Type I Restriction And Modification Systems

Annette J.B. Titheradge

PhD – The University of Edinburgh – 2004

I declare that the work presented in this thesis is my own, unless otherwise stated. This work has not been submitted for any other degree or professional qualification.

To my parents, John Laurenson and Ina B. Bruce.

Acknowledgements

I would like to thank my supervisor, Professor Noreen E. Murray, for her support, advice and encouragement over many years. Without her help this thesis would not have been possible. I would also like to thank past members of the Murray research group for their collaboration and friendship; sadly Gill Cowan died recently.

I am grateful to all the members of the John Hughes Bennett Laboratory for encouraging me and helping me through the times when I didn't feel like writing; especially Bernie Ramsahoye, Julie Buchanan, Helen Taylor, Aileen Leask, Anu Bashamboo and Olga Tura.

Finally, special thanks are due to my husband, Derek, and my daughter, Louise, for their understanding and support over the past year.

Abstract

Bacterial Type I restriction and modification (R-M) systems distinguish between their own DNA and that of "foreign" DNA by the methylation state of a target sequence. Previously, Type I systems had been divided into three discrete families (IA, IB and IC) based on DNA hybridisation, antibody reactions and complementation tests. I have added a fourth family, ID, by characterising the system from *Salmonella enterica* serovar *blegdam*. I identified the *hsd* genes; firstly by sequence analysis, then by both functional and mutational analysis. I helped to determine the target recognition sequence ($CGA(N_6)TACC$). I then extended the ID family to include ECOR9I based on sequence similarity and complementation tests.

I investigated the diversity and distribution of Type I systems by firstly using the ID-specific probe to screen 37 strains from the ECOR collection; a reference set of *Escherichia coli* strains that represents the genetic diversity of this species. I found four probe-positive strains. I cloned the *hsd* genes from one of these strains and found that the specificity of the R-M system was different to that of StySBLI. Secondly, I searched the databases of genomic sequences for amino acid sequences similar to those predicted for the archetypal members of the four Type I families. It was already known that the degree of amino acid identity between families is low (20 – 30%) whereas within families it is >80%. In three of the screens I identified putative Type I R-M systems across a wide range of bacterial species. Each of the four screens identified HsdM polypeptides with >45% identity in different phyla. The hybridisation of DNA from the ECOR strains and the *in silico* screen of the databases showed that the distribution of *hsd* genes among different bacterial populations was not consistent with the relatedness of the strains and therefore suggests horizontal transfer.

HsdR contains motifs, characteristic of DEAD-box proteins, that have been identified in ATP-dependent helicases and translocases. I identified two highly conserved sequences (regions X and Y) in HsdR, in addition to the seven motifs (I – VII) already identified. Region X was later proven to identify the active site for endonuclease activity and region Y was later identified as DEAD-box motif IV. I showed that two conservative mutations in motif II inactivated the HsdR polypeptide. HsdR is also implicated in restriction alleviation (RA) after DNA damage. I sequenced the *hsdR* gene from an *E. coli* strain encoding EcoR124I, but impaired in RA, and found that it contained a mutation. Using site directed mutagenesis I showed that reverting this mutation to wild-type sequence restored the wild-type phenotype. RA is also associated with the acquisition of a new R-M system, whereby the host DNA is protected from restriction until it can be modified with the new specificity. An *hsdC* derivative of *E. coli* C dies after the acquisition of the *hsd* genes encoding either EcoKI or EcoAI. I helped to show that the *hsdC* strain had a mutation in *clpX* but not *clpP*. ClpX and ClpP had previously been shown to be necessary for the efficient transmission of EcoKI (Type IA) and EcoAI (Type IB) R-M systems by conjugation, transformation and transduction, although they do not affect members of the two families in the same way. I also corrected the published *hsdR* sequence of EcoR124I/II. The addition of a C near the 3′-end made a frameshift that added an extra five amino acids to the predicted polypeptide sequence.

Contents

Introduction

In 1953 Bertani and Weigle observed that the growth of bacteriophage λ was sometimes inhibited on its natural host, *Escherichia coli* K-12. Phages which had been previously grown on *E.coli* K-12 would grow equally well on both *E. coli* C and *E. coli* K-12 whereas those previously grown on *E.coli* C would only grow well if they were plated back onto *E. coli* C. i.e. their growth was restricted on K-12. This suggested that the phages were being modified in some way and that this modification was determined by the host. Bertani and Weigle also noticed that the restriction was alleviated after the bacteria had been exposed to ultra-violet light. These observations laid the groundwork for the discovery of restriction enzymes and their regulation by the host.

Later it was found that *E.coli* K-12 specifies a restriction endonuclease (REase), EcoKI (Meselson & Yuan, 1968), whereas *E.coli* C does not. Classically, a REase is accompanied by a methyltransferase (MTase) to protect the host DNA. Bacterial restriction and modification (R-M) systems enable their hosts to distinguish between their own DNA and that of "foreign" DNA. They do this according to the methylation state of a target sequence. If the DNA is hemi-methylated, as would be the case for newly synthesised DNA, then the unmethylated strand will be methylated by the modification component of the system. If the sequence is unmethylated, as for "foreign" DNA, then the restriction enzyme will cut the DNA. These R-M systems have been divided into 3 different Types based on: their complexity, their requirement for cofactors, the nature of their target sequence, and the cleavage position. Type I systems are the most complex of the three systems and comprise three subunits (HsdR, HsdM, HsdS) encoded by host specificity of DNA (*hsd*) genes. The M and S polypeptides make up the methylation component and all three subunits are required for restriction. HsdS dictates sequence specificity for both the restriction and modification activities of the enzyme. S-adenosylmethionine (AdoMet) is required for both the REase and the MTase activities, although only the REase requires ATP. The MTase methylates the adenine on both strands of the target sequence. The REase cleaves the DNA some distance away from the target following

ATP-dependent translocation of the DNA through the bound protein. Type II systems are the simplest. They usually recognise a palindromic sequence and comprise two separate enzymes, a REase and a MTase. The REase requires Mg^{2+} whereas the MTase requires AdoMet. Type II REases are widely used as DNA manipulators because the DNA is usually cut within the target sequence. Type III R-M systems have 2 subunits, but both are required for the REase. The MTase requires AdoMet but only one strand of the target sequence is methylated, in contrast to both the Type I and Type II systems. Type III systems require two recognition sequences, in opposite orientations, but the REase only acts if both are unmodified (Meisel *et al.*, 1992). The REase requires ATP and cuts the DNA close to the target sequence.

When DNA probes were made from the *hsd* genes of *E. coli* K-12 and *E. coli* B they were each found to hybridise to the DNA of E. coli K-12 and E. coli B but not to that of E. coli 15T $^-$ (Murray et al., 1982). This led to the separation of enteric R-M systems into different families based on three criteria. Within a family, *hsd* genes will cross-hybridise and antibodies raised to their Hsd proteins will cross-react to each other. In addition a subunit from one R-M complex can be substituted with that of the respective subunit of another complex to make a chimeric enzyme, the specificity of which is defined by the HsdS polypeptide. In contrast, the genes, proteins or subunits will not behave in this manner between members of any other family. EcoKI and EcoBI are founder members of Type IA and EcoAI (from *E. coli* 15T $^-$) is the founder member of Type IB. Later, the Type IC family was added (Price *et al.*, 1987). The *hsd* genes of different families show exceptionally high diversity. Only 20-30% amino acid identity would be expected between members of two different families when comparing their polypeptide sequences (Murray *et al.*, 1993; Sharp *et al.*, 1992). There is also sequence diversity within a family that is due to differences in the regions of HsdS that confer sequence specificity. When the sequences of two *hsdS* genes are aligned there are two regions of non-identity (variable regions) separated by a conserved centre (Gough & Murray, 1983). Each of the variable regions of HsdS recognises one half of the bipartite target sequence, hence they are called target recognition domains (TRDs) (Fuller-Pace & Murray, 1986; Gann *et al.*, 1987; Nagaraja *et al.*, 1985). The length of the conserved centre

determines the length of the intervening spacer in the target sequence; the spacer being defined as the distance between the two adenosyl residues, which are the targets for methylation. The central conserved sequence of HsdS of EcoR124I has a duplicate sequence (TAEL) whereas the HsdS of EcoR124II has a triplicate TAEL (Price *et al.*, 1989). Both HsdS subunits recognise the same bipartite sequence but EcoR124I has only a 7 base pair (bp) spacer whereas EcoR124II has an 8 bp spacer. An exception to the variability of the two TRDs is when a component of the target sequence is shared. This is true even for members of different families. For example, the N-terminal TRDs of StyLTIII (Type IA) and EcoAI (Type IB) have high identity (44%) and they each recognise the sequence GAG (Cowan *et al.*, 1989). In contrast to the relatively high identity that can be found between the TRDs of HsdS sequences of different families, the HsdM and R sequences only share identity at shorter regions. The HsdM polypeptide has the motifs characteristic of adenine MTases whereas HsdR has seven conserved sequence motifs, characteristic of the DEAD-box family of proteins. These motifs are found in many ATP-dependent helicases and are a feature common to all known Type I HsdR subunits (Gorbalenya & Koonin, 1991). It was suggested that these motifs may be associated with helicase activity and that this activity might be involved in the ATP-dependent translocation of DNA (Gorbalenya & Koonin, 1991; Murray *et al.*, 1993).

Type IA and IB R-M systems are chromosomally encoded and their genes behave as alleles in genetic analyses (Arber & Wauters-Willems, 1970), whereas those of IC are generally plasmid-borne. In both *E. coli* and *Salmonella* the chromosomally encoded *hsd* genes can usually be found within a region referred to by Raleigh (1992) as the immigration control region (ICR), which is close to *serB*. I took advantage of this common location by using flanking probes to locate the *hsd* genes of *Salmonella enterica* serovar *blegdam*, a strain suspected of containing a new Type I system (Ryu *et al.*, 1988; Titheradge *et al.*, 1996). After sequencing the blegdam-specific DNA, I noted that the gene order was different to that of the IA and IB systems; namely *serB, hsdM, S and R,* instead of *serB, hsdR, M* and *S* (Titheradge *et al.*, 1996). Also there was no DNA sequence similarity with known members of the Type I A, B and C families. On the basis of the sequence information and

complementation tests, I placed the R-M system of *S. blegdam* (StySBLI) into a 4th family, ID. I identified the recognition sequence of StySBLI as CGA (N_6) TACC and noted that the number of bp between the two methylated bases is six. This number is smaller than that found in the other 3 families. I then investigated the diversity and distribution of Type I systems. I used a ID-specific probe to screen 37 members of the ECOR collection of wild-type *E. coli*. More putative ID members were found, the first in *E. coli* (Barcus *et al.*, 1995). I also used the amino acid sequence data from all 4 families to screen the databases of genomic sequences. I showed that Type I systems of known families are not confined to enteric bacteria but can also be found across other phlya (Titheradge *et al.*, 2001). Lastly, I investigated the mechanism of restriction activity, focusing on the HsdR subunit (Makovets *et al.*, 1998; and 2004; Webb *et al.*, 1996). By aligning the sequence of HsdR from 4 representatives from each of the families I discovered two conserved sequences (initially designated X and Y) (Titheradge *et al.*, 1996) in addition to the seven previously identified (Gorbalenya & Koonin, 1991). The X motif was later proven to identify the active site for endonuclease activity (Davies *et al.*, 1999b), and the Y motif was identified as a replacement for the DEAD-box motif IV previously described by Gorbalenya & Koonin (1991). A mutation made in this motif gives a restriction-deficient phenotype (Davies *et al.*, 1998).

Characterisation of the R-M system from *S. blegdam*

2.1 Cloning the hsd *region of* S. blegdam

S. blegdam was shown to contain a *serB*-linked R-M system that was different to those currently known (Bullas *et al.*, 1980), and subsequently Ryu *et al.* (1988) suggested that *S. blegdam* may represent a new family of Type I enzymes. My aim was to characterise this R-M system and then compare it to other Type I systems. Firstly, I cloned the *hsd* genes from *S. blegdam* in a lambda vector after digesting the genomic DNA with either EcoRI or Sau3A. The latter enzyme was chosen because it recognises a 4-base target sequence, which has many sites in genomic DNA, therefore a partial digest would yield a population of nearly random fragments. It was expected that if the piece of DNA carried by the phage contained both $hsdM_{StySBLI}$ and $hsdS_{StySBLI}$ then the DNA would be modified with *blegdam* specificity and therefore the phage would have a higher efficiency of plating (e.o.p.), compared to the λ vector, on a strain restricting with *blegdam* specificity. I assayed the phages on an *E. coli* / *Salmonella* hybrid (LB4037) (Bullas *et al.*, 1980) which was proficient in restricting with *blegdam* specificity. This hybrid strain was used because *S. blegdam*, like all wild-type *Salmonella* strains, was lambda resistant. Although this method of screening for mod^+ phages had worked well in the past, with $mod_{EcoKI}{}^+$ and $mod_{EcoAI}{}^+$ phages being easily found (Borck *et al.*, 1976; Fuller-Pace *et al.*, 1985), this time $mod_{StySBLI}{}^+$ phages were not found. Since this was a novel Type I system a DNA probe was not available, so I used a flanking probe, containing DNA from *S. enterica* serovar *typhimurium*, to identify phages that were carrying DNA next to the *hsd* region. A λ recombinant was found that contained both flanking sequence and *blegdam*-specific DNA. The latter DNA was then used as a probe and this "chromosome walking" continued until I found a λ recombinant that hybridised to flanking DNA on the other side of the $hsd_{StySBLI}$ genes (see Fig. 2.1). I wanted to ensure that the restriction map deduced from the *blegdam*-specific DNA, (~12kb) contained within the phage recombinants, correlated with that of the chromosomal DNA. To do this, I probed digested chromosomal DNA with probes made from the λ recombinants and showed that the fragment sizes were equivalent.

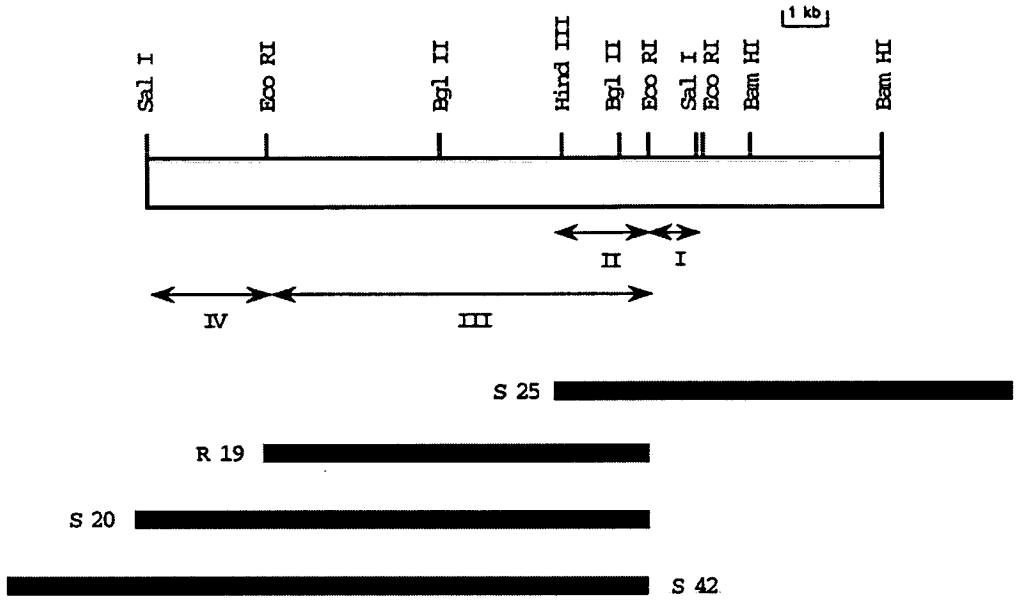


Figure 2.1 Restriction map of the *hsd* region of *S. blegdam*

Roman numerals identify the probes. Probes I – III are *blegdam*-specific. Probe IV hybridises to both *S. blegdam* and *S. typhimurium* DNA. S25, R19, S20 and S42 identify the inserts of λ clones isolated from partial Sau3A (S) or EcoRI (R) libraries. S25 was identified by hybridisation to a flanking probe (containing DNA from *S. typhimurium*); only the BamHI fragment hybridised to this probe. R19 was identified by hybridisation to probe II (made from S25). S20 and S42 were identified by hybridisation to probe III (made from R19).

2.2 Identification of hsd genes by functional analysis

The next step was to confirm that this *blegdam*-specific DNA did indeed contain *hsd* genes. Since none of the λ phages was mod$^+$, I made a mutation, using transposon mutagenesis (Tn*5* *supF*, Phadnis *et al.*, 1989), in the phage carrying the largest *blegdam*-specific insert (λR19). I transferred the mutation to the chromosome of LB4037 where it conferred a res$^-$ phenotype. This confirmed that λR19 contained at least part of the *hsd* region of *S. blegdam*. To discover which *hsd* genes lay within λR19, I transferred the 8.7kb insert to an integration-proficient vector (λNM1151) and made lysogens in NM679, an *E. coli* strain deleted for the ICR region. These lysogens were res$^-$ and some of them were able to modify λ*vir* with *blegdam* specificity, but the reason for the phenotypic variability is unknown. The mod$^+$ lysogens, upon induction, produced phages that retained their protection against restriction by StySBLI. This confirmed the presence of both *hsdM* and *hsdS* within λR19 despite the fact that putative λ*hsdMS* phages fail to modify their DNA on lytic infection. It is known that, despite having many targets, λ is efficiently modified after a single round of growth in strains that contain the other Type I families of R-M systems (Makovets *et al.*, 2004).

2.3 Identification of hsd genes by sequence and mutational analysis

I determined the sequence of the insert in λR19 on both strands and then checked for any frameshifts and open reading frames (ORFs) by using the GeneMark™ program (Borodovsky & McIninch, 1993). No frameshifts and six ORFs, exceeding 200 codons, were found within the EcoRI fragment of λR19; see Fig. 2.2. Four ORFs were complete. However two were incomplete, being interrupted by the EcoRI targets, and the sequences of these were completed from adjacent clones and PCR products. I used the predicted polypeptide sequences deduced from the six ORFs to screen the NCBI database using the BLAST program (Borodovsky *et al.*, 1994). The BLAST program is a local alignment program that searches for the best region of

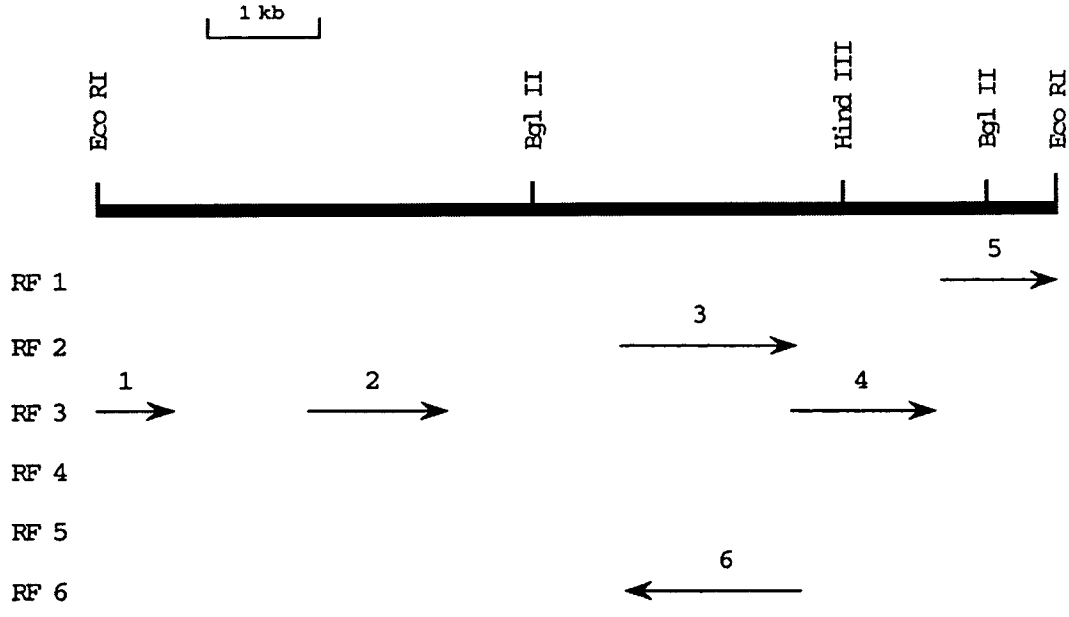


Figure 2.2 ORF's within the EcoRI insert of λR19.

The DNA sequence was translated in all six reading frames (RF 1 – 6). Only ORF's of ≥200 codons are shown, numbered 1 – 6. The arrowheads denote the 3′ ends of the ORF's.

similarity between two sequences. I aligned sequences that shared significant similarity using the PileUp program (Devereux *et al.*, 1984), which aligns sequences along their entire length. Apart from ORF6, all sequences showed similarities indicative of putative functions. ORF1 shared similarities with DNA helicases and ORF5 contained DEAD-box motifs. ORF2 and 3 both had sequences in common with Type I HsdM polypeptides and ORF4 had similarities with some HsdS subunits. Since Type I R-M systems are encoded by three adjacent genes, with *hsdM* and *S* overlapping, *orf3* is a better candidate for the *hsdM* gene because it overlaps the beginning of *orf4*. It was expected that the adjacent *orf5* is *hsdR*. To check that *orf3*, *4* and *5* were the *hsd* genes of *S. blegdam* I did functional tests after transferring the insert of λR19 to a plasmid, pAC18, then subcloning the BglII fragment (with only *orf 3* and *4*) into pUC19. Transformation of an *hsd*Δ strain with this plasmid conferred proficiency in *blegdam*-specific modification, demonstrating that *orf2* is not necessary for modification. I made two mutations in pAC18, one in each of *orf3* and *orf4* and both led to loss of modification. I therefore concluded, in conjunction with the similarities detected by the BLAST program, that *orf3* and *orf4* were *hsdM* and *S* respectively. To further support the evidence that *orf4* is *hsdS*, I compared this nucleotide sequence with that from the putative *hsdS* of EcoR9I. Earlier, I had discovered that DNA from ECOR9, a member of the ECOR collection of *E. coli* strains (Ochman & Selander, 1984), hybridised to my ID-specific probe (Barcus *et al.*, 1995). I made a partial Sau3A library in a λ vector from the chromosomal DNA of ECOR9 (Titheradge *et al.*, 2001). I identified phages that were positive with a *blegdam*-specific probe and then mapped the DNA of some of them by using other *blegdam*-specific probes; one containing just *hsdM*, another containing mostly *hsdS*. D. Ternent subcloned a 4.5kb fragment, which was positive with both probes, and sequenced some of the DNA. When I aligned the putative *hsdS* sequences from both StySBLI and EcoR9I I found that there was a central conserved region and also a conserved C-terminus (see Fig. 2.3). In addition there were two regions of dissimilar sequence that would correlate with the presence of two target recognition domains. This information supports the designation of *orf4* as the *hsdS* of StySBLI. I tested *orf5* by transferring the 1.3kb EcoRI fragment from within *orf5* to a λ phage and then making lysogens in LB4037. If *orf5* was *hsdR* then the λ phage would integrate into

```
StySBLI                                                 MAFEKT IPLNEFITLQ
EcoR9I  MGSNGFKLPL GWNCKKLVDC TKEGNISYGI VQPGQHQEDG IGIIRVNNIQ

StySBLI RGFDLPQDKR VMGDIPVVAS TGVVGYHNEE KVLAPGVVIG RSGSIGGGQY
EcoR9I  NG.......N IYIDDVLKVS HEIESKFAKT RLEGGEVLLT LVGSTGISAI

StySBLI ITTNF..WPL .NTTLWVKDF KGHHPRFVYY LLRSIDFSQF NVGSGV....
EcoR9I  TTKALQGWNV ARAVAVIKPC DEISAEWIHI CLQS.PFTKY FLDSRANTTV

StySBLI .PTLNRNHLS GILVADTSYS YEKEASDIIG ILDDKIKLNK ELNHTLEQIS
EcoR9I  QKTLNLKDVK EIPLPIPPHE ERVSLEKIYF NFENRINLNI KINKILEEMS

StySBLI QTLFKSWFVD FDPVIDNALD AGNPIPEALQ SRAELRQKIR NSADFKPLPA
EcoR9I  QNLFKSWFVD FDPVVDNALD AGNPIPEALQ SRAELRQKVR NSADFKPLPA

StySBLI DIRALFPAEF EETELGWMPK GWITTSFNDL IELIGG.GTP KTSVEEFWNG
EcoR9I  EIRSLFPSEF EETELGWMPK GWQIKSLDHI ANFQNGLALQ KFRPKNMEDD

StySBLI DIPWFSVVDA PSESDVYVLT TEKKITIEGL NNSSAKLLRK GTTIISARGT
EcoR9I  YLPVLKIADL RAGQ....IT NDERARTD.. ISDSCKVY.D GDMIFSWSGT

StySBLI VGKCAMVAVP MAMNQSCYGV IGKNNISDEY IYFQLKNAVQ TLQQMGHGSV
EcoR9I  LMIDIWTGGN AALNQHLYKV TSKK..YPQS FYFMW..TIQ HLSRFQHIAE

StySBLI FNTITRDTFK ..NIKVPFC. ..NEELTNSY SLLVKNYFSK ILNNNYQNIA
EcoR9I  AKAVTMGHIK KGDLSNSFCL IPTSSLITKY DNIVGGYLAK IKNQRLLNNQ

StySBLI LTNLRDTLLP KLISGELSLE DLPNLAKQTE PA
EcoR9I  MTALRDTLLP KLISGELSLD DIPDLNTDTE AA
```

Figure 2.3 An alignment of the amino acid sequences of HsdS from StySBLI and EcoR9I.

Shading indicates conserved amino acids. The alignment was made using the PILEUP program (Wisconsin Package version 10, Genetics Computer Group, Madison, WI, USA).

the chromosome by homologous recombination, creating two incomplete copies of *hsdR*. LB4037 lysogens were found to be restriction-deficient and on excision of the prophage could be restored to wild type. This genetic evidence supports the identification of *orf5* as *hsdR*. Therefore *orf3 orf4* and *orf5* (*hsdM*, *S*, and *R* respectively) encode StySBLI, and although they share a common chromosomal location identified by the ICR, their gene order differs from those members of the IA and IB families found in *E. coli* and *S. enterica*; *hsdR* is downstream of *hsdS*, as found in the plasmid-borne IC family. In addition, the Hsd sequence similarity with representatives of the other Type I families (IA, IB and IC) was between 17 – 29%, consistent with StySBLI being in a different Type I family.

2.4 Complementation between subunits of different enzymes

One of the criteria for the classification of Type I R-M systems into families relies on the ability of subunits from one complex to substitute for those in another to produce a functional chimeric restriction enzyme. As described in the previous section, D. Ternent made a plasmid (pECOR9) that was predicted to contain the modification component of EcoR9I (Titheradge *et al.*, 2001). I used this plasmid for my complementation tests. To do these tests I transformed two derivatives of LB4037 (the *E. coli* strain specifying StySBLI) with pECOR9. In the first derivative, NM856, I made a mutation in *hsdS* by the insertion of *supF* (Titheradge *et al.*, 1996) and in the second, NM867, I made a mutation in *hsdM* (Titheradge *et al.*, 2001). The latter was made by site-directed mutagenesis (SDM) and changed the sequence of MTase motif IV from NPPF to NPPC. The SDM was made within a plasmid and then the mutation was transferred to the chromosome, via a λ*hsd* phage. The complementation results are shown in Table 2.1. The HsdR subunit of StySBLI can substitute for the HsdR subunit of EcoR9I because in both NM856(pECOR9) and NM867(pECOR9) a functional restriction enzyme was made. Similarly, the HsdM subunit of EcoR9I can substitute for the HsdM subunit of StySBLI because NM867(pECOR9) also has restriction activity. In addition, the HsdS subunit of EcoR9I can substitute for the HsdS of StySBLI. When both functional HsdS subunits are present then the strain, NM867(pECOR9), restricts with the specificity of both

Table 2.1 Complementation between the subunits of StySBLI and EcoR9I

Strain	Functional Hsd subunits					e.o.p. of λvir which is unmodified (v.0) or modified against StySBLI (v.SBLI) or EcoR9I (v.EcoR9I)			Relevant phenotype
	StySBLI (on the chromosome)			EcoR9I (on the plasmid)					
	R	M	S	M	S	v.0	v.SBLI	v.EcoR9I	
NM679						1	1	1	r^-m^-
LB4037	+	+	+			$(1.8\pm0.7) \times 10^{-5}$	0.8 ± 0.02	$(4.3\pm5.8) \times 10^{-5}$	$r^+m^+_{StySBLI}$
NM856	+	+				0.7 ± 0.06	0.7 ± 0.1	0.8 ± 0.3	r^-m^-
NM867	+		+			0.9 ± 0.2	0.9 ± 0.1	0.9 ± 0.3	r^-m^-
NM679(pECOR9)				+	+	1.1 ± 0.2	1.0 ± 0.3	1.0 ± 0.4	$r^-m^+_{EcoR9I}$
NM856(pECOR9)	+	+		+	+	$(4.0\pm3.7) \times 10^{-4}$	$(1.2\pm0.7) \times 10^{-4}$	0.5 ± 0.1	$r^+m^+_{EcoR9I}$
NM867(pECOR9)	+		+	+	+	$(3.1\pm1.7) \times 10^{-4}$	$(9.9\pm3.5) \times 10^{-4}$	$(6.4\pm2.9) \times 10^{-4}$	$r^+m^+_{StySBLI}$ $r^+m^+_{EcoR9I}$

Modified from Titheradge *et al*, 2001.

StySBLI and EcoR9I. I also made another derivative of LB4037, NM857, which has a mutation in *hsdR*. It was also made by SDM; the K in the ATP-binding motif is changed to T (Titheradge *et al.*, 2001). J. King & J. Ryu used this derivative and the other two derivatives, NM856 and NM867, for complementation tests using KpnAI (Titheradge *et al.*, 2001). They found that each of the subunits of KpnAI could substitute for the respective subunits of StySBLI. Also if both functional *hsdS* genes were present then the strain would restrict with both KpnAI and StySBLI specificity. These results demonstrate that the three R-M systems (StySBLI, EcoR9I and KpnAI), from different genera, meet the most demanding criteria for membership of the same Type I family.

2.5 The recognition sequence for StySBLI

An *in vivo* method has been used to identify the target site for a Type I R-M system using M13 phages (Gann *et al.*, 1987). These phages were chosen because M13 libraries from previous sequencing projects were available, which had inserts of known sequence. Therefore the DNA sequence of those phages that were restricted by the R-M system could be searched for a common target sequence. Using this method, it was expected that those phages containing a target site for StySBLI would be restricted by LB4037. However, M13 phage can only infect bacteria that have pili, coded by genes contained on the F′ plasmid, therefore I had to make an F′ derivative of LB4037 by transferring the F′::Tn*5* donor from EH22 (Hansen *et al.*, 1984). Six M13 vectors were tested (by M. Ali) and two of them, mp18 and mp19, were restricted by LB4037F′ (Titheradge *et al.*, 2001). They had an e.o.p. of ~0.1 which suggested that they may contain a single target for StySBLI. M13 mp18 and mp19 only differ in the orientation of their polylinker sequence which is longer than that found in the related vectors, mp8 and mp10, see Fig. 2.4. Since neither mp8 nor mp10 is restricted by LB4037F′, the target sequence for StySBLI must reside within the polylinker sequence of mp18 (and mp19), particularly within either the KpnI or SphI site. To discover which of the two sites is relevant, M. Ali selected mutant derivatives of mp18 that were not restricted by StySBLI, after amplification on the *mutD*5 strain, RP526 (Fowler *et al.*, 1974). This strain has a mutation in the

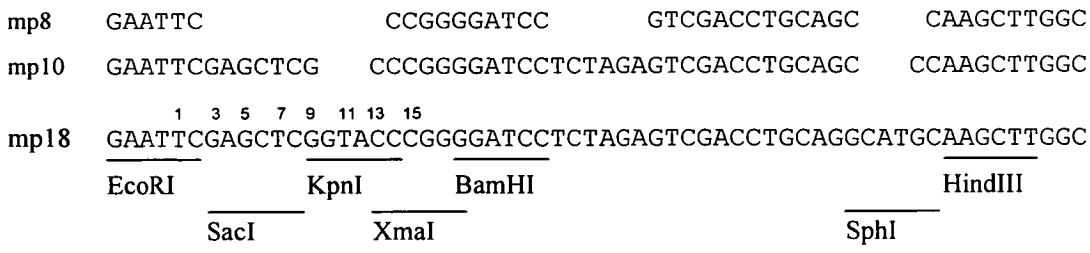


Figure 2.4 Polylinker sequences of M13 vectors.

The sequences of mp8, 10 and 18 are aligned. The numbers (1 – 15) correspond to the bases 1 – 15 in Figure 2.5. Modified from Titheradge *et al*, 2001.

proofreading subunit of DNA polymerase III, therefore misincorporated nucleotides are not corrected and mutations can accumulate. Ali sequenced the DNA of two mutants and found a base substitution in each, within the KpnI target. This indicated that the SphI site was not part of the StySBLI target and therefore the region around the KpnI site should be investigated. D. Ternent cut M13 mp18 independently with the four enzymes that surround the KpnI site. Mutations were made by either removing or filling in the 5′ overhangs. Only changes made within the SacI target removed the StySBLI recognition sequence. The next step was to make changes by SDM in order to identify the StySBLI target sequence, but which bases should be chosen? So far, all Type I recognition sequences are bipartite; the first part being 3 or 4 bp separated by 6 – 8 non-specific bases from the second part that is 4 or 5 bp. Each part includes an adenine residue (methylation substrate), on opposite strands, approximately ten bp apart (Dryden, 1999; Redaschi & Bickle, 1996). When I considered the results then only the A at positions 4 and 11 (opposite strand) fit these constraints, so I decided to choose these nucleotides, and their neighbours to change by SDM. Since it is possible that the StySBLI R-M complex recognises a degenerate sequence, each nucleotide was replaced by the three alternatives. Changing the nucleotides at positions 3, 4, 11 and 12 abolished StySBLI restriction whereas changing the nucleotide at position 10 retained it[1]. Therefore the nucleotide at position 10 must be part of the non-specific sequence. When I changed the nucleotide at position 5 from G to T the restriction was blocked. It was predicted that this change had created the target sequence for the Dam methylase, GATC, and therefore methylation of the adenosine residue would block restriction. To confirm this, I grew the phages on a *dam*$^-$ strain and consequently the StySBLI restriction was restored. Therefore the nucleotide at position 5 is also part of the non-specific sequence. Changes made at positions 2, 13 and 14 also abolished StySBLI restriction whereas changes made at positions 1 and 15 did not[2]. The results are summarised in Fig. 2.5 and show that the target for StySBLI is CGA(N_6)TACC. The 6 bp spacer

[1] D. Ternent made one of the substitutions at position 4. A. Daniel made the three substitutions at positions 11 and 12

[2] D. Ternent made one of the substitutions at position 2. M. Ali had already made substitutions, using a *mutD* strain; one at position 13 and two at position 14.

	1	2	3	4	5	6 – 9	10	11	12	13	14	15
	t	**C**	**G**	**A**	g	ctcg	g	**T**	**A**	**C**	**C**	c
A	√	x	x	n.a.	√	n.d.	√	x	n.a.	x	x	√
C	√	n.a.	x	x	√	n.d.	√	x	x	n.a.	n.a.	n.a.
G	√	x	n.a.	x	n.a.	n.d.	n.a.	x	x	x	x	√
T	n.a.	x	x	x	√*	n.d.	√	n.a.	x	x	x	√

Figure 2.5 Base substitutions made to identify the target recognition sequence of StySBLI.

The nucleotides are numbered 1 – 15, as in Fig. 2.4. The target recognition sequence is in bold, uppercase letters. Substitutions were made for the bases numbered 1- 5 and 10 – 15. √ denotes a base change that has no effect on the target recognition sequence of StySBLI. x denotes a base change that destroys the target sequence. n.a., not applicable. n.d, not done. * When G is replaced with T at this position, the A residues within the sequence GATC become the substrate for the Dam methylase; methylation of the A residues blocks the restriction by StySBLI. The substitutions were made by several workers; the shaded symbols indicate the work that I carried out.

sequence between the adenine residues, the targets for methylation, is shorter than those found for other Type I R-M systems; the IA family has 8 bp, IB has 9 bp, and IC has 7 or 8 bp. The 8 bp spacer within the target site of the IC family corresponds to an extra TAEL sequence in the central conserved sequence.

StySBLI is the first member of a fourth family of Type I R-M systems, ID. The four families have representatives in *E. coli*, *Salmonella*, *Citrobacter*, and *Klebsiella*. Three of these families are *serB*- linked and the fourth is plasmid-borne. My aim was to use the DNA and sequence of StySBLI, together with representatives from the other families, as a probe to increase the current information relating to the diversity of Type I systems. Firstly, DNA probes were used to screen selected members of the ECOR collection of wild-type *E. coli* strains (Barcus *et al.*, 1995). Then the amino acid sequence of the *hsd* region from *S. blegdam*, and a representative from each of the two other families of allelic *hsd* genes, was used to search genomic databases for similar sequences (Titheradge *et al.*, 1996; 2001).

3.1 Hybridisation, cloning and analysis of the ECOR collection

3.1.1 Hybridisation

The ECOR collection is a reference set of 72 wild-type *E. coli* strains that represents the genetic diversity of this species (Ochman & Selander, 1984). Fig. 3.1 shows the dendrogram derived by Herzer *et al* (1990) based on multilocus enzyme electrophoresis (MLEE) of 38 polymorphic enzymes. Classification by MLEE is believed to give a reliable estimate of genetic relatedness of strains and generally agrees with classification based on restriction analyses and DNA sequence comparisons (Milkman & Bridges, 1990; 1993; Ochman *et al.*, 1983). 37 of the 72 strains of the ECOR collection were chosen for the hybridisation studies; all 25 members of group A, since they are the most closely related to *E. coli* K-12, and a few representatives from each of the other groups. The work was shared between myself and V. A. Barcus; my contribution is illustrated in Fig. 3.2. The genomic DNA of the strains was digested with EcoRI before hybridisation to the family-specific probes (see Fig. 3.3). DNA from 12 of the 25 members of group A hybridised to one of the probes; nine to the IA probe, two to the IB probe and one to the ID probe. DNA from five of the 12 remaining strains hybridised to one of the probes; one to IA, one to IB and three to ID. No DNA hybridised to more than one

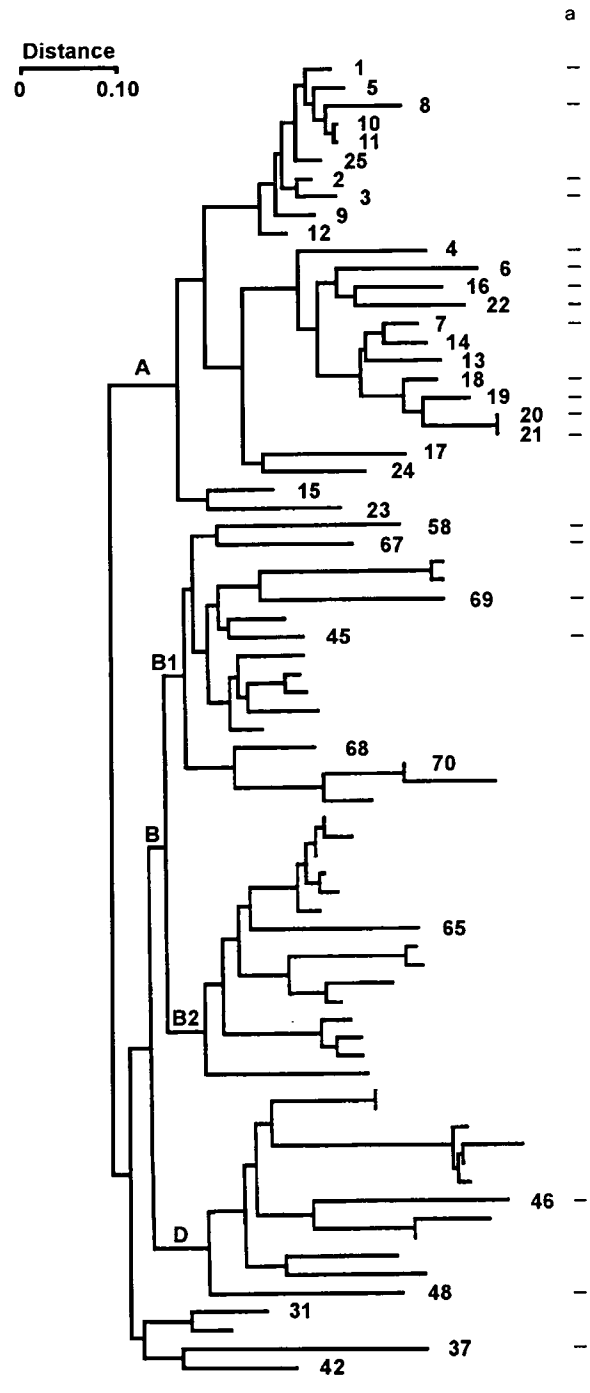


Figure 3.1. Dendrogram of the ECOR collection showing the 37 strains that were chosen for this study.

The dendrogram was derived by Herzer *et al* (1990) based on multilocus enzyme electrophoresis of 38 polymorphic enzymes. Column a indicates the strains that were negative when probed with the family-specific probes. See Table 3.1 for information about those strains that were positive with the probes.

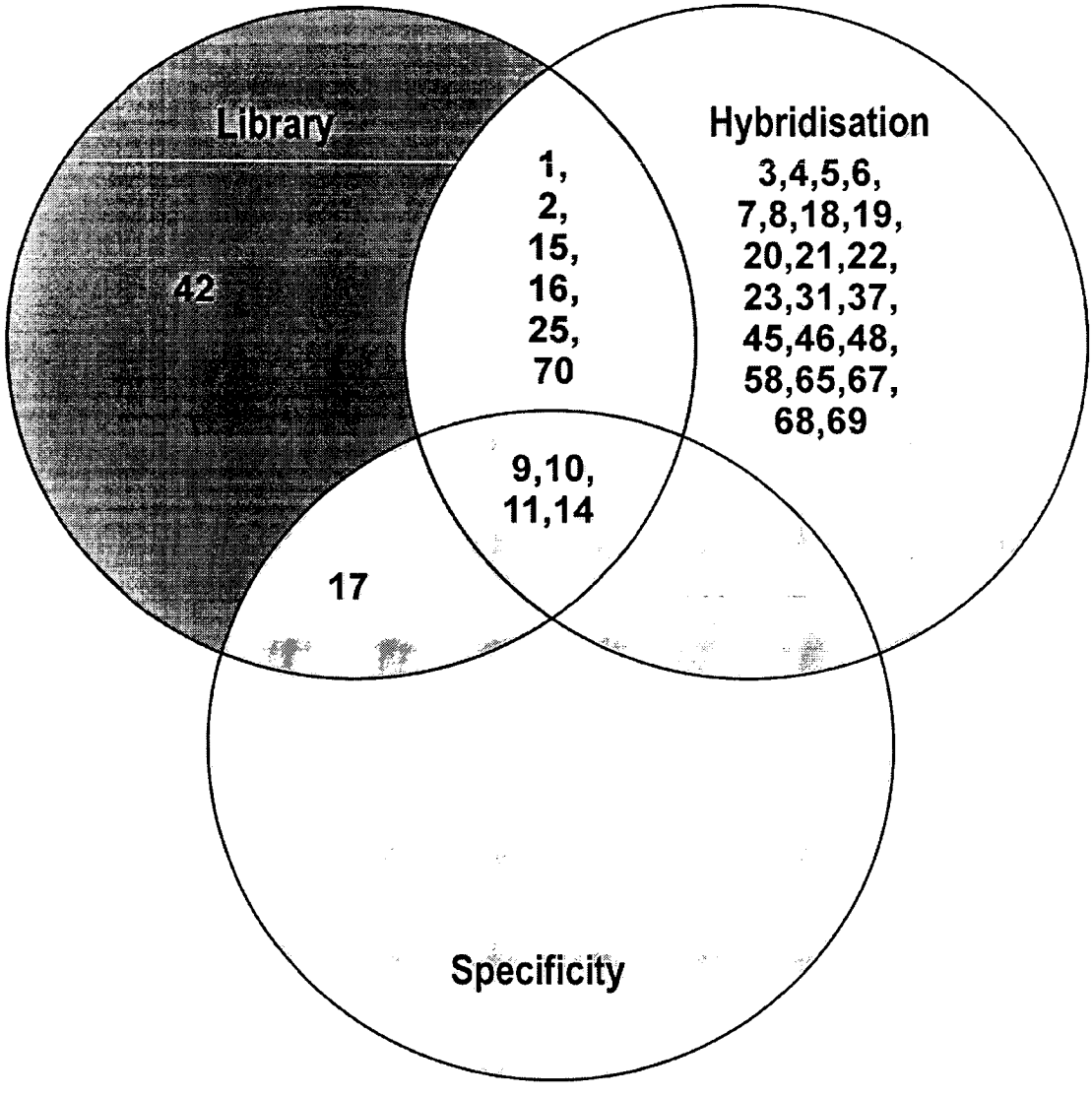


Figure 3.2. My contribution to the ECOR collection analysis, illustrated by a Venn diagram

1 - 69: ECOR strain numbers.

The numbers in the circles refer to the strains:

- I used to construct the libraries (Library)
- I analysed to determine the specificity of the host R-M system (Specificity)
- I used in the hybridisation studies (Hybridisation). All were hybridised with the ID-specific probe. Some of these were also hybridised to the IB-specific probe (31,37,45,46,48,58,65,67,68,69) and some (2, 5, 11, 14, 25, 70) to three probes (IA, IB, and *mcr*).

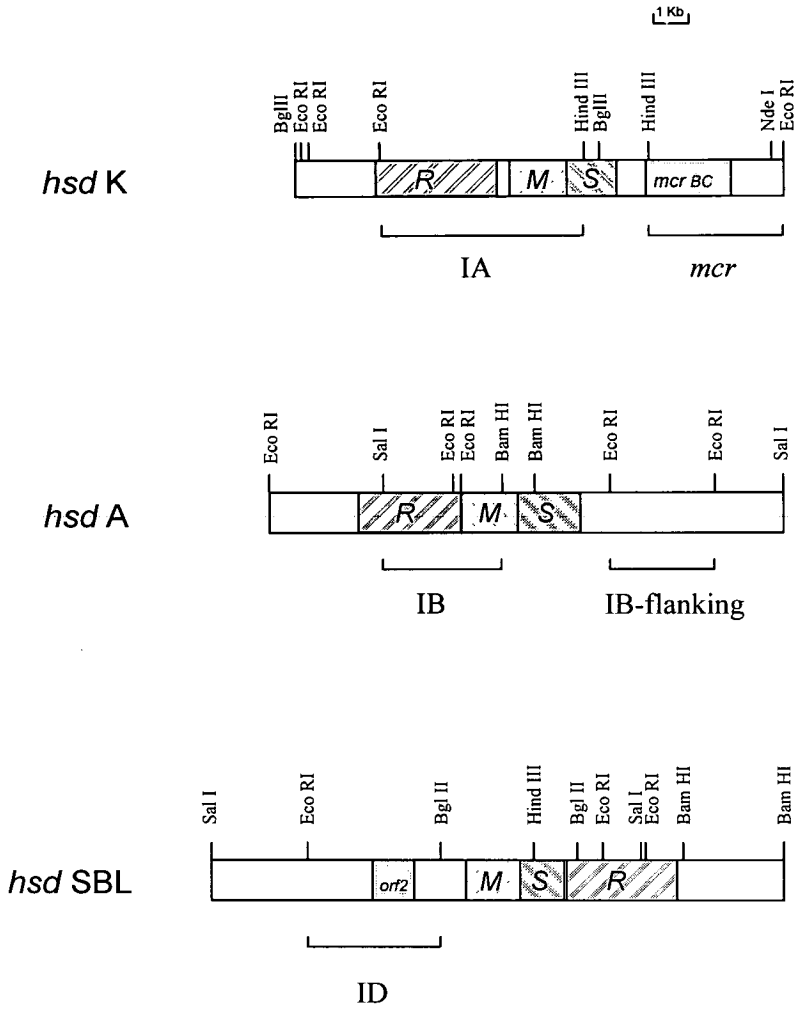


Figure 3.3. Probes used in the hybridisation analysis of the ECOR collection. The maps show the organisation of the *hsd* genes in *E. coli* K-12 (*hsd* K), E.coli 15T⁻ (*hsd* A) and *S. blegdam* (*hsd* SBL). IA: IA family-specific probe and is derived from plasmid pRH1 (Murray *et al.*, 1982). *mcr*: spans *mcrBC* and is derived from plasmid pRH2 (Daniel *et al.*, 1988). IB: IB family-specific probe and is derived from plasmid pFFP32 (Fuller-Pace *et al.*, 1985). IB-flanking: derived from plasmid pFFP31 (Fuller-Pace *et al.*, 1985). ID: ID family-specific probe and is derived from plasmid pAC18 (Titheradge *et al.*, 1996).

probe. In addition, DNA from the ECOR strains was tested with the *mcr* probe (see Fig. 3.3). The *mcr* probe contains a partial *mcrB* gene and a complete *mcrC* gene; genes which encode the McrBC restriction system (Raleigh, 1992). Only DNA from ECOR strains that hybridised to the IA probe also hybridised to the *mcr* probe. The exception to this was DNA from ECOR5, which didn't hybridise to the *mcr* probe despite hybridising to the IA probe. These results agree with previous studies where the *mcrBC* genes have only been found in strains encoding Type IA R-M systems (Daniel *et al.*, 1988). This suggests that the Type IA *hsd* and *mcrBC* genes have been acquired together. However, as is the case for ECOR5, not all strains that have Type IA *hsd* genes also have *mcrBC* genes.

3.1.2 Cloning and analysis

The IA probe hybridises to a single EcoRI fragment in *E. coli* K-12, which contains all of *hsdMS* and most of *hsdR* (Sain & Murray, 1980). This probe also hybridises to a large (11 – 12.5 kb), single EcoRI fragment in five of the ECOR strains (5, 12, 13, 23 and 24). It was expected that this fragment would also contain the respective *hsd* genes, so V.A. Barcus cloned the EcoRI fragment from each of the strains in a λ vector and analysed the specificity of the R-M systems (see Table 3.1). As described in the previous chapter, I transferred the *hsd* genes of ECOR9, which hybridised to the ID probe, into a λ vector. I determined that the specificity of the R-M system in ECOR9 (ECOR9I) was different to that of StySBLI. The DNA from ECOR31, 65 and 68, which also hybridised to the ID probe, was not cloned. The DNA of the remaining eight ECOR strains (10, 11, 14, 15, 17, 25, 42 and 70), that were positive with a family-specific probe, was partially digested with Sau3A prior to a library being made in SuperCos1. I made all these libraries (see Table 3.1 for a summary of the results). This cosmid vector was chosen because relatively large DNA fragments could be inserted (30 - 42 kb) and then efficiently recovered; therefore the whole *hsdRMS* segment could theoretically be cloned making the analysis of novel R-M systems much easier. I screened five of the libraries (ECOR10, 11, 14, 15 and 17) for colonies containing *hsd* genes by hybridising with family-specific probes; IA for ECOR10, 11 and 14 or IB for ECOR15 and 17. I then tested these positive colonies

Table 3.1. Summary of the analyses of libraries made from the DNA of ECOR strains that were positive with a family-specific probe

DNA source (ECOR strain No.)	Vector[a]	Probe used to screen library	Specificity
5	λ	IA	new
9	λ	ID	different to StySBLI
10	cosmid	IA	new
11	cosmid	IA	same as EcoBI
12	λ	IA	same as EcoKI
13	λ	IA	new
14	cosmid	IA	same as EcoBI
15	cosmid	IB	no information
17	cosmid	IB	*
23	λ	IA	new
24	λ	IA	same as EcoKI
25	cosmid	IA	same as EcoR10I
31	n/d	ID	n/d
42	cosmid	IB	new
65	n/d	ID	n/d
68	n/d	ID	n/d
70	cosmid	IA	same as EcoBI

[a] λ libraries were made from EcoRI digested DNA. Cosmid libraries were made from Sau3A partially digested DNA. * P. Thorpe found this specificity to be new. n/d: not determined. ☐ The analyses were shared between myself and V.A. Barcus; the shaded areas indicate the work that I carried out.

from ECOR10, 11, 14 and 17 for their ability to restrict λ*vir*.0. Only ECOR10 and 14 gave res^+ colonies; two out of a total of ten for ECOR10 and one out of nine for ECOR14. I tested all the positive colonies from ECOR10 and ECOR14 for their ability to modify DNA. I did this by propagating λ*vir* phages on each of the strains, and then checking their e.o.p. on a non-restricting strain (DL795) versus a restricting strain; containing either ECOR10I or ECOR14I as appropriate. I found that only those strains that were res^+ modified their DNA. The final step of the analysis was to determine the specificity of ECOR10I and ECOR14I. I did this by plating λ phages, which had been modified to each of the known specificities, onto the strains containing ECOR10I and ECOR14I. If the query strain has the same specificity as that imparted to one of the test phage then the latter would have an e.o.p. of 1 on the query strain. However if the query strain has a new specificity then all the test phages would be restricted on the query strain. The result was that ECOR10I has a new specificity whereas ECOR14I has the same specificity as EcoBI. None of the positive clones picked from the ECOR11 and ECOR17 libraries were restriction proficient using phage λ as a test phage. However, a λ phage containing *hsd* DNA from ECOR11 was able to plate on a strain containing EcoBI with an e.o.p. of 1. Therefore ECOR11 has the same specificity as EcoBI.

To summarise: 17 out of the 37 strains of the ECOR collection hybridised to family-specific probes; 10 to IA, three to IB and four to ID. The 20 strains that didn't hybridise may contain novel R-M systems that couldn't be detected by the probes. Five new specificities were added to the IA and IB families. The strains that hybridised to the ID probe are the first to be identified in *E. coli*. The presence, or absence, of any R-M system didn't correlate with the relatedness of the bacterial strains and therefore suggests horizontal transfer of the *hsd* genes.

3.1.3 Chromosomal location of the hsd genes detected by hybridisation

Although *hsd* genes were detected in 17 strains of the ECOR collection, their position on the chromosome was unknown. Are they like the IA, IB and ID families;

i.e. linked to *serB*? To answer this question I probed the positive clones from 14 of the strains with two flanking probes (Barcus *et al.*, 1995). The DNA for the remaining three strains (ECOR31, 65 or 68), which had hybridised to the ID probe, hadn't been cloned therefore their chromosomal location couldn't be determined. The probes contained DNA downstream of the *hsd* region and are shown in Fig. 3.2; *mcr*, IB-flanking. Nine of the ten clones that were positive with the IA probe also hybridised to the *mcr* probe. The tenth clone, ECOR5, hybridised to the IB-flanking probe. The three clones that were positive with the IB probe, and the one clone that was positive with the ID probe, hybridised to the IB-flanking probe. These results show that at least 14 members of the ECOR collection have *hsd* genes that share a common chromosomal location with the archetypal members of the IA, IB and ID families.

3.2 Database searches

I searched the databases of completed & incomplete genomic sequences for amino acid sequences similar to those predicted for the archetypal subunits of the Type IA, IB, IC and ID families (i.e. EcoKI, EcoAI, EcoR124II and StySBLI) (Titheradge *et al.*, 2001). The aim was to discover if sequences of putative R-M systems could be found in other genera or phyla of bacteria and, if so, were they divided into families such as are found within enteric bacteria. Prior to the searches I corrected the sequence of HsdR of EcoR124I/II. On sequencing *hsdR* I found that there was an additional C at position 3064. This changes the C-terminal sequence from FRKSSRLLRSLKA to FQKIVSFIEKFKGVGGKI.

