Molecular Microbiology (2001) **41**(5), 1199–1210

Analysis of type I restriction modification systems in the Neisseriaceae: genetic organization and properties of the gene products

Andrzej Piekarowicz,[1]* Aneta Kłyż,[1] Agnieszka Kwiatek[1] and Daniel C. Stein[2]*
[1]*Institute of Microbiology, University of Warsaw, 02-096 Warsaw, Poland.*
[2]*Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, MD 20742, USA.*

Summary

The *hsd* locus (host specificity of DNA) was identified in the *Neisseria gonorrhoeae* genome. The DNA fragment encoding this locus produced an active restriction and modification (R/M) system when cloned into *Escherichia coli*. This R/M system was designated *Ngo*AV. The cloned genomic fragment (7800 bp) has the potential to encode seven open reading frames (ORFs). Several of these ORFs had significant homology with other proteins found in the databases: ORF1, the *hsdM*, a methylase subunit (HsdM); ORF2, a homologue of *dinD*; ORF3, a homologue of *hsdS*; ORF4, a homologue of *hsdS*; and ORF5, an endonuclease subunit *hsdR*. The endonuclease and methylase subunits possessed strongest protein sequence homology to the *Eco*R124II R/M system, indicating that *Ngo*AV belongs to the type IC R/M family. Deletion analysis showed that only ORF3 imparted the sequence specificity of the RM.*Ngo*AV system, which recognizes an interrupted palindrome sequence ($GCAN_8$-TGC). The genetic structure of ORF3 (208 amino acids) is almost identical to the structure of the 5′ truncated *hsdS* genes of *Eco*DXXI or *Eco*R124II R/M systems obtained by *in vitro* manipulation. Genomic sequence analysis allowed us to identify *hsd* loci with a very high homology to RM.*Ngo*AV in two strains of *Neisseria meningitidis*. However, significant differences in the organization and structure of the *hsdS* genes in both these systems suggests that, if functional, they would possess recognition sites that differ from the gonococcus and from themselves.

Accepted 18 June, 2001. *For correspondence. E-mail ds64@umail.umd.edu; Tel. (+1) 301 405 5448; Fax (+1) 301 314 9489.

Introduction

Restriction and modification (R/M) systems are found in a wide variety of prokaryotes and are thought to protect the host bacterium from the uptake of foreign DNA (Bickle and Krüger, 1993; Noyer-Weidner and Trautner, 1993). R/M systems have been categorized on the basis of their subunit structures, cofactor requirements, substrate specificity and some other properties (Bickle and Krüger, 1993; Noyer-Weidner and Trautner, 1993). The type I R/M systems consist of three polypeptide subunits, HsdM, HsdS and HsdR (encoded by their corresponding genes), which combine to form a complex enzyme exerting DNA methylation, DNA restriction and DNA-dependent ATPase activities (Bickle and Krüger, 1993). Until recently, only a few type I R/M systems were described and characterized, and most of these systems originated from Enterobacterial species. The data summarized in the latest issue of REBASE (Roberts and Macelis, 2001; http://rebase.neb.com/rebase) indicate that DNA sequences that are predicted to encode type I R/M systems have been identified from more than 30 different bacterial species.

The type I R/M systems can be divided into four distinct families: types IA, IB, IC and ID (Murray *et al*., 1982; Fuller-Pace *et al*., 1985; Suri and Bickle, 1985; Price *et al*., 1987; Price and Bickle, 1988; Titeradge *et al*., 1996). Members within each family display allelic complementation. Type IA and IB R/M systems are chromosomally encoded and allelic (Barcus and Murray, 1995). The type IC systems may be encoded by plasmids (Bannister and Glover, 1968; Hedges and Datta, 1972; Firman *et al*., 1985; Skrzypek and Piekarowicz, 1989; Schouler *et al*., 1998) or may be chromosomally encoded (Tyndall *et al*., 1994; Sitaraman and Dybvig, 1997; Kong *et al*., 2000).

Three genes are necessary to account for all activities of type I R/M systems (reviewed by Bickle and Krüger, 1993; Murray, 2000). The genes *hsdM* and *hsdS* encode the subunits required for a functional modification methylase (MTase) and are transcribed from a single promoter. All three subunits, HsdM, HsdS and HsdR, are needed for restriction activity. The type I R/M systems characterized to date generally recognize non-palindromic sequences composed of two defined half-sites, 3–5 bp in length, separated by a non-specific spacer of defined length. A single peptide, HsdS, is responsible for recognition of both

DNA half-sites as well as for spacer length determination and for interaction with the other peptides of the enzyme complex. HsdS is composed of two independent DNA-binding domains (Fuller-Pace *et al*., 1984; 1985; Nagaraja *et al*., 1985; Fuller-Pace and Murray, 1986; Gubler *et al*., 1992). The amino-proximal domain (ARD) recognizes the 5′ half-site, and the carboxyl-proximal domain (CRD) recognizes the 3′ half-site (Chen *et al*., 1995). These variable regions are separated by amino acid segments that are conserved between members of a particular type I family (Kneale, 1994). Sequence comparisons among the type IC enzymes reveal three conserved regions, one at either end of the peptide and a third centrally located region (Meister *et al*., 1993; Tyndall *et al*., 1994). The sequence in the central conserved region is incompletely repeated at the carboxyl-terminus, with the remainder of the repeat located at the amino-terminus (MacWilliams and Bickle, 1996; Murray, 2000). The number of times the amino acid sequence TAEL is present within the central conserved region governs the length of the recognition site spacer; two repeats result in a 6 bp spacer, whereas three repeats result in a 7 bp spacer (Price *et al*., 1989; Gubler and Bickle, 1991; Gubler *et al*., 1992). This type of structure for the HsdS proteins has given rise to the proposal that the present-day *hsdS* genes arose from a duplication event (Argos, 1985) This proposal is supported by data indicating that truncated HsdS peptides, composed only of the amino-terminal or carboxyl-terminal half of the native peptide, can still function as a recognition subunit (Abadijeva *et al*., 1993; Meister *et al*., 1993). In the case of the *Eco*DXXI 5′, the native enzyme recognizes $TCAN_7RTTC$. When the carboxyl-terminal DNA-binding domain of *hsdS* is deleted, the mutant enzyme recognizes $TCAN_8TGA$ (Meister *et al*., 1993). When the 3′ region of *hsdS* is expressed, the mutant enzyme recognizes an interrupted palindrome, $GAAYN_5RTTC$ (MacWilliams and Bickle, 1996).

In this paper, we report the cloning and characterization of type I R/M systems from *Neisseria gonorrhoeae* FA1090 that shows a very high level of identity at the amino acid sequence level to the HsdM and HsdR proteins of the *Eco*R124II R/M system, but differs significantly in the structure of the *hsdS* gene.

Results

Sequence analysis and cloning of the R/M system of N. gonorrhoeae *strain FA1090*

DNA sequence analysis of the region surrounding the gonococcal genes *lsi-6* and *lsi-7* from *N. gonorrhoeae* strains MS11 and FA19 identified DNA sequences that possessed homology with *hsdM* genes from a variety of organisms, suggesting that the gonococcus possessed a type I R/M system (Levin and Stein, 1996). As the complete DNA sequence of *N. gonorrhoeae* strain FA1090 is available from GenBank, accession no. AE004969, we used this sequence to identify a group of open reading frames (ORFs) that seemed to encode a complete type I R/M system and then amplified the appropriate fragment via polymerase chain reaction (PCR; bp 2 120 613–2 128 413 of the FA1090genome). The location and orientation of these ORFs on the cloned DNA fragment is shown in Fig. 1A. Of the seven ORFs identified, five were encoded on the same DNA strand. The two ORFs encoded on opposite strands (ORFs 6 and 7) did not possess homology with any protein present in various protein databases.

ORF1, predicted to encode a protein of 513 amino acids, shares identity with the putative HsdM protein of the type I R/M system of *Klebsiella pneumoniae* (78%) (accession no. AJ010745), *Eco*prrI (75%) (accession no. X75452), *Eco*R124II (75%) (accession no. X13145), *Helicobacter pylori* strains 26659 (accession no. AE000596) and J99(accession no. AE001509) (53%) and *Xylella fastidiosa* (accession no. AE004079) (53%). This ORF contained all 10 motifs characteristic of N6-adenine MTases (Malone *et al*., 1995) and two additional motifs, GQE-7x-LARMNM and APKsKADfaF, conserved in type IC MTases (Schouler *et al*., 1998), suggesting that the *Ngo*AV R/M system belongs to the type IC family and should methylate an adenine residue.

