Cloning and analysis of the genes encoding the type IIS restriction–modification system *HphI* from *Haemophilus parahaemolyticus*

Arvydas Lubys, Judita Lubienė, Saulius Kulakauskas⁺, Kornelijus Stankevicius, Albertas Timinskas and Arvydas Janulaitis*

Institute of Biotechnology, Graiciuno 8, 2028 Vilnius, Lithuania

Received February 28, 1996; Revised and Accepted May 28, 1996

EMBL accession no. X85374

ABSTRACT

The genomic region encoding the type IIS restriction–modification (R–M) system *HphI* (enzymes recognizing the asymmetric sequence 5′-GGTGA-3′/5′-TCACC-3′) from *Haemophilus parahaemolyticus* were cloned into *Escherichia coli* and sequenced. Sequence analysis of the R–M *HphI* system revealed three adjacent genes aligned in the same orientation: a cytosine 5 methyltransferase (gene *hphIMC*), an adenine N6 methyltransferase (*hphIMA*) and the *HphI* restriction endonuclease (gene *hphIR*). Either methyltransferase is capable of protecting plasmid DNA *in vivo* against the action of the cognate restriction endonuclease. *hphIMA* methylation renders plasmid DNA resistant to *R.HindIII* at overlapping sites, suggesting that the adenine methyltransferase modifies the 3′-terminal A residue on the GGTGA strand. Strong homology was found between the N-terminal part of the m6A methyltransferase and an unidentified reading frame interrupted by an incomplete *galE* gene of *Neisseria meningitidis*. The *HphI* R–M genes are flanked by a copy of a 56 bp direct nucleotide repeat on each side. Similar sequences have also been identified in the non-coding regions of *H.influenzae* Rd DNA. Possible involvement of the repeat sequences in the mobility of the *HphI* R–M system is discussed.

INTRODUCTION

The majority of restriction endonucleases operate within short symmetrical nucleotide sequences (1). A subclass of the type II enzymes, so-called type IIS ENases (2) recognize asymmetrical nucleotide sequences and cleave DNA at a specified distance from it. The spatial separation of the DNA recognition and cleavage sites suggests that the type IIS ENases are likely organized as two-domain enzymes, where one domain is responsible for DNA recognition and the other one carries the catalytic center that makes a double strand break in DNA. This suggestion was supported experimentally by analysis of the

domain structure of the *FokI* restriction endonuclease (3). This finding opens up new prospects for efforts to change the specificity of restriction enzymes that might be more feasible than in the case of type II restriction endonucleases, which are organized as single domain homodimeric proteins (4). However, until now the sequences of only three type IIS restriction–modification (R–M) systems, *FokI*, *MboII* and *StsI* (5–7), have been published. Here we report cloning and sequence analysis of the genomic region encoding the type IIS R–M system *HphI* from *Haemophilus parahaemolyticus* (EMBL accession no. X85374). This R–M system recognizes an asymmetrical sequence, 5′-GGTGA-3′/5′-TCACC-3′. The restriction endonuclease makes a staggered cut at the eighth base pair downstream of its recognition sequence on the upper strand, producing single base 3′ protruding ends (8,9). Earlier studies of this system reported the identification of only one type of methylation within the recognition sequence, namely modification of the inner C on the lower strand, 5′-T^{m5}CACC-3′ (10). Our results provide evidence for the presence of a second methylase that modifies an A base within the complementary strand, yielding 5′-GGTG^{m6}A-3′.

MATERIALS AND METHODS

Bacterial strains, plasmids and phages

Haemophilus parahaemolyticus (ATCC no. 49700) was used as a source of DNA for cloning of the *HphI* R–M system. *Escherichia coli* strain ER2267 of genotype *e14*[–](*mcrA*[–]) *endA1 supE44 thi-1 Δ(mcrC-mrr)114::IS10 Δ(argF-lac)U169 recA1/F*⁺ *proA*⁺*B*⁺ *lacI*^q *Δ(lacZ)M15 zzf::mini-Tn10* (Km^R) was used as host for all cloning and subcloning procedures. *Citrobacter freundii* RFL9 (Ap^RCm^S) isolated in this laboratory is a wild-type strain producing the *Cfr9I* restriction endonuclease (11). It was used as host for the construction and propagation of the positive selection vector pBR-R. The T7 expression system (12) employed for expression of *R.HphI* was purchased from Novagen and includes *E.coli* BL21(DE3) (*E.coli* B strain *F*[–] *ompT* [*lon*] *hdsB* r_B[–] m_B[–] with a λ prophage DE3 carrying the T7 RNA polymerase gene) and plasmid pET-21b. Plasmids pBR322 (13), pBR329 (14), pUC19 (15) and pACYC184 (16) were used as

* To whom correspondence should be addressed

⁺Present address: Station de Recherches Laitières, Domaine de Vilvert, 78352 Jouy-en-Josas Cedex, France

vectors in cloning and subcloning experiments. pCfr9IRM2.3 (17) was used as the source of the gene of the *Cfr9I* restriction endonuclease. Phage λ_{vir} stocks were prepared according to Sambrook *et al.* (18).