Prior to 2001, the reported levels of identity between families ranged from 25 to 33% within HsdM, and 17 to 26% within HsdR (Murray *et al.*, 1993; Sharp *et al.*, 1992; Titheradge *et al.*, 1996); these values reflecting the shared sequences of MTases and DEAD-box motifs respectively. I used the TBLASTN program (Altschul *et al.*, 1997) to screen the databases and then I used the PILEUP program to align similar sequences along their whole length. The BESTFIT program calculated the percentage identity. I analysed the results in two steps: firstly, for each archetypal HsdM sequence, I looked for sequences that matched with an identity higher than

45%. This percent was chosen because it is higher than that previously found in an interfamily comparison, but lower than any reported for an intrafamily comparison. All of the HsdM alignments, with identities >45%, were > 95% of the length of the query sequence, with the exception of *Salmonella paratyphi* (which had frameshifts in its sequence). Secondly, for those genomic sequences that matched an archetypal HsdM sequence with an identity of >45%, I looked for the presence of a complete putative R-M system i.e. a putative HsdS and HsdR close to the HsdM I had found. See Table 3.2 for a summary of the results. Only those strains in which all three polypeptide sequences were found are listed. I found 21 strains, representing four different phyla, which matched with the archetypal subunits of the Type I families; five with IA, six with IB, eight with IC and two with ID. No sequence is present in more than one group, and I found that comparisons between the HsdM sequences in the same group shared >42% identity whereas comparisons made between groups shared <35% identity. This supports a family grouping of the putative Type I R-M systems. Also the order of the *hsd* genes on the chromosome correspond to that expected of their group; *hsdR*, *hsdM*, *hsdS* for IA and IB; *hsdM*, *hsdS*, *hsdR* for IC and ID. The levels of identity found for the HsdR sequences, within a group, were lower than those found for the HsdM sequences; the lowest being 37%. This was not unexpected because HsdR (~1,000 aa) is approximately twice as long as HsdM (~500 aa) and the majority of the HsdR sequence is not within a predicted catalytic domain (Murray *et al.*, 1993). However 37% is still a higher identity than that previously reported for interfamily comparisons (17 – 26%). The comparisons made between the HsdS sequences are more difficult to analyse because HsdS has two TRDs, each recognising one half of the bipartite target sequence. Comparisons made between TRD sequences, both within and between families, only show high identity (~ ≥ 40%) if the TRDs confer the same sequence specificity (Cowan *et al.*, 1989). In contrast, the HsdS sequence separating the two TRDs (central conserved region) is highly conserved within a family, therefore I used this sequence to screen the databases. The central conserved regions for each HsdS sequence have been defined by Sturrock and Dryden (1997) and are relatively short; EcoKI is 56aa (shortest) and EcoAI is 155aa (longest). I found that the HsdS central conserved sequences, like the HsdR sequences, could be placed in the same group as their respective HsdM

Table 3.2 Sequences identified by TBLASTN

(A)Percent identity with EcoKI polypeptide sequences (Type IA)

Classification	Bacterial strain	HsdR[a] (1188aa)	HsdM[a] (529aa)	S[b] (56aa)	Reference[c]
Proteobacteria; *Gamma* subdivision	*Salmonella typhimurium* LT2[d]	74(81)[e]	92(95)	71	StyLTIII, WUGS 99287 contig 1424
Proteobacteria; *Gamma* subdivision	*Salmonella typhi* CT18	91(95)	93(94)	80	Sanger ORFS STY4884, 3 & 1
Proteobacteria; *Gamma* subdivision	*Salmonella paratyphi* A	65(73)[e]	90(91)[e]	80	WUGSC 32027
Proteobacteria; *Gamma* subdivision	*Shewanella putrefaciens*	39(57)	54(69)	71	TIGR24 6431
Firmicutes	*Bacillus stearothermophilus*	37(55)	49(63)	34	UOKN03 1422 contig 715

(B)Percent identity with EcoAI polypeptide sequences (Type IB)

Classification	Bacterial strain	HsdR[a] (810aa)	HsdM[a] (489aa)	S[b] (155aa)	Reference[c]
Proteobacteria; *Gamma* subdivision	*Escherichia coli* 0157:H7EDL933	99(99)	98(98)	88	M.EcoO157 ORF 5947P
Proteobacteria; *Gamma* subdivision	*E. coli* A58	77(87)	90(94)	85	EcoEI (L18759, J03162)
Proteobacteria; *Gamma* subdivision	*Pseudomonas putida* KT2440	61(77)	67(78)	52	TIGR10787
Proteobacteria; *Gamma* subdivision	*Acidithiobacillus ferrooxidans*	56(71)	63(78)	62	TIGR6149
Firmicutes	*Streptococcus pneumoniae*	48(65)	49(66)	35	TIGR3836
Actinobacteria	*Mycobacterium avium*	39(55)	48(66)	i.d.	TIGR332

(C)Percent identity with EcoR12411 polypeptide sequences (Type IC)

Classification	Bacterial strain	HsdR[a] (1033aa)	HsdM[a] (520aa)	S[b] (72aa)	Reference[c]
Chlorobi	*Chlorobium tepidum*	73(84)	84(90)	82	TIGR3499. J.Eisen (pers. comm.)
Proteobacteria; *Beta* subdivision	*Neisseria gonorrhoea*	74(85)	75(87)	49	AE004969
Proteobacteria; *Beta* subdivision	*Neisseria meningitidis* serotype A	72(84)[e]	75(86)	31	M.NmeA ORF1038P
Proteobacteria; *Gamma* subdivision	*Haemophilus influenzae* Rd	76(86)[e]	66(72)[e]	53	M.Hind ORF 215P
Firmicutes	*Streptococcus equi*	67(82)	66(70)	61	Sanger 1336 contig 445
Proteobacteria; *Gamma* subdivision	*Xyella fastidiosa*	42(60)	55(70)	54	M.Xfa ORF2728P[f]
Proteobacteria; *Epsilon* subdivision	*Hylicobacter pylori* J99	42(62)	52(69)	46	M.Hpy99 ORF786P
Proteobacteria; *Epsilon* subdivision	*H. pylori* 26695	43(62)	54(70)	50	M.HpyA ORF 850P

(D)Percent identity with StySBLI polypeptide sequences (Type ID)

Classification	Bacterial strain	HsdR[a] (1088aa)	HsdM[a] (539aa)	S[b] (75aa)	Reference[c]
Proteobacteria; *Gamma* subdivision	*Salmonella enteritidis*	96(96)[e]	100(100)	100[e]	UIUC 592 contigs 1881 & 2214
Proteobacteria; *Gamma* subdivision	*Klebsiella pneunioniae*	94(96)	97(98)	95	KpnAI (U93843)

[a] Percent similarity indicated in brackets. [b] Only the central conserved sequence of HsdS was compared. [c] See REBASE (Roberts & Macelis, 2001) for systems identified as enzyme or protein (P) sequence. [d] Serovar of *S. enterica*. [e] Sequence alignment impaired by putative frame shifts. The numbers given are for the longest alignment. [f] The M and S coding sequences are separated by a short ORF (ORF2727), it is not known whether this ORF is an artifact of cloning, or indicates a natural insertion within the coding sequence. i.d. insufficient data.

sequences (See Table 3.2). Also I found a repeat sequence (within the HsdS central conserved region) in some of the new members of the Type IC family, similar to that in the archetypal subunits; TAEL is present in duplicate in EcoR124I and in triplicate in EcoR124II (Price *et al.*, 1989). TAEL is also present in duplicate in *Chlorobium tepidum* and *Streptococcus equi*. A variation of this sequence (TSEL) is found in triplicate in *Haemophilus influenzae* Rd and similarly: EAEL in *Xyella fastidiosa* and NTEL in *Helicobacter pylori* 26695. Also, *Neisseria gonorrhoea* contains the sequence EATL in duplicate followed by EAEL (Piekarowicz *et al.*, 2001). The repeats in EcoR124I/II affect the spacing between the adenosyl residues in the target recognition sequence; EcoR124I has a spacer of 7 bp (TAEL in duplicate) whereas EcoR124II has a spacer of 8 bp (TAEL in triplicate) (Price *et al.*, 1989). It would be expected that the new members of this group would also have variable spacers in their target sequences depending on the number of their repeat sequences.

In summary: Each of the four screens identified HsdM polypeptides with >45% identity in different phyla and these HsdM sequences are associated with HsdS and R sequences; identifying putative R-M systems. The HsdM sequences in the same group share >42% identity but <35% identity with HsdM sequences in other groups. In addition, no sequence is present in more than one group. This supports the grouping of the putative R-M systems into families and shows that Type I systems are not confined to enteric bacteria but can also be found across other phyla.

Lastly, I searched the databases using the whole HsdS sequence (434 aa) of StySBLI (Titheradge *et al.*, 2001). I found eight strains that had HsdS sequences with >30% identity. They were examined further by looking for identity in each of three regions of the HsdS subunit; N-TRD, central conserved region and the C-TRD. The highest level of identity (50%) in a TRD was found in the C-TRD of *C. tepidum*, however its N-TRD has only 10% identity (see Table 3.3). Other HsdS subunits with low identity (≤ 12%) in the N-TRD were found in *E. coli* ECOR9 and *Pasteurella multocida* PM70, but in contrast they had relatively low identity with the C-TRD (21% and 34% respectively). Four of the remaining five alignments had low identity in the C-TRD (15 – 24%) but a higher identity of 41 – 46% in the N-TRD. An identity of

Table 3.3 Sequence comparisons based on HsdS of StySBLI

Bacterial strain	Percent identity	Length of alignment	Percent identity in N-TRD	Centre	C-TRD	Reference[a]
S. enteritidis	100	434[b]	100	100	100	UIUC 592 contig 2214
E. coli ECOR9	43	300	[c]	89	21	EcoR9I
K. pneumoniae	44	437	41	95	21	KpnAI
C. tepidum[d]	44	327	10	34	50	TIGR 3499
Pasteurella multocida PM70	36	344	12	48	34	CBU MN747 AE006190
Ps. syringae pv tomato	34	433	44	43	21	TIGR 323
A. ferrooxidans[d]	32	427	46	38	15	TIGR 6154
A. actinomycetemcomitans	32	421	43	24	24	OUACGT714

[a] For enzymes and proteins see Rebase (Roberts & Macelis, 2001). [b] Sequence alignment impaired by putative frameshift. [c] Insufficient identity in the N-TRD for the TBLASTN program to make an alignment. [d] The subunit is not that identified in Table 3.2.

>40% between two TRDs may indicate that they recognise the same component of a bipartite target sequence. An alignment between the N-TRDs of HsdS of EcoAI and StyLTIII has 44% identity and correlates with their recognition of the trinucleotide GAG. The remaining alignment, from *S. enteritidis*, has 100% identity with HsdS of StySBLI. This correlates with the biological information that *S. blegdam* and *S. enteritidis* have an R-M system with the same sequence specificity (Bullas *et al.*, 1980).

Mechanism of restriction activity

4.1 Dead-box motifs

HsdR of StySBLI contains motifs, characteristic of DEAD-box proteins, which have been identified in ATP-dependent helicases and translocases. I made a full-length alignment of the HsdR polypeptide sequence predicted for StySBLI with the HsdR sequences predicted for the other archetypal Type IA, IB and IC families (i.e. EcoKI, EcoAI and EcoR124I) (Titheradge *et al.*, 1996). I then scanned the alignment for the seven DEAD-box motifs identified by Gorbalenya and Koonin (1991). Six of these were readily found (See Fig. 4.1) but motif IV was unconvincing. However Gorbalenya and Koonin (1991) did highlight that this motif was difficult to identify in EcoR124I. In addition to the six putative DEAD-box motifs, I identified two highly conserved sequences (see Fig. 4.1); regions X and Y. Region X precedes the seven DEAD-box motifs and was later proven to identify the active site for endonuclease activity (Davies *et al.*, 1999a). Region Y is between motif III and motif IV and was later identified as DEAD-box motif IV; a mutation made in region Y (F to Y) of *hsdR* of *E. coli* K-12, but not that made in the original motif IV (F to S), conferred a restriction-deficient phenotype (Davies *et al.*, 1998).

Conservative changes were made in each of three motifs (I, II and III) in plasmids containing only the HsdR of EcoKI (Webb *et al.*, 1996). In motif I (see Fig.4.1) the K was changed to R, I or T. In motif II the H was changed to D, N or Y. In motif III the A in TAT was changed to D, V or G. The plasmids were tested for their ability to compensate for a chromosomal *hsdR* mutation and it was found that only one, containing the mutation TAT to TGT in motif III, was able to do so, i.e. the other eight mutations inactivated the HsdR polypeptide. However quantification of the restriction phenotype is only possible if the mutation is on the chromosome and not in a multicopy plasmid. I transferred two of the mutations in motif II (H to D, and to Y) to the chromosome via a λ phage (Webb *et al.*, 1996). First I cut the plasmid with SmaI and EcoRI; then I ligated the fragment to the arms of λ1265 (Whittaker *et al.*, 1988), cut with SmaI, and λ1347 (a *c*I857 derivative of λ1266 (Whittaker *et al.*,

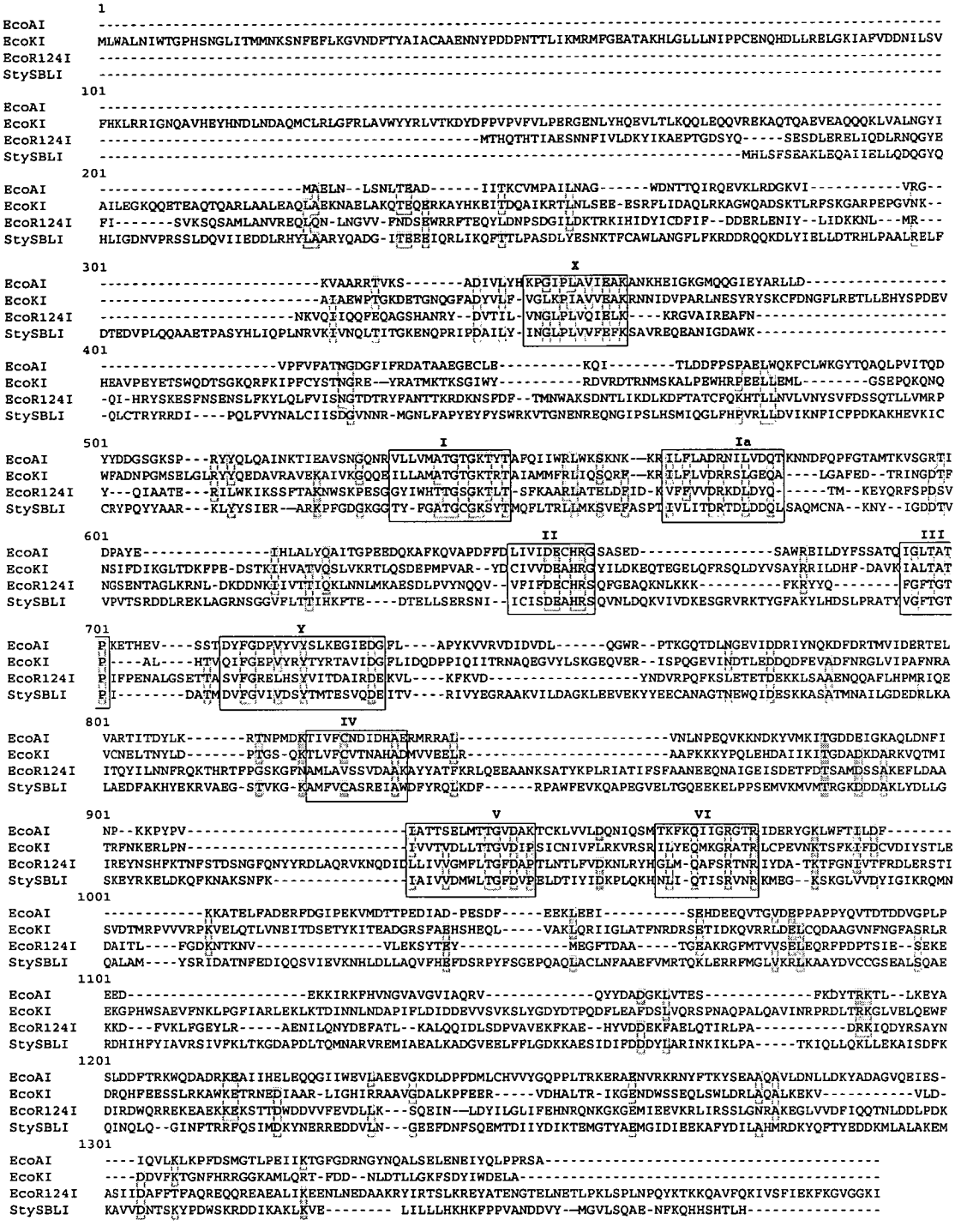


Figure 4.1. Alignment of the HsdR polypeptides of EcoAI, EcoKI, EcoR124I and StySBLI using the PileUp program with a gap weight of 3.0 and a gap length weight of 0.1. Adjustments have been made, from the original alignment found in Titheradge *et al.*, 1996, to start the StySBLI sequence at the first methionine and to end the ECOR124I sequence with the corrected sequence. Shaded regions denote amino acids conserved in at least three of the four polypeptides. Boxed regions I – VI denote the seven putative DEAD-box motifs previously identified in EcoKI (Gorbalenya & Koonin, 1991; Webb *et al.*, 1996). Boxed regions X and Y denote two additional conserved regions

1988)), cut with EcoRI. I recovered the recombinant phage on a *sup°* *hsdΔ* host, which selected against the parent phage (carrying amber mutations). I then used these recombinant phages, containing all but the first few codons of *hsdR*, to infect a restriction proficient *E. coli* K-12 strain. These phages encode a temperature-sensitive repressor (*c*I857) and since they are defective in their attachment site (*b*527), integration is usually achieved by homologous recombination. I selected lysogens at 32°C and tested for EcoKI restriction. If the phages were carrying a mutation that didn't affect the activity of the HsdR polypeptide then lysogens made in a restriction proficient strain would all be res^+. However lysogens made with phages carrying an *hsdR* mutation would be either res^+ or res^- depending on whether the crossover occurred upstream or downstream of the mutation. I found that both DEAH to DEAD and DEAH to DEAY gave both res^+ and res^- lysogens, indicating that the two mutations inactivate the HsdR polypeptide. In the same way, one of the mutations in motif I (K to R) and two in motif III (TAT to either V or G) were transferred to the chromosome. Only TAT to TGT failed to give res^- lysogens. However this mutation gave reduced levels of restriction of phages with only one or two targets, compared to λ*vir* with its full complement of five targets. Mutations made in the other motifs (Ia, V and VI) of EcoKI also result in a restriction-deficient phenotype (Davies *et al.*, 1998).

4.2 Is the hsdC *mutation in a* clp *gene?*

Bacteria can acquire a new R-M system without the host DNA being restricted. This implies that there is a mechanism that protects the DNA from restriction until the DNA can be modified with the new specificity. ClpX and ClpP are necessary for the efficient transmission of the genes specifying EcoKI (Type IA) and EcoAI (Type IB) R-M systems by conjugation, transformation and transduction, although they do not affect members of the two families in the same way (Makovets *et al.*, 1998). An *hsdC* mutation has been found in *E. coli* C that results in the death of the bacteria after the acquisition of the *hsd* genes encoding either EcoKI ((Prakash-Cheng *et al.*, 1993) or EcoAI (Kulik & Bickle, 1996). Perhaps the *hsdC* derivative of *E. coli* C has a defect in *clpX*, *clpP* or both. To answer this question complementation tests were

carried out using four λ*clp* phages and two host strains ($r^-m^-Clp^-$ derivative of *E. coli* K-12 and the *hsdCrecA*$^+$ derivative of *E. coli* C). I constructed three of the phages:

1) λ*clpP*$^+$*X*$^+$ was made by transferring the 6.2 kb BamHI fragment (carrying *clpP* and *clpX*) from the Kohara phage 148 (Kohara *et al.*, 1987) to the λ vector NM1151 (Murray, 1983).
2) λ*clpP*$^+$*X*$^-$ was made by transferring the equivalent BamHI fragment from NM840*clpX* to λNM1151: I prepared DNA from NM840*clpX*, digested it with BamHI and then ligated it into the arms of λNM1151. I identified the plaques containing the required insert by probing.
3) λ*clpP*$^-$*X*$^+$ was made by transferring the HindIII – BamHI fragment from pWPC16 (Maurizi *et al.*, 1990) to λNM1151.

The fourth phage, λ*clpP*$^-$*X*$^-$, was made by S. Makovets. I prepared DNA from all four phages using the new λ miniprep method that I developed (Titheradge *et al.*, 1996). Then I checked that the DNA matched with the expected restriction map and that the phages carried the expected drug resistance. S. Makovets used these four phages to make lysogens, by homologous recombination, of the two strains. She then used these lysogens as recipients for conjugative transfer of F′ plasmids that included the *hsd* genes from *E. coli* K-12. She found that both λ*clpP*$^+$*X*$^+$ and λ*clpP*$^-$*X*$^+$, but neither λ*clpP*$^+$*X*$^-$ nor λ*clpP*$^-$*X*$^-$, complemented the *hsdC* mutation, i.e. the *hsdC* strain had a mutation in *clpX* but not *clpP*.

4.3 Restriction alleviation after DNA damage

In 1953 Bertani and Weigle noticed that restriction was alleviated after the bacteria had been exposed to ultra violet light. The current interpretation of their observation is that there is a mechanism that regulates restriction in response to DNA damage. In the Type IA and IB families a model has been proposed for restriction alleviation (RA); HsdR becomes a substrate for the protease ClpXP after the endonuclease has bound to an unmodified target sequence but before restriction of the DNA has occurred (Makovets *et al.*, 1999). In the Type IC family RA is not dependent on the presence of ClpXP therefore there must be an alternative means of control. To investigate this control a strain encoding EcoR124I was treated with the mutagen

nitrosoguanidine (Makovets *et al.*, 2004) to create mutants that have an impaired regulation of the restriction activity of EcoR124I. Colonies were then screened for their sensitivity to the base analogue 2-aminopurine (2-AP). Treatment with 2-AP causes transitions when the 2-AP pairs with cytosine and this leads to unmodified target sequences in the bacterial chromosome (Makovets *et al*, 1999), therefore bacteria that are unable to regulate their restriction will die from autorestriction (Cromie & Leach, 2001; Makovets *et al.*, 1999). Out of 3000 colonies that were screened, 26 were sensitive to 2-AP (2-AP^s) but for only one of these was the 2-AP^s phenotype dependent on restriction proficiency. A mutation in this strain (NK445) showed tight linkage to a marker close to the *hsd* genes, which suggests that the mutation is in an *hsd* gene rather than an unlinked gene involved in regulation. Since it was a mutation that affected restriction activity it was thought that the likely location would be within *hsdR*. I cloned the 4.9kb EcoRI – BamHI fragment that includes the *hsdR* gene of NK445 in pUC19 (pAT43, see Makovets *et al.*, 2004). I selected the colonies containing pAT43 by hybridisation. A plasmid (pNK19) including the wild-type sequence was available. I sequenced the *hsdR* gene from both pAT43 and the wild type plasmid (pNK19), and found that the nucleotide sequence of *hsdR* in pAT43 differed from the wild type by the change in codon 957 from GCT to GTT. I reverted codon 957 to the wild-type sequence by site-directed mutagenesis and used both pAT43 and the *hsdR*$^+$ revertants to transform an *hsdR*$^-$ M^+S^+ strain. All transformants were restriction proficient, but only those with the mutant *hsdR* sequence were sensitive to 2-AP. Therefore the mutation in codon 957 correlates with the 2-AP^s phenotype, a phenotype consistent with impaired restriction alleviation. Other experiments confirmed that the mutation in *hsdR* impairs restriction alleviation (Makovets *et al.*, 2004) and therefore the phenotype of NK445 resembles that of a *clpX* derivative of a restriction-proficient *E.coli* K-12 (Makovets *et al.*, 1999). These experiments identify the first mutation in *hsdR* that affects the alleviation of restriction.

Discussion

5.1 Characterisation of the R-M system from S. blegdam

Bacteria encoding Type I R-M systems distinguish between their own DNA and that of "foreign" DNA by the methylation state of a target sequence. Previously, Type I systems had been divided into three discrete families (IA, IB and IC) based on DNA hybridisation, antibody reactions and complementation tests. I have added a fourth family, ID, by characterising the R-M system from *S. blegdam*. Initially I tried a screening method to isolate $\lambda mod_{StySBLI}^{+}$ phages, because this method had worked well for the isolation of phages modifying with EcoKI and EcoAI specificity. However mod^{+} phages were not isolated from either an EcoRI or a Sau3A λ library. Later, I made lysogens with a phage known to contain both *hsdM* and *hsdS*, λR19. Only some of these lysogens were able to modify λ*vir* with StySBLI specificity, but the reason for the phenotypic variability is unknown. However the mod^{+} lysogens, upon induction, produced phages that retained their protection against restriction by StySBLI. This variability in the phenotype of the phages may explain why the screen for mod^{+} phages failed.

Since *S. blegdam* was thought to represent a new family of Type I enzymes (Ryu *et al.*, 1988), no *hsd*-specific probe was available to locate the genes. The *hsd* genes were already known to be *serB* linked (Bullas *et al.*, 1980) and Raleigh (1992) reported that the chromosomally encoded *hsd* genes of *E. coli* and *Salmonella* could usually be found within the ICR region, near *serB*. Therefore I used flanking probes to locate the *hsd* genes. However, although these genes share a common location with those of the IA and IB families, they have a different gene order; *hsdR* is located downstream of *hsdS*. This is the same order as found in the plasmid-borne IC family. Since it has been suggested that all Type I R-M systems have a common origin, even those of the IC family (Sharp *et al.*, 1992), it might be expected that *hsd* genes with a common chromosomal location would also have a common gene order. Since the R-M system of *S. blegdam* does not have the same gene order as the IA and IB members then horizontal transfer may be implicated. Recently Sibley and Raleigh (2004) have discovered the boundaries of the ICR region to be the *yjiS* and *yjiA*

genes. They describe the area bounded by the *yjiS* and *yjiA* genes as containing "migratory" genes and suggest that genetic exchange in this region may be accomplished by a site-specific recombinase.

I extended the ID family to include EcoR9I based on sequence similarity and complementation tests. The latter is the most demanding criterion for membership of the same Type I family. The complementation tests were performed using strains harbouring plasmids that contained the *hsd* genes. This was not ideal because plasmid loss may give rise to false negative results. However since the e.o.p. of λ*vir* was compared to reference strains and since positive results were obtained it was assumed that plasmid loss was minimal.

I determined the target recognition sequence of StySBLI (CGA(N_6)TACC) and showed that methylation of the first adenine residue by the Dam methylase prevents attack by StySBLI. All target recognition sequences of Type I R-M systems include two adenosyl residues, which are the targets for methylation, but the spacers between the residues vary in a family specific manner; 9 bp in the IB family, 8 bp in the IA family, 7 or 8 bp in the IC family and only 6 bp in the first member of the ID family (Titheradge *et al.*, 2001). Two members of the IC family have either 7 (EcoR124I) or 8 (EcoR124II) bp spacers depending on whether a tetrapeptide sequence (TAEL) within the central conserved sequence is present in duplicate or triplicate (Price *et al.*, 1989). Recently, Kasarjian *et al.* (2004) elucidated the target recognition sequences, and identified the adenosyl residue targets for methylation, for StySGI (TAAC(N_7)RTCG), KpnAI (GAA(N_6)TGCC), StySEAI (ACA(N_6)TYCA) and StySENI (CGA(N_6)TACC); encoded by *S. gelsenkirchen*, *K. pneumoniae*, *S. eastbourne*, and *S. enteritidis* respectively. They provide evidence that StySGI and StySKI (both Type IB) have a similar specificity; the StySGI target is a degenerate version of the StySKI target. StySGI has a spacer of 9 bp, which is the same as IB family members. KpnAI, known to be a member of the ID family (Lee *et al.*, 1997; Titheradge *et al.*, 2001), has a spacer of 6 bp. StySEAI and StySENI both have a spacer of only 6 bp, with StySENI being an isoschizomer of StySBLI. Since the spacers between adenosyl residues generally vary in a family specific manner,

perhaps StySGI is a member of the IB family and StySEAI and StySENI are members of the ID family.

5.2 Diversity and distribution of R-M systems

Biological evidence indicates that the four families of the Type I R-M systems have representatives in *E. coli*, *Salmonella*, *Citrobacter*, and *Klebsiella*. Three of these families are *serB*- linked and the fourth is plasmid-borne. My aim was to screen the DNA of additional *E. coli* strains, with family-specific probes, to investigate the diversity and distribution of R-M systems. I used the ID-specific probe to screen 37 strains from the ECOR collection (Barcus *et al.*, 1995); a reference set of 72 wild-type *E. coli* strains that represents the genetic diversity of this species (Ochman & Selander, 1984). I found four strains that were probe-positive, the first representatives of this family in *E. coli*. One of these R-M systems (ECOR9I) had a different specificity to StySBLI; the others remain unknown. I (and V. A. Barcus) also used probes from the IA and IB families and found 13 probe-positive strains; 10 with IA and 3 with IB. In total 17 probe-positive strains were found amongst the 37 strains and of these 14 were tested for the location of their *hsd* genes. All of them share a common chromosomal location with the archetypal members of the IA, IB and ID families. The ECOR collection has been divided into groups based on the levels of relatedness (Herzer *et al.*, 1990), however I found no correlation between membership of a family and membership of a group; three different families were found in one group. This is further evidence for the horizontal transfer of *hsd* genes at a common chromosomal location. 20 strains from the ECOR collection didn't hybridise and it could be that they have alternative families of *hsd* genes, or they may be like *E. coli* C and lack *hsd* genes (Daniel *et al.*, 1988). To find the answer, their DNA could be probed with flanking genes and then the intervening DNA analysed; in the same manner as I characterised StySBLI.

I then extended the screen by using family-specific, Hsd, polypeptide sequences to look for identity in databases of completed and incomplete genomic sequences (Titheradge *et al.*, 2001). My aim was to discover if sequences of putative R-M

systems could be found in other genera of bacteria and, if so, were they divided into families. I used the TBLASTN program (Altschul *et al.*, 1997) to search for sequences with significant identity. This program was chosen because it compensates for sequencing errors; a query protein sequence is compared to a nucleotide sequence that has been translated in all six reading frames. However this program only shows alignments between the well-matched areas of the sequence, therefore the level of identity between two sequences, along their entire length, cannot be calculated (unless there is very high identity). To make this calculation the newly identified sequence was aligned with the query sequence using PILEUP and then the BESTFIT program calculated the percentage of identity. For an initial screen I chose the archetypal HsdM sequence from each of the four Type I families. HsdM was chosen because it is smaller than HsdR, and all HsdM subunits share MTase motifs. Although all HsdR sequences have the DEAD-box motifs they show lower levels of identity than HsdM sequences. HsdS sequences have even lower levels of identity because of the two highly variable TRDs. I chose to analyse only those sequences that matched with the query HsdM sequence with an identity >45%; higher than that previously found in an interfamily comparison (20 – 30%), but lower than any reported for an intrafamily comparison (>80%). Those genomic sequences were compared with the respective family HsdS and HsdR sequences i.e. to look for the presence of a complete putative R-M system. Each of the four screens identified HsdM polypeptides with >45% identity in different phyla, with no sequence being identified by more than one probe. In three of the screens I identified putative Type I R-M systems across a wide range of bacterial species. This work suggests new members for each of the four Type I families. In addition, the distribution of *hsd* genes among different bacterial populations was not consistent with the relatedness of the strains and therefore suggests horizontal transfer.

In analysing results obtained by computer searches several points need to be considered:

- An R-M system identified by a screen remains putative until the *hsd* genes have been cloned and demonstrated to be functional.

- Caution needs to be exercised if attempting to assign an R-M system to a particular family based on only sequencing information. In *Lactococcus* the HsdR and HsdM sequences from two plasmids have ~90% identity with each other but only 40% identity with those specified by the chromosome (Schouler *et al.*, 1998). On this basis the chromosomal genes might have been placed in one family and the plasmid encoded genes in another. However, the plasmid encoded HsdS subunits were found to complement the chromosomally encoded HsdM subunits and also their central conserved sequences were highly conserved. This indicates that, by the genetic test, both chromosomal and plasmid encoded genes are actually members of the same family.
- Some of the sequences in the genomic databases may have errors or be incomplete. In my screen of the databases I found that the putative HsdM sequence from *Ureoplasma urealyticum* was interrupted by a rearrangement and the sequence data from the HsdS of *Mycobacterium avium* was insufficient to make an alignment of the central conserved region (Titheradge *et al.*, 2001).
- It is possible to find different levels of identity depending on which computer program is chosen. During the characterisation of the StySBLI R-M system I used the StySBLI polypeptide sequences predicted from the three *hsd* genes to search the databases for similar sequences. I found the highest identity (49.8%) with sequences of *Haemophilus influenzae* Rd (Titheradge *et al.*, 1996). However this strain doesn't appear in the results of the screen for the distribution and diversity of Type I R-M systems (Titheradge *et al.*, 2001). This apparent discrepancy merely highlights the differences between the two computer programs: TBLASTN and PILEUP. Although 49.8% identity was reported using PILEUP, only 43% identity was found using TBLASTN. PILEUP cannot be used to screen databases because it makes an alignment (global) over the whole length of sequences, and cannot just pick out the areas of high identity. Therefore TBLASTN is used first, to locate the genes, and then the whole sequence can be extracted for analysis by PILEUP. Since 43% identity is lower than the cut-off of 45% in the screen for distribution and diversity *H. influenzae* Rd was not included.

Bearing in mind these considerations, an *in silico* screen is still a good starting point for identifying putative R-M systems since the genomic sequences are readily available and the method is easy to use. The information obtained in these comparative analyses can then form the basis of further investigation; e.g. to discover if the putative R-M systems are functional.

The screen of both the ECOR collection and the genomic databases revealed that Type I R-M systems are diverse and are not distributed according to the relatedness of the bacterial strains. However, both screens were designed to look for DNA or sequences with identity to known families of R-M systems. As already discussed, flanking probes could be used to locate the *hsd* genes of ECOR strains that didn't hybridise to the family-specific probes. In the case of the *in silico* screen, sequences that have <45% identity with the query HsdM sequence could be analysed. If a putative *hsdR* and *hsdS* can be found nearby then the polypeptide sequences of this R-M system can be used as a probe to find other matching sequences. By using this itinerative method perhaps other groups/families will be found. Currently, based on sequence information, a phylogenetic tree with four major branches, representing the four families, could be predicted. However as already discussed, caution needs to be exercised if attempting to assign an R-M system to a particular family based on only sequencing information. The classical definition of a family requires that a subunit from one R-M complex can be substituted with that of the respective subunit of another complex to make a chimeric enzyme, the specificity of which is defined by the HsdS polypeptide. Perhaps a comparison between two *hsdS* sequences, excluding the TRDs, could provide a clue about the ability of the HsdS subunits to substitute for each other. High identity, such as that found in the *hsdS* genes of Lactococcus (Schouler *et al.*, 1998) might indicate the ability to complement.

5.3 Mechanism of restriction activity

Type I R-M systems comprise three subunits (HsdR, HsdM and HsdS), with the M and S polypeptides making up the methylation component and HsdR being required

for restriction activity. HsdR contains motifs, characteristic of DEAD-box proteins, which have been identified in ATP-dependent helicases and translocases. HsdR can be divided into several domains. Near the N-terminus there is an endonuclease domain that includes region X, the motif I identified (Titheradge *et al.*, 1996). Region X has been shown to be the active site for endonuclease activity (Davies *et al.*, 1999a) and mutations in this region can translocate but not cut the DNA. The central region of HsdR shares sequence identity with domains 1A and 2A of DNA and RNA helicases (Gorbalenya & Koonin, 1991). Within these domains can be found the seven DEAD-box motifs (I – VI) identified by Gorbalenya and Koonin (1991), although the region Y that I identified (Titheradge *et al.*, 1996) has been shown to be motif IV (Davies *et al.*, 1998). The DEAD-box motifs are associated with ATP binding and hydrolysis, and DNA translocation. I helped to correlate mutations in the DEAD-box motifs with deficiencies in restriction; I demonstrated that two conservative mutations in motif II (DEAH to DEAD and DEAD to DEAY) affected ATP hydrolysis. Mutations have been made in each of the other motifs, with mutations in motif I affecting ATP binding and mutations in III – VI affecting translocation. Recent evidence suggests that the endonuclease translocates unmodified, double-stranded DNA with no helicase activity (Singleton & Wigley, 2002).

Efficient acquisition of a new R-M system, with a specificity different to any R-M system in the host, requires that the host DNA is first modified with the new specificity before restriction activity takes place. Indeed it has been shown that such restriction alleviation (RA) continues for ~15 generations after transmission of *hsd* genes by conjugation (Prakash-Cheng & Ryu, 1993). This could be achieved if the methylase of the new system was produced before the endonuclease, however there is no evidence for the transcriptional control of *hsdR* (Loenen *et al.*, 1987; Prakash-Cheng *et al.*, 1993). Dryden *et al* (1997) have proposed that the assembly of Hsd subunits is under a control, and that both inactive intermediates and HsdR could be susceptible to proteases. *E. coli* K-12 mutants, deficient in proteases, were screened for their ability to acquire R-M genes (Makovets *et al.*, 1998). It was found that both ClpX and ClpP are necessary for the efficient transmission of EcoKI (Type IA) and

EcoAI (Type IB) R-M systems by conjugation, transformation and transduction, although they do not affect members of the two families in the same way. In addition, an *hsdC* mutation has been found in *E. coli* C that results in the death of the bacteria after the acquisition of the *hsd* genes encoding either EcoKI (Prakash-Cheng *et al.*, 1993) or EcoAI (Kulik & Bickle, 1996), but not EcoR124I (Kulik & Bickle, 1996). I helped to show that the *hsdC* strain had a mutation in *clpX* but not *clpP*. It is not only during the acquisition of new *hsd* genes that the host DNA has unmodified targets and is therefore susceptible to restriction, e.g. DNA can be damaged by UV light, or unmodified targets can be created by replication errors. A model has been proposed to explain the phenomenon of RA in the Type IA and IB families, whereby HsdR becomes a substrate for ClpXP after the endonuclease has bound to an unmodified target sequence but before restriction of the DNA has occurred (Makovets *et al.*, 1999).

Since the transfer of a Type IC system is not dependent on the presence of ClpXP in the recipient, there must be an alternative means of control. The RA of EcoR124I was investigated in strains carrying each of eight different protease mutations; none of the mutations prevented the induction of RA in response to 2-AP. Incorporation of the base analogue 2-AP into DNA can cause transitions because 2-AP can pair with cytosine, and therefore new targets for a R-M system can be created which will be unmodified (Makovets *et al.*, 1999); in this situation RA protects the genomic DNA. Since these eight proteases were not involved in RA, a strain encoding EcoR124I was treated with nitrosoguanidine (Makovets *et al.*, 2004) and then colonies were screened for their sensitivity to 2-AP. Nitrosoguanidine creates mutations in the chromosome and then the 2-AP screen would select those defective in RA. Although this screen found 26 2-AP^{s} colonies, in only one of these was the phenotype dependent on restriction proficiency. However this mutation did have tight linkage to a marker close to the *hsd* genes. Since *hsdR* seemed a likely candidate for the location of the mutation, I sequenced this gene and found that it contained a mutation. Using SDM I showed that reverting this mutation to wild-type sequence restored the wild-type phenotype. This was the first time that a mutation in HsdR was shown to impair restriction alleviation. The mutation is in the carboxyl-terminal

region of HsdR therefore perhaps the change affects the interaction of HsdR with the MTase (Davies *et al.*, 1999b).

Abbreviations

AdoMet	S-adenosylmethionine
2-AP	2-aminopurine
ATP	adenosine triphosphate
BLAST	basic local alignment search tool
bp	base pair(s)
DNA	deoxyribonucleic acid
e.o.p.	efficiency of plating
hsd / Hsd	host specificity determinant
ICR	immigration control region
λ	lambda
λvir	lambda virulent
Mg^{2+}	magnesium
MLEE	multilocus enzyme electrophoresis
mod	modification
MTase	methyltransferase
NCBI	National Centre for Biotechnology Information
ORF	open reading frame
PCR	polymerase chain reaction
RA	restriction alleviation
REase	restriction endonuclease
res	restriction
R-M	restriction and modification
SDM	site directed mutagenesis
TIGR	The Institute for Genomic Research
Tn	transposon
TRD	target recognition domain
UOKN	University of Oklahoma
UIUC	University of Illinois at Urbana-Champaign
WUGSC	The Washington University Genome Sequencing Centre

List of figures and tables

Figures

Tables

References

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389-3402.

Arber, W. & Wauters-Willems, D. (1970). Host specificity of DNA produced by Escherichia coli. XII. The two restriction and modification systems of strain 15T. *Mol Gen Genet* **108**, 203-217.

Barcus, V. A., Titheradge, A. J. & Murray, N. E. (1995). The diversity of alleles at the hsd locus in natural populations of Escherichia coli. *Genetics* **140**, 1187-1197.

Bertani, G. & Weigle, J. J. (1953). Host controlled variation in bacterial viruses. *J Bacteriol* **65**, 113-121.

Borck, K., Beggs, J. D., Brammar, W. J., Hopkins, A. S. & Murray, N. E. (1976). The construction in vitro of transducing derivatives of phage lambda. *Mol Gen Genet* **146**, 199-207.

Borodovsky, M. & McIninch, J. D. (1993). GeneMark: parallel gene recognition for both DNA strands. *Comput Chem* **17**, 123-133.

Borodovsky, M., Rudd, K. E. & Koonin, E. V. (1994). Intrinsic and extrinsic approaches for detecting genes in a bacterial genome. *Nucleic Acids Res* **22**, 4756-4767.

Bullas, L. R., Colson, C. & Neufeld, B. (1980). Deoxyribonucleic acid restriction and modification systems in Salmonella: chromosomally located systems of different serotypes. *J Bacteriol* **141**, 275-292.

Cowan, G. M., Gann, A. A. & Murray, N. E. (1989). Conservation of complex DNA recognition domains between families of restriction enzymes. *Cell* **56**, 103-109.

Cromie, G. A. & Leach, D. R. (2001). Recombinational repair of chromosomal DNA double-strand breaks generated by a restriction endonuclease. *Mol Microbiol* **41**, 873-883.

Daniel, A. S., Fuller-Pace, F. V., Legge, D. M. & Murray, N. E. (1988). Distribution and diversity of hsd genes in Escherichia coli and other enteric bacteria. *J Bacteriol* **170**, 1775-1782.

Davies, G. P., Kemp, P., Molineux, I. J. & Murray, N. E. (1999a). The DNA translocation and ATPase activities of restriction-deficient mutants of Eco KI. *J Mol Biol* **292**, 787-796.

Davies, G. P., Powell, L. M., Webb, J. L., Cooper, L. P. & Murray, N. E. (1998). EcoKI with an amino acid substitution in any one of seven DEAD-box motifs has impaired ATPase and endonuclease activities. *Nucleic Acids Res* **26**, 4828-4836.

Davies, G. P., Martin, I., Sturrock, S. S., Cronshaw, A., Murray, N. E. & Dryden, D. T. (1999b). On the structure and operation of type I DNA restriction enzymes. *J Mol Biol* **290**, 565-579.

Devereux, J., Haeberli, P. & Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res* **12**, 387-395.

Dryden, D. T. (1999). Bacterial methyltransferases. In *S-Adenosylmethionine-Dependent Methyltransferases*, pp. 283 - 340. Edited by X. Cheng & R. M. Blumenthal. Singapore: World Scientific Publishing.

Dryden, D. T., Cooper, L. P., Thorpe, P. H. & Byron, O. (1997). The in vitro assembly of the EcoKI type I DNA restriction/modification enzyme and its in vivo implications. *Biochemistry* **36**, 1065-1076.

Fowler, R. G., Degnen, G. E. & Cox, E. C. (1974). Mutational specificity of a conditional Escherichia coli mutator, mutD5. *Mol Gen Genet* **133**, 179-191.

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

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

Gann, A. A., Campbell, A. J., Collins, J. F., Coulson, A. F. & Murray, N. E. (1987). Reassortment of DNA recognition domains and the evolution of new specificities. *Mol Microbiol* **1**, 13-22.

Gorbalenya, A. E. & Koonin, E. V. (1991). Endonuclease (R) subunits of type-I and type-III restriction-modification enzymes contain a helicase-like domain. *FEBS Lett* **291**, 277-281.

Gough, J. A. & Murray, N. E. (1983). Sequence diversity among related genes for recognition of specific targets in DNA molecules. *J Mol Biol* **166**, 1-19.

Hansen, E. B., Atlung, T., Hansen, F. G., Skovgaard, O. & von Meyenburg, K. (1984). Fine structure genetic map and complementation analysis of mutations in the dnaA gene of Escherichia coli. *Mol Gen Genet* **196**, 387-396.

Herzer, P. J., Inouye, S., Inouye, M. & Whittam, T. S. (1990). Phylogenetic distribution of branched RNA-linked multicopy single-stranded DNA among natural isolates of Escherichia coli. *J Bacteriol* **172**, 6175-6181.

Kasarjian, J. K., Hidaka, M., Horiuchi, T., Iida, M. & Ryu, J. (2004). The recognition and modification sites for the bacterial type I restriction systems KpnAI, StySEAI, StySENI and StySGI. *Nucleic Acids Res* **32**, e82.

Kohara, Y., Akiyama, K. & Isono, K. (1987). The physical map of the whole E. coli chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. *Cell* **50**, 495-508.

Kulik, E. M. & Bickle, T. A. (1996). Regulation of the activity of the type IC EcoR124I restriction enzyme. *J Mol Biol* **264**, 891-906.

Lee, N. S., Rutebuka, O., Arakawa, T., Bickle, T. A. & Ryu, J. (1997). KpnAI, a new type I restriction-modification system in Klebsiella pneumoniae. *J Mol Biol* **271**, 342-348.

Loenen, W. A., Daniel, A. S., Braymer, H. D. & Murray, N. E. (1987). Organization and sequence of the hsd genes of Escherichia coli K-12. *J Mol Biol* **198**, 159-170.

Makovets, S., Titheradge, A. J. & Murray, N. E. (1998). ClpX and ClpP are essential for the efficient acquisition of genes specifying type IA and IB restriction systems. *Mol Microbiol* **28**, 25-35.

Makovets, S., Doronina, V. A. & Murray, N. E. (1999). Regulation of endonuclease activity by proteolysis prevents breakage of unmodified bacterial chromosomes by type I restriction enzymes. *Proc Natl Acad Sci U S A* **96**, 9757-9762.

Makovets, S., Powell, L. M., Titheradge, A. J., Blakely, G. W. & Murray, N. E. (2004). Is modification sufficient to protect a bacterial chromosome from a resident restriction endonuclease? *Mol Microbiol* **51**, 135-147.

Maurizi, M. R., Clark, W. P., Katayama, Y., Rudikoff, S., Pumphrey, J., Bowers, B. & Gottesman, S. (1990). Sequence and structure of Clp P, the proteolytic component of the ATP-dependent Clp protease of Escherichia coli. *J Biol Chem* **265**, 12536-12545.

Meisel, A., Bickle, T. A., Kruger, D. H. & Schroeder, C. (1992). Type III restriction enzymes need two inversely oriented recognition sites for DNA cleavage. *Nature* **355**, 467-469.

Meselson, M. & Yuan, R. (1968). DNA restriction enzyme from E. coli. *Nature* **217**, 1110-1114.

Milkman, R. & Bridges, M. M. (1990). Molecular evolution of the Escherichia coli chromosome. III. Clonal frames. *Genetics* **126**, 505-517.

Milkman, R. & Bridges, M. M. (1993). Molecular evolution of the Escherichia coli chromosome. IV. Sequence comparisons. *Genetics* **133**, 455-468.

Murray, N. E. (1983). Phage lambda and molecular cloning. In *Lambda II*, pp. 395 - 432. Edited by R. Hendrix R.W., J.W., Stahl, L.W. & Weisberg, R.A. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

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

Murray, N. E., Daniel, A. S., Cowan, G. M. & Sharp, P. M. (1993). Conservation of motifs within the unusually variable polypeptide sequences of type I restriction and modification enzymes. *Mol Microbiol* **9**, 133-143.

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

Ochman, H. & Selander, R. K. (1984). Standard reference strains of Escherichia coli from natural populations. *J Bacteriol* **157**, 690-693.

Ochman, H., Whittam, T. S., Caugant, D. A. & Selander, R. K. (1983). Enzyme polymorphism and genetic population structure in Escherichia coli and Shigella. *J Gen Microbiol* **129 (Pt 9)**, 2715-2726.

Phadnis, S. H., Huang, H. V. & Berg, D. E. (1989). Tn5supF, a 264-base-pair transposon derived from Tn5 for insertion mutagenesis and sequencing DNAs cloned in phage lambda. *Proc Natl Acad Sci U S A* **86**, 5908-5912.

Piekarowicz, A., Klyz, A., Kwiatek, A. & Stein, D. C. (2001). Analysis of type I restriction modification systems in the Neisseriaceae: genetic organization and properties of the gene products. *Mol Microbiol* **41**, 1199-1210.

Prakash-Cheng, A. & Ryu, J. (1993). Delayed expression of in vivo restriction activity following conjugal transfer of Escherichia coli hsdK (restriction-modification) genes. *J Bacteriol* **175**, 4905-4906.

Prakash-Cheng, A., Chung, S. S. & Ryu, J. (1993). The expression and regulation of hsdK genes after conjugative transfer. *Mol Gen Genet* **241**, 491-496.

Price, C., Lingner, J., Bickle, T. A., Firman, K. & Glover, S. W. (1989). Basis for changes in DNA recognition by the EcoR124 and EcoR124/3 type I DNA restriction and modification enzymes. *J Mol Biol* **205**, 115-125.

Price, C., Pripfl, T. & Bickle, T. A. (1987). EcoR124 and EcoR124/3: the first members of a new family of type I restriction and modification systems. *Eur J Biochem* **167**, 111-115.

Raleigh, E. A. (1992). Organization and function of the mcrBC genes of Escherichia coli K-12. *Mol Microbiol* **6**, 1079-1086.

Redaschi, N. & Bickle, T. A. (1996). DNA restriction and modification systems. In *Escherichia coli and Salmonella*, pp. 773 - 781. Edited by F. C. Neidhart, R. Curtiss, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schnaechter & H. E. Umbarger. Washington, DC: American Society for Microbiology.

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

Ryu, J., Rajadas, P. T. & Bullas, L. R. (1988). Complementation and hybridization evidence for additional families of type I DNA restriction and modification genes in Salmonella serotypes. *J Bacteriol* **170**, 5785-5788.

Sain, B. & Murray, N. E. (1980). The hsd (host specificity) genes of E. coli K 12. *Mol Gen Genet* **180**, 35-46.

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

Sharp, P. M., Kelleher, J. E., Daniel, A. S., Cowan, G. M. & Murray, N. E. (1992). Roles of selection and recombination in the evolution of type I restriction-modification systems in enterobacteria. *Proc Natl Acad Sci U S A* **89**, 9836-9840.

Sibley, M. H. & Raleigh, E. A. (2004). Cassette-like variation of restriction enzyme genes in Escherichia coli C and relatives. *Nucleic Acids Res* **32**, 522-534.

Singleton, M. R. & Wigley, D. B. (2002). Modularity and specialization in superfamily 1 and 2 helicases. *J Bacteriol* **184**, 1819-1826.

Sturrock, S. S. & Dryden, D. T. (1997). A prediction of the amino acids and structures involved in DNA recognition by type I DNA restriction and modification enzymes. *Nucleic Acids Res* **25**, 3408-3414.

Titheradge, A. J., Ternent, D. & Murray, N. E. (1996). A third family of allelic hsd genes in Salmonella enterica: sequence comparisons with related proteins identify conserved regions implicated in restriction of DNA. *Mol Microbiol* **22**, 437-447.

Titheradge, A. J., King, J., Ryu, J. & Murray, N. E. (2001). Families of restriction enzymes: an analysis prompted by molecular and genetic data for type ID restriction and modification systems. *Nucleic Acids Res* **29**, 4195-4205.

Webb, J. L., King, G., Ternent, D., Titheradge, A. J. & Murray, N. E. (1996). Restriction by EcoKI is enhanced by co-operative interactions between target sequences and is dependent on DEAD box motifs. *Embo J* **15**, 2003-2009.

Whittaker, P. A., Campbell, A. J., Southern, E. M. & Murray, N. E. (1988). Enhanced recovery and restriction mapping of DNA fragments cloned in a new lambda vector. *Nucleic Acids Res* **16**, 6725-6736.

The Diversity of Alleles at the *hsd* Locus in Natural Populations of *Escherichia coli*

Victoria A. Barcus,[1] Annette J. B. Titheradge and Noreen E. Murray

Institute of Cell and Molecular Biology, University of Edinburgh, Edinburgh EH9 3JR, United Kingdom

Manuscript received January 2, 1995
Accepted for publication May 17, 1995

ABSTRACT

In enteric bacteria three discrete families of type I restriction and modification systems (IA, IB and ID) are encoded by alleles of the *serB*-linked *hsd* locus. Probes specific for each of the three familes were used to monitor the distribution of related systems in 37 of the 72 wild-type *Escherichia coli* strains comprising the ECOR collection. All 25 members of group A in this collection were screened; 12 were probe-positive, nine have *hsd* genes in the IA family, two in the IB and one in the ID. Twelve strains, representing all groups other than A, were screened; five were probe-positive, one has *hsd* genes in the IA family, one in the IB and three in the ID. The type ID genes are the first representatives of this family in *E. coli*, the probe-negative strains could have alternative families of *hsd* genes. The type IA and IB systems added at least five new specificites to the five already identified in natural isolates of *E. coli*. The distribution of alleles is inconsistent with the dendrogram of the bacterial strains derived from other criteria. This discrepancy and the dissimilar coding sequences of allelic *hsd* genes both imply lateral transfer of *hsd* genes.

EXCEPTIONALLY high intraspecific allelic diversity has been described for a number of loci in both eukaryotes and prokaryotes. This extreme genetic variability often correlates with a need to differentiate "foreign" from "self". In eukaryotes, examples of such systems are the MHC class II alleles in mammals (FIGUEROA *et al.* 1988; LAWLOR *et al.* 1988), mating-type loci in fungi (KÜES and CASSLETON 1992), and the self-incompatibility loci of certain plants (IOERGER *et al.* 1990). For these systems, selection for variation has resulted in the maintenance of a large number of alleles and high intraspecific sequence divergence consistent with gene lineages that predate speciation.

In bacteria, a high degree of variation among the genes encoding a variety of surface antigens is seen as a means of improving the bacterium's chances of escaping the host's immune system. For example, the somatic O lipopolysaccharide is a polymorphic surface antigen encoded by the *rfb* gene cluster; ~60 forms of the O antigen have been identified in Salmonella, and >160 in *Escherichia coli* (see REEVES 1993). In different antigenic groups of *Salmonella enterica*, *rfb* genes of limited sequence similarity are flanked by well-conserved DNA sequences. Recombination may replace one set of *rfb* genes with *rfb* alleles of dissimilar sequence (WANG *et al.* 1992). Other highly polymorphic systems include the flagellin genes of *S. enterica* (SMITH *et al.* 1990) and the genes concerned with capsular serotypes in *E. coli* (DRAKE *et al.* 1993).

It is often argued that bacteria need to defend themselves against invasion by foreign DNA. Restriction and modification (R-M) systems enable bacteria to distinguish "foreign" DNA from their own (for a recent review, see BICKLE and KRÜGER 1993). The modification component of the system monitors the methylation state of the cell's own DNA and methylates specific bases within a recognition sequence, ensuring that newly replicated, hemimethylated, DNA will be fully modified. DNA with unmodified target sequences will be recognized as foreign and cleaved by the restriction component of the system. Considerable evidence already indicates allelic diversity for the genetic determination of type I R-M systems of enteric bacteria (reviewed in BARCUS and MURRAY 1995).

The type I R-M enzymes each comprise three subunits, encoded by the *hsdR*, *M* and *S* genes. The S and M subunits together form a DNA methyltransferase that methylates adenine residues, one on each strand within an asymmetric, bipartite recognition sequence. The methyltransferase component of some type I R-M systems has a preference for hemimethylated DNA. When all three subunits are present, the alternative activities of restriction and modification are dictated by the methylation state of the target sequence; hemimethylated targets are modified, unmethylated targets elicit restriction.

In *E. coli* K-12 the chromosomal genes encoding the type I system *Eco*KI are flanked on one side by *mrr* and on the other by *mcrBC* (see RALEIGH 1992). The *mrr* and *mcr* genes encode two additional restriction systems, but

Corresponding author: Noreen E. Murray, Institute of Cell and Molecular Biology, Darwin Building, The King's Buildings, Mayfield Road, Edinburgh EH9 3JR, UK.

[1] *Present address:* School of Biological Sciences, University of Sussex, Falmer, Brighton, BN1 9QG, UK.

TABLE 1

Target sequences of type I R-M systems of enteric bacteria

Family	Enzyme	Target	Reference
IA	*Eco*KI	ACC (N_6) GTGC	KAN *et al.* (1979)
	*Eco*BI	TGA (N_8) TGCT	LAUTENBERGER *et al.* (1978) RAVETCH *et al.* (1978)
	*Eco*DI	TTA (N_7) GTCY	NAGARAJI *et al.* (1985a)
	*Sty*LTII[a]	GAG (N_6) RTAYG	NAGARAJA *et al.* (1985b)
	*Sty*SPI	AAC (N_6) GTRC	NAGARAJA *et al.* (1985b)
IB	*Eco*AI	GAG (N_7) GTCA	SURI *et al.* (1984)
	*Eco*EI	GAG (N_7) ATGC	COWAN *et al.* (1989)
	*Cfr*AI	GCA (N_8) GTGG	KANNAN *et al.* (1989)
IC	*Eco*R124I	GAA (N_6) RTCG	PRICE *et al.* (1989)
	*Eco*DXXI	TCA (N_7) RTTC	GUBLER *et al.* (1992)
	*Eco*prrI	CCA (N_7) RTGC	TYNDALL *et al.* (1994)

[a] *Sty*LTIII previously known as *Sty*SB.

ones that lack a corresponding modification component and are activated, not blocked, by the presence of methylated bases. The short (~15kb) region of the bacterial chromosome that includes *mrr*, *hsdRMS* and *mcrBC* has been referred to as the immigration control region (ICR), and is hypervariable in *E. coli* (see RALEIGH 1992). Alternative *hsd* genes within the ICR confer diagnostic specificities to the laboratory strains *E. coli* K-12, B and 15T$^-$ (BOYER 1964; ARBER and WAUTERS-WILLEMS 1970). Complementation tests indicated that the *hsd* genes of *E. coli* strains B and K-12 are sufficiently similar that their polypeptide products are exchangeable and that the subunit encoded by *hsdS* confers sequence specificity to the R-M complex. *Eco*KI and *Eco*BI became founder members of a family of closely related type I R-M systems, subsequently called type IA.

A first hint that allelic genes might also encode sufficiently dissimilar type I R-M systems to warrant their separation into a different family came from hybridization screens of bacterial DNAs and serological screens of cell extracts. It was found, as expected, that the nucleotide sequences of the *hsd* genes for *Eco*KI and *Eco*BI hybridized to each other, and antibodies raised against *Eco*KI reacted with *Eco*BI. In contrast, DNA probes made from the *Eco*KI genes failed to hybridize with those of *E. coli* 15T$^-$ which encoded *Eco*AI; similarly antibodies against *Eco*KI did not cross-react with *Eco*AI (MURRAY *et al.* 1982). The *hsd* genes in these two strains are of very different nucleotide sequence, but they behave as alleles in genetic tests dependent on recombination.

Eleven naturally occurring type I R-M systems, each recognizing a unique DNA sequence have been allocated to three discrete families (see Table 1). The members of the IA and IB, but not IC, families are encoded by allelic genes. Two of the three IC representatives are specified by plasmid-borne genes. Interfamily comparisons of the predicted amino acid sequences of the polypeptides of representatives of the three families suggest that the genes are homologous (SHARP *et al.* 1992; MURRAY et al. 1993).

