of the genes encoding the *HincII* restriction-modification system. The amino acid sequence deduced for the *HincII* methylase (*M.HincII*) was compared with that of other type II methylases.

MATERIALS AND METHODS

Bacterial strains, plasmids, and phages

Haemophilus influenzae Rc (28) was obtained from the Institute for Chemical Research, Kyoto University. *E. coli* strains HB101 [F^- , *hsdS20*(rB⁻, mB⁻), *supE44*, *ara14*, λ^- , *galK2*, *lacY1*, *proA2*, *rpsL20*, *xyl-5*, *mtl-1*, *recA13*, (29)], JM109 [*recA1*, *endA1*, *gyrA96*, *thi*, *hsdR17*(rK⁻, mK⁻), *supE44*, *relA1*, λ^- , Δ (*lac-proAB*), *F'*, *traD36*, *proAB*, *lacI^qZ Δ M15*, (30)], and RR1 [same as HB101 except *recA*⁺, (31)], and plasmid pBR322 [*Tc^r*, *Ap^r*, (31,32)], pUC19 [*Ap^r*, (30)], pACYC184 [*Cm^r*, *Tc^r*, (33)], pPL-lambda [*Ap^r*, (34)] (which was purchased from Pharmacia), and pNT203 [*Tc^r*, (35)] were the host-vector systems used for cloning and expression. M13mp18 and mp19 (30) were used for DNA sequencing. *E. coli* competent cells were produced by the procedure of Hanahan (36).

Enzymes and chemicals

Restriction endonucleases and 7-deaza sequencing, kilo-sequence deletion, Megalabel, and ligation kits were products of Takara Shuzo Co., Ltd., and were used according to the manufacturer's instructions. [α -³²P]dCTP (14.7 TBq/mmol) and [γ -³²P]ATP (185 TBq/mmol) were purchased from Amersham International.

Phage and DNA preparations

A stock of λ virulent phage (λ gt10c) and DNA from this phage were prepared from plate lysates (37). Small-scale preparations of *E. coli* plasmid DNA were obtained by a modification of the method of Birnboim and Doly (38). *H. influenzae* Rc chromosomal DNA was purified by the procedure of Thomas et al. (39).

Construction of the *H. influenzae* Rc genomic library

A partial *Sau3AI* digest of 5 μ g of *H. influenzae* Rc chromosomal DNA was separated by agarose gel electrophoresis, and 3–7 kilobase pairs (kb) were recovered from the gel by adsorption to DE81 paper. These *Sau3AI* fragments (0.5 μ g) were ligated to *Bam*HI-digested and dephosphorylated pBR322 with the use of a ligation kit, and *E. coli* RR1 was transformed with the resulting plasmids.

Selection of clones expressing the *HincII* modification system

The transformants obtained were inoculated into and grown to saturation in 20 ml of L-broth [1 g of tryptone, 0.5 g of yeast extract, and 0.5 g of NaCl in a volume of 100 ml (pH 7.5)] (37) that contained 100 µg/ml ampicillin. Plasmid DNAs were prepared by the alkaline method (38). Then 5 µg of purified plasmid DNA was digested for 1 hr at 37°C with 100 units of *HincII* endonuclease (*R.HincII*), followed by dephosphorylation for 1 hr at 55°C with 1.5 units of bacterial alkaline phosphatase. *E. coli* RR1 was transformed with the digested plasmids. The transformants were reinoculated and grown to saturation. These steps from the preparation of plasmid DNA to transformation were done again once, and the transformants obtained were plated onto L-agarose plates containing 100 µg/ml ampicillin. Plasmids were prepared from the surviving bacterial cells, and then analyzed.

Assay of enzyme activities

The *R.HincII* activity *in vivo* was estimated from the plating efficiency of λ virulent phage. *In vitro*, this activity was assayed by the incubation at 37°C of 25 µl of a reaction mixture containing 20 mM Tris-HCl (pH 7.5), 10 mM MgCl₂, 100 mM NaCl, 0.01% bovine serum albumin, and 0.5 µg of λ phage DNA. The *M.HincII* activity *in vivo* was estimated from non-digested plasmids, chromosomal DNA, and λ virulent phage DNA by *R.HincII*. *In vitro*, this activity was assayed by incubation for 1 hr at 37°C of 10 µl of reaction mixture containing 50 mM Tris-HCl (pH 7.5), 5 mM 2-mercaptoethanol, 10 mM EDTA, 1 µg of λ phage DNA, 80 µM S-adenosylmethionine, and the enzyme at various concentrations. Then 40 µl of a reaction mixture containing *R.HincII* was added, and the new mixture was incubated at 37°C until digestion was complete. The *R.HincII* and *M.HincII* activities of various fragments were assayed as follows. Fragments were prepared from the plasmids pAH1 or pAH2, subcloned into appropriate sites of the pBR322 vector, and then introduced into *E. coli* RR1. The activities of the two enzymes from the clones were assayed *in vivo* and *in vitro* as described above.

Analysis of N-terminal amino acid sequence of *R.HincII*

R.HincII was partially purified from *H. influenzae* Rc cells. The sample (30,000 U) was purified on sodium dodecyl sulfate (SDS) polyacrylamide gel (12.5%) and blotted from the gel onto a polyvinylidene difluoride (PVDF) membrane (Millipore Corp.) by the method of Matsudaira (40). This membrane was stained with Coomassie brilliant blue R-250 and destained with 0.7% acetic acid. The band containing the *R.HincII* [*M_r* = approx. 29,000, (41)] was cut out from the membrane and then put on a protein sequencer. The N-terminal sequence was determined with an Applied Biosystems 477A gas-phase protein sequencer. The phenylthiohydantoin (PTH) derivatives of the amino acids were identified with an Applied Biosystems 120A PTH analyzer connected to the sequencer.