Media and transformation

Haemophilus parahaemolyticus cells were grown as described by Kleid *et al.* (8). *Escherichia coli* and *C. freundii* cells were grown in LB medium containing ampicillin (Ap, 60 μ g/ml), kanamycin (Km, 50 μ g/ml) and chloramphenicol (Cm, 30 μ g/ml) as required. Cells were transformed using the CaCl_2 -heat shock method (18). Transformants were selected by plating onto LB agar supplemented with appropriate antibiotics.

Enzymes and chemicals

Restriction enzymes, T4 DNA ligase, DNA polymerase I large fragment (Klenow), Bal31, bacterial alkaline phosphatase (BAP), the ExoIII/S1 Deletion Kit, the DNA Labeling Kit (version 2.0), *Hind*III or *Eco*RI-digested and BAP-dephosphorylated pBR322 and oligonucleotides were products of MBI Fermentas. The DNA Sequencing Kit was purchased from Pharmacia. All enzymes and kits were used according to manufacturers' recommendations. [α - ^{32}P]dATP was obtained from Izotop (St Petersburg, Russia).

DNA preparation and manipulation

Haemophilus parahaemolyticus genomic DNA was extracted and purified as described by Marmur (19). Plasmid DNAs were prepared by the alkaline lysis procedure (20) and were further purified by binding to glass powder (21). Restriction and deletion mapping, agarose gel electrophoresis, isolation of individual DNA restriction fragments from agarose gels, subcloning of DNA fragments and Bal31 deletions were carried out by standard procedures (18). ExoIII/S1 deletions were performed following the manufacturer's recommendations.

Construction of the positive selection cloning vector pBR-R

The 1.2 kb *Bst*XI-*Bam*HI fragment of pCfr9IRM2.3 (17), containing the *cfr9IR* gene, was blunt-ended with Klenow fragment and purified by electrophoresis in an agarose gel. It was ligated into the pBR329 vector, which was digested with *Eco*32I and *Eco*88I, blunt-ended with Klenow fragment and dephosphorylated with BAP. The ligation mix was used to transform competent *C. freundii* RFL9 cells. Plasmids were isolated from Cm^R transformants and screened with restriction endonucleases for inserts of appropriate size and orientation.

Construction and analysis of the *H. parahaemolyticus* library

Haemophilus parahaemolyticus genomic DNA was fragmented by sonication. The sonication products were fractionated on a low melting point agarose gel and 1–8 kb fragments were isolated. Aliquots of 15 μ g of the recovered DNA fragments were blunt-ended with Klenow fragment and ligated with T4 DNA ligase at 8°C for 24 h in a 100 μ l reaction volume to 5 μ g *Eco*47III-cleaved, dephosphorylated pBR-R vector DNA. The ligation mixture was used to transform competent *E. coli* ER2267 cells. Plasmid DNA (10 μ g) isolated from the pooled 10 000 Ap^RKm^R transformants was digested with excess *Hph*I (30 U for

Figure 1. Restriction maps of the cloning vector pBR-R (A) and recombinant plasmid pHphIM2.0 (B). The thin line represents DNA from pBR329, the bold line DNA subcloned from pCfr9IRM2.3 (17). The arrows in the circular map indicate genes for the *Cfr9I* ENase (*cfr9IR*) and ampicillin and chloramphenicol resistance (Ap^R and Cm^R respectively). ori, origin of replication. Direction of transcription from the P_{tet} promoter is shown as a bold arrow. (B) Plasmid clone of *hphIMA*. The *H. parahaemolyticus* DNA fragment (shaded line) encoding the *Hph*I m6A MTase (*hphIMA*, marked by an arrow) and part of the *Hph*I m5C MTase gene (*hphIMC*, marked by a dotted arrow) has been cloned into pBR-R at the *Eco*47III site.

3 h) and then transformed back into ER2267. Plasmid DNA of the resulting 19 individual transformants was then screened for resistance to *R.Hph*I digestion.

Mapping and cloning of the *H. parahaemolyticus* genomic R-M locus

Electrophoresis of *H. parahaemolyticus* DNA fragments obtained after digestion with various restriction enzymes (single and double digests) on agarose gels and Southern transfer were performed as described (18). A DNA Labeling Kit was used for the preparation of two radioactive DNA probes (1.05 kb, *Eco*32I-*Eco*RI and 1 kb, *Eco*RI-*Bgl*II fragments) from the previously cloned 2 kb fragment which encodes the functionally active *Hph*I methyltransferase (MTase) (Fig. 1B). Fragments overlapping with the 2 kb insert were identified and used to construct a partial genomic map. The cloned *hphIM* gene mapped to a 7.2 kb *Hind*III fragment of the *H. parahaemolyticus* genomic DNA, which seemed large enough to contain the complete R-M system. This fragment was cloned into pBR322 as follows. Genomic DNA was digested with *Hind*III and the reaction products resolved on an agarose gel, where 6–8 kb DNA fragments were isolated. Aliquots of 3 μ g of the fragments were ligated at 16°C for 16 h in a 50 μ l reaction volume with 3 μ g *Hind*III-cleaved and BAP-dephosphorylated pBR322. Plasmid DNA isolated from the 5000 pooled Ap^RKm^R transformants was subjected to digestion with *R.Hph*I, followed by re-transformation back into ER2267. Eighty five transformants were obtained, which were tested for resistance to phage infection *in vivo* (see below). Twenty four of these clones were picked randomly for individual screening of their plasmid DNA with the *Hph*I restriction endonuclease *in vitro*.