Unpublished evidence (A. J. B. TITHERADGE) shows that the *serB*-linked *hsd* genes of *S. enterica* serovar bledgdam (BULLAS *et al.* 1980) encode the first member of a fourth family of type I R-M genes (ID); these *hsd* genes do not cross-hybridize with those encoding the other three families of enzymes.

In this paper, we use probes specific for the three families of allelic *hsd* genes to monitor the distribution of related type I R-M systems in *E. coli*. Representatives of the ID family were identified in *E. coli*, where previously this family was only known to occur in Salmonella. The distribution of *hsd* genes among different populations of *E. coli* is not consistent with the relatedness of the bacterial strains based on other criteria, including multilocus enzyme electrophoresis (MLEE). This is so for the different families of enzymes and for alternative specificities within one family.

MATERIALS AND METHODS

Bacteria, bacteriophages and plasmids: Members of the ECOR collection were generally obtained from two sources—the American Type Culture Collection (ATCC) and either Dr. T. S. WHITTAM or Dr. H. OCHMAN. Discrepancies were encountered for only two strains, ECOR18 (ATCC No. 35337) and ECOR23 (ATCC No. 35342). The lyophile ATCC No. 35337 was a mixture of two strains; one included *hsd* genes of the type IB family, the other failed to hybridize with all *hsd* probes. A negative result was obtained with the strain from T.S.W. The lyophile ATCC No. 35342 was negative with the *hsd* probes, while that from T.S.W. included *hsd* genes of the type IA family. The strain from the ATCC lyophile did not have the enzyme profile of ECOR23 (T. S. WHITTAM, personal communication).

Other bacterial strains used were the $r_K^+m_K^+$ strain C600 (APPLEYARD 1954), an *hsd*Δ strain NM679 (KING and MURRAY 1995), a *recA hsd*Δ strain DL795 (D. R. LEACH, unpublished results), ED8654 an $r_K^-m_K^+$ host (BORCK *et al.* 1976), and K803, which is *hsdS*$^-$ and therefore $r_K^-m_K^-$ (WOOD 1966).

λ*vir* was generally used as a test for restriction and modification. Phages grown on strains lacking a modification system are denoted by the symbol .0, *e.g.*, λ*vir*.0. Modifications imposed on phages grown on R-M-proficient strains are identified by the name of the R-M system, *e.g.*, λ*vir*.*Eco*KI and

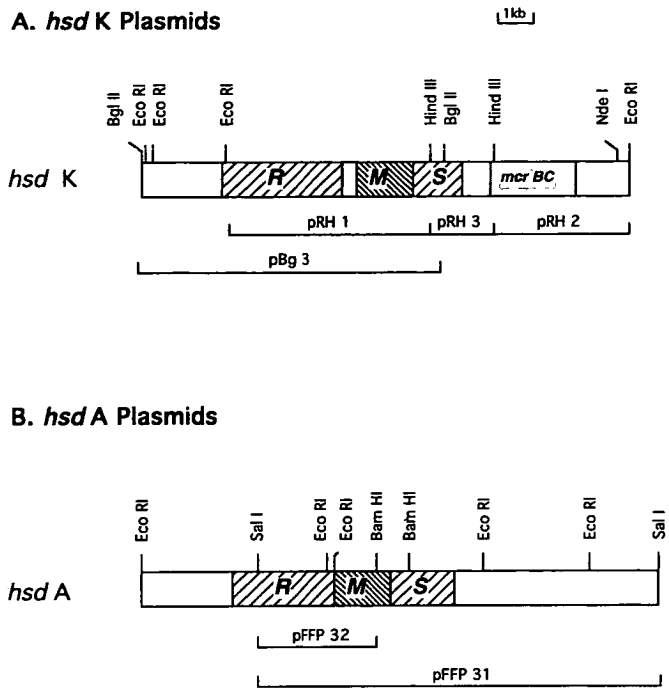


FIGURE 1.—Plasmids containing *hsd* and *mcr* genes. All are derivatives of pBR322. The type IA-specific probe, pRH1, has the *hsdM* gene, as well as parts of *hsdR* and *hsdS*, from *E. coli* K-12. The type IB-specific probe, pFFP32, contains a similar region from *E. coli* 15T⁻. pRH3 was used as a probe for type IA *hsdS* genes and pRH2 as a probe for *mcrBC*. The 3-kb *Eco*RI fragment downstream of *hsd*S, in pFFP31, and the *Hind*III-*Nde*I fragment spanning *mcrBC* in pRH2 were purified and used for probing cosmids. pBg3 was the source of HsdR and M polypeptides in complementation tests.

λ*vir*.*Eco*R5I, where *Eco*KI is the type I R-M system of *E. coli* K-12 and *Eco*R5I denotes the type I R-M system identified in the ECOR5 strain. Phages P3 (BULLAS *et al.* 1980) and the P1 *darA* strain P1CmTc1 (IIDA *et al.* 1987) were also used as substrates to test for restriction.

The plasmids used, derivatives of pBR322, have been described (SAIN and MURRAY 1980; FULLER-PACE *et al.* 1985) and are shown in Figure 1.

Media and microbial techniques: Media and general methods (MURRAY *et al.* 1977) and tests for estimating restriction and modification (FULLER-PACE *et al.* 1985) have been described. The complementation tests for a functional *hsdS* gene were done using the multicopy plasmid pBg3 (see Figure 1) as the source of HsdM and HsdR (KELLEHER *et al.* 1991).

Enzymes and chemicals: Restriction enzymes were purchased from Boehringer Mannheim UK except for *Sau*3A, which was obtained from Northumbria Biologicals Ltd. HK™ phosphatase was purchased from Epicentre Technologies.

Preparation, analysis and ligation of DNA: Bacterial DNA was extracted using the miniprep procedure of REDFIELD and CAMPBELL (1987). Phage λ DNA was extracted according to a miniprep method devised by A. J. B. TITHERADGE (unpublished data). Vector DNA, however, was always isolated from phages that had been purified in a CsCl gradient. Restriction endonuclease digestion was done using the buffers and conditions recommended by the supplier, ligations by standard methods (SAMBROOK *et al.* 1989).

Genomic libraries: Libraries of *Eco*RI fragments were made in the λ replacement vector NM574 (BORCK *et al.* 1976). A library of fragments generated by partial digestion of ECOR9 DNA with *Sau*3A was made in λNM1249, a derivative of EMBL3 in which the *d*857 gene replaces the deletion *d*KH54. The recombinant phages were recovered by *in vitro* packaging using extracts from Amersham International plc. Generally, *hsd* deletion hosts were used for the recovery, propagation and analysis of the recombinant phages.

Libraries of fragments generated by partial digestion of bacterial DNAs with *Sau*3A were made in the cosmid vector Supercos I (Stratagene Ltd) according to the supplier's instructions. Gigapack II Gold *in vitro* packaging extracts (Stratagene Ltd) were used to recover recombinant cosmids by infection of DL795.

Detection of DNA sequences by hybridization: Chromosomal DNA fragments from agarose gels and recombinant DNA molecules from plaques or colonies on agar plates were transferred to nylon filters (Hybond N, Amersham International plc) following standard procedures (see SAMBROOK *et al.* 1989).

The Boehringer DIG labeling and detection system was used as indicated by the suppliers, most recently using the Easy Hyb buffer recommended for the DIG system. Hybridization was carried out at 37° and washes were done at room temperature. For the analysis of bacterial DNA, *Hin*dIII digests of λ DNA labeled with DIG (Boehringer Mannheim UK) were used to provide size markers.

Plasmid probes were linearized or fragments were excised and purified (see Figure 1) before labeling. The ID-specific probe was a purified DNA fragment containing *hsd* genes from *S. enterica* serovar blegdam (A. J. B. TITHERADGE, unpublished results).

RESULTS

The ECOR collection consists of 72 wild-type *E. coli* strains (OCHMAN and SELANDER 1984), chosen on the basis of MLEE to represent the genetic diversity of *E. coli* as a species and, in particular, of the major subspecific groups (OCHMAN *et al.* 1983; WHITTAM *et al.* 1983; SELANDER *et al.* 1986). Classification by MLEE is believed to give a reliable estimate of genetic relatedness of strains and generally agrees with classification based on restriction analyses and DNA sequence comparisons (OCHMAN *et al.* 1983; MILKMAN and BRIDGES 1990, 1993).

A dendrogram showing the relatedness of members of the ECOR collection, based on MLEE of 35 polymorphic enzymes (HERZER *et al.* 1990), provides a framework upon which to superimpose and analyze the distribution of other biological traits in evolutionary terms, in this case the *hsd* genes that encode type I R-M systems (see Figure 2). A total of 37 strains from the ECOR collection was probed for the presence of *hsd* genes from each of three families (IA, B and D). Of the 37 strains, 25 comprise group A, a section of the dendrogram including strains closely related to *E. coli* K-12 (HERZER *et al.* 1990). The others are distributed throughout the collection and include eight commonly used in other surveys of genetic diversity of *E. coli* (MILKMAN and CRAWFORD 1983; DUBOSE *et al.* 1988; DYKHUIZEN and GREEN 1991). The results of all the hybridizations are summarized in Table 2.

DNA from 12 (48%) of the strains in group A of the ECOR collection hybridized to one of the family-

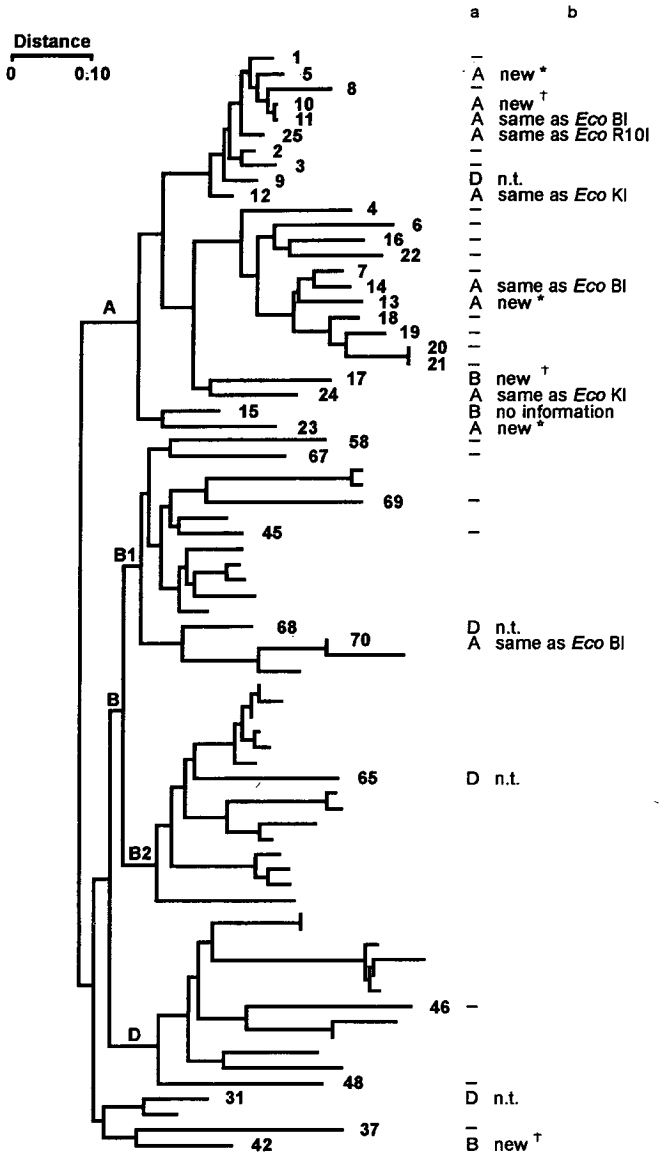


FIGURE 2.—Dendrogram of the ECOR strains. This tree was derived by HERZER *et al.* (1990) on the basis of polymorphisms at 38 enzyme-encoding loci. Only those strains tested with *hsd* probes are identified. Column a shows the results from probing genomic DNA. A, hybridization with the IA-specific probe; B, hybridization with the IB-specific probe; D, hybridization with the ID-specific probe; "—" indicates no hybridization detected with any of the three *hsd*-specific probes. Column b, summarizes the information on the specificity of the R-M systems identified. The sources of the ECOR strains are given in MATERIALS AND METHODS. *, see Table 3 for additional information; †, see Table 4 for additional information; n.t., not tested—the type ID members are the first representatives of this family in *E. coli*.

specific probes (see Table 2); nine to the IA-specific probe (pRH1, see Figure 1), two to the IB probe (pFFP32, see Figure 1), and one to the ID probe. No strain hybridized to more than one probe. Outside of group A, five of 12 strains tested (42%) had DNA that hybridized to one of the probes; one to IA, one to IB, and three to ID.

The ECOR strains used in this survey were also screened for the presence of the genes encoding the McrBC restriction system (RALEIGH 1992). The *mcrBC* genes have been found previously only in strains encoding type IA R-M systems (DANIEL *et al.* 1988), as if the IA family *hsd* genes and the *mcrBC* genes have been acquired concomitantly. However, *mcrBC* genes were not detected in all strains with type IA *hsd* genes; both E166 (encoding *Eco*DI) and *S. enterica* serovar typhimurium (*Sty*LTIII) lacked these genes. The results of the current survey reinforce the conclusions of DANIEL *et al.* (1988). Only strains that hybridized to the IA-specific probe hybridized to the *mcrBC* probe, and ECOR5, though positive with the type IA-specific probe, was negative with the *mcrBC*-specific probe (Table 2).

Cloning and analysis of type IA *hsd* genes in λ vectors: In *E. coli* K-12, the IA-specific probe (pRH1, see Figure 1) hybridizes to a single *Eco*RI fragment encompassing all of *hsdM* and *S* and most of *hsdR* (SAIN and MURRAY 1980). Additionally, the *mcrBC* genes are located in this fragment (ROSS *et al.* 1987; RALEIGH *et al.* 1989). Several of the ECOR strains (ECOR5, 12, 13, 23 and 24) resembled *E. coli* K-12 in that a single, large, *Eco*RI fragment hybridized to the *hsd*-specific probe. For each strain, this *Eco*RI fragment was cloned in an *att*$^-$ *ral*$^-$ *c*I857 replacement vector, λNM574. The origin of the *hsd* region in the various recombinant phages is identified by the ECOR strain number (*e.g.*, λ574–12 carries the *hsd* region of ECOR12). Each cloned fragment also hybridized to the *hsdS* probe (pRH3, see Figure 1) and, with the exception only of that in λ574–5, to the *mcrBC* probe (pRH2, see Figure 1). The latter result is anticipated since the ECOR5 strain itself was negative with this probe. It is likely that all five phages include *hsdM* and *S*, though only a functional *hsdS* gene is essential to check specificity, since the M and R polypeptides can be provided in *trans*.

The specificity polypeptide encoded by each λ derivative was tested for its ability to form a restriction-proficient complex in the presence of the products of the *hsdR* (R) and *hsdM* (M) genes. Previous experiments (KELLEHER *et al.* 1991) have shown that a λ*hsd* phage encoding a functional S polypeptide has a low efficiency of plating (e.o.p.) on a nonmodifying strain carrying a multicopy plasmid (pBg3, see Figure 1) encoding the R and M polypeptides. A low e.o.p. is presumed to depend on the production of a functional endonuclease within a cell that lacks appropriately modified DNA. On the basis of this complementation test, all five phages encode a functional S polypeptide (see Table 3).

The differentiation of new specificities from those already characterized is permitted by an alternative test that also depends on the interchange of subunits between enzymes within the same family. In this test restriction-proficient lysogens are made by replacing the defective *S* gene of an *hsdR*$^+$*M*$^+$*S*$^-$ bacterium (K803) with the functional *S* gene of the λ*hsdS* phage. This is

TABLE 2

Sizes of *EcoRI* fragments identified by hybridization

ECOR No.	Group[a]	Probe			
		ID specific	IB specific	IA specific	*mcr*BC
5	A	—	—	12.5 kb	—
9	A	6 kb	—	—	—
10	A	—	—	5.5 kb	7 kb
11	A	—	—	6 kb	8 kb
12	A	—	—	13 kb	13 kb
13	A	—	—	11 kb	11 kb
14	A	—	—	5.5 kb	11.5 kb
15	A	—	3.5 kb	—	—
17	A	—	6 kb	—	—
23	A	—	—	11 kb	11 kb
24	A	—	—	12.5 kb	12.5 kb
25	A	—	—	6 kb	11 kb
31	*[b]	4 kb	—	—	—
42	*[b]	—	5.5, 4.5 kb	—	—
65	B2	4 kb	—	—	—
68	B1	4 kb	—	—	—
70	B1	—	—	5.5 kb	8 kb

The sizes given are approximate. DNA from the following ECOR strains did not hybridize to any of the four probes: 1, 2, 3, 4, 6, 7, 8, 16, 18, 19, 20, 21, 22, 37, 45, 46, 48, 58, 67, 69. ECOR 18, obtained from the American Type Culture Collection, was found to be a mixture of two strains. Only the contaminant strain hybridized to the IB-specific probe.

[a] This is the group assigned on the dendrogram of HERZER et al. (1990).

[b] ECOR31 and 42 are placed in the same group, but this was unassigned by HERZER *et al.* (1990).

readily achieved by the homology-dependent integration of the phage (GOUGH and MURRAY 1983).

The results of restriction tests are summarized in Table 3. The λ574–5, λ574–12, and λ574–24 lysogens [K803(5), K803(12) and K803(24), respectively], restrict λ*vir*.0 with e.o.ps ranging from 1×10^{-3} to 2×10^{-4}. Neither the λ574–13 nor the λ574–23 lysogen restrict λ*vir*.0. The K803(23) lysogen did, however, restrict P1*dar*.0 and P3.0 ~100-fold, indicating that the ECOR23 *hsdS* gene imparts a specificity that has no target sequences in λ. As all known type I R-M systems recognize targets in λ, the ECOR23 system has a novel specificity. On the basis of the complementation test, ECOR13 includes a functional *S* gene, but none of our tester phages was restricted by K803(13). The evidence for a new specificity in ECOR13, therefore, is circumstantial.

λ*vir* modified by propagation on *E. coli* strains encoding type I R-M systems of different specificities were used to determine whether the ECOR5, 12 and 24 systems had the same specificity as any of those previously identified. The e.o.p. of various appropriately modified

TABLE 3

Analysis of λ*hsd* phages

Phage[a]	Complementation of $hsdR^+M^+$ plasmid	E.o.p. of phages on lysogen[b]				Specificity
		λ*vir*.0	λ*vir*.K	P3.0	P1*dar*.0	
λ574-5	+	2×10^{-4}	2×10^{-4}	nt	nt	New[c]
λ574-12	+	7×10^{-4}	1	nt	nt	*Eco*KI[d]
λ574-13	+	1	nt	1	1	New[e]
λ574-23	+	1	nt	4×10^{-2}	1×10^{-2}	New[f]
λ574-24	+	10^{-3}	1	nt	nt	*Eco*KI[d]

nt, not tested.

[a] The number after the dash identifies the ECOR strain from which the donor fragment came.

[b] Phage (λ 574-5, -12, etc.) integrated into the $hsdS^-$ derivative of *E. coli* K-12, K803, to give K803(5), *etc.*

[c] λ*vir* modified by any known chromosomally-encoded type I R-M system is restricted by K803(5).

[d] See text for distinction between the specificities of *Eco*KI and *Sty*SPI.

[e] Presumed to be a new specificity since it appears to restrict *E. coli* but not λ*vir*. All previously characterized type IA R-M systems have targets in λ DNA.

[f] Differs from K803(13) in that it restricts P3 and P1*dar*.

λ phages was checked on the restriction-proficient lysogens. Both λ*vir.Eco*KI and λ*vir.Sty*SPI plated with an e.o.p. of 1 on lysogens specifying either the ECOR12 or ECOR24 restriction system (Table 3). The *Sty*SPI R-M system recognizes a degenerate form of the *Eco*KI recognition sequence (NAGARAJA *et al.* 1985b), so that phages modified by *Eco*KI will still be restricted by *Sty*SPI. λ phages recovered after propagation on K803(12) and K803(24) were checked for modification against both *Eco*KI and *Sty*SPI restriction. The modification was only active against *Eco*KI restriction (Table 3), confirming that both ECOR12 and ECOR24 encode type I R-M systems with K-specificity.

K803(5) lysogens restricted λ*vir*, modified by propagation on strains encoding methyltransferases with specificities of all known type IA and IB R-M systems, ~10^3-fold (data not shown). This implies that ECOR5, like ECOR23, encodes a type I R-M system with a specificity previously not identified in the IA family.

Cloning and analysis of type IA and IB *hsd* genes in cosmid vectors: λ libraries have provided a reliable route for cloning *hsd* genes (BORCK *et al.* 1976; FULLER-PACE *et al.* 1985). The DNA of λ phages that include *hsdM* and *S* is modified with the specificity dictated by the cloned *hsdS* gene. However, the effective detection and analysis of different specificities need restriction-proficient bacteria and for new systems this will rely on the integration of a λ *hsdMS* phage into an *hsdS*$^-$ (R^+M^+) strain (see Table 3). The requirement for making lysogens can be obviated by a bacterium harboring a stably maintained, *hsd*$^+$ plasmid or cosmid, because this provides a direct test for both restriction and modification. The *hsd* region encoding *Eco*AI was transferred from a λ phage to a plasmid and the resulting transformants were restriction proficient (FULLER-PACE *et al.* 1985); no barrier to the acquisition of these type IB *hsd* genes was detected. For these reasons, the remaining type IA and IB *hsd* genes were cloned by making libraries of cosmids containing fragments generated from partial *Sau*3A digests of bacterial DNAs. Cosmids rather than plasmids were chosen to provide efficient recovery of large inserts.

The libraries were recovered in an *hsd*Δ*recA* host (DL795) and colonies harboring *hsd* genes were detected by their hybridization to probes specific for type IA or type IB systems. Positives from each library were tested for their ability to restrict λ*vir*.0. Restriction-proficient clones were isolated for five of the eight ECOR strains. These clones were also tested for their ability to restrict a collection of modified lysates of λ*vir* chosen to document the known repertoire of type IA and IB modification specificities. The five strains included two with previously unidentified specificities, one member of type IA and one of type IB (see Table 4).

No restriction-proficient clones were found for three ECOR strains (see Table 4), despite the expectation that the majority of cosmids selected would include the entire *hsd* region. It is probable, though not proven, that some recombinant cosmids are not maintained sufficiently well for the host to be scored as restriction proficient. The library from ECOR70 gave the most favourable results: 10/14 clones were modification proficient and six of these were restriction proficient. However, even for these recombinants, if a culture was to be scored as restriction proficient, it was essential to grow the bacteria in medium supplemented with kanamycin (50 μg ml^{-1}), rather than ampicillin. Kanamycin enforces better maintenance of the cosmid than ampicillin. This observation stresses the importance of maintaining the cosmid because cultures in which only half of the cells have maintained the *hsd*$^+$ cosmid would score as restriction deficient. In the absence of a restriction-proficient strain there is no biological test for modification.

Another explanation for the absence of restriction-proficient clones is that the ECOR strains themselves have mutations that make their *hsd* genes defective. Alternatively there could be selection against the recovery of genes encoding a restriction enzyme in the absence of prior establishment of modification. This has been observed for some R-M genes encoding type II and type III systems (DEBACKER and COLSON 1991).

The chromosomal location of three families of *hsd* genes in *E. coli*: The archetypal members of the IA, IB and ID families of R-M systems are encoded by *serB*-linked *hsd* genes (BOYER 1964; ARBER and WAUTERS-WILLEMS 1970; BULLAS *et al.* 1980). It is likely that the *hsd* genes detected by the three family-specific probes (Table 2) share a common chromosomal location. Chromosomal segments of DNA known to be adjacent to *hsdS* in their respective *E. coli* strains, K-12 and 15T$^-$, were used to probe the clones derived from the ECOR strains. The probe from *E. coli* K-12 spans *mcrB* and *C* (see Figure 1A and legend), that from 15T$^-$ is the 3-kb *Eco*RI fragment adjacent to *hsdS* (see Figure 1B and legend). The clones tested included those for the 13 type IA and IB *hsd* genes and one representative of the ID family. The data are summarized in Table 5.

Each clone derived from a chromosomal DNA shown to cross-hybridize with pRH2 was shown to include DNA that hybridized with the probe derived from pRH2, which includes most of *mcrB* and *C*. This result is consistent with the conservation of close linkage between the type IA *hsd* genes and the neighbouring *mcrB* and *C* genes of the immigration control region (RALEIGH 1992). All of the five clones derived from the DNA of ECOR strains that failed to cross-hybridize with pRH2 were found to hybridize with the flanking probe derived from *E. coli* 15T$^-$. Three of the cosmids that include type IA *hsd* genes were tested with both probes and were positive with both as expected if the chromosomal location of the immigration control region is conserved in enteric bacteria.

TABLE 4

Analysis of cosmid clones

Donor DNA	Probe	No. of restriction proficient clones[a]	Modification that protects against restriction	Specificity
ECOR10	IA	2 (10)	*Eco*R25I	New
ECOR11	IA	0 (13)[b]	nt[c]	Same as *Eco*BI
ECOR14	IA	1 (9)	*Eco*BI	Same as *Eco*BI
ECOR15	IB	0 (18)	nt[c]	No information
ECOR17	IB	0 (18)	nt[c]	[d]
ECOR25	IA	4 (7)	*Eco*R10I	Same as *Eco*R10I
ECOR42	IB	2 (6)	None	New
ECOR70	IA	6 (14)	*Eco*BI	Same as *Eco*BI

[a] Values in parentheses are number of clones tested.
[b] Two clones modify with the same specificity as *Eco*BI.
[c] In the absence of restriction, these tests cannot be done.
[d] Specificity known to be new (P. H. THORPE, personal communication)

DISCUSSION

Occurrence: The utility of restriction enzymes in molecular biology has motivated extensive searches for type II R-M enzymes with different specificities. JANULAITIS and coworkers (1988; A. JANULAITIS, personal communication) have screened the Enterobacteriaceae for sequence-specific endonucleases by incubating cell extracts with suitable DNA substrates, then looking for discrete fragments following separation by electrophoresis through agarose gels. Sequence-specific nucleases were detected in nearly 25% of 1000 strains tested. Many type II and type III R-M systems, however, are plasmid-encoded and, given the transmissibility of plasmids, chromosomally located genes are the preferred choice for analysis in the context of the dendrogram of ECOR strains.

TABLE 5

Cocloning of *hsd* genes and the adjacent chromosomal DNA

ECOR strain	Vector	Family of *hsd* genes	Hybridization to chromosomal sequence	
			From K-12[a]	From 15T^{-}[b]
5	λ	A	−[c]	+
9	λ	D	−[c]	+
10	cosmid	A	+	+
11	cosmid	A	+	nt
12	λ	A	+	nt
13	λ	A	+	nt
14	cosmid	A	+	nt
15	cosmid	B	−[c]	+
17	cosmid	B	−[c]	+
23	λ	A	+	nt
24	λ	A	+	nt
25	cosmid	A	+	+
42	cosmid	B	−[c]	+
70	cosmid	A	+	+

nt, not tested.
[a] See Figure 1A and legend.
[b] See Figure 1B and legend.
[c] Bacterial DNA used to make clones does not hybridize with this probe.

The majority of known type I R-M systems are encoded by chromosomal genes, but these enzymes do not generate discrete fragments of DNA and their detection has often relied on the restriction of bacteriophages *in vivo* (BERTANI and WEIGLE 1953; BULLAS *et al.* 1980). This biological screen for R-M systems in natural isolates of bacteria is limited to strains that are sensitive to phage infection. Unfortunately, many wild-type *E. coli* strains are resistant to some aspect of productive infection by common laboratory phages, including even P1 (BARCUS 1993). Genes coding for members of known families of type I R-M systems, however, can be detected by hybridization to DNA probes derived from representative *hsd* genes (DANIEL *et al.* 1988). We have used this approach with probes that identify three discrete families of type I R-M enzymes, IA, IB and ID. The *hsd* genes for the representatives of each family have been shown to be allelic, although their nucleotide sequences are so dissimilar that no cross-hybridization has been detected between them (DANIEL *et al.* 1988; A. J. B. TITHERADGE, unpublished observations).

Of the 37 ECOR strains tested, 17 have DNA that hybridizes to one of the three family-specific probes. Because our analysis is limited by the available probes, we have no information for the remaining 20 strains, and it is quite possible that they have alternative families of *hsd* genes. Present experiments are exploring this possibility. BULLAS *et al.* (1980) screened a number of Salmonella serotypes for their ability to restrict phages and found evidence for eight R-M systems encoded by genes linked, like those for *Eco*KI, to *serB*. One of these is the first representative of the ID family and others could identify alternative families.

The numbers of families of type I R-M systems in the ECOR strains remains unknown and there is no indication how many, if any, of the strains will be like

E. coli C and lack DNA in the region of the *hsd* locus (DANIEL *et al.* 1988).

Diversity: The specificity (S) polypeptide of type I R-M systems dictates sequence specificity for both the restriction and modification activities of the enzyme. Type I R-M systems, therefore, have a greater potential for rapid diversification of sequence specificity than do type II systems in which the restriction and modification enzymes are encoded by separate genes. Indeed, type I systems with new specificities have arisen more than once by chance in the laboratory (BULLAS *et al.* 1976; HUGHES 1977; MEISTER *et al.* 1993).

The HsdS polypeptides of the three well-investigated families have a central conserved region bounded by two recognition domains, each of which is responsible for recognition of one of the two specific components of the target sequence (FULLER-PACE and MURRAY 1986; GANN *et al.* 1987; NAGARAJA *et al.* 1985c). All new specificities isolated in the laboratory have arisen by major changes in the specificity gene, resulting from recombination (FULLER-PACE *et al.* 1984; PRICE *et al.* 1989; GUBLER *et al.* 1992), transposition (MEISTER *et al.* 1993) or deletion (ABADJIEVA *et al.* 1993), and never, so far, from simple amino acid substitutions. For members of the same family of enzymes, diversity of sequence specificity resides in differences in the two recognition domains of the S polypeptides, including their combinations and spacing. Additional natural diversity correlates with alternative families of *hsd* genes that encode HsdR and M polypeptides that share only 25–35% identity (SHARP *et al.* 1992; MURRAY *et al.* 1993).

Four ECOR strains were shown to have *hsd* genes belonging to the type ID family. These are the first representatives of the type ID family in *E. coli.* Members of this new family are still to be analysed, but members of both the IA and IB families have been subdivided on the basis of their sequence specificity. The present analysis of type IA and IB genes identifies at least five, and possibly as many as seven, new specificities. This brings the total of naturally occurring specificities documented for type I enzymes in enteric bacteria to a minimum of 18, 14 of which have been found in *E. coli.* At present no specificity is common to *E. coli* and Salmonella, although *Sty*SPI is a degenerate version of *Eco*KI.

Type IA-specific *hsd* genes have been detected in 16 strains of *E. coli*, including 10 members of the ECOR collection, and the 16 strains include seven different specificities. Three strains have the same specificity as *Eco*KI, five are like *Eco*BI, while ECOR10 and 25 confer the same new specificity. The present pool of specificities appears to be small compared with the theoretical number for a heptanucleotide target sequence. It may be significant that at present no two members of different families have the same sequence specificity, although *Sty*LTIII, *Eco*AI and *Eco*EI share one component of their bipartite target sequence (COWAN *et al.* 1989). This finding implies that new combinations of recognition domains are generated in natural populations.

Distribution of *hsd* genes on the ECOR dendrogram: Probe-positive strains were identified in all the major subgroups of the ECOR collection except group D, from which only two isolates were tested (Figure 2). Type IA strains were mainly, but not exclusively, in group A (nine of 10 identified), a group that would include *E. coli* K-12 had it been a member of the ECOR collection (HERZER *et al.* 1990). As might be expected, some pairs of strains in group A, which are believed to be closely related on the basis of MLEE (*e.g.*, ECOR10 and 11, and ECOR13 and 14), have DNA that hybridized to the same (type IA-specific) probe. However, in each case the *hsd* genes encode enzymes with different specificities and, furthermore, other pairs of closely related strains apparently encode type I R-M systems belonging to different families. Whereas ECOR24 hybridized to the IA family-specific probe, its closest relative, ECOR17, was positive with the IB family-specific probe. Similarly, ECOR23 has type IA-like sequences, and ECOR15 type IB-like. The sole type ID-like representative in group A, ECOR9, is in a cluster of closely related strains of which half hybridized to the IA-specific probe. While the apparent predominance of the IA family *hsd* genes in group A is consistent with the results from MLEE and DNA sequence analysis (MILKMAN and BRIDGES 1990, 1993), there are many discrepancies. Indeed, ECOR2, which is reported to have an identical electromorph type to *E. coli* K-12 (R. K. SELANDER, cited in BISERCIC *et al.* 1991), lacks DNA sequences with similarity to the IA family probe. The discrepancies between relatedness of type I R-M systems and the classification of strains based on MLEE is readily resolved by inferring the horizontal transfer of *hsd* genes.

Horizontal transfer: SHARP *et al.* (1992) reported that the intraspecific divergence between members of the IA and IB families is so great that it is difficult to explain the presence of both in *E. coli* without invoking horizontal transfer from some distantly related species. This view is supported by our results; if MLEE provides a reliable classification of bacteria, it is unlikely that enzymes from three distinct families could be found in closely related species in the absence of horizontal transfer.

Other studies of the ECOR collection also provide strong evidence for horizontal transfer of DNA. These include surveys for multicopy single-stranded DNA (msDNA) (HERZER *et al.* 1990), and for the P pili genes associated with the capacity to cause urinary tract infections (MARKLUND *et al.* 1992). As was the case with families of *hsd* genes, pairs of closely related strains were found to differ with respect to the relevant characteristic, that is, the ability to produce msDNA or the class of P pili they encode. Here, too, the pattern of distribution of the genes among members of the ECOR collection was attributed to recruitment of genes by hori-

zontal transfer from other species subsequent to the divergence of *E. coli* into the present distinct strains.

The regions of the *hsdS* gene that encode the recognition domains have a lower G+C content than the adjacent conserved regions and lower, in fact, than the G+C content of the *E. coli* genome in general (DILA *et al.* 1990; RALEIGH 1992; P. H. THORPE, personal communication). This base composition more typical of distantly related species such as Proteus and Providencia, is found in *hsdS* genes of the types IA, B and C families, implying that the sequences encoding the recognition domains of type I systems have been acquired from other species (DILA *et al.* 1990; RALEIGH 1992).

Although sequence comparisons have provided evidence for the transfer of DNA segments between *E. coli* strains (DUBOSE *et al.* 1988; STOLTZFUS *et al.* 1988; MILKMAN and BRIDGES 1990, 1993; BISERCIC *et al.* 1991; DYKHUIZEN and GREEN 1991; MARKLUND *et al.* 1992), evidence of interspecific recombination involving *E. coli* is limited. NELSON and colleagues (NELSON *et al.* 1991; NELSON and SELANDER 1992) failed to identify any recombination events between Salmonella and *E. coli* in the *gapA* and *putP* loci. However, there is evidence that the unusually divergent *E. coli gnd* locus includes sequences recruited from other species of bacteria, such as Citrobacter and Klebsiella (NELSON and SELANDER 1994), though this may reflect the linkage of *gnd* to the highly variable *rfb* gene cluster. SHARP *et al.* (1992) found evidence for the horizontal transfer of a short sequence of the *hsdM* gene between *S. enterica* serovar potsdam and *E. coli.* Recombination between the *gnd* genes of *E. coli* and *S. enterica* serovar typhimurium LT2 has also been reported, based on sequences with a high degree of similarity shared between LT2 and certain *E. coli* clones (BARCAK and WOLF 1988; BISERCIC *et al.* 1991).

The role of diversity: S. LEDERBERG (RADDING 1973) and others (ENDLICH and LINN 1985; PRICE and BICKLE 1986) have suggested that restriction may stimulate recombination. This may seem difficult to reconcile with data demonstrating that restriction depresses gene transfer by both conjugation (ARBER and MORSE 1965) and P1 transduction (IIDA *et al.* 1987). The major route of DNA transfer in natural populations is uncertain but, irrespective of the route, restriction may be a source of recombinogenic DNA, particularly small DNA fragments. Some understanding of how the RecBCD enzyme, alias ExoV, may serve to degrade the products of restriction, or alternatively use them as substrates for recombination, is emerging (KUZMINOV *et al.* 1994). Recent evidence implicates restriction in the incorporation of small DNA segments into a recipient *E. coli* genome after P1 transduction (MCKANE and MILKMAN 1995). Type I R-M systems are inhibited by the Dar functions of phage P1 (IIDA *et al.* 1987), but this block may be incomplete and the residual activity would generate additional ends of DNA molecules that could serve as substrates for recombination. The importance of restriction in the context of both recombination and defence needs further documentation. Experimental procedures to monitor any influence of restriction on recombination now exist (MCKANE and MILKMAN 1995).

Our work supports the conclusions of SHARP and his colleagues (SHARP *et al.* 1992; MURRAY *et al.* 1993) that selective pressure for variation has resulted in the occurrence of different specificities among closely related strains, even to the extent of the recruitment of parts or all of *hsd* regions from other bacteria. As SHARP and coworkers pointed out, such variation closely resembles the extremely high intraspecific allelic diversity seen in the class II MHC complex in mammals and the self-incompatibility complex of the Solanaceae family of plants (LAWLOR *et al.* 1988; IOERGER *et al.* 1990), maintaining ancient alleles and selecting new ones. Frequency dependent selection could be the evolutionary force. The only explanation offered has been that of LEVIN (LEVIN 1986, 1988; KORONA and LEVIN 1993), who proposed that such selection could be imposed in the environment by bacteriophages, so that bacteria with rare specificities are favoured. Many naturally occurring coliphages are sensitive to restriction by type I systems (KORONA *et al.* 1993), and bacteria colonizing a new habitat may have an advantage if they possess a R-M system with a rare specificity, although this advantage will be short lived. Is this advantage sufficient to provide the selective pressure for the extraordinary allelic diversity of type I R-M systems?

We thank colleagues for their helpful criticisms of the manuscript, particularly MELISSA MCKANE, ROGER MILKMAN, LYNN POWELL, NOEL SMITH, LISE RALEIGH, JOHN MAYNARD SMITH and PETER THORPE. We are indebted to TOM WHITTAM and HOWARD OCHMAN for providing ECOR strains and ISOBEL BLACK and JOSIE PRIESTLEY for skilled preparation of the manuscript. The research was funded by the Medical Research Council (UK).

LITERATURE CITED

ABADJIEVA, A., J. PATEL, M. WEBB, V. ZINKEVICH and K. FIRMAN, 1993 A deletion mutant of the type IC restriction endonuclease *Eco*R124I expressing a novel DNA specificity. Nucleic Acids Res. **21:** 4435–4443.

APPLEYARD, R. K., 1954 Segregation of lambda lysogenicity during bacterial recombination in *Escherichia coli* K-12. Genetics **39:** 440–452.

ARBER, W., and M. L. MORSE, 1965 Host specificity of DNA produced by *Escherichia coli* VI. Effects on bacterial conjugation. Genetics **51:** 137–148.

ARBER, W., and D. WAUTERS-WILLEMS, 1970 Host specificity of DNA produced by *E. coli.* XII The two restriction and modification systems of strain 15T⁻. Mol. Gen. Genet. **180:** 203–217.

BARCAK, G. J., and R. E. WOLF, JR., 1988 Comparative nucleotide sequence analysis of growth-rate-regulated *gnd* alleles from natural isolates of *Escherichia coli* and from *Salmonella typhimurium* LT-2. J. Bacteriol. **170:** 372–379.

BARCUS, V. A., 1993 The type I restriction systems of *Escherichia coli.* Ph.D Thesis, University of Edinburgh, Scotland.

BARCUS, V. A., and N. E. MURRAY, 1995 Barriers to restriction: recombination, pp. 31–58 in *Population Genetics of Bacteria,* edited by S. BAUMBERG, J. P. W. YOUNG, E. M. H. WELLINGTON *et al.* Society for General Microbiology Symposium 52, Cambridge University Press, Cambridge.

BERTANI, G., and J. J. WEIGLE, 1953 Host controlled variation in bacterial viruses. J. Bacteriol. **65:** 113–121.

BICKLE, T. A., and D. H. KRÜGER, 1993 Biology of restriction. Microbiol. Rev. **57:** 434–450.

BISERCIC, M., J. Y. FEUTRIER and P. R. REEVES, 1991 Nucleotide sequences of the *gnd* genes from nine natural isolates of *Escherichia coli*: evidence of intragenic recombination as a contributing factor in the evolution of the polymorphic *gnd* locus. J. Bacteriol. **173:** 3894–3900.

BORCK, K., J. D. BEGGS, W. J. BRAMMAR, A. S. HOPKINS and N. E. MURRAY, 1976 The construction *in vitro* of transducing derivatives of phage lambda. Mol. Gen. Genet. **146:** 199–207.

BOYER, H. W., 1964 Genetic control of restriction and modification in *E. coli*. J. Bacteriol. **88:** 1652–1660.

BULLAS, L. R., C. COLSON and A. VAN PEL, 1976 DNA restriction and modification systems in Salmonella. SQ, a new system derived by recombination between the SB system of *Salmonella typhimurium* and the SP system of *Salmonella potsdam*. J. Gen. Microbiol. **95:** 166–172.

BULLAS, L. R., C. COLSON and B. NEUFELD, 1980 Deoxyribonucleic acid restriction and modification systems in *Salmonella*: chromosomally located systems of different serotypes. J. Bacteriol. **141:** 275–292.

COWAN, G. M., A. A. F. GANN and N. E. MURRAY, 1989 Conservation of complex DNA recognition domains between families of restriction enzymes. Cell **56:** 103–109.

DANIEL, A. S., F. V. FULLER-PACE, D. M. LEGGE and N. E. MURRAY, 1988 Distribution and diversity of *hsd* genes in *Escherichia coli* and other enteric bacteria. J. Bacteriol. **170:** 1775–1782.

DEBACKER, D., and C. COLSON, 1991 Transfer of the genes for the *Sty*LTI restriction-modification system of *Salmonella typhimurium* to strains lacking modification ability results in death of the recipient cells and degradation of their DNA. J. Bacteriol. **173:** 1328–1330.

DILA, D., E. SUTHERLAND, L. MORAN, B. SLATKO and E. A. RALEIGH, 1990 Genetic and sequence organization of the *mcrBC* locus of *Escherichia coli* K-12. J. Bacteriol. **172:** 4888–4900.

DRAKE, C. R., G. J. BOULNOIS and I. S. ROBERTS, 1993 The *Escherichia coli serA*-linked capsule locus and its flanking sequences are polymorphic, genetic evidence for the existence of more than two groups of capsule gene clusters. J. Gen. Microbiol. **139:** 1707–1714.

DUBOSE, R. F., D. E. DYKHUIZEN and D. L. HARTL, 1988 Genetic exchange among natural isolates of bacteria: recombination within the *phoA* gene of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **85:** 7036–7040.

DYKHUIZEN, D. F., and L. GREEN, 1991 Recombination in *Escherichia coli* and the definition of biological species. J. Bacteriol. **173:** 7257–7268.

ENDLICH, B., and S. LINN, 1985 The DNA restriction endonuclease of *Escherichia coli* B. J. Biol. Chem. **260:** 5729–5738.

FIGUEROA, F., E. GÜNTHER and J. KLEIN, 1988 MHC polymorphism pre-dating speciation. Nature **335:** 265–267.

FULLER-PACE, F. V., and N. E. MURRAY, 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., L. R. BULLAS, H. DELIUS and N. E. MURRAY, 1984 Genetic recombination can generate altered restriction specificity. Proc. Natl. Acad. Sci. USA **81:** 6095–6099.

FULLER-PACE, F. V., G. M. COWAN and N. E. MURRAY, 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.

GANN, A. A. F., A. J. B. CAMPBELL, J. F. COLLINS, A. F. W. COULSON and N. E. MURRAY, 1987 Reassortment of DNA recognition domains and the evolution of new specificities. Mol. Microbiol. **1:** 13–22.

GOUGH, J. A., and N. E. MURRAY, 1983 Sequence diversity among related genes for recognition of specific targets in DNA molecules. J. Mol. Biol. **166:** 1–19.

GUBLER, M., D. BRAGUGLIA, J. MEYER, A. PIEKAROWICZ and T. A. BICKLE, 1992 Recombination of constant and variable modules alters DNA sequence recognition by type IC restriction-modification enzymes. EMBO J. **11:** 233–240.

HERZER, P. J., S. INOUYE, M. INOUYE and T. S. WHITTAM, 1990 Phylogenetic distribution of branched RNA-linked multicopy single-stranded DNA among natural isolates of *Escherichia coli*. J. Bacteriol. **172:** 6175–6181.

HUGHES, S. G., 1977 Studies of plasmid encoded restriction and modification systems. Ph.D Thesis, University of Edinburgh, Scotland.

IIDA, S., M. B. STREIGG, T. A. BICKLE and W. ARBER, 1987 Two DNA antirestriction systems of bacteriophage P1, *darA*, and *darB*: characterization of *darA*⁻ phages. Virology **157:** 156–166.

IOERGER, T. R., A. G. CLARK and T.-H. KAO, 1990 Polymorphism at the self-incompatability locus in *Solanaceae* predates speciation. Proc. Natl. Acad. Sci. USA **87:** 9732–9735.

JANULAITIS, A., R. KAZLAUSKIENE, L. LAZAREVICIUTE, R. GILVONAUSKAITE, D. STEPONAVICIENE *et al.*, 1988 Taxonomic specificity of restriction-modification enzymes. Gene **74:** 229–232.

KAN, N. C., J. A. LAUTENBERGER, M. EDGELL and C. A. HUTCHISON III, 1979 The nucleotide sequence recognized by the *Escherichia coli* K12 restriction and modification enzymes. J. Mol. Biol. **130:** 191–209.

KANNAN, P., G. M. COWAN, A. S. DANIEL, A. A. F. GANN and N. E. MURRAY, 1989 Conservation of organization in the specificity polypeptides of two families of type I restriction enzymes. J. Mol. Biol. **209:** 335–344.

KELLEHER, J. E., A. S. DANIEL and N. E. MURRAY, 1991 Mutations that confer *de novo* activity upon a maintenance methyltransferase. J. Mol. Biol. **221:** 431–440.

KING, G., and N. E. MURRAY, 1995 Restriction alleviation and modification enhancement by the Rac prophage of *Escherichia coli* K-12. Mol. Microbiol. **16:** 769–777.

KORONA, R., and B. R. LEVIN, 1993 Phage-mediated selection and the evolution and maintenance of restriction-modification. Evolution **47:** 556–575.

KORONA, R., B. KORONA and B. R. LEVIN, 1993 Sensitivity of naturally occurring coliphages to type I and type II restriction and modification. J. Gen. Microbiol. **139:** 1283–1290.

KÜES, U., and L. A. CASSLETON, 1992 Fungal mating type genes—regulators of sexual development. Mycol. Res. **96:** 993–1006.

KUZMINOV, A., E. SCHABTACH and F. W. STAHL, 1994 χ sites in combination with RecA protein increase the survival of linear DNA in *Escherichia coli* by inactivating exoV activity of RecBCD nuclease. EMBO J. **13:** 2764–2776.

LAUTENBERGER, J. A., N. C. KAN, D. LACKEY, S. LINN, M. H. EDGELL *et al.*, 1978 Recognition site of *Escherichia coli* B restriction enzyme on ϕXsB1 and simian virus 40 DNAs: an interrupted sequence. Proc. Natl. Acad. Sci. USA **75:** 2271–2275.

LAWLOR, D. A., F. E. WARD, P. D. ENNIS, A. P. JACKSON and P. PARHAM, 1988 HLA-A and B polymorphisms predate the divergence of humans and chimpanzees. Nature **335:** 265–267.

LEVIN, B. R., 1986 Restriction-modification immunity and the maintenance of genetic diversity in bacterial populations, pp. 669–688 in *Evolutionary Processes and Theory*, edited by S. KARLIN and E. NERO. Academic Press, New York.

LEVIN, B. R., 1988 Frequency-dependent selection in bacterial populations. Philos. Trans. R. Soc. Lond. B Biol. Sci. **319:** 459–472.

MCKANE, M., and R. MILKMAN, 1995 Transduction, restriction and recombination patterns in *Escherichia coli*. Genetics **139:** 35–43.

MARKLUND, B.-I., J. M. TENNENT, E. GARCIA, A. HAMERS, M. BÅGA *et al.*, 1992 Horizontal gene transfer of the *E. coli pap* and *prs* pili operons as a mechanism for the development of tissue-specific adhesive properties. Mol. Microbiol. **6:** 2225–2242.

MEISTER, J., M. MACWILLIAMS, P. HÜBNER, H. JÜTTE, E. SKRZYPEK *et al.*, 1993 Macroevolution by transposition: drastic modification of DNA recognition by a type I restriction enzyme following Tn*5* transposition. EMBO J. **12:** 4585–4591.

MILKMAN, R., and M. M. BRIDGES, 1990 Molecular evolution of the *Escherichia coli* chromosome. III. Clonal frames. Genetics **126:** 505–517.

MILKMAN, R., and M. M. BRIDGES, 1993 Molecular evolution of the *Escherichia coli* chromosome. IV. Sequence comparisons. Genetics **133:** 455–468.

MILKMAN, R., and I. P. CRAWFORD, 1983 Clustered third-base substitution among wild strains of *Escherichia coli*. Science **221:** 378–380.

MURRAY, N. E., W. J. BRAMMAR and K. MURRAY, 1977 Lambdoid phages that simplify the recovery of *in vitro* recombinants. Mol. Gen. Genet. **150:** 53–61.

MURRAY, N. E., J. A. GOUGH, B. SURI and T. A. BICKLE, 1982 Structural homologies among type I restriction-modification systems. EMBO J. **1:** 535–539.

MURRAY, N. E., A. S. DANIEL, G. M. COWAN and P. M. SHARP, 1993

Conservation of motifs within the unusually variable polypeptide sequences of type I restriction and modification enzymes. Mol. Microbiol. **9:** 133–143.

NAGARAJA, V., M. STEIGER, C. NAGER, S. M. HADI and T. A. BICKLE, 1985a The nucleotide sequence recognised by the *Escherichia coli* D type I restriction and modification system. Nucleic Acids Res. **13:** 389–399.

NAGARAJA, V., J. C. W. SHEPHERD, T. PRIPFL and T. A. BICKLE, 1985b Two type I restriction enzymes from *Salmonella* species: purification and DNA recognition sequences. J. Mol. Biol. **182:** 579–587.

NAGARAJA, V., J. C. W. SHEPHERD and T. A. BICKLE, 1985c A hybrid recognition sequence in a recombinant restriction enzyme and the evolution of DNA sequence specificity. Nature **316:** 371–372.

NELSON, K., and R. K. SELANDER, 1992 Evolutionary genetics of the proline permease gene (*putP*) and the control region of the proline utilization operon in populations of *Salmonella* and *Escherichia coli*. J. Bacteriol. **174:** 6886–6895.

NELSON, K., T. S. WHITTAM and R. K. SELANDER, 1991 Nucleotide polymorphism and evolution in the glyceraldehyde-3-phosphate dehydrogenase gene (*gapA*) in natural populations of *Salmonella* and *Escherichia coli*. Proc. Natl. Acad. Sci. USA **88:** 6667–6671.

OCHMAN, H., and R. K. SELANDER, 1984 Standard reference strains of *E. coli* from natural populations. J. Bacteriol. **157:** 690–693.

OCHMAN, H., T. S. WHITTAM, D. A. CAUGANT and R. K. SELANDER, 1983 Enzyme polymorphism and genetic population structure in *Escherichia coli* and *Shigella*. J. Gen. Microbiol. **129:** 2715–2726.

PRICE, C., and T. A. BICKLE, 1986 A possible role for DNA restriction in bacterial evolution. Microbiol. Sci. **3:** 296–299.

PRICE, C., J. C. W. SHEPHERD and T. A. BICKLE, 1987 DNA recognition by a new family of type I restriction enzymes: a unique relationship between two different specificities. EMBO J. **6:** 1493–1497.

PRICE, C., J. LINGNER, T. A. BICKLE, K. FIRMAN and S. W. GLOVER, 1989 Basis for changes in DNA recognition by the *Eco*R124 and *Eco*R124/3 Type I DNA restriction and modification enzymes. J. Mol. Biol. **205:** 115–125.

RADDING, C. M., 1973 Molecular mechanisms in genetic recombination. Annu. Rev. Genet. **7:** 87–111.

RALEIGH, E. A., 1992 Organization and function of the *mcrBC* genes of *Escherichia coli* K-12. Mol. Microbiol. **6:** 1079–1086.

RALEIGH, E. A., R. TRIMARCHI and H. REVEL, 1989 Genetic and physical mapping of the *mcrA* (*rglA*) and *mcrB* (*rglB*) loci of *Escherichia coli* K-12. Genetics **122:** 279–296.

RAVETCH, J. V., K. HORIUCHI and N. D. ZINDER, 1978 Nucleotide sequence of the recognition site for the restriction-modification enzyme of *Escherichia coli* B. Proc. Natl. Acad. Sci. USA **75:** 2266–2270.

REDFIELD, R. J., and A. M. CAMPBELL, 1987 Structure of cryptic lambda prophages. J. Mol. Biol. **198:** 383–404.

REEVES, P., 1993 Evolution of *Salmonella* O antigen variation by interspecific gene transfer on a large scale. Trends Genet. **9:** 17–22.

ROSS, T. K., E. C. ACHBERGER and H. D. BRAYMER, 1987 Nucleotide sequence of the McrB region of *Escherichia coli* K-12 and evidence for two independent translation initiation sites at the *mcrB* locus. J. Bacteriol. **171:** 1974–1981.

SAIN, B., and N. E. MURRAY, 1980 The *hsd* (host specificity) genes of *E. coli* K12. Mol. Gen. Genet. **180:** 35–46.

SAMBROOK, J., E. F. FRITSCH and T. MANIATIS, 1989 *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

SELANDER, R. K., D. A. CAUGANT, H. OCHMAN, J. M. MUSSER, M. N. GILMOUR *et al.*, 1986 Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. Appl. Environ. Microbiol. **51:** 873–884.

SHARP, P. M., J. E. KELLEHER, A. S. DANIEL, G. M. COWAN and N. E. MURRAY, 1992 Roles of selection and recombination in the evolution of type I restriction-modification systems in enterobacteria. Proc. Natl. Acad. Sci. USA **89:** 9836–9840.

SMITH, N. H., P. BELTRAN and R. K. SELANDER, 1990 Recombination of *Salmonella* phase I flagellin genes generates new serovars. J. Bacteriol. **172:** 2209–2216.

STOLTZFUS, A., J. F. LESLIE and R. MILKMAN, 1988 Molecular evolution of the *Escherichia coli* chromosome. I. Analysis of structure and natural variation in a previously uncharacterized region between *trp* and *tonB*. Genetics **120:** 345–358.

SURI, B., J. C. W. SHEPHERD and T. A. BICKLE, 1984 The *Eco*A restriction and modification system of *Escherichia coli* 15T$^-$: enzyme structure and DNA recognition sequence. EMBO J. **3:** 575–579.

TYNDALL, C., J. MEISTER and T. A. BICKLE, 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.

WANG, L., L. K. ROMANA and P. R. REEVES, 1992 Molecular analysis of a *Salmonella enterica* group E1 *rfb* gene cluster: O antigen and the genetic basis of the major polymorphism. Genetics **130:** 429–443.

WHITTAM, T. S., H. OCHMAN and R. K. SELANDER, 1983 Multilocus genetic structure in natural populations of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **80:** 1751–1755.

WOOD, W. B., 1966 Host specificity of DNA produced by *Escherichia coli*: bacterial mutations affecting the restriction and modification of DNA. J. Mol. Biol. **16:** 118–133.

Communicating editor: M. M. HOWE

The EMBO Journal vol.15 no.8 pp.2003–2009, 1996

Restriction by *Eco*KI is enhanced by co-operative interactions between target sequences and is dependent on DEAD box motifs

Julie L.Webb, Gareth King[1], Diane Ternent, Annette J.B.Titheradge and Noreen E.Murray[2]

Institute of Cell and Molecular Biology, University of Edinburgh, The Darwin Building, The King's Buildings, Edinburgh EH9 3JR, UK

[1]Present address: Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, MA 02115, USA

[2]Corresponding author

Julie L.Webb and Gareth King contributed equally to this work

One subunit of both type I and type III restriction and modification enzymes contains motifs characteristic of DEAD box proteins, which implies that these enzymes may be DNA helicases. This subunit is essential for restriction, but not modification. The current model for restriction by both types of enzyme postulates that DNA cutting is stimulated when two enzyme complexes bound to neighbouring target sequences meet as the consequence of ATP-dependent DNA translocation. For type I enzymes, this model is supported by *in vitro* experiments, but the predicted co-operative interactions between targets have not been detected by assays that monitor restriction *in vivo*. The experiments reported here clearly establish the required synergistic effect but, in contrast to earlier experiments, they use *Escherichia coli* K-12 strains deficient in the restriction alleviation function associated with the Rac prophage. In bacteria with elevated levels of *Eco*KI the co-operative interactions are obscured, consistent with co-operation between free enzyme and that bound at target sites. We have made changes in three of the motifs characteristic of DEAD box proteins, including motif III, which in RecG is implicated in the migration of Holliday junctions. Conservative changes in each of the three motifs impair restriction.

Keywords: DEAD box motifs/DNA translocation/helicase/Rac prophage/restriction enzymes

Introduction

Restriction and modification (R–M) systems are remarkable for their diversity. They are separated commonly into three major types on the basis of their organization and properties (for reviews, see Wilson and Murray, 1991; Bickle, 1993; Bickle and Krüger, 1993; Halford *et al.*, 1993; Heitman, 1993; King and Murray, 1994; Barcus and Murray, 1995). The first R–M systems studied, now designated type I, are the most complex; they comprise three different subunits, HsdR, HsdM and HsdS. Each system recognizes its specific target sequence, the methylation status of which determines whether the R–M complex responds by modifying (methylating) or restricting (cutting) the DNA. Hemimethylated target sequences elicit methylation of the unmodified strand. Restriction, the response to unmodified target sequences, is ATP dependent and involves the cutting of non-specific DNA sequences at variable distances from target sites.