ORF5, predicted to encode a protein of 921 amino acids, shares 73% identity with the HsdR subunit of the *Eco*R124II R/M system (accession no. X13145) and 43% identity with the putative HsdR protein of the *X. fastidiosa* R/M system (accession no. AE04079). These data strongly suggested that ORF5encodes the ENase.

Although ORFs 3and 4 are encoded in different reading frames, they both share 30% identity with the *hsdS* subunit of the *Eco*R124II or *Eco*DXXI R/M systems. However, if either of these ORFs were to encode an active protein, the protein would be much smaller than any native HsdS type IC subunit characterized to date (208 amino acids and 131 amino acids, respectively, compared with an average of 450–500 amino acids). On the other hand, ORF3 and ORF4 would encode proteins that are very similar to the structure of the truncated *hsdS* subunits of the *Eco*R124II R/M systems, in which their 5′ and 3′ truncated *hsdS* subunits obtained by *in vitro* manipulation retained biological function.

Sequence analysis of ORF3 and ORF4 allowed us to identify the presence of an analogue of the conserved central region (designated as B1 and B2 in Fig. 1B) that is incompletely repeated at the central and carboxyl-terminal ends of the subunit in the native IC *hsdS* genes, with the remainder of the repeat located at the amino-terminus

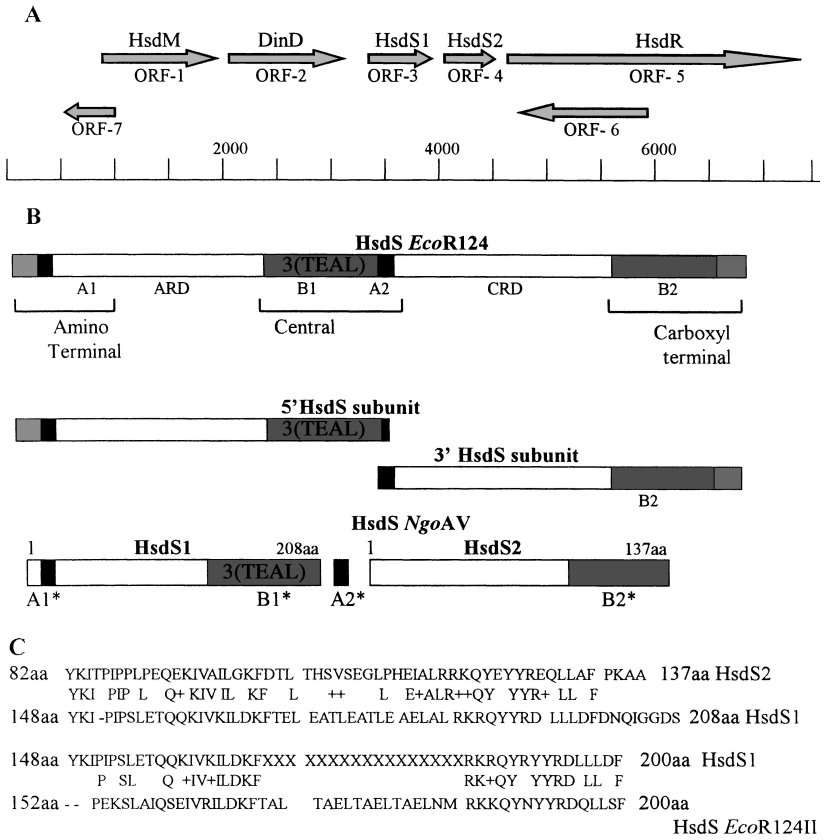

Fig. 1. Analysis of the *hsd* loci of *N. gonorrhoeae* and *E. coli*.
A. The organization of the *hsd* locus of *N. gonorrhoeae* FA1090. The lower part depicts the DNA fragment size in bp. The location of the ORFs (> 400 bp) depict reading frames that initiate with an ATG and continue to a stop codon.
B. Top. The domains found in the HsdS protein of the *Eco*R124 R/M system. The peptide can be divided into two DNA-binding domains (ARD and CRD, white areas) that are flanked by regions conserved among the IC family members. The location of the three intrafamily conserved regions (amino-terminal, central and carboxyl-terminal) are indicated. A portion of the amino-terminal region, A1, has a high degree of similarity to part of the central conserved region, A2, and the approximate location of these regions is indicated by the solid black boxes. The remainder of the central region, B1, is similar to part of the carboxyl-terminal region, B2, and is indicated by the speckled box.
Middle. Each half of the HsdS, when expressed as individual peptides, can substitute separately for the full-length HsdS protein (Meister *et al.*, 1993; MacWilliams and Bickle, 1996) (this part of the figure is adapted from MacWilliams and Bickle, 1996).
Bottom. The genomic organization of HsdS1 and HsdS2 from *N. gonorrhoeae* strain FA1090. The carboxyl-terminal ends of both parts of HsdS1 (B1*) and Hsd2 (B2*) have a high degree of similarity. A portion of the amino-terminal region A1* is similar to the region located between HsdS1 and HsdS2 (A2*). This sequence is in the same reading frame as ORF4 but before the beginning of the HsdS2 subunit.
C. The amino acid sequence alignments of the carboxyl-terminal end of HsdS1, HsdS2 and the central region of the *hsdS* gene of *Eco*R124II. The carboxyl-terminal B1* region of HsdS1 extending from 148 to 208 aligns with the carboxyl-terminal B2* region of HsdS2 from amino acid 85 to 137 and the central region of *Eco*R124II from amino acid 152 to 200. Sequence alignments were performed with the BLASTP program.

(designated as A1 and A2 in Fig. 1B; MacWilliams and Bickle, 1996; Murray, 2000). The same central region is also present at the carboxyl-terminal portion of the 5′ and 3′ truncated form of the *hsdS* genes. In ORF3 and ORF4, an equivalent of the central region is present at the carboxyl-terminal ends of both genes (designated as B1* and B2* in Fig. 1C) showing a very high level of similarity [region B1* for amino acids from 148 to 208 in *hsdS1*, and region B2* for amino acids from 82 to 137 in *hsdS2*; 95% similarity (40% identity and 55% positive)] as well as a high level of similarity between the carboxyl-terminal end of the *hsdS1* (region B1*) and the central region of the HsdS subunit of the *Eco*R124II (amino acids from 152 to 200) (Fig. 1D). The carboxyl-terminal end of ORF3 contains the sequence LEAT LEAT LEAE (or EATL EATL EAEL) that could be an equivalent of TAEL, the characteristic sequence of the type IC *hsdS* genes present in their central regions (Hadi *et al.*, 1979; Abadijeva *et al.*, 1993; Meister *et al.*, 1993; Kneale, 1994; Tyndall *et al.*, 1994; MacWilliams and Bickle, 1996). An equivalent of the conserved sequence of amino acids present in the central and amino-terminal regions of type IC *hsdS* genes (designated as A1 and A2 in Fig. 1B) is present only in the carboxyl-terminal sequence of ORF3 (Fig. 1C). It is also found in the same reading frame as ORF4 but would not be part of this ORF because of the presence of an intervening stop codon.

The differences in the structure of the *hsdS* region

of the *Ngo*AV R/M system compared with the native and truncated forms of HsdS type IC R/M systems (MacWilliams and Bickle, 1996; Murray, 2000) led us to question whether the *hsd* loci present in the *N. gonorrhoeae* genome actually encoded a biologically active R/M system. To characterize this system in further detail, the DNA sequence encoding these ORFs was amplified by PCR, and the 7800-bp-long amplicon was cloned into pBluescript KS+. The resulting recombinant transformants were tested for their ability to restrict plaque formation of phage λ. A clone was found, designated as pRMNgoAV-12, that was able to restrict λ with an efficiency of 10^{-4} (Table 1). We determined that this specificity is different from the specificity of some of the other members of the type IC family of R/M systems (*Eco*R124I, *Eco*R124II and *Eco*DXXI) because λ propagated on *E. coli* strains carrying these systems was restricted when plated on an *E. coli* strain possessing pRMNgoAV-12 (Table 1). Furthermore, propagation of λ on an *E. coli* strain possessing pRMNgoAV-12 did not protect it from restriction by *E. coli* strains expressing the *Eco*R124I, *Eco*R124II and *Eco*DXXI systems.