Analysis of the *Hph*I endonuclease and methylase activities

To determine restriction activity of individual clones *in vivo*, plating efficiency of λ_{vir} on these clones was measured and compared with that of *E. coli* ER2267 cells carrying the pBR322

vector alone (22). Screening of restriction-proficient cells was carried out by replicating the transformants onto top-layer agar containing 10^6 phage particles per plate. The endonuclease activity *in vitro* was tested by incubating serial dilutions of cell-free extracts prepared as described (23) with 1 μ g λ DNA at 37°C for 1 h in a 40 μ l reaction mixture containing 10 mM Tris-HCl, pH 7.5 at 37°C, 10 mM MgCl₂ and 0.1 mg/ml bovine serum albumin. Reaction products were analyzed by electrophoresis in 1% agarose gels. To determine the *HphI*-specific modification *in vivo*, plasmid DNA was challenged with an excess of the *HphI* restriction endonuclease followed by agarose gel electrophoresis.

Determination of the base modified by the *HphI* m6A MTase

To test for methylation by the *hphIMA* gene product, we first designed a plasmid, pBR-H, which contains the *HphI* recognition sequence (bold letters) overlapping by 1 bp with the recognition sequence of *HindIII* (boxed). The double-stranded linker

was introduced into *SalI*/*PaeI*-digested pBR322. The resulting construct was verified by DNA sequencing. To determine the effect of *HphI* methylation on cleavage of the overlapping *HindIII* site by *R.HindIII*, the plasmid pBR-H was transformed into *E.coli* ER2267 harboring plasmid pAC-HphIA3 (Fig. 2). The plasmid DNA isolated from Ap^RCm^RKm^R transformants was tested with *HphI* and *HindIII* endonucleases.

Expression system for *R.HphI*

To provide the genes for *HphI* methylases *in trans*, these genes (*hphIMA* and *hphIMC*) were subcloned into a compatible vector, pACYC184. The 2.75 kb *Psp1406I*-*VspI* fragment of pHphIMM1, containing both methylase genes, was isolated (Fig. 2) and blunt-ended by treatment with Klenow polymerase. The fragment was ligated to *Eco32I*-digested and dephosphorylated pACYC184. After transformation of *E.coli* ER2267 several of the Cm^RKm^R transformants were tested for the presence of *HphI*-specific methylation and for the desired orientation of the subcloned fragment. The resulting 7.0 kb plasmid, designated pAC-HphIMM, was introduced into the recipient cells and was expressed *in trans* during subsequent endonuclease expression experiments. Subcloning of the *hphIR* gene into the expression vector pET-21b was carried out as follows. A 1.3 kb *TaqI* fragment of pHphIMM1 was gel purified, treated with Klenow fragment and ligated to *XbaI*-digested, blunt-ended and dephosphorylated pET-21b. The ligate was used to transform *E.coli* ER2267 cells that contained the methylation-proficient plasmid pAC-HphIMM. Individual Ap^RCm^RKm^R clones were screened for the correct size and orientation of the *hphIR* gene. One isolate (pET-R1) was used in the expression experiments. Expression of the *hphIR* gene was induced by adjusting the culture [*E.coli* BL21(DE3) pre-transformed with plasmids pET-R1 and pAC-HphIMM; see Fig. 2] to 1 mM isopropyl- β -D-thiogalactoside (IPTG) at an OD₆₀₀ of ~0.5. After 3 h induction the cells were harvested by centrifugation and lysed (23). Crude cell extracts were used for determination of *R.HphI* activity.

Figure 2. Restriction endonuclease and deletion mapping of genes encoding the *HphI* R-M system. Restriction endonuclease sites are abbreviated as follows: H, *HindIII*; M, *MluI*; P, *Psp1406I*; R, *EcoRI*; T, *TaqI*; V, *VspI*. Only a portion of DNA originating from *H.parahaemolyticus* is shown. The hatched box represents the sequenced part of the cloned fragment, the hatched arrows indicate the ORFs which were truncated during the cloning or not completely sequenced. Additionally, the *EcoRI* site of pBR322 is shown in pHphIMM1 and pHphIMM2. Designations for genetic markers are: *orfX*, 3'-terminal part of the unknown ORF; *hphIMC*, gene encoding the *HphI* m5C MTase; *hphIMA*, gene encoding the *HphI* m6A MTase; *hphIR*, gene encoding the *HphI* restriction endonuclease; *menB*, gene encoding dihydroxynaphthoate synthase (not completely sequenced). Direction of transcription from the known vector promoters adjacent to the cloned fragments is indicated by bold arrows. P1 is the so-called 'anti-tet' promoter in pBR322 located upstream of the gene for Tc^R (47). Construction of plasmids pHphIMM1 and pHphIMM2 is described in Materials and Methods. Plasmids pAC-HphIA3 and pAC-HphIA4 were obtained after subcloning the 2.5 kb *Eco32I* fragment from pHphIMM2.0 into *Eco32I*-digested pACYC184. To obtain the plasmids pBR-HphIC1 and pBR-HphIC2, the 2.2 kb *EcoRI* fragment from pHphIMM2 was inserted into *EcoRI*-cleaved pBR322. R-M phenotypes were determined as described in Materials and Methods. Construction of plasmids pAC-HphIMM and pET-R1 is described (Materials and Methods). To provide the genes for *HphI* methylases *in trans*, the plasmid pAC-HphIMM was introduced into *E.coli* BL21 (DE3) before transformation with the plasmid pET-R1. The asterisk and parentheses denote plasmid provided *in trans* (pAC-HphIMM)^{*} and its phenotype (M⁺)^{*}.