Analysis of the nucleotide sequences

Nucleotide sequences were determined by the dideoxy chain termination method (37,42), except that 7-deaza-dGTP was used instead of dGTP. Various fragments (0.2–0.4 kb) of the *HincII* restriction-modification genes were prepared from the plasmid pAH2 and cloned into appropriate sites of the vectors M13mp18 and mp19 (30) to provide templates; the sequences were deduced from the results for both strands. The DNA sequences and the predicted amino acid sequences were analyzed with a DNASIS

computer (Hitachi Soft Engineering). The amino acid sequences were studied with use of the Gene/Protein sequence database of DNASIS (Gen Bank R60.0, compiled by National Institutes of Health, USA, and EMBL R19.0, compiled by European Molecular Biology Laboratory).

Southern hybridization

To isolate the *HincII* restriction-modification system, Southern hybridization was carried out (37,43). After *H. influenzae* Rc chromosomal DNA (each 2 µg) was digested with various restriction endonucleases and put on 0.7% agarose gel, the digested DNA was transferred onto a nitrocellulose membrane (Nitro Plus 2000, MSI Corp.) from the agarose gel by use of a Vacublot (American Bionetics Corp.). The transferred membrane was denatured and neutralized. The membrane was treated with a ³²P-labeled probe (0.5 µg) in hybridization buffer [6×SSC solution (1×SSC is 0.15 M NaCl plus 0.015 M sodium citrate), 5×Denhardt solution, 0.1% SDS, and 0.1 mg/ml denatured salmon DNA] at 50°C, rinsed with 6×SSC solution, and used to expose to radiographic film. The *Sau3AI* probe (1.0 kb, 0.5 µg) was prepared from plasmid pAH1 and then phosphorylated with 0.37 MBq of [γ-³²P]ATP with use of a Megalabel kit.

Purification of *R.HincII*

About 126 g of *E. coli* RR1-*Hinc* cells was suspended in 250 ml of 20 mM potassium phosphate buffer (pH 7.5) containing 10 mM 2-mercaptoethanol and 0.15% Triton X-100, and sonicated. The cell debris was removed by centrifugation for 1 hr at 100,000×g. To the supernatant was added a 10% solution of streptomycin to the final concentration of 2%. After removal of the precipitate by centrifugation, solid ammonium sulfate was added to 30% saturation. After removal of the precipitate by centrifugation, solid ammonium sulfate was added to 75% saturation. The resulting precipitate was collected by centrifugation, dissolved in 20 mM potassium phosphate buffer (pH 7.5) containing 10 mM 2-mercaptoethanol, 5% glycerol, and 0.15% Triton X-100 (buffer A) plus 0.15 M KCl and dialyzed against the same buffer. The dialysate was put onto a phosphocellulose column (4.6×12 cm; Whatman P-11) and chromatographed with 1,500 ml of a linear KCl gradient from 0.15 to 1.0 M in buffer A. The *R.HincII* activity was eluted at 0.3–0.4 M KCl. The active fraction was dialyzed against buffer A, put on a DEAE-cellulose column (2.5×15 cm; Whatman DE-52), and chromatographed with 400 ml of a linear KCl gradient from 0 to 1.0 M of KCl in buffer A. The peak of *R.HincII* activity, which was eluted at 0.25–0.35 M of KCl, was dialyzed against buffer A, put on a heparin-Sepharose column (1.2×17 cm; Pharmacia), and eluted with 200 ml of a linear gradient of KCl from 0–1.0 M in buffer A. The active fraction was concentrated with a Centriflo filter (Amicon Corp.) and filtered through a Sephadex G-100 column (2×70 cm; Pharmacia). The fraction with the peak of activity was added to the same volume of 50% ammonium sulfate and put on a phenyl-Sepharose column (1.2×17 cm; Pharmacia). After the column was washed with 10 mM potassium phosphate buffer (pH 7.5) containing 25% ammonium sulfate, the enzyme was eluted with a linear gradient of ammonium sulfate (25 to 0%) and a linear gradient of ethylene glycol (0 to 50%) at the same time. The active fractions, which were obtained at 45–50% ethylene glycol, were dialyzed against buffer A. The dialysate obtained was stored at 4°C until use and used as the purified *R.HincII* enzyme.

Figure 1. Restriction map of plasmid pAH1. (A): Plasmid pAH1, which contained only the *M.HincII* gene, was digested with appropriate restriction endonucleases. Agarose gel electrophoresis was used for analysis. A *Sau3AI* fragment (1 kb) was used as the probe for Southern hybridization. Ba-*Bam*HI; Af-*Afa*I; D-*Dra*I; X-*Xba*I; ER-*Eco*RI; Hd-*Hind*III; Pm-*Pma*CI; Sa-*Sau*3AI; Ap^r-resistance gene of ampicillin. (B): The enzyme activities of *R.HincII* (R) and *M.HincII* (M) fragments of various sizes were assayed as described in 'Materials and Methods'.

