CHAPTER V

Characterization of Type I and Type II Restriction-Modification Systems of *Histophilus somni*

ABSTRACT

Restriction-modification systems are ubiquitous among eubacteria and archaea. Besides their general role in protecting the host from bacteriophage attack, these systems may also play a role in evolution and speciation. Type I, Type II, and Type III restriction-modification systems have been well characterized and exploited as tools in the laboratory for genetic and molecular analyses. In this chapter, the genomic location, genetic comparison, and functional identification of the methyltransferase components of the Type I and Type II restriction-modification systems from *H. somni* are described. The Type II restriction-modification system of *H. somni* consists of the classical endonuclease-methyltransferase components whose recognition sequence is predicted to be 5'-GCGC-3'. The Type I restriction-modification system of *H. somni* is comprised of the classical HsdR-HsdS-HsdM components whose recognition sequence remains to be identified. The genetic and evolutionary significance of these systems in *H. somni* and other members of the *Pasteurellaceae* is discussed.

Introduction

Several types of restriction-modification (RM) systems have been described from prokaryotes. According to Kobayashi (2004) they may;

(i) Be an example of the simplest unique forms of life (like transposons and viruses) whose loss leads to host cell death (the "*selfish gene hypothesis*")

(ii) Provide defense against bacteriophage attack and maintain the integrity of the bacterial genome (the "*cellular defense hypothesis*")

(iii) Generate genome diversity and drive evolution and speciation by facilitating gene loss/gain (the "*variation hypothesis*")

Type I and Type II RM systems comprise the most frequently encountered DNA modifying enzymes among eubacteria. Since the purification of the first restriction enzyme from *Haemophilus influenzae* in 1970, hundreds of these enzymes have been identified and characterized (Murray, 2000; Roberts et al., 2003).

The Type I RM systems generally consists of a restriction endonuclease (REase) subunit encoded by the *hsdR* gene, a methyltransferase (MTase) subunit encoded by the *hsdM* gene, and a specificity subunit encoded by the *hsdS* gene. In the Type I RM system, the "enzymes are multisubunit proteins that function as a single protein complex and usually contain two HsdR subunits, two HsdM subunits, and one HsdS subunit" (Roberts et al., 2003). The HsdS subunit determines the specificity of the DNA sequence to be either cut by HsdR subunit or methylated by the HsdM subunit (Roberts et al., 2003). When the Type I RM complex encounters unmethylated DNA as a substrate, the subunits usually function as an endonuclease and cut DNA at sites distant from their recognition sequence using ATP as a phosphate donor. However, when the Type I RM complex encounters hemimethylated DNA as a substrate, the subunits usually

function as a methyltransferase and methylate DNA using *S*-adenosyl-L-methionine (AdoMet) as a methyl donor (Roberts et al., 2003). The usefulness of Type I RM systems as reagents for molecular cloning is limited owing to their complex nature.

The Type II RM systems generally consist of two open reading frames (ORFs) that encode a REase and a MTase. These ORFs form a tight locus and are usually regulated by a single promoter unit. In the Type II RM system, the REase "recognizes specific DNA sequences and cleaves at constant positions at or close to that sequence to produce 5'-phosphates and 3'-hydroxyls". The Type II REases "act as monomers, dimmers or even tetramers and usually act independently of their corresponding MTases" (Roberts et al., 2003). In contrast to the Type I RM systems, Type II RM systems are very useful as reagents for molecular cloning because of their ability to cut DNA at exact sites and produce 5' and/or 3'termini that can be recombined.

Since their first discovery in *H. influenzae*, several Type I and Type II RM systems have been identified in members of the *Pasteurellaceae*. A Type II RM system (PhaI) from *Mannheimia haemolytica* has been shown to recognize the 5 bp nonpalindromic sequences 5'-GCATC-3' and 5'-GATGC-3' (Briggs et al., 1994). Two Type II RM systems, HhaI from *Haemophilus haemolyticus* and HinP1I from *H. influenzae*, have been shown to recognize the 4 bp palindromic sequence 5'-GCGC-3' (Barsomian et al., 1988). Whereas R.HhaI produces a CG-3′ extension, R.HinP1I produces a 5′-CG extension. A Type II REase (R.HsoI), which recognizes the sequence 5'-GCGC-3' and cleaves it asymmetrically to leave a 5′-CG extension, has been purified from *H. somni* strain 2336 (Briggs and Tatum, 2005). Therefore, R.HsoI appears to be an isoschizomer of R.HhaI and R.HinP1I, and enzymatically identical to R.HinP1I. Briggs and Tatum (2005) also showed that methylation of DNA using HhaI MTase (M.HhaI) affords protection against the endonuclease activity of R.HsoI.

Among *H. somni* strains, the preputial isolates (especially strain 129Pt) have been shown to be relatively easily transformed by shuttle vectors compared to the pathogenic isolates (Sanders et al., 1997). This difference in the ability of *H. somni* strains to uptake foreign DNA has been thought to be due to the differences in the restriction-modification systems among them (Sanders et al., 1997; Wu et al., 2000). Heterologous *in vivo* methylation of DNA by passing it through *H. influenzae* has been shown to facilitate DNA uptake by some pathogenic *H. somni* strains (Sanders et al., 2003). Heterologous *in vitro* methylation of DNA by treating it with M.HhaI has also been shown to facilitate DNA uptake by some pathogenic *H. somni* strains (Wu et al., 2000; Briggs and Tatum, 2005). These results are not surprising since most of the commercial cloning vectors that could be used as suicide vectors in *H. somni* as well as some shuttle vectors contain a number of 5'-GCGC-3' sequences.

Although several researchers, including Briggs and Tatum (2005), have biochemically characterized REases from members of the *Pasteurellaceae*, very little is known about the genetics and composition of Type I and Type II RM systems among different species in this bacterial family. In view of these considerations, and the importance of RM systems in cellular physiology and evolution, the objective of this study was to identify and characterize the RM systems from *H. somni*. In this chapter, the positioning of the Type I and Type II RM loci on the chromosomes of *H. somni* strains 2336 and 129Pt has been compared using genome sequence data. The loci have been cloned and the putative MTase components of the two systems have been functionally identified. The possibility of introducing plasmid DNA methylated by Type I and Type II MTases into *H. somni* pathogenic strains remains to be examined. The expression and purification of REases from these systems using suitable *Escherichia coli* strains as hosts need to be further explored since these enzymes may be of biological interest.

Materials and Methods

Identification of Type I and Type II RM homologs in *H. somni*

The genomes of *H. somni* strains 2336 and 129Pt were searched for homologs of genes involved in Type I and Type II RM using NCBI blastn and blastx programs. The genes, and their translated proteins, were compared with each other and their homologs from other species using blastp, clustalW, and BOXSHADE. Genetic maps of the Type I and Type II RM loci were drawn using BioEdit. Prophage-related sequences and insertion sequence-like elements were identified by manual curation of the genomes and comparing them to analogous genetic elements.

Cloning and expression of *H. somni* Type I and Type II RM genes in *E. coli*

Eight primers (Table 5.01) were designed for PCR amplification of regions containing different genes of the Type I and Type II RM systems from *H. somni*.

Table 5.01: Primers used in this study

Primer Name	Primer Sequence	Description
SS I En L	5′-GGAACAACGTTTAATGGGAGAG-3′	To amplify 3,500 bp region containing *hsdR*
SS I En R	5′-GGGAATTTTGCTCATCTTTTTG-3′	
SS I Me L	5′-TTTTCTTGAAAGGGATTGAACA-3′	To amplify 1,793 bp region containing *hsdM*
SS I Me R	5′-TTTTCTCAAAAGTGCGGTAACA-3′	
SS II En L	5′-GGTTGAAACGATCTTCCATTGT-3′	To amplify 872 bp region containing *R.hsoI*
SS II En R	5′-CGGCTCAATTATTAAAATCGATAG-3′	
SS II Me L	5′-CAGTTTGTGAGCCTAGGTTGG-3′	To amplify 1,127 bp region containing *M.hsoI*
SS II Me R	5′-TGACAAGCGGTTAGGTTTAAGA-3′	

Primers were purchased from Sigma Genosys, The Woodlands, TX, and were diluted to 1 nmol/µl using sterile double distilled water. Appropriate primers were combined to amplify the genes from *H. somni* strain 2336 genomic DNA by PCR utilizing LA Taq DNA Polymerase (Takara Mirus Bio, Madison, WI), as outlined in the next page.

