JOURNAL OF BACTERIOLOGY, Apr. 2003, p. 2296–2305 Vol. 185, No. 7
0021-9193/03/$08.00+0 DOI: 10.1128/JB.185.7.2296–2305.2003
Copyright © 2003, American Society for Microbiology. All Rights Reserved.

Evidence for Horizontal Transfer of the *Eco*T38I Restriction-Modification Gene to Chromosomal DNA by the P2 Phage and Diversity of Defective P2 Prophages in *Escherichia coli* TH38 Strains

Keiko Kita,[1,2]* Hideaki Kawakami,[1] and Hiroaki Tanaka[1]

Department of Biotechnology, Tottori University, Tottori,[1] and Division of Applied Life Sciences, Graduate School of Agriculture, Kyoto University, Uji, Kyoto,[2] Japan

Received 17 September 2002/Accepted 7 January 2003

A DNA fragment carrying the genes coding for a novel *Eco*T38I restriction endonuclease (R.*Eco*T38I) and *Eco*T38I methyltransferase (M.*Eco*T38I), which recognize G(A/G)GC(C/T)C, was cloned from the chromosomal DNA of *Escherichia coli* TH38. The endonuclease and methyltransferase genes were in a head-to-head orientation and were separated by a 330-nucleotide intergenic region. A third gene, the C.*Eco*T38I gene, was found in the intergenic region, partially overlapping the R.*Eco*T38I gene. The gene product, C.*Eco*T38I, acted as both a positive regulator of R.*Eco*T38I gene expression and a negative regulator of M.*Eco*T38I gene expression. M.*Eco*T38I purified from recombinant *E. coli* cells was shown to be a monomeric protein and to methylate the inner cytosines in the recognition sequence. R.*Eco*T38I was purified from *E. coli* HB101 expressing M.*Eco*T38I and formed a homodimer. The *Eco*T38I restriction (R)-modification (M) system (R-M system) was found to be inserted between the *A* and *Q* genes of defective bacteriophage P2, which was lysogenized in the chromosome at *loc*I, one of the P2 phage attachment sites observed in both *E. coli* K-12 MG1655 and TH38 chromosomal DNAs. Ten strains of *E. coli* TH38 were examined for the presence of the *Eco*T38I R-M gene on the P2 prophage. Conventional PCR analysis and assaying of R activity demonstrated that all strains carried a single copy of the *Eco*T38I R-M gene and expressed R activity but that diversity of excision in the *ogr*, *D*, *H*, *I*, and *J* genes in the defective P2 prophage had arisen.

Restriction (R)-modification (M) systems (R-M systems) serve as bacterial immune systems that destroy foreign DNA entering a cell. Typically, type II R-M systems consist of two enzymes: an endonuclease that recognizes and cleaves a specific DNA sequence and a methyltransferase that modifies the same sequence to protect the host chromosome from cleavage. More than 3,000 restriction endonucleases and methyltransferases have been isolated from many species of bacteria and have been subjected to biochemical and genetic studies (30). *E. coli* strains produce a variety of restriction endonucleases, including those of type I and III R-M systems. Type I and III systems are located on the chromosome; in contrast, most of the structural genes of type II R-M systems are located on a plasmid (5, 20, 27, 38, 39), and quite a few are located on chromosomal DNA (16, 21). *Escherichia coli* TH38, isolated from a pig, produces a type II restriction endonuclease, R.*Eco*T38I, that recognizes and cleaves the nucleotide sequence 5′-G(A/G)GC(C/T) ↓C-3′ (23). The occurrence of Hsd (host specificity for DNA) plasmids in *E. coli* TH38 has been investigated; one large plasmid was isolated from *E. coli* TH38, but Hsd^+ transformants were not detected by use of a plasmid (38). These findings strongly suggest the possibility of chromosomal localization of the *Eco*T38I R-M system. Restriction endonucleases that show the same specificity as R.*Eco*T38I, such as R.*Ban*II (14), have been isolated from a variety of bacteria, but none of the gene structures has been determined yet.

The *Eco*O109I R-M system has been cloned from chromosomal DNA, and it was discovered that a P4 integrase-like gene is present in the region adjacent to the R-M system (16). Genes encoding proteins involved in DNA mobility, such as transposases, integrases, and invertases, are sometimes found in the vicinity of R-M systems located on chromosomal DNA (1, 7, 12, 15, 21, 33). These genes might facilitate the transfer of R-M genes among different bacterial strains, one of the indices of the chromosomal location of R-M systems. In addition to the integrase gene, the complete P4 phage genome, except for the *c*II, β, and *gop* genes, and the *leuX* gene from *E. coli* K-12 chromosomal DNA were found in the flanking region; this is the first evidence of the horizontal transfer of a type II R-M system to *E. coli* chromosomal DNA (16).

In addition to these genes involved in DNA mobility, a small open reading frame (ORF) (C) that is known to regulate the expression of the R and/or M genes has been found next to some R-M genes (31, 35). For the *Eco*O109I (16) and *Pvu*II (34) R-M systems, it has been shown that the C and R genes form an operon and that each product of the C gene binds to a specific site upstream of its translational start site and triggers the expression of each C-R operon. Both R-M systems have been found to be transferred to cells on a mobile genetic element: the *Pvu*II system on a plasmid (24, 27) and the *Eco*O109I system on the P4 phage (16). As C proteins are assumed to generate a timing delay, allowing a methyltrans-

* Corresponding author. Mailing address: Division of Applied Life Sciences, Graduate School of Agriculture, Kyoto University, Gokasho, Uji, Kyoto 611-0011, Japan. Phone: 81-774-38-3729. Fax: 81-774-38-3730. E-mail: kita@kais.kyoto-u.ac.jp.

ferase to appear before a cognate restriction endonuclease in new host cells, the C gene is closely associated with the mobility of the R-M system.

In this article, we report the cloning and characterization of the *Eco*T38I R-M system and indicate the location of the system. The nucleotide sequence adjacent to the R-M system leads to horizontal transfer of the *Eco*T38I R-M system through P2 phage integration and indicates the diversity of imprecise P2 excision events.

MATERIALS AND METHODS

Bacterial strains, phages, and plasmids. The *E. coli* strains used in this study were TH38 (23), HB101 (6), and JM109 (37). *E. coli* cultures were incubated at 37°C in Luria-Bertani medium containing 1% Bacto Tryptone (Difco Laboratories, Detroit, Mich.), 0.5% Bacto Yeast Extract (Difco), and 1% NaCl (pH 7.0). When needed, ampicillin (100 μg/ml) and chloramphenicol (25 μg/ml) were added to the cultures.

Enzymes and chemicals. Restriction enzymes and λ DNA were purchased from Takara Shuzo Co., Ltd. (Kyoto, Japan); Toyobo (Osaka, Japan); and Nippon Gene (Toyama, Japan).

Assays of enzyme activities. R.*Eco*T38I activity was assayed by adding 2 μl of enzyme solution to a 15-μl reaction mixture containing 10 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 1 mM dithiothreitol, 50 mM NaCl, and 0.5 μg of λ DNA. Incubation was performed at 37°C for 1 h. Restriction fragments were separated by electrophoresis on a 1% agarose gel. M.*Eco*T38I activity was assayed by adding 2 μl of enzyme solution to a 20-μl reaction mixture containing 10 mM Tris-HCl (pH 8.0), 10 mM 2-mercaptoethanol, 10 mM EDTA, 80 μM *S*-adenosyl-L-methionine, and 0.5 μg of λ DNA. After incubation at 37°C for 1 h, the reaction was terminated by phenol-chloroform extraction; the DNA fragments then were precipitated with ethanol. Following the addition of 15 μl of 10 mM Tris-HCl (pH 7.5)–10 mM $MgCl_2$–1 mM dithiothreitol–50 mM NaCl–4 U of R.*Eco*T38I, the solution was incubated at 37°C for 1 h; aliquots then were analyzed by gel electrophoresis. Modification was judged to have occurred from the lack of digestion. M.*Eco*T38I activity in vivo was assayed as the susceptibility to R.*Eco*T38I of plasmid DNA prepared from the cells being tested.

Determination of the N-terminal amino acid sequence of R.*Eco*T38I. To determine the N-terminal amino acid sequence of R.*Eco*T38I, wild-type R.*Eco*T38I was purified from *E. coli* TH38 cells by chromatography on DEAE-Sephacel (Pharmacia), heparin-Sepharose (Pharmacia), HiTrap Q (Pharmacia), and HiTrap heparin (Pharmacia) columns. The active fractions were pooled, concentrated with Centricon (Amicon), and stored at −20°C. Enzymes at the final purification step were blotted from the sodium dodecyl sulfate (SDS)-polyacrylamide gel onto a polyvinylidene difluoride membrane (Millipore) (21). Sequential degradation of a protein of interest was performed with an ABI491 protein sequencer.

Selection of *Eco*T38I R-M clones. On the basis of the amino acid sequence of R.*Eco*T38I, oligonucleotide N-ter [5′-GTIAA(C/T)CA(C/T)GA(A/G)CA(A/G)GCITA(C/T)AA(C/T)GTIAT-3′] was synthesized as a probe for Southern hybridization analysis of genomic DNA and screening of a library. Hybridization was performed at 37°C for 16 h with modified Church-Gilbert buffer (9), comprising 0.5 M phosphate buffer (pH 7.2), 7% SDS, 10 mM EDTA, 0.05 mg of denatured herring sperm DNA per ml, and ^{32}P-labeled oligonucleotides. Washes were carried out at 50°C with 5× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–0.1% SDS and then with 2× SSC–0.1% SDS. Hybridization was carried out with a PCR product encoding a portion of the M.*Eco*T38I gene, which was labeled with a PCR digoxigenin probe synthesis kit (Boehringer Mannheim) by using two oligonucleotides (5′-GTCGACGTAAACGACTAGGTAC-3′ and 5′-CGTTCCTCAGAATAGGGAACG-3′) as primers. PCR was carried out according to the manual for the kit.