SDS-Polyacrylamide gel electrophoresis (PAGE)

For the expression of *R.HincII* in *E. coli* RR1-Hinc, the estimation of its molecular weight, or, the determination of the N-terminal amino acid sequence, SDS-PAGE was done by the method of Laemmli (44) with a 12.5–15.0% polyacrylamide slab gel.

RESULTS

Isolation of the gene coding for *M.HincII*

The selection of the gene coding for *M.HincII* was based on the resistance of self-modifying recombinant plasmids to digestion by *R.HincII* (45). A plasmid library of *H. influenzae* Rc was digested with an excess of *R.HincII* for the selection of clones expressing the *M.HincII* gene. About 1×10^4 colonies carrying plasmids were obtained.

A mixed population of surviving recombinant plasmids was prepared and then re-digested with *R.HincII*. The re-digested DNAs were dephosphorylated and used to transform *E. coli* RR1 cells. The recombinant plasmids of randomly selected surviving

clones were analyzed by digestion with *R.HincII*. Finally plasmids from a total of 11 clones were not digested by *R.HincII*. In general, the two genes of restriction-modification systems are in the same vicinity. The 11 clones, not affected by *R.HincII*, were analyzed for their resistance to infection by λ virulent phage. Unfortunately, none of the clones were resistant. But λ phage DNAs prepared from these infected plaques and the chromosomal DNAs from these 11 clones could not be digested by *R.HincII* (data not shown). Plasmids from these clones contained only the *M.HincII* gene, or, if they also contained the *R.HincII* gene, it was not expressed. The plasmid containing the longest insert of DNA was named pAH1 and analyzed by digestions with various restriction endonucleases (Fig. 1A). This plasmid contained about 3.5 kb of DNA from the *H. influenzae* Rc chromosome. Various fragments of the plasmid were prepared (Fig. 1B) and introduced into host cells; the resulting *R.HincII* and *M.HincII* activities were assayed as described in 'Materials and Methods'. The two activities were not detected together in any one fragment *in vitro*. As mentioned above, the plasmid pAH1 expressed only the *M.HincII* gene, to judge from the enzyme activities found *in vivo*.

Figure 2. Southern hybridization of *H. influenzae* Rc chromosomal DNA. DNA was transferred from the agarose gel to a membrane and Southern hybridization was done with a ^{32}P -labeled *Sau*3AI probe. Position markers of molecular weights are shown on the left (lane M). Lane 1, *Acc*II; lane 2, *Eco*RI; lane 3, *Eco*T22I; lane 4, *Hind*III; lane 5, *Pma*CI; lane 6, *Sna*BI; lane 7, *Eco*RI & *Bam*HI; lane 8, *Eco*RI & *Eco*T22I; lane 9, *Eco*RI & *Pma*CI; lane 10, *Eco*RI & *Sna*BI; lane 11, *Eco*T22I & *Hind*III; lane 12, *Hind*III & *Pma*CI; lane 13, *Pma*CI & *Sna*BI.

Analysis of the N-terminal amino acid sequence of *R.Hinc*II and identification of part of the *R.Hinc*II gene

The N-terminal amino acid sequence of *R.Hinc*II, which partially purified *R.Hinc*II (30,000 U) from *H. influenzae* Rc, was determined as described in 'Materials and Methods'. Twenty-nine amino acids, Ser-Phe-Ile-Lys-Pro-Ile-Tyr-Gln-Asp-Ile-Asn-Ser-Ile-Leu-Ile-Gly-Gln-Lys-Val-Lys-Arg-Pro-Lys-Ser-Gly-Thr-Leu-Ser-Gly, were detected. The nucleotide sequence of the 3.5 kb fragment inserted in plasmid pAH1 was also identified (not shown). The N-terminal amino acid sequence of *R.Hinc*II was consistent with the amino acid sequences deduced from the DNA sequence downstream from the *M.Hinc*II gene. Thus, plasmid pAH1 contained the *M.Hinc*II gene and part of the *R.Hinc*II gene.

Isolation of the genes coding for the *Hinc*II restriction-modification system

To isolate the two genes, the 1 kb *Sau*3AI fragment shown in Fig. 1 was prepared from plasmid pAH1 and used as a probe in Southern hybridization with *H. influenzae* Rc chromosomal DNA digested with various restriction endonucleases (Fig. 2). The 2.8 kb *Hind*III fragment (Fig. 2; lane 4) was cloned into the *Hind*III site of the pBR322 vector, and the 3.5 kb *Eco*RI-