Identification of the genes needed for R/M function

Because ORF3 and ORF4 had homology to other type IC *hsd* genes, it suggested that both these proteins could play a role in determining the specificity of the system. In order to determine whether either or both proteins were essential for recognition specificity, as well as to verify the role of *hsdM* and *hsdR* in mediating the R/M activity seen by the clone, a series of deletion and insertion mutants was constructed and tested for R/M properties (Fig. 2). In order to determine whether either *hsdS* gene encoded an active protein, various deletions and subclones were made. The primers used for the construction of plasmid pMS5 were designed in such way that amplified DNA should encode a complete version of *hsdS1*, but lack the sequences needed to encode any of *hsdS2*. Plasmid pMS8 was constructed in such a way that the coding sequence for HsdS1 was deleted, but the ability to express HsdS2 was retained

Phage λ, propagated on an *E. coli* strain carrying the plasmid pMS5, was not restricted when plated on an *E. coli* strain containing pRMNgoAV-12, indicating that HsdS1 is able to provide the *Ngo*AV R/M system with a function that results in protection from restriction when plated on strains that express a complete system. Phage λ, grown on mutants that failed to express ORF3 because of deletions (plasmids pMS3 and pMS8), was restricted when it was tested for its ability to form plaques on a strain expressing *Ngo*AV. A similar pattern of restriction was seen when λ was propagated on *E. coli* containing plasmid pMS6, which carries a deletion in both ORFs. As neither pMS3 nor pMS8 could provide protection from host-mediated restriction by *Ngo*AV, it strongly suggested that ORF4 was not functional in the parent R/M system. This observation was confirmed by analysing the properties of the insertion mutant pRMS2, a derivative of pNgoAV-12, containing an insertion within ORF3 (in the *Eco*RV site). This mutant lacked both the MTase and ENase activities (Fig. 2). When the insertion was introduced into ORF4 (in the *Stu*I site; pRMS1 mutant), the mutant retained the MTase and ENase phenotype of the *Ngo*AV R/M system. From these experiments, we conclude that only ORF3 is needed for the recognition of the specific DNA sequence characteristic of the *Ngo*AV and for binding with HsdM and HsdR subunits, making the active HsdMS and HsdRMS complexes able to modify or restrict and modify the substrate DNA. The deletion of ORF5 (plasmid pMS2) resulted in the loss of the restriction activity without influencing the modification properties, whereas the deletion of ORF1 (pRMS3) resulted in the loss of both restriction and modification properties.

Identification of the base that is methylated by Ngo*AV*

The nature of the modified products that resulted after methylation of the DNA with HsdMS complex was determined by thin-layer chromatography (TLC). DNA was methylated with [^{3}H]-AdoMet, using a partially purified

Table 1. Restriction and modification properties of the cloned *Ngo*AV Hsd system.

	Efficiency of plating[a] of phage λ propagated on strain				
Indicator strain	DH5αmcr (pUC19)	DH5αmcr [(pNgoAV)-12]	DH5αmcr (pES14)	DH5αmcr (pUNG20)	DH5αmcr (pUNG30)
DH5αmcr(pUC19)	1	1	1	1	1
DH5αmcr(pNgoAV)-12	1×10^{-4}	1	1×10^{-4}	1×10^{-4}	1×10^{-4}
DH5αmcr(pES14)	1×10^{-4}	1×10^{-4}	1	1×10^{-4}	1×10^{-4}
DH5αmcr(pUNG20)	1×10^{-4}	1×10^{-4}	1×10^{-4}	1	1×10^{-4}
DH5αmcr(pUNG30)	1×10^{-4}	1×10^{-4}	1×10^{-4}	1×10^{-4}	1

a. Efficiency of plating: pfu ml^{-1} formed on an indicator strain divided by pfu ml^{-1} on the propagating strain.

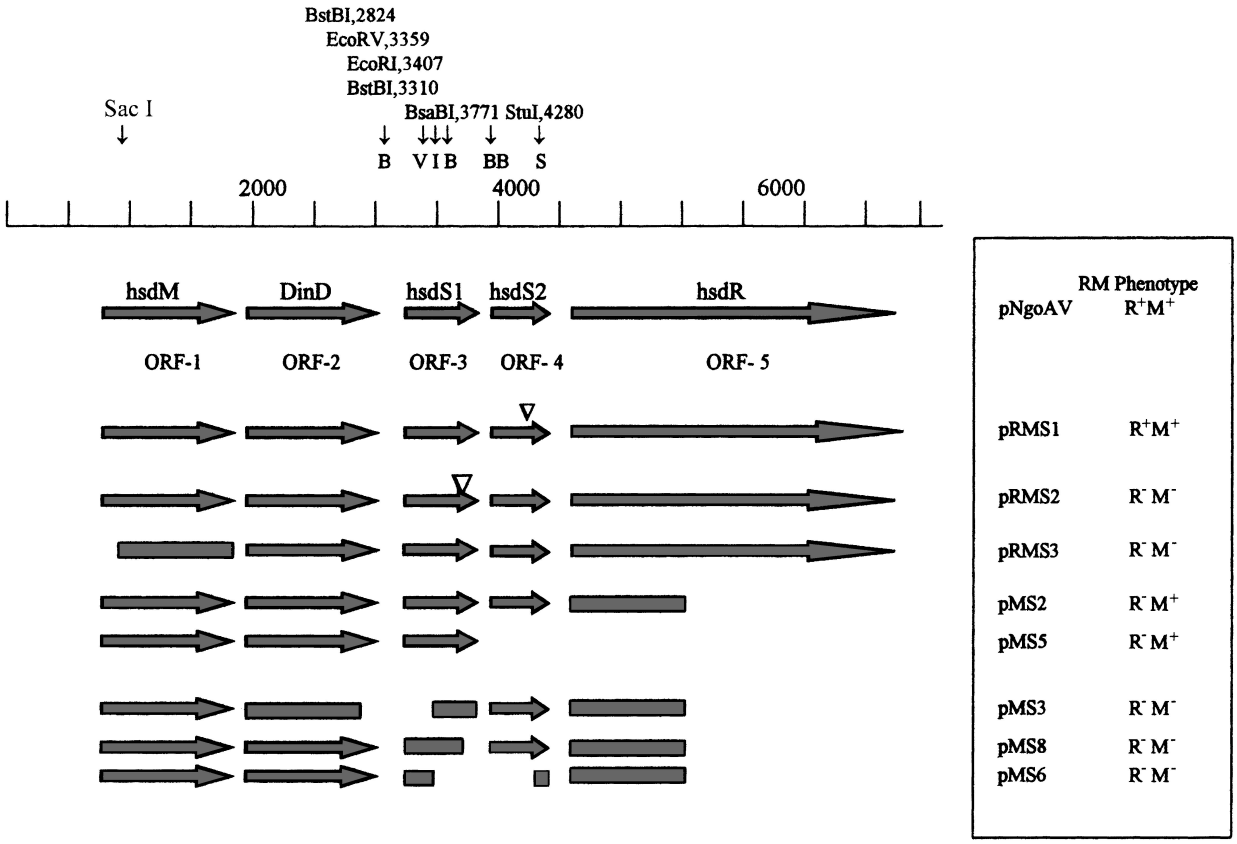

Fig. 2. R/M phenotypes of various plasmid constructs. Plasmid pNgoAV encodes a functional R/M system. The various mutants were obtained as described in *Experimental procedures*. R/M phenotypes were determined by plating unmodified or modified λ phage on particular strains carrying appropriate plasmids and measuring the plating efficiencies (EOP) of bacteriophage compared with the EOP on control strain DH5αmcr cells carrying pUC19 plasmid. Plasmid pRMS1 expresses both restriction and modification activities. Plasmids pMS2 and pMS5 express modification activity but lack restriction activity. Plasmids pRMS1, pRMS3, pM3, pMS8 and pMS6 all lack both restriction and modification activity. A triangle shows the site of insertion of the chloramphenicol resistance cassette.

enzyme isolated from an *E. coli* strain carrying the plasmid pMS2. The DNA was digested with bacterial alkaline phosphatase and snake venom phosphodiesterase, and the products were analysed by TLC. The data obtained indicated that the tritium label was present only in the position corresponding to the m6A standard (Table 2). This indicates that *Ngo*AV methylates adenine yielding N6-methyladenine.