PCR mixture

Left primer 1 µl (10 pmol/µl), Right primer 1 µl (10 pmol/µl),

Genomic DNA template 8 µl (60 ng/µl), 10X LA Taq PCR buffer 5 µl,

dNTPs 8 µl (2.5 mM each), LA Taq 0.5 µl (5 units/ µl), Sterile double distilled water 24.5 µl

PCR protocol

(using 'Eppendorf Matercycler Gradient' thermocycler)

96^0 C for 3 min (one cycle)

96^0 C for 1 min, 68^0 C for 22 min (35 cycles)

72^0 C for 10 min (once cycle)

TOPO cloning

The PCR products were TOPO cloned into StrataClone™ PCR cloning vector pSC-A (Strategene, La Jolla, CA) using *E. coli* SoloPack® competent cells (Strategene, La Jolla, CA), according to the protocol supplied with the kit. The cloned inserts in pSC-A were confirmed by restriction enzyme digestion and DNA sequencing.

Bisulfite treatment

Plasmids with or without the Type II MTase cloned into them were treated with sodium bisulfite to convert unmethylated cytosines to uracils. For this purpose, the EZ DNA Methylation Kit™ (Zymo Research, Orange, CA) was used. The protocol supplied by the manufacturer was used for bisulfite treatment and DNA bound to the column matrix was eluted using water instead of the supplied M-Elution Buffer. The eluted DNA was used as the template in PCR and/or sequencing reactions.

Results

Characteristics of Type I RM genes from *H. somni*

In *H. somni* strain 2336, *hsdR, hsdM*, and *hsdS* are located on the minus strand in the region from 1,714,942 bp to 1,722,690 bp of the 2,263,870 bp chromosome. A gene whose protein product is putatively involved in cell division is present between the *hsdM* and *hsdS* homologs. Another gene, whose protein product is putatively involved in resistance to bacteriophage infection, is present between the *hsdS* and *hsdR* homologs (Fig. 5.01). A 417 bp sequence that is found only in *H. somni* 2336 is present upstream of the *hsdM* gene. Upstream of this 417 bp unique sequence is a gene whose protein product is a putative phage integrase.

In *H. somni* strain 129Pt, a full length homolog of *hsdR* is located on the minus strand in the region from 593,223 bp to 596,423 bp of the 2,008,359 bp chromosome. *H. somni* strain 129Pt HsdR is 98% identical to *H. somni* strain 2336 HsdR. In strain 129Pt, two truncated versions of *hsdS* are present upstream of *hsdR*, but there are no homologs of *hsdM* (Fig. 5.01). In between these two truncated versions of *hsdS* is a gene whose protein product is 99% identical to the phage integrase of strain 2336 mentioned above. Downstream of one of these truncated versions of *hsdS* is a 153 bp sequence that is found only in *H. somni* 129Pt. Two genes, whose protein products are related to proteins of bacteriophage P1, are present immediately upstream of *hsdR*. Interestingly these two bacteriophage P1 related proteins are absent in strain 2336. Furthermore, the genes that are present between the *hsdR, hsdM*, and *hsdS* homologs in strain 2336 are absent in strain 129Pt (Fig. 5.01).

In *H. somni*, *hsdR* is one of the largest genes and makes up about 0.14% and 0.16% of the chromosome in strains 2336 and 129Pt, respectively. The predicted molecular weights of *H. somni* HsdR, HsdM, and HsdS are 122.76 kDa, 60.6 kDa, and 47.7 kDa, respectively.

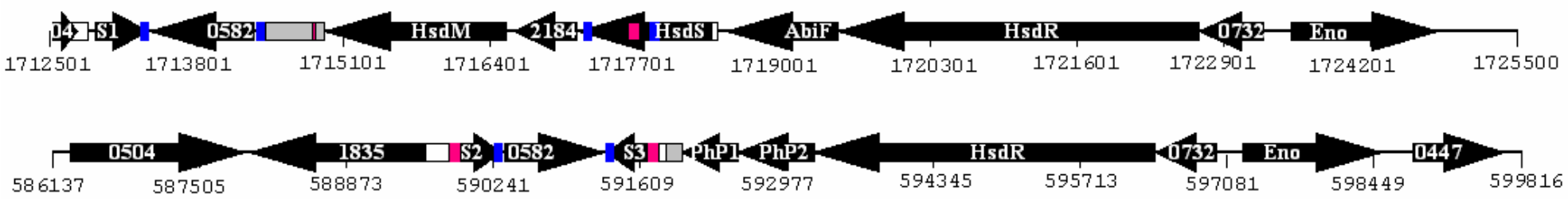

Fig. 5.01: Map of *H. somni* chromosomal region containing the Type I RM locus

Top; *H. somni* strain 2336, 04; CTP synthase (UTP-ammonia lyase) truncated, White box; 221 bp sequence found only in *H. somni* 2336 and 129Pt, S1; Truncated version of HsdS, Blue box; 69 bp sequence found only in *H. somni* 2336 and 129Pt (occurs four times in this strain, as indicated on the map), 0582; Phage integrase, Gray box; 417 bp sequence found only in *H. somni* 2336, HsdM; Type I restriction-modification system methyltransferase subunit (not found in *H. somni* 129Pt), 2184; Protein involved in cell division (not found in *H. somni* 129Pt), Red Box; 99 bp sequence found only in *H. somni* 2336 and 129Pt, HsdS; Type I restriction-modification system specificity subunit, White box; 62 bp sequence found only in *H. somni* 2336 and 129Pt, AbiF; Abortive infection bacteriophage resistance protein (not found in *H. somni* 129Pt), HsdR; Type I restriction-modification system restriction subunit, 0732; Restriction endonuclease S subunits (unrelated to HsdS), Eno; Enolase.

Bottom; *H. somni* strain 129Pt, 0504 (PyrG); CTP synthase (UTP-ammonia lyase), 1835; Predicted acyltransferase, White box; 221 bp sequence found only in *H. somni* 2336 and 129Pt, Red Box; 99 bp sequence found only in *H. somni* 2336 and 129Pt (occurs twice in this strain, as indicated on the map), S2; Truncated version of HsdS, Blue box; 69 bp sequence found only in *H. somni* 2336 and 129Pt (occurs twice in this strain, as indicated on the map), 0582; Phage integrase, S3; Truncated version of HsdS, White box; 62 bp sequence found only in *H. somni* 2336 and 129Pt, Gray box; 153 bp sequence found only in *H. somni* 129Pt, PhP1 and PhP2; phage P1-related proteins (not found in *H. somni* 2336), PhP2, HsdR, 0732, and Eno; as in strain 2336.

Characteristics of Type II RM genes from *H. somni*

In *H. somni* strain 2336, *R.hsoI* (which encodes a putative Type II REase) and *M.hsoI* (which encodes a putative Type II MTase) are located on the minus strand in the region from 400,804 bp to 402,529 bp of the 2,263,870 bp chromosome. The genes form a tight locus and a 10 bp 5'-CTTCACCTCT-3' sequence separates them. The predicted molecular weights of R.HsoI and M.HsoI are 29.08 kDa and 37.24 kDa, respectively. A gene whose protein product is putatively involved in valyl-tRNA synthesis is present immediately upstream of *R.hsoI*. Downstream of *M.hsoI* are a 47 bp sequence and a 17 bp sequence that are found only in *H. somni* 2336. A gene whose protein product is a putative response regulator and a 254 bp sequence found only in *H. somni* 2336 are also present downstream of *M.hsoI* (Fig. 5.02).