Amplification of DNA fragments by PCR. Oligonucleotides were synthesized according to the nucleotide sequences of the target sites and used as PCR primers. PCR was carried out with an LA-PCR kit (Takara Shuzo Co.) as recommended by the manufacturer. Amplification of DNA fragments flanking the *Eco*T38I R-M system was performed by inverse PCR.

Nucleotide sequencing and analysis. DNA strands were sequenced with a model 3700 DNA sequencer (Applied Biosystems). Oligonucleotide primers were then synthesized and used to walk along the DNA template. Amino acid sequences were compared with all of the sequences in the GenBank database by using the BLAST program.

Expression and purification of M.*Eco*T38I and R.*Eco*T38I in recombinant *E. coli* cells. M.*Eco*T38I was purified from *E. coli* HB101 cells carrying pUCEV by chromatography on HiTrap Q and HiTrap heparin columns. The active fractions were pooled and stored at −20°C.

To express R.*Eco*T38I in recombinant *E. coli* cells, the M.*Eco*T38I gene was subcloned into pACYC184. The M.*Eco*T38I gene was digested with *Sma*I and *Sph*I from pUCEV and then ligated into pACYC184 cleaved with *Eco*RV and *Sph*I; the resulting plasmid, p184 M38, was transformed into *E. coli* HB101. The R.*Eco*T38I gene was amplified by PCR with two primers (5′-GAATTCTTACTAAAGGACACCTATGAAAG-3′ and 5′-CCCGGGAGTATTAATTTTTAAATATGG-3′) and then cloned into the *Eco*RI and *Hin*cII sites of pUC118; the resulting plasmid, pUCR38, was transformed into *E. coli* HB101 carrying p184 M38. R.*Eco*T38I was purified from recombinant *E. coli* HB101 cells carrying pUCR38 and p184 M38 by chromatography on HiTrap Q and HiTrap heparin columns.

Expression of C.*Eco*T38I and measurement of promoter activity. pBSCEcoT38I-His_6 was constructed as follows. Primers C38N (5′-TCTAGATTTTGGAATAAAACCAATGATAGG-3′) and C38C (5′-GGATCCTCAGTGGTGGTGGTGGTGGTGTGATTTACTAATACTTTCATAGGTGTC-3′), which was flanked by a *Bam*HI site and which added a six-amino-acid (His_6) tag to the C-terminal end of C.*Eco*T38I, were used to subclone the His_6-tagged C.*Eco*T38I gene. The PCR-generated DNA fragment was digested with *Xba*I and *Bam*HI, ligated into pBluescript II SK cleaved with *Xba*I and *Bam*HI, and then transformed into *E. coli* JM109. Ni-nitrilotriacetic acid–alkaline phosphatase conjugates (Qiagen) were used for detection of the His_6-tagged protein according to the manufacturer's instructions.

The DNA fragment to be assayed for promoter activity was amplified by PCR with selected primers and then ligated into vector pGEM-T (Promega). The DNA fragment was excised with *Apa*I and *Spe*I and then ligated into *Apa*I/*Spe*I-digested pMCLTerAR (17). The activity of NADPH-dependent aldehyde reductase (AR1) (EC 1.1.1.2) at 37°C was determined by measuring the rate of decrease in the absorbance at 340 nm as described previously (17).

Nucleotide sequence accession number. The GenBank accession number for the DNA sequence of the gene encoding the *Eco*T38I R-M system is AF545861.

RESULTS

Purification of R.*Eco*T38I from *E. coli* TH38. R.*Eco*T38I was purified to homogeneity from *E. coli* TH38, giving a major band of 40 kDa (data not shown). The band was blotted onto a polyvinylidene difluoride membrane and then subjected to N-terminal amino acid sequence analysis. The first 26 amino acids of R.*Eco*T38I obtained by Edman degradation were Met-Lys-Val-Leu-Val-Asn-His-Glu-Gln-Ala-Tyr-Asn-Val-Ile-Ile-Asn-Ala-Xxx-Asn-Asp-Ala-Lys-Lys-Ile-Thr-Asp (Xxx, not identified).