DNA sequence recognized by R/M Ngo*AV*

As the molecular weight and the structure of the potential HsdS subunits encoded by ORF3 corresponds more to the truncated form of the known HsdS proteins, we predicted that *Ngo*AV would recognize an interrupted palindrome sequence, characteristic of this type of HsdS protein. The method used previously for determination of the specificity of MTases (Hadi *et al.*, 1979) was used to establish the specificity of the MTase activity. Plasmid DNA of known sequence, pMPMT4, was methylated *in vitro* with partially purified *Ngo*AV in the presence of [^{3}H]-AdoMet. Methylated DNA was digested with a variety of ENases and electrophoresed on a polyacrylamide gel. After drying, the gel was subjected to fluorography. An example of such an analysis is presented in Fig. 3. Digestion with a variety of ENases, combined with autoradiography, allowed us to determine that M.*Ngo*AV has at least four recognition sites

Table 2. Identification of the base methylated by M.*Ngo*AV.

	Solvent G		Solvent D	
	m6Ad	m4C + m5C	m6Ad + m4C	m5C
Relative mobility	1.3	1.1	1.0	0.9
^{3}H incorporation	2150	10	2250	20

Partially purified M.*Ngo*AV enzyme was incubated with [methyl-^{3}H]-AdoMet and pMPMT4 DNA. After methylation, the DNA was digested to mononucleotides, mixed with unlabelled methyldeoxynucleosides and separated by TLC. Two different solvents were used to distinguish all possible methylated bases. The spots visible under ultraviolet light were scraped from the TLC plate and counted for radioactivity. All radioactivity co-migrated with m6Ad. Mobility shown is relative to thymidine; 4mC and 5mC effectively co-migrate in solvent G; m6Ad and 4mC co-migrate in solvent D. Data are given in c.p.m.

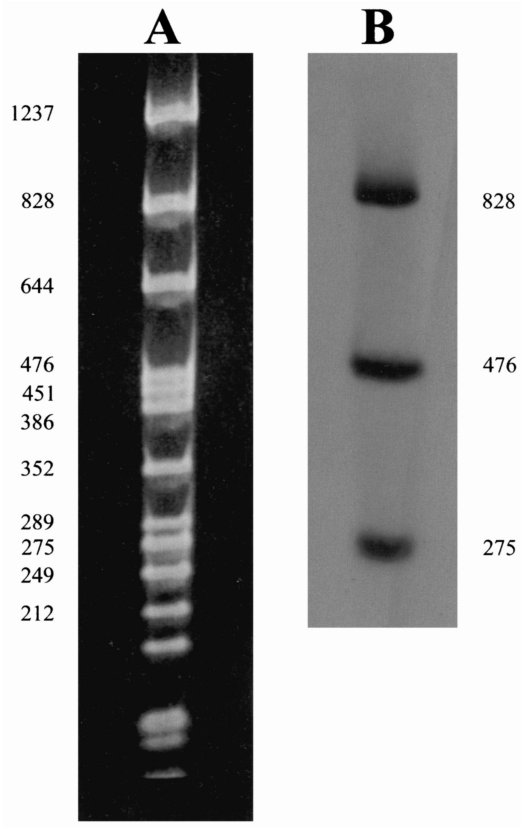

Fig. 3. Mapping of methyltransferase M.*Ngo*AV recognition sites in pMPMT4Ω DNA. Plasmid DNA was *in vitro* methylated with M.*Ngo*AV and [^{3}H]-AdoMet, digested with restriction enzymes and the resulting fragments separated by electrophoresis on a 3% polyacrylamide gel. The gel was stained with ethidium bromide, photographed, dried and fluorographed.
A. The stained gel.
B. The fluorogram for DNA digested with *Cfr*13I.

on pMPMT4, located between bp 2601 and 2700, bp 3362 and 3479, bp 3885 and 4033 and bp 6002 and 6238 (see Fig. 4). When pBR322 was methylated, label was incorporated into one site also present in pMPMT4. Neither pUC19 DNA nor ΦX174 DNA was methylated by this enzyme. Computer analysis identified the sequence 5′-GCAN$_8$TGC-3′ as being present in all methylated fragments but missing in all unlabelled fragments, suggesting that this is the sequence that is recognized by *Ngo*AV. To confirm this result, we used the fact that, in the plasmid pMS2, a plasmid that expressed the MTase *Ngo*AV activity, the sequence 5′-GCAN$_8$TGC-3′ overlaps the recognition sequence of *Sfa*NI at position 2675. This enzyme will not cleave DNA if it contains a methylated adenine in the recognition sequence 5′-GCATC/GATGC-3′ (McClelland *et al*., 1994). Plasmid pMS2 was isolated and subjected to *Sfa*NI digestions. The cleavage products were analysed by agarose gel electrophoresis. Cleavage by *Sfa*NI should generate 20 fragments, with all but three of them being smaller than 400 bp in size: the exceptions being fragments of 2941, 1052 and 914 bp. If the site at 2675 is cleaved, the resulting fragments should be 914 and 305 bp in size. Protection from *Sfa*NI cleavage at this site should result in the loss of these fragments and the appearance of a 1296 bp fragment. The data presented in Fig. 5 indicate that the 914 and 305 bp fragments are missing and that a new fragment of 1296 bp appears. These results agree with the fluorography results, indicating that *Ngo*AV recognizes the sequence 5′-GCAN$_8$TGC-3′ and methylates at least one adenine residue (the first 5′ adenine in a bottom strand).

Genomic organization of type I R/M systems in the Neisseria

The type I family of R/M systems is typically organized with the *hsdS* and *hsdM* genes occurring as part of an operon, with the *hsdR* gene forming a separate transcriptional unit (Wilson, 1991). In the *N. gonorrhoeae* strains FA1090, MS11 and FA19, the *hsdM* gene and the *hsdS* gene(s) are separated by an ORF (ORF2) that we predict is not directly connected to R/M function, as it shares 80% identity with the *dinD* gene of *E. coli* (See Fig. 1A). These data suggest that the transcriptional organization of the gonococcal type I R/M system is different from that of other systems.

As the genomic sequences of two *N. meningitidis* strains, MC58 and Z2491, are available, we performed homology searches to determine whether these strains possessed genes that would produce a type I R/M system homologous to the *Ngo*AV system. These data showed that both strains possessed *hsd* loci closely related to the *Eco*R124II R/M system. The genomic organization of these two meningococcal systems (Fig. 6) resembled that of *Ngo*AV, with some strong similarity and differences. In all three cases, the *hsdM* gene is separated from the *hsdS* gene by a gene that is not directly connected to the R/M systems. The HsdM proteins in all three loci show over 97% identity with each other. In *Ngo*AV and *N. meningitidis* strain Z2491, the DNA sequences encoding the ENase are located directly after the *hsdS* DNA sequence, whereas in *N. meningitidis* strain MC58, this sequence is separated from the *hsdS* sequence by three additional ORFs, one of which is a truncated form of *hsdR*. In *N. meningitidis* strain Z2491, the *hsdR* gene carries several stop signals, indicating that this ORF can encode only for a protein 297 amino acids long; it probably does not encode an active ENase.