In *H. somni* strain 129Pt, there are no homologs of *R.hsoI* and *M.hsoI*. However, the genes that are up- and down-stream of *R.hsoI* and *M.hsoI* in strain 2336 are present in strain 129Pt in the region from 1,442,400 bp to 1,450,800 bp of the 2,008,359 bp chromosome, in the same orientation (Fig. 5.02). Within this region is a gene, whose translation has been annotated as 'nuclease of the RecB family', that is not found in strain 2336. Downstream of this gene are a 115 bp sequence and a 119 bp sequence that are found only in *H. somni* 129Pt (Fig. 5.02).

Further comparison of the regions described above revealed that a 98 bp sequence and a 144 bp sequence are found in both strains 2336 and 129Pt, but not in the GenBank database (Fig. 5.02). Interestingly, the 98 bp sequence in strain 2336 appears to contain an insertion sequence-like element that reads 5'-CCCACACGAGCCGTGTGGG-3' (with 8 bp inverted repeats underlined). The 98 bp sequence in strain 129Pt also contains an insertion sequence-like element which reads 5'-AACCCCACACGAGCCGTGTGGGGTT-3' (with 11 bp inverted repeats underlined).

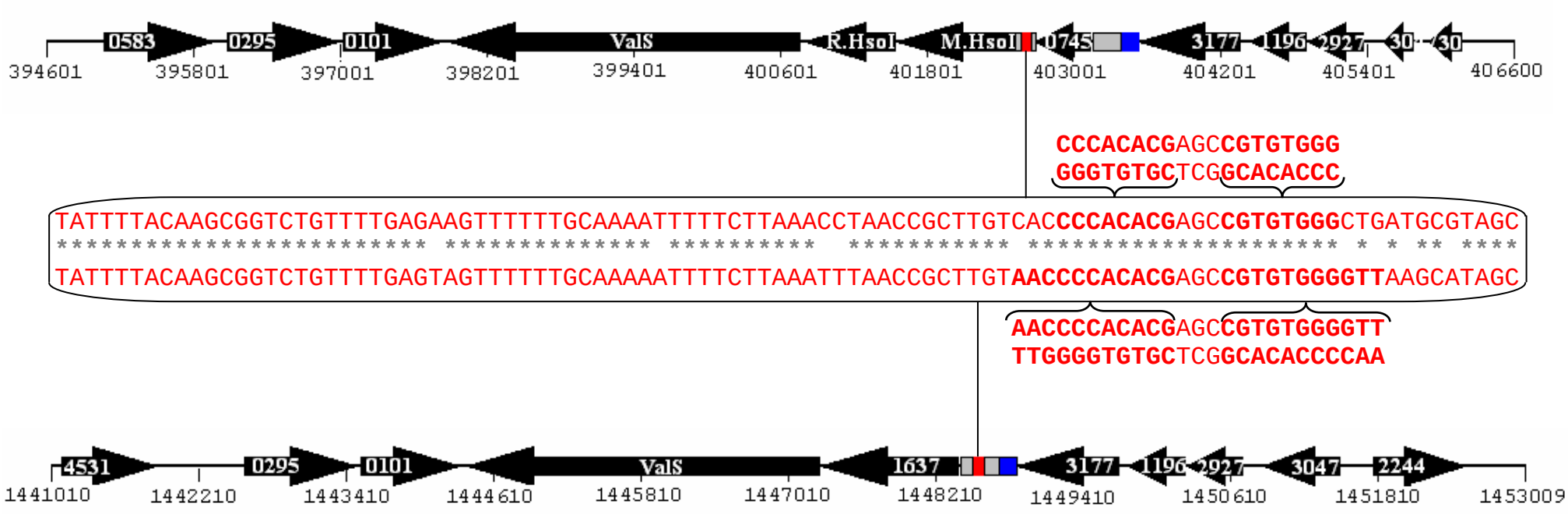

Fig. 5.02: Map of the *H. somni* chromosomal region containing the Type II RM locus (detailed legend appears in the next page)

Fig. 5.02: Map of *H. somni* chromosomal region containing the Type II RM locus

Top; *H. somni* strain 2336, 0583; Transcriptional regulator, 0295; Cytidine deaminase, 0101; Pseudouridylate synthase, ValS; Valyl-tRNA synthetase, R.HsoI; Type II endonuclease (not found in *H. somni* 129Pt), M.HsoI; Type II methyltransferase (not found in *H. somni* 129Pt), Gray box; 47 bp sequence found only in *H. somni* 2336, Red box; 98 bp sequence found only in *H. somni* 2336 and 129Pt, Gray box; 17 bp sequence found only in *H. somni* 2336, 0745; Response regulator, Gray box; 254 bp sequence found only in *H. somni* 2336, Blue box; 144 bp sequence found only in *H. somni* 2336 and 129Pt, 3177; Conserved hypothetical protein, 1196; Chromosome segregation ATPases, 2927; DNA polymerase III, chi subunit, 30; Outer membrane protein W (truncated), 30; Outer membrane protein W (truncated).

Bottom; *H. somni* strain 129Pt, 4531; periplasmic zinc transporter/adhesin B precursor, 0295, 0101, and ValS; as in strain 2336, 1637; Predicted nuclease of the RecB family [a DNA replication, recombination, and repair protein that is not found in *H. somni* 2336], Gray box; 115 bp sequence found only in *H. somni* 129Pt, Red box; 98 bp sequence found only in *H. somni* 2336 and 129Pt, Gray box; 119 bp sequence found only in *H. somni* 129Pt, Blue box; 144 bp sequence found only in *H. somni* 2336 and 129Pt, 3177, 1196, and 2927; as in strain 2336, 3047; Outer membrane protein W.

The sequences in the center of the two maps show a 98 bp region found only in *H. somni* 2336 and 129Pt (marked with red boxes in the maps). The asterisks indicate identical nucleotides. The sequence above the asterisks is from strain 2336. The sequence below the asterisks is from strain 129Pt. This region contains an insertion sequence-like element (shown above and below the 98 bp sequences).

PCR amplification and cloning of the *H. somni* strain 2336 Type I and Type II RM genes

Different genes of the Type I and Type II RM systems from *H. somni* strain 2336 genomic DNA were amplified by PCR. The products of PCR amplification are shown in fig. 5.03. The genes were TOPO cloned into PCR cloning vector pSC-A using *E. coli* SoloPack® cells to create plasmids shown in table 5.02.

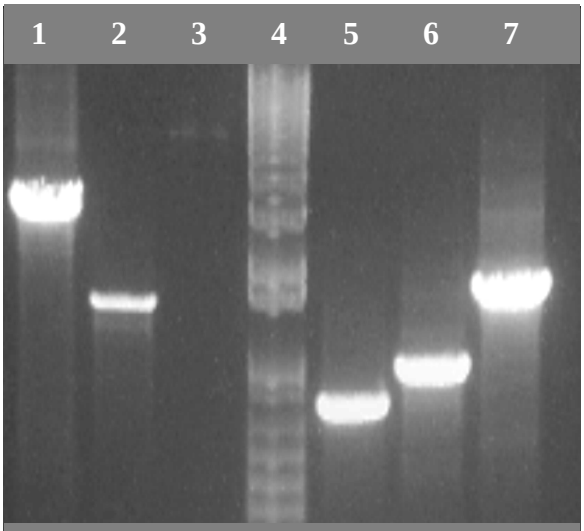

Fig. 5.03: PCR amplification of DNA containing *H. somni* 2336 Type I and Type II RM genes