Type I systems, despite their similarities, are sufficiently heterogeneous to warrant their subdivision into different families. Within each family, polypeptide and gene sequences are well conserved, though members that recognize different target sequences have notable differences within HsdS, the single subunit that confers on the multimeric complex specificity for a particular target sequence. These regions of extensive differences, originally referred to as variable regions (Gough and Murray, 1983), identify target recognition domains (TRDs) (see any of the reviews cited above and Noyer-Weidner and Trautner, 1993).

The sequences of the *hsd* genes for representatives of three different families of type I R–M systems have been determined (Loenen *et al.*, 1987; Price *et al.*, 1989; Sharp *et al.*, 1992; Murray *et al.*, 1993) and comparisons of the inferred polypeptide sequences detect quite limited similarities, only 20–35% identities in pairwise interfamily comparisons (Sharp *et al.*, 1992; Murray *et al.*, 1993). The conserved sequences are predicted to include amino acids of key functional importance for the activities common to type I R–M enzymes. In HsdS, the similar sequences include candidate regions for subunit interactions (Kannan *et al.*, 1989; Cooper and Dryden, 1994; Kneale, 1994), or entire TRDs when two HsdS subunits share a common component in their DNA target sequence (Fuller-Pace and Murray, 1986; Cowan *et al.*, 1989). For the HsdM polypeptides two motifs have been implicated in the transfer of methyl groups to the target sequences. Analyses of mutants indicate that one of these motifs is involved in the binding of the methyl donor, *S*-adenosylmethionine, and the other in the methylation reaction itself (Willcock *et al.*, 1994).

The third subunit, HsdR, is essential for restriction, but is not required for modification. The *hsdR* genes for all three families of enzymes encode polypeptides containing motifs characteristic of ATP binding proteins, consistent with the ATP dependence of restriction. In addition, each HsdR subunit includes other motifs previously identified in ATP-dependent helicases or putative helicases (Linder *et al.*, 1989; Gorbalenya and Koonin, 1991; Murray *et al.*, 1993; Fuller-Pace, 1994). One of these, a common variant of the element Asp–Glu–Ala–Asp, or DEAD in a single letter code, has given its name to this family of proteins (Linder *et al.*, 1989).

It has been tempting to speculate that the DEAD box motifs are relevant to the mechanism of restriction, in particular to the ATP-dependent translocation of DNA that precedes the cutting of DNA (Murray *et al.*, 1993).

One attractive model for restriction by type I R–M systems postulates that cutting of DNA is stimulated when two enzyme complexes bound to their target sequences meet as the consequence of ATP-dependent DNA translocation (Studier and Bandyopadhyay, 1988). Type III R–M systems, like type I, are DEAD box proteins (Gorbalenya and Koonin, 1991; Dartois *et al.*, 1993). A common role for DEAD box motifs in the restriction subunits of type I and type III enzymes could be satisfied if ATP-dependent DNA translocation were involved in restriction by both classes of enzymes. It is known that the target necessary for restriction by a type III system consists of two recognition sequences, each of the same six nucleotides but in inverse orientation (Meisel *et al.*, 1992). The distance between the two components of the target sequence may be large, and one obvious means of communication between two distant sites would be ATP-dependent DNA translocation (Murray *et al.*, 1993). Very recent data show that DNA restriction is dependent on the hydrolysis of ATP (Meisel *et al.*, 1995; Saha and Rao, 1995), and translocation is implicated by the finding that Lac repressor bound at an operator site located between the two components of a restriction target inhibits cutting by the type III enzyme *Eco*P15I (Meisel *et al.*, 1995). A role for DNA translocation in restriction by type III systems is readily appreciated where the two components of a restriction target are some distance apart. The target for a type I R–M enzyme, however, is a single bipartite sequence which in itself is sufficient to elicit restriction either *in vivo* (Murray *et al.*, 1973a) or *in vitro* (Horiuchi and Zinder, 1972; Murray *et al.*, 1973b). Here we report evidence from *in vivo* experiments for co-operative interactions between two widely separated targets for the type I restriction enzyme *Eco*KI. We also report the isolation of mutants in which each of three of the so-called 'DEAD box' motifs of the HsdR subunit of the type I system of *Escherichia coli* K-12 (*Eco*KI) have been changed. Conservative changes in each of these three motifs affect restriction activity.

Results

The cloned hsdR gene enhances restriction

Plasmid pBg3 includes the *hsdR* gene of *E.coli* K-12 (Sain and Murray, 1980), but it provides poor amplification of HsdR and is not an ideal substrate for mutational analysis of *hsdR*. A plasmid including the T7 promoter, pT7-7, was chosen for the expression of *hsdR*. The *Bam*HI target in the polylinker of this vector was deleted to permit the subsequent manipulations dependent on the *Bam*HI target in the *hsdR* gene (pT7-7Δ, see Figure 1). The 3.6 kb *Eco*RI–*Sma*I fragment including all but the 5′ end of *hsdR* was cloned in pT7-7Δ (pSB1, Figure 1). The N-terminal sequence of HsdR (D.T.F.Dryden and L.Cooper, personal communication) is consistent with initiation of translation at the second methionine codon of the open reading frame that identified *hsdR* (Loenen *et al.*, 1987). This N-terminal sequence defined the length of the short oligonucleotide linker (Figure 1) which, on insertion into pSB1, completed the coding sequence for HsdR. The resulting plasmid, pSB2, includes the *hsdR* gene with the upstream translational initiation region and promoter of T7 gene 10 and it contains unique *Kpn*I and *Bam*HI sites flanking the three sequences targeted for site-directed mutagenesis.

Plasmid pSB2 was recovered in the *hsd*Δ strain NM679 in the absence of the λDE3 prophage, and consequently in the absence of T7 RNA polymerase. The phenotypic test for a functional HsdR polypeptide requires the presence of HsdM and HsdS to generate the *Eco*KI complex, therefore transformants of *hsdR*Δ4M^+S^+ strains were isolated and checked for restriction. Both NM526 and a derivative lysogenic for λDE3, which includes the T7 gene encoding RNA polymerase under the control of the *lac* UV5 promoter, were used. Transformants of both strains were r_k^+, indicating expression of the *hsdR* gene even in the absence of T7 RNA polymerase. The combined presence of the λDE3 prophage and pSB2 resulted in strains that grew slowly even in the absence of inducer. Since T7 RNA polymerase was unnecessary for the expression of *hsdR*, all quantitative phenotypic tests were made in the absence of the λDE3 prophage.

Transformants of a number of *hsd*$^+$ and *hsdR*$^-$ M^+S^+ strains were isolated and checked for restriction of λ*vir*.0. All were r_k^+. However, for transformants of either AB1157 or Ymel, *hsd*$^+$ strains lacking the Rac prophage, restriction was greatly enhanced.

The Rac prophage present in most *E.coli* K-12 strains reduces the maximal level of restriction by *Eco*KI (Loenen and Murray, 1986; King and Murray, 1995). It is assumed that in rare cells ($\sim 10^{-4}$) the Rac prophage spontaneously induces and, as a consequence, its restriction alleviation gene is expressed and restriction is blocked. In the absence of the Rac prophage, the presence of multiple copies of *hsdR* in strains transformed with pSB2 leads to elevated levels of restriction. In a Rac$^+$$r_k^+$ strain, the efficiency of plating (e.o.p.) of λ*vir*.0 is 2×10^{-4}, in a Rac$^-$ strain, such as AB1157 or Ymel, it is reduced to 10^{-5} while the presence of pSB2 in a Rac$^-$ strain reduces it to as little as 10^{-8}. We conclude that in wild-type cells restriction is limited by a deficiency of HsdR. This conclusion is consistent with estimates of the relative levels of Hsd subunits in minicells carrying a plasmid encoding *Eco*KI (Weiserova *et al.*, 1993).

Co-operative interactions between widely separated targets

The realization that the restriction alleviation function of the Rac prophage obscures the maximal level of restriction led us to re-examine the correlation between target number and level of restriction. The Rac$^-$ strains AB1157 and AB1157 transformed with pSB2 were used.

Hybrid phages ($h^{80}imm^{\lambda}$) with zero, one or two targets for restriction for *Eco*KI (Murray *et al.*, 1973a; Brammar *et al.*, 1974) were grown on a non-modifying strain (NM679) and their e.o.p. on a Rac$^-$$r_k^+$ strain (AB1157) determined relative to a congenic *hsdR*$^-$ derivative (NM795). Derivative phages that lacked targets for *Eco*KI were used to monitor any other factors affecting e.o.p. In the absence of co-operative effects between targets, the expected restriction of a phage with two targets was calculated as the product of the values found for each of the two phages with single targets. Two phages with two targets were tested; one has two λ targets (skλ1 and skλ2) separated by ~12 kb and the other, a *trp*-transducing

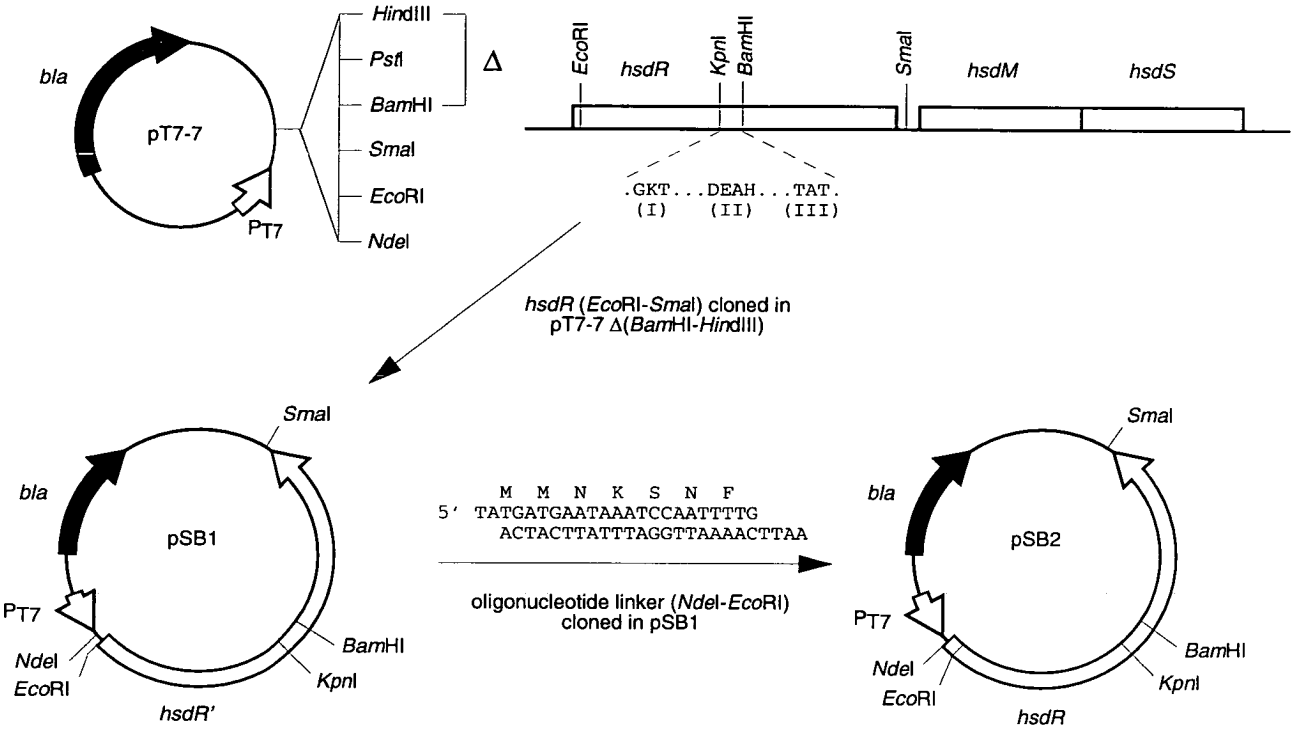


Fig. 1. Origin of pSB2. A derivative of the expression vector pT7-7 was made which lacks the *Bam*HI target in the polylinker (pT7-7Δ). The 3.6 kb *Eco*RI–*Sma*I fragment containing most of *hsdR* was cloned in pT7-7Δ (pSB1), and the 5′ end of *hsdR* was added as an *Nde*I–*Eco*RI synthetic linker (pSB2). The only *Kpn*I and *Bam*HI targets in pSB2 are those that flank the sequence encoding the DEAD box motifs I–III. *bla*, gene coding for β-lactamase; pT7, the promoter and upstream translational initiation region from phage T7 gene 10; GKT, DEAH and TAT denote the positions of the DEAD box motifs I, II and III, respectively. The maps are not drawn to scale.

Table I. *Eco*KI targets co-operate to enhance restriction

Targets in $h^{80}imm^{\lambda}$ phages	Host	No. of experiments	'Enhanced restriction' = Observed value[a] / Expected value[b]		
			Minimum	Maximum	Average
skλ1 + skλ2 (targets ~12 kb apart)	AB1157	10	7	18	15
sk*trp* + skλ2 (targets ~2 kb apart)	AB1157	9	8	20	13
skλ1 + skλ2	AB1157 + pSB2	5	0.6	1.6	1.1
skλ*trp* + skλ2	AB1157 + pSB2	4	1.4	1.6	1.5

[a]Observed value determined from e.o.p. of phage with two targets. The e.o.p. on AB1157 was determined relative to NM795, the *hsdR* derivative of AB1157, and on AB1157 + pSB2 relative to NM795 harbouring the vector plasmid pT7-7.
[b]Expected value based on the product of the values for the two phages with one target.

phage, has skλ2 and a second target in *trpE* (sk*trp*) separated by only a few kilobases.

In all experiments using AB1157, restriction of phages with two targets was much in excess over the value predicted in the absence of co-operative interactions between targets. The lowest estimate of enhanced restriction for AB1157 was 7-fold, the highest 20-fold (Table I). In contrast, in the presence of the plasmid pSB2, when restriction of phages with even single targets is greatly elevated (400-fold rather than 20-fold in AB1157), the observed restriction for two targets was as expected in the absence of co-operative interactions between targets (Table I).

Mutations in the DEAD box motifs

The sequences in which the mutations were made are shown in Figure 2. Motifs I and II are as identified by Gorbalenya and Koonin (1991). These authors, however, noted that motif III was difficult to identify in the *hsd*R gene of a type IA system. The recently corrected nucleotide sequence for *hsdR* of *E.coli* K-12 (Blattner *et al.*, unpublished, GenBank accession No. U14003) introduces a short frame-shifted segment within which motif III is readily recognized. This sequence, TAT, aligns with TAT and TGT respectively, in type IB and IC HsdR polypeptides (Figure 2). TAT was the third target for mutagenesis.

Mutations in *hsdR* were made by site-directed mutagenesis using PCR. The 550 bp fragment, generated by PCR, was cut with *Bam*HI and *Kpn*I, or its neoschizomer *Asp*718, and substituted for the wild-type sequence of pSB2. The nine mutations shown in Figure 2 were identified by DNA sequence. All plasmids contain the expected sequence in the PCR-derived region and no other changes.

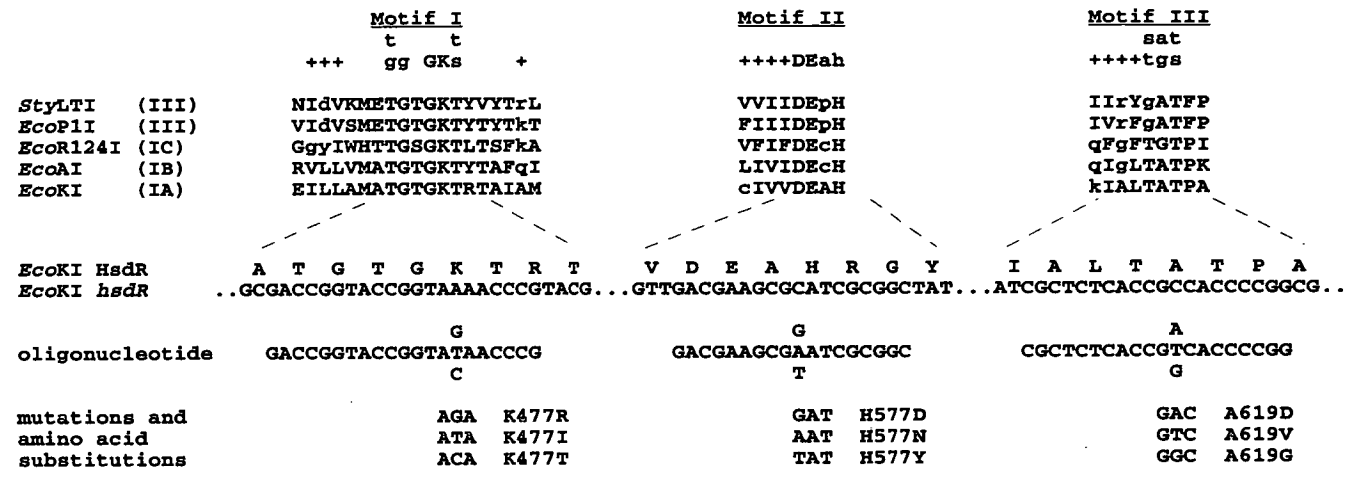


Fig. 2. Alignment of DEAD box motifs in type I and III R–M systems and identification of the substitutions made in *Eco*KI. The consensus sequence for DEAD box proteins is shown above an alignment of the predicted amino acid sequences of *Sty*LTI, *Eco*P1I (type III R–M systems), *Eco*R124I, *Eco*AI and *Eco*KI (type IC, IB and IA R–M systems, respectively) in the regions of motif I, II and III. Shown below is the nucleotide sequence of *hsdR* from *E.coli* K-12 in the region of each motif, the sequence of one of each pair of complementary oligonucleotides used for the site-directed mutagenesis, and the amino acid substitutions generated. The numbers identify the location of the amino acid in the HsdR polypeptide. In the DEAD box consensus, upper case letters indicate invariant residues, lower case letters denote preferred residues and the symbol '+' is for hydrophobic residues. In the alignments, lower case letters are used to indicate a deviation from the consensus sequence.

Table II. The effects of changes in the DEAD box motifs on the restriction phenotype

Motif	Sequence	Mutation on plasmid		Mutation on chromosome[b]
		in *hsdR*$^-$ host[a]	in *hsdR*$^+$ host[a]	
(I)	GTGRT	r_k^-	r_k^-	r_k^-
GTGKT	GTGIT	r_k^-	r_k^-	nt
	GTGTT	r_k^-	r_k^-	nt
(II)	DEAD	r_k^-	r_k^-	r_k^-
DEAH	DEAN	r_k^-	r_k^-	nt
	DEAY	r_k^-	r_k^+	r_k^-
(III)	TDT	r_k^-	r_k^-	nt
TAT	TVT	r_k^-	r_k^-	r_k^-
	TGT	r_k^+	r_k^+	r_k^{+c}

[a]The *hsdR*$^-$ strain, NM795, is a derivative of the *hsdR*$^+$ strain, AB1157.
[b]The mutations were transferred to the chromosome of AB1157 via a λ*hsdR*$^-$ phage.
[c]See Table III for quantification.
r_k^- is inferred from an e.o.p. of ~1 relative to the r_k^- strain NM795; nt = not tested.

Phenotypes of mutants

Tests dependent on multicopy plasmids. The parent plasmid, pSB2, when maintained in an *hsdR* strain, compensates for the chromosomal *hsdR* mutation; the restriction-proficient phenotype was recognized by the low e.o.p. of λ*vir*.0. The phenotypes associated with the presence of the mutant derivatives of pSB2 were tested in the same way; a low e.o.p. of unmodified phages would be indicative of a functional HsdR polypeptide. On the basis of this test, only one mutation (TAT to TGT in motif III) fails to inactivate the HsdR polypeptide (Table II).

This test, however, is not ideal for at least two reasons. First, any cell lacking the plasmid will fail to restrict λ*vir*.0 and progeny phages will be modified by the *Eco*KI methyltransferase. It is, therefore, critical that the vast majority of cells retain the plasmid, otherwise a false indication of restriction deficiency will result. Second, the level of HsdR polypeptide produced by the multicopy plasmid will be very different from that found in a cell with only a chromosomal copy of the *hsdR* gene. The first of these problems is avoided if the mutant polypeptide can compete with wild-type and, particularly when present in excess, can convert an *hsd*$^+$ host into an r_k^- phenocopy.

Each of the nine plasmids carrying a mutant *hsdR* gene was used to transform the r_k^+ strain AB1157, and the restriction phenotypes of the transformed strains were tested. Seven of the transformed strains were clearly r_k^- (see Table II), a phenotype that requires the presence of an inactive HsdR polypeptide. Two were r_k^+, consistent with either a functional HsdR polypeptide, or an inactive polypeptide that fails to displace the wild-type HsdR subunit.

Test dependent on mutations in the bacterial chromosome. Quantification of the restriction phenotype of any *hsdR* mutant is only possible if the mutation is stably maintained in single copy on the bacterial chromosome, preferably in its normal location under the control of its resident promoter. Five of the mutations have been transferred to the chromosome of the Rac$^-$ strain AB1157 and restriction quantified using λ*vir*.0. All but one of the mutants tested, including the conservative change of DEAH to DEAD,

Table III. Changing motif III from TAT to TGT reduces restriction

Phage (targets for *Eco*KI)	Efficiency of plating (e.o.p.)		Restriction[b]	
	Wild-type[a]	Mutant (TGT)[a]	Wild-type	Mutant
λNM105	2.7×10^{-4}	1.1×10^{-2}	3227 ± 1007	121 ± 51
(skλ1 + skλ2)	2.9×10^{-4}	9.6×10^{-3}		
	5.7×10^{-4}	1.1×10^{-2}		
	2.5×10^{-4}	5.1×10^{-3}		
λNM175	1.6×10^{-2}	8.6×10^{-2}	38 ± 23	8.3 ± 2.4
(skλ1)	1.6×10^{-2}	1.2×10^{-1}		
	7.7×10^{-2}	1.4×10^{-1}		
	3.8×10^{-2}	1.1×10^{-1}		
	4.3×10^{-2}	1.9×10^{-1}		
λNM176	2.3×10^{-2}	3.1×10^{-1}	33 ± 10	3.5 ± 1.5
(skλ2)	2.8×10^{-2}	4.8×10^{-1}		
	2.9×10^{-2}	3.1×10^{-1}		
	5.0×10^{-2}	1.8×10^{-1}		

[a]The wild-type strain AB1157 and the congenic mutant derivative were used. The e.o.p. was calculated relative to an r_k^- derivative of AB1157. In all experiments a hybrid phage with no targets for *Eco*KI gave an e.o.p. of ~1.
[b]A value for restriction was estimated from the inverse of each e.o.p. The means and standard deviations are shown.

were completely deficient in restriction. The remaining mutant (TAT to TGT) gave reduced levels of restriction, only readily detected for phages with one or two targets (Table III), rather than for λ*vir* with its complement of five targets (Table II). All of the nine mutations affect restriction, eight of them severely, and seven of these make polypeptides that compete with wild-type for interaction with the methyltransferase. As expected, the chromosomal *hsdR* mutations affected only restriction and not modification.

Discussion

Many enzymes containing the purine NTP binding motif have been identified as members of a 'super family' sharing a number of motifs including GxGKS/T, an abridged version of the 'A' component of the ATP binding motif (Linder *et al.*, 1989). Within this super family of 'DEAD box' proteins, there are subgroups which include either the motif 'DEAD' or 'DExH', (where x is commonly A or C), and at least five other conserved regions (Linder *et al.*, 1989). Some of these proteins are known to be helicases, the remainder are regarded as putative helicases (Schmid and Linder, 1992).

Mutational analyses of DEAD box proteins have been reported, the most extensive being for an RNA helicase, the mammalian translation initiation factor eIF-4A (Pause and Sonenberg, 1992). The results suggest that the highly conserved regions in the DEAD box proteins are critical for helicase activity. Mutations in motifs I and II affect ATP binding and ATP hydrolysis respectively, while those in motif III can affect helicase activity in the absence of any major effect on ATP hydrolysis.

Some members of the DEAD box protein family have been shown to be DNA, rather than RNA, helicases. These include RAD 3 of yeast (Sung *et al.*, 1988), and RecG (Sharples *et al.*, 1994) and RecQ (Umezu *et al.*, 1990) of *E.coli*, all involved in recombination and repair processes. RecG includes all seven motifs characteristic of the DExH family (Lloyd and Sharples, 1991) and catalyses branch migration of Holliday junction intermediates in a reaction that requires ATP hydrolysis. A mutation in motif III, in which the sequence TAT was changed to TVT, was shown to reduce ATP hydrolysis and to block branch migration (Sharples *et al.*, 1994). For RecG, it is concluded that branch migration of Holliday junctions is related to an ATP-dependent helicase activity.

A search of the inferred HsdR polypeptide sequences of *Eco*KI (type IA) and *Eco*R124I (type IC) for the seven motifs identified in one family of DEAD box proteins identified six of the seven in each HsdR polypeptide (Gorbalenya and Koonin, 1991). The counterpart to motif III was only identified with any confidence in *Eco*R124I and that to motif IV only in the HsdR polypeptide of *Eco*KI. More recently, two items of sequence data add further support to justify the identification of type I R–M systems as DEAD box proteins. First, within the inferred sequences of the HsdR polypeptides of *Eco*AI and *Eco*EI, two members of another family (type IB) of type I R–M systems, all seven motifs were identified (Murray *et al.*, 1993). Second, the correction to the sequence of the *hsdR* gene of *E.coli* K-12, made by the Blattner group (GenBank accession No. U14003), introduces a short frame-shift within which is a convincing motif III, one of the targets of our analysis. The 'DEAD box' motifs I, II and III in the HsdR polypeptides of three type I systems are shown in Figure 2, together with those identified in two type III R–M systems (Gorbalenya and Koonin, 1991; Dartois *et al.*, 1993).

Comparisons of the amino acid sequences of the HsdR polypeptides of the three families of type I R–M systems with those of the restriction subunits of type III enzymes emphasize the DEAD box motifs. It has been suggested that the implicated helicase activity may be involved in the local unwinding of DNA at the cleavage site (Gorbalenya and Koonin, 1991; Dartois *et al.*, 1993). There are some type II enzymes (type IIs) which do not cut DNA within their target sequence. These enzymes are not ATP dependent, they are not DEAD box proteins, and DNA translocation is not implicated. Rather than a role for DEAD box motifs in local unwinding of DNA at a cleavage site removed from the target sequence, we favour a role in the ATP-dependent translocation that has been proposed to precede DNA cutting, by both type I and type

III R–M enzymes. Recent data for type III enzymes (Meisel *et al.*, 1995), and the experiments reported in this paper, emphasize DNA translocation as the feature common to type I and type III R–M systems.

The model proposed for *Eco*KI (Studier and Bandyopadhyay, 1988), in which the cutting of DNA is stimulated when two *Eco*KI enzyme complexes bound to their target sequences meet as the result of ATP-dependent DNA translocation, can be adapted for type III systems (Murray *et al.*, 1993; Meisel *et al.*, 1995). The target sequence necessary to elicit restriction by the type III enzyme *Eco*P15I comprises two identical, but inversely oriented, sequences separated by as much as 3 kb; DNA translocation would provide the necessary means of communication between the two components of the target.

Early experiments showed that *Eco*KI remains bound to its target sequence but cuts the DNA elsewhere (Bickle *et al.*, 1978). The model of Studier and Bandyopadhyay (1988) was based on the demonstration that, under appropriate conditions, DNA cutting does not occur randomly, but mid-way between two target sites. Although this model is well supported by the analysis of the products of *in vitro* reactions, previous quantification of restriction *in vivo* has not indicated the predicted co-operative interactions between pairs of target sequences (Murray *et al.*, 1973a). Furthermore, the observation that a single target within a phage genome is sufficient to elicit restriction requires significant modification of the model. The *in vivo* experiments presented in this paper now document the anticipated co-operative interactions between pairs of well separated target sequences. They also show that at higher concentrations of enzyme co-operative interactions are less evident, in agreement with the suggestion that unbound enzyme can co-operate with bound complex and cut DNA substrates that have only one target sequence (Studier and Bandyopadhyay, 1988). The signal necessary to elicit restriction by *Eco*KI *in vivo*, when a phage genome has only a single target, could be dependent on the substrate being a covalently closed, circular molecule and the consequent constraints imposed on DNA translocation. *In vitro*, cutting linear λ genomes with only single targets for *Eco*KI was inefficient and needed 40-fold more enzyme than the cutting of substrates with two or more targets (Murray *et al.*, 1973b).

It has been suggested that the DNA translocation mediated by *Eco*KI is associated with topoisomerase activity, but the nicking and joining of DNA strands by *Eco*KI has not been documented (Yuan *et al.*, 1980). Similarly type I and III R–M enzymes remain *putative* DNA helicases; our genetic data indicate that the motifs characteristic of DEAD box proteins are relevant to restriction but their role, even if indicative of helicase activity, remains unknown. Of particular interest is the change of the TAT sequence to TGT, the sequence found in the type IC enzyme *Eco*R124I. This motif has been correlated with helicase activity in other systems (Sharples *et al.*, 1994). For *Eco*KI, as is the case for RecG, the change from TAT to TVT leads to the loss of activity, whereas the very conservative change to TGT leads to only a modest reduction in the efficiency of restriction without a loss in co-operativity. Conservative changes in motifs I and II (GTGKT to GTGRT and DEAH to DEAD) resulted in subunits that appear to compete with wild-type HsdR for the other components of the complex and yield enzymes defective in restriction. Mutations in the HsdR polypeptide should facilitate the dissection of the complex, ATP-dependent steps that intervene between the specific binding of *Eco*KI to its target sequence and the cutting of DNA at sites remote from the target sequence.

The expression of *hsdR* in pT7-7 did not provide effective yields of HsdR, as assayed by immunological tests of polypeptides following their separation on polyacrylamide gels (data not shown), although the restriction phenotype associated with the presence of pSB2 in the *hsd*$^+$ strain AB1157 indicated enhanced levels of *Eco*KI. Even the levels of HsdR provided by pSB2 were detrimental to the bacterial cell in the presence of the other subunits needed to make the restriction enzyme. This suggests that an excess of modification enzyme is normally required to maintain viability of the bacteria. Since the endonuclease is active only on unmethylated DNA, some newly synthesized DNA, perhaps the product of DNA repair, may be the target for restriction.

Materials and methods

Bacterial strains, phages and plasmids

Three restriction-proficient (r_k^+), modification-proficient (m_k^+) *E.coli* K-12 strains, C600 (Appleyard, 1954), Ymel (Rickenberg and Lester, 1955) and AB1157 (Low, 1973), were used. C600 generally has been the standard $r_k^+m_k^+$ host, but both Ymel (King and Murray, unpublished observations) and AB1157 (Kaiser and Murray, 1979) are preferred hosts for quantifying restriction because they lack the restriction alleviation function associated with the Rac prophage. The phenotypes of the *hsdR* mutants were first examined following transformation of *hsdR*(r_k^-) derivatives of Ymel (NM526), and AB1157 (NM795), with plasmids carrying mutated *hsdR* genes. NM526 has an extensive deletion (Δ4) in the *hsdR* gene; NM795 is a missense mutant. Checks for negative complementation were made following transformation of AB1157. NM526 lysogenic for λDE3 (Studier and Moffatt, 1986; King and Murray, 1995) was used to elicit transcription of *hsdR* by T7 RNA polymerase.

DNA from λNM1050, a phage including the complete coding sequence for *Eco*KI (Sain and Murray, 1980), was the template for amplification by the PCR. λ virulent (laboratory collection), either unmodified (λ*vir*.0) following propagation on NM679 (Δ*hsdRMS*; King and Murray, 1995), or *Eco*KI-modified (λ*vir*.K) by growth on C600, was used to measure restriction activity. The plasmid pBg3 (Sain and Murray, 1980) was the source of the truncated *hsdR* gene cloned in the expression vector pT7-7 (Tabor, 1990). Phages λNM1265 (Whittaker *et al.*, 1988) and λNM1347, a *c*I857 derivative of λNM1266 (Whittaker *et al.*, 1988), were used to clone mutant *hsdR* genes and facilitate their transfer from the plasmid to the host chromosome.

Hybrid phages with the left arm of *phi*80 substituted for that of λ have reduced numbers of targets (sk) for *Eco*KI (Franklin and Dove, 1969). Hybrid phages with either one or two targets were used as substrates to monitor restriction *in vivo*. Phages of the genotype h^{80} att^{λ} *int*29 sk$\lambda 2^+$ *c*I857 sk$\lambda 1^+$ and mutant derivatives that have lost skλ1, skλ2 or both targets were those described by Murray *et al.* (1973a). Hybrid *trp*-transducing phages with a target in *trpE* (sk*trp*) and skλ2 together with mutant derivatives lacking either one or the other site have been described by Brammar *et al.* (1974).

DNA manipulation

Preparation and manipulation of DNA, nucleotide sequencing and site-directed mutagenesis using the PCR were performed as described previously (King and Murray, 1995). The origin of the plasmid, pSB2, that encodes the HsdR polypeptide is shown in Figure 1.

The three motifs analysed by site-directed mutagenesis are between unique *Kpn*I and *Bam*HI targets in the *hsdR* gene of plasmid pSB2. The oligonucleotide primers used in the PCR to amplify this 550 bp region were 5′CGATGGCGACCGGTACCG and 5′GGCGGATCCTGGTCG-ATC, in which the *Kpn*I (*Asp*718) and *Bam*HI targets respectively are underlined. The *hsdR* fragment was excised by *Bam*HI in combination with either *Kpn*I or *Asp*718.

The site-directed mutagenesis of motifs II and III used the recombinant PCR technique as described by Higuchi (1990). One of the primers used to change the coding sequence for each motif is given in Figure 2; the second primer was the complement. In each case, degeneracies were introduced at one position (see Figure 2). For motif I the mutations were sufficiently close to the *Kpn*I site that the oligonucleotide primer included both the target site and the sequence to be mutated. The sequence of this primer, and the degeneracies associated with the codon for the lysine (K) residue, are also shown in Figure 2. Vent$_R$ DNA polymerase (New England BioLabs) was used, and for each mutation the nucleotide sequence of the entire *Kpn*I–*Bam*HI fragment was determined to identify the mutation and confirm the absence of other mutations.

The mutations were transferred to a λ vector to facilitate their transfer to the *E.coli* chromosome. The 3.6 kb *Eco*RI–*Sma*I fragment containing all but the first few codons of *hsdR* (see Figure 1) was excised from mutant derivatives of pSB2 and used as a linker to join the left arm of the *Pam* phage λNM1265 to the right arm of the *Eam* phage λNM1347. The genome of the former phage was cut with *Sma*I, and that of the latter with *Eco*RI. Phage genomes were recovered by *in vitro* packaging (Epicentre Technologies). The required recombinants, which in contrast to both parental phages lack amber mutations, were selected on a *sup*° *hsd*Δ host NM679. The presence of the insert was confirmed by analysis of digests of phage DNA.

Transfer of mutations to the bacterial chromosome

The λ phages including *hsdR* encode a temperature-sensitive repressor (*c*I857) and are defective in their attachment site (*b*527). Lysogens selected as immune colonies at 32°C frequently result from homology-dependent recombination, but they occasionally lose their prophage by homologous recombination. These 'cured' clones may be selected at 42°C. The *b*527 mutation greatly reduces the frequency of site-specific integration but does not entirely prevent it. Lysogens resulting from site-specific recombination only rarely produce cured clones and liberate no or few progeny phage.

λ phages in which the *hsdR* mutation has inactivated restriction activity gave both r_k^+ and r_k^- lysogens of AB1157, depending on the location of the integrative cross-over with respect to the mutation. Roughly 50% of the cured colonies selected at 42°C have replaced the chromosomal *hsdR* allele with that of the phage. The presence of the required mutation on the chromosome was confirmed by determination of the nucleotide sequence using the PCR. One mutant (A619G) failed to give r_k^- lysogens in an *hsdR*$^+$ host but, as expected, some cured derivatives acquired the mutation.

Acknowledgements

We are indebted to Simon Boulton for pSB2 and to Laurie Cooper and David Dryden for making available the N-terminal sequence of the HsdR polypeptide. We thank all our colleagues for their constructive comments and Josie Priestley for her patient help in the preparation of this manuscript. The work was supported by the MRC and studentships to G.K. (MRC) and J.L.W. (BBSRC).

References

Appleyard,R.K. (1954) *Genetics*, **39**, 440–452.
Barcus,V.A. and Murray,N.E. (1995) In Baumberg,S., Young,J.P.W., Wellington,E.M.H. and Saunders,J.R. (eds), *Population Genetics of Bacteria*. Cambridge University Press, Cambridge, pp. 31–58.
Bickle,T.A. (1993) In Linn,E.M., Lloyd,R.S. and Roberts,R.J. (eds), *Nucleases*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 89–109.
Bickle,T.A. and Krüger,D.H. (1993) *Microbiol. Rev.*, **57**, 434–450.
Bickle,T.A., Brack,C. and Yuan,R. (1978) *Proc. Natl Acad. Sci. USA*, **75**, 3099–3103.
Brammar,W.J., Murray,N.E. and Winton,S. (1974) *J. Mol. Biol.*, **90**, 633–647.
Cooper,L.P. and Dryden,D.T.F. (1994) *J. Mol. Biol.*, **236**, 1011–1021.
Cowan,G.M., Gann,A.A.F. and Murray,N.E. (1989) *Cell*, **56**, 103–109.
Dartois,V., De Backer,O. and Colson,C. (1993) *Gene*, **127**, 105–110.
Franklin,N.C. and Dove,W.F. (1969) *Genet. Res.*, **14**, 151–157.
Fuller-Pace,F.V. (1994) *Trends Cell Biol.*, **4**, 271–274.
Fuller-Pace,F.V. and Murray,N.E. (1986) *Proc. Natl Acad. Sci. USA*, **83**, 9368–9372.
Gorbalenya,A.E. and Koonin,E.V. (1991) *FEBS Lett.*, **291**, 277–281.
Gough,J.A. and Murray,N.E. (1983) *J. Mol. Biol.*, **166**, 1–19.
Halford,S.E., Taylor,J.D., Vermonte,C.L.M. and Vipond,I.B. (1993) In Ecstein,F. and Lilley,D.M.J. (eds), *Nucleic Acids and Molecular Biology*. Springer-Verlag, Berlin, Vol. 7, pp. 47–69.
Heitman,J. (1993) In Setlow,J.K. (ed.), *Genetic Engineering*. Plenum Press, New York, Vol. 15, pp. 57–108.
Higuchi,R. (1990) In Innis,M.A., Gelfand,D.H., Sninsky,J.J. and White,T.J. (eds), *PCR Protocols: A Guide to Methods and Applications*. Academic Press, Inc., San Diego, CA, pp. 177–183.
Horiuchi,K. and Zinder,N.D. (1972) *Proc. Natl Acad. Sci. USA.*, **69**, 3220–3232.
Kaiser,K. and Murray,N.E. (1979) *Mol. Gen. Genet.*, **175**, 159–174.
Kannan,P.R., Cowan,G.M., Daniel,A.S., Gann,A.A.F. and Murray,N.E. (1989) *J. Mol. Biol.*, **209**, 335–344.
King,G. and Murray,N.E. (1994) *Trends Microbiol.*, **2**, 465–469.
King,G. and Murray,N.E. (1995) *Mol. Microbiol.*, **16**, 769–777.
Kneale,G.G. (1994) *J. Mol. Biol.*, **243**, 1–5.
Linder,P., Lasko,P.F., Ashburner,M., Leeroy,P., Nielsen,P.J., Nishi,K., Schnier,J. and Slonimski,P.P. (1989) *Nature*, **337**, 121–122.
Lloyd,R.G. and Sharples,G.J. (1991) *J. Bacteriol.*, **173**, 5414–5418.
Loenen,W.A.M. and Murray,N.E. (1986) *J. Mol. Biol.*, **190**, 11–22.
Loenen,W.A.M., Daniel,A.S., Braymer,H.D. and Murray,N.E. (1987) *J. Mol. Biol.*, **198**, 159–170.
Low,B. (1973) *J. Bacteriol.*, **113**, 798–812.
Meisel,A., Bickle,T.A., Krüger,D.H. and Schroeder,C. (1992) *Nature*, **355**, 467–469.
Meisel,A., MacKeldanz,P., Bickle,T.A., Krüger,D.H. and Schroeder,C. (1995) *EMBO J.*, **14**, 2958–2966.
Murray,N.E., Manduca de Ritis,P. and Foster,L.A. (1973a) *Mol. Gen. Genet.*, **120**, 261–281.
Murray,N.E., Batten,P.L. and Murray,K. (1973b) *J. Mol. Biol.*, **81**, 395–407.
Murray,N.E., Daniel,A.S., Cowan,G.M. and Sharp,P.M. (1993) *Mol. Microbiol.*, **9**, 133–143.
Noyer-Weinder,M. and Trautner,T. (1993) In Jost,J.P. and Saluz,H.P. (eds), *DNA Methylation: Molecular Biology and Biological Significance*. Birkhauser Publishers, Basel, pp. 40–108.
Pause,A. and Sonenberg,N. (1992) *EMBO J.*, **11**, 2643–2654.
Price,C., Linger,J., Bickle,T.A., Firman,K. and Glover,S.W. (1989) *J. Mol. Biol.*, **205**, 115–125.
Rickenberg,H.V. and Lester,G. (1955) *J. Gen. Microbiol.*, **13**, 279–284.
Saha,S. and Rao,D.N. (1995) *J. Mol. Biol.*, **247**, 559–567.
Sain,B. and Murray,N.E. (1980) *Mol. Gen. Genet.*, **180**, 35–46.
Schmid,S.R. and Linder,P. (1992) *Mol. Microbiol.*, **6**, 283–292.
Sharp,P.M., Kelleher,J.E., Daniel,A.S., Cowan,G.M. and Murray,N.E. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 9836–9840.
Sharples,G.J., Whitby,M.C., Ryder,L. and Lloyd,R.G. (1994) *Nucleic Acids Res.*, **22**, 308–313.
Studier,F.W. and Bandyopadhyay,P.K. (1988) *Proc. Natl Acad. Sci. USA*, **85**, 4677–4681.
Studier,F.W. and Moffatt,B.A. (1986) *J. Mol. Biol.*, **189**, 113–130.
Sung,P., Higgins,D., Prakash,L. and Prakash,S. (1988) *EMBO J.*, **7**, 3263–3269.
Tabor,S. (1990) In Ausubel,F.A., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl,K. (eds), *Current Protocols in Molecular Biology*. Greene Publishing and Wiley-Interscience, New York, pp. 16.2.1–16.2.11.
Umezu,K., Nakayama,K. and Nakayama,H. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 5363–5367.
Weiserova,M., Janscak,P., Benada,O., Zinkevich,V.E., Glover,S.W. and Firman,K. (1993) *Nucleic Acids Res.*, **21**, 373–379.
Whittaker,P.A., Campbell,A.J.B., Southern,E.M. and Murray,N.E. (1988) *Nucleic Acids Res.*, **16**, 6725–6736.
Willcock,D.F., Dryden,D.T.F. and Murray,N.E. (1994) *EMBO J.*, **13**, 3902–3908.
Wilson,G.G. and Murray,N.E. (1991) *Annu. Rev. Genet.*, **25**, 585–627.
Yuan,R., Hamilton,D.L. and Burckhardt,J. (1980) *Cell*, **20**, 237–244.

Received on November 14, 1995; revised on December 14, 1995

Molecular Microbiology (1996) 22(3), 437–447

A third family of allelic *hsd* genes in *Salmonella enterica*: sequence comparisons with related proteins identify conserved regions implicated in restriction of DNA

Annette J. B. Titheradge, Diane Ternent and Noreen E. Murray*

Institute of Cell and Molecular Biology, Darwin Building, King's Buildings, University of Edinburgh, Mayfield Road, Edinburgh, EH9 3JR, UK.

Summary

***Salmonella enterica* serovar *blegdam* has a restriction and modification system encoded by genes linked to *serB*. We have cloned these genes, putative alleles of the *hsd* locus of *Escherichia coli* K-12, and confirmed by the sequence similarities of flanking DNA that the *hsd* genes of *S. enterica* serovar *blegdam* have the same chromosomal location as those of *E. coli* K-12 and *Salmonella enterica* serovar *typhimurium* LT2. There is, however, no obvious similarity in their nucleotide sequences, and while the gene order in *S. enterica* serovar *blegdam* is *serB hsdM*, *S* and *R*, that in *E. coli* K-12 and *S. enterica* serovar *typhimurium* LT2 is *serB hsdR*, *M* and *S*. The *hsd* genes of *S. enterica* serovar *blegdam* identify a third family of *serB*-linked *hsd* genes (type ID). The polypeptide sequence predicted from the three *hsd* genes show some similarities (18–50% identity) with the polypeptides of known and putative type I restriction and modification systems; the highest levels of identity are with sequences of *Haemophilus influenzae* Rd. The HsdM polypeptide has the motifs characteristic of adenine methyltransferases. Comparisons of the HsdR sequence with those for three other families of type I systems and three putative HsdR polypeptides identify two highly conserved regions in addition to the seven proposed DEAD-box motifs.**

Introduction

High levels of intraspecific allelic diversity are unusual but are characteristic of some loci in many, if not all, species. Examples include the MHC class II alleles in mammals (Figueroa *et al.*, 1988; Lawlor *et al.*, 1988), mating-type loci in fungi (Kües and Cassleton, 1992), self-incompatibility loci of some plants (Ioerger *et al.*, 1990), and the surface antigens encoded by the *rfb* gene cluster in *Escherichia coli* and *Salmonella enterica* (see Reeves, 1993). For these systems, selection for variation has resulted in the maintenance of a large number of alleles. This is also true for the chromosomally encoded restriction and modification (RM) genes of *E. coli* (Sharp *et al.*, 1992; Barcus and Murray, 1995). In all these examples, sequence divergence may be sufficiently great to indicate that gene lineages predate speciation. Of the five systems cited above, four participate in the differentiation of 'foreign' from 'self'.

RM systems enable bacteria to distinguish 'foreign' DNA from their own (for a recent review, see Bickle and Krüger, 1993). The modification component of the system monitors the methylation state of the cell's own DNA and methylates specific bases within a recognition sequence, ensuring that newly replicated, hemimethylated DNA will be fully modified. DNA that includes unmodified target sequences will be recognized as foreign and cleaved by the restriction component of the system.

The type I RM enzymes each comprise three subunits, encoded by the *hsdR*, *M* and *S* genes. The HsdS and M subunits together form a DNA methyltransferase that methylates adenine residues, one on each strand within an asymmetric, bipartite recognition sequence. When all three subunits are present, the alternative activities of restriction and modification are dictated by the methylation state of the target sequence; hemimethylated targets are modified, non-methylated targets elicit restriction.

The 15 kb segment of the *E. coli* K-12 chromosome referred to by Raleigh (1992) as the immigration control region (ICR) includes *hsd* genes specifying the type I RM system *Eco*KI and genes encoding the *mrr* and *mcrBC* restriction systems. Alternative *hsd* genes within the ICR were known to confer different restriction specificities to the *E. coli* strains K-12, B and 15T$^-$ (Boyer, 1964; Arber and Wauters-Willems, 1970). Complementation tests indicate that the *hsd* genes of *E. coli* strains B and K-12 are sufficiently similar that their products are exchangeable, and that the subunit encoded by *hsdS* confers sequence

Received 9 July, 1996; revised 15 August, 1996; accepted 21 August, 1996. *For correspondence. E-mail Noreen.Murray@ed.ac.uk; Tel. (0131) 6505374; Fax (0131) 6683870.

specificity to its RM system (Boyer and Roulland-Dussoix, 1969; Arber and Linn, 1969). *Eco*KI and *Eco*BI are members of a family of closely related type I RM systems, currently called type IA. Similarly, allelic *hsd* genes in *Salmonella enterica* serovars *typhimurium* LT2 and *potsdam* encode members of the IA family of RM systems, as shown by genetic tests (Bullas and Colson, 1975) and sequence similarity (Murray *et al.*, 1982).

The first indication that some allelic *hsd* genes might encode sufficiently dissimilar type I RM systems to warrant their separation into a second family (IB) came from hybridization screens of DNA and serological screens of cell extracts from different strains of *E. coli*. It was found, as expected, that the nucleotide sequences of the *hsd* genes for *Eco*KI and *Eco*BI hybridized to each other, and antibodies raised against *Eco*KI reacted with *Eco*BI. In contrast, DNA probes made from the *Eco*KI genes failed to hybridize with the *hsd* genes of *E. coli* 15T$^-$ that encoded *Eco*AI; similarly, antibodies against *Eco*KI did not cross-react with *Eco*AI (Murray *et al.*, 1982). The *hsd* genes in these two strains are of very different nucleotide sequence, but they behave as *serB*-linked alleles in genetic tests (Arber and Wauters-Willems, 1970), and their common chromosomal location is supported by physical evidence based on the nucleotide sequence of a flanking region (Kannan *et al.*, 1989).

The *hsd* genes of 12 natural isolates of *E. coli* were identified recently (Barcus *et al.*, 1995). Data based on the use of flanking probes indicated that these genes, irrespective of whether they were members of the IA or IB family, were located within the ICR. Eight serovars of *S. enterica* including *kaduna* and *blegdam*, in addition to *typhimurium* and *potsdam*, were shown by Bullas *et al.* (1980) to have *serB*-linked *hsd* genes. It has been suggested (Ryu *et al.*, 1988) that some of these might represent new families of type I RM systems. The nucleotide sequence of the *serB*-linked *hsd* genes of *S. enterica* serovar *kaduna* identifies this system as a member of the type IB family (P. H. Thorpe, D. Ternent and N. E. Murray, unpublished). We have determined the nucleotide sequence of the *hsd* genes of *S. enterica* serovar *blegdam*. These *hsd* genes identify the *blegdam*-specific RM system (*Sty*SBLI) as the first member of a third family of chromosomally encoded type I RM systems. This family has been referred to as type ID (Barcus *et al.*, 1995) because the plasmid-encoded systems *Eco*R124I and *Eco*DXXI were previously designated type IC (Firman *et al.*, 1985). As with all known type I systems, two genes are essential for modification and an additional gene is required for restriction. The predicted polypeptide sequence of one of the subunits of *S. enterica* serovar *blegdam* essential for modification includes the motifs characteristic of an adenine methyltransferase and is presumed to be HsdM. The sequence of the subunit required only for restriction includes the DEAD-box motifs characteristic of many ATP-dependent helicases, a feature common to all known type I HsdR subunits (Gorbalenya and Koonin, 1991). The predicted sequences of the Hsd polypeptides of *S. enterica* serovar *blegdam* are more similar to those of the putative Hsd polypeptides of *Haemophilus influenzae* Rd than to those of known families of type I RM enzymes.

Results

Cloning the hsd *region of* S. enterica *serovar* blegdam

An *in vivo* selection procedure has been a reliable method of finding λ*hsd* phages in genomic libraries. Phages including the *hsdM* and *S* genes of *E. coli* K-12 (IA family) were selected on an *E. coli* K-12 strain because the modification enzyme encoded by the *hsd* genes methylates target sequences within the vector, thus conferring protection against restriction by the cognate restriction system, *Eco*KI (Borck *et al.*, 1976). The same procedure was used to recognize λ clones including the genes encoding *Eco*AI (Fuller-Pace *et al.*, 1985).

Two λ libraries were made from the DNA of a derivative of *E. coli* K-12 (LB4037) in which the resident *hsd* genes had been replaced by those from *S. enterica* serovar *blegdam* (Bullas *et al.*, 1980); one library was derived from *Eco*RI fragments, the other from partial digestion products generated by *Sau*3A. The enrichment of these libraries for phages resistant to restriction by the *blegdam* system of strain LB4037 failed to identify λ*hsd* clones. In the absence of an *hsd*-specific probe, a DNA fragment containing DNA from immediately downstream of the *hsdS* gene of *S. enterica* serovar *typhimurium* was used to identify clones with homologous inserts from the chromosome of LB4037. A fragment of DNA isolated from one such phage (λS25, see Fig. 1) hybridized to the DNA of *S. enterica* serovar *blegdam*, but not to that of *typhimurium*, consistent with the anticipated presence of *blegdam*-specific genes. This *blegdam*-specific probe was used to isolate additional phages. These provided the next probe and the process was continued until phages were isolated that included sequences which hybridized to the DNA upstream of the *hsdR* gene of *S. enterica* serovar *typhimurium*. A restriction map (Fig. 1) was derived from the analysis of the inserts of many phages, and probes were used to confirm that fragment sizes deduced from the clones were consistent with those found in the DNA of *S. enterica* serovar *blegdam*. The restriction map defines a segment of '*blegdam*-specific' DNA of ≈12 kb which is expected to include the *hsd* genes of *S. enterica* serovar *blegdam*.

Identification of an hsd *gene in cloned DNA*

In the absence of evidence for the expression of *hsd* genes in any of the λ phages that included '*blegdam*-specific'

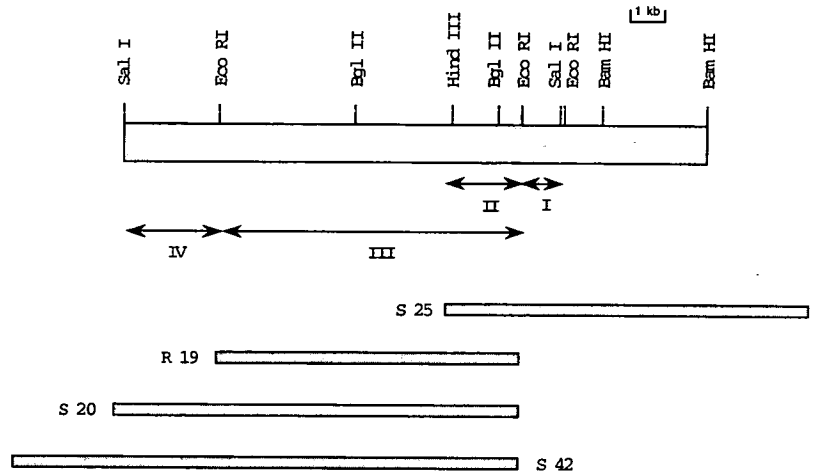


Fig. 1. Restriction map of the *hsd* region of *S. enterica* serovar *blegdam*. The map was determined from the analyses of overlapping λ clones. Roman numerals I–IV identify the probes used to detect overlaps. S20, S25, S42 and R19 identify the inserts of λ clones isolated from partial *Sau*3A (S) or *Eco*RI (R) libraries. Many overlapping clones were analysed but only those mentioned in the text are shown. The first clone isolated (S25) was identified by hybridization to a probe made from pAB3, a plasmid including sequence downstream of *hsdS* in *S. enterica* serovar *typhimurium* LT2; the hybridization detected was confined to the *Bam*HI fragment found only in λS25. A restriction map of the bacterial DNA in λS25 was made, and two fragments were used to make probes I and II. Probe II, but not probe I, detected λR19 in the *Eco*RI library. A probe made from the *Eco*RI fragment in λR19 (probe III) identified λS20 and λS42. Probe III, similar to probe I and II, is specific for *blegdam* DNA. Probe IV from λS20 hybridizes to both *blegdam* and *typhimurium* LT2 DNA, identifying DNA upstream of *hsdR* in *S. enterica* serovar *typhimurium*.

DNA, an *hsd* gene was identified by 'reverse genetics'. A mutation was made within a cloned segment of DNA and transferred to the chromosome of LB4037 to assess its effect on the restriction phenotype of the bacterium. The λ clone including the 8.74 kb *Eco*RI fragment (λR19 in Fig. 1) was used as the substrate for transposon mutagenesis by the mini-transposon Tn*5supF* (Phadnis *et al.*, 1989). A derivative of this phage carrying Tn*5supF* was selected, DNA was prepared and it was shown that the transposon was within the *Eco*RI fragment. The *supF*-tagged sequence was transferred to the bacterial chromosome of LB4037. This was achieved as the consequence of two homology-dependent cross-overs. The first cross-over integrates the prophage into the *hsd* region of LB4037; the second excises it. Bacteria cured of the prophage substitute the segment of '*blegdam*-specific' sequence between the two cross-overs for the one originally present in the host. The lysogens were restriction proficient, but those cured derivatives that acquired *supF* in place of the wild-type sequence were defective in the *blegdam*-specific RM system. We conclude that the *Eco*RI fragment includes at least part of the *hsd* region of the *blegdam* serovar.

The 8.74 kb *Eco*RI fragment of λR19 was transferred to an integration-proficient (*att*$^+$ *int*$^+$) vector (λNM1151) and lysogens were made in an *E. coli* K-12 strain deleted for the ICR region (NM679). These lysogens were always restriction deficient, but some were able to modify λ*vir* with the specificity characteristic of the *blegdam* system. Furthermore, the phages that emerged following induction of a modification-proficient lysogen retained protection against restriction by the *blegdam* system. These results indicate the presence of both the *hsdM* and *S* genes of *S. enterica* serovar *blegdam* within the *Eco*RI fragment, despite the fact that the putative λ*hsdMS* phages fail to modify their DNA on lytic infection. An explanation for the variable phenotypes has not been sought.

The nucleotide sequence of the 8.74 kb Eco*RI fragment and flanking sequences*

The functional tests for *Sty*SBLI indicate that the 8.74 kb *Eco*RI fragment in λR19 specifies the modification enzyme but not the restriction endonuclease. On the assumption that *Sty*SBLI resembles known type I RM systems, two open reading frames (ORFs) within the *Eco*RI fragment should equate with *hsdM* and *hsdS*, and an adjacent, incomplete ORF would be part of *hsdR*.

The nucleotide sequences of both strands of the *Eco*RI fragment in λR19 were determined. The GENEMARK program (Borodovsky and McIninch, 1993) was used to identify ORFs and to indicate possible frame-shifts in the nucleotide sequence. Only those ORFs in excess of 200 codons are identified in Fig. 2.