In all three cases, the nucleotide sequence of the putative *hsdS* gene is built from two blocks of amino acid sequences making two separate ORFs. DNA sequence analysis of the sequences found in the databases suggests that the *hsdS1* genes from both meningococcal strains would be non-functional. *N. meningitidis* strain MC58 contains a point mutation that changes the first E of

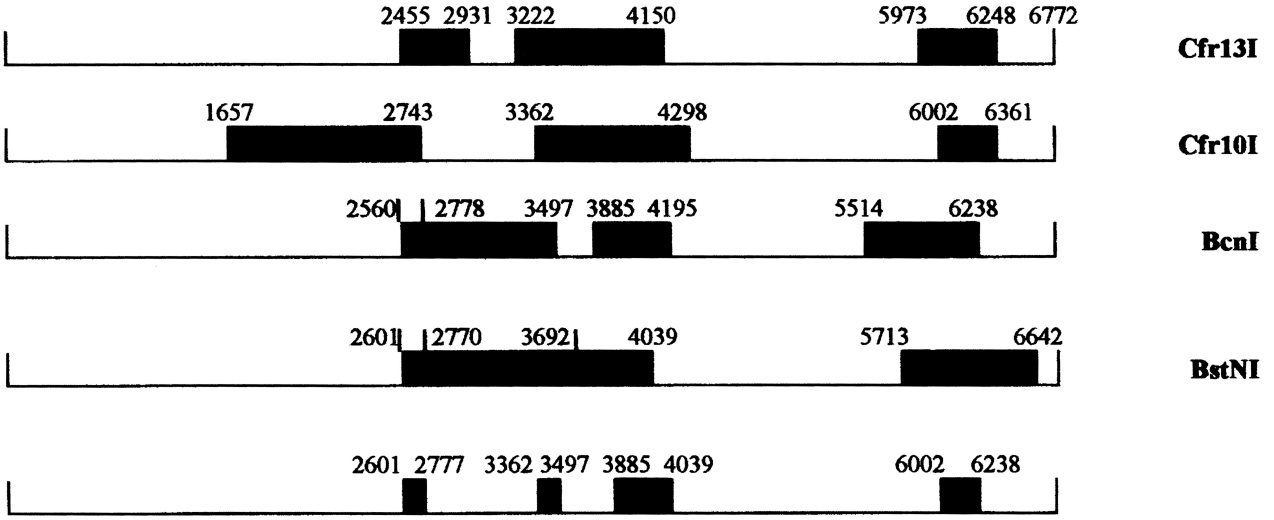

Fig. 4. Schematic representation of the regions of pMPMT4Ω (methylated by M.NgoAV). The black region indicates those restriction enzyme fragments that were methylated, whereas the white region indicates regions that were not methylated. The bottom line shows the four regions of plasmid DNA containing the *Ngo*AV recognition sites.

the LEAT repeat to a stop codon. *N. meningitidis* strain Z2491 contains a small deletion that deletes the LEAT repeated region. However, amino acid sequence alignments of the appropriate regions indicate that the *hsdS1* gene from two meningococcal strains could once have encoded a functional protein. The data provided in Fig. 7 demonstrate the degree of homology between the meningococcal and gonococcal *hsdS1* genes. These proteins possess significant homologies at the amino-terminus and carboxy-terminus, but diverge significantly in the central portion.

Discussion

The specificity of the type IC R/M systems is provided by the HsdS subunit, whose coding sequence is found within the *hsd* locus. This subunit provide two functions; recognition of the specific target sequence and binding to the HsdM and HsdR subunits to form an active R/M complex (Bickle and Krüger, 1993; Murray, 2000). Some of these functions are provided by the conserved amino acid domains located in various regions with the Hsd protein (Abadijeva *et al*., 1993; MacWilliams and Bickle, 1996). In wild-type IC R/M systems, the HsdS subunit recognition of the target sequence depends on the presence of the two variable regions (TRDs) located on both ends of the protein and the tetra amino acid repeated sequence located in the central conserved regions that governs the length of the spacer in the recognition sequence. This gives rise to a non-palindromic, interrupted recognition sequence (Meister *et al*., 1993; MacWilliams and Bickle, 1996). Analysis of truncated HsdS subunits, made by *in vitro* manipulation of the *hsdS* coding sequence or obtained as the products of transposition events, indicated that the functions of recognition of the target sequence (palindromic and interrupted) can be provided by the halves of the wild-type subunits on the condition that they contain one TRD and both the internally repeated regions A and B, as shown in Fig. 1 (Meister *et al*., 1993; MacWilliams and Bickle, 1996).

In this study, we have shown that *N. gonorrhoeae* strain FA1090 encodes a functional type I R/M system, *Ngo*AV. The specificity of the active type IC R/M system from *N. gonorrhoeae* is provided by a truncated form of HsdS subunit, similar to that seen in *in vitro*-manipulated systems. Although the region of DNA coding for the determination of specificity shows the presence of two potential *hsdS* genes (*hsdS1* and *hsdS2*), our genetic analysis indicates that only the first of these genes, *hsdS1*, encodes an active HsdS subunit. The size of this subunit (208 amino acids), the sequence structure and amino acid homology to the *Eco*R124 HsdS subunit clearly indicate that the *hsd1* gene is homologous to the truncated form of genes of *Eco*R124II or *Eco*DXXI R/M systems (Abadijeva *et al*., 1993; Meister *et al*., 1993; MacWilliams and Bickle, 1996).

The Neisserial HsdS subunit has the highest sequence homology with the type IC subunits of enteric bacteria. However, the HsdS gene of enteric bacteria possess a characteristic TAEL repeat, whereas the Neisserial gene contains the repeated sequence LEAT or EATL. In this context, it should be noted that some HsdS homologues of the type IC R/M possess different types of tandem repeats: *Mycoplasma pneumoniae* possesses a SEAL repeat (NCBI ID no. 16742248); *Haemophilus influenzae* Rd a TSEL repeat (U32706); and *H. pylori* J99 a TELN repeat (AE001504). In each of these cases, the repeated amino acids are always a tetra amino acid sequence that can

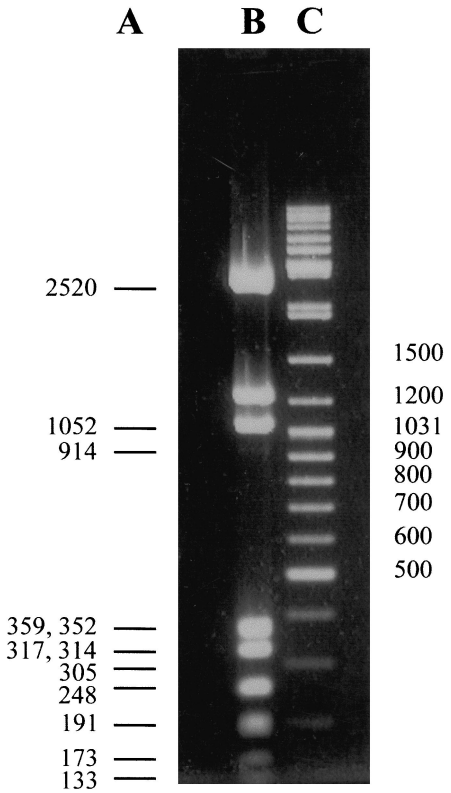

Fig. 5. Possible outcomes of *Sfa*NI restriction digestion at an overlapping recognition site. Unmodified at this site, pMS2 DNA would be cleaved by *Sfa*NI at site 2679 as well as at the flanking sites at 1765 and 2983. This would result in two fragments of 914 bp and 305 bp. If cleavage at this site is inhibited by M.*Ngo*AV methylation, a single 1219 bp fragment would result. The computer analysis indicates that unmodified pMS2 DNA cleaved by *Sfa*NI would generate 20 fragments of 2520, 1052, 914, 359, 352, 317, 314, 305, 228, 191, 173, 133, 52, 40, 36; 7 and 2 bp that allowed us to distinguish a fused fragment from all others. The theoretical digest of pMS2 is shown in lane A. To analyse the *Sfa*NI cleavage products, pMS2 DNA was digested with *Sfa*NI, and the digests were electrophoresed on a 1.8% agarose gel (lane B); Fermentas ladder molecular weight standards (lane C). Molecular weight standards (bp): 10 000, 8000, 6000, 5000, 4000, 3500, 3000, 2500, 2000, 1500, 1200, 1031, 900, 800, 700, 600, 500, 400, 300, 200 and 100 bp. The lack of *Sfa*NI 914 bp and the presence of the 1219 bp fragment confirm that this enzyme cannot cleave at site 2679 that overlaps the recognition sequence of *Ngo*AV. The arrow shows the predicted localization of the missing *Sfa*NI 914 bp fragment.

form an α-helix structure. It is not known whether any of these sequences can be present in the particular *hsdS* subunit or if the sequence depends on the whole HsdS subunit sequence.

