1694–1700 *Nucleic Acids Research, 1997, Vol. 25, No. 9*

The specificity of *Sty*SKI, a type I restriction enzyme, implies a structure with rotational symmetry

Peter H. Thorpe, Diane Ternent and Noreen E. Murray*

Institute of Cell and Molecular Biology, Darwin Building, Kings Buildings, University of Edinburgh, Mayfield Road, Edinburgh EH9 3JR, UK

Received February 13, 1997; Accepted March 13, 1997

ABSTRACT

The type I restriction and modification (R–M) enzyme from *Salmonella enterica* serovar *kaduna* (*Sty*SKI) recognises the DNA sequence 5′-CGAT(N)$_7$GTTA, an unusual target for a type I R–M system in that it comprises two tetranucleotide components. The amino target recognition domain (TRD) of *Sty*SKI recognises 5′-CGAT and shows 36% amino acid identity with the carboxy TRD of *Eco*R124I which recognises the complementary, but degenerate, sequence 5′-RTCG. Current models predict that the amino and carboxy TRDs of the specificity subunit are in inverted orientations within a structure with 2-fold rotational symmetry. The complementary target sequences recognised by the amino TRD of *Sty*SKI and the carboxy TRD of *Eco*R124I are consistent with the predicted inverted positions of the TRDs. Amino TRDs of similar amino acid sequence have been shown to recognise the same nucleotide sequence. The similarity reported here, the first example of one between amino and carboxy TRDs, while consistent with a conserved mechanism of target recognition, offers additional flexibility in the evolution of sequence specificity by increasing the potential diversity of DNA targets for a given number of TRDs. *Sty*SKI identifies the first member of the IB family in *Salmonella* species.

INTRODUCTION

The type I restriction and modification (R–M) systems of enteric bacteria, the first restriction systems to be characterised, comprise a diverse group of complex enzymes that possess both endonuclease and methyltransferase (MTase) activities (for reviews see 1–7). The importance of these enzymes *in vivo* as a host barrier to phage infection is highlighted by the variety of phage-encoded anti-restriction systems specific to type I enzymes (3), but DNA breakage may also promote homology-dependent recombination between the incoming DNA and the bacterial chromosome (6,8).

The type I R–M systems are encoded by three genes, *hsdR*, *hsdM* and *hsdS*; the HsdS and HsdM subunits together form an active MTase which with the addition of the HsdR subunit forms the complete R–M system (1–5). These enzymes bind to characteristic bipartite DNA target sequences, typically a 3 bp component separated from a 4–5 bp component by a non-specific spacer of 6–8 bp (1–5). The MTase transfers a methyl group from *S*-adenosyl methionine (AdoMet), the methyl donor and co-factor, to the N6 position of a specific adenine residue on each strand of the DNA target. DNA methylation by the type II Mtase M.*Hha*I has been shown to involve a base-flipping mechanism (9). The similarities seen between the predicted secondary structure of the AdoMet-binding domain of *Eco*KI (10) and those of M.*Hha*I (9) and other type II Mtases (11,12) suggest that base-flipping is a prerequisite to DNA methylation by *Eco*KI. Upon binding to unmethylated DNA targets, the R–M enzyme initiates the complex restriction reaction which involves ATP-dependent DNA translocation followed by double strand cleavage at a non-specific site distant from the enzyme-bound target sequence (13–17).

Unlike type II systems, which have separate MTases and endonucleases, the multi-subunit complexes of type I systems are endowed with both activities. One subunit, HsdS, determines the DNA specificity of the system and consequently changes within the *hsdS* gene that confer a new specificity concomitantly affect both restriction and modification. In contrast, an alteration in the specificity of a type II R–M system requires changes in both the endonuclease and the MTase. The known type I systems of enteric bacteria have been divided into four families (IA, IB, IC and ID) on the basis of a number of characteristics, most notably high sequence conservation within but not between families (18). The *hsd* genes for members of the IA, IB and ID families are closely linked to the *serB* locus, at 98.5 minutes on the chromosome of *Escherichia coli* K-12, and they behave as alleles in tests dependent on genetic recombination (19,20). The IA and ID families have members from both *E.coli* and *Salmonella*, the IB family has a member from *Citrobacter freundii* in addition to those from *E.coli* (18,21,22). The majority of the type IC systems characterised to date are encoded by plasmids that are known to be transferred readily between *E.coli* and *Salmonella* (2).