DNA sequence determination

DNA sequencing was done in both directions from a series of nested deletions generated using the ExoIII/S1 or Bal31 methods (24). Sequencing reactions were carried out using a DNA Sequencing Kit, [α -³²P]dATP, M13/pUC (direct, reverse) or pBR322 (*SalI* site, ccw) standard sequencing primers and double-stranded, supercoiled plasmid DNA as templates. The reaction products were resolved by electrophoresis on wedge-shaped polyacrylamide gels.

Comparison of deduced amino acid sequences

The comparison of deduced amino acid sequences with the EMBL and GenBank databases translated in six reading frames was done using the BLAST (25) program to identify similar, potentially related sequences.

RESULTS AND DISCUSSION

Construction of the positive selection vector pBR-R

The lethality of a restriction endonuclease that is expressed without its accompanying methylase has been used as the basis for the construction of positive selection systems in several

laboratories (26,27). The pBR-R vector described in this work (Fig. 1A) permits cloning of any blunt-ended fragments (*Eco*47III site) as well as fragments generated by the restriction endonucleases *Bam*HI, *Bcl*II, *Bgl*III, *Bst*YI, *Mbo*I, *Sau*3A (*Bgl*III site), *Nla*III, *Nsp*I or *Pae*I (*Pae*I site). Our system possesses certain advantages over the other known systems: (i) propagation of this plasmid in *C.freundii* RFL9, which lacks any natural plasmids, does not require subsequent separation of a companion methylase-producing plasmid, as in the case of pKG2 (27); (ii) pBR-R is considerably smaller than the positive selection vectors pLV57 and pLV59 and it does not require a temperature shift for inactivation of a companion methylase, as do pLV57 and pLV59 (26); (iii) despite of its similar size to the vectors pKGW and pKGS (27), two antibiotic resistance genes are located on pBR-R (Ap and Cm). Expression of the *cfi9IR* gene from the constitutive P_{tet} promoter allows transformation of common *E.coli* strains and, in contrast with pKGW and pKGS, requires no addition of IPTG (27). The survival frequency of cells transformed with this plasmid was found to be 10^{-4} as compared with that of cells transformed with a control vector, pBR329. The efficiency of pBR-R as a cloning vector was demonstrated by the construction of a *H.parahaemolyticus* library using DNA fragments generated by sonication.

Cloning of the genes encoding *Hph*I methyltransferases

Selection of the gene coding for M.*Hph*I was based on the resistance of self-modifying recombinant plasmids to digestion by R.*Hph*I (28). Nineteen transformants obtained after the selection procedure described in Materials and Methods were further screened for the presence of the *Hph*I-specific modification. Eight modification-positive clones contained the same plasmid with a 2 kb insert (designated pHphIM2.0; Fig. 1B). None of these clones showed any *Hph*I endonuclease activity as assayed both *in vivo* and *in vitro*. The 2.5 kb *Eco*32I fragment from pHphIM2.0 was subcloned into pUC19 for sequence determination (not shown). Sequence analysis revealed the presence of a complete ORF potentially encoding a m6A methylase. This conclusion was based on the analysis of conserved amino acid motifs typical for DNA MTases (29). The ORF corresponding to the presumed m6A methylase is flanked by two truncated ORFs. The upstream ORF (Fig. 1B) contained conserved motifs VI–X characteristic of m5C methylases (30), while the downstream ORF did not show any similarities with known protein sequences. As restriction endonucleases usually share little conservation, we assumed that the latter ORF might encode the N-terminal part of the *Hph*I ENase. It therefore seemed likely that the entire *Hph*I R-M system is located in DNA regions extending in both directions around the cloned fragment.

Cloning of the entire R-M system

To select for a DNA fragment large enough to include the entire m5C methylase and *Hph*I restriction endonuclease genes a restriction map of the *H.parahaemolyticus* genomic R-M locus was determined (not shown). Two DNA fragments prepared from pHphIM2.0 (1.05 kb *Eco*32I–*Eco*RI and 1 kb *Eco*RI–*Bgl*III) were used as probes in Southern hybridization. The 7.2 kb genomic *Hind*III fragment was cut with R.*Eco*RI into two pieces. The 2.2 kb subfragment hybridized with the 1.05 kb probe, while the 5.0 kb subfragment hybridized with the 1 kb probe (not shown).