Computer analysis of the *hsd* loci from a variety of organisms suggests that the truncation of the Hsd protein can occur naturally. The *hsd* loci of *Mycoplasma pulmonis* and *H. pylori* possess small HsdS subunits; however, no experimental evidence exists showing whether they are biologically active or not (Himmelreich *et al*., 1997; Dybvig *et al*., 1998; http://rebase.neb.com/rebase). As the gonococcal system is functional and only requires the presence of a single truncated HsdS protein, these studies provide the first experimental evidence that these naturally occurring truncations can provide a specificity function. However, the diversity in *hsdS* DNA sequences in the three *Neisseria* strains examined suggests a wide diversity of specificities within this species. Sequence analysis suggests that only the gonococcal strain actually produces a functional R/M system; MC 58 would be non-functional as a result of a point mutation in the *hsdS1* gene, and Z2491 would be non-functional because of mutations in the *hsdS1* gene and a deletion of part of the *hsdR* gene.

Although analysis of the truncated forms of the HsdS subunit of type IC R/M systems strongly suggested that the activity of the gene depends on the presence of the repeated sequence present at the amino-terminal and carboxyl-terminal ends (Abadijeva *et al*., 1993; Meister *et al*., 1993; MacWilliams and Bickle, 1996), this is not true for the HsdS1 subunit of the *Ngo*AV R/M system. There are no repeated amino acid sequences found at either end of the protein. These results show that the biological activity of the HsdS subunit only requires the presence of the central domain of the wild-type HsdS protein without any amino-terminal sequence repeated at the carboxyl-terminal end.

Our results show that the HsdS2 subunit is non-functional. The most probable reason is that it lacks specific amino acid domains needed to carry out a required function. It could be that the *hsdS* subunit activity for this ORF would be restored when the protein contained the amino acid sequence present before the start codon. The structure and the functional analysis of the cloned region coding for ORF3 and ORF4 suggest that their integrity as a presumptive wild-type HsdS subunit of type IC RM systems was lost as a result of the generation of the frameshift mutations.

It is interesting to note the significant divergence in sequence organization of the *hsd* loci for the three Neisserial strains. Although the *hsdR* and *hsdM* genes are almost identical, the *hsdS* genes only possess homology at the amino- and carboxy-termini of the proteins. As we only looked at the sequence in three strains, and the *hsdS* sequences were highly divergent, it suggests that the *Neisseria* might be an ideal genus to use to study how type I R/M systems recognize DNA.

Experimental procedures

Bacterial strains, phages and media

The *E. coli* K-12 strains DH5αmcr [*supE*44, *recA1*, *gyrA* (NalR) *thi*-1 *hsdR17* Δ*mcr*] and GM2163 (dam$^-$, dcm$^-$) were used throughout this work. These strains and their derivatives were grown at 37°C in Luria–Bertani (LB) medium (Sambrook *et al*., 1989). *N. gonorrhoeae* strain FA1090 was obtained from Dr J. Cannon, University of North Carolina, Chapel Hill (USA). Strain DH5αmcr containing the appropriate plasmids

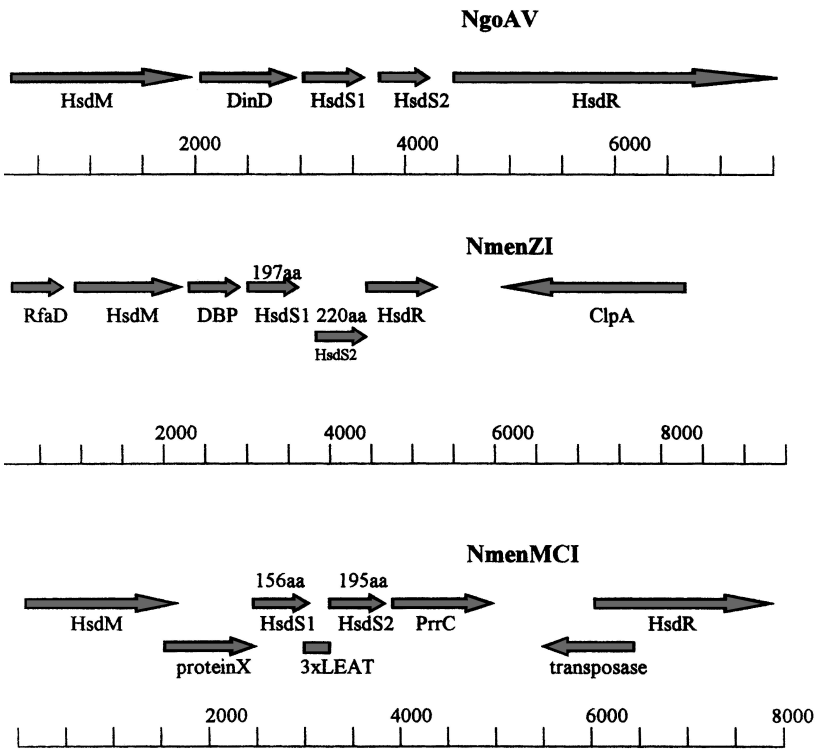

Fig. 6. Organization of the *hsd* locus of the *N. gonorrhoeae* strain FA1090 and *N. meningitidis* strains Z2491 and MC58 encoding type IC *hsd* genes. Note that the HsdS region in all three strains is encoded by two separate ORFs translated in two separate reading frames. In *N. meningitidis*, the central region containing the repeated LEAT sequence (repeated three times) is separated from the putative HsdS1 subunit by a stop codon (TAA).

was used for bacteriophage λvir *in vivo* restriction and modification assays.

Cloning of N. gonorrhoeae *DNA fragment carrying the* hsd *locus and plasmid constructions*

Oligonucleotides used to amplify the FA1090 chromosome fragment encoding the *hsd* locus were: oligo 1 (5′-TTCGATG TCAAAAGCAGGTGTCAAC-3′) and oligo 2 (5′-TTCGATGT CAAAGCAGGTGTCAACCGCTAC-3′). All PCR reactions were carried out using an Expand Long Template PCR kit (Boehringer Mannheim) using the recommended reaction conditions. Primers for PCR amplification were obtained from BioServe Biotechnologies. The resulting fragment of 7800 bp was cloned into the *Hin*dIII site of pUC19 DNA. Plasmid pMS2 was constructed by amplifying the chromosomal fragment corresponding to bp 1–4883 of the originally cloned fragment using primers oligo 1 and oligo 5 (5′-CTGGTTGATAAGCTGC ACATGGTT-3′). The subsequent amplicon was cloned into the *Hin*dIII site of pUC19 DNA. Plasmid pMS5 was constructed by amplifying the chromosome fragment corresponding to bp1–3880 of the originally cloned chromosome fragment using oligo 1 and oligo 3 (5′-TTTTCAGACGG TTTTCAGACGGCATTGATAGCGA-3′), followed by cloning into the *Hin*dIII site of pUC19. Plasmid pMS3 is a derivative of pMS2, derived by deleting the internal *Bst*BI fragment present in the cloned chromosomal DNA of pMS2. Plasmid pMS8 is a derivative of pMS2 constructed by deleting the internal *Eco*RV and *Bsa*BI fragments. Plasmid pMS6 is a derivative of pMS2 obtained by deleting the internal *Eco*RV and *Stu*I fragments. A chloramphenicol resistance cassette was inserted into pNgoAV at the *Stu*I site located at bp 4283 giving pRMS1 and into *Eco*RV located at bp 3359 giving pRMS2. Plasmid pRMS3 is a derivative of pNgoAV constructed by deleting the internal *Sac*I fragment. All routine cloning procedures were carried out in accordance with protocols described by Sambrook *et al.* (1989). Plasmids pUNG20 and pUNG30 (Firman *et al.*, 1985) were the source of the R/M systems *Eco*R124 and *Eco*R124/3 respectively. Plasmid pES14 (Skrzypek and Piekarowicz, 1989) was the source of the R/M system *Eco*DXXXI. Plasmid pMPMT4Ω (Mayer, 1995) was used for determination of the recognition site of the *Ngo*AV R/M system.