Whilst the DNA and amino acid sequences of *hsd* systems within a family are highly conserved, the *hsdS* genes contain two large regions of sequence (~450 bp) which are usually quite different even in members of the same family of type I systems (22–25), and new combinations of these so-called variable sequences within the *hsdS* gene generate different DNA specificities

* To whom correspondence should be addressed. Tel: +44 131 650 5374; Fax: +44 131 668 3870; Email: noreen.murray@ed.ac.uk

(24–27). Each variable sequence encodes a domain that is responsible for recognising half of the DNA target (27), and these domains are referred to as target recognition domains (TRDs). The domain boundaries predicted from sequence comparisons are reinforced by analyses of the products of limited proteolysis of both *Eco*KI (29) and *Eco*R124I (30) Mtases. As new type I systems have been discovered and their DNA specificity determined, it has become clear that the amino acid sequences of TRDs may show significant similarity if an identical target is recognised. Two members of the IA family share a common trinucleotide component in their specific recognition sequence; *Eco*KI recognises 5′-AAC(N)$_6$GTGC and *Sty*SPI recognises 5′-AAC(N)$_6$GTRC. The amino TRDs which are responsible for the recognition of the 5′-AAC share 90% identity (31). Also, two members of the IB family *Eco*AI and *Eco*EI share a trinucleotide target sequence with *Sty*LTIII of the IA family. These three amino TRDs, which all recognise the sequence 5′-GAG, have 40% amino acid identity, identifying possible regions within the TRD that are involved in DNA recognition (28).

Only a small number of amino acids within each TRD is expected to be critical for determining the DNA specificity of a target half-site. The remaining amino acids may serve a structural role for which some changes may be tolerated. Analyses of TRDs that recognise similar or identical DNA targets, and consequently have amino acid identities, should reveal regions of importance for specific interactions with DNA. This expectation has stimulated the search for new type I R–M systems. Thirty-seven of the ECOR collection of wild-type *E.coli* strains (32) have been screened using probes specific for members of the IA, IB or ID families, and 17 of them were positive (18). There has been no comparable screen of *Salmonella* species for type I systems, although sequence analysis of the characterised IA and IB family *hsd* genes implies horizontal transfer of type I R–M genes between *E.coli* and *Salmonella* (33). Bullas *et al.* (19) have screened many strains of *Salmonella* for the restriction of bacteriophages, and they identified 12 *Salmonella* strains with previously unknown restriction specificities. The *hsd* genes identified were found to be linked to the *serB* locus and included those from *Salmonella enterica* serovar *kaduna* (*S.enterica* ser *kaduna*) encoding the *Sty*SKI R–M system.

We have cloned the genes encoding the MTase component of *Sty*SKI, and the nucleotide sequences of the *hsd* genes identify *Sty*SKI as a member of the IB family. The amino TRD of the HsdS subunit of *Sty*SKI shows 36% identity with a carboxy TRD sequence from an IC family member, *Eco*R124I, implying that these two target recognition domains might recognise the same DNA target. The target sequences determined by these two TRDs are not identical, but rather one is a degenerate complement of the other. The complementarity is consistent with the suggestion that these two TRDs orient inversely on their DNA targets.

MATERIALS AND METHODS

Bacterial strains, phages, plasmids and microbial techniques

LB4023 (19), an *E.coli* K-12 derivative in which the *hsd* region of the chromosome has been replaced with that of *S.enterica* serovar *kaduna*, was used as the source of the *kaduna*-specific *hsd* genes and as a λ-sensitive strain proficient in *kaduna*-specific restriction. A derivative sensitive phage to M13 was made by the acquisition of F′*kan* from EH55 (34). Mutations were induced in phages by propagation in NM720, an *hsd*Δ *mutD*5 strain harbouring F′*kan*. The hsdΔ F′ strain NM522 (23) was used as a standard strain for M13, and ED8654 (35) for λ phages. NM782 is a derivative of NM522 in which the *hsd* region of *E.coli* K-12 has been replaced with that encoding *Eco*AI, and NM789 is a derivative of NM782 in which *hsdS* has been inactivated by a deletion.