Figure 3. Restriction map of plasmid pAH2. (A) Plasmid pAH2, which contained both the *R.Hinc*II and *M.Hinc*II genes, was digested with appropriate restriction endonucleases. Agarose gel electrophoresis was used for analysis. Hd-*Hind*III; Af-*Afa*I; D-*Dra*I; X-*Xba*I; Bs-*Bsp*HI; Ps-*Pst*I; Ap^r-resistance gene of ampicillin. (B): *R.Hinc*II (R) and *M.Hinc*II (M) activities of fragments of various sizes were assayed as described in 'Materials and Methods'.

```

AAGCTTACGACTCTGAATTCCTTGAGGGAGAACAAGTCGAAGTGGCTCGCGTGAAGTGGTAAATTCGCAACGTAAGCAGAAGGCAAAACCACCGTTGA 100
ATTTGAACGTGAATTTATGGGTATCACCAAGCGCTCTTGGCACTGAATCCTTCATCTCTGCGCATCGTTCAGAAACACTACTCGTGTCTTACCGAA 200
GCAGCAGTGGCAGGTAACGTGATGAATACGTGGCTTGAAGAGAACGTAATAGTAGGTCGATTAATTCCTGCGGGTACAGGTTTTGCGTATCATCAA 300
ATCGTCAACAACATCGCTTAGTTGATGATGTCGAGCGAAATATCTGAAGAAGATGAAGCGGTATCGCTGATGAGTGTACAAACGGCAGATGATGAAC 400
GCAAAATCTTGAACATTGTTAAATCTGAAATAGAAGATTAAATTTGTATCTTATGTTTAAACCTCAGAATGAGAGTTCTGAGGGTTTTATTAGAA 500

      M D E S K K I S L G Q F F T P T H I V K Y M I G L M T K N 29
AAATAGGATAAAATATGGATGAATCAAGAAAATACACTAGGTGAGTTTTTACTCCTACACATATTGTTAAATATATGATTGGACTAATGACGAAA 600
      K N A S I L E P S S G N G V F L D S L I Q L G Y T N L T S Y E I D 62
ATAAAATGCTAGTATTTTGAACCTAGTAGTGGGAATGGTGTATTCTAGATAGTTAATTCATAGGTTATACAAATTTAACATCGTATGAAATAGA 700
      G D I I S H P F V I N S S F I T S Y D K P Q Y D S I I G N P P Y V 95
TGGTGATATTATTTCTCATCCATTTGTAATTAACCAAGTTTATAACATCATATGAAGCCTCAATATGATTGATCATTTGGAAATCTTATGTT 800
      R W K N L S E L Q K K E L K D N S I W K M Y C N S L C D Y F Y I F I 129
CGTTGGAAAAACCTAAGTGAGCTTCAGAAAGAGCTTAAAGATAATCTATATGGAAATGATTGTAATAGTCTTTGTGATTATTCTATATTTTTA 900
      I K S I L Q L K V G G E L I F I C P D Y F F S T K N A E G L R K F 162
TTATTAATCTATACTTCAATAAAGTTGGAGGAGAAATTAATTTATTTGCGCTGATTATTTTTTCCACTAAGAACGCTGAGGGTTAAGGAAAT 1000
      L I N G S F E K I I L F N E S K V F H G V S S S V V I F K Y I K 195
TTTGATAAATAAGGAGTTTGAAGAAATATATTTAATGAAGTAAAGTATTCATGAGGTTTCTTCTCAGTTGTTATTTTAAATATATAAAA 1100
      G K N I D N I N I I I D S K S P I K S E D I E S L G E S Y Y I P R 229
GGAAAAATATTGATAATATAATATTATTAATAGATTCAAAAAGCCCAATAAAATCTGAGGATATAGAATCTTTAGGTGAATCATATTATTTCCAA 1200
      F S S S D V W V T S P N H I K V A L D K F E S Y C K T I K K V Q Q 262
GATTTCTAGCTCAGATGTTTGGGTACTTCGCCCAATCATATTAAGTTGCATTAGATAAATTTGAGAGTTACTGCAAAACTATCAAAAAGTTCAAGCA 1300
      K S L F D D L S F L D K A F Q M N D I N Y S E L E L L N S I C V A 295
AAAATCATTTTGTGATCTTATGTTTCTAGATAAGGCTTTTCAGATGAATGATATAAATATTTCAGAAATAGAAATGCTTAATTCATTTGTGTAGCT 1400
      K A K H L D A F C F S G Y T R Y K L I L D D N E D K L I T Y F P N 329
AAAGCTAAACACTTAGACGCATTTTGTCTCTGGATACACCAGATACAAGTTAATATTAGATGATAAATGAGGATAAATTAATCACTATTTCCTCAA 1500
      F F Y E F N N Y K D Y L L K R Y S Y N K Y L P Y W E V A F L R N F 362
ATTTCTTTTATGAGTTTAAATATAAGATTATCTATTAAGAGATATAGTTATAAATATTTCCTTATTTGGGAAGTGGCTTTTCTTAGAAATTT 1600
      S L F S K N E K K I F V P C K E R I S K K S N F R F S L V D E F I 395
TTCTCTATTACGCAAGATGAGAAGAAATATTTGTCCGTGCAAGAAAGAAATTTCAAAAAAAGTAATTTAGATTAGCTTAGTTGATGAATTTATA 1700
      Y P T Q D V T A L Y K K E N V K E S I E Y I T A Y L N S K A V F L W 429
TATCCAACACAGGATGTTACTGCATTATATAAGAAAGAAATGTAAGGAGAGATTGAGTATATTACAGCATATTTAAACTCTAAAGCTGTGTTTTAT 1800
      M K Y K G V V K G N V G E F S E K P L T N I P F R R I D W Q L K S 462
GGATGAAGTATAAGGGTGTGTCAGAGTAATGATGTTGAATTTTCAGAAAAACCATTAACAAATATTCCTTTTAGAAGAATAGATTGGCAATTAAGAA 1900
      E K K I H D D I T N L V R K Y L S N K E F S I L H E I N L N L E K 495
CGAAAAAAATTCATGATGATATACTAATTTAGTAAGAAATATTTATCAACAAAGAAATTTCAATTTTACATGAGATTAACCTTAAATTTAGAAAA 2000
      L G I K V E I ★ 502
TTAGGTATAAAGGTGGAATATGAGTTTCATAAAACCTATTTATCAGGATATTAATTCATATTAATCGGGCAAAAAAGTGAACGCTCTAAATCAGGTAC 2100
      M S F I K P I Y Q D I N S I L I G Q K V K R P K S G T 27
TCTGTCAGGTCTGCTGCAGGGGAACCATTTGAAAAATTAGTATATAAGTTTTTGAAGAAAACCTGTCAGATTTAACATTTAAGCAATATGAATCTT 2200
      L S G H A A G E P F E K L V Y K F L K E N L S D L T F K Q Y E Y L 60
AATGATTTATTTATGAAGAACCTGCGATAATTTGGACATGAAGCTAGATATAAATTTATTTTCAACATGCTTTTTTGTTAAGTAGAGGTAAG 2300
      N D H L F M K N P A I I G H E A R Y K L F N S P T L L L F L L S R G K A 94
CTGCAACTGAAAAATGGAGCATAGAAAATTTTGGAGGAGAAACAAATGATACTGCAGATATTTTATTAGTAAGGATCAGTTCTATGAATTTGTAGA 2400
      A T E N W S I E N L F E E K Q N D T A D I L L V K D Q F Y E L L D 127
TGTCAAAAAGAAGAAATATTAGTAATCAGCTCAAGCACCAATATTATTTTCAGCATATAAATAGTCAGACGTGTGCAAAATGATTGATAATAAGAA 2500
      V K R R N I S K S A Q A P N I I S A Y K L A Q T C A K M I D N K E 160
TTTGATTTATTTGATATTAATTTTGAAGTAGACTCGGAACCTAATGGTGAAGATCTAGTTTGTGTTGCTACTCTTTTGTGTAACATTTAAATCTG 2600
      F D I I N Y L E V D S E L N G E D L V C V S T S F A E L F K S E 194
AACCTAGTGAACATATATTAACCTGGGCTGCAGCTGCAAAATTCAGTTTCATGTAAGAGATTTAGATCAGGGGTTAATGGAACATAGAGAATGGGC 2700
      P S E L Y I N W A A A M Q I Q F H V R D L D Q G F N G T R E E W A 227
AAAATCTTATCTAAACATTTTGTACTCAAGCAGAGCAAGAGCCATATCTATGATAGATAAGTTTGTAAAGCCATTTAAGAAATATATCTTTAGTT 2800
      K S Y L K H F V T Q A E Q R A I S M I D K F V K P F K K Y I L ★ 258
CAAGTTCAAAAGATATTATAAGCTT