This indicated that the *Hind*III fragment includes the cloned 2 kb region of pHphIM2.0 and flanking sequences. Cloning of the 7.2 kb *Hind*III fragment and selection of R.*Hph*I-resistant clones was carried out as described in Materials and Methods. Eighty five transformants were analyzed, but none of them demonstrated resistance to λ_{vir} infection. Nine out of 24 clones tested were resistant to cleavage *in vitro* with R.*Hph*I. Restriction mapping revealed the presence of two types of recombinant plasmids that differed in the orientation of the same insert (designated pHphIMM1 and pHphIMM2; Fig. 2). These clones were tested for endonuclease activity both *in vivo*, using the plating efficiency test (Materials and Methods) and *in vitro*. In neither case was endonuclease activity detected, indicating that the R.*Hph*I gene was either incomplete or inactive in *E.coli* cells. To analyze the regions surrounding the MTase genes, sequencing of the 4.8 kb *Hind*III–*Mlu*I fragment from pHphIMM1 and pHphIMM2 (Fig. 2) was carried out.

Gene structure of the *Hph*I system and flanking DNA sequences

A 4790 bp region of the cloned 7.2 kb *Hind*III DNA fragment encompassing the genes for *Hph*I MTases as well as adjacent regions was sequenced on both strands. Five ORFs (three complete, one truncated and one partially sequenced), all oriented in the same direction, were identified (Fig. 2). The first ORF (ORFX), 258 bp long (nt 3–260, termination codon included) is truncated and may represent the 3'-terminal part of an unknown gene. The next three ORFs encode the *Hph*I R-M enzymes: ORF2 (*hphIMC*), m5C MTase; ORF3 (*hphIMA*), m6A MTase; ORF4 (*hphIR*), *Hph*I restriction endonuclease (see next paragraph). The position of a translation initiation codon in *hphIMC* could not be unambiguously inferred from the nucleotide sequence alone due to the presence of four potential initiation sites upstream of the first conserved motif of the m5C MTase (30) (ATG at position 562, ATG at position 634, TTG at position 640 and ATG at position 688); none of the start codons is preceded by a putative Shine–Dalgarno sequence either. This ORF (*hphIMC*) ends with translation termination codon TAA at position 1680. ORF3 (*hphIMA*) is 1011 bp long and extends from nt 1673 to 2683 (336 amino acids). This ORF overlaps by 8 nt with the preceding ORF and by 1 nt with the downstream one. The fourth ORF, 1140 bp long (nt 2683–3822, the *hphIR* gene), encodes the *Hph*I ENase of 379 amino acid residues. The last ORF (*menB*) was not completely sequenced; it extends from nt 3983 to 4789 and represents the 5'-terminal part of the *menB* gene (see below). A putative Shine–Dalgarno sequence was found only upstream of the last gene, *menB* (AAGGA at positions 3969–3973).

Further analysis of the sequence revealed two copies of a 56 bp stretch that flanked the *Hph*I R-M system on both sides (Fig. 3). Restriction–modification systems, although not essential for cell viability, are widespread in bacteria (1). Dissimilar systems are often present in related bacterial strains and, *vice versa*, homologous systems can be found in taxonomically distant microorganisms (17,31,32). These observations suggest that many R-M genes may have enjoyed a certain degree of mobility, at least in the recent evolutionary past. Our finding of a 56 bp long direct repeat flanking the *Hph*I R-M genes may be important in this context. Its presence suggests models for R-M mobility involving recombination of a circular DNA containing the 56 bp sequence and the *Hph*I R-M gene cassette with a corresponding (although

DRI	297-	AAATTTGTAAATATTCTGCAAAATTTAACCGCTTTTATTTCACAGCGTACAAAA-352
DRII	3879-	AAACTTGC AAATATTCTGCAAAATTTGACCGCTTTTATTTCACAGCGTACAAAA-3934
U32824	445-	GAA GTGTAAATATTCTGCAAAATTTAACCGCTTATTTT-485
U32827	4160-	AAATTTGT AAATATTCTGCAAAATTTGACCGCTTGTAGTT-4121
CONSENSUS		RAAnTTGYRAAATATTCTGCAAAATTTACCGCTTnTnTT
		G A

Figure 3. Comparison of the direct repeats I (DRI) and II (DRII) with homologous sequences found in *H. influenzae* chromosomal fragments U32824 and U32827 (36). Nucleotides identical to those in DRI are in bold. The R in the consensus sequence stands for a purine base (A,G), Y for a pyrimidine (C,T), n for 3 nt found at this position. The U32824 sequence is located between genes encoding hypothetical protein HI1458 and sigma factor (HI1459, *algU*). The sequence for U32827 lies between genes encoding hypothetical proteins HI1505 and HI1506.

somewhat divergent) 56 bp sequence in the chromosome. One class of models invokes site-specific recombination, such as with a temperate phage (i.e. λ) (33) and with the 56 bp sequence constituting (part of) the attachment site; a related model would be one based on 'integrins', a recently discovered class of mobile DNA segments that underlie site-specific insertion of certain antibiotic resistance genes into new plasmids or transposons (34). There is no obvious similarity between the *HphI* 56 bp repeat element and the 59 bp repeats of currently known integrins however and thus any possible 'integron' involving the *HphI* R-M system would differ in specificity from the drug resistance integrins described to date. It is of course formally possible that the 56 bp repeats are not components of a site-specific recombination system. Rather, since the repeats exceed the minimum length needed for RecA protein-mediated DNA pairing (35), this suggests generalized recombination for both insertion of the *HphI* R-M genes into particular chromosomal sites or their excision and loss from the cell. Of note, two sequences highly related to the *HphI* repeats (Fig. 3) are present in intergenic regions of *H. influenzae* strain Rd, whose genome has been entirely sequenced (36).