Enzymes and chemicals

Restriction enzymes and T4 DNA ligase were purchased from Fermentas. All chemicals used were reagent grade or better and were obtained from Sigma, unless otherwise noted.

Protein purification

The partial purification of M.*Ngo*AV MTase was carried out as follows: frozen cells (15 g wet weight) were thawed and suspended in 50 ml of buffer A (20 mM Tris-HCl, pH 7.5, 1 mM EDTA, 7 mM 2-mercaptoethanol). Cells were disrupted by sonication, and cellular debris was removed by centrifugation at 40 000 *g* for 1 h. The nucleic acids were precipitated by the addition of polyethylenimine solution (10%, neutralized) to a final concentration of 1% in the presence of 0.2 M NaCl. The precipitate was removed by centrifugation at 10 000 r.p.m. for 15 min. Soluble proteins were precipitated from solution by the addition of $(NH_4)_2SO_4$ (80% saturation) and removed by centrifugation. The pellet was dissolved in 50 ml of buffer B

```
FA1090      1    MDMQSKAKKLIEMIQTAPVEWKPLGEVLVRTKGTKITAGQMKEMHKDNAPLKIFAGGKTF
MC58        1    MDMQNKAKKLIEMIQTAPVEWKPLGEVAKVLRGKRLTKKELIEGGK----YPVFHG-
Z2491       1    MDMQSKAKKLIEMIQTAPVEWKPLGGENGIAIIKTGQAVSKQKISNNIGSYPVINS----

FA1090     61    ALVDFDDVPDKDIHREPS-IIVKSRGIIEFEYYDKPFSHKNEMWSYHSVNK-HIYIKYVY
MC58       53    GLIPLGWFDQFNRRANQT-MIINTGSIGEVIWSGVDFWSSDGTFVIQTPN--YLDDKFIF
Z2491      57    GKEPLGYIDEWNTENDPIGITTRGAGVGSITWQEGRYFRGNLNYAVTIKNRTELDVRFLY

FA1090    119    YFLKTQENYFRNIGSKMQMPQIATPDTDNYKIPIPSLETQQKIVKILDKFTELEATLEAT
MC58      110    YFLKTREGYIKSQKRVGGVPTIDRLVVENISIPIPPLETQQKIVKILDKFTELXATLEAT
Z2491     117    HILLEFEQEIHALCTFTGIPALNASNLKKLLIPIPPLETQQKIVKILDKFTELEA-----

FA1090    179    LEAELALRKRQYRYYRDLLLDFDNQIGGDS
MC58      170    LEAELTLRKRQYRYYRDFLLDFNNQIGGDS
Z2491     172    -ELALRKRQYRYYRDFLLDFDNQIGGG-
```

Fig. 7. Amino acid sequence alignment of *hsdS1* of *N. gonorrhoeae* strain FA1090 and *N. meningitidis* strains Z2491 and MC58. The predicted amino acid sequences were generated from the DNA sequences available from GenBank and then aligned using the program CLUSTAL. When two sequences contain an identical base, the amino acid is bold and shadowed. When two or more sequences contain a conserved amino acid, the sequence is bold.

(20 mM KPO_4, pH 7.5, 1 mM EDTA, 10 mM 2-mercapto-ethanol, 5% glycerol) and dialysed overnight against 2 l of the same buffer. The dialysate was applied to a phosphocellulose column (1.5 cm × 20 cm) that had been equilibrated with buffer B. The column was washed with 150 ml of buffer B, and the proteins were eluted with a 200 ml linear NaCl gradient (0–1.0 M) in the same buffer. Fractions (3 ml) were collected and assayed for MTase activity. MTase activity eluted between 0.4 and 0.5 M NaCl. The peak fractions were pooled and dialysed against buffer C (20 mM Tris-HCl, pH 7.5, 1 mM EDTA, 20 mM NaCl, 7-mM 2-mercaptoethanol). The dialysed fractions were applied to an Accel QMA column (1 cm × 10 cm). The column was washed with 50 ml of the same buffer and eluted with a 60 ml linear NaCl gradient (0.02 M to 1.0 M). Fractions (1.5 ml) were collected and assayed for MTase activity. The MTase activity eluted from the column at about 0.2–0.25 M. These fractions were used for all subsequent experiments.

MTase assays

DNA was incubated with the partially purified enzyme at 37°C

Table 3. Sequence comparison of *N. gonorrhoeae* strain FA1090, *N. meningitidis* strains Z2491 and MC58 Hsd proteins and their closest homologue *E. coli* EcoR124II.

N. gonorrhoeae strain FA1090	*N. meningitidis* Z2491		*N. meningitidis* MC58		*Eco*R124II	
	% Similarity	% Identity	% Similarity	% Identity	% Similarity	% Identity
HsdM	97	97	98	97	87	75
HsdR	96	95	96	95	83	73
HsdS1[a]	52	38	57	42	43	30
HsdS2[a]	97[b]	95[b]	69	58	56	40

a. Gives the percentage similarity or identity between corresponding proteins, i.e. HsdS1/HsdS2 of *N. gonorrhoeae* and HsdS1/HsdS2 of both *N. meningitidis* strains.
b. These are the values for the 40 amino acid residues of the carboxyl-terminal parts of both proteins. The remaining parts show no homology.

in buffer containing 50 mM Tris-HCl (pH 7.5), 10 mM EDTA, 10 mM 2-mercaptoethanol and [methyl-^{3}H]-AdoMet (2 μM). After 6 h of incubation at 37°C, reaction mixtures were spotted onto DE81 paper (Whatman), washed as described by Sambrook *et al*. (1989), dried and counted in a liquid scintillation counter (Instagel; Packard). In some experiments, the reaction mixtures were extracted with phenol, and the aqueous phase was passed through a Sephadex G-100 column equilibrated with 10 mM Tris-HCl (pH 7.0) buffer to separate the labelled DNA from low-molecular-mass radioactive material.

Determination of the MTase specificity

The method originally developed for the determination of type III restriction enzyme sequences (Hadi *et al.*, 1979) was used to establish the specificity of the MTase. Plasmid DNAs were methylated with [methyl-^{3}H]-AdoMet, digested with a variety of restriction enzymes, the fragments separated by gel electrophoresis, and those containing radioactivity identified by fluorography. DNA sequences unique to the labelled fragments and missing from the unlabelled fragments were identified by computer analysis developed at our Institute.

Identification of the nucleotides methylated by RM. Ngo*AV*

The methylated substrate DNA (1×10^4 c.p.m.) was digested with bacterial alkaline phosphatase (1 μl, 100 units μl^{-1}) in 50 μl of 20 mM Tris-HCl (pH 8.4), 10 mM $MgCl_2$ at 37°C for 2 h. The digests were desalted on a 10 ml Sephadex G-10 gel filtration column that was equilibrated with deionized water. The column was eluted with water, and fractions (0.5 ml) containing the ^{3}H peak were concentrated by vacuum centrifugation. This sample was divided in half, and each half was spotted onto a 20 cm × 20 cm fluorescent cellulose TLC plate (Merck). Unlabelled methylated deoxynucleosides (0.1 A_{260} unit of each) were spotted on top of the sample to allow for the location of the compound after TLC. One plate was developed in solvent D [80:20 (v/v) ethanol to water] and the second plate in solvent G [66:33:1 (by volume) isobutyric acid–water–ammonium hydroxide]. The plates were illuminated at 254 nm, and the spots in the sample lane formed by m5C, m4C and m6A were scraped and dissolved in 1 ml of water. After the addition of 10 ml of scintillation cocktail, the samples were counted in a scintillation counter.

Acknowledgements

This work was supported by KBN grant no. 6 PO4A 037 18. We acknowledge the Gonococcal Genome Sequencing Project supported by USPHS/NIH grant no. AI38399, and B. A. Roe, L. Song, S. P. Lin, X. Yuan, S. Clifton, Tom Ducey, Lisa Lewis and D. W. Dyer at the University of Oklahoma. The GenBank accession number for the completed *N. gonorrhoeae* genome is AE004969, and this link gives the GenBank sequence file containing the sequence deposited into GenBank.