Isolation of *Eco*T38I R-M genes. To isolate the R-M genes, oligonucleotide N-ter was synthesized from the N-terminal amino acid sequence of R.*Eco*T38I and used as a probe for Southern hybridization with *E. coli* TH38 chromosomal DNA digested with various restriction endonucleases. The 1.4-kb *Hin*cII fragment was cloned into the *Hin*cII site of the pUC118 vector to obtain pUC14. The nucleotide sequence of the 1.4-kb fragment indicated that of the three ORFs, one was identical to the sequence encoding the N-terminal region of R.*Eco*T38I and the others were located upstream of the R.*Eco*T38I gene, one on the same DNA strand and the other on the opposite DNA strand. As the deduced amino acid sequence of the ORF located on the opposite DNA strand exhibits significant identity with those of 5-methylcytosine methyltranferases, we assumed that the ORF encodes M.*Eco*T38I. In order to find large DNA fragments carrying the complete M.*Eco*T38I gene within *E. coli* TH38 chromosomal DNA, the 2.2-kb *Eco*RV fragment was cloned into the *Hin*cII site of pUC118 to obtain pUCEV, in which the M.*Eco*T38I gene was under the control of the *lac* promoter. The finding that pUCEV showed resis-

```
1     TACCCACATCAATTCTGGTCAAGCTAAGCAGTACAGCTTATGATTTCAAGATGCAAACTCATGCAGCAATGTTAAATTCTGCATCCATCG
      ATGGGTGTAGTTAAGACCAGTTCGATTCGTCATGTCGAATACTAAAGTTCTACGTTTGAGTACGTCGTTACAATTTAAGACGTAGGTAGC

91    GTTTGAAAATGTGATCCGTTTTACGAATTCAAACTCAACTCTGTTTGACTCTTTTGAACATTAATAGTTTTACCAGTAAGGACATCCAAA
      CAAACTTTTACACTAGGCAAAATGCTTAAGTTTGAGTTGAGACAAACTGAGAAAACTTGTAATTATCAAAATGGTCATTCCTGTAGGTTT
                                  *  S  N  L  S  L  E  T  Q  S  K  Q  V  N  I  T  K  G  T  L  V  D  L      342

181   ACCCCTGCAGCAACAGCCTGCGCCAATTTACAAGGCACAGCATTCCCTATCTGACGCCAAATTGACGAATTTTTCCCCACAAAGACATAA
      TGGGGACGTCGTTGTCGGACGCGGTTAAATGTTCCGTGTCGTAAGGGATAGACTGCGGTTTAACTGCTTAAAAAGGGGTGTTTCTGTATT
      V  G  A  A  V  A  Q  A  L  K  C  P  V  A  N  G  I  Q  R  W  I  S  S  N  K  G  V  F  V  Y      312

271   TTACTCGGGAATGTTTGAATCAAAAGAGCCTCATCTATCGTTAGTCGACGTAAACGACTAGGTACATCGTGCATATCTTTTGGCTCACCA
      AATGAGCCCTTACAAACTTAGTTTTCTCGGAGTAGATAGCAATCAGCTGCATTTGCTGATCCATGTAGCACGTATAGAAAACCGAGTGGT
      N  S  P  F  T  Q  I  L  L  A  E  D  I  T  L  R  R  L  R  S  P  V  D  H  M  D  K  P  E  G      282

361   CCATTCATTAGATGATTATGATAAGTCTCAACATAACTTGTTGAGCCATCATAAAGGTGTTCTTCATCAATAATTGGTGTCCTATTCCCC
      GGTAAGTAATCTACTAATACTATTCAGAGTTGTATTGAACAACTCGGTAGTATTTCCACAAGAAGTAGTTATTAACCACAGGATAAGGGG
      G  N  M  L  H  N  H  Y  T  E  V  Y  S  T  S  G  D  Y  L  H  E  E  D  I  I  P  T  R  N  G      252

451   CCCATAGAAGCAGCTATAGTTGAAGCATACCCATTCGGGTTAATTGGTCTTCCCTGCCCGTTAAACATCATTCCCGCGTAAGGTGACTTC
      GGGTATCTTCGTCGATATCAACTTCGTATGGGTAAGCCCAATTAACCAGAAGGGACGGGCAATTTGTAGTAAGGGCGCATTCCACTGAAG
      G  M  S  A  A  I  T  S  A  Y  G  N  P  N  I  P  R  G  Q  G  N  F  M  M  G  A  Y  P  S  K      222

541   CGTAATACTGGCTTAGCAGCATTTGTAATCCTTGCATTACAGACTCTTGAATTATTTGGATTACCCGCTCGCCCTAAAGGCCGAATAATC
      GCATTATGACCGAATCGTCGTAAACATTAGGAACGTAATGTCTGAGAACTTAATAAACCTAATGGGCGAGCGGGATTTCCGGCTTATTAG
      R  L  V  P  K  A  A  N  T  I  R  A  N  C  V  R  S  N  N  P  N  G  A  R  G  L  P  R  I  I      192

631   TCACCAACTGTTGGCGAAAGATATTGGTATTTTTTATCGCCGCCTCAACTCGTTTCACATCATTTTCGGAACGAAACCCTATAAAAAAC
      AGTGGTTGACAACCGCTTTCTATAACCATAAAAAATAGCGGCGGAGTTGAGCAAAGTGTAGTAAAAGCCTTGCTTTGGGATATTTTTTG
      E  G  V  T  P  S  L  Y  Q  Y  K  K  I  A  A  E  V  R  K  V  D  N  E  S  R  F  G  I  F  F      162

721   ATACGTTCCCTATTCTGAGGAACGCCAAAATCTTTAGAATTCAGCACCAGCAAGGCTGAACGATAGCCTGACTTTTCAGCCATCTTAAAC
      TATGCAAGGGATAAGACTCCTTGCGGTTTTAGAAATCTTAAGTCGTGGTCGTTCCGACTTGCTATCGGACTGAAAAGTCGGTAGAATTTG
      M  R  E  R  N  Q  P  V  G  F  D  K  S  N  L  V  L  L  A  S  R  Y  G  S  K  E  A  M  K  F      132

811   AGTTCACACCTAACTGGTTCAAATTTGGACAACTGTGCTAAAGCCTTAACATTCTCCATAACAAAGCATTCTGGTTGAATTATATCCACC
      TCAAGTGTGGATTGACCAAGTTTAAACCTGTTGACACGATTTCGGAATTGTAAGAGGTATTGTTTCGTAAGACCAACTTAATATAGGTGG
      L  E  C  R  V  P  E  F  K  S  L  Q  A  L  A  K  V  N  E  M  V  F  C  E  P  Q  I  I  D  V      102

901   GCCCGCATGAATGACATAACCAGCTTACTACGTGGATCATGAGCGTCCATTTTCCCTGCCACAGAAAATCCTTGGCATGGAGGCCCACCA
      CGGGCGTACTTACTGTATTGGTCGAATGATGCACCTAGTACTCGCAGGTAAAAGGGACGGTGTCTTTTAGGAACCGTACCTCCGGGTGGT
      A  R  M  F  S  M  V  L  K  S  R  P  D  H  A  D  M  K  G  A  V  S  F  G  Q  C  P  P  G  G      72

991   AAAACACAACCTATATTTTTCAGCCCAGACAATTCACTTAGCATCTCATCAATATCACCACAAAAAACAGGTGATTCATGATTAAGCTTA
      TTTTGTGTTGGATATAAAAAGTCGGGTCTGTTAAGTGAATCGTAGAGTAGTTATAGTGGTGTTTTTTGTCCACTAAGTACTAATTCGAAT
      F  V  C  G  I  N  K  L  G  S  L  E  S  L  M  E  D  I  D  G  C  F  V  P  S  E  H  N  L  K      42

1,081 AATGTATCACATGCATCCTTATCTATATCATTGGCCCAAACTGTTCGGAACCCGGCATTAGAGAAACCGACATCCATCCCTCCTGCCCCA
      TTACATAGTGTACGTAGGAATAGATATAGTAACCGGGTTTGACAAGCCTTGGGCCGTAATCTCTTTGGCTGTAGGTAGGGAGGACGGGGT
      F  T  D  C  A  D  K  D  I  D  N  A  W  V  T  R  F  G  A  N  S  F  G  V  D  M  G  G  A  G      12
                                                               ... ecoT38IC  M  I  G  K  D  R  H     7
1,171 GTAAATAAGCTCACCGCTGAAATTTTGTTCACTTCTTTAGCACCAATTATTCGATTTTGGAATAAAACCAATGATAGGTAAGGATCGGCA
      CATTTATTCGAGTGGCGACTTTAAAACAAGTGAAGAAATCGTGGTTAATAAGCTAAAACCTTATTTTGGTTACTATCCATTCCTAGCCGT
      T  F  L  S  V  A  S  I  K  N  M  ecoT38IM                                                     1

      T  I  N  H  L  L  N  H  M  A  K  R  Q  P  T  I  H  D  H  R  Y  R  C  L  V  Q  L  L  T  K      37
1,261 CACAATCAATCACCTTCTAAACCACATGGCAAAACGACAACCAACTATACATGATCATAGATACAGGTGCCTAGTCCAGCTTCTCACCAA
      GTGTTAGTTAGTGGAAGATTTGGTGTACCGTTTTGCTGTTGGTTGATATGTACTAGTATCTATGTCCACGGATCAGGTCGAAGAGTGGTT

      L  R  K  E  A  S  L  S  Q  S  E  L  A  I  F  L  G  L  S  Q  S  D  I  S  K  I  E  S  F  E      67
1,351 ACTACGCAAAGAAGCTTCTCTTTCGCAGTCAGAATTGGCCATTTTTCTAGGATTATCACAATCAGATATATCTAAAATCGAAAGCTTTGA
      TGATGCGTTTCTTCGAAGAGAAAGCGTCAGTCTTAACCGGTAAAAAGATCCTAATAGTGTTAGTCTATATAGATTTTAGCTTTCGAAACT

      R  R  L  D  A  L  E  L  F  E  L  L  E  V  V  A  S  R  L  G  L  P  M  D  I  L  L  K  D  T      97
1,441 AAGGCGATTAGATGCTTTAGAACTATTTGAATTATTAGAGGTGGTTGCAAGCCGATTAGGTCTACCAATGGATATCTTACTAAAGGACAC
      TTCCGCTAATCTACGAAATCTTGATAAACTTAATAATCTCCACCAACGTTCGGCTAATCCAGATGGTTACCTATAGAATGATTTCCTGTG

ecoT38IR M  K  V  L  V  N  H  E  Q  A  Y  N  V  I  I  N  A  I  N  D  A  K  K  L  T  D  Y  K  T      29
          Y  E  S  I  S  K  S  *                                                                    104
1,531 CTATGAAAGTATTAGTAAATCATGAGCAAGCCTATAATGTAATTATAAATGCAATCAACGATGCAAAAAAATTAACTGATTACAAAACAA
      GATACTTTCATAATCATTTAGTACTCGTTCGGATATTACATTAATATTTACGTTAGTTGCTACGTTTTTTTAATTGACTAATGTTTTGTT

      N  N  Q  W  V  S  I  Q  N  V  I  L  G  T  H  L  T  Y  R  Y  I  L  I  T  G  L  L  A  K  A      59
1,621 ATAATCAATGGGTCTCCATTCAAAATGTAATTTTAGGCACCCACCTTACATACAGATATATACTTATTACTGGTTTACTTGCAAAAGCAA
      TATTAGTTACCCAGAGGTAAGTTTTACATTAAAATCCGTGGGTGGAATGTATGTCTATATATGAATAATGACCAAATGAACGTTTTCGTT

      T  D  P  R  V  N  P  L  A  L  Q  A  N  A  P  V  D  G  A  Y  D  A  R  S  L  C  H  S  V  I      89
1,711 CAGACCCTAGAGTTAACCCTCTGGCTTTACAAGCTAATGCCCCCGTAGACGGAGCTTATGATGCAAGGAGTTTATGTCATTCGGTAATAG
      GTCTGGGATCTCAATTGGGAGACCGAAATGTTCGATTACGGGGGCATCTGCCTCGAATACTACGTTCCTCAAATACAGTAAGCCATTATC

      V  G  K  V  E  G  P  F  L  E  G  K  L  G  A  S  N  E  P  F  L  N  K  P  A  R  Y  M  L  H      119
1,801 TAGGCAAAGTAGAAGGCCCATTTTTAGAGGGTAAGTTAGGCGCATCAAATGAGCCATTTCTAAATAAACCAGCTCGTTACATGCTCCACT
      ATCCGTTTCATCTTCCGGGTAAAAATCTCCCATTCAATCCGCGTAGTTTACTCGGTAAAGATTTATTTGGTCGAGCAATGTACGAGGTGA

      S  S  D  N  P  V  R  R  G  N  D  K  V  L  Q  Q  L  S  I  D  I  L  H  A  A  T  T  Q  T  L      149
1,891 CCTCTGATAACCCTGTGCGCAGAGGAAATGATAAGGTTTTACAACAACTTTCAATTGATATATTACATGCAGCAACAACCCAAACCTTAG
      GGAGACTATTGGGACACGCGTCTCCTTTACTATTCCAAAATGTTGTTGAAAGTTAACTATATAATGTACGTCGTTGTTGGGTTTGGAATC

      A  Y  E  M  L  V  I  A  L  Y  F  T  L  Q  R  T  N  R  V  I  T  P  N  S  I  N  F  D  F  H      179
1,981 CTTATGAAATGCTTGTTATTGCATTATATTTTACCCTACAGCGAACCAATAGAGTTATAACGCCGAACTCAATTAATTTTGATTTCCATA
      GAATACTTTACGAACAATAACGTAATATAAAATGGGATGTCGCTTGGTTATCTCAATATTGCGGCTTGAGTTAATTAAAACTAAAGGTAT

      K  I  I  Y  N  I  I  S  H  P  C  D  G  E  T  C  A  I  A  A  A  I  S  L  H  L  L  G  E  Q      209
2,071 AGATTATTTATAATATTATATCTCATCCTTGTGATGGAGAAACATGCGCGATTGCTGCCGCAATTTCGTTACATTTATTAGGTGAACAAA
      TCTAATAAATATTATAATATAGAGTAGGAACACTACCTCTTTGTACGCGCTAACGACGGCGTTAAAGCAATGTAAATAATCCACTTGTTT

      R  G  W  I  I  K  A  H  P  V  N  Q  A  G  S  S  S  K  E  I  L  D  I  D  V  Y  H  D  D  I      239
2,161 GAGGATGGATAATAAAAGCACATCCCGTCAATCAGGCCGGTAGCTCCTCAAAAGAGATACTGGACATAGATGTATATCATGATGATATAG
      CTCCTACCTATTATTTTCGTGTAGGGCAGTTAGTCCGGCCATCGAGGAGTTTTCTCTATGACCTGTATCTACATATAGTACTACTATATC

      V  F  L  S  I  E  V  K  D  K  P  F  N  Y  Q  D  V  N  H  A  V  S  K  A  S  A  S  G  I  S      269
2,251 TTTTCCTTTCTATAGAAGTGAAAGACAAACCATTTAATTATCAAGATGTTAATCACGCCGTTTCAAAAGCATCTGCATCCGGGATTTCAA
      AAAAGGAAAGATATCTTCACTTTCTGTTTGGTAAATTAATAGTTCTACAATTAGTGCGGCAAAGTTTTCGTAGACGTAGGCCCTAAAGTT

      K  V  I  F  L  K  G  P  R  A  T  N  L  D  I  D  E  S  L  A  I  E  N  A  A  T  K  G  V  S      299
2,341 AAGTTATCTTCCTTAAAGGGCCAAGAGCAACAAACTTAGATATTGATGAGTCTCTTGCGATTGAGAACGCTGCGACAAAAGGTGTTTCAC
      TTCAATAGAAGGAATTTCCCGGTTCTCGTTGTTTGAATCTATAACTACTCAGAGAACGCTAACTCTTGCGACGCTGTTTTCCACAAAGTG

      L  S  F  S  D  V  M  T  F  T  T  T  C  Y  A  L  S  P  L  L  S  N  D  R  I  I  D  F  I  N      329
2,431 TAAGCTTTAGTGACGTCATGACATTCACAACAACATGTTATGCGTTATCGCCACTATTAAGTAATGACAGAATAATAGATTTCATAAATA
      ATTCGAAATCACTGCAGTACTGTAAGTGTTGTTGTACAATACGCAATAGCGGTGATAATTCATTACTGTCTTATTATCTAAAGTATTTAT

      N  T  L  K  D  I  R  A  K  D  S  T  I  E  Y  I  Q  S  I  F  K  N  *                          351
2,521 ACACTTTAAAAGACATCAGAGCTAAAGATAGCACTATAGAATACATACAATCCATATTTAAAAATTAATACTCCCGGGACAATATAGTTC
      TGTGAAATTTTCTGTAGTCTCGATTTCTATCGTGATATCTTATGTATGTTAGGTATAAATTTTTAATTATGAGGGCCCTGTTATATCAAG

2,611 ACAACTACGTTGTCCCTTTTAAAACAGTTAAGCAATTGCATTATATTAATTATTTACTTTTTATCACAATCTCATAATTCATACGCTCTC
      TGTTGATGCAACAGGGAAAATTTTGTCAATTCGTTAACGTAATATAATTAATAAATGAAAAATAGTGTTAGAGTATTAAGTATGCGAGAG
```