Expression of the *HphI* R-M system in *E. coli*

The deletion and subcloning experiments were carried out to investigate the expression of *HphI* R-M genes (Fig. 2). We found that both MTases are expressed in *E. coli* and protect plasmid DNA against *R.HphI* when cloned together or subcloned separately. Moreover, a similar degree of DNA protection was detected when the plasmids carrying the MTase genes in different orientations were isolated, treated with *R.HphI* and analyzed. These findings indicate that the endogenous promoters operate efficiently in *E. coli*. No activity of the restriction endonuclease was detected in crude extracts from cells carrying the initial plasmids, pHphIMM1 and pHphIMM2. We subcloned the presumed *hphIR* gene (ORF4) into the expression vector pET-21b using cells harboring the methylase genes in *trans* as recipient. In this system, the restriction endonuclease activity was readily detectable upon induction with IPTG, indicating that translation initiation functions normally and that there are no strong transcription termination signals immediately upstream of the *hphIR* gene. Apparently, expression of this gene from its own promoter in *E. coli* is so weak that it could not be detected by the methods used.

Analysis and comparison of deduced amino acid sequences

No amino acid sequences similar to that of *R.HphI* were found in the EMBL and GenBank databases. The sequence motif PD(X)₁₇EGK

was found in the central part of *R.HphI* (amino acids 140–161), which matches well the consensus P(E/D)X_{9–18}(E/D)XK motif essential for catalytic activity and Mg²⁺ binding in the *EcoRI*, *EcoRV* and *PvuII* restriction endonucleases (4).

Similar analysis of *hphIMC* revealed the presence of all 10 motifs common to m5C MTases (30). This MTase shares the greatest degree of similarity with *DdeI* (37) MTase (28% identical and additionally 11% similar amino acids) and *AquI* (38) MTase (33% identical and 11% similar amino acids). The translation product of *hphIMA* contains conserved motifs typical for N6-methyladenine MTases. This MTase belongs to the D₁₂ class of m6A MTases (29), a group where the conserved motifs appear in the order 'F-G-G' and then 'DPPY'. Similarly to some other members of the D₁₂ group, the *HphI* m6A MTase contains a slightly scrambled version of the first motif: F-GG instead of F-G-G (29). This MTase shares the greatest, although marginal, degree of similarity with the N-terminal part of *FokI* and *SstI* MTases (21 and 20% identity was observed, which rises to 38 and 36% respectively when conservative substitutions are taken into account). In addition, strong homology was found between the N-terminal part of this *HphI* MTase and an unidentified ORF located upstream of the truncated copy of the *galE* gene in *Neisseria meningitidis* (39,40) (67% identity; Fig. 4). Curiously, the conservation is not only observed within the uncharacterized ORF, but extends downstream in the same reading frame after the termination codon. The N-terminal part of the *HphI*-like MTase of *N.meningitidis* and the C-terminal part of the truncated *GalE* protein overlap by the KMPYT pentapeptide, downstream of which no further homology is observed. This suggests that the *HphI*-like MTase in *N.meningitidis* was disrupted during duplication of the *galE* locus.

We found that the protein encoded by the first ORF matched the C-terminal part of the (p)ppGpp 3'-pyrophosphohydrolase (gene *spoT*) from *H. influenzae* Rd (36) and *E. coli* (41) and contains 29% identical and 24% functionally compatible amino acids in the case of *SpoT* from *H. influenzae* Rd and 26% identical and 29% functionally compatible amino acids with *SpoT* from *E. coli*. Such a degree of homology was insufficient to predict a function for the truncated protein, so this ORF was designated *orfX*. The last ORF hit two entries in the database, dihydroxynaphthoate synthases (gene *menB*) from *H. influenzae* Rd (36) and *E. coli* (42) (89 and 88% identity respectively). This suggests that the ORF is possibly the 5'-end of the *menB* gene in *H. parahaemolyticus*. The sequence context surrounding the *HphI* R-M system differs from those in the other three R-M systems identified in *Haemophilus* strains. Namely, *HindIII* and *HinPII* from, respectively, *H. influenzae* Rd and *H. influenzae* P1 and *HpaII* from *H. parainfluenzae*, all being found located in the neighborhood of the *valS* gene (43,44).

```

M.HphI 1 MNNPKYKVMYIGNKEKIAEWICQLFVDVRTIADVFSGGCSFSFEAKKNGYQVIANDILNINYQLALALIVNNQEILTACDVFDFSNPPK 92
      N +YPK+NYIGNKEKIA WIC+QLF DV T+ADVFSGGCSF++EAKKNGY+VI NDIL INYQ+ALALI NN E L DV IFS P
L09189 246 MNNQYFKINYIGNKEKIASWICQLFSDVDVTADVFSGGCSFAYEAKKNGYKRVITNDILAINYQIALALIIENNHTLNDDVAMIFSGSPH 518

M.HphI 93 SGFTNTNTSDVFFFECEKELDAIRANILKLNNTYKQALAFALMRRAMIRKMPYS 147
      +GFM++ Y++ F+F +E ++LD R NI KL+N YK+ALAF LMRAMIRKMPY+
L09189 519 AGFMSQRYAEKTFYHDEYQQLDL*RNKIGKLDNQYKRALAFTLMRRAMIRKMPYT 683