Lane 1; 3,500 bp PCR product containing *hsdR*

Lane 2; 1,793 bp PCR product containing *hsdM*

Lane 3; 7,944 bp PCR product containing *hsdR*, *hsdS*, and *hsdM* as well as genes in between them

Lane 4; 1 kb plus DNA ladder

Lane 5; 872 bp PCR product containing *R.hsoI*

Lane 6; 1,127 bp PCR product containing *M.hsoI*

Lane 7; 1,922 bp PCR product containing *R.hsoI* and *M.hsoI*

Table 5.02: pSC-A plasmids containing *H. somni* strain 2336 RM genes

Plasmid Name	Description
pSC-A-Hsd-R	Contains 3,500 bp *hsdR* insert (from lane 1 in fig. 5.03)
pSC-A-Hsd-M	Contains 1,793 bp *hsdM* insert (from lane 2 in fig. 5.03)
pSC-A-Hsd-RSM	Contains 7,944 bp insert with *hsdR*, *hsdS*, and *hsdM* as well as genes in between them (from lane 3 in fig. 5.03)
pSC-A-HsoI-R	Contains 872 bp *R.hsoI* insert (from lane 5 in fig. 5.03)
pSC-A-HsoI-M	Contains 1,127 bp *M.hsoI* insert (from lane 6 in fig. 5.03)
pSC-A-HsoI-RM	Contains 1,922 bp insert with *R.hsoI* and *M.hsoI* (from lane 7 in fig. 5.03)

Functional characterization of *H. somni* strain 2336 Type I and Type II methyltransferase genes

To test if *H. somni* strain 2336 *hsdM* and *M.hsoI* genes function as methyltransferases, *E. coli* strain AP1-200-9 was used as a reporter. This strain is temperature-sensitive for the Mcr and Mrr systems and contains a *lacZ* gene fused to the damage-inducible *dinD* locus. Plasmids encoding DNA methyltransferases can be identified using *E. coli* strain AP1-200-9 since DNA methylation within this strain engenders the DNA-damage inducible SOS response by the Mcr and Mrr systems (Piekarowicz et al., 1991).

Plasmids containing *H. somni* strain 2336 *hsdM* or *M.hsoI* genes were electroporated into the *E. coli* reporter strain AP1-200-9 using standard procedures. One drop (~10 µl) of transformed cells was placed on a LB agar plate containing 100 µg/ml ampicillin and 30 µg/ml X-gal. The plate was incubated at 37^0C for 48 hours and the colony mass was blotted to a nitrocellulose membrane to visualize color development. Only *E. coli* strain AP1-200-9 transformed with plasmids containing Type I or Type II RM loci developed blue color, indicating that these loci contain functional methyltransferase genes (Figs. 5.04, 5.05, and 5.06).

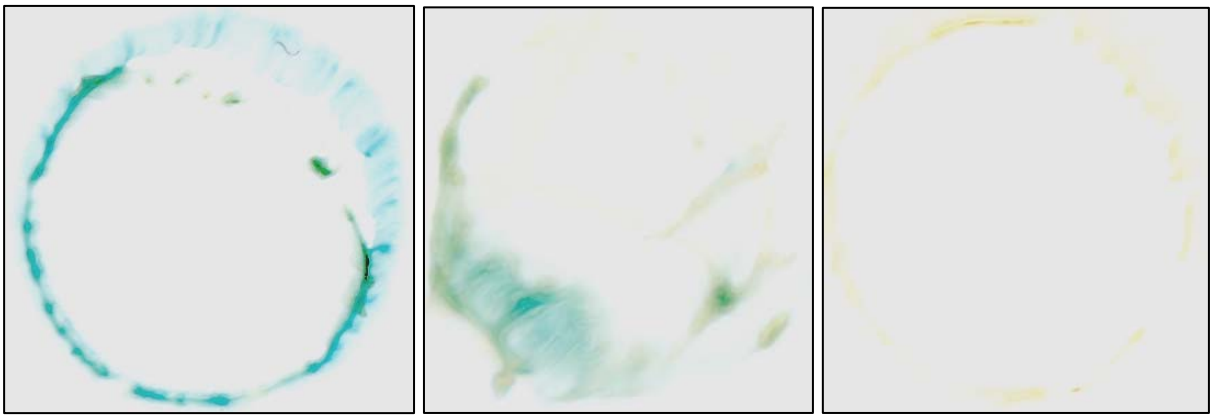

Fig. 5.04 (left): *E. coli* strain AP1-200-9 transformed with plasmid pSC-A-HsoI-RM

Fig. 5.05 (center): *E. coli* strain AP1-200-9 transformed with plasmid pSC-A-Hsd-RSM

Fig. 5.06 (right): *E. coli* strain AP1-200-9 transformed with plasmid pSC-A (negative control)

Methylation of DNA by *H. somni* strain 2336 Type II MTase

Previous studies have shown that *in vitro* methylation of DNA using M.HhaI affords protection against R.HsoI (Briggs and Tatum, 2005). Therefore, it is possible that M.HsoI is an isoschizomer of M.HhaI, which is one of the most well characterized enzymes and a paradigm for studying prokaryotic MTases (Bheemanaik et al., 2006). A blastp search revealed that M.HsoI was 67% identical to M.HinP1I and 31% identical to M.HhaI. A type II MTase named MspI, which recognizes the sequence 5'-CCGG-3', was also found to be 32% identical to M.HsoI (data not shown). Although these proteins appear to be distantly related, comparative analyses according to Bujnicki (2000) using the ClustalW-BOXSHADE multiple sequence alignment tools revealed the presence of several conserved regions that are predicted to interact with substrate DNA target sites (5'-GCGC-3 or 5'-CCGG-3) and AdoMet (Fig. 5.07).

Based on the above observations, it was hypothesized that methylation by M.HsoI would protect DNA against the endonuclease activity of R.HhaI. To test this hypothesis, various plasmids with or without *H. somni* strain 2336 Type II MTase inserted into them were purified from *E. coli* SoloPack® cells and digested with R.HhaI (New England Biolabs). Only plasmids pSC-A-HsoI-M and pSC-A-HsoI-RM, or those that were maintained in *E. coli* SoloPack® cells in the presence of pSC-A-HsoI-M or pSC-A-HsoI-RM, were protected from R.HhaI (Fig. 5.08). To test the hypothesis further, genomic DNA from *H. somni* strains 2336 and 129Pt was purified and digested with R.HhaI. It was found that R.HhaI cleaves *H. somni* strain 129Pt genomic DNA, but not *H. somni* strain 2336 genomic DNA (Fig. 5.09).

These results indicate that *H. somni* strain 2336 Type II MTase is capable of DNA methylation *in vivo* and can protect chromosomal and plasmid DNA molecules from the endonuclease activity of R.HhaI *in vitro*.