References

Abadijeva, A., Webb, M., Patel, J., Zinkevich, V., and Firman, K. (1993) Deletions within the DNA recognition subunit of M.EcoR124I that identify a region involved in protein–protein interaction between HsdS and HsdM. *J Mol Biol* **241**: 403–410.

Argos, P. (1985) Evidence for repeating domain in type I restriction enzymes. *EMBO J* **4**: 1351–1355.

Bannister, D., and Glover, S.W. (1968) Restriction and modification of bacteriophages by R^+ strains of *Escherichia coli* K12. *Biochem Biophys Res Commun* **30**: 735–738.

Barcus, V.A., and Murray, N. (1995) Barriers to recombination: restriction. In *Population Genetics of Bacteria.* Baumberg, S., Young, J.P.W. Wellington, E.M.H., and Saunders, R. (eds). Cambridge: Society for General Microbiology, pp. 31–38.

Bickle, T., and Krüger, D.H. (1993) Biology of DNA restriction. *Microbiol Rev* **57**: 434–450.

Chen, A., Powell, L.M., Dryden, D.T.F., Murray, N., and Brown, T. (1995) Tyrosine 27 of the specificity polypeptide of *Eco*KI can be UV crosslinked to a bromodeoxyuridine substituted DNA target sequence. *Nucleic Acids Res* **23**: 1177–1183.

Dybvig, K., Sitaraman, R., and French, C.T. (1998) A family of phase-variable restriction enzymes with differing specificities generated by high-frequency gene rearrangements. *Proc Natl Acad Sci USA* **95**: 13923–13928.

Firman, K., Price, C., and Glover, S.W. (1985) The *Eco*R124 and *Eco*R124/3 restriction and modification systems: cloning the genes. *Plasmid* **14**: 145–153.

Fuller-Pace, F.V., and Murray, N.E. (1986) Two DNA recognition domains of the specificity polypeptides of a family of type I restriction enzymes. *Proc Natl Acad Sci USA* **83**: 9368–9372.

Fuller-Pace, F.V., Bullas, L.R., Delius, H., and Murray, N.E. (1984) Genetic recombination can generate altered restriction specificity. *Proc Natl Acad Sci USA* **81**: 6095–6099.

Fuller-Pace, F.V., Cowan, G.M., and Murray, N.E. (1985) Eco*A and* Eco*E: alternatives to the* Eco*K family of type I restriction and modification systems of* Escherichia coli. *J Mol Biol* **186**: 65–75.

Gubler, M., and Bickle, T.A. (1991) Increased protein flexibility leads to promiscuous protein–DNA interactions in type IC restriction-modification systems. *EMBO J* **10**: 951–957.

Gubler, M., Braguglia, D., Meyer, J., Piekarowicz, A., and Bickle, T.A. (1992) Recombination of constant and variable modules alters DNA sequence recognition by type IC restriction- modification enzymes. *EMBO J* **11**: 233–240.

Hadi, S.M., Bachi, B., Shepherd, J.C.V., Yuan, R., Ineichen, K., and Bickle, T.A. (1979) DNA recognition and cleavage by *Eco*P15 restriction endonuclease. *J Mol Biol* **134**: 655–666.

Hedges, R.W., and Datta, N. (1972) R124, an fi R factor of a new compatibility class. *J Gen Microbiol* **71**: 403–405.

Himmelreich, R., Plagnes, H., Hilbert, H., Reiner, B., and Herrmann, R. (1997) Comparative analysis of the genomes of the bacteria *Mycoplasma pneumoniae* and *Mycoplasma genitalium*. *Nucleic Acids Res* **25**: 701–712.

Kneale, G.G. (1994) A symmetrical model for the domain

structure of type I DNA methyltransferases. *J Mol Biol* **14**: 1–5.

Kong, H., Lin, L.F., Porter, N., Stickel, S., Byrd, D., Posfai, J., and Roberts, R. (2000) Functional analysis of putative restriction-modification system genes in the *Helicobacter pylori* J99 genome. *Nucleic Acids Res* **28**: 3216–3223.

Levin, J.C., and Stein, D.C. (1996) Cloning, complementation, and characterization of an *rfa*E homolog from *Neisseria gonorrhoeae*. *J Bacteriol* **178**: 4571–4575.

McClelland, M., Nelson, M., and Raschke, E. (1994) Effect of site-specific modification on restriction endonucleases and DNA modification methyltransferases. *Nucleic Acids Res* **22**: 3640–3659.

MacWilliams, M., and Bickle, T.A. (1996) Generation of new DNA binding specificity by truncation of the type IC *Eco*DXXI *hsdS* gene. *EMBO J* **15**: 4775–4783.

Malone, T., Blumenthal, R.M., and Cheng, X. (1995) Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyl-transferases, and suggests a catalytic mechanism for these enzymes. *J Mol Biol* **253**: 618–632.

Mayer, M.P. (1995) A new set of useful cloning and expression vectors derived from pBlueScript. *Gene* **163**: 41–46.

Meister, J., MacWilliams, M., Hubner, P., Jutte, H., Skrzypek, E., Piekarowicz, A., and Bickle, T.A. (1993) Macroevolution by transposition: drastic modification of DNA recognition by a type I restriction enzyme following Tn*5* transposition. *EMBO J* **12**: 4585–4591.

Murray, N.E. (2000) Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). *Microbiol Mol Biol Rev* **64**: 412–434.

Murray, N.E., Gough, A., Suri, B., and Bickle, T.A. (1982) Structural homologies among type I restriction-modification systems. *EMBO J* **1**: 535–539.

Nagaraja, V., Shepherd, J.C., and Bickle, T.A. (1985) A hybrid recognition sequence in a recombinant restriction enzyme and the evolution of DNA sequence specificity. *Nature* **316**: 371–372.

Noyer-Weidner, M., and Trautner, T. (1993). In Methylation of DNA in procaryotes. In *DNA Methylation: Molecular Biology and Biological Significance*. Jost, J.P., and Saluz, H.P. (eds). Basel: Birkhäuser, pp. 39–108.

Price, C., and Bickle, T.A. (1988) Evolution of DNA sequence specificity in type I restriction enzymes. *Biochem Soc Trans* **16**: 942–943.

Price, C., Pripfl, T., and Bickle, T.A. (1987) *Eco*R124 and *Eco*R124/3: the first members of a new family of type I restriction and modification systems. *Eur J Biochem* **167**: 111–115.

Price, C., Lingner, J., Bickle, T.A., Firman, K., and Glover, S.W. (1989) Basis for changes in DNA recognition by the R. *Eco*124 and *Eco*R124/3 type I DNA restriction and modification enzymes. *J Mol Biol* **205**: 115–125.

Roberts, R.J., and Macelis, D. (2001) REBASE – restriction enzymes and methylases. *Nucleic Acids Res* **29**: 268–269.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Schouler, C., Gautier, M., Ehrlich, S.D., and Chopin, M.C. (1998) Combinational variation of restriction modification specificities in *Lactococcus lactis*. *Mol Microbiol* **28**: 169–178.

Sitaraman, R., and Dybvig, K. (1997) The *hsd* loci of *Mycoplasma pulmonis*: organization, rearrangements and expression of genes. *Mol Microbiol* **26**: 109–120.

Skrzypek, E., and Piekarowicz, A. (1989) The *Eco*DXXI restriction and modification system: cloning the genes and homology to type I restriction modification systems. *Plasmid* **21**: 195–204.

Suri, B., and Bickle, T. (1985) *Eco*A: The first member of a new family of type I restriction and modification systems. *J Mol Biol* **186**: 77–85.

Titeradge, A.J., Ternent, B., and Murray, N.E. (1996) A third family of allelic *hsd* genes in *Salmonella enterica*: sequence comparisons with related protein identify conserved regions implicated in restriction of DNA. *Mol Microbiol* **22**: 437–447.

Tyndall, C., Meister, J., and Bickle, T.A. (1994) The *Escherichia coli prr* region encodes a functional type IC DNA restriction system closely integrated with an anticodon nuclease gene. *J Mol Biol* **237**: 266–274.

Wilson, G.G. (1991) Organization of restriction-modification systems. *Nucleic Acids Res* **19**: 2539–2566.