```

Figure 4. Comparison of the m6A MTase *HphI* with amino acid sequences derived from an unidentified genomic fragment of *N.meningitidis* (GenBank accession no. L09189) (40). Numbers represent amino acid residues in *M.HphI* or nucleotide positions in L09189 corresponding to the deduced polypeptides. Identical amino acids are printed between the two sequences, conservative substitutions are shown with 'plus' signs (+). The position of the termination codon is indicated by an asterisk (*). The pentapeptide KMPYT, which marks the start of homology between the truncated GalE protein and its native form, is in bold (40).

Specificity of the *HphI* MTases

Many sequence-specific MTases, a constituent part of the type II R-M systems, are able to modify bases of a certain single type (either A or C) positioned symmetrically in the complementary strands of a palindromic recognition sequence (10). More complex and poorly investigated are the type IIS enzymes, which interact with asymmetrical sequences. In such a case two different sequences in complementary DNA strands are recognized. There have been only a few reports where specificity of type IIS modification methylases were analyzed in detail (5,45,46). These include: a single monomeric enzyme modifying adenine residues in both strands of the target DNA (*FokI*) (5); two separate cytosine MTases (*HgaI*), each responsible for the methylation of different DNA strands (45); DNA methylases yielding m6A or m5C on complementary strands in the reaction catalyzed by a single tandemly arranged enzyme or two separate m6A and m5C MTases (46; J. Bitinaite, personal communication). The *HphI* R-M system is likely to belong to the latter type. Sequence analysis of p*HphIMM1* revealed two separate genes encoding proteins resembling, respectively, m5C and m6A DNA modification methylases. *hphIMC* is most likely responsible for the reported earlier (10) methylation of the 5'-proximal C base in the bottom strand of the recognition sequence 5'-GGTGA-3'/5'-T^m5CACC-3'. The unique 3' A residue in the upper strand seemed a plausible target base for the *HphI* m6A MTase encoded by *hphIMA*. To verify this assumption, we constructed a plasmid, pBR-H (see Materials and Methods), incorporating the overlapping *HphI*-*HindIII* site GGTGAAGCTT. It was reported that the methylated sequence 5'-m⁶AAGCTT-3' is resistant to *R.HindIII*, while this site is readily cleaved in the unmodified state (10). We found that *R.HindIII* was unable to cleave the diagnostic site when pBR-H was isolated from ER2267[pAC-*HphIM3*] cells, while the cleavage of a second *HindIII* site originating from pBR322 occurred as expected (not shown). These results indicate that N⁶-methyladenine-specific *M.HphI* modifies the 3'-terminal A base in the upper (5'-GGTGA-3') strand. Additional experiments are necessary to define whether this methylation is specific for the upper strand or whether modification of the bottom strand (5'-TCACC-3') also occurs.

ACKNOWLEDGEMENTS

We thank Elisabeth A. Raleigh for *E.coli* strain ER2267. We are also grateful to Douglas E. Berg and Saulius Klimasauskas for valuable discussions and linguistic help and to Jean-Francois Tomb for providing data on *H.influenzae* Rd U32824 and U32827 sequences.