At present, the smallest gene encoding a component of a type I RM comprises more than 400 codons. Within the

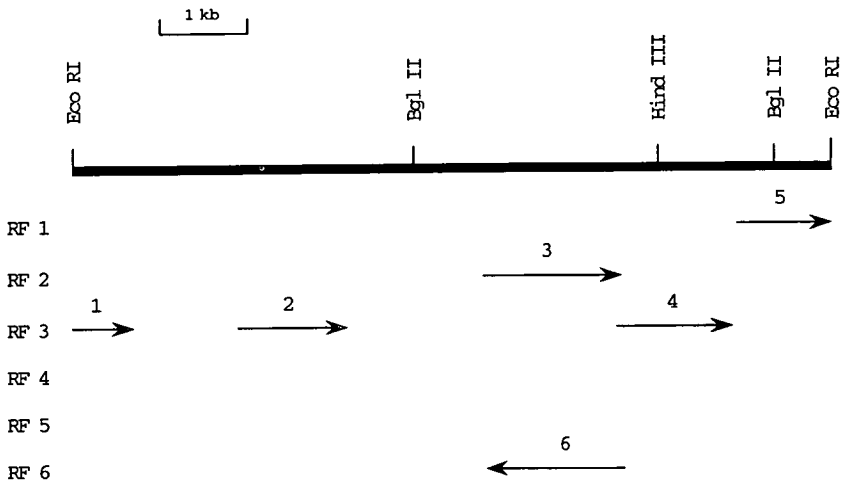


Fig. 2. ORFs within the 8.74 kb *Eco*RI fragment present in λR19 and plasmid pAC18. The nucleotide sequence of the *Eco*RI fragment was translated in all six reading frames (RF1–6). Only ORFs of 200 or more codons are shown, and these are numbered 1 to 6. The arrowheads denote the 3′ ends of the ORFs. The GeneMark program was used to validate the ORFs as coding sequences.

sequences of the *Eco*RI fragment there are 6 ORFs in excess of 200 codons; four are complete (ORF2, 3, 4 and 6; see Fig. 2) and two are incomplete ORFs interrupted by the *Eco*RI targets that define the insert (ORFs 1 and 5 in Fig. 2). The sequence of *orf1* was completed from clones of the adjacent 2.7 kb *Sal*I–*Eco*RI fragment isolated from λS20 and λS42 (see Fig. 1). A polymerase chain reaction (PCR) product generated from the chromosomal DNA of LB4037 provided sequence to join the two contiguous fragments. The 3.1 kb *Bgl*II–*Bam*HI fragment derived from λS25 (see Fig. 1) was used to complete the sequence of *orf5*. The sequence of a 12.4 kb segment of DNA from *S. enterica* serovar *blegdam* has been deposited in the EMBL nucleotide sequence database (Accession No. X99719).

The codon usage of *orf1*, *3*, *4* and *5* is consistent with that of *E. coli* Class I sequences, while those of *orf2* and *orf6* were determined to be Class III, and are probable coding sequences using the less-constrained parameters for transmissible elements (Medigue *et al.*, 1993). The predicted polypeptide sequences deduced from the six ORFs were compared with the sequences in the NCBI database using the BLAST program (Borodovsky *et al.*, 1994). Sequence similarities were assessed and those considered, using the BLAST program, to share significant similarity to sequences in the database were aligned using the PILEUP program (Devereux *et al.*, 1984). All sequences, other than that predicted from *orf6*, showed similarities indicative of putative functions. The sequence predicted from *orf1* included ATP/GTP-binding motifs and shared similarities with many DNA helicases. Those for *orf2* and *orf3* included the motifs characteristic of the S-adenosylmethionine (Ado–Met)-binding domain of adenine methyltransferases and therefore they both have sequences in common with the HsdM polypeptides of type I RM systems. The similarities detected with the ORF4 polypeptide were with a number of proven and putative HsdS subunits. Finally, the polypeptide specified by *orf5* shared DEAD-box motifs with type I HsdR polyeptides. Similar to known HsdR subunits, the polypeptide is large, with a molecular mass of ≈100 kDa.

All type I RM systems characterized to date are encoded by three adjacent genes, and in each case *hsdM* and *S* are not merely adjacent, but overlapping. On this basis, *orf3*, which precedes and overlaps *orf4*, is preferred to *orf2* as the candidate *hsdM* gene. The adjacent *orf5* is anticipated to be *hsdR*.

Genetic evidence for hsdM *and* S

The BLAST program identified *orf2* and *orf3* in the 8.74 kb *Eco*RI fragment (Fig. 2) as candidates for *hsdM*, and *orf4* as a probable *hsdS*. Two approaches were made to define the *hsd* genes by functional tests. One was to subclone a segment from within the *Eco*RI fragment excluding one of the candidate *hsdM* genes; the other was to produce mutations within those ORFs that were considered to be probably *hsdM* and *hsdS*. These manipulations were facilitated by the transfer of the *Eco*RI fragment from λR19 to a plasmid (pBluescript). The resulting plasmid was designated pAC18.

Transformation of an *hsd*Δ strain (NM679) with pAC18 confers proficiency in *blegdam*-specific modification, consistent with the expression of *hsdM* and *S*. The phenotypic test using a plasmid, in contrast to the earlier tests relying on lysogens, gave reproducible results. We suspect that increasing the number of *hsd* genes leads to enhanced levels of the modification enzyme, and a more reliable test for modification.

A second plasmid was made in which the *Bgl*II fragment excised from pAC18 (see Fig. 2) was transferred to pUC19. This plasmid, which includes *orf3* and *orf4*, but

Table 1. Comparisons of predicted amino acid sequences.

Strain	Enzyme/Diagnostic information	Per Cent Identity HsdR	HsdM	HsdS	Reference
E. coli 15T⁻	*Eco*AI	–	–	21.8	Cowan *et al.* (1989)
		–	26.4	–	Sharp *et al.* (1992)
		17.7	–	–	Murray *et al.* (1993)
E. coli K-12	*Eco*KI	–	–	19.5	Gough and Murray (1983)
		–	26.6	–	Loenen *et al.* (1987)
		20.6	–	–	Webb *et al.* (1996)
(plasmid R124)[c]	*Eco*R124I	24.2	28.5	22.5	Price *et al.* (1989)
M. pulmonis	*Mpu*?	24.2	28.8	22.0[a]	Dybvig and Yu (1994)
		–	–	20.2	
K. pneumoniae	*Kpn*BI	25.8	–	–	Valinluck *et al.* (1995)
H. influenzae Rd	HI 1285[b]	33.4	–	–	Fleischmann *et al.* (1995)
	HI 1286	–	–	27.4	
	HI 1287	–	49.8	–	
	HI 0215	–	30.0	–	
	HI 0216	–	–	26.6	

The predicted polypeptide sequences encoded by the *hsd* genes of *S. enterica* serovar *blegdam* were aligned with those for other Hsd and putative Hsd polypeptides using PILEUP. The BESTFIT (Devereux *et al.*, 1984) program was used to determine the per cent identity. PILEUP and BESTFIT are part of the Wisconsin Package, Version 8, of the Genetics Computer Group, Wisconsin, USA.

a. The sequence for *M. pulmonis* identifies two *hsdS* genes.

b. These numbers identify the genes described in Fleischmann *et al.* (1995). All the genes, other than HI 1286, were identified as putative *hsd* genes; the sequences indicate two separated *hsd* regions.

c. Bacteria transformed with plasmid R124.

not *orf2*, resembles pAC18 in its ability to confer *blegdam*-specific modification. We conclude that *orf2* is not necessary for modification.

Two mutations were made in pAC18, one in *orf3*, and the other in *orf4*. Either mutation leads to loss of modification. The mutation in *orf3* was made by cutting the plasmid with *Nde*I and extending the resulting 3′ recessed strands to generate blunt ends prior to ligation. The second mutation was made in a similar way using *Nco*I to interrupt *orf4*. The predicted sequences of the mutations were confirmed.

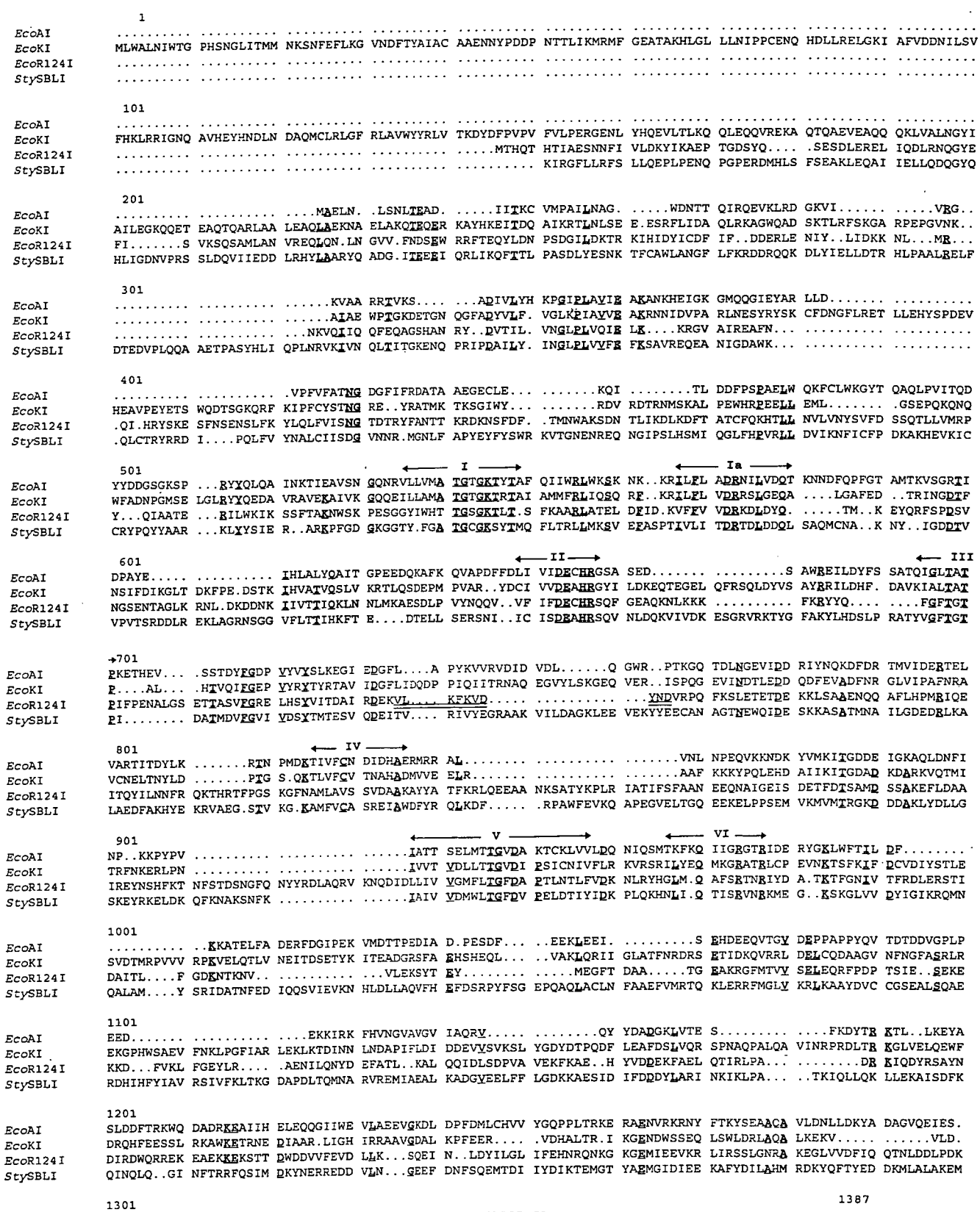
Both mutations resulted in frame-shifts and both were associated with a modification-deficient phenotype. These data support the conclusion that *orf3* and *orf4* are necessary and sufficient for *blegdam*-specific modification, although it remains possible that the phenotype of the *orf3* mutant results from a polar effect on *orf4*. On the basis of the genetic evidence, combined with the similarities detected by the BLAST program, we conclude that *orf3* and *orf4* are *hsdM* and *S*, respectively.

If *orf5* is *hsdR*, the integration, by homologous recombination, of a λ phage including a 1.3 kb *Eco*RI fragment from within *orf5* is predicted to inactivate *hsdR*. Lysogens of the restriction-proficient strain LB4037 would have two incomplete copies of *hsdR*, one lacking a beginning and one lacking the end. A restriction-deficient lysogen made in this way was restored to wild type on excision of the prophage. This genetic evidence supports the identification of *orf5* as *hsdR*.

Analysis of sequences

We will refer to *orf3*, *4* and *5*, the three ORFs essential for *blegdam*-specific restriction and modification, as *hsdM*, *S* and *R*, respectively. These designations are those dictated by sequence comparisons with other families of type I RM enzymes. Our genetic tests identify *hsdR*, but are unable to discriminate between *hsdM* and *hsdS*. The recognition of *hsdS* by complementation tests requires mutants from two members of the same family which confer different specificities. We do, however, have some nucleotide sequence of the *hsd* region of the wild-type strain EcoR9 (our unpublished data); these genes identify *Eco*R9I, the second member of the ID family (Barcus *et al.*, 1995). The putative *hsdS* genes of *Sty*SBLI and *Eco*R9I include two regions of dissimilar sequences, which would correlate with the presence of two target recognition domains. This information reinforces the designation of *hsdS* based on sequence organization within *hsdS* genes from other families.

The sequences of the predicted Hsd polypeptides were aligned with those of their putative homologues using the


EcoAI
EcoKI
EcoR124I
StySBLI
I
Ia
II
III
IV
V
VI

PILEUP program (Devereux *et al.*, 1984). The estimated amino acid identities are given in Table 1. The levels of identity range between 18 and 50%, but show no simple relationship with phylogenetic relatedness of the bacterial species as determined by 16S RNA sequence comparisons. The highest levels of identity detected, 50% for HsdM and 33% for HsdR, were with putative Hsd polypeptides of *H. influenzae* Rd. The order of the genes in *S. enterica* serovar *blegdam* is that found in *H. influenzae* Rd, rather than that of members of the 1A and 1B family in *E. coli* and *S. enterica*. The G+C contents of the *hsd* genes of *S. enterica* serovar *blegdam* are not dissimilar to those of other *E. coli* and *Salmonella* systems (39% for *hsdS* and ≈55% for *hsdR* and *M*). The G+C content for the entire genome of *H. influenzae* Rd is close to 38%; the values of 43% for *hsdR* and 49% for *hsdM* are, therefore, high. These values contrast with a G+C content for *hsdS* of only 35%, but the lower value of 35% for *hsdS* is characteristic of *hsdS* genes (Dila *et al.*, 1990).

The predicted polypeptide sequence specified by *orf2*, the alternative candidate for *hsdM*, showed the highest level of identity (29%) with that of *Mycoplasma pulmonis*. The predicted polypeptide encoded by an ORF of 124 codons immediately preceding *orf3* has regions of significant similarity with the Mrr polypeptide of *E. coli* K-12. However, the Mrr polypeptide is more than 300 codons in length.

The alignment of the predicted polypeptide sequence of HsdR with those of the three other families of type I systems were scanned for the DEAD-box motifs characteristic of helicases, and putative helicases. The comparisons emphasize six of these seven motifs (Figs 3 and 4). The remaining motif, motif IV, is not always conspicuous even in proven DNA helicases, e.g. RecG (Lloyd and Sharples, 1993). Gorbalenya and Koonin (1991) originally pointed out that motif IV was difficult to identify in *Eco*R124I, as was motif III in *Eco*KI. The latter motif was easily found when a short frame-shifted region was identified in *Eco*KI (Burland *et al.*, 1995; Webb *et al.*, 1996). The identification of motif IV remains tentative. The original candidates for motif IV (Gorbalenya and Koonin, 1991) in both *Eco*KI and *Eco*R124I are indicated in Fig. 5. The sequences of *Eco*AI and *Sty*SBLI show some similarities with *Eco*KI in this region. There is, however, an additional conserved region between motifs III and V that is found in all four polypeptides. This sequence in *Eco*KI would not have been considered by Gorbalenya and Koonin (1991) as it lies within the frame-shifted region mentioned above.

In Fig. 4, nine regions of similarity were identified from comparisons of the seven known and putative HsdR sequences. These include the seven DEAD-box motifs identified by Gorbalenya and Koonin (1991) in *Eco*KI and two other conserved regions (X and Y); region X precedes the seven DEAD-box motifs and region Y is between motifs III and IV. The region of the HsdR polypeptide of *Kpn*BI aligned with motif II does not include the anticipated DEXH sequence. Valinluck *et al.* (1995) suggested that motifs II and III are in inverse order in *Kpn*BI, although in the present alignment the sequence they proposed for motif II, DEAV rather than DEXH, falls within region Y (see Fig. 4), a region that is an alternative candidate for motif IV.

The relevance of the DEAD-box motifs, and other conserved sequences, is being tested by mutations. Motifs I, II and III have already been shown to be relevant to restriction by *Eco*KI (Webb *et al.*, 1996).

Discussion

Two families of chromosomally encoded type I RM systems have been described and studied in some detail. Members of the IA family were identified in *E. coli* and *S. enterica* many years ago (for a review, see Barcus and Murray, 1995). Representatives of the type IB family have been identified in *E. coli* and *Citrobacter freundii* and are now known to occur in *S. enterica* (P. H. Thorpe, D. Ternent and N. E. Murray, unpublished). All known IA and IB systems are encoded by genes that have a common chromosomal location identified by the ICR (Barcus *et al.*, 1995). The gene order, where known, is always *hsdS*, *M* and *R* with *hsdR* closest to *serB* (Raleigh *et al.*, 1989; Kannan *et al.*, 1989). The original representatives of the IC family are plasmid borne. In this family, the *hsdM* and *S* genes are co-transcribed from a promoter upstream of *hsdM* as in the other three families, but *hsdR* is located downstream of *hsdS* rather than upstream of *hsdM*.

Our strategy for cloning the *hsd* genes of *S. enterica* serovar *blegdam* depended on the genetic evidence which maps the *hsd* genes to the same chromosomal location as those of the IA family in *S. enterica* serovar *typhimurium* LT2. A probe derived from the chromosomal DNA that includes sequence downstream of the *hsdS* gene of *S. enterica* serovar *typhimurium* LT2 was used to identify λ clones containing *blegdam*-specific DNA. This DNA includes the *hsd* genes encoding *Sty*SBLI. Although these genes share a common location with those of the IA and IB families, they have a different

Fig. 3. Alignment of the HsdR polypeptides of *Eco*AI, *Eco*KI, *Eco*R124I and *Sty*SBLI using the PILEUP Program with a gap weight of 3.0 and a gap-length weight of 0.1. Amino acids conserved in at least three of the four polypeptides are underlined. The sequences identified between arrows indicate the seven putative DEAD-box motifs previously identified in *Eco*KI (Gorbalenya and Koonin, 1991; Webb *et al.*, 1996). An alternative candidate for motif IV in *Eco*R124I (Gorbalenya and Koonin, 1991) is identified by a double underline.

Region X

*Eco*AI	KPGIPLAVIEAK
*Eco*KI	VGLKPIAVVEAK
*Eco*R124I	VNGLPLVQIELK
Mpu?	INGFPLILFEFK
*Sty*SBLI	INGLPLVVFEFK
Hind?	VNGLPLVVFELK
*Kpn*BI	VNGIALGVIELK

Ia

*Eco*AI	ILFLADRNILVDQT
*Eco*KI	ILFLVDRRSLGEQA
*Eco*R124I	VFFVVDRKDLDYQT
Mpu?	VVFLVDRNDLNDQT
*Sty*SBLI	IVLITDRTDLDDQL
Hind?	IVVVTDRNDLDGQL
*Kpn*BI	VLIITDRTELDEQI
Consensus	++++p r +
	k

III

*Eco*AI	IGLTATP
*Eco*KI	IALTATP
*Eco*R124I	FGFTGTP
Mpu?	IGFSGTP
*Sty*SBLI	VGFTGTP
Hind?	IGFTGTP
*Kpn*BI	IGFTGTP
Consensus	+++tat
	sgs

IV

*Eco*AI	TIVFCNDIDHAE
*Eco*KI	TLVFCVTNAHAD
*Eco*R124I	AMLAVSSVDAAK
Mpu?	SIIAFDTIQDAL
*Sty*SBLI	AMFVCASREIAW
Hind?	AMMVVSSRQICV
*Kpn*BI	AMLVCSSVYQAC
Consensus	+++f t
	y s

VI

*Eco*AI	TKFKQIIGRGTR
*Eco*KI	ILYEQMKGRATR
*Eco*R124I	HGLMQAFSRTNR
Mpu?	HNLIQAFSRTNR
*Sty*SBLI	HNLIQTISRVNR
Hind?	HNLMQAIARVNR
*Kpn*BI	HNLFQAICRVNR
Consensus	q +GR+gr
	a

I

*Eco*AI	VLLVMATGTGKTYT
*Eco*KI	ILLAMATGTGKTRT
*Eco*R124I	GYIWHTTGSGKTLT
Mpu?	AYIWHTTGSGKTLT
*Sty*SBLI	GTYFGATGCGKSYT
Hind?	GVMWHTQGSGKSIS
*Kpn*BI	GIIWHTQGSGKSLT
Consensus	+++ tg GKt
	s s

II

*Eco*AI	LIVIDECHRG
*Eco*KI	CIVVDEAHRG
*Eco*R124I	VFIFDECHRS
Mpu?	VFIIDECHRS
*Sty*SBLI	ICISDEAHRS
Hind?	IVISDEAHRS
*Kpn*BI	MKRNIPADFS
Consensus	++++DEah

Region Y

*Eco*AI	DYFGDPVYVYSLKEGIEDG
*Eco*KI	QIFGEPVYRYTYRTAVIDG
*Eco*R124I	SVFGRELHSYVITDAIRDE
Mpu?	KIFGNEIDSYNMKDAILDK
*Sty*SBLI	DVFGVIVDSYTMTESVQDE
Hind?	DVFGRYVSIYDLQDAVEDG
*Kpn*BI	EVFGPYIHTYKFDEAVNDG

V

*Eco*AI	IATTSELMTTGVDAK
*Eco*KI	IVVTVDLLTTGVDIP
*Eco*R124I	LLIVVGMFLTGFDAP
Mpu?	IVIVVDMLLTGFDSP
*Sty*SBLI	IAIVVDMWLTGFDVP
Hind?	VVIVRDMWLTGFDAP
*Kpn*BI	LLIVVDKLLTGFDAP
Consensus	++++t + g+ +
	s

Fig. 4. Alignment of conserved sequences in known and putative HsdR polypeptides. *Mpu*? denotes sequence for *M. pulmonis* and *Hind*?, sequence for *H. influenzae* Rd. The references for the published sequences are given in Table 1. Amino acids conserved in at least four of the seven sequences are shown in bold. The consensus sequences for the seven DEAD-box motifs, identified by Roman numerals, are taken from Gorbalenya and Koonin (1991). Upper-case letters identify conserved amino acids; lower-case letters indicate alternative deviations; and plus signs (+) indicate conserved hydrophobic residues.

gene order; *hsdR* is downstream of *hsdS*, as found in the plasmid-borne IC family (see Fig. 5).

On the basis of sequence similarities, it has been proposed that all type I RM systems have a common origin, even those of the plasmid-borne members of the IC family (Sharp *et al.*, 1992). The simple expectation that *hsd* genes with a common chromosomal location and a common ancestral origin would have the same gene order has not been met. The hypervariability of the nucleotide sequence of chromosomal *hsd* genes in *E. coli* and *Salmonella* has been used as evidence for horizontal transfer of *hsd* genes. The divergence in gene order present for *S. enterica* serovar *blegdam* is again indicative of this. Plasmid-mediated transfer of *hsd* genes has been suggested, but the physical basis for the conserved chromosomal location is not obvious. The divergence of gene order may suggest the evolutionary establishment of a modification system prior to the acquisition of *hsdR* and the evolution of the restriction complex.

The three known families of chromosomally encoded type I RM systems have members in both *E. coli* and *S. enterica* (Barcus *et al.*, 1995; Thorpe *et al.*, unpublished). It is not improbable that other families of enzymes will have representatives in a variety of species. Possible candidates have been identified in *Klebsiella pneumoniae* (Daniel *et al.*, 1988; Valinluck *et al.*, 1995) and in *Bacillus subtilis*

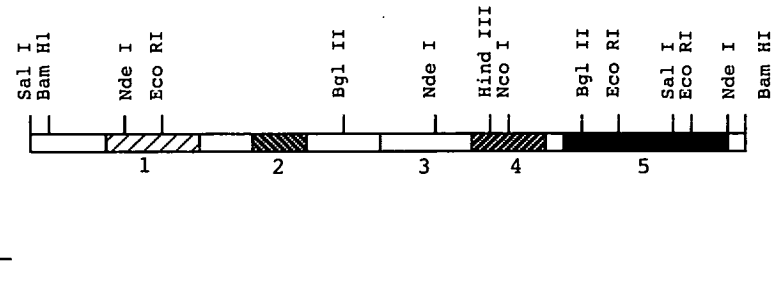


Fig. 5. Map of the *hsd* region of *S. enterica* serovar *blegdam*. The numbers below the map identify the ORFs; *orf3* (560 codons) is *hsdM*, *orf4* (452 codons) is *hsdS* and *orf5* (1114 codons) is *hsdR*. *orf3* and *orf4* overlap by 61 bp, the first methionine codon of *orf4* overlapping the penultimate codon of *orf3*. The orientation of the sequence with respect to *serB* is indicated by the arrow below the map. The order differs from that in *E. coli* K-12 and *S. enterica* serovar *typhimurium*, where *hsdR* is the closest *hsd* gene to *serB*.

(Xu *et al.*, 1995). A reliable indication of the diversity of type I RM systems encoded by allelic genes awaits the analysis of more strains of *E. coli* and other bacteria.

Experimental procedures

Bacteria, phages and plasmids

LB4037 (Bullas *et al.*, 1980), an *E. coli* K-12 derivative in which the *mrr hsd mcr* region of the chromosome has been replaced with the *hsd* region of *S. enterica* serovar *blegdam*, was used as the source of *blegdam*-specific *hsd* genes and as a λ-sensitive strain proficient in *blegdam*-specific restriction. NM679 (King and Murray, 1995), a (*mrr hsd mcrBC*) Δ*mcrA* Δ derivative of W3110 was used as a restriction- and modification-deficient derivative of *E. coli* K-12. DB4496 (Phadnis *et al.*, 1989) was used as a source of Tn*5supF* for mutagenesis. The strain is *dam*$^-$ and the transposable element is maintained on plasmid pBRG1310. DK21 (Kurnit and Seed, 1990), a *sup*° strain with an amber mutation in *dnaB*, but lysogenic for λ*imm*^{21}P1*ban*, was used to select λ derivatives that had incorporated Tn*5supF*. ED8654 (Borck *et al.*, 1976) was used as a standard λ- sensitive *hsdR* strain.

Libraries of DNA fragments generated by *Eco*RI were made in λNM1048, a derivative of λNM781 including the *hsd* genes of *E. coli* K-12 (Sain and Murray, 1980). This phage is deleted for *att int* and *red*, but is *c*I857; libraries were recovered on an *hsd* Δ strain, NM555 (Fuller-Pace *et al.*, 1985). λNM1151 (Murray, 1983) was used as an integration-proficient vector for *Eco*RI fragments. λNM461 (phage IV in Murray and Murray, 1974) was used as an integration-deficient vector for *Eco*RI fragments. Libraries of DNA fragments generated by partial digestion with *Sau*3A were recovered in EMBL3*cos* (Whittaker *et al.*, 1988). These *att*$^-$ *int*$^-$ *red*$^-$ *gam*$^-$ *c*I$^-$ phages were recovered on an *hsdR recD* strain NM621 (Whittaker *et al.*, 1988).

λ*vir*, either unmodified (λ*vir*.0) following propagation on NM679, or modified by the *blegdam*-specific system by growth on LB4037, was used to measure modification and restriction.

Lysogens with the immunity of phage λ, which were the result of homology-dependent integration, were selected using λ*c*I26 and a hybrid phage *h*^{82}*imm*$^{\lambda}$*c*I26.

M13mp18 and 19, pBR322, pUC19, pACYC184 and pBluescript were used as vectors for the subcloning of DNA fragments (see Sambrook *et al.*, 1989). pAB3, the plasmid used to provide a probe for a sequence close to the *hsd* region in *S. enterica* serovar *blegdam*, is a derivative of pBR322 that includes a 4.5 kb *Bam*HI fragment from *S. enterica* serovar *typhimurium* LT2. This fragment begins in the 3′ end of *hsdM*, and includes *hsdS* and ≈2 kb of the downstream sequence (Fuller-Pace and Murray, 1986).

Media and microbial techniques

Media and general methods (Murray *et al.*, 1977) and tests for estimating restriction and modification (Fuller-Pace *et al.*, 1985) have been described.

Enzymes and reagents

Restriction enzymes, T4 DNA ligase, deoxynucleoside triphosphates and Klenow polymerase were purchased from Boehringer Mannheim. *Sau*3A was obtained from Northumbria Biologicals Ltd., and HK phosphatase was from Epicentre Technologies. The Red Hot Taq polymerase from Advanced Biotechnologies Ltd. was used to amplify DNA. The products were purified using the Qiaquick PCR Purification Kit (Qiagen). Sequenase Version 2.0 DNA Sequencing Kit was supplied by Amersham, and the ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit was from Perkin Elmer. λ packaging extracts were obtained from Promega.

Preparation, analysis and ligation of DNA

Bacterial DNA was extracted using the miniprep procedure of Redfield and Campbell (1987).

Phage λ DNA was prepared according to the following miniprep method. A clarified phage lysate (4 ml) was treated with nucleases (5 μl 10 mg ml^{-1} RNase and DNase) before the phages were precipitated overnight at 4°C following the addition of an equal volume of polyethylene glycol (PEG) solution (20% w/v PEG 6000 and 2 M NaCl in phage buffer). The phage particles were sedimented by centrifugation (13 000 × *g* for 20 min at 4°C). The pellet was drained of PEG solution and any remaining PEG was removed with tissue before the pellet was resuspended in 500 μl of phage buffer. Chloroform was added (500 μl) to remove remaining PEG and debris, the mixture was clarified by centrifugation (12 000 × *g* for 1 min), and the aqueous phase transferred to a tube containing an equal volume of Tris-equilibrated phenol to which 100 μl of TE buffer was added. The DNA was extracted and then re-extracted, first by a 1:1 mixture of phenol and chloroform and then by

chloroform. The DNA in 400 μl of aqueous phase was precipitated during incubation at 0°C following the addition of 800 μl of ethanol. The precipitate was sedimented by centrifugation (10 min at 10 000 × *g*), and the supernatant removed. The DNA pellet was washed with 70% ethanol, dissolved in 400 μl of TE buffer, then reprecipitated with ethanol (800 μl) in the presence of 40 μl of 3M NaAc (pH 5.4). The precipitate was air dried before dissolving in 100 μl of TE. The yields were 250–500 μg.

λ vector DNA was isolated from phages that had been purified in a CsCl gradient. Restriction-endonuclease digestion was performed using the buffers and conditions recommended by the supplier. Ligations were carried out by standard methods (Sambrook *et al.*, 1989).

Detection of DNA sequences by hybridization

The reagents and methods were as described by Barcus *et al.* (1995).

Nucleotide sequence determination

Sequence was determined from single-stranded M13 templates and double-stranded plasmids. Initially, sequence was determined by the dideoxy chain-termination method (Sanger *et al.*, 1977) using deoxyadenosine 5-[α^{35}S]-thiotriphosphate and T7 DNA polymerase (Tabor and Richardson, 1987). An Applied Biosystems automated sequencer (ABI 377) was used with the Dye Terminator Cycle Sequencing method. Plasmid DNA for sequence determination was prepared using the Biotech Flexiprep kit (Sephaglas; Pharmacia) or spin columns (silica-gel membrane; Qiagen).

Oligonucleotide primers were obtained from Oswel and Perkin Elmer. The sequence was always determined on both strands. The sequences of DNA fragments were compiled by Staden (1982) programs, and the ABI Sequence Navigator (Perkin Elmer).

Acknowledgements

We wish to thank our colleagues for their helpful discussions and Josie Priestley for her expert assistance in the preparation of this manuscript. The work was supported by the Medical Research Council.

References

Arber, W., and Linn, S. (1969) DNA modification and restriction. *Annu Rev Biochem* **38:** 467–500.

Arber, W., and Wauters-Willems, D. (1970) Host specificity of DNA produced by *Escherichia coli* XII. The two restriction and modification systems of strain 15T$^-$. *Mol Gen Genet* **108:** 203–217.

Barcus, V.A., and Murray, N.E. (1995) Barriers to recombination: restriction. In *Population Genetics of Bacteria.* Baumberg, S., Young, J.P.W., Saunders, S.R., and Wellington, E.M.H. (eds). Cambridge, UK: Cambridge University Press, pp. 31–58.

Barcus, V.A., Titheradge, A.J.B., and Murray, N.E. (1995) The diversity of alleles at the *hsd* locus in natural populations of *Escherichia coli. Genetics* **140:** 1187–1197.

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

Borck, K., Beggs, J.D., Brammar, W.J., Hopkins, A.S., and Murray, N.E. (1976) The construction *in vitro* of transducing derivatives of phage lambda. *Mol Gen Genet* **146:** 199–207.

Borodovsky, M., and McIninch, J.D. (1993) GeneMark: parallel gene recognition for both DNA strands. *Comput Chem* **17** (2)**:** 123–133.

Borodovsky, M., Rudd, K.E., and Koonin, E.V. (1994) Intrinsic and extrinsic approaches for detecting genes in a bacterial genome. *Nucl Acids Res* **22:** 4756–4767.

Boyer, H.W. (1964) Genetic control of restriction and modification. *J Bacteriol* **88:** 1652–1660.

Boyer, H.W., and Roullard-Dussoix, D. (1969) A complementation analysis of the restriction and modification of DNA in *Escherichia coli. J Mol Biol* **41:** 459–472.

Bullas, L.R., and Colson, C. (1975) DNA restriction and modification systems in *Salmonella*. III. SP, a *Salmonella potsdam* system allelic to the SB system in *Salmonella typhimurium. Mol Gen Genet* **139:** 177–188.

Bullas, L.R., Colson, C., and Neufeld, B. (1980) Deoxyribonucleic acid restriction and modification systems in *Salmonella*: chromosomally located systems of different serotypes. *J Bacteriol* **141:** 275–292.

Burland, V., Plunkett, III, G., Sofia, H.J., Daniels, D.L., and Blattner, F.R. (1995) Analysis of the *Escherichia coli* genome VI: DNA sequence of the region from 92.8 through 100 min. *Nucl Acids Res* **23:** 2105–2119.

Cowan, G.M., Gann, A.A.F., and Murray, N.E. (1989) Conservation of complex DNA recognition domains between families of restriction enzymes. *Cell* **56:** 103–109.

Daniel, A.S., Fuller-Pace, F.V., Legge, D.M., and Murray, N.E. (1988) Distribution and diversity of *hsd* genes in *Escherichia coli* and other enteric bacteria. *J Bacteriol* **170:** 1775–1782.

Devereux, J., Haeberli, P., and Smithies, O. (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucl Acids Res* **12:** 387–395.

Dila, D., Sutherland, E., Moran, L., Slatko, B., and Raleigh, E.A. (1990) Genetic and sequence organisation of the *mcrBC* locus of *Escherichia coli* K-1. *Bacteriology* **172:** 4888–4900.

Dybvig, K., and Yu, H. (1994) Regulation of a restriction and modification system via DNA inversion in *Mycoplasma pulmonis. Mol Microbiol* **12:** 547–560.

Figueroa, F., Günther, E., and Klein, J. (1988) MHC polymorphisms predating speciation. *Nature* **335:** 265–267.

Firman, K., Price, C., and Glover, S.W. (1985) The *Eco*R124I and *Eco*R124/3 restriction and modification systems: cloning the genes. *Plasmid* **14:** 224–234.

Fleischmann, R.D., *et al.* (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269:** 496–512.

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., 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.

Gorbalenya, A.E., and Koonin, E.V. (1991) Endonuclease (R) subunits of type-I and type-III restriction-modification enzymes contain a helicase-like domain. *FEBS Lett* **291:** 277–281.

Gough, J.A., and Murray, N.E. (1983) Sequence diversity among related genes for recognition of specific targets in DNA molecules. *J Mol Biol* **166:** 1–19.

Ioerger, T.R., Clark, A.G., and Kao, T.-H. (1990) Polymorphisms at the self-incompatibility locus in *Solanaceae* predates speciation. *Proc Natl Acad Sci USA* **87:** 9732–9735.

Kannan, P., Cowan, G.M., Daniel, A.S., Gann, A.F., and Murray, N.E. (1989) Conservation of organization in the specificity polypeptides of two families of type I restriction enzymes. *J Mol Biol* **209:** 335–344.

King, G., and Murray, N.E. (1995) Restriction alleviation and modification enhancement by the Rac prophage of *Escherichia coli* K-12. *Mol Microbiol* **16:** 769–777.

Kües, U., and Cassleton, L.A. (1992) Fungal mating type genes – regulators of sexual development. *Mycol Res* **96:** 993–1006.

Kurnit, D.M., and Seed, B. (1990) Improved genetic selection for screening bacteriophage libraries by homologous recombination *in vivo*. *Proc Natl Acad Sci USA* **87:** 3166–3169.

Lawlor, D.A., Ward, F.E., Ennis, P.D., Jackson, A.P., and Parham, P. (1988) HLA-A-B polymorphisms predate the divergence of humans and chimpanzees. *Nature* **335:** 265–267.

Lloyd, R.G., and Sharples, G.J. (1993) Processing of recombination intermediates by the RecG and RuvAB proteins of *Escherichia coli*. *Nucl Acids Res* **21:** 1719–1725.

Loenen, W.A.M., Daniel, A.S., Braymer, H.D., and Murray, N.E. (1987) Organization and sequence of the *hsd* genes of *Escherichia coli* K-12. *J Mol Biol* **198:** 159–170.

Medigue, C., Viari, A., Hénaut, A., and Danchin, A. (1993) Colibri: a functional data base for the *Escherichia coli* genome. *Microbiol Rev* **57:** 623–654.

Murray, N.E. (1983) Phage lambda and molecular cloning. In *Lambda II*. Hendrix, R.W., Roberts, J.W., Stahl, F.W., and Weisberg, R.A. (eds). Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press, pp. 395–432.

Murray, N.E., and Murray, K. (1974) Manipulation of restriction targets in phage λ to form receptor chromosomes for DNA fragments. *Nature* **251:** 476–481.

Murray, N.E., Brammar, W.J., and Murray, K. (1977) Lambdoid phages that simplify the recovery of *in vitro* recombinants. *Mol Gen Genet* **150:** 53–61.

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

Murray, N.E., Daniel, A.S., Cowan, G.M., and Sharp, P.M. (1993) Conservation of motifs within the unusually variable polypeptide sequences of type I restriction and modification enzymes. *Mol Microbiol* **9:** 133–143.

Phadnis, S.H., Huang, H.V., and Berg, D.E. (1989) Tn*5supF*, a 264-base-pair transposon derived from Tn*5* for insertion mutagenesis and sequencing DNAs cloned in phage λ. *Proc Natl Acad Sci USA* **86:** 5908–5912.

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

Raleigh, E.A. (1992) Organization and function of the *mcrBC* genes of *Escherichia coli* K-12. *Mol Microbiol* **6:** 1079–1086.

Raleigh, E.A., Trimarchi, R., and Revel, H. (1989) Genetic and physical mapping of the *mcrA* (*rglA*) and *mcrB* (*rglB*) loci of *Escherichia coli* K-12. *Genetics* **122:** 279–296.

Redfield, R.J., and Campbell, I.A.M. (1987) Structure of cryptic lambda prophages. *J Mol Biol* **198:** 383–404.

Reeves, P. (1993) Evolution of *Salmonella* O antigen variation by interspecific gene transfer on a large scale. *Trends Genet* **9:** 17–22.

Ryu, J.-I., Rajadas, P.T., and Bullas, L.R. (1988) Complementation and hybridization evidence for additional families of type I DNA restriction and modification genes in *Salmonella* serotypes. *J Bacteriol* **170:** 5785–5788.

Sain, B., and Murray, N.E. (1980) The *hsd* (host specificity) genes of *E. coli* K-12. *Mol Gen Genet* **180:** 35–46.

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

Sanger, F., Nicklem, S., and Coulsen, A.R. (1977) DNA sequencing and chain-terminating inhibitors. *Proc Natl Acad Sci USA* **74:** 5463–5465.

Sharp, P.M., Kelleher, J.E., Daniel, A.S., Cowan, G.M., and Murray, N.E. (1992) Roles of selection and recombination in the evolution of type I restriction-modification systems in enterobacteria. *Proc Natl Acad Sci USA* **89:** 9836–9840.

Staden, R. (1982) Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. *Nucl Acids Res* **10:** 4731–4751.

Tabor, S., and Richardson, C.C. (1987) DNA sequence analysis with a modified T7 DNA polymerase. *Proc Natl Acad Sci USA* **84:** 4767–4771.

Valinluck, B., Lee, N.S., and Ryu, J. (1995) A new restriction-modification system, *Kpn*BI, recognized in *Klebsiella pneumoniae*. *Gene* **167:** 59–62.

Webb, J.L., King, G., Ternent, D., Titheradge, A.J.B., and Murray, N.E. (1996) Restriction by *Eco*KI is enhanced by co-operative interactions between target sequences and is dependent on DEAD box motifs. *EMBO J* **15:** 2003–2009.

Whittaker, P.A., Campbell, A.J.B., Southern, E.M., and Murray, N.E. (1988) Enhanced recovery and restriction mapping of DNA fragments cloned in a new λ vector. *Nucl Acids Res* **16:** 6725–6736.

Xu, G., Willert, J., Kapfer, W., and Trautner, T. A. (1995) *Bsu*CI, a type-I restriction-modification system in *Bacillus subtilis*. *Gene* **157:** 59.

Molecular Microbiology (1997) 23(4), 851

Corrigendum

A third family of allelic *hsd* genes in *Salmonella enterica*: sequence comparisons with related proteins identify conserved regions implicated in restriction of DNA

Annette J. B. Titheradge, Diane Ternent & Noreen E. Murray*
Institute of Cell and Molecular Biology, Darwin Building, King's Buildings, University of Edinburgh, Mayfield Road, Edinburgh EH9 3JR, UK.

This article was published in *Mol Microbiol* (1996) **22**(3), 437–447.

In Fig. 3 (alignment of the HsdR polypeptides of *Eco*AI, *Eco*KI, *Eco*R124I and *Sty*SBLI using the PILEUP program), the sequence for *Sty*SBLI is anticipated to initiate at codon 27 if translation begins at the first methionine.

In the legend to Fig. 5, the lengths of the open reading frames are in error. The legend is corrected below.

Fig. 5. Map of the *hsd* region of *S. enterica* serovar *blegdam*. The numbers below the map identify the ORFs; *orf3* (540 codons) is *hsdM*, *orf4* (435 codons) is *hsdS*, and *orf5* (1089 codons) is *hsdR*. The first methionine codon of *orf4* overlaps the penultimate codon of *orf3*. The orientation of the sequence with respect to *serB* is indicated by the arrow below the map. The order differs from that in *E. coli* K-12 and *S. enterica* serovar *typhimurium* where *hsdR* is the closest *hsd* gene to *serB*.

Received 20 November, 1996. *For correspondence. E-mail Noreen.Murray@ed.ac.uk; Tel. (0131) 6505374; Fax (0131) 6683870.

Molecular Microbiology (1998) **28**(1), 25–35

ClpX and ClpP are essential for the efficient acquisition of genes specifying type IA and IB restriction systems

Svetlana Makovets, Annette J. B. Titheradge and Noreen E. Murray*
Institute of Cell and Molecular Biology, Darwin Building, King's Buildings, University of Edinburgh, Mayfield Road, Edinburgh EH9 3JR, UK.

Summary

Efficient acquisition of genes that encode a restriction and modification (R–M) system with specificities different from any already present in the recipient bacterium requires the sequential production of the new modification enzyme followed by the restriction activity in order that the chromosome of the recipient bacterium is protected against attack by the restriction endonuclease. We show that ClpX and ClpP, the components of ClpXP protease, are necessary for the efficient transmission of the genes encoding *Eco*KI and *Eco*AI, representatives of two families of type I R–M systems, thus implicating ClpXP in the modulation of restriction activity. Loss of ClpX imposed a bigger barrier than loss of ClpP, consistent with a dual role for ClpX, possibly as a chaperone and as a component of the ClpXP protease. Transmission of genes specifying *Eco*KI was more dependent on ClpX and ClpP than transmission of the genes for *Eco*AI. Sensitivity to absence of the protease was also influenced by the mode of gene transfer; conjugative transfer and transformation were more dependent on ClpXP than transduction. In the absence of either ClpX or ClpP transfer of the *Eco*KI genes by P1-mediated transduction was impaired, transfer of the *Eco*AI genes was not.

Introduction

There are two regions within the chromosome of *Escherichia coli* where effective recombinational replacement produces hypervariable, or polymorphic, loci (Milkman, 1997). One includes the O-antigen gene complex at 45 min (see Hobbs and Reeves, 1994) and the other includes the restriction and modification (R–M) genes at 98.5 min (see Barcus and Murray, 1995). In *E. coli* and *Salmonella enterica*, the latter region is already known to specify as many as 16 alternative type I R–M systems, each conferring a different sequence specificity. The efficient acquisition of R–M genes conferring specificities different from that of the recipient requires the sequential production of the new modification and restriction activities, and the consequent protection of the recipient chromosome against attack by the restriction endonuclease. Failure to modify the targets in the recipient chromosome, before the production of the restriction enzyme, would result in cell death. This paper identifies two genes of *E. coli* K-12 that are essential for the efficient transmission of chromosomal genes encoding type I R–M systems; the products of these genes may modulate restriction activity.

At least three families of type I R–M systems are encoded by alleles within the hypervariable locus in *E. coli* (Barcus *et al.*, 1995), although members of only two of these (IA and IB) have been well documented. In this paper, the transmission of genes representing both of these families is investigated.

Three genes, *hsd* (for host specificity of DNA) *R*, *M* and *S*, encode the subunits of each type I R–M system. All three subunits combine to make a large oligomeric protein that, in response to the pattern of adenine methylation within specific DNA sequences, functions to maintain the methylation pattern of the resident DNA and to cut foreign DNA that is devoid of methylated bases within the specific target sequences. HsdS confers sequence specificity to both the modification and restriction activities. For *Eco*KI, the R–M system found in *E. coli* K-12, the stoichiometry of the subunits is 2HsdR, 2HsdM and 1HsdS (Dryden *et al.*, 1997), commonly abbreviated to $R_2M_2S_1$, and this complex coexists *in vivo* with a smaller one, M_2S_1, endowed with only modification activity (Dryden *et al.*, 1993). A similar relationship is expected for *Eco*AI.

The *hsd* genes of *E. coli* K-12 have two promoters (Loenen *et al.*, 1987), one (*pmod*) for the transcription of *hsdM* and *S* and the consequent production of the modification enzyme, the other (*pres*) for transcription of *hsdR* to provide HsdR(R), which on association with the modification enzyme generates the restriction endonuclease ($R_2M_2S_1$). Experiments have failed to find evidence that transcriptional control of the *hsd* genes accounts for the sequential production of the modification and restriction activities. Transcription from *pres* is unaffected by the modification phenotype of the bacteria (Loenen *et al.*, 1987) and transcription

Received 16 October, 1997; revised 23 December, 1997; accepted 29 December, 1997. *For correspondence. E-mail Noreen.Murray@ed.ac.uk; Tel. (0131) 6505374; Fax (0131) 6508650.

from *pmod* and *pres* occurs simultaneously after conjugative transfer to a modification-deficient recipient (Prakash-Cheng *et al.*, 1993). Other experiments have shown a lag of many generations before transconjugants become restriction proficient, despite transcription of the *hsd* genes (Prakash-Cheng and Ryu, 1993). Post-translational control, possibly at the level of subunit assembly, could be a critical factor in causing a delay in the production of the restriction enzyme, and Dryden *et al.* (1997) have used their data from an *in vitro* assembly pathway as the basis of such a mechanism. In the assembly pathway they propose, both inactive intermediates in the assembly pathway and the HsdR polypeptide could be susceptible to proteases. A mutation isolated in *E. coli* C, which results in cell death after the acquisition of *hsd* genes (Prakash-Cheng *et al.*, 1993), has identified a gene, *hsdC*, the product of which is a candidate for influencing the assembly pathway of *Eco*KI (Dryden *et al.*, 1997).

In this paper, we screen *E. coli* K-12 mutants deficient in proteases for their ability to acquire R–M genes. We include those mutants currently shown to lack ATP-dependent proteases (*clpP*, *A*, *X*, *Q*, *Y*, *hflA* and *lon*) and a mutant now known to have low amounts of the protease encoded by *hflB*; this gene, also known as *ftsH*, is an essential gene and it is therefore impossible to use null mutations (see Kihara *et al.*, 1997; Shotland *et al.*, 1997). We find that the products of both *clpP* and *clpX* affect the efficient transmission of *hsd* genes by conjugation, transformation and transduction, although they do not affect members of the two families in the same way.

Results

Do proteases affect the introduction of type I R–M genes by conjugation?

The allelic *hsd* genes that encode *Eco*KI and *Eco*AI, the best studied representatives of two families of type I R–M systems, can be transferred by conjugation to recipients lacking appropriate modification of their DNA without any detectable killing of the potentially vulnerable cells. *E. coli* C is a natural isolate of *E. coli* that lacks *hsd* genes and is commonly used as a restriction- and modification-deficient (r^-m^-) strain. A derivative of *E. coli* C (*hsdC*) has been isolated that, unlike the wild type, dies after the receipt of the *hsd* genes encoding either *Eco*KI (Prakash-Cheng *et al.*, 1993) or *Eco*AI (Kulik and Bickle, 1996).

It has been suggested that when *hsd* genes enter a recipient cell proteases could contribute to a post-transcriptional mechanism for ensuring the sequential production of the modification methylase before the restriction endonuclease (Kulik and Bickle, 1996; Dryden *et al.*, 1997; J. Ryu, personal communication). As a test of this idea, a series of $r_K^-m_K^-$ recipients of *E. coli* K-12 was made in which each strain was defective in a structural component of a known protease, and these strains were monitored for the efficiency with which they acquired F′ plasmids carrying functional *hsd* genes. The donors of the F′ plasmids were *recA* to prevent recombination between the F′ and the donor chromosome as recombination could in some cases result in early conjugative transfer of the wild-type allele of the protease gene to the recipient. Pairs of F′ plasmids, one $hsdR^+M^+S^+$ and the other $hsdR^-M^+S^+$, were used for both the *Eco*KI and *Eco*AI systems. The titres of recipients and of conjugants acquiring a selectable plasmid marker (Tc^r or Cm^r, see Tables 1 and 2) were determined after 2.5 h. When the plasmid conferred an r^+ phenotype, the titres of both conjugants and recipients that survived conjugation were severely depressed for two of the eight *E. coli* K-12 mutants, *clpX* and *clpP*, as well as for the *hsdC* strain used as a positive control (see Table 1 for *Eco*KI and Table 2 for *Eco*AI). For both families of systems, therefore, *clpP* and *clpX* are implicated as relevant to the sequential establishment of the modification and restriction activities.

In the case of *clpP*, *clpX* and *hsdC*, in which the mutations imposed a barrier to the acquisition of the drug resistance marker present on the F′, a sample of transconjugants was tested for restriction and modification. Similar checks were made on progeny from the clp^+ and $hsdC^+$ control experiments. For *Eco*KI, r_K^+ derivatives were rare among the survivors of either the *clpX* recipient (2 out of 100) or the *hsdC* strain (0 out of 57). In contrast, half (10 out of 20) of the *clpP* transconjugants were r_K^+ and all the transconjugants tested from the negative control experiments (NM840 and JR300 *gyrA*) were restriction proficient. The presence of the F′ plasmid in the $r_K^-m_K^-$ transconjugants was confirmed by their sensitivity to the male-specific phage, M13, and by their ability to transfer Tc^r by conjugation. The $r_K^-m_K^-$ transconjugants behaved as if they had mutations in the *hsd* genes of the F′ plasmid; they could transfer an F′, but the plasmid no longer conferred an r_K^+ phenotype to the recipient cell. The barrier against the *Eco*AI system was less effective, irrespective of the genotype of the recipient, and most (19 out of 20) of the conjugants displayed the r_A^+ phenotype. Therefore, the titres of Tc^r conjugants (Table 2) closely reflect the titres of r_A^+ conjugants. In contrast, for *clpP*, *clpX* and *hsdC*, the titres of Tc^r conjugants overestimate the titres of r_K^+ conjugants; the corrected figures are given in the penultimate column (Table 1).

The low titre of recipients obtained when *clpX*, *clpP* and *hsdC* strains receive an hsd^+, rather than an $hsdR^-M^+S^+$ plasmid, is consistent with cell death after the premature production of the restriction enzyme (Table 1). The kinetics of the killing effect, following transfer of the hsd_K^+ plasmid to various recipient strains is shown in Fig. 1A. No killing was detected if the F′ plasmid had a defect in *hsdR* (Fig. 1B).

The products of *clpX* and *clpP* together make a complex

Table 1. Conjugative transfer of hsd_K genes to r^-m^- recipients.

Recipient strains	Donor strains: JC9935 F′ $r_K^-m_K^{+a}$		JC9935 F′ $r_K^+m_K^{+a}$			
	Titre of recipients[b] after conjugation (ml^{-1})	Titre of Tc^r (Cm^r) conjugants[b,c] (ml^{-1})	Titre of recipients[b] after confugration (ml^{-1})	Titre of Tc^r (Cm^r) conjugants[b,c] (ml^{-1})	Titre of Tc^r (Cm^r) r^+m^+ conjugants (ml^{-1})	Relative frequency (%) of survival of recipients[d]
NM840 (C600 *Δhsd gyrA*)	2.1×10^8	2.1×10^8	2.1×10^8	2.1×10^8	2.1×10^8	100.0
NM840 *clpP*	1.6×10^8	1.6×10^8	1.8×10^5	7.4×10^4	3.7×10^4	0.11
NM840 *clpX*	1.4×10^8	1.4×10^8	3.5×10^4	4.8×10^3	9.6×10^1	0.03
NM840 *clpA*	1.6×10^8	1.6×10^8	1.6×10^8	1.4×10^8	NT	100.0
NM840 *clpQ*	9.9×10^7	9.8×10^7	9.7×10^7	9.7×10^7	NT	98.0
NM840 *clpY*	6.9×10^7	6.8×10^7	8.5×10^7	8.5×10^7	NT	123.2
NM840 *lon*	5.2×10^7	4.5×10^7	6.7×10^7	4.8×10^7	NT	128.8
NM840 *hflA*	7.1×10^7	3.2×10^7	1.0×10^8	9.5×10^7	NT	140.8
NM840 *hflB*	6.2×10^7	2.5×10^7	6.2×10^7	4.4×10^7	NT	100.0
JR300 *gyrA* (*E. coli* C)	1.5×10^8	1.5×10^8	1.5×10^8	1.5×10^8	1.5×10^8	100.0
NM820 (*E. coli* C *hsdC*)	2.0×10^8	2.0×10^8	2.6×10^4	4.8×10^3	$<8.4\times10^1$	0.01

a. F′ $r_K^-m_K^+$ (F′101-202) and F′ $r_K^+m_K^+$ (F′101-201) are derivatives of F′101-102 and F′101-101, respectively (Prakash-Cheng *et al.*, 1993), with miniTn*5*-Cm insertions (de Lorenzo *et al.*, 1990). The former includes $hsd_KR^-M^+S^+$ the latter $hsd_KR^+M^+S^+$.
b. Nalidixic acid was used to select against the donor, with the exception of NM820 in which kanamycin was used.
c. Tetracycline was used to select for transconjugants. When the recipient was Tc^r (NM840 *lon*, NM840 *hflA* and NM840 *hflB*) chloramphenicol was used.
d. Frequency (%) with which recipient cells survived conjugation with the donor of F′101-201 is expressed relative to that for F′101-202 (the ratio of the titre of recipient cells after conjugation with JC9935 F′101-201 to the titre of recipient cells after conjugation with JC9935 F′101-202 as a percentage).
NT, Conjugants have not been tested for restriction modification.

that functions as a protease, ClpX itself is a chaperone (Levchenko *et al.*, 1995; Gottesman, 1996); either or both activities could be relevant to the delayed production of the restriction enzyme after transcription of the *hsd* genes. The finding that the effect of *clpX* is more severe than that of *clpP* supports a dual role for ClpX.

Does the ClpXP complex affect the acquisition of hsd *genes by transformation?*

Bacteria were made competent in the uptake of DNA and scored for the efficiency of uptake of plasmid DNA carrying *hsd* genes (Table 3). In each series of experiments,

Table 2. Conjugative transfer of hsd_A genes to r^-m^- recipients.