```

Figure 4. Nucleotide sequence of the 2.8 kb *Hind*III fragment from plasmid pAH2. DNA was sequenced as described in 'Materials Methods'. The amino acid sequences deduced from the *M.Hinc*II and *R.Hinc*II genes are also indicated. The one-letter amino acid notation is used. Dotted sequences indicate those SD-like sequences of *E. coli*. The underlined sequence is palindromic.

*Sna*BI fragment (Fig. 2; lane 10) was cloned into the vector between the *Eco*RI and *Eco*RV sites. Purified plasmid DNAs from these clones were not digested with the *R.Hinc*II, and these clones were not infected by λ virulent phage (titer about 1×10^{-4} to 1×10^{-3}). The plasmid containing the 2.8 kb *Hind*III fragment was named pAH2 and analyzed in detail. First, a physical map of this plasmid was made after its digestion with various restriction endonucleases (Fig. 3A). Second, the *R.Hinc*II and *M.Hinc*II activities resulting when various fragments of the plasmid were introduced into host cells were assayed. *In vitro* activities were not detected, but both activities were detected *in vivo* with same fragments (Fig. 3B).

Nucleotide sequences of the *Hinc*II restriction-modification system from plasmid pAH2

Nucleotide sequences of the entire 2.8 kb *Hind*III fragment inserted in plasmid pAH2 were determined by the method of Sanger et al. (46), with 7-deaza sequencing kit (Fig. 4). Two open reading frames (ORF) that were the *R.Hinc*II and *M.Hinc*II

genes were found. The *M.Hinc*II gene was 1,506 bp long (coding for positions 516 to 2,022), corresponding to a protein of 502 amino acids. The *R.Hinc*II gene was 774 bp long (coding for positions 2,022 to 2,794), corresponding to a protein of 258 amino acids. A palindromic sequence (underlined in Fig. 4) of 12 bp was found upstream of the *M.Hinc*II gene. The *R.Hinc*II gene overlapped by 1 bp with the *M.Hinc*II gene.

Preparation of a clone overproducing the *R.Hinc*II

In general, it is difficult to overproduce methylase, because its turnover is rapid. Hence, we tried to prepare a clone that overproduces only *R.Hinc*II. As cells that overproduce *R.Hinc*II necessarily express the *M.Hinc*II gene effectively, the palindromic sequence, which would control the expression of *M.Hinc*II, was deleted with use of a kilo-sequence deletion kit. First, a 1.8 kb *Afa*I fragment was prepared from plasmid pAH2 and subcloned into the *Sma*I site of pUC19, giving a plasmid named pUC-M.Hc 1. This plasmid was digested with *Bam*HI and *Sph*I, and then treated with exonuclease III, mung bean nuclease, and Klenow