The library of DNA fragments was made in the λ vector EMBL3*cos* (36). The *Sal*I fragment encoding the modification component of *Eco*AI was transferred from the λ*hsd* phage described by Fuller-Pace *et al.* (37) to a derivative of EMBL3 (NM1249) in which the *c*I$^-$ allele has been replaced by *c*I857. This enabled lysogens to be made at 32° following integration of λ*hsd* phages by homology-dependent recombination. A derivative of the λ*hsdMS* phage was made in which a *Bam*HI fragment containing the 5′ half of the *hsdS* gene was replaced with the 5′ half of the *hsdS* gene conferring the *Sty*SKI specificity. The hybrid *hsdS* gene was transferred to the chromosome of strain NM789 following the homology-dependent integration and excision of the genome of the λ*hsdc*I857 phage. The specificity of the resulting strain (NM821) is determined by the chimaeric HsdS polypeptide comprising the amino TRD of *Sty*SKI and the carboxy TRD of *Eco*AI.

The plasmid pAB3 used as a probe specific for a sequence close to the *hsd* genes of *S.enterica* serovar *kaduna* includes DNA from immediately downstream of the *hsdS* gene of serovar *typhimurium* LT2 (20).

Restriction was estimated from the reduced efficiency of plating (e.o.p.) of unmodified phages on a restricting strain relative to that on a nonrestricting strain; modified phages plated with equal efficiency on both strains. Modification was assessed by the protection against the relevant restriction system e.g. *Sty*SKI (LB4023), *Eco*AI (NM782) or the hybrid derivative (NM821).

Deletion derivatives of λ phages were selected on BBL agar supplemented with 0.2 mM EDTA (37). Deletion and substitution derivatives of λ used to locate targets for either *Sty*SKI or the hybrid system included deletions in the *b*2 region (e.g. λ*b*527, λ*b*538 and λ*b*221) and deletions or substitutions in the immunity region (λ*imm*^{434}, λ*imm*^{21} and λ*c*IΔKH54). The coordinates for these deletions and substitutions are given in Appendix I of Lambda II (38).

Mutagenesis of target sequences

Lysates of λ and M13 were made on an *hsd*Δ strain (the *mutD* strain NM720 for λ and NM789 for M13), and then transferred to the appropriate restricting strain, either LB4023 F′*kan* encoding *Sty*SKI, or NM821 specifying the hybrid R–M system. After five cycles of enrichment in which phages were grown on an *hsd*Δ strain and then challenged by transfer to a restricting strain, phages were isolated that were resistant to restriction.

Enzymes, reagents and DNA manipulations

These were as described in reference 33. The –40 primer was supplied with the Amersham Sequence® kit, other oligonucleotides, purchased from Oswell DNA Service (University of Southampton), were as follows. 5′-CACTGACCGGAACGCCAA (T3941) and 5′-GCCAGAATATCCCTGCCA (T3942) were used to amplify and sequence the region of λ to the right of gene *J* which contains a target for *Sty*SKI. 5′-GGCGAGCCTGGCTAACCG (T5482) and 5′-GTAAGCGCATTGGCCCGC (T5483)

were used to amplify and sequence the region of λ in gene *O* including a second target for *Sty*SKI. 5′-AATTGTCGANCTGC-TCCGTTAGGC (T6731) and its complementary oligonucleotide 5′-AATTGCCTAACGGAGCAGNTCGAC (T6732) were used to generate a putative target sequence for *Sty*SKI with a single degeneracy at position 4 of the target sequence. 5′-AATTGACCK-ATACTGAGAGTTACGC (T10012) and its complementary oligonucleotide 5′-AATTGCGTAACTCTCAGTATMGGTC (T10013) generated a putative target sequence for *Sty*SKI, but with a single degeneracy at position 2. Both pairs of complementary oligonucleotides have 5′ extensions that base pair with vector cut with *Eco*RI.