Recipient strains	Donor strains: JC9935 F′ $r_A^-m_A^{+a}$		JC9935 F′ $r_A^+m_A^{+a}$		
	Titre of recipients[b] after conjugation (ml^{-1})	Titre of Tc^r (Cm^r) conjugants[b,c] (ml^{-1})	Titre of recipients[b] after confugration (ml^{-1})	Titre of Tc^r (Cm^r) conjugants[b,c] (ml^{-1})	Relative frequency (%) of survival of recipients[d]
NM840 (C600 *Δhsd gyrA*)	6.0×10^7	5.9×10^7	6.0×10^7	6.0×10^7	100.0
NM840 *clpP*	3.3×10^7	3.1×10^7	6.2×10^5	1.6×10^5	1.9
NM840 *clpX*	5.1×10^7	4.0×10^7	1.3×10^5	2.8×10^4	0.3
NM840 *clpA*	1.4×10^8	1.1×10^8	1.5×10^8	1.0×10^8	107.1
NM840 *clpQ*	9.9×10^7	9.3×10^7	9.7×10^7	8.4×10^7	98.0
NM840 *clpY*	5.5×10^7	3.8×10^7	4.8×10^7	4.8×10^7	87.3
NM840 *lon*	3.8×10^7	3.8×10^7	4.3×10^7	3.2×10^7	113.2
NM840 *hflA*	4.3×10^7	4.2×10^7	4.9×10^7	4.6×10^7	114.0
NM840 *hflB*	5.1×10^7	3.3×10^7	4.2×10^7	3.9×10^7	82.4
JR300 *gyrA* (*E. coli* C)	1.7×10^8	1.7×10^8	1.0×10^8	7.1×10^7	58.8
NM820 (*E. coli* C *hsdC*)	1.8×10^8	1.7×10^8	9.1×10^6	5.5×10^6	5.1

a. F′ $r_A^-m_A^+$ (F′101-302) and F′ $r_A^+m_A^+$ (F′101-301) are derivatives of F′101 (Low, 1972) in which *Eco*KI genes were replaced by $hsd_AR^-M^+S^+$ and $hsd_AR^+M^+S^+$ respectively. Both plasmids have *zjj*::Tn*10* and miniTn*5*-Cm insertions as described in *Experimental procedures*.
b. Nalidixic acid was used to select against the donor, with the exception of NM820 in which kanamycin was used.
c. Tetracycline was used to select for transconjugants. When the recipient was Tc^r (NM840 *lon*, NM840 *hflA* and NM840 *hflB*) chloramphenicol was used.
d. Frequency (%) with which recipient cells survived conjugation with the donor of F′101-301 is expressed relative to that for F′101-302 (the ratio of the titre of recipient cells after conjugation with JC9935 F′101-301 to the titre of recipient cells after conjugation with JC9935 F′101-302 as a percentage).

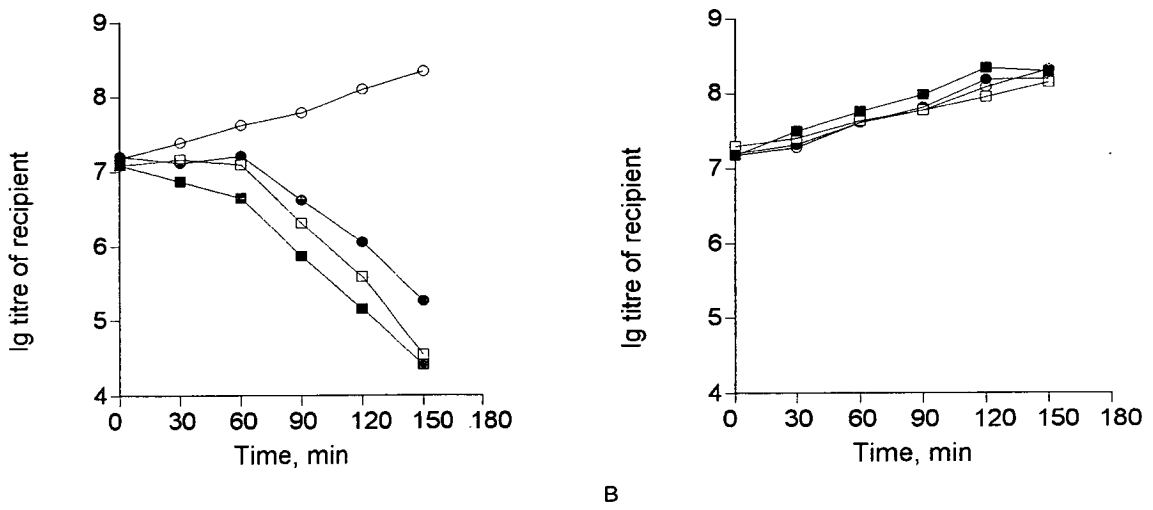


Fig. 1. Conjugative transfer of (A) $hsd_KR^+M^+S^+$ and (B) $hsd_KR^-M^+S^+$ genes to NM840 (○), NM840 *clpP* (●), NM840 *clpX* (□) and NM820 *E. coli* C *hsdC* (■). Cultures of donor (JC9935 F'101-201 and JC9935 F'101-202) and recipient bacteria, grown to mid-log phase, were mixed at a ratio of 10:1 incubated at 37°C and samples were plated at 30 min intervals on media selective for recipients (nalidixic acid for NM840, NM840 *clpP* and NM840 *clpX* and kanamycin for NM820).

a marker plasmid was mixed with the test plasmid. The marker was pBRK, a derivative of pBR322 in which the *bla* gene had been inactivated by the insertion of the *kan* gene from Tn*903* at the *Pst*I site. The test plasmid was a Tc^s Ap^r derivative of pBR322 including *hsd* genes, either $R^+M^+S^+$ or $R^-M^+S^+$. The ratios of Ap^r/Tc^r transformants were monitored; ratios lower than those obtained for clp^+ or $hsdC^+$ bacteria indicated a reduced efficiency in the recovery of the plasmid carrying the *hsd* genes. Significant variations were obtained only for plasmids including $hsdR^+$, M^+ and S^+. The *clpP*, *clpX* and *hsdC* mutations depressed the recovery of hsd^+ plasmids by transformation, although the effect on the plasmid encoding the *Eco*AI system was smaller than that on the one encoding *Eco*KI. For *Eco*KI particularly, the *hsdC* strain had a bigger barrier than either *clpX* or *clpP* derivatives of *E. coli* K-12. No transformant of the *hsdC* strain was obtained for the *hsd* plasmid carrying the *Eco*KI genes, whereas among the Ap^r transformants of *clpX* and *clpP* strains appreciable numbers (20% and 70%, respectively) were r_K^+.

The effect of clpP *and* clpX *on the transfer of* hsd *genes by P1 transduction*

The efficiency of co-transduction of *hsd* with *dnaC*, a closely linked marker, was assessed for $hsdR^+M^+S^+$ (r^+m^+) and $hsdR^-M^+S^+$ (r^-m^+) donors. The *dnaC325* mutation confers a temperature-sensitive lethal phenotype. P1 lysates made on hsd_K and hsd_A donors were used to transduce r^-m^- *dnaC* recipient strains. $DnaC^+$ transductants were selected at 42°C and scored for the acquisition of *hsd* genes and the loss of Tn*10*, a marker for Tc^r on the opposite side of *dnaC* from the *hsd* deletion (Tables 4 and 5). The *hsdC* mutation in *E. coli* C has been shown to prevent the co-transduction of the hsd^+ genes encoding *Eco*DI, like *Eco*KI a member of the type IA family, but not if the donor is $hsdR^-M^+S_D^+$ (O'Neill *et al.*, 1997), and for this reason the experiment using an *hsdC* host was omitted for *Eco*KI. The effect of *clpX* on *Eco*KI (Table 4) resembled that of *hsdC* on the acquisition of *Eco*DI (O'Neill *et al.*, 1997); *clpP* has a smaller effect than either *clpX* or *hsdC*. A very different result was obtained for the type IB

Table 3. The effect of *clpP* and *clpX* on uptake of *hsd* genes by transformation.

Recipient strains	Ratio of transformants with *hsd* plasmids to control plasmid			
	p*hsdRMS*$_K$ (p*hsd*$^+$)	p*hsdMS*$_K$ (p*hsdR*$^-$)	p*hsdRMS*$_A$ (pFFP30)	p*hsdMS*$_A$ (pFFP31)
NM840 (C600 Δ*hsd gyrA*)	3.0	10.9	9.5	6.9
NM840 *clpP*	6.7×10^{-2}	6.6	1.7	5.6
NM840 *clpX*	6.6×10^{-3}	10.3	2.3	7.3
JR300 (*E. coli* C)	2.4	6.8	4.3	5.8
NM820 (*E. coli* C *hsdC*)	$<10^{-4}$ [a]	5.3	6.0×10^{-1}	8.7

a. No transformants isolated.

Table 4. The effect of *clpP* and *clpX* on uptake of hsd_K genes by P1 transduction.

	Donor strains					
	C600 ($hsd_KR^+M^+S^+$ *dnaC*$^+$)			5K ($hsd_KR^-M^+S^+$ *dnaC*$^+$)		
Recipient strains	Sample size	*dnaC*$^+$–*hsd*$^+$ linkage	*dnaC*$^+$–Tn*10*° linkage	Sample size	*dnaC*$^+$–*hsdM*$^+S^+$ linkage	*dnaC*$^+$–Tn*10*° linkage
NK125 (NM840 *dnaC zjj*::Tn*10*)	50	0.32	0.90	50	0.42	1.00
NK125 *clpP*	100	0.09	0.96	100	0.31	0.96
NK125 *clpX*	100	0.00	0.94	50	0.46	0.98

representative; none of the three mutations had a demonstrable effect on the acquisition of the *hsd* genes of *Eco*AI (Table 5). In the tests based on either conjugation or transformation the barrier to the transmission of the *hsd* genes encoding *Eco*AI was weaker than that against *Eco*KI, nevertheless it was demonstrable even for *clpP*.

Do clpP *and* clpX *affect the level of restriction?*

It has been shown that increasing the number of copies of *hsdR* relative to *hsdM* and *S* decreases the efficiency of plating (e.o.p.) of unmodified phage λ (Webb *et al.*, 1996). It seems probable that increasing the number of copies of *hsdR* increases the concentration of the restriction endonuclease ($R_2M_2S_1$) relative to that of the modification methylase (M_2S_1). If the ClpXP protease normally modulates the level of restriction endonuclease, perhaps by attacking HsdR, then in its absence the balance between the levels of the endonuclease and the modification enzyme might be shifted in favour of the endonuclease.

The effect of the *clpP* and *clpX* mutations on the restriction of unmodified phage λ was assessed. Restriction by both *Eco*KI and *Eco*AI was increased in the absence of either ClpX or ClpP (Table 6). This is consistent with an elevated level of the active endonuclease in cells lacking the ClpXP protease. Similar elevated levels of restriction were also detected for *clpP* and *clpX* derivatives of AB1157, a strain of *E. coli* K-12 that already shows high levels of restriction because it lacks a cryptic prophage (Rac) including a gene that alleviates restriction (see Webb *et al.*, 1996). Elevated levels of restriction were not found for $hsd^+{}_K$ derivatives of the *hsdC*$^-$ *E. coli* C (data not shown). Kulik and Bickle (1996) report that they fail to detect a change in restriction by *Eco*AI in an r_A^+ derivative of an *hsdC* derivative of *E. coli* C. This apparent difference between *E. coli* C and *E. coli* K-12 is not explicable at the present time.

Do clpP *and* clpX *affect the loss of* hsd *genes?*

Some type II R–M systems have been shown to stabilize the maintenance of the plasmid that encodes them (Kulakauskas *et al.*, 1995; Naito *et al.*, 1995). It has been argued that when R–M genes are lost and the restriction and modification enzymes cease to be made, the cells eventually lose the capacity to modify the many target sequences within their chromosomes and hence lose the capacity to protect themselves against residual restriction enzyme. Unmodified target sequences will elicit restriction and the cell will be killed. Some R–M systems therefore have been considered as 'plasmid addiction systems' (Naito *et al.*, 1995). *Eco*KI, however, does not behave in this way; genes encoding *Eco*KI are readily lost and replaced (O'Neill *et al.*, 1997). Nevertheless, if ClpXP affects the establishment and severity of the restriction-proficient phenotype, it might be anticipated that it could

Table 5. The effect of *clpP* and *clpX* on uptake of hsd_A genes by P1 transduction.

	Donor strains					
	WA2899 ($hsd_AR^+M^+S^+$ *dnaC*$^+$)			NM863 ($hsd_AR^-M^+S^+$ *dnaC*$^+$)		
Recipient strains	Sample size	*dnaC*$^+$–*hsd*$^+$ linkage	*dnaC*$^+$–Tn*10*° linkage	Sample size	*dnaC*$^+$–*hsdM*$^+S^+$ linkage	*dnaC*$^+$–Tn*10*° linkage
NK125 (NM840 *dnaC zjj*::Tn*10*)	60	0.48	0.83	50	0.54	0.90
NK125 *clpP*	100	0.63	0.88	49	0.51	0.81
NK125 *clpX*	100	0.65	0.80	49	0.63	0.84
NM824 (*E. coli* C *dnaC zjj*::Tn*10*)	48	0.31[a]	0.75	48	0.33[a]	0.83
NM822 (*E. coli* C *hsdC dnaC zjj*::Tn*10*)	50	0.42[a]	0.90	48	0.42[a]	0.90

a. Lower frequencies of co-transduction were observed in similar transduction experiments between *E. coli* K-12 and *E. coli* C (O'Neill *et al.*, 1997).

Table 6. The effect of *clpP* and *clpX* on expression of restriction activity.

Strains	Efficiency of plating (e.o.p.)[a]	
	λ*vir*.0	λ*vir*.K
*Eco*KI system[b]		
C600 *gyrA*	$(1.43 \pm 0.02) \times 10^{-4}$	0.75 ± 0.05
C600 *gyrA clpP*	$(2.03 \pm 0.85) \times 10^{-5}$	0.86 ± 0.29
C600 *gyrA clpX*	$(1.27 \pm 0.24) \times 10^{-5}$	0.95 ± 0.21
*Eco*AI system[c]		
WA2899	$(1.25 \pm 0.11) \times 10^{-2}$	1.00 ± 0.25
WA2899 *clpP*	$(2.89 \pm 0.32) \times 10^{-3}$	0.87 ± 0.08
WA2899 *clpX*	$(1.84 \pm 0.17) \times 10^{-3}$	0.90 ± 0.10

a. The data represented are based on three independent experiments.
b. E.o.p. was calculated relative to NM840 Δ*hsdRM* derivative of C600 *gyrA*.
c. E.o.p. was calculated relative to NM863, $hsd_A R^- M^+ S^+$ derivative of WA2899.

affect the relative stabilities of the restriction enzyme ($R_2M_2S_1$) and the modification component (M_2S_1). If this were so, in the absence of ClpXP the loss of *hsd* genes, like the loss of type II R–M genes, could lead to 'programmed cell death'.

The fate of *clpP* and *clpX* cells after the loss of functional *hsd* genes by P1 transduction was assessed. Donor strains with deletions in *hsd* and a closely linked, selectable marker (*zjj*::Tn*10* or *dnaC*$^+$) were used. Tc^r, or DnaC$^+$, transductants were selected and scored for their restriction phenotype. If ClpXP was needed to prevent cell death, r^-m^- transductants of *clp*$^-$ recipients would be rare. A donor (NK231) with a deletion of *hsdM* and *S* was used for the *Eco*KI system and the frequency of r^-m^- transductants was unaffected by mutations in either *clpP* or *clpX*; the frequencies were 0.24, 0.22 and 0.24, respectively, for *clpP* (NK115), *clpX* (NK116) and *clp*$^+$ (NK31). In the experiment for the *Eco*AI system the donor (NM789) had a deletion in *hsdS*. The frequency of r^-m^- transductants was unaffected by either of the *clp* mutations; the frequencies of co-transduction were 0.26, 0.22 and 0.24 for *clpP* (NK219), *clpX* (NK220) and *clp*$^+$ (NM858) respectively. These data show that the *clp* mutations do not impede the loss of functional R–M genes for either *Eco*KI or *Eco*AI, despite the fact that the deletion in each donor strain left the *hsdR* gene functional. Even in the absence of ClpXP, the residual *Eco*KI and *Eco*AI do not lead to cell death when the coding sequences for the modification enzymes are lost.

Does the hsdC *derivative of* E. coli *C have a defect in* clpX *or* clpP*?*

Complementation tests were carried out using λ phages including wild-type and mutant *clp* genes. Four λ*clp* phages (*clpP*$^+$*X*$^+$, *clpP*$^+$*X*$^-$, *clpP*$^-$*X*$^+$ and *clpP*$^-$*X*$^-$) were used to lysogenize the r^-m^-Clp$^-$ derivatives of *E. coli* K-12 and the *hsdCrecA*$^+$ derivative of *E. coli* C. Integration was dependent on homologous recombination as the phages were *int*$^-$. The lysogens were used as recipients for conjugative transfer of the F′ plasmids that included the *hsd* genes from *E. coli* K–12. The expected complementation of known *clp* mutations (*clpP*$^-$*X*$^+$ + *clpP*$^+$*X*$^-$) was recognized by the efficient acquisition of F′ plasmids encoding functional *Eco*KI. Tests on lysogenic derivatives of the *hsdC* strain were consistent with a defect in *clpX* but not *clpP*; both λ*clpP*$^+$*X*$^+$ and λ*clpP*$^-$*X*$^+$, but neither λ*clpP*$^+$*X*$^-$ nor λ*clpP*$^-$*X*$^-$, complemented the *hsdC* lesion. These results agree with experiments carried out using multicopy plasmids as vectors for the *clp* genes (J. Ryu, personal communication).

Discussion

A sequential production of the modification and restriction enzymes is imperative when the genes encoding these activities are transferred to a recipient that lacks the appropriate protective modification of susceptible target sequences within its genome. For type II systems, in which the modification and restriction activities reside in different enzymes rather than both sharing a common component, there is evidence for the existence of regulatory genes, candidates for the transcriptional control of gene expression (Ives *et al.*, 1992; Tao and Blumenthal, 1992). There is no evidence for transcriptional control of expression of genes for type I and type III systems, regardless of whether they are encoded by plasmid, phage or chromosomal genes. Nevertheless, their transmission is efficient. One exception, where transmission was not detected, is the chromosomally encoded type III system, *Sty*LTI (de Backer and Colson, 1991). Consistent with resistance to transmission, and presumably therefore to change of specificity, no allelic diversity has been detected for the locus including the genes specifying *Sty*LTI. This contrasts with the *hsd* genes encoding the type IA system *Sty*LTIII; these *hsd* genes are alleles of those specifying *Eco*KI and *Eco*AI.

The role of proteases as an alternative mechanism for the control of gene expression is currently becoming well documented in both prokaryotes and eukaryotes. Early lines of evidence for this type of post-transcriptional control were provided from studies of the developmental pathways of phages (see, for example, Gottesman *et al.*, 1981). Recent biochemical experiments have shown that the CII regulatory protein of λ is degraded by the essential host protease FtsH, alias HflB (Kihara *et al.*, 1997; Shotland *et al.*, 1997), and that the Mu repressor is a target for the ClpXP protease (Laachouch *et al.*, 1996). The complex,

hetero-oligomeric type I and type III R–M systems offer specific targets that would enable proteases to elicit temporal control of the two activities. The simplest model would require that one subunit of a complex restriction endonuclease, a subunit essential for restriction but not modification, is the target for proteolysis. The association of this subunit with the component essential for modification – the Mod subunit of a type III enzyme or M_2S_1 complex of a type I system – would protect the sensitive subunit – Res for a type III system and HsdR or R for a type I – from proteolysis.

Redaschi and Bickle (1996) used Western blots to monitor the production of the Res subunit of the type III systems *Eco*P1I and *Eco*P15I. They detected very little Res in the absence of Mod and suggested that the Mod subunit may regulate the amount of Res by protecting it from degradation by proteases. Some mechanism of control for type I systems was anticipated from early observations that the genes encoding the modification component of a type IB system, *Eco*AI, could not be transferred to a recipient containing a plasmid expressing the *hsdR* gene (Fuller-Pace *et al.*, 1985; Suri and Bickle, 1985). More recently, it has been demonstrated that the efficient establishment of the genes encoding a type IA system (Prakash-Cheng *et al.*, 1993) or a type IB system (Kulik and Bickle, 1996) in a new host is dependent on the product of the *hsdC* gene of *E. coli*, whereas the plasmid-encoded genes for a member of the IC family are efficiently transferred even to an *hsdC*$^-$ recipient (Kulik and Bickle, 1996). Our experiments identify both polypeptides of the ClpXP protease as necessary for the efficient acquisition of the genes encoding either *Eco*KI or *Eco*AI. A simple explanation of this result is that the ClpXP protease, or a component of this complex, competes with the methylase (M_2S_1) for interaction with HsdR, thereby delaying the production of the complex with endonuclease activity.

The *clpX* and *P* alleles used in our experiments are well-characterized null mutations. Our data show that ClpX and ClpP are both relevant to the transmission of *hsd* genes for members of both the type IA and IB families of R–M systems but that the *clpX* mutation imposes a stronger barrier than *clpP*, implicating a dual role for ClpX, possibly as a chaperone and as a component of the ClpXP protease. Although efficient transmission of the hsd_K and hsd_A genes to other strains requires ClpX and ClpP, the *hsd* genes for *Eco*KI are more dependent than those for *Eco*AI. Furthermore, the strength of the barrier imposed by *clp* mutations is influenced by the mode of transmission: conjugation and transformation being more sensitive than P1 transduction. The relevance of the mode of gene transfer and the difference between families is accentuated by the P1 transduction experiments. In the case of *Eco*AI, no barrier was detectable even for a *clpX* recipient (Table 5). Our data, particularly for transformation (see Table 3) indicate a difference between the phenotype of the *clpX* mutation in *E. coli* K-12 and the phenotype of the *hsdC* mutation in the *recA*$^+$ derivative of JR302 (NM820). The difference may reflect an additional mutation or the different genetic backgrounds. Genetic analysis of JR302 is hindered by the fact that JR302 is a hybrid of *E. coli* C and *E. coli* K-12.

The hsd_A genes may be more amenable to transfer because the modification component of this system is active on unmethylated target sequences (Suri and Bickle, 1985), whereas that of *Eco*KI has a very strong preference for hemimethylated DNA (Dryden *et al.*, 1993). The relevance of the mode of gene transfer is not known, but it is possible that cells receiving DNA from P1 capsids are less sensitive to restriction, less vulnerable to DNA degradation, or both. It is known that the Dar function of P1 protects the donor DNA from restriction enzymes present in the recipient cell (Iida *et al.*, 1987). The same, or another, function could provide some protection against a restriction activity encoded by donor genes.

The current thinking about ATP-dependent proteases, including ClpXP, is one in which a structural component of the protease has intrinsic chaperone activity and it has been suggested that the initial steps in energy-dependent protein degradation may be similar to those of chaperone-dependent protein folding. Gottesman *et al.* (1997) have proposed a general model for handling misfolded proteins *in vivo* in which either swift refolding of proteins with functional potential is achieved or irreversibly denatured and damaged proteins are rapidly degraded. If this model is applied to type I R–M systems, on the assumption that HsdR subunits are the substrate for the ClpXP protease, a failure to stabilize an inactive polypeptide would lead to loss of restriction potential. In fact, our experiments show that restriction is enhanced in a *clpX* bacterium, consistent with increased levels of functional *Eco*KI ($R_2M_2S_1$). This result implies that ClpX chaperone activity is not essential for endonuclease activity, rather that ClpX may play the opposite role of inducing a change from the functional form of HsdR to one that becomes a substrate for degradation by the ClpXP protease. This role could be one in which HsdR is destabilized, making it a substrate for protease activity, or one in which HsdR is either prevented from interaction with or made unsuitable for interaction with the modification methylase. Association of HsdR with the methylase could block access of ClpX to its target sequence. Precedence for explanations of this sort have been provided recently for the MuA transposase in which target sequences for ClpX and MuA overlap (Levchenko *et al.*, 1997). Although potential target sequences in the HsdR polypeptide can be identified, it is premature to speculate. It remains to be determined whether HsdR itself is a substrate for the ClpXP protease, but the identification of ClpXP as a host factor facilitating the transmission of type

Table 7. Bacterial strains.

Strain	Relevant genotype or phenotype	Source or origin
E. coli K-12		
AB1157	$r_K^+m_K^+$ *rac*-0	De Witt and Adelberg (1962)
C600	$r_K^+m_K^+$	Appleyard (1954)
5K	C600 *hsdR514*	Hubacek and Glover (1970)
JC9935	AB1157 *recA13 sup*^{o}	A. J. Clark
KL719	F′ 101 ($hsd_KR^+M^+S^+$)	Low (1968)
NM477	C600 *ΔhsdMS*	Gough and Murray (1983)
NM654	C600 *ΔhsdRM*	Loenen *et al.* (1987)
NM789	$\Delta hsdS_A$	Thorpe *et al.* (1997)
TPC48	*dnaC325 zjj*::Tn*10*	Masters *et al.* (1989)
WA2552	*hsdR* ($r_A^-m_A^+$)	Arber and Wauters-Willems(1970)
WA2899	C600 $r_K^-m_K^-r_A^+m_A^+$	Fuller-Pace *et al.* (1985)
XL1-Blue	*recA1 endA1 gyrA96* (Nal^r) *thi hsdR17* ($r_K^-m_K^+$) *supE44 relA1 lac*F′ [Tn*10* (Tc^r) *proAB lacI*^{q}*ΔlacZM15*]	Bullock *et al.* (1987)
SG20252	MC4100 *lon-100 zba-3000*::Tn*10*	Trisler and Gottesman (1984)
SG21173	MC4100 *ΔclpA*::*kan*	Gottesman (1990)
SG22007	MC4100 *ΔclpP1*::*cat*	Maurizi *et al.* (1990)
SG22080	MC4100 *ΔclpX1*::*kan*	Gottesman *et al.* (1993)
SG22129	MC4100 *ΔclpP1*::*cat ΔclpX1*::*kan*	S. Gottesman
SG22192	MC4100 *ΔclpQ*::*cat*	W.-F.Wu and S.Gottesman
SG22193	MC4100 *ΔclpY*::*cat*	W.-F.Wu and S.Gottesman
MA156	*hflA150* Tn*10**	M. A. Hoyt
WA8304	*hflA150 hflB29 zgj25*::Tn*10*	Banuett *et al.* (1986)
NK31	C600 *gyrA96*	C600×P1(XL1-Blue)
NM840	C600 *gyrA96 ΔhsdRM* ($r_K^-m_K^-$)	NK31×P1(NM654)
NM858	WA2899 *dnaC325 zjj*::Tn*10*	WA2899×P1(TPC48)
NM863	WA2899 *hsdR*	NM858×P1(WA2552)
NK121	NM840 *ΔclpP1*::*cat*	NM840×P1(SG22007)
NK123	NM840 *ΔclpX1*::*kan*	NM840×P1(SG22080)
NK152	NM840 *lon-100 zba-3000*::Tn*10*	NM840×P1(SG20252)
NK188	NM840 *ΔclpA*::*kan*	NM840×P1(SG21173)
NK190	NM840 *ΔclpQ*::*cat*	NM840×P1(SG22192)
NK191	NM840 *ΔclpY*::*cat*	NM840×P1(SG22193)
NK228	NM840::Tn*10 hflA150**	NM840×P1(MA156)
NK229	NM840 *hflB29 zgj25*::Tn*10*	NM840×P1(WA8304)
NK113	AB1157 *ΔclpP1*::*cat*	AB1157×P1(SG22007)
NK114	AB1157 *ΔclpX1*::*kan*	AB1157×P1(SG22080)
NK115	C600 *gyrA96 ΔclpP1*::*cat*	NK31×P1(SG22007)
NK116	C600 *gyrA96 ΔclpX1*::*kan*	NK31×P1(SG22080)
NK125	NM840 *dnaC325 zjj*::Tn*10*	NM840×P1(TPC48)
NK122	NK125 *ΔclpP1*::*cat*	NK125×P1(SG22007)
NK124	NK125 *ΔclpX1*::*kan*	NK125×P1(SG22080)
NK167	NM840 $hsd_AR^+M^+S^+$ *zjj*::Tn*10*	NK125×P1(WA2899)
NK170	NM840 $hsd_AR^-M^+S^+$ *zjj*::Tn*10*	NK125×P1(NM863)
NK219	NM858 *ΔclpP1*::*cat*	NM858×P1(SG22007)
NK220	NM858 *ΔclpX1*::*kan*	NM858×P1(SG22080)
NK231	NM477 *gyrA96 zjj*::Tn*10*	NM477 *gyrA*×P1(TPC48)
NK233	WA2899 *ΔclpP1*::*cat*	WA2899×P1(SG22007)
NK234	WA2899 *ΔclpX1*::*kan*	WA2899×P1(SG22080)
E. coli C		
JR300	wild type r^{o}m^o	Prakash-Cheng *et al.* (1993)
JR302	JR300 *recA hsdC* Km^r	Prakash-Cheng *et al.* (1993)
NK38	JR300 *gyrA96*	JR300×P1(XL1-Blue)
NM820	JR302 $recA^+$	O'Neill *et al.* (1997)
NM822	NM820 $hsd_KR^-M^+S^+$ *dnaC325 zjj*::Tn*10 hsdC*	O'Neill *et al.* (1997)
NM824	JR300 $hsd_KR^-M^+S^+$ *dnaC325 zjj*::Tn*10*	O'Neill *et al.* (1997)

* Where Tn*10* is linked to *hflA150*.

I R–M genes enables investigation of the role of ClpX and ClpP in the sequential control of the modification and restriction activities at a molecular level.

Experimental procedures

Bacterial strains

These are listed in Table 7.

Plasmids

F'101-201 and F'101-202, respectively, are hsd_K^+ and $hsdR^-$ Cm^r derivatives of F'101-101 and F'101-102 (Prakash-Cheng *et al.*, 1993). Chloramphenicol resistance, conferred by miniTn*5*-Cm, was transposed from the mobilizable vector pUT/Cm as described by de Lorenzo *et al.* (1990).

F' plasmids with hsd_A^+ genes (F'101-301) and an $hsdR^-$ derivative (F'101-302) were isolated following the conjugative transfer of F'101 (Low, 1972) to NK167 ($r_A^+m_A^+$) and NK170 ($r_A^-m_A^+$) respectively, recipient strains with a Tc^r marker (*zjj*::Tn*10*) linked to the chromosomal hsd_A genes. F'101 includes a chromosomal segment of the *E. coli* K-12 genome that spans *hsd* and *zjj*::Tn*10*. Recombinant derivatives in which the hsd_K genes on the F' had been replaced by the hsd_A genes were detected by transfer of the F' plasmids to the *recA* strain JC9935. Tc^r Sm^r conjugants were tested for restriction and modification by the *Eco*AI system and for their ability to co-transfer *tet* and the hsd_A genes. MiniTn*5*-Cm derivatives of the new plasmids were selected as described above.

Other plasmids were derivatives of pBR322. One pair ($phsd^+$ and $phsdR^-$) includes the *hsd* genes of *E. coli* K-12 differing only by a missense mutation in *hsdR* (O'Neill *et al.*, 1997). Similarly, pFFP30 includes the three functional *hsd* genes encoding *Eco*AI, and pFFP31 lacks the *hsdR* gene (Fuller-Pace *et al.*, 1985). pBRK has a *kan* gene inserted at the *Pst*I site in the *bla* gene, thereby conferring an Ap^sKm^r phenotype. pWPC16 includes $clpX^+$ and *clpP*::*cat* (Maurizi *et al.*, 1990).

Phages

λ*vir* was used as a test phage for restriction and modification systems. Other phages were the following *clp* derivatives of the λ vector NM1151 (Murray, 1983). The $\lambda clpP^+X^+$ (λNM1357) includes a 6.2kb *Bam*HI fragment from the Kohara phage 148 (Kohara *et al.*, 1987), a *clpX*::*kan* derivative (λNM1361) has the homologous *Bam*HI fragment from NM840*clpX*, and a *clpP*::*cat* derivative (λNM1359) was made by inserting the *Hin*dIII–*Bam*HI fragment from pWPC16 (Maurizi *et al.*, 1990); this $\lambda clpP^-X^+$ phage, like pWPC16, has a deletion that extends from within the *tig* gene, upstream of *clpP* into *bolA*, leaving *clpX* as the only functional gene within the cloned DNA fragment. λ*clpP*::*cat clpX*::*kan* (λNM1362) was made *in vivo* by excision of the prophage from the *clpP*::*cat clpX*::*kan* double mutant (SG22129) lysogenic for λ*clpP*::*cat* (λNM1359).

Microbial methods

Media and general methods have been described (Murray *et al.*, 1977). P1 transductions were carried out according to Miller (1992) and conjugation experiments followed the procedures described by Prakash-Cheng *et al.* (1993). Cells were made competent in the uptake of DNA by electroporation using a Gene Pulser (Bio-Rad). Some lysogens could be selected as Cm^r or Km^r colonies; if not they were selected as immune colonies at 32°C, using λ*b2* $imm^{21}cI^-$ and h^{82} *b522* imm^{21} cI^- phages for selection.

Acknowledgements

We thank Mary O'Neill and David Dryden for helpful comments and Natalie Honeyman for her patient preparation of the manuscript. We are indebted to Susan Gottesman, Ariane Toussaint and David Leach for providing strains and the Medical Research Council for continued support. Svetlana Makovets thanks the Darwin Trust for the award of a scholarship.

References

Appleyard, R.K. (1954) Segregation of lambda lysogenicity during bacterial recombination in *Escherichia coli* K-12. *Genetics* **39:** 440–452.

Arber, W., and Wauters-Willems, D. (1970) Host specificity of DNA produced by *Escherichia coli*. XII. The two restriction and modification systems of strain 15T. *Mol Gen Genet* **108:** 203–217.

Banuett, F., Hoyt, M.A., McFarlane, L., Echols, H., and Herskovitz, I. (1986) *hflB*, a new *Escherichia coli* locus regulating lysogeny and the level of bacteriophage lambda cII protein. *J Mol Biol* **187:** 213–224.

Barcus, V.A., and Murray, N.E. (1995) Barriers to recombination: restriction. In *Population Genetics of Bacteria*. Cambridge, UK: Cambridge University Press, pp. 31–58.

Barcus, V.A., Titheradge, A.J.B., and Murray, N.E. (1995) The diversity of alleles at the *hsd* locus in natural populations of *Escherichia coli*. *Genetics* **140:** 1187–1197.

Bullock, W.O., Fernandez, J.M., and Short, J.M. (1987) XLI-Blue: A high efficiency plasmid transforming *recA Escherichia coli* strain with beta-galactosidase selection. *BioTechniques* **5:** 376–379.

De Backer, D., and Colson, C. (1991) Transfer of genes for the *Sty*LTI restriction-modification system of *Salmonella typhimurium* to strains lacking modification ability results in death of the recipient cells and degradation of their DNA. *J Bacteriol* **173:** 1328–1330.

De Lorenzo, V., Herrero, M., Jakubzik, U., and Timmis, K.N. (1990) Mini-Tn*5* transposon derivatives for insertion mutagenesis, promoter probing, and chromosomal insertion of cloned DNA in gram-negative eubacteria. *J Bacteriol* **172:** 6568–6572.

De Witt, S.K., and Adelberg, E.A. (1962) The occurrence of a genetic transposition in a strain of *Escherichia coli*. *Genetics* **47:** 577.

Dryden, D.T.F., Cooper, L.P., and Murray, N.E. (1993) Purification and characterisation of the methyltransferase from the type I restriction and modification system of *Escherichia coli* K12. *J Biol Chem* **268:** 13228–13236.

Dryden, D.T.F., Cooper, L.P., Thorpe, P.H., and Byron, O. (1997) The *in vitro* assembly of the *Eco*KI type I DNA restriction/modification enzyme and its *in vivo* implications. *Biochemistry* **5:** 1065–1076.

Fuller-Pace, F.V., Cowan, J.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.

Gottesman, S. (1990) Minimizing proteolysis in *Escherichia coli*: genetic solutions. *Methods Enzymol* **185:** 119–129.

Gottesman, S. (1996) Proteases and their targets in *Escherichia coli*. *Annu Rev Genet* **30:** 465–506.

Gottesman, S., Gottesman, M.E., Shaw, J.E., and Pearson, M.L. (1981) Protein degradation in *E. coli*: the *lon* mutation and bacteriophage lambda N and CII protein stability. *Cell* **24:** 225–233.

Gottesman, S., Clark, W.P., de Crecy-Lagard, V., and Maurizi, M.R. (1993) ClpX, an alternative subunit for the ATP-dependent Clp protease of *Escherichia coli*. *J Biol Chem* **268:** 22618–22626.

Gottesman, S., Wickner, S., and Maurizi, M.R. (1997) Protein quality control: triage by chaperones and proteases. *Genes Dev* **11:** 815–823.

Gough, J.A., and Murray, N.E. (1983) Sequence diversity among related genes for recognition of specific targets in DNA molecules. *J Mol Biol* **166:** 1–19.

Hobbs, M., and Reeves, P.R. (1994) The JUMPstart sequence: a 39 bp element common to several polysaccharide gene clusters. *Mol Microbiol* **12:** 855–856.

Hubacek, J., and Glover, S.W. (1970) Complementation analysis of temperature-sensitive host specificity mutations in *Escherichia coli*. *J Mol Biol* **50:** 111–127.

Iida, S., Streigg, M.B., Bickle, T.A., and Arber, W. (1987) Two DNA antirestriction systems of bacteriophage P1, *darA*, and *darB*: characterization of *darA*$^-$ phages. *Virology* **157:** 156–166.

Ives, C.L., Nathan, P.D., and Brooks, J.E. (1992) Regulation of the *Bam*HI restriction modification system by a small intergenic ORF, *bamHIC*, in both *Escherichia coli* and *Bacillus subtilis*. *J Bacteriol* **174:** 7194–7201.

Kihara, A., Akiyama, Y., and Ito, K. (1997) Host regulation of lysogenic decision in bacteriophage λ: transmembrane modulation of FtsH (HflB), the cII degrading protease, by HflKC (HflA). *Proc Natl Acad Sci USA* **94:** 5544–5549.

Kohara, Y., Akiyama, K., and Isono, K. (1987) The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. *Cell* **50:** 495–508.

Kulakauskas, S., Lubys, A., and Ehrlich, S.D. (1995) DNA restriction-modification systems mediate plasmid maintenance. *J Bacteriol* **177:** 3451–3454.

Kulik, E.M., and Bickle, T.A. (1996) Regulation of the activity of the type IC *Eco*R124I restriction enzyme. *J Mol Biol* **264:** 891–906.

Laachouch, J.E., Desmet, L., Geuskens, V., Grimaud, R., and Toussaint, A. (1996) Bacteriophage Mu repressor as a target for the *Escherichia coli* ATP-dependent Clp protease. *EMBO J* **15:** 437–444.

Levchenko, I., Luo, L., and Baker, T.A. (1995) Disassembly of the Mu transposase tetramer by the ClpX chaperone. *Genes Dev* **9:** 2399–2408.

Levchenko, I., Yamauchi, M., and Baker, T.A. (1997) ClpX and MuB interact with overlapping regions of Mu transposase: implications for control of the transposition pathway. *Genes Dev* **11:** 1561–1572.

Loenen, W.A.M., Daniel, A.S., Braymer, H.D., and Murray, N.E. (1987) Organization and sequence of the *hsd* genes of *Escherichia coli K-12*. *J Mol Biol* **198:** 159–170.

Low, K.B. (1968) Formation of merodiploids in matings with a class of Rec$^-$ recipient strains of *Escherichia coli* K12. *Proc Natl Acad Sci USA* **60:** 160–167.

Low, K.B. (1972) *Escherichia coli* K-12 F-prime factors, old and new. *Bacteriol Rev* **36:** 587–607.

Masters, M., Paterson, T., Popplewell, A.G., Owen-Hughes, T., Pringle, J.H., and Begg, K.J. (1989) The effect of DnaA protein levels and the rate of initiation at *oriC* on transcription originating in the *ftsQ* and *ftsA* genes: *in vivo* experiments. *Mol Gen Genet* **216:** 475–483.

Maurizi, M.R., Clark, W.P., Katayama, Y., Rudikoff, S., Pumphrey, J., Bowers, B., and Gottesman, S. (1990) Sequence and structure of ClpP, the proteolytic component of the ATP-dependent Clp protease of *Escherichia coli*. *J Biol Chem* **265:** 12536–12545.

Milkman, R. (1997) Recombination and population structure in *Escherichia coli*. *Genetics* **146:** 745–750.

Miller, J.H. (1992) *A Short Course in Bacterial Genetics: A Laboratory Manual and Handbook for Escherichia coli and Related Bacteria*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Murray, N.E. (1983) Phage lambda and molecular cloning. In *Lambda II*. Hendrix, R.W., Roberts, J.W., Stahl, F.W., and Weisberg, R.A. (eds). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, pp. 395–432.

Murray, N.E., Brammar, W.J., and Murray, K. (1977) Lambdoid phages that simplify the recovery of *in vitro* recombinants. *Mol Gen Genet* **150:** 53–61.

Naito, T., Kusano, K., and Kobayashi, I. (1995) Selfish behaviour of restriction-modification systems. *Science* **267:** 897–899.

O'Neill, M., Chen, A., and Murray, N.E. (1997) The restriction-modification genes of *E. coli* K-12 may not be selfish; they do not resist loss and are readily replaced by alleles conferring different specificities. *Proc Natl Acad Sci USA* **94:** 14596–14601.

Prakash-Cheng, A., and Ryu, J. (1993) Delayed expression of *in vivo* restriction activity following conjugal transfer of *Escherichia coli* hsd_K genes. *J Bacteriol* **175:** 4905–4906.

Prakash-Cheng, A., Chung, S.S., and Ryu, J.-I. (1993) The expression and regulation of *hsdK* genes after conjugative transfer. *Mol Gen Genet* **241:** 491–496.

Redaschi, N., and Bickle, T.A. (1996) Posttranscriptional regulation of *Eco*P1I and *Eco*P15I restriction activity. *J Mol Biol* **257:** 790–803.

Shotland, Y., Koby, S., Teff, D., Mansur, N., Oren, D.A., Tatematsu, K., *et al.* (1997) Proteolysis of the phage (CII regulatory protein by FtsH (HflB) of *Escherichia coli*. *Mol Microbiol* **24:** 1303–1310.

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

Tao, T., and Blumenthal, R.M. (1992) Sequence and

characterization of *pvuIIR*, the *Pvu*II endonuclease gene, and of *pvuIIC*, its regulatory gene. *J Bacteriol* **174:** 3395–3398.

Thorpe, P.H., Ternent, D., and Murray, N.E. (1997) The specificity of *Sty*SKI, a type I restriction enzyme, implies a structure with rotational symmetry. *Nucleic Acids Res* **25:** 1694–1700.

Trisler, P., and Gottesman, S. (1984) *lon* transcriptional regulation of genes necessary for capsular polysaccharide synthesis in *Escherichia coli* K-12. *J Bacteriol* **160:** 184–191.

Webb, J.L., King, G., Ternent, D., Titheradge, A.J.B., and Murray, N.E. (1996) Restriction by *Eco*KI is enhanced by co-operative interactions between target sequences and is dependent on DEAD box motifs. *EMBO J* **15:** 2003–2009.

Nucleic Acids Research, 2001, Vol. 29, No. 20 4195–4205

Families of restriction enzymes: an analysis prompted by molecular and genetic data for type ID restriction and modification systems

Annette J. B. Titheradge, Jonathan King[1], Junichi Ryu[1] and Noreen E. Murray*

Institute of Cell and Molecular Biology, University of Edinburgh, King's Buildings, Edinburgh EH9 3JR, UK and [1]Department of Microbiology and Molecular Genetics, Loma Linda University, Loma Linda, CA 92350, USA

Received June 14, 2001; Revised and Accepted August 17, 2001

DDBJ/EMBL/GenBank accession no. AJ132566

ABSTRACT

Current genetic and molecular evidence places all the known type I restriction and modification systems of *Escherichia coli* and *Salmonella enterica* into one of four discrete families: type IA, IB, IC or ID. *Sty*SBLI is the founder member of the ID family. Similarities of coding sequences have identified restriction systems in *E.coli* and *Klebsiella pneumoniae* as probable members of the type ID family. We present complementation tests that confirm the allocation of *Eco*R9I and *Kpn*AI to the ID family. An alignment of the amino acid sequences of the HsdS subunits of *Sty*SBLI and *Eco*R9I identify two variable regions, each predicted to be a target recognition domain (TRD). Consistent with two TRDs, *Sty*SBLI was shown to recognise a bipartite target sequence, but one in which the adenine residues that are the substrates for methylation are separated by only 6 bp. Implications of family relationships are discussed and evidence is presented that extends the family affiliations identified in enteric bacteria to a wide range of other genera.

INTRODUCTION

In *Escherichia coli* K-12, a 15 kb segment of the bacterial chromosome, referred to as the immigration control region (ICR), specifies three restriction endonucleases: one classical type I restriction and modification (R–M) system and two methylation-dependent restriction enzymes (1). Three genes (*hsdR*, *M* and *S*) encode the classical R–M system (*Eco*KI) first identified by Bertani and Weigle (2). The ICR, however, is hypervariable in *E.coli* and its close relatives (3–5). In different strains of *E.coli*, alternative *hsd* genes specify type I R–M systems with different specificities. *Escherichia coli* K-12 and *E.coli* B, for example, have R–M systems specified by allelic genes and complementation tests showed that the subunits of one system, *Eco*KI, can associate with those of the other, *Eco*BI, to make functional chimaeric enzymes (6,7). *Eco*KI and *Eco*BI differ significantly in only one of their three subunits, HsdS, the subunit that confers sequence specificity to the heterooligomeric complex. *Eco*KI and *Eco*BI each comprise one specificity subunit and two of each of the other subunits, HsdM and HsdR. The alternative activities of the R–M complex are dictated by the methylation state of the target sequence. Unmethylated targets evoke endonuclease activity and hemimethylated targets are the substrates for methylation (8–11).

*Eco*KI and *Eco*BI are the founder members of a family of restriction and modification strains referred to as type IA. This family also includes members from strains of *Salmonella enterica* (12) and a variety of natural isolates of *E.coli* (for reviews see 11,13). The strictest requirement for membership of a family depends on relatedness as demonstrated by complementation tests in which subunits from different enzymes associate to make a functional enzyme. These tests require partial diploids made in bacterial strains sensitive to tester phages and, therefore, have seldom been extended to different genera. More generally applicable tests rely on molecular evidence derived from hybridisation screens of bacterial DNA using *hsd* sequences as probes, and serological screens of cell extracts with antibodies raised against a representative of a known family of enzymes (5,14).

Escherichia coli 15T$^-$ has chromosomal *hsd* genes that behave as alleles of those in *E.coli* K-12 (15) but they share very little sequence similarity as evidenced initially from hybridisation screens (14). The *hsd* genes of *E.coli* 15T$^-$ specify *Eco*AI, the first member of a second family of type I R–M systems (IB) in which different HsdS subunits confer specificities for different nucleotide sequences (16,17). A third family (type IC) was identified by plasmid representatives (8). More recently an R–M system originally identified in the *blegdam* serovar of *S.enterica* (12) was identified as the first member of a new family (type ID) of chromosomally encoded type I R–M systems (18). This system, *Sty*SBLI, is encoded by genes that may be alleles of those for the type IA or IB systems (12). Physical evidence obtained using flanking DNA

*To whom correspondence should be addressed. Tel: +44 131 650 5374; Fax: +44 131 650 8650; Email: noreen.murray@ed.ac.uk
Present address:
Annette J. B. Titheradge, John Hughes Bennett Laboratories, Department of Oncology, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK

sequences as probes identifies a very similar location on the chromosome but no hybridisation was detected between the *hsd* genes of either *E.coli* K-12 or *E.coli* 15T$^-$ and those of *S.enterica* serovar *blegdam* (18).

Currently, molecular tests place all the known type I R–M systems of *E.coli* and *S.enterica* into one of four discrete groups or families; those in type IC may be plasmid-borne and those of type IA, IB and ID are specified by genes linked to *serB* in the bacterial chromosome. The subdivision of type I R–M systems is an empirical one. High levels of identity at the level of nucleotide sequences are indicative of relatively recent divergence and conservation at the level of protein subunits. Any comparison between representatives of two families of systems identifies little sequence similarity even at the level of amino acid sequence; commonly only 20–30% amino acid identity. The known exceptions are two target recognition domains (TRDs) that recognise the same nucleotide sequences (19). To date classifications dependent on molecular probes are consistent with the limited information from complementation tests. In this paper we identify the recognition sequence of *Sty*SBLI and provide evidence based on complementation for the extension of the ID family documented in *Salmonella* to include members from *E.coli* and *Klebsiella pneumoniae*. We then review the status of the family concept as assessed by comparisons of amino acid sequences, an approach more relevant to the present era.

MATERIALS AND METHODS

Bacteria, phages and plasmids

The *hsd* genes specifying *Eco*R9I were cloned from the DNA of ECOR9 (ATCC no. 35328) (20). LB4037 (12), an *E.coli* K-12 derivative, in which the *mcr hsd mrr* region has been replaced with the *hsd* region of *S.enterica* serovar *blegdam* (hereafter abbreviated to *S.blegdam*), was used as a λ-sensitive strain proficient in the *blegdam* (*Sty*SBLI) R–M system. Mutant derivatives of LB4037 were made for use in complementation tests; in NM856 (*hsdS*), the *hsdS* gene was inactivated by the insertion of *supF* (18), in NM867 (*hsdM*), a mutation changed the sequence of the methyltransferase motif IV from NPPF to NPPC and in NM857 (*hsdR*), a mutation changed the K in the ATP-binding motif to T.

For experiments using M13, F′ derivatives of LB4037, the *mutD*5 strain RP526 (21) and the *dam*$^-$ strain CB51 (provided by Dr A. C. Boyd, Medical Genetics, Western General Hospital, University of Edinburgh, UK) were made by transferring the F′::Tn*5* donor from EH55 (22). XL1-Blue MRF′ (Stratagene) was used as an r$^-$m$^-$*supE* host for M13 phages. DH5α (23) and XL1-Blue (Stratagene) were hosts for the recovery and amplification of plasmid DNA. ED8654 (24) was the standard λ-sensitive, restriction-deficient strain for the recovery and amplification of λ phages. Either NM679 (25), a Δ (*mcr hsd mrr*) Δ *mcrA* derivative of W3110 or JR300 (26), an *E.coli* C strain that is naturally deficient in *hsd* genes, were used as λ-sensitive, r$^-$m$^-$ hosts.

Libraries of ECOR9 DNA were made in NM1249, a *c*I857 derivative of EMBL3 (27). A 4.5 kb *Eco*RI–*Sal*I fragment from a λ*hsd* phage (isolate #8) subcloned in pUC19 generated pECOR9, a plasmid that conferred *Eco*R9I modification proficiency, but not restriction proficiency, to *E.coli* K-12 strains. The plasmids containing the *hsd* genes of *K.pneumoniae* were pJR41 (*hsdM*$^+$*S*$^+$), pJR43 (*hsdM*$^+$), pNL3 one of four (*hsdR*$^+$) clones pNL1-4 (28) and pJR51 (*hsdR*$^+$*M*$^+$*S*$^+$). pJR51 contains two *Pst*I fragments in pUC19: one (6.7 kb) from pJR31 includes *hsdM* and *hsdS* and a part of *hsdR*, and the other, a 1.4 kb fragment from pNL3, provides the remainder of *hsdR*. The plasmids containing the *hsd* genes from *S.blegdam* were described by Titheradge *et al.* (18). pAC18 includes the *hsdM* and *hsdS* genes within an *Eco*RI fragment and pAT29, the *hsdR* gene within a *Bgl*II fragment in pUC9. These plasmids were the substrates for site-directed mutagenesis to generate mutations in *hsdM* and *hsdR*, respectively. In each case an *Eco*RI fragment was then subcloned in NM461 (λ*b*538*c*I857srI4°*nin* srI5°) an integration-defective, temperature-inducible λ vector (29), and transferred to the bacterial chromosome to generate the *hsdM* and *hsdR* derivatives of LB4037.

λ*vir*, used to test restriction and modification, was either unmodified (λ*vir*.0) by propagation in NM679 or JR300, or appropriately modified by propagation in LB4037 (λ*vir*.*Sty*SBLI), *E.coli* C/pJR41 (λ*vir*.*Kpn*AI) or NM679 lysogenic for λ*hsd* clone #8 (λ*vir*.*Eco*R9I). λ*vir*.0 after propagation on a test strain was checked for modification. All tests dependent on plasmids used freshly transformed strains.

Media and microbial techniques

Media and general methods (28,30) and the use of λ*hsd* phages to transfer mutations to the chromosome have been described (31).

Enzymes, reagents and reactions

Restriction enzymes, T4 DNA ligase, mung bean nuclease and Klenow polymerase were purchased from Boehringer Mannheim or New England Biolabs. Red Hot *Taq* polymerase from Advanced Biotechnologies Ltd was used to amplify DNA to make probes. Site-directed mutagenesis was done using the Quikchange™ mutagenesis kit of Stratagene. λ packaging extracts were from Promega.

DNA preparations and manipulations used standard methods (32). The reagents and methods for the detection of DNA sequences by hybridisation have been described (5). The ABI PRISM dRhodamine Terminator Cycle Sequencing Ready Reaction Kit from PE Biosystems was used to prepare samples for an automated sequencer (AB1 PRISM 377). Plasmid DNA for sequence determination was prepared using the Biotech Flexiprep Kit (Pharmacia). Oligonucleotides were obtained from Oswel or MWG-Biotech UK Ltd. The sequence of the *hsd* genes of ECOR9 was always determined on both strands. The sequences of DNA fragments were compiled using GeneJockey.

Sequence comparisons

Alignments were made initially using the TBLASTN program (33) available on the National Center for Biotechnology Information web site (http://www.ncbi.nlm.nih.gov/BLAST). The predicted amino acid sequences of the subunits of *Eco*KI, *Eco*AI, *Eco*R124I/II and *Sty*SBLI were used to screen databases. The nucleotide sequences specifying these subunits are in the EMBL/GenBank/DDBJ databases under the following accession numbers: *Eco*KI (U14003); *Eco*AI HsdR (L18758), HsdM (L02505), HsdS (J03150); *Eco*R124I/II (X13145) and

```
mp8  GAATTC              CCGGGGATCC       GTCGACCTGCAGC     CAAGCTTGGC

mp10 GAATTCGAGCTCG    CCCGGGGATCCTCTAGAGTCGACCTGCAGC    CCAAGCTTGGC

                1  3  5  7  9  11 13 15
mp18 GAATTCGAGCTCGGTACCCGGGGATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTGGC
     EcoRI      SacI     KpnI     XmaI     BamHI                SphI      HindIII
```

Figure 1. Polylinker sequences of M13 vectors. The sequences of the polylinkers of mp8, mp10 and mp18 are aligned. The targets for type II endonucleases are identified in the sequence of mp18 and the numbers (1–15) correspond to bases 1–15 in Figure 2.

*Sty*SBLI (X99719). The sequence for the HsdR of *Eco*R124 I/II (X13145) was corrected by the addition of a C at position 3064; this changes the C-terminal sequence from FRKSSRLLRSLKA to FQKIVSFIEKFKGVGGKI. In addition to the databases of published sequences, preliminary sequence data were obtained from the NCBI unfinished Microbial Genomes BLAST website at http://www.ncbi.nlm.nih.gov/Microb_blast/unfinishedgenome.html; http://www.sanger.ac.uk/Projects/S_typhi; http://www.sanger.ac.uk/Projects/S_equi.

The multiple sequence alignments for the comparisons in Table 5 used the CLUSTAL W(1.5) program.

RESULTS

The recognition sequence for *Sty*SBLI

Gann *et al.* (34) introduced an *in vivo* strategy in which M13 clones were used as substrates to define the target sequence for a type I R–M system; only those phages that are susceptible to restriction contain a target sequence. When experiments were initiated to determine the target sequence recognised by *Sty*SBLI, it was found that two commonly used M13 vectors, mp18 and mp19, that differ in the orientation of their polylinker sequence, had an efficiency of plating (e.o.p.) of ~0.1 on an *E.coli* strain in which the *hsd* genes for *Eco*KI had been replaced by those specifying *Sty*SBLI (LB4037F′). This implied a single target for *Sty*SBLI within these M13 vectors. Related vectors, mp8 and mp10, differing from mp18 in their shorter polylinker sequences (Fig. 1), were not restricted (e.o.p. of 1). These results indicate that the polylinker sequence of M13mp18 includes a recognition sequence for *Sty*SBLI, part of which is within either the *Kpn*I or the *Sph*I target; targets that are absent in mp8 and mp10 (Fig. 1).

Two methods were used to refine the boundaries of the predicted recognition sequence. First, mutant derivatives of mp18 that have an e.o.p. of 1 on LB4037F′ were selected after amplification on a *mutD* strain. Two mutants with base substitutions were identified, and each base change was within the *Kpn*I target (residues 13 and 14 in Fig. 1). Secondly, mutations were made within the polylinker sequence after cutting the vector with a type II restriction endonuclease. The projecting 5′ single-stranded regions were either removed or used as templates to make double-stranded ends, and circular genomes were restored by DNA ligase. Changes within the *Eco*RI, *Bam*HI and *Xma*I target sequences failed to remove the *Sty*SBLI recognition sequence, while that within the *Sac*I sequence generated a restriction-resistant derivative. These experiments identify the relevance of bases within the *Sac*I and *Kpn*I target sequences.

Known type I recognition sequences are bipartite, consisting of a 3 or 4 bp component, separated by a non-specific spacer of 6–8 bp from a second component of 4 or 5 bp. Each component includes an adenine residue, the substrate for methylation. The adenine residues are on opposite strands and generally ~10 bp apart (8,9). Residue 4 within the *Sac*I target and residue 11 within the *Kpn*I target are the only candidates for methylation, within the defined region (Fig. 1).

Eleven nucleotides that included the *Sac*I and *Kpn*I target sequence (nucleotides 1–5 and 10–15) were subjected to site-directed mutagenesis. Each nucleotide was replaced by the three alternatives to determine whether degeneracies at any position provided a sequence that could be recognised by *Sty*SBLI. The results, summarised in Figure 2, define a target sequence, CGA(N_6)TACC comprising 7 bp with no degeneracies and a spacer of 6 bp. They also show that methylation of the first adenine residue (position 4 in Fig. 2) by the Dam methylase prevents attack by *Sty*SBLI.