tance to cleavage by R.*Eco*T38I indicated that the 2.2-kb region is essential for encoding of the methyltransferase.

Sequence analysis of the complete R-M system. In order to obtain the complete nucleotide sequence of the *Eco*T38I R-M system, the 1.8-kb DNA fragment downstream of the R.*Eco*T38I gene was amplified by inverse PCR, and then its nucleotide sequence was analyzed. The DNA sequence of the 2,700-bp region that covers the entire *Eco*T38I R-M system in the 4,127-bp *Eco*RV/*Pst*I fragment analyzed is shown in Fig. 1. The R.*Eco*T38I and M.*Eco*T38I genes were aligned head to head, and putative palindromic sequences that were seen downstream of each gene could be the transcriptional termination sites for individual genes.

In the ORF assigned to the endonuclease gene, an ATG codon at nucleotide position 1,533 and a termination codon at nucleotide position 2,586 were found. In addition, an appropriate ribosome-binding sequence, AGGA, was present 9 bp upstream of the ATG codon. The ORF consisted of 1,053 bp and encoded a 351-residue polypeptide. The predicted mass, 38,898 Da, was close to the value estimated for the enzyme purified from *E. coli* TH38 by SDS-polyacrylamide gel electrophoresis (PAGE). The deduced amino acid sequence of R.*Eco*T38I exhibited 28% identity with that of R.*Sac*I (36), which recognizes GAGCT↓C, one of the variants of the recognition sequence of R.*Eco*T38I.

In the ORF assigned to the methylase gene, an ATG codon at nucleotide position 1,157 and a termination codon at nucleotide position 113 were found. In addition, an appropriate ribosome-binding sequence, AGGA, was present 7 bp upstream of the ATG codon. The ORF consisted of 1,044 bp and encoded a 348-residue polypeptide. The predicted sequence of the M.*Eco*T38I protein included 10 sequence motifs characteristic of all prokaryotic 5-methylcytosine methyltransferases (29) but did not show significant similarity to that of M.*Sac*I, as expected from the results obtained for the cognate restriction enzyme.

As shown in Fig. 1, a third ORF was discovered upstream of the R.*Eco*T38I gene. In the ORF, the ATG codon was located at nucleotide position 1,241 and the termination codon was located at nucleotide position 1,553, partially overlapping the R.*Eco*T38I gene. In addition, an appropriate ribosome-binding sequence, GGA, was present 12 bp upstream of the ATG codon. The ORF consisted of 312 bp and encoded a 104-residue polypeptide. As shown in Fig. 2, the central region (36 to 74 amino acid residues) of the deduced amino acid sequence exhibited significant similarity not only to those of members of the helix-turn-helix 3 and helix-turn-helix XRE (xenobiotic response element) families of DNA-binding proteins, including Cro and *c*I, but also to that of C.*Eco*RV, an activator protein for the R.*Eco*RV gene (25). The ORF therefore was designated the C.*Eco*T38I gene, which might produce a control protein for the *Eco*T38I R-M gene. Neither a conserved DNA sequence element termed a "C box," which is found immediately upstream of some C genes and is at least one target of C.*Pvu*II binding (31), nor a C.*Eco*O109I-binding sequence (17) was found upstream of the C.*Eco*T38I gene translational start site. The G+C contents of the R.*Eco*T38I, M.*Eco*T38I, and C.*Eco*T38I genes were 35, 42, and 38%, respectively. There was no significant similarity between the nucleotide and deduced amino acid sequences of R.*Eco*T38I and M.*Eco*T38I.

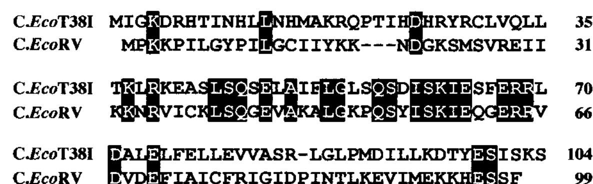

FIG. 2. Alignment of C.*Eco*T38I and C.*Eco*RV. Each line represents one sequence. Shading indicates conserved amino acid residues. Gaps in the aligned sequences are indicated by dashes.

Expression of the R.*Eco*T38I and M.*Eco*T38I genes in recombinant *E. coli* and characterization of the gene products. We inserted the R.*Eco*T38I gene under the control of the *lac* promoter, and the resulting plasmid, pUCR38, was transformed into *E. coli* HB101 cells carrying p184 M38, which expresses M.*Eco*T38I. To characterize R.*Eco*T38I, *E. coli* HB101 cells carrying p184 M38 and pUCR38 were cultured in the presence of 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG), and then the enzyme was purified from the cells. The specific activity of R.*Eco*T38I in extracts of recombinant cells was 100 times higher than that in extracts of wild-type cells. The purified sample gave a single protein band corresponding to a molecular mass of 40 kDa. The molecular mass of the native enzyme, which was measured by HiPrep Sephacryl S-100 gel filtration, was estimated to be 73 kDa. Like other type II R enzymes, R.*Eco*T38I is a homodimeric protein. The sequence of the first five amino acids of the enzyme, Met-Lys-Val-Leu-Val, obtained by Edman degradation exactly corresponded to that of wild-type R.*Eco*T38I and that predicted from the nucleotide sequence. The restriction pattern obtained on digestion with recombinant R.*Eco*T38I was the same as that obtained with the wild-type enzyme.

To characterize M.*Eco*T38I, we purified the enzyme from *E. coli* HB101 cells carrying pUCEV. The purified sample gave a single protein band corresponding to a molecular mass of 41 kDa. The molecular mass of the native enzyme, which was measured by HiPrep Sephacryl S-100 gel filtration, was estimated to be 39 kDa; this value was consistent with the value determined by SDS-PAGE. Like other type II DNA methyltransferases, M.*Eco*T38I is a monomeric protein. The start of the methylase gene was confirmed by N-terminal amino acid analysis. The sequence of the first 10 amino acids of the enzyme, Met-Gln-Lys-Ile-Ser-Ala-Val-Ser-Leu-Phe, obtained by Edman degradation exactly corresponded to that predicted

FIG. 1. Nucleotide sequence of the 2,700-bp region. The amino acid sequences assigned to the C.*Eco*T38I (*ecoT38IC*) and R.*Eco*T38I (*ecoT38IR*) genes are given below the nucleotide sequences, and that assigned to the M.*Eco*T38I (*ecoT38IM*) gene is given above the nucleotide sequence. The nucleotide sequence is numbered from the leftmost end, and the deduced amino acid sequences of R.*Eco*T38I, M.*Eco*T38I, and C.*Eco*T38I are numbered from the initiation codon of each gene. The potential ribosome-binding sequence is indicated by dots. Pairs of arrows indicate palindromic sequences characteristic of the termination signal.

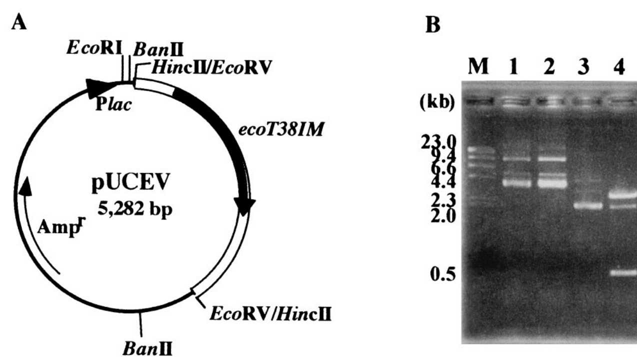

FIG. 3. Analysis of the methylation specificity of M.*Eco*T38I (*ecoT38IM*). (A) Structure of pUCEV, in which the 2.2-kb *Eco*RV fragment carrying the M.*Eco*T38I gene was cloned into the *Hin*cII site of pUC118. (B) Resistance to R.*Ban*II of M.*Eco*T38I-modified DNA. pUCEV and pUC118 were incubated with R.*Ban*II. The products were separated by 1% agarose gel electrophoresis. Lane 1, pUCEV; lane 2, pUCEV plus R.*Ban*II; lane 3, pUC118; lane 4, pUC118 plus R.*Ban*II. λ *Hin*dIII digests were used for molecular size calibration (lane M).