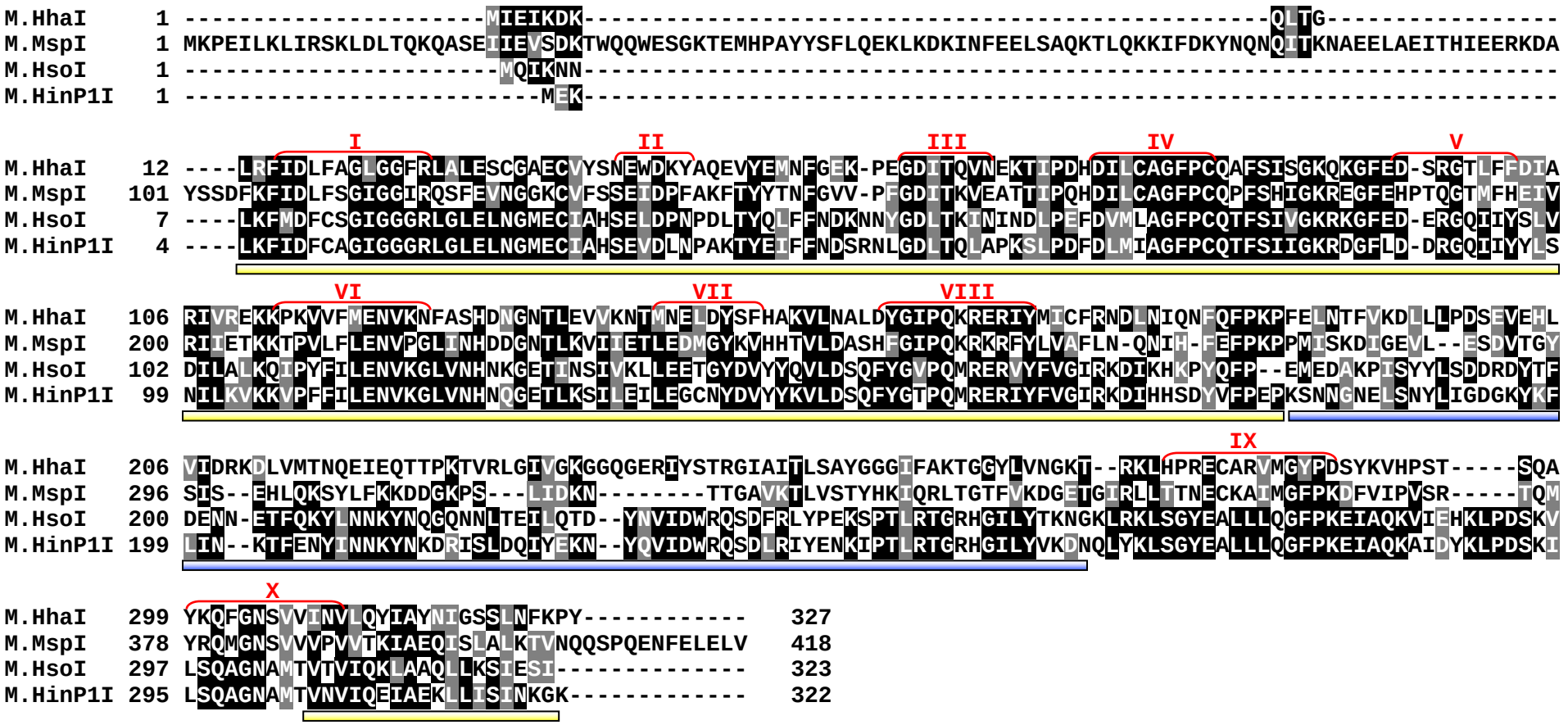

Fig. 5.07: ClustalW-BOXSHADE multiple sequence alignment of M.HsoI, M.HhaI, M.HinP1I, and M.MspI homologs

Homologous regions are box-shaded dark (identical amino acid residues) and gray (conserved amino acid substitutions). The regions of M.HsoI and M.HinP1I that correspond to domain 1 and domain 2 of M.HhaI are marked with yellow and blue bars, respectively. I–III; predicted motifs that form the AdoMet binding subunit (along with motif X), IV–VIII; predicted motifs that form the catalytic domain, IX; predicted motif that forms the small DNA-binding domain (along with the 'target recognizing domain', which is the region between motifs VIII and IX).

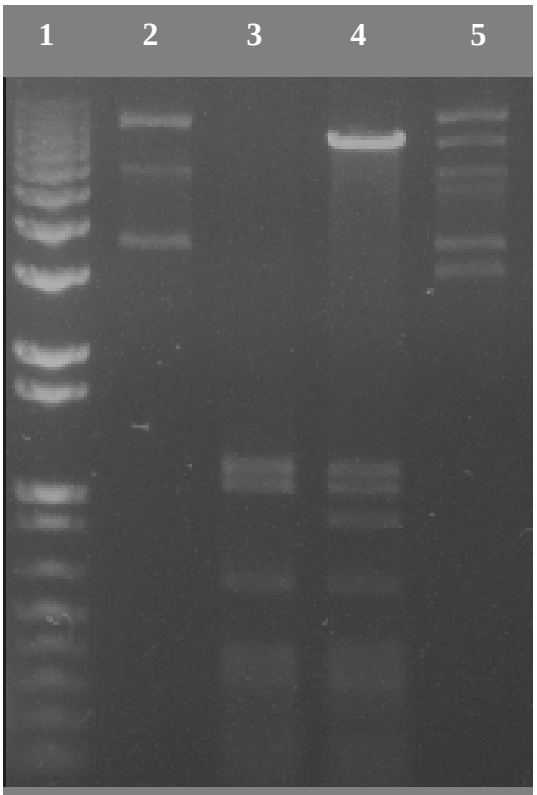

Fig. 5.08: M.HsoI methylation and R.HhaI digestion of DNA

Lane 1; 1 Kb plus DNA ladder

Lane 2; pSC-A-HsoI-RM digested with HhaI (plasmid remains uncut)

Lane 3; pBS*luxS*Kan digested with R.HhaI (plasmid has been digested completely)

Lane 4; pBS*luxS*Kan maintained in the presence of pSC-A-Hsd-RSM, plasmids were purified together and digested with HhaI (both plasmids have been digested completely, the top bright band is 7944 bp insert containing *hsdR*, *hsdS*, and *hsdM* genes)

Lane 5; pBS*luxS*Kan maintained in the presence of pSC-A-HsoI-RM, plasmids were purified together and digested with HhaI (both plasmids remain undigested)

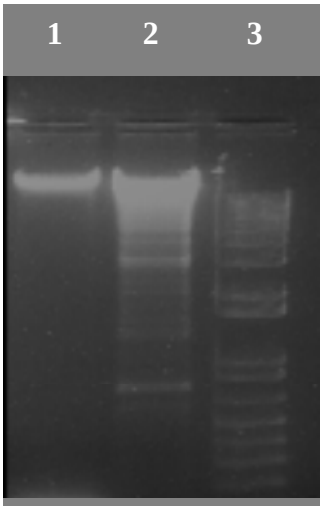

Fig. 5.09: R.HhaI digestion of genomic DNA

Lane 1; *H. somni* strain 2336 genomic DNA digested with R.HhaI (DNA remains undigested)

Lane 2; *H. somni* strain 129Pt genomic DNA digested with R.HhaI (DNA has been digested)

Lane 3; 1 Kb plus DNA ladder

In silico analyses of *H. somni* strain 2336 Type II REase

Based on the studies by Briggs and Tatum (2005), R.HsoI is predicted to be an isoschizomer of R.HhaI and R.HinP1I and functionally identical to R.HinP1I, which is yet another well characterized restriction enzyme (Yang et al., 2005). A blastp search revealed that R.HsoI was 71% identical to R.HinP1I. Unlike M.HsoI, which is 32% identical to M.MspI, R.HsoI showed very little homology to R.MspI by the ClustalW-BOXSHADE method (Fig. 5.10). This is not surprising since Type II REases, with the exception of a few isoschizomers, display little or no sequence homology (Pingoud and Jeltsch, 2001). Nevertheless, the structures of most of these enzymes have a core made of a 'five-stranded mixed β-sheet flanked by α-helices' (Pingoud and Jeltsch, 2001).

For characterization of R.HsoI, its protein sequence was compared to those of R.HinP1I and R.MspI since the crystal structures and domain features of these enzymes have been elucidated (Xu et al., 2004; Yang et al., 2005). Yang et al. (2005) showed that R.HinP1I and R.MspI share 14% identity and, despite the lack of sequence homology, the two enzymes are structurally similar. Although not immediately apparent in the ClustalW-BOXSHADE alignment (Fig. 5.10), 28 identical amino acid residues and 35 conserved substitutions were identified based on the comparison of R.HsoI with the structure-sequence alignment of R.HinP1I and R.MspI by Yang et al. (2005). The identical residues include those involved in forming the DNA-binding and catalytic motifs (Fig. 5.11). The sequence comparison also facilitated the identification of putative core β-sheets flanked by α-helices in R.HsoI (data not shown).