Cloning the *hsd* genes of ECOR9

DNA from ECOR9, a member of the ECOR collection of *E.coli* strains (20), has been shown to hybridise with a probe specific to the ICR of *S.blegdam* (5). ECOR9 is presumed to include *hsd* genes, but the strain is insensitive to the phages commonly available in the laboratory; therefore, no biological tests could be made. Evidence for a functional restriction system required that the putative *hsd* genes were cloned in

	1	2	3	4	5	6-9	10	11	12	13	14	15
	t	**C**	**G**	**A**	**g**	**ctcg**	**g**	**T**	**A**	**C**	**C**	**c**
A	√	x	x	n.a.	√	n.d.	√	x	n.a.	x	x	√
C	√	n.a.	x	x	√	n.d.	√	x	x	n.a.	n.a.	n.a.
G	√	x	n.a.	x	n.a.	n.d.	n.a.	x	x	x	x	√
T	n.a.	x	x	x	√*	n.d.	√	n.a.	x	x	x	√

Figure 2. Base substitutions made to identify the nucleotide sequence recognised by *Sty*SBLI. The upper case letters in bold define the nucleotide sequence recognised by *Sty*SBLI and the nucleotides are numbered as in Figure 1. Substitutions were made for the bases identified by numbers 1–5 and 10–15. √ identifies a base change that is without effect on the recognition of the nucleotide sequence by *Sty*SBLI. x identifies a base change that destroys the target sequence. n.a., not applicable; n.d., not done. *When G is replaced with T at this position, the A residues within the sequence GATC become the substrate for the Dam methylase; methylation of the A residues blocks restriction by *Sty*SBLI.

E.coli K-12. A library of DNA fragments, generated as the products of partial digestion by *Sau*3AI, was recovered in the λ vector NM1249, a *c*I857 derivative of EMBL3 (27). Plaques identified by the *Sty*SBLI-specific probe were purified and tested with probes derived from the sequences that flank the *hsd* genes in serovars of *S.enterica*. These probes also hybridise to DNA in *E.coli* K-12. One *hsd* phage (clone 8) included DNA that hybridised to the probe made from sequence extending ~2 kb downstream of *hsdS* in *S.enterica* serovar *typhimurium* LT2 (pAB3 in ref. 18), while another λ*hsd* phage (clone 9) hybridised to the probe made from a DNA fragment located ~5 kb upstream of *hsdM* in *S.blegdam* (18). The type IA *hsd* genes of serovar *typhimurium* and the type ID *hsd* genes of serovar *blegdam* were known to have the same chromosomal location despite their different gene order (18). The physical evidence from the λ clones supports the same location for the type ID *hsd* genes of ECOR9.

A 4.5 kb *Eco*RI–*Sal*I fragment shown to hybridise to the *hsdS* gene of *Sty*SBLI was transferred from λ clone 8 to pUC19. The nucleotide sequence of part of the chromosomal DNA fragment within this plasmid (pECOR9) was determined (EMBL nucleotide sequence database, accession no. AJ132566). The sequence from *E.coli* ECOR9 was readily aligned with the *hsd* gene sequences of *S.blegdam*. It is consistent with the presence of *hsdM* and *hsdS*, followed by the beginning of *hsdR*; the same gene order as the first member of the ID family. Plasmid pECOR9 is predicted to specify the modification component of the R–M system *Eco*R9I. The comparison of the *hsdS* genes of *Sty*SBLI and *Eco*R9I identifies two variable regions as anticipated for two TRDs that specify a bipartite target sequence, flanking a central conserved region (Fig. 3). There is a conserved sequence at the C-terminus, but no well-conserved sequence was identified at the N-terminus.

```
          1                                                   50
StySBLI                                          MAFEKT IPLNEFITLQ
EcoR9I    MGSNGFKLPL GWNCKKLVDC TKEGNISYGI VQPGQHQEDG IGIIRVNNIQ

         51                                                  100
StySBLI   RGFDLPQDKR VMGDIPVVAS TGVVGYHNEE KVLAPGVVIG RSGSIGGGQY
EcoR9I    NG.......N IYIDDVLKVS HEIESKFAKT RLEGGEVLLT LVGSTGISAI

        101                                                   150
StySBLI   ITTNF..WPL .NTTLWVKDF KGHHPRFVYY LLRSIDFSQF NVGSGV....
EcoR9I    TTKALQGWNV ARAVAVIKPC DEISAEWIHI CLQS.PFTKY FLDSRANTTV

        151                                                   200
StySBLI   .PTLNRNHLS GILVADTSYS YEKEASDIIG ILDDKIKLNK ELNHTLEQIS
EcoR9I    QKTLNLKDVK EIPLPIPPHE ERVSLEKIYF NPENRINLNI KINKILEEMS

        201                                                   250
StySBLI   QTLFKSWFVD FDPVIDNALD AGNPIPEALQ SRAELRQKIR NSADFKPLPA
EcoR9I    QNLFKSWFVD FDPVVDNALD AGNPIPEALQ SRAELRQKVR NSADFKPLPA

        251                                                   300
StySBLI   DIRALFPAEF EETELGWMPK GWITTSFNDL IELIGG.GTP KTSVEEFWNG
EcoR9I    EIRSLFPSEF EETELGWMPK GWQIKSLDHI ANFQNGLALQ KFRPKNMEDD

        301                                                   350
StySBLI   DIPWFSVVDA PSESDVYVLT TEKKITIEGL NNSSAKLLRK GTTIISARGT
EcoR9I    YLPVLKIADL RAGQ....IT NDERARTD.. ISDSCKVY.D GDMIFSWSGT

        351                                                   400
StySBLI   VGKCAMVAVP MAMNQSCYGV IGKNNISDEY IYFQLKNAVQ TLQQMGHGSV
EcoR9I    LMIDIWTGGN AALNQHLYKV TSKK..YPQS FYFMW..TIQ HLSRFQHIAE

        401                                                   450
StySBLI   FNTITRDTFK ..NIKVPFC. ..NEELTNSY SLLVKNYFSK ILNNNYQNIA
EcoR9I    AKAVTMGHIK KGDLSNSFCL IPTSSLITKY DNIVGGYLAK IKNQRLLNNQ

        451                                  482
StySBLI   LTNLRDTLLP KLISGELSLE DLPNLAKQTE PA
EcoR9I    MTALRDTLLP KLISGELSLD DIPDLNTDTE AA
```

Figure 3. An alignment of the amino acid sequences of the S subunits of *Sty*SBLI and *Eco*R9I. The alignment was made using PILEUP [Wisconsin Package Version 10, Genetics Computer Group (GCG), Madison, WI, USA]. Conserved amino acids are indicated by bold type. The alignments identify two variable regions (TRDs) flanking a central conserved region and a conserved C-terminus.

Complementation between subunits of different enzymes

The close relatedness of two type I R–M systems was first demonstrated for *Eco*KI and *Eco*BI; in these experiments functional R–M systems resulted when a subunit of one enzyme replaced the subunit of the other, with the HsdS subunit determining the sequence specificity of a chimaeric complex (6). These tests for restriction and modification were done *in vivo* using a partial diploid in which a second set of *hsd* genes was provided on an F′ plasmid. Such tests are now more generally applicable to genes cloned in a plasmid vector.

The *hsd* genes for *Eco*R9I were previously identified by hybridisation screens, but those in *K.pneumoniae*, specifying *Kpn*AI have already been cloned. Sequence comparisons identified the latter as a likely member of the ID family (28); the predicted HsdM sequences of *Kpn*AI and *Sty*SBLI share 97% identity and those for HsdR 94% identity. The *hsd* genes of *S.blegdam* had been transferred to *E.coli* (12) and a mutation in *hsdS* made by the insertion of *supF* (18). Site-directed mutagenesis of *hsdM* (in pAC18) and *hsdR* (in pAT29) was used to generate substitutions in conserved motifs within the active sites for the methyltransferase and ATPase activities, respectively. These mutations were transferred via λ*hsd* phages to the chromosome of LB4037, the *E.coli* strain specifying *Sty*SBLI, to make *hsdR*, *hsdM* and *hsdS* mutants of a λ-sensitive strain (see Materials and Methods). Partial diploids were made by transforming the mutant strains with a plasmid carrying the *hsdM* and *hsdS* genes of ECOR9 or plasmids carrying one or more of the *hsd* genes that specify *Kpn*AI. The HsdR subunit of *Sty*SBLI could substitute for the HsdR subunit of *Eco*R9I, and the HsdM subunit of *Eco*R9I for that of *Sty*SBLI as seen by the presence of two specificities when the *hsdM* strain (NM867) was transformed with pECOR9 (Table 1). Similarly, the subunits of *Kpn*AI could substitute for those of *Sty*SBLI (Table 2). Wherever a strain included functional *hsdR*, *hsdM* and *hsdS* genes, a R–M-proficient strain was obtained. Additional complementation tests using two plasmids confirmed the functional association of HsdR and HsdM of *Sty*SBLI with those of *Kpn*AI (data not shown).

Our results demonstrate that the three R–M systems from different genera maintain sufficient similarity to meet the most demanding requirement for membership of the same family of type I R–M systems.

DISCUSSION

Sequence specificity

All type I R–M systems are likely to have a common origin (35), those allocated to a family being very closely related but illustrating significant diversification only within the specificity gene. The present organisation of an *hsdS* gene that specifies two TRDs is believed to have arisen by either duplication

Table 1. Complementation between subunits of *Sty*SBLI and *Eco*R9I

Strain	Functional Hsd subunits *Sty*SBLI (on the chromosome)			*Eco*R9I (on the plasmid)		e.o.p. of λvir which is unmodified (v.0) or modified against *Sty*SBLI (v.SBLI) or *Eco*R9I (v.R9I)			Relevant phenotype
	R	M	S	M	S	v.0	v.SBLI	v.R9I	
NM679						1	1	1	r^-m^-
LB4037	+	+	+			$(1.8 \pm 0.7) \times 10^{-5}$	0.8 ± 0.02	$(4.3 \pm 5.8) \times 10^{-5}$	$r^+m^+_{StySBLI}$
NM856	+	+				0.7 ± 0.06	0.7 ± 0.1	0.8 ± 0.3	r^-m^-
NM867	+		+			0.9 ± 0.2	0.9 ± 0.1	0.9 ± 0.3	r^-m^-
NM679(pECOR9)				+	+	1.1 ± 0.2	1.0 ± 0.3	1.0 ± 0.4	$r^-m^+_{EcoR9I}$
NM856(pECOR9)	+	+		+	+	$(4.0 \pm 3.7) \times 10^{-4}$	$(1.2 \pm 0.7) \times 10^{-4}$	0.5 ± 0.1	$r^+m^+_{EcoR9I}$
NM867(pECOR9)	+		+	+	+	$(3.1 \pm 1.7) \times 10^{-4}$	$(9.9 \pm 3.5) \times 10^{-4}$	$(6.4 \pm 2.9) \times 10^{-4}$	$r^+m^+_{StySBLI}$ $r^+m^+_{EcoR9I}$

Table 2. Complementation between subunits of *Sty*SBLI and *Kpn*AI

Strain	Functional Hsd Subunits *Sty*SBLI (on the chromosome)			*Kpn*AI (on the plasmid)			e.o.p. of λvir which is unmodified (v.0) or modified against *Sty*SBLI (v.SBLI) or *Kpn*AI (v.KpnAI)			Relevant phenotype
	R	M	S	R	M	S	v.0	v.SBLI	v.KpnAI	
E.coli C							1	1	1	r^-m^-
LB4037	+	+	+				$(1.2 \pm 0.7) \times 10^{-4}$	0.8 ± 0.6	$(3.5 \pm 0.9) \times 10^{-4}$	$r^+m^+_{StySBLI}$
DH5α(pJR51)				+	+	+	$(3.6 \pm 1.6) \times 10^{-8}$	$(5.7 \pm 4.2) \times 10^{-8}$	0.8 ± 0.4	$r^+m^+_{KpnAI}$
NM857		+	+				1.2 ± 0.1	1.4 ± 0.5	1.4 ± 0.5	$r^-m^+_{StySBLI}$
NM867	+		+				1.5 ± 0.1	1.4 ± 0.4	1.6 ± 0.6	r^-m^-
NM857(pNL3)		+	+	+			$(1.0 \pm 0.6) \times 10^{-5}$	1.5 ± 0.5	$(9.3 \pm 5.8) \times 10^{-5}$	$r^+m^+_{StySBLI}$
NM867(pJR43)	+		+		+		$(2.8 \pm 2.2) \times 10^{-4}$	1.2 ± 0.2	$(6.5 \pm 4.6) \times 10^{-4}$	$r^+m^+_{StySBLI}$
NM867(pJR41)	+		+		+	+	$(5.8 \pm 3.4) \times 10^{-5}$	$(6.3 \pm 1.3) \times 10^{-3}$	$(6.6 \pm 4.3) \times 10^{-4}$	$r^+m^+_{StySBLI}$ $r^+m^+_{KpnAI}$

within the gene or gene duplication followed by gene fusion. The determinants of the N- and C-terminal TRDs would, therefore, be derived from a common ancestor (36). Gene duplication may have occurred more than once. The TRDs have evolved to recognise a variety of tri-, tetra- and even pentanucleotide sequences, but all known target sequences for type I R–M systems comprise two components separated by a non-specific sequence of 6–8 bp.

Currently, all the sequences recognised by type I R–M systems include two adenosyl residues, one in each strand of DNA, which are the targets for methylation. Early experiments indicated that the adenosyl residues targeted by type I modification enzymes were ~10 bp apart, separated by 8 or 9 bp, consistent with the two TRDs of HsdS making interactions within two successive major grooves of the DNA helix (37). However the two adenine residues that are methylated by *Eco*R124I (type IC) are separated by 7 bp (38) and our experiments determine the sequence recognised by *Sty*SBLI as $CGA(N)_6TACC$ within which the two available adenosyl residues that identify each strand of the recognition sequence are separated by only 6 bp. Methylation of the adenosyl residue in the CGA sequence was shown to protect the target sequence from restriction by *Sty*SBLI.

The emergence of families of enzymes, which differ in the distance between the bases that are the targets for methylation, has enhanced the potential for the diversification of target sequences. The present data, summarised in Table 3, are consistent with the methylation of adenosyl residues separated by 9 bp in the IB family, 8 bp in the IA family, 7 or 8 bp in the IC family and 6 bp in the first member of the ID family. The variability in the IC family is dictated by whether a tetrapeptide sequence (TAEL) within the central conserved region is present in duplicate or in triplicate, the additional four amino acids increasing the separation of the target adenines by 1 bp (39). The importance of the correct spacing between the adenine residues is emphasised by the target sequences for *Eco*R124IΔ (40) and *Eco*DXXIΔ (41). In these systems, where the 3′ half of *hsd*S is lost and the two truncated HsdS subunits associate symmetrically, an additional base pair within the spacer sequence maintains the distance between the target adenines (Table 3).

The genetic concept of families

The separation of type I R–M systems into families originally relied on genetic tests for complementation, first demonstrated for *Eco*KI and *Eco*BI (6,7). Complementation requires sufficient sequence conservation to permit subunits from one

Table 3. Family-specific distance between target adenines

Family	Enzyme	Recognition sequence	Reference
IB	*Eco*AI	G**A**GNNNNNNNG**T**CA	37
	*Eco*EI	G**A**GNNNNNNNA**T**GC	19
	*Cfr*AI	GC**A**NNNNNNNNG**T**GG	36
IA	*Eco*BI	TG**A**NNNNNNNNN**T**GCT	37
	*Eco*KI	**A**ACNNNNNNG**T**GC	37
	*Eco*DI	TT**A**NNNNNNNG**T**CY	37
	StyLTIII	G**A**GNNNNNNR**T**AYG	37
	*Sty*SPI	**A**ACNNNNNNG**T**RC	37
IC	*Eco*R124I	GA**A**NNNNNNR**T**CG	38,55
	*Eco*R124IΔ	GA**A**NNNNNNNN**T**TC	40
	EcoR124II[a]	GA**A**NNNNNNNR**T**CG	37
	*Eco*DXXI[a]	TC**A**NNNNNNNR**T**TC	56
	*Eco*prrI[a]	CC**A**NNNNNNNR**T**GC	43
	*Eco*DXXIΔ[a]	TC**A**NNNNNNNN**T**GA	41
ID	*Sty*SBLI	CG**A**NNNNNN**T**ACC	This work

Where N = any nucleotide, R is either purine and Y is either pyrimidine and the bold type identifies either the adenine that is the target for methylation or the thymine complementary to the target adenine. For *Eco*EI, *Cfr*AI and *Sty*SBLI the relevant adenine residues are not defined by experiments, but are the sole candidates within the target sequences.
[a]These type IC members have four more amino acids within the central conserved region, the region that links the TRDs, than *Eco*R124I.

complex (e.g. *Eco*AI) to substitute for those in another (e.g. *Eco*EI) (16). The principal differences between two members of one family reside in the TRDs, the regions of HsdS that confer sequence specificity to the enzyme (31). Experimental tests for family relationships have come to rely on the more generally applicable approaches of DNA hybridisation screens for similar gene sequences, or serological tests for similar HsdM and HsdR subunits, rather than genetic tests (5,14). The *hsd* genes for *Eco*R9I were identified by hybridisation with a probe derived from the *hsd* genes of *Sty*SBLI, the archetypal member of the ID family (5). The complementation tests reported in this paper confirm the allocation of *Eco*R9I to the type ID family.

The amino acid sequences predicted for the subunits of enzymes within the established families have >80% identity if the TRDs are ignored, whereas those between families generally indicate only 20–30% identity (18,35,42). The levels of identity in these sequence comparisons explain the ease with which type I R–M systems have been allocated to a family. Such sequence comparisons placed *Kpn*AI in the ID family (28). Our genetic tests for complementation provide the classical evidence that *Kpn*AI, like *Eco*R9I, is a member of the ID family. The genes for *Eco*R9I and *Sty*SBLI have a similar location in the chromosome of their respective bacterial species; the location of those for *Kpn*AI remains to be determined. However, allelism is not a necessary requirement for membership of a family; one member of the type IC family is specified by chromosomal genes, apparently as part of a defective prophage (43,44), others are specified by plasmid genes (37,39).

During the past two decades the concept of families of type I restriction enzymes, in which alternative specificity subunits confer different sequence specificities, has provided a pragmatic basis for the description and understanding of these R–M systems. It seems likely that this concept can be extended from enteric bacteria to other genera. In *Helicobacter*, for example, allelic genes specify putative type I R–M systems for which the predicted HsdS subunits seem likely to confer different specificities (11). Similar conclusions can be drawn for *Lactococcus lactis* (45). However, in this case the genetic experiments add a novel complication to the family status. The data from this genus suggest a discrepancy between the family defined by complementation tests and that dependent upon a high level of amino acid identity (>80%) between polypeptides. Schouler and colleagues (45) found for *Lactococcus* that the HsdR and HsdM subunits specified by two plasmids have the high level of identity (~90%) expected for membership of a family, while these sequences have only 40% identity with those specified by the bacterial chromosome. Despite this relatively low level of identity, the plasmid-encoded HsdS subunits were found to interact with the chromosomally encoded HsdM subunits. Type I R–M systems, which by sequence analysis might be separated into two families, were within one family as assessed by complementation tests. A more detailed analysis of the HsdS subunits identified conserved sequences common to all the HsdS subunits, whether of plasmid or chromosomal origin. In addition, all the HsdM subunits share a conserved C-terminus. It was suggested (45) that sequences conserved in all HsdM subunits, and those conserved in all HsdS subunits, identify the sites of interaction between HsdM and HsdS.

On the basis of the *Lactococcus* enzymes, comparisons of the conserved sequences within HsdS could indicate which HsdS subunits might substitute for others, and hence point to the family relationship, but the general similarities detected between the conserved regions of HsdS subunits in different type I families (36,46) suggest that the experimental test for complementation may remain the only test for exchange of subunits. The high level of identity within the C-termini of the chromosomal and plasmid-encoded HsdM subunits could indicate that convergent evolution has enhanced the potential for the generation of enzymes with different specificities. In this way a reservoir of independent, plasmid-borne *hsdS* genes provides an effective 'combinational' system for varying the target specificity of the catalytic subunits provided by the host.

Sequence comparisons

We wished to determine whether the striking subdivision of type I R–M systems found within the enteric bacteria can be extended to other genera, or other phyla, of bacteria. Genetic tests are of limited value in the analysis of putative R–M systems provided by the databases of genomic sequences from an extensive array of bacterial species. Polypeptide sequences provide a general approach and current programs compensate for possible errors in the DNA sequence. We therefore decided to search the databases of completed and incomplete genomic sequences, using the TBLASTN program (33), for amino acid sequences similar to those predicted for the archetypal subunits of the type IA, IB, IC and ID families. All HsdM subunits share the motifs of methyltransferases and all HsdR subunits have

the DEAD-box motifs; the motifs correlate with the catalytic activities of the subunits. In the enteric bacteria the current interfamily levels of identity for HsdM extend from 25 to 33% and those for HsdR from 17 to 26% (18,35,42).

In our first screen we used the amino acid sequences of the HsdM polypeptides of the archetypal representatives of each of the four established families of type I enzymes to identify putative HsdM subunits with identities >45% (Table 4). The figure of >45% was chosen as one that was appreciably higher than any found previously for an interfamily comparison, but it sets a high level of identity for comparisons between widely separated genera, where low levels of identity are found when enzymes involved in basic metabolic pathways are compared. For example, the level of identity for triosephosphate isomerase for *E.coli* and *Xylella fastidiosa* is 48%; that between *E.coli* and *Bacillus stearothermophilus* is only 41%.

The polypeptides identified by our screens are listed in Table 4, each section identifying one of the four screens. The table is simplified by the omission of three sequences identified by the screen with the HsdM polypeptide of *Eco*R124I (Table 4C) and some sequences identified by the screen with *Sty*SBLI (Table 4D). The putative HsdM sequences omitted from Table 4C are a coding sequence in *Ureoplasma urealyticum* which is interrupted by a rearrangement, a coding sequence in *K.pneumoniae* which is within transposon Tn*5708* and the sequence in serotype B of *Neisseria meningitidis* which is nearly identical to that identified in serogroup A. The sequences omitted from Table 4D are identified later.

The sequences of the HsdM polypeptides listed in Table 4 were compared with each other using the TBLASTN program. Comparisons within a group usually showed >45% identity, the lowest level was 43% for that between *Streptococcus pneumoniae* and *Mycobacterium avium*; those between sequences allocated to different groups had <35% identity. On this basis, where comparisons relied on alignments of the major portion (>95%) of the subunit, the HsdM polypeptides identified in Table 4 fall into four discrete groups.

The databases were then screened with each of the four archetypal polypeptide sequences for HsdR and HsdS. In those genomic sequences where the HsdM polypeptides were identified by the IA, IB or IC comparisons, HsdR and HsdS polypeptides were detected (Table 4A–C). The coding sequences for the HsdR, HsdM and HsdS polypeptides identified in Table 4 were closely linked, and in the order consistent with the suggested family affiliation; the gene order for the known members of the IA and IB families is *hsdR*, *hsdM*, *hsdS*, while that for the IC and ID families is *hsdM*, *hsdS*, *hsdR*. The levels of identity found for HsdR were 37% or greater, higher therefore than those (17–26%) reported for interfamily comparisons in *E.coli*. The relatively long (~1000 amino acids) HsdR subunits of enteric bacteria show lower levels of identity than HsdM subunits, probably because much of the HsdR polypeptide sequence is not within a predicted catalytic domain (42). Close relationships between HsdS subunits are obscured by the very variable sequences of the TRDs, but for each of the known families conservation is readily detected within the region of HsdS that separates the two TRDs. This region, the so-called central conserved region, is presumed to be involved in the association of HsdS with the catalytic subunits. With the exception of *M.avium*, for which the sequence data were insufficient, the central conserved region of the HsdS subunit was readily aligned with that of the archetypal representative of the group designated on the basis of identities within HsdM (Table 4).

The screen using the type ID sequences identified a number of HsdM subunits with >45% identity, but some HsdM sequences (*B.stearothermophilus*, *Mycobacterium tuberculosis* and *Mycobacterium bovis*) were not associated with HsdR subunits and others (e.g. *Acidithiobacillus ferrooxidans*, *X.fastidiosa*, *Campylobacter jejuni*, *Pseudomonas syringae* p.v. tomato and a second putative type I R–M system of *Chlorobium tepidum*) were associated with HsdR subunits with <35% identity. These putative type I R–M systems were, therefore, omitted from Table 4. The R–M system of *Pasteurella haemolytica*, referred to as type 1d by Highlander and Garza (47) did not qualify for entry in Table 4D using our stringent criteria; its HsdM subunit has only 42% identity. The sequence comparisons (Table 4D), as predicted, identify *Kpn*AI with scores >90% identity, but they indicate even higher scores (96–100% identity) for three polypeptides in *S.enterica* serovar *enteritidis*. These Hsd sequences in *S.enteritidis* correlate well with the biological information that *S.blegdam* and *S.enteritidis* have an R–M system with the same sequence specificity (12).

Three of the screens identified putative type I R–M systems across a wide range of bacterial species. Each of the four screens identify HsdM polypeptides with >45% identity in different genera of the *Gamma* subdivision of the *Proteobacteria* (e.g. *Shewanella*, *Haemophilus* and *Pseudomonas*) and in the *Firmicutes* (*Bacillus*, *Streptococcus* and *Mycobacterium*). In addition, close relatives of *Eco*R124I (type IC) were identified in the *Beta* (*Neisseria*) and *Epsilon* (*Helicobacter*) subdivisions of the *Proteobacteria* and in a member of the *Green Sulphur Bacteria* (*Chlorobium*). The HsdM and HsdR polypeptides listed in Table 4 were aligned by the CLUSTAL W program and the identities determined after the exclusion of sites that contain a gap in any sequence (48). These comparisons (Table 5) identify groups IA, IB, IC and ID on the basis of identities in HsdM and HsdR. Members of the IA group are more closely related to those of IB than they are to those of IC and ID; similarly members of IC are more closely related to those of ID than they are to those of IA and IB. The extension of sequence comparisons to more distantly related bacteria currently permits the unambiguous association of putative type I R–M systems from a wide range of eubacterial species with one of the known families. Of course, many HsdM sequences in the database showed <45% identity with each of the four test sequences. The simplification of our analyses by selection of only M subunits with >45% identity has excluded some sequences that can be associated with one of the four families (P.M.Sharp, personal communication). The relationships analysed in Table 5 are not obviously complicated by genetic recombination. On the basis of sequence comparisons in Tables 4 and 5, we suggest that the family affiliations extend across the eubacterial kingdom. These affiliations are of evolutionary significance, irrespective of whether the groupings meet the classical genetic criterion for family status.

An examination of the aligned central conserved regions of HsdS polypeptides with that of *Eco*R124II indicated an additional feature that is in accord with the subdivisions based on identities within HsdM and HsdR. The archetypal members of the IC family include a tetrapeptide sequence (TAEL) present

Table 4. Comparisons among sequences identified by TBLASTN

Bacterial strain	HsdR[a]	HsdM[a]	S[b]	Reference[c]
(A) Per cent identity with *Eco*KI polypeptide sequences (type IA)				
S.typhimurium LT2[d]	74(81)[e]	92(95)	71	*Sty*LTIII, WUGS 99287 contig 1424
S.typhi CT18	91(95)	93(94)	80	Sanger ORFS STY4884,3 & 1
S.paratyphi A	65(73)[e]	90(91)[e]	80	WUGSC 32027
S.putrefaciens	39(57)	54(69)	71	TIGR24 6431
B.stearothermophilus	37(55)	49(63)	34	UOKN03 1422 contig 715
(B) Per cent identity with *Eco*AI polypeptide sequences (type IB)				
E.coli O157:H7EDL933	99(99)	98(98)	88	M.*Eco*O157 ORF 5947P
E.coli A58	77(87)	90(94)	85	*Eco*EI
P.putida KT2440	61(77)	67(78)	52	TIGR10787
A.ferrooxidans	56(71)	63(78)	62	TIGR6149
S.pneumoniae	48(65)	49(66)	35	TIGR3836
M.avium	39(55)	48(66)	i.d.	TIGR332
(C) Per cent identity with *Eco*R124II polypeptide sequences (type IC)				
C.tepidum	73(84)	84(90)	82	TIGR3499. J.Eisen (pers. comm.)
N.gonorrhoea	74(85)	75(87)	49	AEOO4969
N.meningitidis serotype A	72(84)[e]	75(86)	31	M.*Nme*A ORF1038P
Haemophilus influenzae Rd	76(86)[e]	66(72)[e]	53	M.*Hin*d ORF 215P
S.equi	67(82)	66(79)	61	Sanger 1336 contig 445
X.fastidiosa	42(60)	55(70)	54	M.*Xfa* ORF2728P[f]
H.pylori J99	42(62)	52(69)	46	M.*Hpy*99 ORF786P
H.pylori 26695	43(62)	54(70)	50	M.*Hpy*A ORF 850P
(D) Per cent identity with *Sty*SBLI polypeptide sequences (type ID)				
S.enteritidis	96(96)[e]	100(100)	100[e]	UIUC 592 contigs 1881 & 2214
K.pneumoniae	94(96)	97(98)	95	*Kpn*AI

Data selected where >45% identity for HsdM. The alignments for HsdR and HsdM include >95% of the length of the respective sequence, with the exception of *S.paratyphi* and *S.enteriditis*. The data are omitted for well-established family members (e.g. *Eco*BI and *Eco*DXXI) if the sequences are not available for all three genes.

[a]Per cent similarity indicated in brackets.

[b]The central conserved sequence of HsdS, as defined by Sturrock and Dryden (50), was compared to avoid the contribution made by the TRDs; the conserved regions are relatively short (varying from 56 amino acids for *Eco*KI to 155 for *Eco*AI). i.d., insufficient data.

[c]See REBASE (57) for systems identified as enzyme or protein (P) sequence.

[d]Serovar of *S.enterica*.

[e]Sequence alignment impaired by putative frame shifts. The numbers given are for the longest alignment.

[f]The M and S coding sequences are separated by a short ORF (ORF2727), it is not known whether this ORF is an artefact of cloning, or indicates a natural insertion within the coding sequence.

in duplicate in *Eco*R124I and in triplicate in *Eco*R124II (39). An identical, or related, repeat sequence was found in enzymes affiliated with the IC family. A duplicated TAEL sequence is present in *C.tepidum* and *Streptococcus equi*. Similar repeat sequences are present in triplicate in *Neisseria gonorrhoeae* (EATL), in *Haemophilus* (TSEL), in *Xylella* (EAEL) and in *Helicobacter pylori* 26695 (NTEL).

The specificity subunits

The HsdS subunits identified by searches using the predicted amino acid sequence of the specificity subunit of *Sty*SBLI provide additional interest (Table 6). Either an N- or C-terminal segment of the polypeptide, each presumed to be a TRD sequence, often enhances the level of identity. This information is lost when the entire HsdS polypeptides are compared. While the predicted carboxy TRD of the HsdS subunit from *C.tepidum* has 50% identity with *Sty*SBLI, the predicted amino TRD has merely 10% identity. This is in contrast to HsdS subunits from *K.pneumoniae*, *P.syringae*, *A.ferrooxidans* and *Actinobacillus actinomycetemcomitans* in which the predicted amino TRDs each have ~45% identity and the carboxy TRDs ~20%. All these examples suggest sequence conservation between the test sequence and one TRD. In an early comparison of the predicted amino acid sequence of *Eco*AI (IB) with those of five members of the IA family, marked similarity in the HsdS subunits was detected only with

Table 5. Sequence comparisons within and between groups

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	*Eco*KI (IA)	–	**95**	**60**	**55**	31	30	28	28	29	22	22	21	22	22	19	19	18	18
2	*Sty*LTIII	**92**	–	**60**	**56**	31	30	28	27	29	22	22	22	23	22	19	19	18	18
3	*S.putrefaciens*	**47**	**48**	–	**55**	34	33	29	29	30	20	21	20	20	21	19	17	20	20
4	*B.stearothermophilus*	**44**	**44**	**51**	–	32	31	29	27	29	21	21	20	22	21	21	20	19	19
5	*Eco*AI (IB)	25	24	23	25	–	**70**	**67**	**52**	**51**	22	24	23	23	24	20	23	19	19
6	*P.putida*	25	25	23	24	**64**	–	**70**	**53**	**50**	21	23	22	22	24	22	22	18	18
7	*A.ferroxidans*	25	24	25	25	**58**	**60**	–	**52**	**52**	20	23	21	22	22	20	21	17	17
8	*S.pneumoniae*	23	23	22	22	**52**	**53**	**51**	–	**41**	20	21	21	20	21	19	22	18	18
9	*M.avium*	23	23	23	23	**39**	**39**	**39**	**37**	–	22	23	21	21	21	18	19	19	18
10	*Eco*R124I (IC)	15	15	15	14	15	15	14	14	14	–	**86**	**82**	**78**	**69**	**56**	**58**	29	29
11	*C.tepidum*	15	15	14	14	15	16	15	14	15	**81**	–	**81**	**78**	**70**	**55**	**57**	30	29
12	*Hin*d ORF215P	15	16	15	15	15	16	14	15	15	**81**	**79**	–	**79**	**73**	**56**	**58**	30	29
13	*N.gonorrhoea*	16	16	14	16	15	16	15	15	15	**80**	**82**	**80**	–	**70**	**54**	**56**	29	28
14	*S.equi*	15	15	14	14	14	15	14	14	15	**74**	**73**	**75**	**75**	–	**59**	**59**	28	28
15	*Hpy*99 ORF786P	15	16	16	15	16	15	16	16	16	**49**	**49**	**48**	**49**	**49**	–	**62**	32	31
16	*Xfa* ORF2728P	15	15	15	16	15	15	16	16	14	**48**	**50**	**48**	**49**	**48**	**58**	–	31	31
17	*Sty*SBLI (ID)	13	14	12	13	15	15	14	13	14	22	22	22	22	20	20	20	–	**98**
18	*Kpn*AI	14	14	12	13	14	15	14	13	14	21	21	21	21	20	19	21	**95**	–

The R–M systems, or bacterial species, identified by numbers 1–18 are listed in the left-hand column; for 12, 15 or 16 the putative R–M system is identified by the ORF for HsdM (57). The values are the per cent identity of aligned sequences after the exclusion, from any pairwise comparison, of any site that contains a gap in any sequence; above the diagonal for HsdM and below the diagonal for HsdR. The sequences were aligned by the CLUSTAL W(1.5) multiple sequence alignment program (48), with a minor adjustment for HsdR. Values for comparisons within groups are given in bold. For each group only one representative of any genus was included. The maximum difference for comparisons between two representatives of the same group from one genus is that for *Eco*AI and *Eco*EI (10% for HsdM and 20% for HsdR), the smallest that for *Eco*AI and the polypeptides of *E.coli* O157 (1 and 0.25%, respectively).

Table 6. Sequence comparisons based on HsdS of *Sty*SBLI

Bacterial strain	Per cent identity	Length of alignment	Per cent identity in[a]			Reference[b]
			N-TRD	Centre	C-TRD	
S.enteritidis	100	434[c]	100	100	100	UIUC 592 contig 2214
E.coli ECOR9	43	300	[d]	89	21	*Eco*R9I (this work)
K.pneumoniae	44	437	41	95	21	*Kpn*AI
C.tepidum[e]	44	327	10	34	50	TIGR 3499
Pasteurella multocida PM70	36	344	12	48	34	CBU MN747 AE006190
Ps.syringae pv tomato	34	433	44	43	21	TIGR 323
A.ferrooxidans[e]	32	427	46	38	15	TIGR 6154
A.actinomycetemcomitans	32	421	43	24	24	OUACGT714

HsdS subunits with >30% identity in HsdS (434 amino acids) and >40% in some region of HsdS.
[a]The per cent identity in each of the three regions was calculated on the basis of the alignments given by TBLASTN for the entire HsdS sequence of *Sty*SBLI; the regions are defined according to Sturrock and Dryden (50).
[b]For enzymes and proteins see REBASE (57).
[c]Sequence alignment impaired by putative frame shift.
[d]There is insufficient identity in the N-TRD for the TBLASTN program to make an alignment. For an alignment by PILEUP see Figure 3.
[e]The subunit is not that identified in Table 4.

*Sty*LTIII (previously called *Sty*SB) (19). Residues within the N-terminal TRDs of *Eco*AI and *Sty*LTIII showed 44% identity. The similarity of the sequences of the TRDs of these enzymes correlates with their recognition of the trinucleotide GAG. This and other evidence (49) indicate that TRDs from different families are of similar sequence if they confer the same

sequence specificity. This similarity between TRDs in distantly related bacteria appears to imply a closer evolutionary relationship than those between dissimilar TRDs in the same family. It could, however, reflect some structural constraint on TRDs that recognise the same nucleotide sequence. A recent analysis suggested that all type I TRDs include a region with a similar conserved tertiary structure at the interface with DNA (50). Whatever the underlying explanation, the sequence similarities imply a conservation of TRDs in type I R–M systems, irrespective of family relationships or bacterial phylogeny.

Comparative analyses of the sequences of type II restriction endonucleases and modification methylases rarely document evidence for the relatedness of amino acid sequences outside of the catalytic domains (51). For many type II enzymes, structures of cocrystals are available in which the enzymes are bound to their target sequence. On the basis of these, diversity in the structure of the polypeptide sequences interfacing with DNA has been emphasised (52). In general, relatively little support for the evolutionary connections between different type II systems has been presented. This contrasts with the information from type I R–M systems where closely related groups, previously obvious for the enteric bacteria, are found in widely different bacterial phyla. This is not unexpected if the type I R–M systems have a common origin. Extremely tight post-translational control of restriction activity in the absence of adequate modification should facilitate the evolution of new specificities (53,54).

ACKNOWLEDGEMENTS

We thank our colleagues for their interest and support, particularly Costel Atanasiu and Erica de Leau. We are indebted to Jonathan Eisen, Rich Roberts, Paul Sharp and Geoff Wilson for their comments on the manuscript, Paul Sharp for generating Table 5, and to Natalie Honeyman and Alix Fraser for their tireless efforts in the preparation of this manuscript. We wish to acknowledge the many contributors to REBASE and the databases of bacterial genomes including the following supplied by personal communication: The Pathogen Sequencing Unit at The Sanger Centre for the sequences of *Salmonella typhi* and *S.equi*; The Institute for Genomic Research for the genomes *C.tepidum*, *M.avium*, *Pseudomonas putida*, *P.syringae*, *Shewanella putrefaciens*, *S.pneumoniae*, *A.ferrooxidans*; the sequencing centres at the University of Washington for WUGS 99287 (*Salmonella typhimurium*), 3207 (*Salmonella paratyphi*); The University of Illinois for UIUC592 (*S.enteritidis*); The University of Oklahoma for OUKN03 1442 (*B.stearothermophilus*) and Ngon 485 (*N.gonorrhoea*). The Gonococcal Genome Sequencing Project was supported by USPHS/NIH grant #AI38399. The work in Edinburgh was supported by the Medical Research Council, that in Loma Linda by grant number NMTBO315-8853-01.

REFERENCES

1. Raleigh,E.A. (1992) Organization and function of the *mcrBC* genes of *Escherichia coli* K-12. *Mol. Microbiol.*, **6**, 1079–1086.
2. Bertani,G. and Weigle,J.J. (1953) Host-controlled variation in bacterial viruses. *J. Bacteriol.*, **65**, 113–121.
3. Sain,B. and Murray,N.E. (1980) The *hsd* (host specificity) genes of *E.coli* K-12. *Mol. Gen. Genet.*, **180**, 35–46.
4. Daniel,A.S., Fuller-Pace,F.V., Legge,D.M. and Murray,N.E. (1988) Distribution and diversity of *hsd* genes in *Escherichia coli* and other enteric bacteria. *J. Bacteriol.*, **170**, 1775–1782.
5. Barcus,V.A., Titheradge,A.J.B. and Murray,N.E. (1995) The diversity of alleles at the hsd locus in natural populations of *Escherichia coli*. *Genetics*, **140**, 1187–1197.
6. Boyer,H.W. and Roulland-Dussoix,D. (1969) A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. *J. Mol. Biol.*, **41**, 459–472.
7. Glover,S.W. and Colson,C. (1969) Genetics of host-controlled restriction and modification in *Escherichia coli*. *Genet. Res.*, **13**, 227–240.
8. Redaschi,N. and Bickle,T.A. (1996) DNA restriction and modification systems. In Neidhart,F.C., Curtiss,R., Ingraham,J.L., Lin,E.C.C., Low,K.B., Magasanik,B., Reznikoff,W.S., Riley,M., Schaechter,M. and Umbarger,H.E. (eds), *Escherichia coli and Salmonella*, 2nd Edn. American Society for Microbiology, Washington, DC, pp. 773–781.
9. Dryden,D.T.F. (1999) Bacterial methyltransferases. In Cheng,X. and Blumenthal,R.M. (eds), *S-Adenosylmethionine-Dependent Methyltransferases*. World Scientific Publishing, Singapore, pp. 283–340.
10. Rao,D.N., Saha,S. and Krishnamurthy,V. (2000) The ATP-dependent restriction enzymes. *Prog. Nucleic Acid Res. Mol. Biol.*, **64**, 1–63.
11. Murray,N.E. (2000) Type I restriction systems: sophisticated molecular machines. *Microbiol. Mol. Biol. Revs*, **64**, 412–434.
12. Bullas,L.R., Colson,C. and Neufeld,B. (1980) Deoxynucleic acid restriction and modification systems in *Salmonella*: chromosomally-located systems of different serotypes. *J. Bacteriol.*, **141**, 275–292.
13. Barcus,V.A. and Murray,N.E. (1995) Barriers to recombination. In Baumberg,S., Young,J.P.W., Saunders,S.R. and Wellington,E.M.H. (ed.) *Population Genetics of Bacteria*. Society for General Microbiology, Cambridge University Press, pp. 31–58.
14. Murray,N.E., Gough,J.A., Suri,B. and Bickle,T.A. (1982) Structural homologies among type I restriction–modification systems. *EMBO J.*, **1**, 535–539.
15. Arber,W. and Wauters-Willems,D. (1970) Host specificity of DNA produced by *Escherichia* coli. XII. The two restriction and modification systems of strain 15T⁻. *Mol. Gen. Genet.*, **108**, 203–217.
16. Fuller-Pace,F.V., Cowan,G.M. and Murray,N.E. (1985) EcoA and EcoE: alternatives to the EcoK family of type I restriction and modification systems of *Escherichia coli*. *J. Mol. Biol.*, **186**, 65–75.
17. Suri,B. and Bickle,T.A. (1985) EcoA: the first member of a new family of type I restriction modification systems. Gene organization and enzymatic activities. *J. Mol. Biol.*, **186**, 77–85.
18. Titheradge,A.J.B, Terment,D. and Murray,N.E. (1996) A third family of allelic hsd genes in *Salmonella enterica*: sequence comparisons with related proteins identify conserved regions implicated in restriction of DNA. *Mol. Microbiol.*, **22**, 437–447.
19. Cowan,G.M., Gann,A.A.F. and Murray,N.E. (1989) Conservation of complex DNA recognition domains between families of restriction enzymes. *Cell*, **56**, 103–109.
20. Ochman,H. and Selander,R.K. (1984) Standard reference strains of *E.coli* from natural populations. *J. Bacteriol.*, **157**, 690–693.
21. Fowler,G., Degnen,E. and Cox,E.C. (1974) Mutational specificity of a conditional *Escherichia coli* mutator *mut*D5. *Mol. Gen. Genet.*, **133**, 179–191.
22. Hansen,E.B., Atlung,T., Atlung,F.G., Skovgaard,O. and Van Mayenberg,K. (1984) Fine structure genetic map and complementation analysis of mutations in the *dnaA* gene of *Escherichia coli*. *Mol. Gen. Genet.*, **196**, 387–396.
23. Grant,S.G.N., Jesse,J., Bloom,F.R. and Hanahan,D. (1990) Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation restriction mutants. *Proc. Natl Acad. Sci. USA*, **87**, 4645–4649.
24. Bork,K., Beggs,J.D., Brammar,W.J., Hopkins,A.S. and Murray,N.E. (1976) The construction *in vitro* of transducing derivatives of Phage Lambda. *Mol. Gen. Genet.*, **146**, 199–207.
25. King,G. and Murray,N.E. (1995) Restriction alleviation and modification enhancement by the Rac prophage. *Mol. Microbiol.*, **16**, 769–777.
26. Prakash-Cheng,A., Chung,S.S. and Ryu,J. (1993) The expression and regulation of *hsd*K genes after conjugative transfer. *Mol. Gen. Genet.*, **241**, 491–496.
27. Frischauf,A.-M., Lehrach,H., Poustka,A. and Murray,N.E. (1983) Lambda replacement vectors carrying polylinker sequences. *J. Mol. Biol.*, **170**, 827–842.

28. Lee,N.S., Rutebuka,O., Arakawa,T., Bickle,T.A. and Ryu,J. (1997) *Kpn*AI, a new type I restriction–modification system in *Klebsiella pneumoniae*. *J. Mol. Biol.*, **271**, 342–348.
29. Murray,N.E. and Murray,K. (1974) Manipulation of restriction targets in phage λ to form receptor chromosomes for DNA fragments. *Nature*, **251**, 476–481.
30. Murray,N.E., Brammar,W.J. and Murray,K. (1977) Lambdoid phages that simplify the recovery of *in vitro* recombinants. *Mol. Gen. Genet.*, **150**, 53–61.
31. Gough,J.A. and Murray,N.E. (1983) Sequence diversity among related genes for recognition of specific targets in DNA molecules. *J. Mol. Biol.*, **166**, 1–9.
32. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edn. Cold Spring Harbor Laboratory Press, Plainview, NY.
33. Altschul,S.F., Madden,T.L., Schäffer,J.Z., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3384–3402.
34. Gann,A.A.F., Campbell,A.J.B., Collins,J.F., Coulson,A.F.W. and Murray,N.E. (1987) Reassortment of DNA recognition domains and the evolution of new specificities. *Mol. Microbiol.*, **1**, 13–22.
35. Sharp,P.M., Kelleher,J.E., Daniel,A.S., Cowan,G.M. and Murray,N.E. (1992) Roles of selection and recombination in the evolution of type I restriction–modification systems in enterobacteria. *Proc. Natl Acad. Sci. USA*, **89**, 9836–9840.
36. Kannan,P., Cowan,G.M., Daniel,A.S., Gann,A.A.F. and Murray,N.E. (1989) Conservation of organization in the specificity polypeptides of two families of type I restriction enzymes. *J. Mol. Biol.*, **209**, 335–344.
37. Bickle,T.A. (1987) Restriction and modification systems. In Neidhardt,F.C. (ed.), *Escherichia coli and Salmonella*, Edn 1. ASM Press, Washington DC, pp. 692–696.
38. Taylor,I., Watts,D. and Kneale,G. (1993) Substrate recognition and selectivity in the type IC DNA modification methylase M.*Eco* R124I. *Nucleic Acids Res.*, **21**, 4929–4935.
39. Price,C., Lingner,J., Bickle,T.A., Firman,K. and Glover,S.W. (1989) Basis for changes in DNA recognition by the *Eco*R124 and *Eco*R124/3 type I DNA restriction and modification enzymes. *J. Mol. Biol.*, **205**, 115–125.
40. Abadjieva,A., Webb,M., Patel,J., Zinkevich,V. and Firman,K. (1993) Deletions within the DNA recognition subunit of M.*Eco*R124I that identify a region involved in protein-protein interactions between HsdS and HsdM. *J. Mol. Biol.*, **241**, 35–43.
41. Meister,J., MacWilliams,M., Hubner,P., Jutle,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.
42. Murray,N.E., Daniel,A.S., Cowan,G.M. and Sharp,P.M. (1993) Conservation of motifs within the unusually variable polypeptide sequences of type I restriction and modification enzymes. *Mol. Microbiol.*, **9**, 133–143.
43. 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.
44. Tyndall,C., Lehnherr,H., Sandmeier,U., Kulik,E. and Bickle,T.A. (1997) The type IC *hsd* loci of the enterobacterial arre flanked by DNA with high homology to the phage P1 genome: implications for the evolution and spread of DNA restriction systems. *Mol. Microbiol.*, **23**, 729–736.
45. Schouler,C.M., Gautier,M., Ehrlich,S.D. and Chopin,M.-C. (1998) Combinational variation of restriction–modification specificities in *Lactococcus lactis*. *Mol. Microbiol.*, **28**, 169–178.
46. Kneale,G.G. (1994) A symmetrical model for the domain structure of type I DNA methyltransferases. *J. Mol. Biol.*, **243**, 1–5.
47. Highlander,S.K. and Garza,D. (1996) The restriction–modification system of *Pasteurella haemolytica* is a member of a new family of enzymes. *Gene*, **178**, 89–96.
48. Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4672–4680.
49. Thorpe,P.H., Ternent,D. and Murray,N.E. (1997) The specificity of *Sty*SKI, a type I restriction enzyme, implies a structure with rotational symmetry. *Nucleic Acids Res.*, **25**, 1694–1700.
50. Sturrock,S.S. and Dryden,D.T.F. (1997) A prediction of the amino acids and structures involved in DNA recognition by type I DNA restriction and modification enzymes. *Nucleic Acids Res.*, **25**, 3408–3414.
51. Wilson,G.G. and Murray,N.E. (1991) Restriction and modification systems. *Annu. Rev. Genet.*, **25**, 585–627.
52. Lukacs,C.M. and Aggarwal,A.K. (2001) *Bg*lII and *Mun*I: what a difference a base makes. *Curr. Opin. Struct. Biol.*, **11**, 14–18.
53. Makovets,S., Doronina,V.A. and Murray,N.E. (1999) Regulation of endonuclease activity by proteolysis prevents breakage of unmodified bacterial chromosomes by type I restriction enzymes. *Proc. Natl Acad. Sci. USA*, **96**, 9757–9762.
54. O'Neill,M., Powell,L.M. and Murray,N.E. (2001) Target recognition by *Eco*KI; the recognition domain is robust and restriction-deficiency commonly results from the proteolytic control of enzyme activity. *J. Mol. Biol.*, **307**, 951–963
55. Price,C., Shepherd,J.C.W. and Bickle,T.A. (1987) DNA recognition by a new family of type I restriction enzymes: a unique relationship between two different DNA specificities. *EMBO J.*, **6**, 1493–1497.
56. Gubler,M., Graguglia,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.
57. Roberts,R.J. and Macelis,D. (2001) REBASE: restriction enzymes and methylases. *Nucleic Acids Res.*, **29**, 268–269.

Molecular Microbiology (2004) 51(1), 135–147 doi:10.1046/j.1365-2958.2003.03801.x

Is modification sufficient to protect a bacterial chromosome from a resident restriction endonuclease?

Svetlana Makovets,† Lynn M. Powell, Annette J. B. Titheradge, Garry W. Blakely* and Noreen E. Murray
Institute of Cell and Molecular Biology, Darwin Building, University of Edinburgh, King's Buildings, Edinburgh EH9 3JR, UK.

Summary

It has been generally accepted that DNA modification protects the chromosome of a bacterium encoding a restriction and modification system. But, when target sequences within the chromosome of one such bacterium (*Escherichia coli* K-12) are unmodified, the cell does not destroy its own DNA; instead, ClpXP inactivates the nuclease, and restriction is said to be alleviated. Thus, the resident chromosome is recognized as 'self' rather than 'foreign' even in the absence of modification. We now provide evidence that restriction alleviation may be a characteristic of Type I restriction–modification systems, and that it can be achieved by different mechanisms. Our experiments support disassembly of active endonuclease complexes as a potential mechanism. We identify amino acid substitutions in a restriction endonuclease, which impair restriction alleviation in response to treatment with a mutagen, and demonstrate that restriction alleviation serves to protect the chromosome even in the absence of mutagenic treatment. In the absence of efficient restriction alleviation, a Type I restriction enzyme cleaves host DNA and, under these conditions, homologous recombination maintains the integrity of the bacterial chromosome.

Introduction

A classical restriction endonuclease is maintained in bacterial cells in association with its cognate modification enzyme. Together, these activities constitute a restriction and modification (R-M) system in which the modification enzyme serves to conceal potential target sequences from its cognate endonuclease by the addition of methyl groups to specific bases, one on each strand of the target sequence for the endonuclease. Newly synthesized DNA normally retains the methylated imprint on one strand and, in this hemimethylated state, the unmethylated daughter strand is a substrate for modification rather than restriction.

Escherichia coli K-12 specifies *Eco*KI, which, like all Type I R-M systems, is a multifunctional complex comprising three different subunits (HsdR, HsdM and HsdS). All three subunits are essential for endonuclease activity (for reviews, see Bickle and Krüger, 1993; Murray, 2000; 2002; Dryden *et al.*, 2001). Although HsdM and HsdS can suffice for modification, the alternative activities of the complete Type I R-M complex are determined by the methylation state of its target sequence. Restriction by Type I R-M systems proceeds via a complex pathway that includes ATP-dependent translocation of DNA before cleavage of the duplex. Recent experiments indicate that this multistep pathway provides the opportunity for host proteins to intervene after the restriction process has been initiated but before DNA is cleaved (Makovets *et al.*, 1999; reviewed by Murray, 2002).

When Bertani and Weigle (1953) discovered the first classical R-M system, they showed that irradiation of the bacteria with ultraviolet (UV) light resulted in the alleviation of restriction, as assessed by enhanced survival of phage λ previously grown on a modification-deficient host. A current interpretation of this observation is that DNA damage induces loss of restriction activity. It was proposed that DNA replication associated with either recombination or repair, as well as that following the incorporation of the wrong base, can generate unmodified target sequences within a bacterial chromosome (Makovets *et al.*, 1999). It was shown that restriction alleviation (RA) was essential for *E. coli* K-12 to grow in the presence of the base analogue 2-aminopurine (2-AP) and that RA was achieved by ClpXP-dependent degradation of HsdR, the component of the *Eco*KI complex essential for restriction but not modification (Makovets *et al.*, 1999). Makovets *et al.* (1999) proposed that mutagenesis can lead to unmodified target sequences in the bacterial chromosome. Experiments demonstrate that, when a *clpX* mutant of *E. coli* K-12 is grown in the presence of 2-AP, double-

Accepted 28 August, 2003. *For correspondence. E-mail Garry.Blakely@ed.ac.uk; Tel. (+44) 131 650 5359; Fax (+44) 131 650 8650. †Present address: Department of Biochemistry and Biophysics, University of California at San Francisco, CA 94143-2200, USA.
This paper is dedicated to the memory of Professor Jeff Schell who made early contributions to the field of restriction before his pioneering work on the plasmids of *Agrobacterium*.

strand breaks (DSBs) are made in the bacterial chromosome, and this DNA breakage is dependent on *Eco*KI (Cromie and Leach, 2001). RA of *Eco*KI in response to 2-AP, unlike that triggered by UV, is not dependent on the SOS pathway (Efimova *et al.*, 1988).

Four families of distantly related Type I R-M systems have been identified in enteric bacteria (Types IA, IB, IC and ID), each with representatives in *E. coli* (Murray *et al.*, 1982; Price *et al.*, 1987; Barcus *et al.*, 1995; Titheradge *et al.*, 1996; 2001; Lee *et al.*, 1997). *Eco*KI is the archetypal member of the IA family.

Type I R-M genes (*hsdR*, *hsdM*, *hsdS*), specifying a restriction-proficient complex, are readily transferred to recipient bacteria lacking the cognate modification activity (O'Neill *et al.*, 1997; Makovets *et al.*, 1998). A bacterial function is essential to permit the establishment of the *hsd* genes specifying EcoKI (Prakash-Cheng *et al.*, 1993). This function, essential for the efficient acquisition of the *hsd* genes for both the EcoKI (Type IA) and the EcoAI (Type IB) systems, was identified as the ClpXP protease (Makovets *et al.*, 1998). Similarly, ClpXP permits RA for EcoKI and EcoAI in response to 2-AP (Makovets *et al.*, 1999). It was proposed, therefore, that ClpXP is the key to post-translational control of the restriction activities of the Type IA and IB families of R-M systems.

Transfer of the *hsd* genes of EcoR124I, a Type IC system, is not dependent on ClpXP, and this finding suggests an alternative mechanism of control for members of the Type IC family (Kulik and Bickle, 1996). One explanation invokes the delayed assembly of the restriction-proficient EcoR124I complex (Kulik and Bickle, 1996; Firman *et al.*, 2000). Although this hypothesis would explain the ease of transmission of *hsd* genes, it would not explain a loss of restriction activity in response to radiation or chemical mutagens.

In this paper, we provide evidence that R-M systems of the Type IC and ID families have mechanisms that regulate restriction activity in response to DNA damage. For the Type IC system, we show that RA is not associated with extensive degradation of HsdR, and identify mutations in *hsdR* that impair the alleviation of restriction activity. An analysis of these mutations suggests that control of restriction activity is advantageous during growth of a restriction-proficient bacterium, and that the mechanism of control may be dissociation of HsdR from the endonuclease complex. In the absence of RA, homologous recombination is needed to maintain the integrity of the bacterial chromosome.

Results

Restriction alleviation may be a general characteristic of Type I R-M systems

2-AP has been shown to induce RA in *E. coli* strains specifying either a Type IA or a Type IB system (Efimova *et al.*, 1988; Kelleher and Raleigh, 1994). We examined the effect of 2-AP on restriction by EcoR124I and *Sty*SBLI, representatives of the Type IC and ID families respectively. The *hsd* genes specifying EcoR124I, a plasmid-borne system, were transferred to the chromosome of NK311, an *hsd* deletion derivative of *E. coli* K-12 (see *Experimental procedures*). In the resulting strain, NK402, the *hsd* genes are stably maintained, one copy per bacterial chromosome. An *E. coli* K-12/*Salmonella* hybrid, LB4037 (Bullas *et al.*, 1980), was used as a λ-sensitive, *Sty*SBLI strain.