Figure 5. Construction of plasmid pAH-M.Hc. The construction of plasmid pAH-M.Hc from plasmid pAH2 is summarized here. Details are in the text. Plasmid pAH-M.Hc carries the Cm^r gene from the pACYC184 vector and *lac* promoter from the pUC19 vector. Hd-*Hind*III; Af-*Afa*I; Sm-*Sma*I; Ba-*Bam*HI; Sp-*Sph*I; Pv-*Pvu*II; EV-*Eco*RV; Ap^r -resistance gene of ampicillin; Cm^r -resistance gene of chloramphenicol; Tc^r -resistance gene of tetracycline; R-R.*Hinc*II gene; M-M.*Hinc*II gene; PS-palindromic sequence; *lac-lac* promoter.

fragment, in this order, before self-ligation. The plasmid with the palindromic sequence deleted at the most appropriate place (deleted from *Bam*HI site to a position at 481 bp as shown in Fig. 4) was selected and named plasmid pUC-M.Hc 13. A *Pvu*II fragment (1.9 kb) was prepared from this plasmid and cloned into the *Eco*RV site of plasmid pACYC184 (Fig. 5), giving a plasmid named pAH-M.Hc. Then, we tried to create a clone that overproduced R.*Hinc*II. The plasmid pAH2 was digested with *Bsp*HI, treated with T4 DNA polymerase, and digested with *Hind*III to prepare *Bsp*HI-*Hind*III fragments blunt-ended on the *Bsp*HI side and 0.9 kb long. The *Eco*RI-*Hpa*I fragment (0.47 kb) was prepared from pPL-lambda. These fragments were cloned into the *Hind*III and *Eco*RI sites of the pBR322 vector (Fig. 6), giving a plasmid that we named pAH-R.Hc. This plasmid contained the P_L promoter from λ phage DNA, so the plasmid pNT203 was used (35) to control the expression of the R.*Hinc*II gene. This plasmid contained the cl_{I857} and the N genes from λcI_{857} -Sam phage DNA. The three plasmids pAH-M.Hc, pAH-R.Hc, and pNT203 were cloned into the same *E. coli* RR1 cells. The strain created was named *E. coli* RR1-Hinc.

Expression of the R.*Hinc*II polypeptide

E. coli RR1-Hinc was cultured in 20 ml of broth consisting of 2% tryptone, 1% yeast extract, 0.5% NaCl, 0.2% glucose, and 50 mM potassium phosphate buffer (pH 7.5), containing three

antibiotics, 100 μ g/ml ampicillin, 15 μ g/ml tetracycline, and 30 μ g/ml chloramphenicol, at 30°C until the absorbance at 600 nm was 1.0–2.0. Cells were then heated at 42°C for 15 min to activate the promoter before culture at 37°C was continued for 5 hr. Cells were harvested, and about 0.24 g was suspended in 1 ml of buffer A, sonicated for 10 min, and then centrifuged at 25,000 $\times g$ for 30 min. The supernatant obtained was assayed for R.*Hinc*II activity. The total activity of the indicated total supernatant was calculated to be 75,000 U in the conditions used here. This activity was about 1,000-fold that from *H. influenzae* Rc. The supernatant (17 μ l) was checked for expression of R.*Hinc*II by SDS-PAGE. A polypeptide with the molecular weight of 29,000 was overproduced (Fig. 7A); this polypeptide was probably R.*Hinc*II.

Purification of R.*Hinc*II and estimation of its molecular weight

The molecular weight of the R.*Hinc*II purified from *E. coli* RR1-Hinc was estimated from the results of SDS-PAGE shown in Fig. 7B to be 29,000. This value agreed well with the molecular weight (28,490) predicted from the DNA sequence.

Comparison of amino acid sequences of M.*Hinc*II and other adenine methylases

The amino acid sequence of M.*Hinc*II was compared with

Figure 6. Construction of plasmid pAH-R.Hc. Construction of plasmid pAH-M.Hc from plasmid pAH2 is summarized here. Details are in the text. Plasmid pAH-R.Hc carries the Ap^r genes from the pBR322 vector and P_L promoter from λ phage. Hd-HindIII; Bs-BspHI; ER-EcoRI; Hp-HpaI; Ap^r-resistance gene of ampicillin; Tc^r-resistance gene of tetracycline; R-R.HincII gene; M-M.HincII gene; P_L-P_L promoter.

sequences published for other adenine methylases (22,23,27). Homology was found between M.HincII and four methylases: M.TaqI (11), M.PstI (5), M.PaeR7 (7), and M.CviBIII (10). There were at least three blocks that contained consensus sequences (Fig. 8). Blocks I and II have been found in most of the adenine methylases which have been published (27). However, block III has been reported only for the five methylases named above.

DISCUSSION

The host's restriction system, Mrr (20), would restrict DNA (5'-G^mAC-3') modified by M.HincII, if the HincII system was cloned into *E. coli* with the Mrr⁺ phenotype. Therefore, *E. coli* HB101 or RR1, which lack Mrr seem preferable as hosts. With HB101, however, we found that plasmid pAH-M.Hc introduced was partially digested with R.HincII (data not shown). It is not known if this happened because the M.HincII gene was imperfectly expressed, because part of its gene was deleted, because M.HincII was destroyed as soon as it was produced, or

Figure 7. Expression and estimation of the molecular weight of R.HincII from *E. coli* RR1-Hinc. (A); Lane 1, *E. coli* RR1 without induction of promoter; lane 2, *E. coli* RR1 with induction; lane 3, *E. coli* RR1-pAH M.Hc without induction; lane 4, *E. coli* RR1-pAH M.Hc with induction; lane 5, *E. coli* RR1-Hinc without induction; lane 6, *E. coli* RR1-Hinc with induction; lane M, the molecular weight markers phosphorylase b ($M_r=97,400$), bovine serum albumin (66,200), ovalbumin (42,699), carbonic anhydrase (31,000), soybean trypsin inhibitor (21,500), and lysozyme (14,400). The arrow indicates the position for the molecular weight of R.HincII. (B); Lane M, molecular weight markers; lane S, sample of R.HincII purified from *E. coli* RR1-Hinc.

because of some other reason. The functioning Mrr system (20) is not as easily discovered as the Mcr system (15-19), so that investigation of the Mrr system may be needed for better understanding of the *E. coli* restriction systems.