These results indicate that *H. somni* strain 2336 *R.hsoI* may encode the enzyme described by Briggs and Tatum (2005). This enzyme, which cleaves 5'-GCGC-3' identical to R.HinP1I, was named R.HsoI since it was the first REase identified in *H. somni* strain 2336.

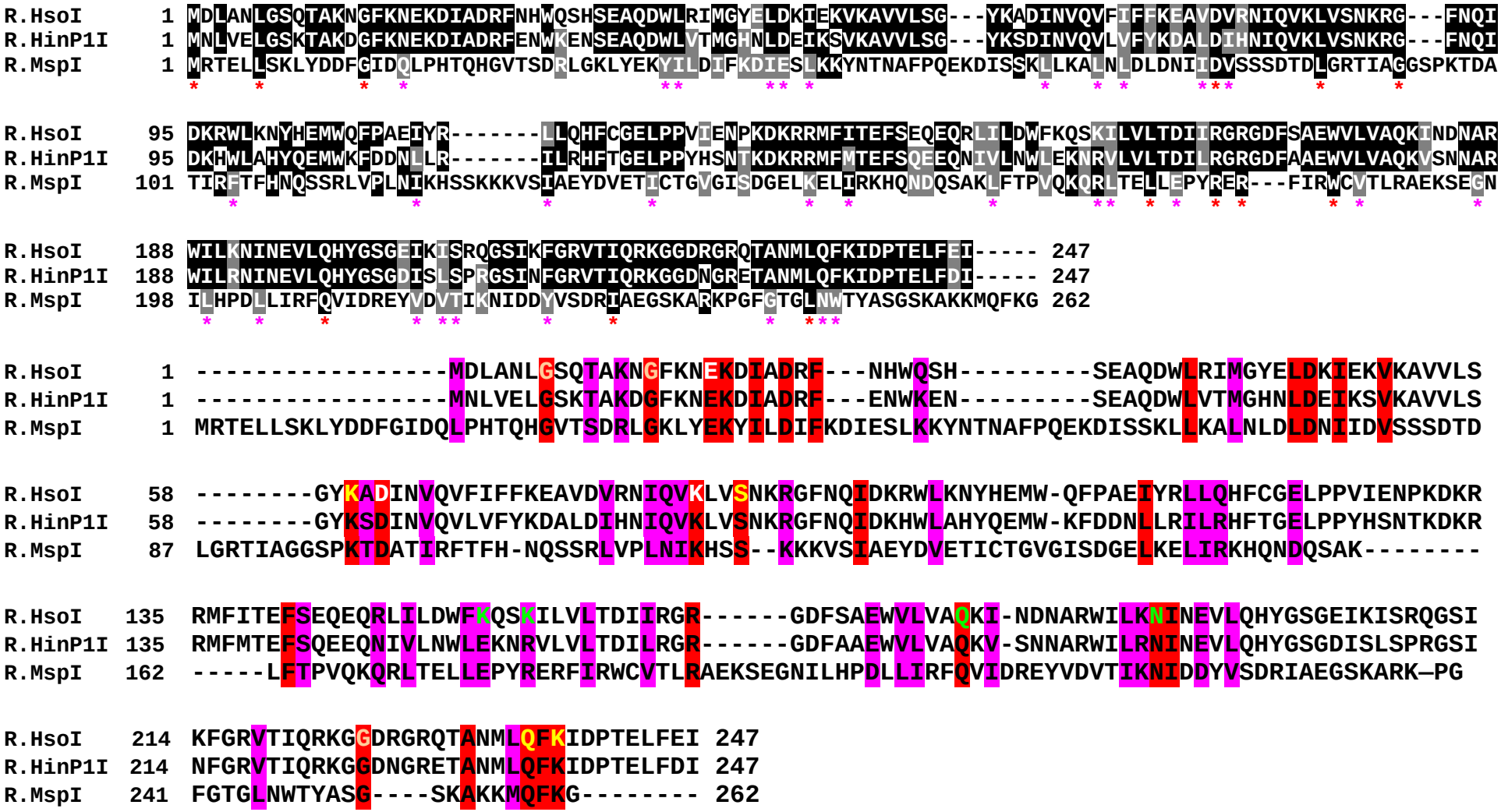

Fig. 5.10 (Top): ClustalW-BOXSHADE multiple sequence alignment of R.HsoI, R.HinP1I, and R.MspI homologs

Fig. 5.11 (Bottom): Comparison of R.HsoI with the structure-based sequence alignment of R.HinP1I and R.MspI

[detailed legends for both figures appear in the next page]

Fig. 5.10: ClustalW-BOXSHADE multiple sequence alignment of R.HsoI, R.HinP1I, and R.MspI homologs

Homologous regions are box-shaded dark (identical amino acid residues) and gray (conserved amino acid substitutions). The alignment reveals that there are only 13 identical amino acid residues (marked by red asterisks) and 32 conserved substitutions (marked by pink asterisks) among the three sequences.

Fig. 5.11: Comparison of R.HsoI with R.HinP1I and R.MspI

[Based on the structure-based sequence alignment of R.HinP1I and R.MspI by Yang et al. (2005)]

In the first step, sequences of R.HinP1I and R.MspI were aligned according to Yang et al. (2005). The sequence of R.HsoI was then matched to that of R.HinP1I. The comparison reveals that there are 28 identical amino acid residues (shaded red) and 35 conserved substitutions (shaded pink) among the three sequences. Conserved glycine residues that may stabilize the protein structure are marked tan. Conserved putative catalytic residues are marked white. Conserved residues that may interact with DNA are marked yellow. Conserved residues that may be involved in intra-molecular interactions are marked green.

Preliminary analyses of *H. somni* strain 2336 Type I RM locus

A blastp search revealed that *H. somni* HsdR was 62% identical to HsdR of *Streptococcus suis*, 52% identical to HsdR of *Lactobacillus reuteri*, 34% identical to HsdR of *E. coli*, 31% identical to HsdR of *Enterococcus faecium*, *Clostridium cellulolyticum*, and *Bacillus licheniformis*, 30% identical to HsdR of *Actinobacillus pleuropneumoniae*, 29% identical to HsdR of *H. influenzae* 86-028NP, *Listeria innocua*, *Vibrio splendidus*, and *Neisseria meningitidis*, 28% identical to HsdR of *Staphylococcus epidermidis*, *N. gonorrhaeae*, and *Campylobacter jejuni*.

A blastp search revealed that *H. somni* HsdS was 44% identical to HsdS of *B. licheniformis*, 39% identical to HsdS of *S. suis*, 38% identical to HsdS of *L. reuteri*, 30% identical to HsdS of *E. faecium*, 27% identical to HsdS of *Staphylococcus aureus* and *L. innocua*, 26% identical to HsdS of *V. vulnificus* and *C. cellulolyticum*, and 24% identical to HsdS of *Yersinia pseudotuberculosis*. However, no homologs of *H. somni* HsdS were found among members of the *Pasteurellaceae*.

A blastp search revealed that *H. somni* HsdM was 70% identical to HsdM of *S. suis*, 61% identical to HsdM of *L. reuteri*, 51% identical to HsdM of *S. epidermidis*, 48% identical to HsdM of *E. faecium* and *C. cellulolyticum*, 45% identical to HsdM of *B. licheniformis*, 42% identical to HsdM of *L. innocua*, 40% identical to HsdM of *Vibrio vulnificus* and *Campylobacter fetus*. As in HsdS, no homologs of *H. somni* HsdM were found among members of the *Pasteurellaceae*.

Furthermore, homologs of the *H. somni* ORF labeled 0732 in fig. 5.01 are also absent in other members of the *Pasteurellaceae*, but are present in *S. suis* and *Bacillus licheniformis* (36% identity). However, homologs of the *H. somni* ORF labeled Eno (Fig. 5.01) are found in several members of the *Pasteurellaceae*, including *H. influenzae* (93% identity).