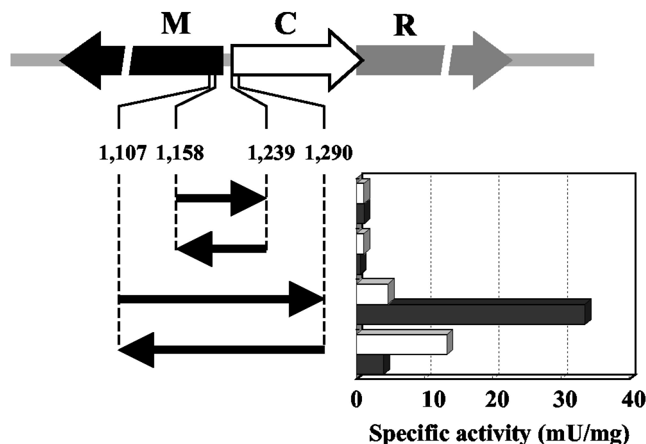

FIG. 4. Effect of C.*Eco*T38I on the promoter activities of the R.*Eco*T38I and M.*Eco*T38I genes. The segments from 1,158 to 1,239 bp and from 1,107 to 1,290 bp shown in Fig. 1 were cloned upstream of a promoterless AR1 gene in vector pMCLTerAR. Promoter activities associated with various DNA fragments were measured for both orientations. Each clone was assayed for AR1 activity in the absence (open bars) and presence (closed bars) of C.*Eco*T38I (graph). R, M, and C represent R.*Eco*T38I, M.*Eco*T38I, and C.*Eco*T38I, respectively.

from the nucleotide sequence starting at a GTG codon appearing at nucleotide position 1,202 but not that starting from the ATG codon described above. The ORF consisted of 1,092 bp and encoded a 364-residue polypeptide. The predicted mass, 39,803 Da, was close to the value determined by SDS-PAGE. No appropriate ribosome-binding sequence was present upstream of the GTG codon.

In order to characterize the methylation specificity of M.*Eco*T38I, pUCEV, carrying the M.*Eco*T38I gene and two *Eco*T38I sites, GAGCTC and GGGCTC, was incubated with R.*Ban*II, an isoschizomer of R.*Eco*T38I which can cut G(A/G) GC(C/T)m^{5}C but not G(A/G)Gm^{5}C(C/T)C (14, 26). Digestion of pUC118 with R.*Ban*II gave 2.7- and 0.45-kb fragments (Fig. 3B, lane 4), in contrast, pUCEV was not digested with R.*Ban*II (Fig. 3B, lane 2). The deduced amino acid sequence of M.*Eco*T38I contained all 10 motifs that are conserved in bacterial 5-methylcytosine methyltransferases. These results suggested that M.*Eco*T38I methylated the inner cytosines in the recognition sequence to give 5′-G(A/G)Gm^{5}C(C/T)C-3′.

Function of C.*Eco*T38I. To determine the C.*Eco*T38I responsiveness of *Eco*T38I promoters, we placed the His$_6$-tagged C.*Eco*T38I gene under the control of the *lac* promoter of pBluescript II SK to supply C.*Eco*T38I in *trans*. The production of C.*Eco*T38I was examined by Western blot analysis with anti-His tag antibodies; a protein band corresponding to a molecular mass of 16 kDa was detected for *E. coli* JM109 carrying pBSCEcoT38I-His$_6$ (data not shown). We cloned the putative promoter region upstream of the promoterless AR1 gene in plasmid pMCLTerAR. Fragments to be assayed for promoter activity were PCR amplified with selected oligonucleotide primers, and DNA products were inserted between the *Spe*I and *Apa*I sites of the promoter screening vector, pMCLTerAR, in both orientations. Cotransformants of *E. coli* JM109 carrying plasmids derived from pMCLTerAR and from pBluescript II SK or pBSCEcoT38I-His$_6$, which generates C.*Eco*T38I, were grown in the presence of IPTG. Reporter gene assays then were carried out for bacterial cell extracts with 4-chloro-3-oxobutanoate ethyl ester as a substrate.

As shown in Fig. 4, the promoter activity of an 82-bp fragment including the intergenic region of the M.*Eco*T38I and C.*Eco*T38I genes was below the limit of detection and was not affected by C.*Eco*T38I in either orientation. The promoter activity of a 184-bp fragment was high in the absence of C.*Eco*T38I and increased to seven times that in the absence of C.*Eco*T38I when the fragment from bp 1107 to 1290 was inserted in the same orientation as the reporter gene. In contrast, when the 184-bp fragment was inserted in the opposite orientation, the promoter activity increased to three times that in the absence of C.*Eco*T38I and decreased in the presence of C.*Eco*T38I. These results suggested that C.*Eco*T38I acts as both a positive regulator of R.*Eco*T38I expression and a negative regulator of M.*Eco*T38I expression.

Nucleotide sequences flanking the R-M system. In the analyzed 4,127-bp region, one partial ORF was identified downstream of the R.*Eco*T38I gene. The product of this ORF showed significant homology to the N-terminal region encoded by the *A* gene from phage P2 (GenBank accession number AF063097). A P2 *cos* sequence, followed by a nucleotide sequence similar to that of the *Q* gene from phage P2, was also found downstream of the M.*Eco*T38I gene. These findings suggest that the *Eco*T38I R-M system was inserted between the genes from the P2 prophage integrated in *E. coli* chromosomal DNA. In order to characterize the gene arrangement of the P2 prophage on *E. coli* TH38 chromosomal DNA, we amplified the DNA fragments flanking the R-M system and determined the entire nucleotide sequence of the 16-kb *Eco*RI region. BLAST searches of the GenBank database revealed that DNA downstream of the R.*Eco*T38I gene exhibited a high level of similarity with the nucleotide sequences of the *A*, *C*, and *int* genes and the *att*L site from phage P2, followed by the sequence from b2084 through *gatR1* of *E. coli* K-12 MG1655 (4). The DNA downstream of the M.*Eco*T38I gene exhibited a high level of similarity with the nucleotide sequences of the *Q*, *O*, *N*, *M*, *L*, *X*, *Y*, *lysB*, *W*, *J*, *I*, *H*, *D*, and *ogr* genes and the *att*R site from phage P2, followed by the *yegQ* sequence of *E. coli* K-12

MG1655. These results indicated that on *E. coli* TH38 chromosomal DNA, a P2 attachment site similar to *loc*I found at 46.7 min on *E. coli* K-12 MG1655 was occupied by a defective P2 prophage carrying the *Eco*T38I R-M system. The *Q*, *O*, *Y*, *lysB*, *W*, *J*, *I*, *D*, and *A* genes seemed to be pseudogenes, as truncation at the 5′ and/or 3′ ends or a frameshift mutation was found in these genes. In contrast, other genes could code for proteins corresponding to phage P2. Gene duplications, rearrangements, inversions, or insertions have not been found in the region analyzed so far. The gene organization is summarized in Fig. 5.

Variability of the P2 prophage in *E. coli* TH38. In order to determine whether or not *E. coli* TH38 has (an)other P2 prophage(s) at (an)other *att*B site(s), as in *E. coli* K-12 MG1655, we amplified possible *att*B sites by using PCR. Four pairs of primers were designed to amplify the *loc*I, *loc*II, *loc*III, and *loc*H sites based on the nucleotide sequence of *E. coli* K-12 MG1655 (Table 1). The lengths of the DNA fragments amplified by using the *loc*I and *loc*III primers from *E. coli* TH38 were the same as those obtained for *E. coli* K-12 MG1655. In contrast, a DNA fragment was not amplified from *E. coli* TH38 with the *loc*H primers (data not shown). These findings indicated the presence of *loc*II and *loc*III sites and the absence of a *loc*H site or a mutation of the *loc*H flanking region. With the *loc*I primers, 0.7- and 12-kb fragments were amplified from *E. coli* K-12 MG1655 and *E. coli* TH38, respectively. These fragments were consistent with the sizes expected from the nucleotide sequences. In addition to the 12-kb fragment, DNA fragments of smaller sizes were amplified from *E. coli* TH38. A sample from a freezer vial was streaked, and 10 colonies were taken from the same streak. The numbers and sizes of these fragments differed from those of the colony from which the template DNA was prepared. We examined samples for the presence of R.*Eco*T38I activity and the P2 prophage carrying the *Eco*T38I R-M system. R.*Eco*T38I activity was detected in all colonies, and the 12-kb DNA fragment was amplified from all samples, but the patterns of DNA fragments of smaller sizes could be categorized into four groups: 1.7 kb, 1.1 kb, 1.0 kb, and none (Fig. 6B). We determined the partial nucleotide sequences of these fragments and found that each fragment carried the P2 *att loc*I through *D* genes, but the nucleotide sequence was not identical to that of *E. coli* TH38 determined so far. These results suggested the variability of defective P2 prophages among *E. coli* TH38 strains. We synthesized additional PCR primers (Table 1 and Fig. 6A) and amplified DNA segments covering the defective P2 prophage integrated at *loc*I. The lengths of the DNA fragments amplified with the E-F, G-H, and I-J primers were similar (data not shown), but those amplified with the A-B and C-D primers were different and could be categorized into four groups (Fig. 6C). These results suggested a difference in genetic organization from the *ogr* through *L* genes in the defective P2 prophage. DNA sequences were determined and compared to those of phage P2, *E. coli* K-12 MG1655, and the *E. coli* TH38 control strain analyzed above (Fig. 6D). The gene order and the orientation were identical to those found for the *E. coli* TH38 control strain. The region from 25 bp upstream of the *D* gene to 455 bp downstream of the initiation codon of the *H* gene was deleted at the same positions in all *E. coli* TH38 strains. In addition, the entire 5′ end of the *W* gene was truncated at the same position. However, the *D* gene showed diversity. The N-terminal part of the *ogr* gene product and the C-terminal part of the *D* gene product were deleted in group 1, the N-terminal half of the *D* gene product was deleted in group 2, and the *ogr* and *D* gene products were conserved in group 3. The 3′ end of the *D* gene was truncated at the same positions in the group 1 and control strains, but the 5′ end of the *D* gene was truncated at different positions in *E. coli* K-12 MG1655 and the control strain.