Nucleotide sequence of M13 single-stranded DNA was obtained using the di-deoxy chain termination sequencing method (39) with 5-[α-^{35}S]thiotriphosphate (40) using reagents and methods as recommended in the Amersham Sequenase® sequencing kit. The nucleotide sequence of PCR products were obtained using the Amersham PCR product sequencing kit; the PCR products were treated with alkaline phophatase and exonuclease III to remove dNTPs and primers respectively. The enzymes were inactivated by incubation at 80°C before the sequencing reactions were carried out. The sequence comparisons and alignments used the GAP and PILEUP programs of the Wisconsin Package, Version 8 of the Genetics Computer Group, WI, USA.

RESULTS

Cloning the *hsd* region of *S.enterica* ser *kaduna*

A derivative of *E.coli* K-12 (LB4023) in which the resident *hsd* genes encoding *Eco*KI have been replaced by those from *S.enterica* ser *kaduna* (19) was used as the source of bacterial DNA, and as a λ-sensitive strain restricting with the specificity of *Sty*SKI. A library of large DNA fragments, generated by digesting the DNA of LB4023 with *Sau*3A, was made in a λ replacement vector. Probes specific for either type IA or IB systems had failed to identify *hsd* genes in digests of DNA isolated from LB4023 (41), therefore a DNA fragment from immediately downstream of the *hsdS* gene of *S.enterica* ser *typhimurium* was used as a probe to identify λ clones with inserts from the homologous region of the chromosome of LB4023 (20).

Three independent clones were isolated and the genome of one of the three (λ4023) was found to be resistant to restriction by LB4023, consistent with the modification of its target sequences for *Sty*SKI. λ4023 was predicted to include the genes encoding the modification component of *Sty*SKI, and was the basis of all subsequent analyses.

Deletion derivatives of λ4023 were selected by isolating plaques on medium containing the chelating agent EDTA. The modification phenotypes of the deletion derivatives were determined. Some derivatives were sensitive to restriction by LB4023, as would be expected if *hsdM* or *hsdS* were inactivated by a deletion. DNA preparations from the parental phage and deletion derivatives were analysed by restriction digests. The deletion mutants localise *hsdM* and *S* to the left-hand part of the restriction map (Fig. 1). Each of the two adjacent *Sal*I fragments (2.3 and 2.6 kb) located in this region was transferred to M13mp18 and mp19 and the nucleotide sequence of the bacterial DNA was determined.

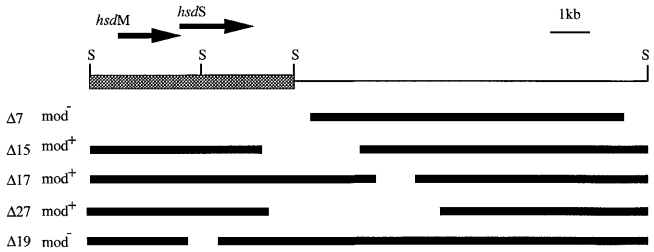

Figure 1. Only the fragment of bacterial DNA within the λ vector is shown, the thick grey bar indicates the region (two *Sal*I fragments) that was sequenced. The positions of the ORFs corresponding to *hsdM* and *hsdS* are shown above the grey bar. Five deletion derivatives are identified. The extent of each deletion is indicated by a gap, the *Sty*SKI modification phenotype is listed in the left-hand column.

The nucleotide sequence of the *S.enterica* ser *kaduna hsd* genes

The sequence obtained (EMBL database accession no. Y11005) includes three ORFs. The amino acid sequences predicted from these ORFs are consistent with the identification of the *hsdM* and *hsdS* genes preceded by 660 bp of the 3′ end of *hsdR*. The *hsdM* gene and the 3′ end of *hsdR* share 70–95% identity with the *hsdM* and *hsdR* genes encoding *Eco*AI and *Eco*EI, consistent with *S.enterica* ser *kaduna* having a type IB R–M system. The *hsdS* gene includes the conserved regions characteristic of the type IB family, whilst two regions predicted to encode the two TRDs are not conserved, as expected if they are responsible for the recognition of different DNA targets. The *hsd* genes of members of the same family have always been sufficiently similar to be identified with family-specific DNA probes, this is so for the cloned *Sty*SKI *hsd* genes (data not shown), despite poor hybridisation of DNA in chromosomal digests (41).