In most Type I RM systems, the methylated base formed is m6A (Roberts et al., 2003). As a first step toward understanding the function of the *H. somni* Type I RM locus, it was hypothesized that plasmids containing this locus would not be cut by some REases sensitive to adenine methylation. To test this hypothesis, plasmid pSC-A-Hsd-RSM was digested with R.HinfI (New England Biolabs), a *H. influenzae* Type II REase that recognizes the sequence 5'-GANTC-3' and is sensitive to M.HinfI catalyzed methylation of the adenine on both strands (Chandrasegaran et al., 1988). However, R.HinfI cut the plasmid and generated fragments of expected size (Fig. 5.12).

To test the hypothesis further, plasmid pSC-A-Hsd-RSM was digested with R.Hin4I (Fermentas, Hanover, MD), another *H. influenzae* Type II REase that is sensitive to adenine methylation and recognizes the sequence shown below (arrows point to sites where DNA is cut):

$$5'\ldots\downarrow\ {}_{8}\mathbf{(N)\,G\,A\,Py(N)_5\,(A/C/G)\,T\,C(N)}_{13\text{-}14}\downarrow\ldots 3'$$
$$3'\ldots\uparrow_{13\text{-}14}\mathbf{(N)\,C\,T\,Pu\,(N)_5\,(T/G/C)\,A\,G(N)}_{8}\ \uparrow\ldots 5'$$

However, R.Hin4I was also found to cut the plasmid as expected (Fig. 5.13).

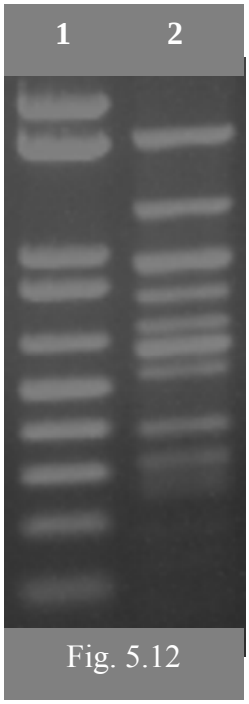

Fig. 5.12

Fig. 5.12: R.HinfI digestion of pSC-A-Hsd-RSM

Lane 1; 1 kb plus DNA ladder (bands shown are from 100 bp to 2,000 bp), Lane 2; pSC-A-Hsd-RSM (purified from *E. coli* SoloPack®) cut with R.HinfI produced expected results (1,767 bp, 1,245 bp, 986 bp, and several fragments smaller than these).

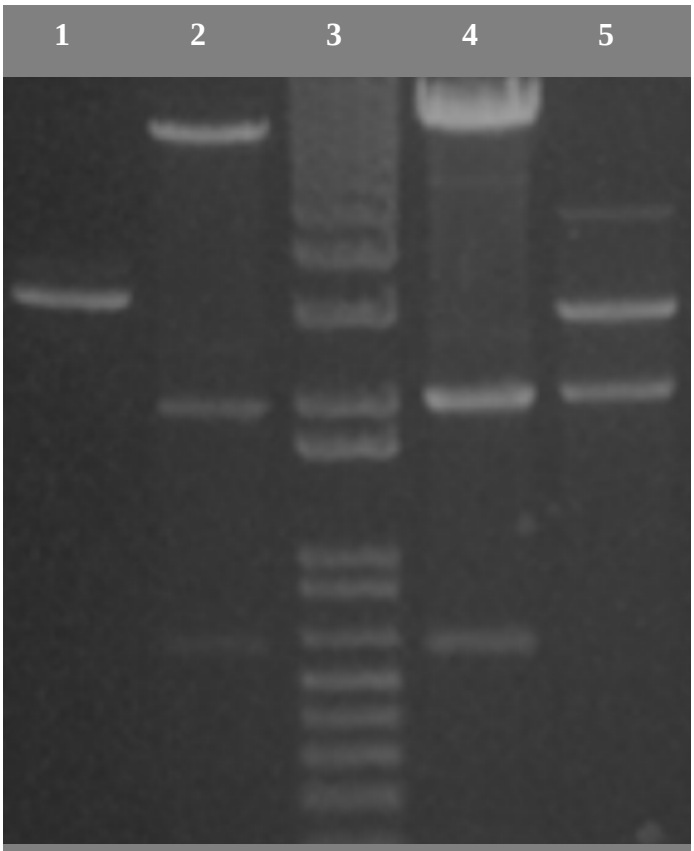

Fig. 5.13:

R.Hin4I digestion of pSC-A-Hsd-RSM

Lane 1; pSC-A (purified from *E. coli* SoloPack®) cut with R.Hin4I (3,500 bp plasmid has been linearized), Lane 2; pSC-A-Hsd-RSM (purified from *E. coli* SoloPack®) cut with R.Hin4I, produced expected results (8,539 bp, 2,040 bp, and 635 bp fragments), Lane 3; 1 kb plus DNA ladder (bands shown are from 200 bp to 12,000 bp), Lane 4; pSC-A-Hsd-RSM (purified from *E. coli* AP1-200-9) cut with R.Hin4I, produced expected results (similar to lane 2), Lane 5; pSC-A-HsoI-RM (purified from *E. coli* AP1-200-9) cut with R.Hin4I, produced expected results (3,380 bp and 2,040 bp fragments).

Discussion

Comparative analyses of the genome sequences of *H. somni* strains 2336 and 129Pt have revealed the presence of several genes whose protein products are related to enzymes involved in restriction-modification of DNA. Some of these genes are in close proximity to each other on the chromosomes of *H. somni* strains and resemble the typical RM loci found in other bacteria. Based on protein homology, two of the *H. somni* strain 2336 loci that contain RM genes have been predicted to encode Type I and Type II RM functions.

The Type I RM locus of strain 2336 contains three genes (*hsdR-hsdS-hsdM*) analogous to other Type I RM systems and is found to be associated with a region that has been designated as prophage region VI (chapter II). The closest homologs of strain 2336 HsdR, HsdS, and HsdM are found in several other Gram-positive bacteria, but not in members of the *Pasteurellaceae* (with the exception of HsdR, whose homologs are found in *A. pleuropneumoniae* and *H. influenzae*). Bacteriophage-associated RM systems have also been found in other bacteria (Jeltsch and Pingoud, 1996; Kobayashi, 2001; Dempsey et al., 2005). Based on these features, it is predicted that the Type I RM system of strain 2336 was acquired by bacteriophage-mediated HGT. Although reporter assays using *E. coli* strain AP1-200-9 indicated that strain 2336 genes encoding Type I RM proteins are functional, their DNA recognition specificity could not be identified by homology and comparison to other Type I RM enzymes.

In *H. somni* strain 129Pt, the locus that is analogous to the Type I RM locus of strain 2336 contains only *hsdR* and truncated *hsdS* genes, but not *hsdM*. Therefore, the Type I RM system is predicted to be inactive in this strain. This locus in strain 129Pt is found to be associated with a region that has been designated as prophage region IV (chapter II). Furthermore, the Type I RM locus in strain 2336 contains a gene that is predicted to encode an

abortive infection bacteriophage resistance protein (AbiF). Abortive infection proteins have been shown to mediate resistance against bacteriophages in members of *Lactococcus* spp., and bacteriophages that are immune to the function of AbiF have also been described (Rince et al., 2000). However, *abiF* is neither found in strain 129Pt nor in *H. influenzae* strains Rd KW20 and 86-028NP. One possible explanation of this variation is that the common ancestor of strains 2336 and 129Pt acquired the Type I RM locus by bacteriophage-mediated HGT, and that some genes in this locus in strain 129Pt got disrupted by a secondary bacteriophage infection during subsequent evolution.