DISCUSSION

We analyzed the gene organization of the novel *Eco*T38I R-M system on the chromosomal DNA of *E. coli* TH38. Many R-M systems recognize the same DNA sequence as *Eco*T38I; however, this is the first report of the gene structure of an R-M system that recognizes [G(A/G)GC(C/T)C]. The 2.7-kb region encompassed all of the R-M genes, which were aligned in a head-to-head orientation. R.*Eco*T38I and M.*Eco*T38I could be expressed in *E. coli* HB101 and purified to homogeneity. The molecular masses of both proteins were consistent with those predicted from the nucleotide sequences. R.*Eco*T38I was a homodimeric protein; in contrast, M.*Eco*T38I was a monomeric protein. The results presented in this report suggest that M.*Eco*T38I methylated the inner cytosines in the recognition sequence to yield 5-methylcytosine.

Between the R and M genes, one small ORF (C.*Eco*T38I gene) was present upstream of the R.*Eco*T38I gene and partially overlapped that gene. The C.*Eco*T38I gene encodes a protein of 104 amino acids with a molecular mass of 12,014 Da, a size which is in good agreement with the predicted sizes of other C proteins that associate with several type II R-M systems and regulate the expression of R-M genes. In this study, we have shown that C.*Eco*T38I acts both as an activator of R.*Eco*T38I expression and as a repressor of M.*Eco*T38I expression. In both the *Eco*O109I and *Pvu*II systems, it has been shown that the C protein binds to DNA upstream of its structural gene and triggers the transcription of its own gene and the cognate restriction endonuclease gene but has little effect on the transcription of the methyltransferase gene. In contrast, C.*Bam*HI acts both as an activator of endonuclease expression and as a repressor of methyltransferase expression in *E. coli* (13). No significant homology of the deduced amino acid sequences of C.*Eco*T38I, C.*Eco*O109I, C.*Pvu*II, and C.*Bam*HI was found. Furthermore, no sequence similar to the binding sites determined for C.*Eco*O109I and C.*Pvu*II was found upstream of C.*Eco*T38I. These results suggest that C.*Eco*T38I may bind to a novel site and regulate the transcription of the *Eco*T38I R-M genes. Clarification of the control mechanism awaits the purification of C.*Eco*T38I as well as in vitro DNA-binding assays. The C protein is assumed to generate a timing delay of R gene expression when the R-M system enters a new host. It is quite reasonable that a gene encoding a putative C protein was found in the *Eco*T38I R-M system, which was shown to be horizontally transferred by phage P2.

The structure of the DNA adjacent to the *Eco*T38I R-M system was analyzed in detail, and we found that almost 30% of the phage P2 DNA was arranged sequentially on both sides of the system. Furthermore, a DNA sequence identical to that of *loc*I of the phage P2 attachment site followed by *E. coli* K-12

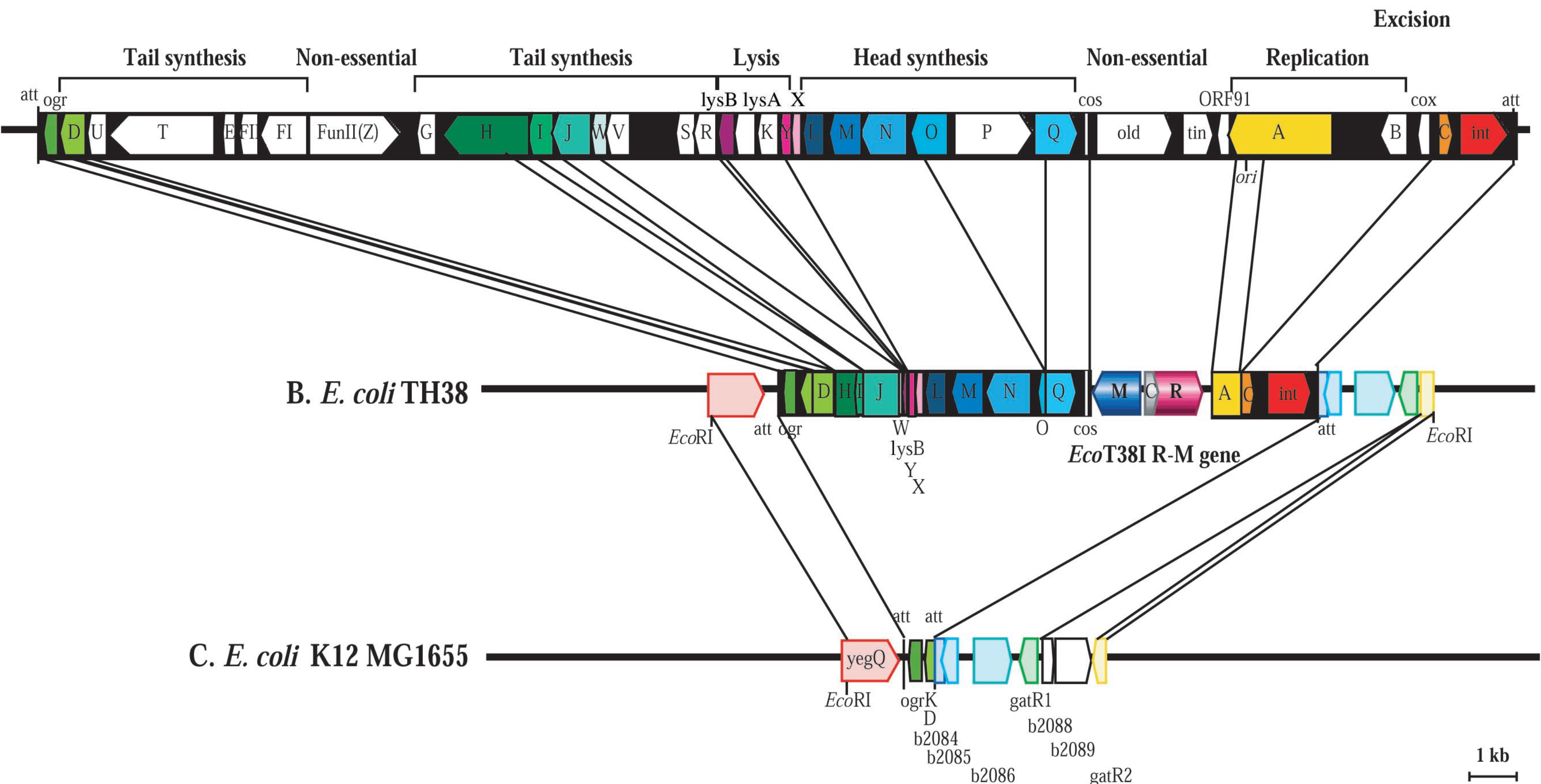

FIG. 5. Gene organization of the *Eco*T38I R-M system and flanking regions on the *E. coli* TH38 chromosome. (A) Gene order of the P2 phage integrated into the *E. coli* chromosome. (B) Gene order of the 16-kb *Eco*RI region flanking the *Eco*T38I R-M genes on the *E. coli* TH38 chromosome. The genes and their directions of transcription, *cos*, and phage attachment sites (*att*) are shown. R, M, and C represent R.*Eco*T38I, M.*Eco*T38I, and C.*Eco*T38I, respectively. (C) Gene order of P2 *att* (*loc*I) and flanking regions in *E. coli* K-12 MG1655. att indicates *att* (*loc*I).

TABLE 1. PCR primers used in this study

Target DNA	Primer	Sequence (5′-3′)	Positions[a]
P2 *att loc*I	I61	TGTACCAGGAACCACCTCCTTAGCCTGTGT	1354–1383[b]
	I62	GAAGACGGTGACCAGGGATAGGGCTTATGC	13398–13369[b]
P2 *att loc*II	II30	GAACTGACACAACCGCTATTTATCGCCGCG	7168–7197[c]
	II31	GTAGAGAATGAGCCACCAAACGCGAATCGC	7256–7227[c]
P2 *att loc*III	III28	ACTGCCTTTATTCAACAACGGAAATTCCAG	2814–2843[d]
	III29	TGGAACCAAAAACAAAAAAACAGCGTTCGC	2902–2873[d]
P2 *att loc*H	H32	TTTGTCGCGAAACAGAAACACTGTGTCAGG	1167–1196[e]
	H33	CATGAGAATCAGACCATTCGCCGTTGCATC	1261–1232[e]
I[f]	A	CAGGAACCACCTCCTTAGCC	1359–1378[b]
	B	GGGTATGACGGGGGCGGG	2713–2696[b]
II[f]	C	GCATCAGCATGTAATCCGGCGTC	2632–2654[b]
	D	ACCGCGTCGCTTTATGAGCG	4766–4747[b]
III[f]	E	AATCAGTGTCGCCTTGCGTTC	3957–3977[b]
	F	CGGCCTTTCGACTTCACCATGTTTTCGCG	7622–7594[b]
IV[f]	G	CGCGAAAACATGGTGAAGTCG	7594–7614[b]
	H	GAGTGGTCGGCATTTACGCGC	11414–11394[b]
V[f]	I	CAATCTCATAATTCATACGCTCTCC	10865–10889[b]
	J	GAAGACGGTGACCAGGGATAG	13398–13378[b]

[a] Nucleotide positions are based on the numbering of the published sequences indicated in footnotes *b* to *e*.
[b] This study.
[c] GenBank accession number AE000461.
[d] GenBank accession number AE000245.
[e] GenBank accession number AE000292.
[f] See Fig. 6A.