The HincII system also recognizes the same sequence (5'-GTCGAC-3') recognized by the SalI system (46). Therefore, SalI site of plasmid carrying the M.HincII gene will be methylated and plasmid will not be digested with R.HincII or R.SalI, provided that the M.HincII gene is expressed. In practice, plasmids pAH1 and pAH2 carrying the M.HincII gene were not digested with R.SalI (not shown).

The M.HincII and R.HincII genes overlapped slightly and had the same orientation. An ORF was found upstream from the M.HincII gene. Some of the nucleotide sequence of this ORF was identified and the Gene Bank was searched for homologous substances. The ORF had 90% homology with the RNA polymerase β chain gene of *E. coli*. Another ORF was found downstream from the R.HincII gene, but could not be further identified. The two genes coding for the HincII system in *H. influenzae* Rc genome DNA were neighboring, with a palindromic sequence upstream from the methylase, separating them from other nearby genes. In both genes, Shine-Dalgarno (SD)-like sequences were found.

There were the gross changes in profile on induction of *E. coli* RR1-Hinc (Fig. 7A, lane 6). In particular, the bands at approx. $M_r=29,000$ and $43,000$ were induced. It was estimated that the former band corresponds to R.HincII. It was considered that the latter band is suited to a product ($M_r=41,000$) of tetracycline resistance gene, an unknown protein ($M_r=43,000$) from plasmid pNT203, or heat induced protein from host cells.

As shown in Fig. 7B, R.HincII was homogeneous on a SDS-PAGE. There was no report on specific activity of R.HincII. That of the commercially available R.HincII was about 10,000 U/mg. Our purified enzyme here was finally 400,000 U/mg. Highly purified preparation, however, was apt to inactivate and needed to coexist with 0.01% BSA and 0.15% triton X-100. In the

Figure 8. Consensus sequences of five adenine methylases. (A); Diagram of blocks of amino acids homologous in the five adenine methylases listed. Homologous blocks are shown as black bars (27). (B): Amino acid sequences are shown, together with their numerical positions in each block. The one-letter amino acid notation is used.

published restriction enzymes, these molecular weights are 25,000 to 30,000 and these active forms are dimer. On the other hand, that of *R.BsuRI*, which this active form is monomer, is 66,000. That of *R.HincII* was estimated to be 29,000. The active form of this enzyme seems to be dimer as judged by the molecular weight.

Many restriction-modification systems have been cloned in *E. coli* and their gene structures have been analyzed. Methylases have been studied to find homology of their nucleotide sequences and of their predicted amino acid sequences (27). From its predicted amino acid sequence, *M.HincII* was found to have the consensus sequences named blocks I and II (Fig. 8). Block I is found in most methylases and block II is found in adenine methylases. Furthermore, *M.HincII* and four other methylases (Fig. 8) had other consensus sequences (block III). Although these functions of three consensus sequences have been not revealed, block I may be a DNA-binding site in adenine and cytosine methylation and block II is probably important for adenine methylation (27). Block III, found in five methylases (Fig. 8), is a previously unknown sequences not reported. Five methylases, *M.HincII*, *M.TaqI*, *M.PstI*, *M.PaeR7*, and *M.CviBIII*, recognize and methylate the sequence GTPyPu^mAC, TCG^mA, CTGC^mAG, CTCG^mAG, and TCG^mA, respectively. The TNN^mA sequence is commonly in these methylases. Therefore, it was assumed that block III takes part in the recognition and methylation in above methylases. Experiments on our assumption are now in progress with mutations in block III.

ACKNOWLEDGMENTS

We thank Mr. M. Ito of Biotechnology Research Laboratories, Takara Shuzo Co., Ltd., for N-terminal amino acid sequence analysis, and Mr. H. Sagawa of Bioproducts Development Center, Takara Shuzo Co., Ltd., for technical assistance.