Although one of the functions of RM systems is to afford protection against bacteriophage attack (the "*cellular defense hypothesis*"), it is interesting to note that RM systems are found associated with bacteriophages. Furthermore, several bacteriophage infections seem to have occurred in strain 2336 despite the presence of RM systems and abortive infection components. In view of these considerations, further functional characterization of the Type I RM locus and the *abiF* gene of strain 2336 is worth pursuing.

The Type II RM locus of strain 2336 contains two genes (*R.hsoI-M.hsoI*) analogous to other Type II RM systems. Although sequencing of DNA (pretreated with sodium bisulfite to convert unmethylated cytosines to uracils) was used as an assay to identify the sites methylated by M.HsoI, no meaningful sequence was obtained and the results of this assay remain inconclusive (data not shown). However, based on *in vivo* DNA methylation and protection assays, as well as homology modeling in relation R.HinP1I and R.MspI, this system is predicted to recognize the sequence 5'-GCGC-3'. Further characterization of the Type II RM locus of strain 2336 is of no commercial value as enzymes that recognize and cleave 5'-GCGC-3' are available in the market.

In addition, akin to the HsoI RM locus of strain 2336, the HindIII RM system of *H. influenzae* Rd and the HinP1I RM system of *H. influenzae* P1 have been shown to occur in the proximity of the *valS* gene that encodes valyl-tRNA synthetase (Yang et al., 2005). However, the HinP1I and HsoI RM systems are unrelated to the HindIII RM system. According to Yang et al. (2005), the presence of genes encoding disparate RM systems close to *valS* in different bacteria may either be a genetic convergence or a coincidence and does not necessarily signify common ancestral origin. Nevertheless, it appears that the HinP1I and HsoI RM systems originated in the common ancestor of *H. influenzae* P1 and *H. somni* 2336, and they could be considered allelic. The ability to recognize and cleave the sequence 5'-GCGC-3' may therefore be a conserved feature among these bacteria despite their divergent evolution with respect to host specificity.

The absence of *R.hsoI* and *M.hsoI* genes in strain 129Pt is of curiosity since this strain has a locus that contains most of the genes that flank the *R.hsoI-M.hsoI* genes of strain 2336, including the *valS* gene. It is tempting to speculate that the presence of an insertion sequence-like element, along with a strain-specific gene that is predicted to encode a RecB protein, within this locus may have been responsible for the loss of the *R.hsoI-M.hsoI* genes in strain 129Pt. Since insertion sequence-mediated inactivation of RM systems has been observed previously (Schoner and Kahn, 1981; Rodicio et al., 1991), it would be interesting to investigate if a similar mechanism caused the loss of the *R.hsoI-M.hsoI* genes in strain 129Pt.

In conclusion, the loss of genes of the Type I and Type II RM systems in strain 129Pt indicates that RM systems are not absolutely essential for cell survival. This loss may also partially explain the relative ease with which this strain can be transformed in the laboratory. However, the functions of a strain-specific RM system acquired via HGT (prophage region II, chapter II), and whether this RM system can compensate for the loss of other RM systems in strain 129Pt, remain to be examined.

References

Barsomian JM, Card CO, Wilson GG (1988). Cloning of the HhaI and HinPI restriction-modification systems. *Gene* **74**: 5-7.

Bheemanaik S, Reddy YV, Rao DN (2006). Structure, function and mechanism of exocyclic DNA methyltransferases. *Biochemistry Journal* **399:** 177-90.

Briggs RE, Tatum FM, Casey TA, Frank GH (1994). Characterization of a restriction endonuclease, PhaI, from *Pasteurella haemolytica* serotype A1 and protection of heterologous DNA by a cloned PhaI methyltransferase gene. *Applied and Environmental Microbiology* **60**: 2006-2010.

Briggs RE and Tatum FM (2005). Generation and molecular characterization of new temperature-sensitive plasmids intended for genetic engineering of *Pasteurellaceae*. *Applied and Environmental Microbiology* **71**: 7187-7195.

Bujnicki JM (2000). Homology modeling of the DNA 5mC methyltransferase M.BssHII. Is permutation of functional subdomains common to all subfamilies of DNA methyltransferases? *International Journal of Biological Macromolecules* **27**: 195-204.

Chandrasegaran S, Lunnen KD, Smith HO, Wilson GG (1988). Cloning and sequencing the HinfI restriction and modification genes. *Gene* **70**: 387-392.

Dempsey RM, Carroll D, Kong H, Higgins L, Keane CT, Coleman DC (2005). Sau42I, a BcgI-like restriction-modification system encoded by the *Staphylococcus aureus* quadruple-converting phage Phi42. *Microbiology* **151**: 1301-1311.

Jeltsch A and Pingoud A (1996). Horizontal gene transfer contributes to the wide distribution and evolution of type II restriction-modification systems. *Journal of Molecular Evolution* **42**: 91-96.

Kobayashi I (2001). Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. *Nucleic Acids Research* **29**: 3742-3756.

Kobayashi I (2004). Restriction-modification systems as minimal forms of life. In: Pingoud A (Editor) *Restriction Endonucleases*. Berlin: Springer**,** pp.19-62.

Murray NE (2000). Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). *Microbiology and Molecular Biology Reviews* **64**: 412-34.

Piekarowicz A, Yuan R, Stein DC (1991). A new method for the rapid identification of genes encoding restriction and modification enzymes. *Nucleic Acids Research* **19**: 1831-1835.

Pingoud A and Jeltsch A (2001). Structure and function of type II restriction endonucleases. *Nucleic Acids Research* **29**: 3705-3727.

Rince A, Tangney M, Fitzgerald GF (2000). Identification of a DNA region from lactococcal phage sk1 protecting phage 712 from the abortive infection mechanism AbiF. *FEMS Microbiology Letters* **182**: 185-91.

Roberts RJ, Belfort M, Bestor T, Bhagwat AS, Bickle TA, Bitinaite J, Blumenthal RM, Degtyarev SK, Dryden DT, Dybvig K, Firman K, Gromova ES, Gumport RI, Halford SE, Hattman S, Heitman J, Hornby DP, Janulaitis A, Jeltsch A, Josephsen J, Kiss A, Klaenhammer TR, Kobayashi I, Kong H, Kruger DH, Lacks S, Marinus MG, Miyahara M, Morgan RD, Murray NE, Nagaraja V, Piekarowicz A, Pingoud A, Raleigh E, Rao DN, Reich N, Repin VE, Selker EU, Shaw PC, Stein DC, Stoddard BL, Szybalski W, Trautner TA, Van Etten JL, Vitor JM, Wilson GG, Xu SY (2003). A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Research* **31**: 1805-1812.

Rodicio MR, Alvarez MA, Chater KF (1991). Isolation and genetic structure of IS112, an insertion sequence responsible for the inactivation of the SalI restriction-modification system of *Streptomyces albus* G. *Molecular and General Genetics* **225**: 142-147.

Sanders JD, Tagawa Y, Briggs RE, Corbeil LB (1997). Transformation of a virulence associated gene of *Haemophilus somnus* into a strain lacking the gene. *FEMS Microbiology Letters* **154**: 251-258.

Sanders JD, Bastida-Corcuera FD, Arnold KF, Wunderlich AC, Corbeil LB (2003). Genetic manipulation of immunoglobulin binding proteins of *Haemophilus somnus*. *Microbial Pathogenesis* **34**: 131-139.

Schoner B and Kahn M (1981). The nucleotide sequence of IS5 from *Escherichia coli*. *Gene* **14**:165-174.

Wu Y, McQuiston JH, Cox A, Pack TD, Inzana TJ (2000). Molecular cloning and mutagenesis of a DNA locus involved in lipooligosaccharide biosynthesis in *Haemophilus somnus.* *Infection and Immunity* **68**: 310-319.

Xu QS, Kucera RB, Roberts RJ, Guo HC (2004). An asymmetric complex of restriction endonuclease MspI on its palindromic DNA recognition site. *Structure* **12:** 1741–1747.