MG1655 chromosomal DNA was found in the adjacent region. P2 is a temperate phage that forms stable lysogens in several enterobacteria, including *E. coli* C and K-12. In the lysogenic stage, P2 has always been found at an integrated prophage. The integration occurs through site-specific recombination between a bacterial attachment site, *att*B, and the attachment site of P2, *att*P. The P2 prophage has been found at at least 10 sites on the *E. coli* genome. The preferred site varies with the strain used, but multiple lysogens are possible, with integration occurring at separate locations. In *E. coli* C, one site, *loc*I, is preferred, being occupied before any of the others (3). This is also true of *E. coli* K-12, in which four *att*B sites, *loc*I, *loc*II, *loc*III, and *loc*H, were found; however, *loc*I was occupied by a cryptic remnant of P2 (Fig. 5C) (2). The core sequence of *loc*I completely matched the 27-nucleotide core sequence of *att*P; the sequences of *loc*II, *loc*III, and *loc*H exhibited 20, 17, and 17 matches, respectively. Both *att*R and *att*L found in *E. coli* TH38 completely matched the core sequence of *loc*I; in addition, nucleotides outside the core region of *att*R and *att*L were almost identical to those in *E. coli* K-12. We showed that *E. coli* TH38 carries additional P2 *att*B sites, *loc*II and *loc*III, but not *loc*H, although it was not determined whether these sites are at the same map positions as in *E. coli* K-12 MG1655. These results strongly support the proposal that a hybrid phage P2, in which the *old*, *tin*, ORF91, and partial *A* genes were substituted by R-M genes, infected *E. coli* TH38 cells that carried multiple P2 *att*B sites and integrated the DNA into chromosomal DNA at the *loc*I site through site-directed recombination catalyzed by P2 integrase, followed by excision of some P2 genes. The G+C content of the genes in the *Eco*T38I R-M system was low, as in other R-M systems. It is quite interesting that the P2 genes *old*, *tin*, and ORF91, which were replaced by the *Eco*T38I R-M system, and nonessential genes for the lytic cycle, such as *Fun*II(Z), which were deleted from the defective prophage on the *E. coli* TH38 genome, have low G+C contents. P2 replication is initiated by a strand-specific nick at *ori*, which is located within the *A* gene, and replication proceeds unidirectionally (8). The position of the nick site, position 29,892 in the P2 complete sequence (GenBank accession number AF063097), was found in the *A* gene of the defective P2 prophage and was located about 450 nucleotides downstream of the junction of phage P2 and the R-M system. These results support the possibility that the single-stranded tail (corresponding to the partial *A* gene, ORF91, *tin*, *old*, and other genes) produced by gene *A* products promotes recombination between phage P2 genes and the *Eco*T38I R-M system.

In addition to *E. coli* TH38, it has been shown that *E. coli* K-12 derivatives, such as MG1655, C600r−m+, K-207, LG102, and the W3110 strain Kohara clone (19), contain a 639-nucleotide cryptic remnant of P2 at a site with a sequence similar to that of *loc*I (2). The P2 remnant consists of the C-terminal part of the *D* gene, the complete *ogr* gene, and *att*R. This finding suggests that an ancestor of *E. coli* K-12 was lysogenized by phage P2 and that an imprecise excision event that removed most of the phage DNA then occurred, leaving only part of the *D* gene, the complete *ogr* gene, and *att*R on the chromosome.

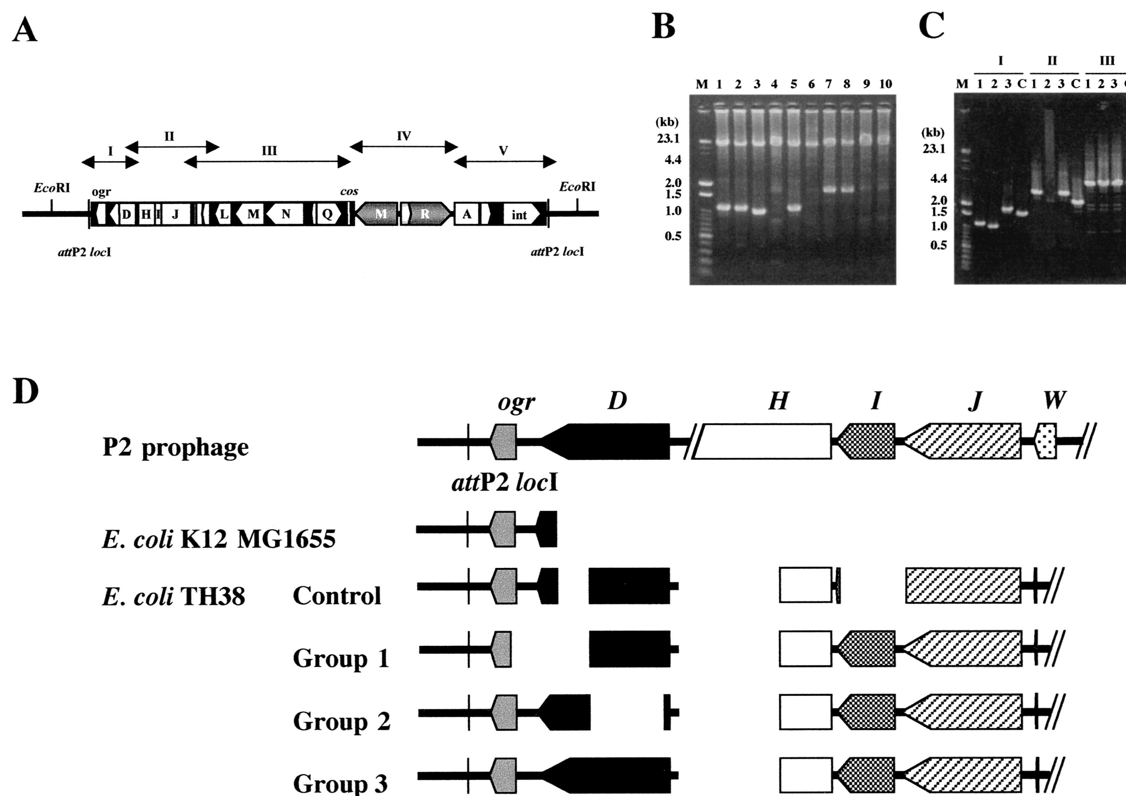

FIG. 6. Variability of defective P2 prophage genomes in *E. coli* TH38 DNA. (A) Locations of target DNAs for PCR amplification. (B) Agarose gel electrophoresis of PCR products obtained from *E. coli* TH38 strains with primers I61 and I62. The PCR mixture (50 μl) comprised 25 pmol of each primer, 400 μM each deoxynucleotide triphosphate, 2.5 mM $MgCl_2$, LA-PCR buffer (Mg^{2+} free), template DNA, and 2.5 U of *LA Taq* DNA polymerase. The reaction mixture was overlaid with mineral oil, and the reaction was carried out with a Perkin-Elmer Cetus thermal cycler. The initial template denaturation step consisted of 1 min at 94°C. The amplification profile (20 s at 98°C and 25 min at 68°C) was repeated for 30 cycles. Lanes 1 to 10 correspond to products from 10 *E. coli* TH38 colonies picked randomly. A mixture of λ *Hin*dIII digests and 100-bp DNA ladder markers (Toyobo) was used for molecular size calibration (lane M). The PCR products were subjected to 1% agarose gel electrophoresis and then stained with ethidium bromide. (C) Agarose gel electrophoresis of *E. coli* DNA fragments amplified by PCR with primers A and B (I), C and D (II), and E and F (III). The PCR mixture (50 μl) comprised 10 pmol of each primer, 200 μM each deoxynucleotide triphosphate, 2.5 mM $MgCl_2$, LA-PCR buffer (Mg^{2+} free), template DNA, and 2.5 U of *LA Taq* DNA polymerase. The reaction mixture was overlaid with mineral oil, and the reaction was carried out with a Perkin-Elmer Cetus thermal cycler. The initial template denaturation step consisted of 1 min at 94°C. The amplification profile (30 s at 94°C, 1 min at 60°C, and 5 min at 72°C) was repeated for 25 cycles. Lanes C correspond to the products from the *E. coli* TH38 control strain, for which the complete nucleotide sequence of the 16-kb *Eco*RI region was determined in this study. Lanes 1 to 3 correspond to products with different molecular sizes from the *E. coli* TH38 strains analyzed in panel B. (D) Schematic diagram of the defective P2 prophage genome in various *E. coli* TH38 strains. Groups 1 to 3 correspond to lanes 1 to 3 in panel C.

We found four types of variants in this region of the *E. coli* TH38 chromosome. The region from *ogr* through the *W* gene in the P2 prophage appeared to be a hot spot for the excision event in *E. coli* TH38. With respect to the region from *ogr* to the *J* gene, the genomic DNA of the control strain suffered from at least one more genetic recombination event than that of the other *E. coli* TH38 strains. Further investigation may provide additional insight into the connection between the presence of the R-M system and the stability of the defective prophage in *E. coli* TH38 strains. The P2 *int* transcript starts at the Pc promoter, located upstream of the *C* repressor gene (2). The −35 and −10 sequences of the Pc promoter and the amino acid sequence deduced from the sequence of the *int* gene on *E. coli* TH38 chromosomal DNA were completely conserved. These results suggested that P2 integrase was expressed in *E. coli* TH38.

In this report, we have provided the first evidence of horizontal transfer of the R-M system by phage P2. In an earlier study, Kita et al. provided evidence that the *Eco*O109I R-M system was inserted between the genes from the P4 prophage on *E. coli* H709c chromosomal DNA (16). In addition to the P4 phage, other prophages are also known to carry type II R-M systems. For example, the *Hin*dIII R-M system was found on a cryptic prophage, ϕflu, in *Haemophilus influenzae* Rd (10); the *Bsu*MI R-M system was found in the prophage 3 region in *Bacillus subtilis* Marburg 168 (28); and the *Sau*42I R-M system was found on the ϕ42 prophage in *Staphylococcus aureus* 42CR3-L (18). *Eco*prrI type I R-M (32) and *Eco*P1 type III (11) systems were found on a P1 prophage in *E. coli*. These systems are assumed to be transferred to the chromosomal DNA by the corresponding phage. It is quite reasonable from the standpoint of the physiological role of the R-M system that there is no target sequence for R.*Eco*T38I in the phage P2 genome, as shown in the *Eco*O109I R-M system. Although more than 150 type II R-M genes have been cloned and their nucleotide sequences have been analyzed, DNA sequencing

has been limited to structural genes. Sequencing of neighboring regions of the R-M system and a variety of bacterial genomes will provide evidence supporting the proposed phage-mediated mobility of the R-M system.