The restriction proficiency of strains NK402 and LB4037 is reflected by 10^3- to 10^4-fold reduction in the plating efficiency of unmodified phage (λ.0). The restriction of unmodified phage on both NK402 and LB4037 was reduced ≈200-fold after treatment with 2-AP, i.e. 2-AP induced RA (Table 1, lines 1 and 2). When NK402 and LB4037 were grown in broth supplemented with 2-AP, they showed no evidence of filamentation, consistent with proficient RA (data not shown). In contrast, the strain specifying EcoRI, a Type II R-M system, was found to filament: an indication of the activation of the SOS response resulting from DNA breakage. RA has not been observed for a Type II system (Efimova *et al.*, 1988). Similarly, when the effect of 2-AP on growth of bacteria on solid medium was monitored, a Type II R-M, but not a Type I R-M system, impaired bacterial growth (Fig. 1). Our data indicate that the restriction activity of a representative of

Table 1. Restriction alleviation in response to 2-AP.

Strain	Restriction of unmodified λ after 100 min treatment[a]		
	–2-AP	+2-AP	RA
EcoR124I (NK402)	$(3.5 \pm 1.7) \times 10^3$	$(1.2 \pm 0.6) \times 10^1$	$(3.8 \pm 1.3) \times 10^2$
StySBLI (LB4037)	$(1.9 \pm 0.3) \times 10^4$	$(1.2 \pm 0.6) \times 10^2$	$(1.5 \pm 0.7) \times 10^2$
EcoR124I *clpX* (NK403)	$(4.3 \pm 1.0) \times 10^3$	8.6 ± 3.0	$(5.2 \pm 1.4) \times 10^2$
EcoR124I *recA* (NM1003b)	$(6.3 \pm 2.3) \times 10^3$	$(1.5 \pm 0.3) \times 10^1$	$(4.2 \pm 1.3) \times 10^2$

a. Cells were exposed to 2-AP for 100 min to ensure the opportunity for at least two rounds of DNA replication to generate new, unmodified target sequences by base substitution.

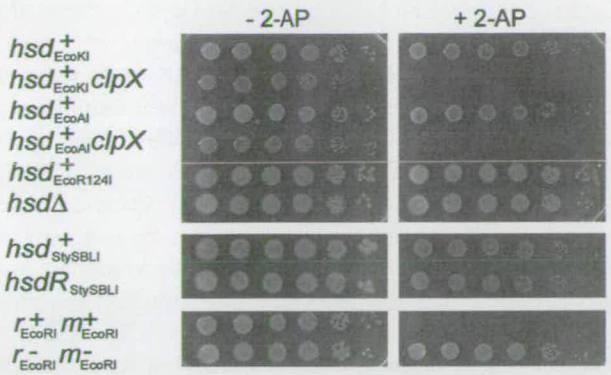


Fig. 1. Type I R-M systems do not confer sensitivity to 2-AP. Serial dilutions of the bacterial strains were spotted onto L agar and L agar supplemented with 2-AP (400 μg ml^{-1}). All strains encoding Type I R-M systems are resistant to 2-AP. A strain encoding EcoKI, or EcoAI, becomes sensitive to 2-AP in the absence of ClpX. The strain encoding EcoRI (r^+_{EcoRI} m^+_{EcoRI}), a Type II R-M system, is sensitive to 2-AP. The strains used in the experiments, in descending order, are: NK301; NK304; NK354; NK355; NK402; NK311; LB4037; NM857; NM146; 5K (see Table 5). LB4037 and its derivative NM857 grow more slowly than NK301 and its derivatives. Tests of LB4037 and NM857 were incubated for 48 h; other tests were for 24 h.

each of the four known families of Type I R-M systems is susceptible to control in response to 2-AP.

The mechanism of regulation of the restriction activity of EcoR124I differs from that for Type IA and IB systems

When a strain encoding EcoKI is grown in the presence of 2-AP the HsdR subunit is degraded. In *clpP* or *clpX* mutants, where this degradation is prevented, the bacteria remain restriction proficient. These mutants form filaments and die if the treatment with 2-AP is prolonged (Makovets *et al.*, 1999).

We examined null mutations in genes known, or predicted, to be relevant to proteolysis for an effect on the RA of EcoR124I in response to 2-AP. None of the mutations (*clpA*, *clpX*, *clpP*, *clpY*, *clpQ*, *lon*, *hflB* and *clpB*) prevented the induction of RA (data for *clpX* shown in Table 1, line 3). When grown in the presence of 2-AP (400 μg ml^{-1}) for 4–5 h, the protease-deficient derivatives of NK402 did not produce filaments, whereas the *clpX* control strain encoding EcoKI (NK304) produced filaments and died.

We looked for evidence of proteolytic control of *Eco*R124I restriction activity in NK402 by monitoring the HsdR polypeptide on Western blots. Treatment of NK402 with 2-AP led to RA, but no effect on the level, or the mobility, of HsdR was detected (Fig. 2). Our experiments found no evidence for the control of restriction activity by degradation of HsdR, although our analyses would fail to detect a minor change in the length of HsdR. The maintenance of HsdR under conditions of RA would be consistent with its loss from the R-M complex.

The isolation of mutants with impaired regulation of restriction activity

Mutations in *clpX* or *clpP* block the RA pathway for Type IA and IB R-M systems and, as a consequence, confer restriction-proficient cells with the phenotype of sensitivity to 2-AP (2-AP^s). Treatment with 2-AP is believed to lead to unmodified target sequences in the bacterial chromosome and, in the absence of RA, bacteria die from autorestriction (Makovets *et al.*, 1999; Cromie and Leach, 2001). A screen for 2-AP^s bacteria should produce mutants deficient in their regulation of restriction activity. To test this idea, we treated the strain encoding EcoKI (NK301) with a mutagen (nitrosoguanidine) and screened colonies (2700) for sensitivity to 2-AP. Twenty-three 2-AP^s mutants were isolated, and five of them lost their 2-AP^s phenotype when restriction activity was destroyed by a mutation in *hsdR* (*hsdR*::*kan*). The 2-AP sensitivity of four of the five strains was caused by mutations in the *clpPX* region, as assessed by P1 co-transduction frequencies of 70–80% between *tsx*::Tn*10* and 2-AP^R. The remaining mutation mapped close to *hsd* on the basis of its linkage to *zjj*::Tn*10*. These results validate this experimental approach.

In our experiment with NK402, in which the *hsd* genes that specify EcoR124I are located within the *lac* operon (see *Experimental procedures*), 3000 colonies were screened. For only one of 26 2-AP^s mutants was the 2-AP^s phenotype dependent on restriction proficiency. The mutation in this strain (NK445) showed tight linkage to *zaj*::Tn*10* in the *lac* region, consistent with its location in an *hsd* gene, rather than an unlinked gene involved in regulation.

Characterization of the NK445 mutant

The most likely location for an *hsd* mutation that disturbs regulation of the restriction activity of the EcoR124I com-

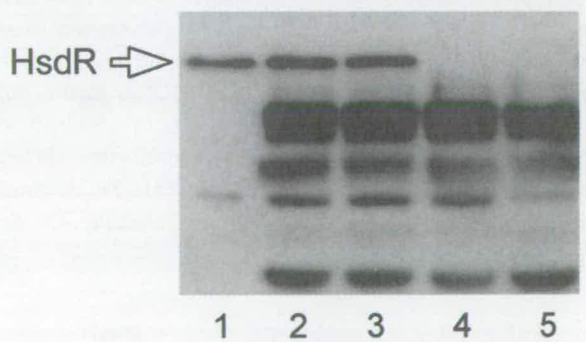


Fig. 2. The steady-state level of HsdR of EcoR124I does not change during treatment with 2-AP. A polyclonal antibody raised against *Eco*R124I was used to detect HsdR. Track (1) purified EcoR124I; the other tracks use the following cell extracts: (2) NK402 (*hsd*$^+$); (3) NK402 + 2-AP; (4) NK413 (*hsdRΔM*$^+$*S*$^+$); (5) NK311 (*hsdΔ*).

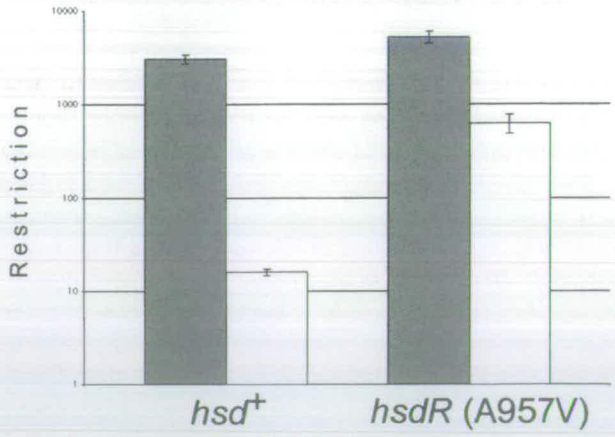


Fig. 3. The amino acid change HsdR A957V impairs RA. The shaded bars give the value for restriction in the absence of 2-AP and the open bars in the presence of 2-AP. Restriction was measured by the reduction in titre of λ*vir*.0 on the test strain relative to that on a restriction-deficient host. Strain NK445 has enhanced restriction and reduced RA compared with NK402.

plex is within *hsdR*. The *hsdR* gene of the RA-deficient strain, NK445, was cloned in pUC19 (pAT43). The nucleotide sequence of *hsdR* in pAT43 was found to differ from that of wild type (pNK19) by the change in codon 957 from GCT to GTT. Codon 957 of pAT43 was reverted to the wild-type sequence by site-directed mutagenesis (SDM). Both pAT43 and revertants with wild-type nucleotide sequence were used to transform the *hsdR*$^-$*M*$^+$*S*$^+$ strain, NK413. All transformants were restriction proficient, but only those with the mutant *hsdR* sequence were sensitive to 2-AP; therefore, the mutation in codon 957, which results in the substitution of valine for alanine, is responsible for the 2-AP^s phenotype.

Strain NK445 restricts unmodified phage λ with a slightly higher efficiency than the parent strain, NK402, and is deficient in its ability to alleviate restriction in response to 2-AP (Fig. 3). In both these respects, the phenotype of NK445 resembles that of a *clpX* derivative of a restriction-proficient *E. coli* K-12 (Makovets *et al.*, 1999). Our experiments indicate that the amino acid substitution in HsdR impairs RA. Codon 957 is in the carboxy-terminal region of HsdR, the region required for interaction between HsdR and the methyltransferase component of EcoKI. The substitution A957V in HsdR could impair RA by enhancing the stability of the EcoR124I complex.

Constitutive RA in a modification-deficient strain

The methyltransferase activity of EcoKI can be abolished by a substitution (F269G) in the NPPF/Y motif of the HsdM subunit, but this mutation has no effect on the binding of the essential cofactor AdoMet (Willcock *et al.*, 1994). The EcoKI complex retains some restriction activity (Doronina and Murray, 2001) but, despite the absence of methyltransferase activity and the consequent unmodified chromosome, the *hsdM* mutant survives in the presence of a wild-type *hsdR* gene (Makovets *et al.*, 1999). In this modification-deficient strain, HsdR is destroyed by ClpXP-dependent degradation to produce a restriction-deficient phenotype (Makovets *et al.*, 1999). This finding implies that RA for EcoKI is independent of an external stimulus, hence 'constitutive' RA. We made a substitution (Y305G) in the NPPF/Y motif of the HsdM subunit of EcoR124I in the *hsdR*$^-$ *M*$^+$*S*$^+$ strain NK413: the mutation was made in the absence of an *hsdR* gene to ensure viability. The resulting modification-deficient strain, LP4, remained viable in the presence of a wild-type *hsdR* gene within a λ prophage, but was restriction deficient, a phenotype consistent with constitutive RA. We checked whether the viability of this *hsdR*$^+$*M*$^-$(Y305G)*S*$^+$ strain was dependent on RA. In this experiment, we used integration-proficient λ phages carrying either mutant or wild-type *hsdR* in the presence of a selective marker conferring resistance to tetracycline. λ phages carrying the mutant *hsdR* (A957V) gene, in contrast to those carrying the wild-type *hsdR*, formed clear, rather than turbid, plaques on the *hsdM*(Y305G)*S*$^+$ strain, LP4, consistent with death of the infected cells. λ*hsdR*$^+$, tetracycline-resistant lysogens were readily formed, whereas only occasionally were tetracycline-resistant lysogens formed using the λ*hsdR* (A957V) phage (Table 2). An analysis of 13 of the λ*hsdR* (A957V) lysogens indicated that each included a suppressor mutation.

We conclude that an EcoR124I complex that includes catalytically inactive HsdM subunits retains endonuclease activity. In the presence of wild-type HsdR the complex is susceptible to RA, whereas one that includes HsdR (A957V) is not and, as a consequence, is able to attack the unmodified bacterial chromosome.

The isolation and characterization of additional mutations in hsdR

Base changes were made in the region identified by the mutation *hsdR* (A957V). The mutations were made by SDM of pNK25, a derivative of pBRINT that includes *hsdR* of EcoR124I in the absence of *hsdM* and *hsdS*. The *hsdR*$^-$ *M*$^+$*S*$^+$ strain NK413 was transformed by pNK25 and by mutant derivatives. All the transformed strains were restriction proficient. They were grown in the presence of ampicillin and tested for their ability to form colonies in the presence of 2-AP. NK413 transformed with pNK25 (*hsdR*$^+$), or derivatives with the substitution S948A or R959D, were resistant to 800 μg ml^{-1} 2-AP, as expected for strains proficient in RA. For the substitution A957K, <1% of viable cells formed colonies on medium supple-

Table 2. The relevance of methyltransferase activity to the establishment and maintenance of λ*hsdR* lysogens.

Bacterium	Phage	Experiment 1[a] % lysogens	Experiment 2[b] Relative titre of lysogens
NK311 (*hsdΔ*)	λvector	1.0	1.0
	λ*hsdR*$^+$	0.5	1.1
	λ*hsdR* (A957V)	1.0	1.0
NK413 (*hsdR*$^-$*M*$^+$*S*$^+$)	λvector	0.5	1.0
	λ*hsdR*$^+$	0.5	0.9
	λ*hsdR* (A957V)	0.5	0.8
LP4[*hsdR*$^-$*M* (Y305G)*S*$^+$][c]	λvector	0.5	0.9
	λ*hsdR*$^+$	0.2	0.9
	λ*hsdR* (A957V)	0.001	0.001
LP4[*hsdR*$^-$*M* (Y305G)*S*$^+$][c]	λvector	0.5	0.6
	λ*hsdR*$^+$	0.5	0.5
	λ*hsdR* (A957V)	0.001	0.002

a. Cells infected at a multiplicity of infection (MOI) of 5, plated in the presence of sodium citrate to prevent adsorption of liberated phage. % lysogens expressed relative to surviving cells.
b. Cells infected at an MOI of 0.5, sodium citrate added after adsorption and at the time of plating. Titres of Tet^R colonies assayed.
c. Different cultures of the same strain.

mented with 400 μg ml^{-1} 2-AP and, for F956A, the transformed cells became sensitive to 2-AP when the concentration was raised to 800 μg ml^{-1}. The sensitivity to 2-AP conferred by substitutions A957K and F956A is consistent with a defect in RA.

A more sensitive test of the effects of the substitutions in HsdR was made taking advantage of LP4, the *hsdRΔ* strain with an additional mutation in *hsdM* that blocks methyltransferase activity. In this assay, the efficiency of establishment of pNK25 and derivative plasmids was monitored after transformation of NK311 (no *hsd* genes), NK413 (*hsdRΔ hsdM*$^+$ *hsdS*$^+$) and LP4 (*hsdRΔ hsdM* (Y305G) *hsdS*$^+$). In each transformation assay, a kanamycin-resistant marker plasmid (pBRK) was mixed with the ampicillin-resistant test plasmid. The ratios of Amp^R to Kan^R transformants were monitored; ratios for NK311 and NK413 were very similar irrespective of the plasmid under test; those for LP4 were reduced for pNK25 derivatives with the substitutions A957K and F956A, most particularly for A957K. For pNK25 *hsdR* (A957K) in each of 10 experiments, the ratio of transformants of strain LP4 was reduced, generally by at least 100-fold. The occasional Amp^R transformants were slow to appear. Of seven transformants tested, six were resistant to 2-AP, a finding consistent with the survival of LP4 pNK25 *hsdR* (A957K) as a consequence of suppressor mutations.

The *hsdR* gene of pNK25 (A957K) was subcloned in λNM1151 (see *Experimental procedures*), and the mutation transferred to the chromosome of NK402 by homologous recombination. The resulting strain, NM996, was more sensitive to 2-AP (400 μg ml^{-1}) than wild type, restricted λ*vir*.0 with a similar efficiency to that of NK445 and showed that RA was reduced from ≈400 (Table 1, line 1) to 3.7 (±1.1). The additional mutations that confer the 2-AP^s phenotype (A957K and F956A) support the concept that the region identified by the mutation in the RA deficient strain, NK445, is relevant to the control of restriction activity.

The transmission of hsd *genes*

When the *hsd* genes of *E. coli* K-12 are transferred to a modification-deficient recipient, ClpXP is essential to alleviate restriction until the recipient chromosome has become modified. In the absence of ClpXP the recipient cell is killed (Makovets *et al.*, 1999).

The restriction phenotype of NK445 supports the hypothesis that the amino acid substitution A957V in HsdR impairs post-translational regulation of the restriction activity of EcoR124I. An extension of this analogy between EcoR124I and EcoKI would predict that this substitution could lead to the loss of control that is essential for the establishment of the *hsd* genes specifying *Eco*R124I after their transmission to a modification-deficient recipient strain. The gain and loss of the *hsd* genes of NK445 was monitored by P1-mediated co-transduction of a closely linked Tn*10* marker gene. The co-transduction frequencies (Table 3) do not support the prediction; the mutation in *hsdR* has no detectable effect on either the transmission or the loss of the genes encoding the EcoR124I complex. The control of restriction activity by assembly may suffice for the establishment of the EcoR124I system in a naïve environment (Kulik and Bickle, 1996; Firman *et al.*, 2000).

The significance of the modification activities of Type I R-M systems

The *hsd* genes of the EcoKI system are rarely established after transmission by P1 transduction to a recipient that lacks ClpXP, whereas the establishment of the *hsd* genes specifying EcoAI is effective in the absence of ClpXP

Table 3. (A) Co-transfer of $hsd_{EcoR124I}$ with *zaj-3053*::Tn*10* by P1 transduction.

Transduction			
Donor	Recipient	No. of Tet^R colonies tested[a]	% Linkage of Tn*10* to *hsd*
hsdR$^-$*M*$^+$*S*$^+$ *zaj*::Tn*10* (NK413)	*hsd*Δ (NK311)	100	26
hsd$^+$ *zaj*::Tn*10* (NK419)	*hsd*Δ (NK311)	85	33
hsdR A957V *zaj*::Tn*10*(NK446)	*hsd*Δ (NK311)	66	27

(B) Loss of $hsd_{EcoR124I}$ assessed by co-transduction of *hsd*Δ with *zaj-3035*::Tn*10*.

Transduction			
Donor	Recipient	Number of Tet^R colonies tested	% Linkage of Tn*10* to *hsd*Δ
*hsd*Δ *zaj*::Tn*10* (NM1002)	*hsd*$^+$ (NK402)	72	18
*hsd*Δ *zaj*::Tn*10* (NM1002)	*hsdR* A957V (NK445)	72	25

a. The restriction phenotype of each Tet^R Cm^R co-transductant corresponded with that of the donor strain (i.e. r$^-$m$^+$, r$^+$m$^+$ or r$^+$m$^+$ 2-AP^s).

(Makovets *et al.*, 1998). A plausible contribution to this difference could be the very different characteristics of their methyltransferase activities. For both systems, methylation can be catalysed by either the R-M complex or the complex that lacks HsdR, but of likely significance is the finding of Suri and Bickle (1985) that, although the complexes of the EcoAI system are efficient methyltransferases if the substrate is unmodified DNA, the complexes of the EcoKI system have a pronounced preference for hemimethylated DNA (Suri and Bickle, 1985). *In vitro* experiments have indicated that the modification component of EcoR124I has a 100-fold preference for hemimethylated DNA (Taylor *et al.*, 1993), whereas the R-M complex has only a modest preference when assayed in the presence of ATP (Price *et al.*, 1987).

We tested the efficiency with which the modification enzyme of the EcoKI, EcoAI and EcoR124I systems methylate unmodified DNA *in vivo*. Our substrate was a *ral* mutant of phage λ. Ral antagonizes restriction and enhances modification by Type IA R-M systems (Zabeau *et al.*, 1980; Loenen and Murray, 1986). Methylation activity was measured after a single round of growth of phages in r$^-$m$^+$ bacteria by comparing the titre of the progeny on restriction-proficient and restriction-deficient bacteria. The modification systems of EcoR124I and EcoAI, but not EcoKI, provide efficient modification of phage genomes in the absence of Ral (Table 4). We conclude that *in vivo* the modification activity of EcoR124I more closely resembles that of EcoAI (Type IB) than that of EcoKI (Type IA). This efficient *de novo* modification of DNA should help in the establishment of *hsd* genes in a naïve environment.

Homologous recombination is required for viability of a strain defective in RA

While measuring the growth rates of strains specifying EcoR124I (see *Experimental procedures*), we observed that the RA-deficient strain, NK445, had a longer doubling time than the RA-proficient strain NK402 (29 min and 23 min respectively). Despite the apparent difference in doubling time between the two strains, the mean cell lengths for exponentially growing cells from both cultures were shown to be the same, 2.9 μm. Note that cells from slower growing cultures are shorter than cells from faster growing cultures (Donachie, 1968). One explanation for our observations is that autorestriction of chromosomal DNA occurs in NK445, and that such DSBs ultimately lead to the death of some cells within the culture. If this explanation is correct, we would expect that the DSBs generated in NK445 would require RecA to effect repair. Consistent with this observation, the introduction of a *recA* allele into NK445 increased the doubling time and reduced the viability of this strain compared with the *recA* derivative of the RA-competent strain NK402 (95 min and 29 min respectively). The detrimental effect of the *recA* mutation in the RA-deficient strain, NK445, can be seen in Fig. 4. At the concentrations of 2-AP used in our plating experiments (400 μg ml^{-1}), NK445 *recA* (NM1004b) fails to grow. The data shown in Table 1, line 4, confirm that RA for *Eco*R124I occurs in response to 2-AP in the absence of RecA.

Table 4. Modification of unmethylated phage DNA *in vivo*.

Bacteria			E.o.p. of test phage[a]		
r$^-$m$^+$	r$^+$m$^+$	System	λ*ral*$^-$	λ*ral*$^+$	E.o.p. of λ*vir*.0[b]
NK351	NK301	*Eco*KI	0.00017	0.90	0.0003
NM863	NK354	*Eco*AI	0.78	0.81	0.01
NK413	NK402	*Eco*R124I	0.98	0.93	0.0005

a. Efficiency of plating (e.o.p.) [titre of phage on an r$^+$ (restriction-proficient) strain relative to titre on r$^-$] was assessed after a single round of growth of unmodified phage in r$^-$m$^+$ bacteria.
b. E.o.p. of λ*vir*.0 (unmodified λ) on the relevant r$^+$m$^+$ strain.
λ has five targets for EcoKI, one for EcoAI and 14 for EcoR124I.

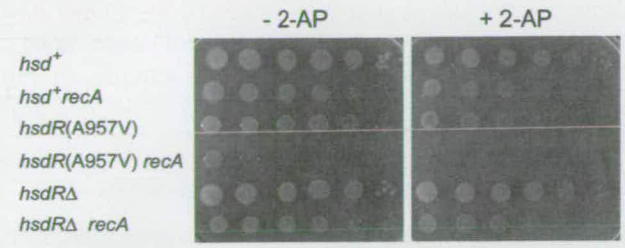


Fig. 4. Recombination rescues an RA-deficient strain. Serial dilutions of bacterial cultures specifying wild-type or mutant EcoR124I were spotted on L agar and L agar + 2-AP (400 μg ml^{-1}) and incubated overnight. The mutation *hsdR* (A957V) increases sensitivity to 2-AP; this is greatly enhanced in a *recA* strain. The bacterial strains used, in descending order, are: NM1003a; NM1003b; NM1004a; NM1004b; NM1008a; NM1008b (see Table 5).

Failure to repair DSBs would be expected to lead to degradation of the chromosome by exonuclease activity (Skarstad and Boye, 1993). To determine whether the reduced viability of NK445 *recA* was caused by degradation of chromosomal DNA, we used fluorescence microscopy to check the integrity of nucleoids in exponentially growing cells. Figure 5 shows examples of cells from cultures of NK445 and the NK445 *recA* derivative.

The majority of cells from cultures of NK445 contained defined, well-stained nucleoids. There was a small proportion of anucleate cells (≈ 1% of the population), which might explain the reduced growth rate of this strain (Fig. 5A). There was no evidence of cell filamentation in NK445, indicating that the SOS response was not induced. However, a striking phenotype was

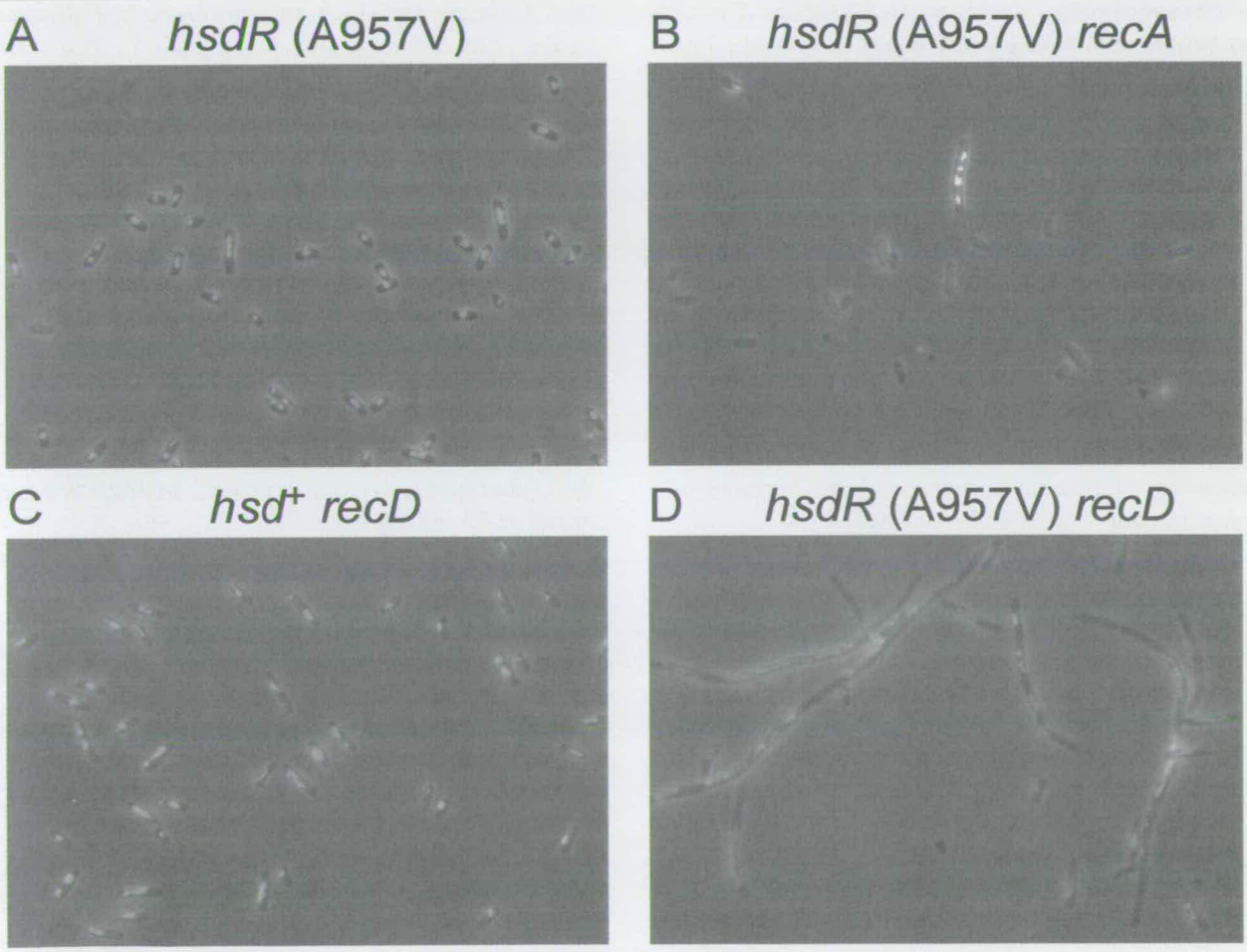


Fig. 5. Combined phase-contrast and fluorescence micrographs of cells from exponentially growing cultures showing nucleoids stained with DAPI.
A. RA-deficient strain NK445.
B. NK445 *recA* (NM1004b).
C. RA-proficient strain NK402 *recD* (NM1034).
D. NK445 *recD* (NM1035).
Note the absence of nucleoid staining in many of the cells from strain NK445 *recA* (B) and the block in cell division that leads to extensive filamentation in strain NK445 *recD* (D).

observed with strain NK445 *recA*; ≈41% of the cells contained no obvious nucleoids, whereas the majority of the remaining cells showed faint and diffuse staining of nucleoids that was consistent with DNA degradation (Fig. 5B). In comparison, cultures of the *recA* derivative of the RA-proficient strain NK402 contained ≈9% anucleate cells, a frequency consistent with previous observations of *recA* strains (Skarstad and Boye, 1993). Autorestriction of chromosomal DNA followed by exonuclease-mediated degradation in many of the cells from strain NK445 *recA* would be consistent with the observed dramatic reduction in growth rate and viability. The catastrophic consequences of irreparable DSBs was further supported by the finding that *recB*::Tn*10* and *recC*::Tn*10* derivatives of NK445 grew even more slowly than NK445 *recA*.

To ascertain whether chromosome degradation in the *recA* derivative of NK445 was a consequence of RecBCD exonuclease function, we made an NK445 *recD* strain. This RA-deficient strain had a doubling time of 47 min compared with a doubling time of 31 min for the RA-proficient NK402 *recD* strain. A longer doubling time for NK445 *recD* was not anticipated because *recD* strains have been reported to be highly viable (Chaudhury and Smith, 1984). They are also hyper-recombinogenic, so there should not have been a deficiency in DSB repair. Indeed, cells from a *recD* derivative of the RA-proficient strain NK402 (Fig. 5C) had normal morphology and a mean length of 3.1 μm, a figure comparable with that of the parental NK402 strain (2.9 μm). Microscopic examination of samples from cultures of NK445 *recD* demonstrated that many of these cells were highly filamentous and contained aberrant nucleoids (Fig. 5D). Therefore, the processing of DSBs in the absence of RecD impairs cell division by inducing the SOS response.

Discussion

Chromosome integrity

Chromosome breakage is caused by a variety of agents that damage DNA. In this paper, our focus is autorestriction – the potential 'self-inflicted' damage resulting from the action of a resident restriction enzyme should the chromosome acquire unmodified target sequences. Vulnerable target sequences may be generated (i) by DNA repair processes that lead to a duplex region of newly synthesized DNA (Makovets *et al.*, 1999; see Murray, 2000); (ii) as the result of a mutation that creates a new target (Makovets *et al.*, 1999); or (iii) should the normal methylation process fail. We find that RA is a means of preventing autorestriction by EcoR124I and, should this fail, recombination serves to restore the integrity of the chromosome. A repair role for recombination has been demonstrated following autorestriction by the EcoKI system after treatment with 2-AP (Cromie and Leach, 2001). Our experiments explore chromosome integrity in the absence of mutagenic treatments.

Control of autorestriction by Type I R-M systems

Type I restriction enzymes share a complex restriction pathway in which they cleave DNA thousands of base-pairs away from an unmodified target sequence, after an ATP-dependent process in which DNA is translocated through the enzyme bound to its target sequence (for recent reviews, see Murray, 2000; 2002; Dryden *et al.*, 2001; Bourniquel and Bickle, 2002). Control of EcoKI restriction activity, in response to DNA-damaging agents, results from the ClpXP-dependent destruction of HsdR (Makovets *et al.*, 1999). The susceptibility of HsdR to proteolysis requires the EcoKI complex to retain its ability to bind its target sequence and hydrolyse ATP (Makovets *et al.*, 1999; Doronina and Murray, 2001). The interval between the recognition of an unmodified target sequence in the bacterial chromosome and breakage of DNA creates an opportunity for the exposure of a signal within HsdR: in consequence, HsdR is destroyed before the chromosome is broken.

EcoR124I and StySBLI are both subject to RA in response to 2-AP (Table 1). Our experiments, using a methyltransferase-deficient mutant of EcoR124I, show that wild-type HsdR is tolerated, but an HsdR subunit that prevents efficient RA is not. For EcoR124I, as for EcoKI, constitutive RA permits the survival of the modification-deficient, restriction-proficient mutant. Therefore, in the absence of modification, both EcoR124I and EcoKI can distinguish 'self' DNA from 'non-self', although the mechanism of RA is different.

The acquisition of genes encoding EcoKI is dependent on post-translational regulation by ClpXP. Two factors other than transcriptional, or post-transcriptional, regulation of expression of *hsd* genes could help the establishment of *hsd* genes in a naïve environment. First, an assembly pathway in which the restriction complex is not made until the recipient chromosome has been methylated (Kulik and Bickle, 1996; Firman *et al.*, 2000). Secondly, a modification enzyme that behaves as an effective *de novo* methyltransferase will greatly enhance the efficiency of the protective process. The modification activity of EcoKI is that of a maintenance methyltransferase, that of EcoAI a *de novo* methyltransferase (Suri and Bickle, 1985). Both the instability of the *Eco*AI complex (Suri and Bickle, 1985) and the efficient methylation of unmethylated targets could contribute to the relative ease of transmission of the *hsd* genes for EcoAI in the absence of *clpX* (Makovets *et al.*, 1998). When *hsdM* and *S* from *E. coli* K-12 are transferred to a modification-deficient strain, it

takes many generations before the descendants of the recipients become capable of methylating phage genomes that enter the cell (Makovets, 1999). It would appear that, during this period of 10 h, the modification activity within the cell is required to methylate the bacterial chromosome. For EcoR124I, the assembly pathway may ensure that the restriction complex is not made until the recipient chromosome has been methylated (Kulik and Bickle, 1996; Firman *et al.*, 2000). The effective *de novo* methyltransferase activity in the absence of HsdR (Table 4) will enhance the efficiency of this protective process.

The genes specifying EcoKI are readily deleted in the presence or absence of ClpXP (O'Neill *et al.*, 1997; Makovets *et al.*, 1998). Experiments for EcoR124I (Table 3B) show that, even when RA is impaired, the *hsd* genes are readily lost. Currently, there is no proven explanation for the survival of bacteria after the loss of *hsd* genes in the absence of RA. EcoR124I readily loses one HsdR subunit *in vitro* (Janscak *et al.*, 1998). Dissociation of HsdR subunits from the Type I R-M complex, as the concentration of protein decreases, may suffice to prevent autorestriction. If the mutation in NK445 enhances the stability of the EcoR124I complex, this effect on stability may be relevant only when the complex is bound to DNA.

An alternative mechanism for RA

Our mutant hunt failed to identify a host function involved in RA of EcoR124I, although a parallel hunt using a strain encoding EcoKI identified four mutations that were candidates for *clpX* and *clpP*. Our failure to identify a regulatory gene for the EcoR124I system can be explained should inactivation of the relevant function be detrimental to the bacterium; alternatively, it may indicate that the mechanism of RA for EcoR124I is independent of external functions. Current evidence implicates the assembly pathway as the means of controlling restriction activity of EcoR124I during the establishment of these *hsd* genes in a naïve environment (Kulik and Bickle, 1996; Firman *et al.*, 2000). If assembly is the mechanism of control for EcoR124I during establishment, then disassembly, i.e. the dissociation of HsdR from the protein–DNA complex, could be the mechanism of RA. In the case of EcoR124I, DNA translocation on the host chromosome could destabilize the R-M complex, whereas for EcoKI, it would expose the signal for ClpXP to attack and degrade HsdR.

Alignment of HsdR polypeptide sequences for representatives of the four families of Type I systems indicates a common organization (Titheradge *et al.*, 1996). For EcoKI, the carboxy-terminal 155 amino acids are necessary for the association of HsdR with the methyltransferase (Davies *et al.*, 1999). The amino acid change A957V is within the carboxy-terminal region of EcoR124I HsdR. If this change was to affect interaction of HsdR with the methyltransferase, it could impair RA by enhancing the stability of EcoR124I. This mechanism implies that the wild-type enzyme is destabilized when it attempts to translocate unmodified DNA of the bacterial chromosome. Dissociation of HsdR from the complex leaves a functional methyltransferase bound to the unmodified target sequence. The mechanism proposed for EcoR124I shares a critical feature with EcoKI, in which dissociation of the protein–DNA complex follows initiation of the ATP-dependent restriction pathway on the bacterial chromosome (Makovets *et al.*, 1999). The means of differentiation between the nucleoid and newly acquired DNA remains to be determined (for a review, see Murray, 2002).

Homologous recombination, chromosome integrity and a role for RecD in preventing induction of the SOS response

Our analysis of strains with reduced RA demonstrates that chromosome integrity and viability are impaired in only a small proportion of cells during exponential growth. The essential role for homologous recombination in maintaining chromosome integrity after autorestriction is clearly shown by the large numbers of anucleate cells (41%) in cultures of the NK445 *recA* strain. Previous studies have shown that *recA* strains show reduced viability with ≈10% of cells being anucleate. Chromosome degradation in *recA* strains is mediated by the exonuclease activity of the RecBCD complex acting at DSBs (Skarstad and Boye, 1993).

A requirement for RecBCD in RA-deficient strains is also clearly shown by the extremely slow growth of strains NK445 *recB* and NK445 *recC*. The role of RecBCD in repairing DSBs predicts that crossing over during recombination will generate chromosome dimers that need to be resolved to monomers, by Xer site-specific recombination, before segregation (Blakely *et al.*, 1993; Steiner and Kuempel, 1998). This prediction was confirmed by the large number of filamentous cells and longer doubling times of RA-deficient *xerC* strains (data not shown).

An unexpected result from our work was the discovery that the absence of RecD had a marked effect on cell division when RA was deficient. We anticipated that removing RecD and the consequent exonuclease activity would prevent the generation of anucleate cells. *recD* mutants have generally been reported to be viable and hyper-recombinogenic (Chaudhury and Smith, 1984). However, the *recD*, RA-deficient strain produced large numbers of filamentous cells that were associated with induction of the SOS response. There are at least two explanations for SOS induction in NK445 *recD*: (i) failure to degrade regressed replication forks could stabilize

unmodified DNA that would then be a substrate for restriction; (ii) restriction of other unmodified chromosomal sites, followed by RecBC-mediated unwinding at DSBs, could generate single-stranded DNA, which would stimulate the co-protease activity of RecA. Both explanations are consistent with the known role of RecBC in generating single-stranded DNA that stimulates the SOS response (Chaudhury and Smith, 1985; Rinken *et al.*, 1992).

Type II R–M systems

The evidence indicates that Type II systems are not subject to RA (Efimova *et al.*, 1988), but such systems comprise two independent enzymes. Transcriptional regulation of the restriction activity is well documented after acquisition of genes specifying a Type II system (Tao *et al.*, 1991; Ives *et al.*, 1992). Both the relative concentration of the methyltransferase and the endonuclease, and their respective affinities for the target sequences, offer scope for the avoidance of chromosome breakage. Current information identifies autorestriction only when either the nature of the endonuclease or its concentration is disturbed.

The significance of DNA ligase in re-establishing the integrity of a DNA strand in response to nicks generated by a Type II endonuclease has been emphasized. In the absence of sufficient DNA ligase, homologous recombination was essential for the maintenance of viability (Heitman *et al.*, 1999). Recombinational repair is consistent with the need to restore replication forks that have stalled when they encounter a discontinuity in the template strand. Heitman *et al.* (1999) suggested that DNA ligase, rather than recombination, may repair DSBs inflicted by *Eco*RI.

Much evidence indicates that bacteria die if they lose their Type II R–M genes: this selfish behaviour ensures their maintenance (see Kobayashi, 2001). When Type II R-M genes are lost, the concentration of both the endonuclease and the methyltransferase will be reduced as the cells grow, some target sequences will fail to acquire modification, and the bacterial chromosome becomes vulnerable to the residual endonuclease. Under these conditions, the SOS response is triggered, and homologous recombination repairs the breaks (Handa *et al.*, 2000).

The idea that RA may be important during growth of bacteria encoding an R–M system is encouraged by current awareness of the relevance of homologous recombination in the repair of stalled replication forks (McGlynn and Lloyd, 2000; Seigneur *et al.*, 2000): the same process that repairs discontinuities in a DNA strand could itself generate unmodified targets and make bacterial chromosomes susceptible to breakage by host restriction enzymes.

Experimental procedures

Bacterial strains, plasmids, phages and general methods

The bacterial strains are listed in Table 5. DH5α was the strain used to recover and amplify recombinant plasmids. *Eco*R124I is specified by plasmid R124 (Hedges and Datta, 1972). The sequence of the *hsd* genes encoding *Eco*R124I has been reported previously (Price *et al.*, 1989), and was used to guide the cloning experiments in this paper. The relevant 14 kb HindIII fragment of plasmid R124 was recovered in a λ vector, amplified and cut with *Bsr*BI. The resulting 8 kb segment that includes *hsdR*, *M* and *S* was inserted between the HindIII and SmaI sites of pBRINT to produce pNK19. pNK22 is a derivative of pNK19 in which *hsdR* has been removed by the deletion of a *Pst*I fragment, and pNK25 is a derivative in which *hsdM* and *S* have been removed by deletion of the DNA between the HindIII and EcoRI targets.

In pBRINT (Balbas *et al.*, 1996), the cloning sites are adjacent to a gene conferring Cm^R; the cloning sites and *cat* interrupt the *lacZ* gene. The flanking *lacZ* sequences provide homology with the bacterial chromosome and, consequently, facilitate co-transfer of cloned genes and the genetic tag from linearized plasmid DNA to the chromosome of the *recB recC sbcC* strain, JC7623. The cloned *hsd* genes were transferred from JC7623 to the chromosome of NK311 by P1 transduction. The hsd^+ strain, NK402, was derived via pNK19 and the r^-m^+ strain, NK413, via pNK22. pBRK is a kanamycin derivative of pBR322 (Makovets *et al.*, 1998).

A mutation was made in *hsdM* of pNK22 by SDM, changing Y305 to G. The mutation was transferred to the chromosome of NK311 to make strain LP4. Mutations in *hsdR* were made in pNK25. A BamHI fragment including one such mutation was transferred to the λ vector NM1151 (Murray, 1983), and the integration-deficient phage was used to replace the wild-type sequence of NK402 with the mutation identified in NM996.

The 4.9 kb EcoRI–BamHI fragment that includes the *hsdR* gene of NK445 was cloned in pUC19 to produce pAT43 for sequence determination and SDM. This same fragment and its homologue from NK402 were transferred to the integration-proficient λ vector TXF97 (St Pierre and Linn, 1996) for genetic experiments dependent on expression of the *hsdR* gene.

The QuikChange™ kit (Stratagene) was used for SDM. Media, general methods for cloning, handling of bacteria and their phages were as described previously (Makovets *et al.*, 1998), except that RA was assessed after 100 min rather than 60 min. λ*ral cI* and λ*cI* (Zabeau *et al.*, 1980) were used to assess modification. Growth rates were based on OD_{600} readings taken at 15 min intervals during exponential growth from an OD of 0.02–0.2.

Analysis of proteins

The methods were as described in a previous paper (Makovets *et al.*, 1999). The polyclonal antibody raised against EcoR124I and the EcoR124I protein used as a marker, were kindly provided by Pavel Janscak and Keith Firman.

Table 5. Bacterial strains (derivatives of *E. coli* K-12).

Strain	Relevant genotype	Reference or origin
LE451	*rac-0 recA srl*::Tn*10*	Diaz *et al.* (1979)
CAG12148	*tsx-247*::Tn*10*	Singer *et al.* (1989)
CAG18091	*zaj-3053*::Tn*10*	Singer *et al.* (1989)
CAG12135	*recD1901*::Tn*10*	Singer *et al.* (1989)
DH5α	*endA hsdR17 recA gyrA*	Hanahan (1983)
JC7623	*recBC sbcC*	Kushner *et al.* (1971)
N2101	*recB268*::Tn*10*	Lloyd *et al.* (1987)
N2103	*recC266*::Tn*10*	Lloyd *et al.* (1987)
5K	*hsdR514*	Hubacek and Glover (1970)
NM146	5K *ecoRIR*$^+$ and *M*$^+$	Brammar *et al.* (1974)
NM863	Δhsd_{EcoKI} hsd_{EcoAI} $R^-M^+S^+$	Makovets *et al.* (1998)
NK301	hsd^+_{EcoKI} *rac-0 gyrA96*	Makovets *et al.* (1999)
NK304	NK301 *clpX*::*kan*	Makovets *et al.* (1999)
NK310	NK301 *hsdR514*	NK301 + P1.5K
NK311	NK301 Δ(*mrr-hsd-mcrBC*)	Makovets *et al.* (1999)
NK351	NK301 *hsdR*(A619V)	Makovets *et al.* (1999)
NK354	Δhsd_{EcoKI} hsd^+_{EcoAI}	Makovets *et al.* (1999)
NK355	NK354 *clpX*::*cat*	Makovets *et al.* (1999)
NK402	Δhsd_{EcoKI} *lac*::(*hsd*$^+_{EcoR124I}$ *cat*)	Derivative of NK311
NK403	NK402 *clpX*::*kan*	This paper
NK413	Δhsd_{EcoKI} *lac*::($hsd_{EcoR124I}M^+S^+$ *cat*)	This paper
NK419	NK402 *zaj-3053*::Tn*10*	NK402 + P1.CAG18091
NK420	NK413 *zaj-3053*::Tn*10*	NK413 + P1.CAG18091
NK445	NK402 *hsdR*(A957V)	Mutagenesis
NK446	NK445 *zaj-3053*::Tn*10*	NK445 + P1.NK420
LP4	NK413 *hsdM*(Y305G)	This paper, by SDM
LB4037	Δhsd_{EcoKI} $hsd^+_{StySBLI}$	Bullas *et al.* (1980)
NM857	LB4037 *hsdR* (K333T)	Titheradge *et al.* (2001)
NM996	NK402 *hsdR* (A957K)	This paper, by SDM
NM1002	NK311 *zaj*::Tn*10*	NK311 + P1.CAG18091
NM1003 a and b	NK402 *srl*::Tn*10* *rec*$^+$ and *recA*	NK402 + P1.LE451
NM1004 a and b	NK445 *srl*::Tn*10* *rec*$^+$ and *recA*	NK445 + P1.LE451
NM1008 a and b	NK413 *srl*::Tn*10* *rec*$^+$ and *recA*	NK413 + P1.LE451
NM1034	NK402 *recD*::Tn*10*	NK402 + P1.CAG12135
NM1035	NK445 *recD*::Tn*10*	NK445 + P1.CAG12135

Nucleoid staining and microscopy

Exponentially growing cells were treated with chloramphenicol (100 μg ml^{-1}) for 15 min at 37°C to condense nucleoids before staining with DAPI using the method of Hiraga *et al.* (1989). Combined phase-contrast and fluorescent images were captured using a Zeiss Axioplan II microscope with a Hamamatsu Orca CCD camera. Cell length measurements were obtained using Improvision OPENLAB software.

Acknowledgements

We are indebted to colleagues for their help and useful discussions, particularly Costel Atanasiu, John Connelly, Angela Dawson, David Dryden, Emmitt Jolly, David Leach, Vladimir Nekrasov and Mary O'Neill, to Alexander Gann and Thomas Bickle for constructive comments on our manuscript, and to Alix Fraser for its careful preparation. The work was supported by MRC (N.E.M.), The Wellcome Trust (G.B.) and a Darwin Trust Scholarship to S.M.

References

Balbas, P., Alexeyev, M., Shokolenko, I., and Bolivar, F. (1996) A pBRINT family of plasmids for integration of cloned DNA into the *Escherichia coli* chromosome. *Gene* **172:** 65–69.

Barcus, V.A., Titheradge, A.J.B., and Murray, N.E. (1995) The diversity of alleles at the *hsd* locus in natural populations of *Escherichia coli*. *Genetics* **140:** 1187–1197.

Bertani, G., and Weigle, J.J. (1953) Host controlled variation in bacterial viruses. *J Bacteriol* **65:** 113–121.

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

Blakely, G., May, G., McCulloch, R., Arciszewska, L.K., Burke, M., Lovett, S.T., and Sherratt, D.J. (1993) Two related recombinases are required for site-specific recombination at *dif* and *cer* in *E. coli* K12. *Cell* **75:** 351–361.

Bourniquel, A.A., and Bickle, T.A. (2002) Complex restriction enzymes: NTP driven molecular motors. *Biochimie* **84:** 1047–1059.

Brammar, W.J., Murray, N.E., and Winton, S. (1974) Restriction of λ*trp* bacteriophages of *Escherichia coli* K. *J Mol Biol* **90:** 633–647.

Bullas, L.R., Colson, C., and Neufeld, B. (1980) Deoxyribonucleic acid restriction and modification systems in *Salmonella*, chromosomally located systems of different serotypes. *J Bacteriol* **141:** 275–292.

Chaudhury, A.M., and Smith, G.R. (1984) A new class of *Escherichia coli recBC* mutants: Implications for the role of

RecBC enzyme in homologous recombination. *Proc Natl Acad Sci USA* **81:** 7850–7854.

Chaudhury, A.M., and Smith, G.R. (1985) Role of *Escherichia coli* RecBC enzyme in SOS induction. *Mol Gen Genet* **201:** 525–528.

Cromie, G.A., and Leach, D.R. (2001) Recombination repair of chromosomal DNA double-strand breaks generated by a restriction endonuclease. *Mol Microbiol* **41:** 873–884.

Davies, G.P., Martin, I., Sturrock, S.S., Cronshaw, A., Murray, N.E., and Dryden, D.T. (1999) On the structure and operation of Type I DNA restriction enzymes. *J Mol Biol* **290:** 565–579.

Diaz, R., Barnsley, P., and Pritchard, R.H. (1979) Location and characterisation of a new replication origin in the *E. coli* K12 chromosome. *Mol Gen Genet* **175:** 151–157.

Donachie, W.D. (1968) Relationships between cell size and time of initiation of DNA replication. *Nature* **219:** 1077–1079.

Doronina, V.A., and Murray, N.E. (2001) The proteolytic control of restriction activity in *Escherichia coli* K-12. *Mol Microbiol* **39:** 416–428.

Dryden, D.T.F., Murray, N.E., and Rao, D.N. (2001) Nucleoside triphosphate-dependent restriction enzymes. *Nucleic Acids Res* **29:** 3728–3741.

Efimova, E.P., Delver, E.P., and Belogurov, A.A. (1988) 2-Aminopurine and 5-bromouracil induce alleviation of Type I restriction in *Escherichia coli*: mismatches function as inducing signals? *Mol Gen Genet* **214:** 317–320.

Firman, K., Dutta, C., Weiserová, M., and Janscák, P. (2000) The role of subunit assembly in the functional control of Type I restriction-modification enzymes. *Mol Biol* **1:** 35–41.

Hanahan, D. (1983) Studies on transformations of *Escherichia coli* with plasmids. *J Mol Biol* **166:** 557–580.

Handa, N., Ichige, A., Kusano, K., and Kobayashi, I. (2000) Cellular responses to postsegregational killing by restriction-modification genes. *J Bacteriol* **182:** 2218–2229.

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

Heitman, J., Ivanenko, T., and Kiss, A. (1999) DNA nicks inflicted by restriction endonucleases are repaired by a RecA- and RecB-dependent pathway in *Escherichia coli*. *Mol Microbiol* **33:** 1141–1151.

Hiraga, S., Niki, H., Ogura, T., Ichinose, C., Mori, H., Ezaki, B., and Jaffe, A. (1989) Chromosome partitioning in *Escherichia coli*: novel mutants producing anucleate cells. *J Bacteriol* **171:** 1496–1505.

Hubacek, J., and Glover, S.W. (1970) Complementation analysis of temperature-sensitive host specificity mutations in *Escherichia coli*. *J Mol Biol* **50:** 111–127.

Ives, C.L., Nathan, P., D., and Brooks, J.E. (1992) Regulation of the BamHI restriction-modification system by a small intergenic open reading frame, *bamHIC*, in both *Escherichia coli* and *Bacillus subtilis*. *J Bacteriol* **174:** 7194–7201.

Janscak, P., Dryden, D.T., and Firman, K. (1998) Analysis of the subunit assembly of the Type IC restriction-modification enzyme EcoR124I. *Nucleic Acids Res* **26:** 4439–4445.

Kelleher, J.E., and Raleigh, E.A. (1994) Response to UV damage by four *Escherichia coli* K-12 restriction systems. *J Bacteriol* **176:** 5888–5896.

Kobayashi, I. (2001) Behaviour of restriction-modification systems as selfish mobile elements and its relationship with genome evolution. *Nucleic Acids Res* **29:** 3742–3756.

Kulik, E.M., and Bickle, T.A. (1996) Regulation of the activity of the Type IC EcoR124I restriction enzyme. *J Mol Biol* **264:** 891–906.

Kushner, S.R., Nagaishi, H., Templin, A., and Clark, A.J. (1971) Genetic recombination in *Escherichia coli*: The role of exonuclease I. *Proc Nat Acad Sci USA* **68:** 824–827.

Lee, N.S., Rutebuka, O., Arakawa, T., Bickle, T.A., and Ryu, J. (1997) KpnAI, a new Type I restriction-modification system in *Klebsiella pneumoniae*. *J Mol Biol* **271:** 342–348.

Lloyd, R.G., Buckman, C., and Benson, F.E. (1987) Genetic analysis of conjugated recombination in *Escherichia coli* K12 strains deficient in RecBCD enzyme. *J Gen Microbiol* **133:** 2531–2538.

Loenen, W.A.M., and Murray, N.E. (1986) Modification enhancement by the restriction alleviation protein (Ral) of bacteriophage λ. *J Mol Biol* **190:** 11–22.

McGlynn, P., and Lloyd, R.G. (2000) Modulation of RNA polymerase by (p)ppGpp reveals a RecG-dependent mechanism for replication fork progression. *Cell* **101:** 35–45.

Makovets, S. (1999) Regulation of the endonuclease activity of Type I restriction-modification systems. PhD Thesis. University of Edinburgh.

Makovets, S., Titheradge, A.J.B., and Murray, N.E. (1998) ClpX and ClpP are essential for the efficient acquisition of genes specifying Type IA and IB restriction systems. *Mol Microbiol* **28:** 25–35.

Makovets, S., Doronina, V.A., and Murray, N.E. (1999) Regulation of endonuclease activity by proteolysis prevents breakage of unmodified bacterial chromosomes by Type I restriction enzymes. *Proc Natl Acad Sci USA* **96:** 9757–9762.

Murray, N.E. (1983) Phage lambda and molecular cloning. In *Lambda II*. Hendrix, R.W., Roberts, J.W., Stahl, F.W., and Weisberg, R.A. (eds). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, pp. 395–432.

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. (2002) Immigration control of DNA in bacteria: self versus non-self. *Microbiology* **148:** 3–20.

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

O'Neill, M., Chen, A., and Murray, N.E. (1997) The restriction-modification genes of *Escherichia coli* K-12 may not be selfish: they do not resist loss and are readily replaced by alleles conferring different specificities. *Proc Natl Acad Sci USA* **94:** 14596–14601.

Prakash-Cheng, A., Chung, S.S., and Ryu, J. (1993) The expression and regulation of *hsdK* genes after conjugative transfer. *Mol Gen Genet* **241:** 491–496.

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 *Eco*R124 and *Eco*R124/3 Type I DNA restriction and modification enzymes. *J Mol Biol* **205:** 115–125.

Rinken, R., Thoms, B., and Wackernagel, W. (1992) Evidence that *recBC*-dependent degradation of duplex DNA in *Escherichia coli recD* mutants involves DNA unwinding. *J Bacteriol* **174:** 5424–5429.

Seigneur, M., Ehrlich, S.D., and Michel, B. (2000) RuvABC-dependent double-strand breaks in *dnaB*ts mutants require *recA*. *Mol Microbiol* **38:** 565–574.

Singer, M., Baker, T.A., Schnitzler, G., Deischel, S.M., Goel, M., Dove, W., *et al.* (1989) A collection of strains containing genetically linked alternating antibiotic resistant elements for genetic mapping of *Escherichia coli*. *Microbiol Rev* **53:** 1–24.

Skarstad, K., and Boye, E. (1993) Degradation of individual chromosomes in *recA* mutants of *Escherichia coli*. *J Bacteriol* **175:** 5505–5509.

St Pierre, R., and Linn, T. (1996) A refined vector system for the *in vitro* construction of single-copy transcriptional or translational fusions to *lacZ*. *Gene* **169:** 65–68.

Steiner, W.W., and Kuempel, P.L. (1998) Sister chromatid exchange frequencies in *Escherichia coli* analyzed by recombination at the *dif* resolvase site. *J Bacteriol* **180:** 6269–6275.

Suri, B., and Bickle, T.A. (1985) *Eco*A: the first member of a new family of Type I restriction modification systems. Gene organization and enzymic activities. *J Mol Biol* **186:** 77–85.

Tao, T., Bourne, J., C., and Blumenthal, R.M. (1991) A family of regulatory genes associated with Type II restriction-modification systems. *J Bacteriol* **173:** 1367–1375.

Taylor, I., Watts, D., and Kneale, G. (1993) Substrate recognition and selectivity in the Type IC DNA modification methylase M.EcoR124I. *Nucleic Acids Res* **21:** 4929–4935.

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

Titheradge, A.J.B., King, J., Ryu, J., and Murray, N.E. (2001) Families of restriction enzymes: an analysis prompted by molecular and genetic data for Type ID restriction and modification systems. *Nucleic Acids Res* **29:** 4195–4205.

Willcock, D.F., Dryden, D.T., and Murray, N.E. (1994) A mutational analysis of the two motifs common to adenine methyltransferases. *EMBO J* **13:** 3902–3908.

Zabeau, M., Friedman, S., Van Montagu, M., and Schell, J. (1980) The *ral* gene of phage λ. *Mol Gen Genet* **179:** 63–73.