REFERENCES

- Walder, R.Y., Hartley, J.L., Donelson, J.E., and Walder, J.A. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 1503-1507.
- Greene, P.J., Gupta, M., Boyer, H.W., Brown, W.E., and Rosenberg, J.M. (1981) *J. Biol. Chem.* **256**, 2143-2153.
- Schoner, B., Kelly, S., and Smith, H.O. (1983) *Gene* **24**, 227-236.
- Bougueleret, L., Schwarzstein, M., Tsugita, A., and Zabeau, M. (1984) *Nucl. Acids Res.* **12**, 3659-3676.
- Walder, R.Y., Walder, J.A., and Donelson, E. (1984) *J. Biol. Chem.* **259**, 8015-8026.
- Kiss, A., Posfai, G., Keller, C.C., Venetianer, P., and Roberts, R.J. (1985) *Nucl. Acids Res.* **13**, 6403-6421.
- Therlaux, 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.
- Lacks, S.A., Mannarelli, B.M., Springhorn, S.S., and Greenberg, B. (1986) *Cell* **46**, 993-1000.
- Szynter, L.A., Slatko, B., Moran, L., O'Donnell, K.H., and Brooks, J.E. (1987) *Nucl. Acids Res.* **15**, 8249-8266.
- Narva, K.E., Wendell, D.L., Skrdla, M.P., and Van Etten, J.L. (1987) *Nucl. Acids Res.* **15**, 9807-9823.
- Slatko, B.E., Benner, J.S., Jager-Quinton, T., Moran, L.S., Simcox, T.G., Van Cott, E.M., and Wilson, G.G. (1987) *Nucl. Acids Res.* **15**, 9781-9796.
- Kita, K., Kotani, H., Sugisaki, H., and Takanami, M. (1989) *J. Biol. Chem.* **264**, 5751-5756.
- Heidmann, S., Seifert, W., Kessler, C., and Domdey, H. (1989) *Nucl. Acids Res.* **17**, 9783-9796.
- Newman, A.K., Rubin, R.A., Kim, S.H., and Modrich, P. (1981) *J. Biol. Chem.* **256**, 2131-2139.
- Ehrlich, M., Gama-Sosa, M.A., Carreira, L.H., Ljungdahl, L.G., Kuo, K.C., and Gehrke, C.W. (1985) *Nucl. Acids Res.* **13**, 1399-1412.
- Noyer-Weidner, M., Diaz, R., and Reiners, L. (1986) *Mol. Gen. Genet.* **205**, 469-475.
- Ehrlich, M., Wilson, G.G., Kuo, K.C., and Gehrke, C.W. (1987) *J. Bacteriol.* **169**, 939-943.
- Raleigh, E.A., Murray, N.E., Revel, H., Blumenthal, R.M., Westaway, D., Reith, A.D., Rigby, P.W.J., Elhai, J., and Hanahan, D. (1988) *Nucl. Acids Res.* **16**, 1563-1575.
- Brooks, J.E., Benner, J.S., Heiter, D.F., Silber, K.R., Szynter, L.A., Jager-Quinton, T., Moran, L.S., Slatko, B.E., Wilson, G.G., and Nwankwo, D.O. (1989) *Nucl. Acids Res.* **17**, 979-997.

20. Heitman, J. and Model, P. (1987) *J. Bacteriol.* **169**, 3243–3250.
21. Stephenson, F.H., Ballard, B.T., Boyer, H.W., Rosenberg, J.M., and Greene, P.J. (1989) *Gene* **85**, 1–13.
22. Brooks, J.E., Blumenthal, R.M., and Gingeras, T.R. (1983) *Nucl. Acids Res.* **11**, 837–851.
23. Lauster, R., Kriebardis, A., and Guschlbauer, W. (1987) *FEBS Letters* **220**, 167–176.
24. Posfai, J., Bhagwat, A.S., Posfai, G., and Roberts, R.J. (1989) *Nucl. Acids Res.* **17**, 2421–2435.
25. Lauster, R. (1989) *Nucl. Acids Res.* **17**, 4402.
26. Lauster, R., Trautner, T.A., and Noyer-Weidner, M. (1989) *J. Mol. Biol.* **206**, 305–312.
27. Lauster, R. (1989) *J. Mol. Biol.* **206**, 313–321.
28. Landy, A., Ruedisueli, E., Robinson, L., Foeller, C., and Ross, W. (1974) *Biochemistry* **13**, 2134–2142.
29. Boyer, H.W. and Roulland-Dussoix, D. (1969) *J. Mol. Biol.* **41**, 459–472.
30. Yanisch-Perron, C., Vieira, J., and Messing, J. (1985) *Gene* **33**, 103–119.
31. Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L., Boyer, H.W., Crosa, J.H., and Falkow, S. (1977) *Gene* **2**, 95–113.
32. Peden, K.W.C. (1983) *Gene* **22**, 277–280.
33. Chang, A.C.Y. and Cohen, S.N. (1978) *J. Bacteriol.* **134**, 1141–1156.
34. Remaut, E., Stanssens, P., and Fiers, W. (1981) *Gene* **15**, 81–93.
35. Shigesada, K., Tsurushita, N., Matsumoto, Y., and Imai, M. (1984) *Gene* **29**, 199–209.
36. Hanahan, D. (1983) *J. Mol. Biol.* **166**, 557–580.
37. Maniatis, T., Fritsch, D.F., and Sambrook, J. (1982) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory. Cold Spring Harbor, New York.
38. Birnboim, H.C. and Doly, J. (1979) *Nucl. Acids Res.* **7**, 1513–1523.
39. Thomas, C.A., Jr., Berns, K.I. and Kelly, T.J., Jr. (1966) in *Procedures in Nucleic Acid Research* (Cantoni, G.L. and Davies, D.R. eds.) pp.535–540, Harper and Row, New York.
40. Matsudaira, P. (1987) *J. Biol. Chem.* **262**, 10035–10038.
41. Smith, H.O. and Wilcox, K.W. (1970) *J. Mol. Biol.* **51**, 379–391.
42. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
43. Southern, E. (1975) *J. Mol. Biol.* **98**, 503–517.
44. Laemmli, U.K. (1970) *Nature* **227**, 680–685.
45. Szomolanyi, E., Kiss, A., and Venetianer, P. (1980) *Gene* **10**, 219–225.
46. Rodicio, M.R. and Chater, K.F. (1988) *Mol. Gen. Genet.* **213**, 346–353.