ACKNOWLEDGMENTS

E. coli TH38 was obtained from Nippon Gene. We are greatly indebted to Eiji Namba and Kaori Adachi, Gene Research Center, Tottori University, for the DNA sequence analysis and to Nobuhiro Mori, The United Graduate School of Agricultural Science, Tottori University, for the amino acid sequence analysis. We are also grateful to Dhruba K. Chattoraj for many helpful suggestions.

REFERENCES

1. **Anton, B. P., D. F. Heiter, J. S. Benner, E. J. Hess, L. Greenough, L. S. Moran, B. E. Slatko, and J. E. Brooks.** 1997. Cloning and characterization of the *Bgl*II restriction-modification system reveals a possible evolutionary footprint. Gene **187:**19–27.
2. **Barreiro, V., and E. Haggard-Ljungquist.** 1992. Attachment sites for bacteriophage P2 on the *Escherichia coli* chromosome: DNA sequences, localization on the physical map, and detection of a P2-like remnant in *E. coli* K-12 derivatives. J. Bacteriol. **174:**4086–4093.
3. **Bertani, G., and E. Six.** 1958. Inheritance of prophage P2 in bacterial crosses. Virology **6:**357–381.
4. **Blattner, F. R., G. Plunkett III, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Man, and Y. Shao.** 1997. The complete genome sequence of *Escherichia coli* K-12. Science **277:**1453–1461.
5. **Bougueleret, L., M. Schwarzstein, A. Tsugita, and M. Zabeau.** 1984. Characterization of the genes coding for the *Eco*RV restriction and modification system of *Escherichia coli*. Nucleic Acids Res. **12:**3659–3676.
6. **Boyer, H. W., and D. Roulland-Dussoix.** 1969. Complementation analysis of the restriction and modification of DNA in *Escherichia coli*. J. Mol. Biol. **41:**459–472.
7. **Brassard, S., H. Paquet, and P. H. Riym.** 1992. A transposon-like sequence adjacent to the *Acc*I restriction-modification operon. Gene **157:**69–72.
8. **Chattoraj, D. K.** 1978. Strand-specific break near the origin of bacteriophage P2 DNA replication. Proc. Natl. Acad. Sci. USA **75:**1685–1689.
9. **Church, G. M., and W. Gilbert.** 1984. Genomic sequencing. Proc. Natl. Acad. Sci. USA **81:**1991–1995.
10. **Hendrix, R. W., M. C. M. Smith, R. N. Burns, M. E. Ford, and G. F. Hatfull.** 1999. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. Proc. Natl. Acad. Sci. USA **96:**2192–2197.
11. **Humbelin, M., B. Suri, D. N. Rao, D. P. Hornby, H. Eberle, T. Pripfl, S. Kenel, and T. A. Bickle.** 1988. Type III DNA restriction and modification systems *Eco*P1 and *Eco*P15. Nucleotide sequences of the *Eco*P1 gene, the *Eco*P15 mod gene and some *Eco*P1 mod mutants. J. Mol. Biol. **200:**23–29.
12. **Inoue, S., M. Sunshine, E. W. Six, and M. Inoue.** 1991. Retronphage ϕR73: an *E. coli* phage that contains a retroelement and integrates into a tRNA gene. Science **256:**969–971.
13. **Ives, C. L., P. D. Nathan, and J. E. Brooks.**1992. Regulation of the *Bam*HI restriction-modification system by a small intergenic open reading frame, *bamHIC*, in both *Escherichia coli* and *Bacillus subtilis*. J. Bacteriol. **174:**7194–7201.
14. **Iwabuchi, M., S. Tajima, T. Inoue, T. Shibata, and T. Ando.** 1992. A restriction enzyme, *Bpu*I, is an isoschizomer of *Ban*II. Nucleic Acids Res. **20:**5850.
15. **Karreman, C., and A. de Waard.** 1988. Cloning and complete nucleotide sequences of the type II restriction-modification genes of *Salmonella infantis*. J. Bacteriol. **170:**2527–2532.
16. **Kita, K., J. Tsuda, T. Kato, K. Okamoto, H. Yanase, and M. Tanaka.** 1999. Evidence of horizontal transfer of the *Eco*O109I restriction-modification gene to *Escherichia coli* chromosomal DNA. J. Bacteriol. **181:**6822–6827.
17. **Kita, K., J. Tsuda, and S. Nakai.** 2002. C.*Eco*O109I, a regulatory protein for production of *Eco*O109I restriction endonuclease, specifically binds to and bends DNA upstream of its translational start site. Nucleic Acids Res. **30:**3558–3565.
18. **Kobayashi, I.** 2001. Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. Nucleic Acids Res. **29:**3742–3756.
19. **Kohara, Y., K. Akiyama, and K. Isono.** 1987. The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. Cell **50:**495–508.
20. **Kossykh, V., A. Repyk, A. Kaliman, and Y. Buuryanov.** 1989. Nucleotide sequence of the *Eco*RII restriction endonuclease gene. Biochim. Biophys. Acta **1009:**290–292.
21. **Lee, K. F., P. C. Shaw, S. J. Picone, G. G. Wilson, and K. D. Lunnen.** 1998. Sequence comparison of the *Eco*HK31I and *Eae*I restriction-modification systems suggests an intergenic transfer of genetic material. Biol. Chem. **379:**437–441.
22. **Matsudaira, P.** 1987. Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. J. Biol. Chem. **262:**10035–10038.
23. **Mise, K., K. Nakajima, N. Terakado, and M. Ishidate.** 1986. Production of restriction endonucleases using multicopy Hsd plasmids occurring naturally in pathogenic *Escherichia coli* and *Shigella boydii*. Gene **44:**165–169.
24. **Naderer, M., J. R. Brust, D. Knowle, and R. M. Blumenthal.** 2002. Mobility of a restriction-modification system revealed by its genetic context in three hosts. J. Bacteriol. **184:**2411–2419.
25. **Nakayama, Y., and I. Kobayashi.** 1998. Restriction-modification gene complexes as selfish entities: role of a regulatory system in their establishment, maintenance, and apoptotic mutual exclusion. Proc. Natl. Acad. Sci. USA **95:**6442–6447.
26. **Nelson, M., C. Christ, and I. Schildkraut.** 1984. Alteration of apparent restriction endonuclease recognition specificities by DNA methylases. Nucleic Acids Res. **12:**5165–5173.
27. **Newman, A. K., R. A. Rubin, S.-H. Kim, and P. Modrich.** 1981. DNA sequences of the structural genes for *Eco*RI DNA restriction and modification enzymes. J. Biol. Chem. **256:**2131–2139.
28. **Ohshima, H., S. Matsuoka, K. Asai, and Y. Sadaie.** 2002. Molecular organization of intrinsic restriction and modification genes *Bsu*M of *Bacillus subtilis* Marburg. J. Bacteriol. **184:**381–389.
29. **Posfai, J., A. S. Bhagwat, G. Posfai, and R. J. Roberts.** 1989. Predictive motifs derived from cytosine methyltransferases. Nucleic Acids Res. **17:**2421–2435.
30. **Roberts, R. J., and D. Macelis.** 2001. REBASE—restriction enzymes and methylases. Nucleic Acids Res. **29:**268–269.
31. **Tao, T., J. C. Bourne, and R. M. Blumenthal.** 1991. A family of regulatory genes associated with type II restriction-modification systems. J. Bacteriol. **173:**1367–1375.
32. **Tyndall, C., H. Hehnherr, U. Sandmeier, E. Kulik, and T. A. Bickle.** 1997. The type IC *hsd* loci of the enterobacteria are flanked by DNA with high homology to the phage genome: implications for the evolution and spread of DNA restriction systems. Mol. Microbiol. **23:**729–736.
33. **Vaisvilla, R., G. Vilkaitis, and A. Janulaitis.** 1995. Identification of a gene encoding a DNA invertase-like enzyme adjacent to the *Pae*R7I restriction-modification system. Gene **157:**81–84.
34. **Vijesurier, R. M., L. Carlock, R. M. Blumenthal, and J. C. Dunbar.** 2000. Role and mechanism of action of C.*Pvu*II, a regulatory protein conserved among restriction-modification systems. J. Bacteriol. **182:**477–487.
35. **Wilson, G., and N. E. Murray.** 1991. Restriction and modification systems. Annu. Rev. Genet. **25:**585–627.
36. **Xu, S.-Y., J.-P. Xiao, L. Ettwiller, M. Holden, J. Aliotta, C. J. Poh, M. Dalton, D. P. Robinson, T. R. Petronzio, L. Moran, M. Ganatra, J. Ware, B. Slatko, J. Benner, B. Slatko, and E. P. Guthrie.** 1998. Cloning and expression of the *Apa*LI, *Nsp*I, *Nsp*HI, *Sac*I, *Sca*I and *Sap*I restriction-modification systems in *Escherichia coli*. Mol. Gen. Genet. **260:**226–231.
37. **Yanisch-Perron, C., J. Vieiera, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33:**103–119.
38. **Yoshida, Y., and K. Mise.** 1986. Occurrence of small Hsd plasmids in *Salmonella typhi*, *Shigella boydii*, and *Escherichia coli*. J. Bacteriol. **165:**357–362.
39. **Zakharova, M. Z., I. V. Beletskaya, A. N. Kravetz, A. V. Pertzev, S. G. Mayorov, M. G. Shlyapnikov, and A. S. Solonin.** 1998. Cloning and sequence analysis of the plasmid-borne genes encoding the *Eco*29kI restriction and modification enzymes. Gene **208:**177–182.