REFERENCES

- Roberts,R.J. and Macelis,D. (1994) *Nucleic Acids Res.*, **22**, 3628-3639.
- Szybalski,W., Kim,S.C., Hasan,N. and Podhajski,A.J. (1991) *Gene*, **100**, 13-26.
- Li,L., Wu,L.P. and Chandrasegaran,S. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 4275-4279.
- Aggarwal,A.K. (1995) *Curr. Opin. Struct. Biol.*, **5**, 11-19.
- Looney,M.C., Moran,L.S., Jack,W.E., Feehery,G.R., Benner,J.S., Slatko,B.E. and Wilson,G.G. (1989) *Gene*, **80**, 193-208.
- Bocklage,H., Heeger,K. and Muller-Hill,B. (1991) *Nucleic Acids Res.*, **19**, 1007-1013.
- Kita,K., Suisha,M., Kotani,H., Yanase,H. and Kato,N. (1992) *Nucleic Acids Res.*, **20**, 4167-4172.
- Kleid,D., Humayun,Z., Jeffrey,A. and Ptashne,M. (1976) *Proc. Natl. Acad. Sci. USA*, **73**, 293-297.
- Kleid,D.G. (1980) *Methods Enzymol.*, **65**, 163-166.
- McClelland,M.C., Nelson,M. and Raschke,E. (1994) *Nucleic Acids Res.*, **22**, 3640-3659.
- Janulaitis,A., Stakenas,P., Petrusyte,M., Bitinaite,J., Klimasauskas,S. and Butkus,V. (1984) *Mol. Biol.*, **18**, 115-129 (in Russian).
- Studier,F.W., Rosenberg,A.H., Dunn,J.J. and Dubendorff,J.W. (1990) *Methods Enzymol.*, **185**, 60-89.
- 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.
- Covarrubias,L. and Bolivar,F. (1982) *Gene*, **17**, 79-89.
- Yanisch-Perron,C., Vieira,J. and Messing,J. (1985) *Gene*, **33**, 103-119.
- Chang,A.C.Y. and Cohen,S.N. (1978) *J.Bacteriol.*, **134**, 1141-1156.
- Lubys,A., Menkevicius,S., Timinskas,A., Butkus,V. and Janulaitis,A. (1994) *Gene*, **141**, 85-89.
- Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Marmur,J. (1961) *J. Mol. Biol.*, **3**, 208-218.
- Birnboim,H.C. and Doly,J. (1979) *Nucleic Acids Res.*, **7**, 1513-1523.
- Marko,M.A., Chipperfield,R., Birnboim,H.C. (1982) *Anal. Biochem.*, **121**, 382-387.
- Mann,M.B., Rao,R.N. and Smith,H.O. (1978) *Gene*, **3**, 97-112.
- Whitehead,P.R. and Brown,N.L. (1985) *Arch. Microbiol.*, **141**, 70-74.
- Chen,E.Y. and Seeburg,P.H. (1985) *DNA*, **4**, 165-170.
- Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) *J. Mol. Biol.*, **215**, 403-410.
- O'Connor,C.D. and Humphreys,G.O. (1982) *Gene*, **20**, 219-229.
- Kuhn,I., Stephenson,F.H., Boyer,H.W. and Greene,P.J. (1986) *Gene*, **44**, 253-263.
- Szomolanyi,E., Kiss,A. and Venetianer,P. (1980) *Gene*, **10**, 219-225.
- Timinskas,A., Butkus,V. and Janulaitis,A. (1995) *Gene*, **157**, 3-11.
- Pósfai,J., Bhagwat,A.S., Pósfai,G. and Roberts,R.J. (1989) *Nucleic Acids Res.*, **17**, 2421-2435.
- Wilson,G.G. and Murray,N.E. (1991) *Annu. Rev. Genet.*, **25**, 585-627.
- Withers,B.E., Ambrosio,L.A. and Dunbar,J.C. (1992) *Nucleic Acids Res.*, **20**, 6267-6273.
- Weisberg,R.A. and Landy,A. (1983) In Hendrix,R.W., Roberts,J.W., Stahl,F.W. and Weisberg,R.A.(eds) *Lambda II*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 211-250.
- Hall,R.M. and Collis,C.M. (1995) *Mol. Microbiol.*, **15**, 593-600.
- Shen,P. and Huang,H.V. (1986) *Genetics*, **112**, 441-457.
- Fleischmann,R.D. et al. (1995) *Science*, **269**, 496-512.

- 37 Sznyter, L.A., Slatko, B., Moran, L., O'Donnell, K.H. and Brooks, J.E. (1987) *Nucleic Acids Res.*, **15**, 8249–8266.
- 38 Karreman, C. and de Waard, A. (1990) *J. Bacteriol.*, **172**, 266–272.
- 39 Jennings, M.P., van der Ley, P., Wilks, K.E., Maskell, D.J., Poolman, J.T. and Moxon, E.R. (1993) *Mol. Microbiol.*, **10**, 361–369.
- 40 Hammerschmidt, S., Birkholz, C., Zähringer, U., Robertson, B.D., van Putten, J., Ebeling, O. and Frosch, M. (1994) *Mol. Microbiol.*, **11**, 885–896.
- 41 Sarubbi, E., Rudd, K.E., Xiao, H., Ikehara, K., Kalman, M. and Cashel, M. (1989) *J. Biol. Chem.*, **264**, 15074–15082.
- 42 Sharma, V., Suvarna, K., Meganathan, R. and Hudspeth, M. (unpublished) M93421.
- 43 Nwankwo, D.O., Moran, L.S., Slatko, B.E., Waite-Rees, P.A., Dorner, L.F., Benner, J.S. and Wilson, G.G. (1994) *Gene*, **150**, 75–80.
- 44 Kulakauskas, S., Barsomian, J.M., Lubys, A., Roberts, R.J. and Wilson, G.G. (1994) *Gene*, **142**, 9–15.
- 45 Sugisaki, H., Yamamoto, K. and Takanami, M. (1991) *J. Biol. Chem.*, **266**, 13952–13957.
- 46 Bitinaite, J., Maneliene, Z., Menkevicius, S., Klimasauskas, S., Butkus, V. and Janulaitis, A. (1992) *Nucleic Acids Res.*, **20**, 4981–4985.
- 47 Balbás, P., Soberón, X., Merino, E., Zurita, M., Lomeli, H., Valle, F., Flores, N. and Bolívar, F. (1986) *Gene*, **50**, 3–40.