The predicted amino acid sequences of the subunits of *Sty*SKI are compared with those of other type I R–M enzymes in Table 1. The high levels of sequence identity found in comparisons between each of two members of the IB family contrasts with the low levels of identity with members of each of the three other families of type I R–M enzymes. The figures are consistent with those of previous comparisons within and between families (33,42).

Table 1. Comparisons of the predicted amino acid sequences, shown as per cent identity

Enzymes	Family	HsdR	HsdM	HsdS
*Eco*AI	IB	90	97	55
*Eco*EI	IB	72	91	50
*Eco*KI	IA	23	30	21
*Eco*R124I	IC	21	24	30
*Sty*SBLI	ID	14	28	17

The amino acid sequences of the Hsd subunits of *Sty*SKI were predicted from the nucleotide sequences and aligned with those for other Hsd polypeptides using PILEUP. The BESTFIT program (43) was used to determine per cent identity. HsdM and HsdS were assumed to begin at the first methionine; the alignments for HsdR are based on only the 220 amino acids at the carboxy terminus of the predicted sequence for *Sty*SKI.

An unexpected similarity was detected between the predicted HsdS sequences of *Sty*SKI and *Eco*R124I (an IC family member). The amino TRD sequence of *Sty*SKI shows 36% identity with the carboxy TRD sequence of *Eco*R124I. This is the first report of extensive similarity between an amino TRD sequence and a carboxy TRD sequence. Similarity of TRD sequences is not usual unless they recognise the same sequence. The level of identity of these TRDs is comparable to the 40% detected for two amino TRD sequences responsible for the recognition of 5′-GAG; *Eco*AI of the IB family and *Sty*LTIII of the IA family.

DNA specificity of *S.enterica* ser *kaduna*

In order to determine whether the identity between the *Eco*R124I carboxy TRD and the *Sty*SKI amino TRD is functionally significant it was essential to determine the target sequence recognised by *Sty*SKI. Deletions and substitutions were used to identify and localise the targets for *Sty*SKI in the genome of phage λ. Unmodified phages were tested for their sensitivity to restriction by LB4023. Wild-type λ has an e.o.p. on LB4023 of ~$10^{-3.}$ Some deletions in the *b*2 region of λ, including *b*538 and *b*527, had no effect on the e.o.p.; others in which the deletion extended closer to gene *J*, including *b*189 and *b*221, resulted in an increased e.o.p. (10^{-1} to 10^{-2}). Similarly, substituting imm^{434} for *imm*λ had no effect, but an imm^{21} substitution increased the e.o.p. (10^{-1} to 10^{-2}). These results are consistent with one *Sty*SKI target in the *b*2 region close to the *J* gene and a second in the immunity region. In subsequent genetic crosses it was found that not all imm^{21} phages had an increased e.o.p., a finding explained by the final location of the second target close to, rather than within, the immunity region.

More precise positions of the predicted target sequences were identified by mutations within the λ*b*527 genome which conferred resistance to attack by *Sty*SKI. Mutations were induced by growth of the phage in a *mutD* strain (NM720), to increase the chance that restriction targets were lost as the result of base substitutions rather than deletions. λ*b*527 was chosen as the starting phage since it lacks a normal attachment site and is therefore less likely than λ^+ to yield deletions as the result of Int-mediated recombination. Mutant phages were isolated following multiple cycles of enrichment for phages with increased e.o.p. on LB4023 (see Materials and Methods). Two classes of mutants were isolated: those that had an increased e.o.p. on LB4023 when compared with the original phage (e.o.p. of 10^{-1} to 10^{-2} rather than 10^{-3}), and those that were completely resistant to restriction by *Sty*SKI, as indicated by an e.o.p. of 1 on LB4023. Mutants that had lost their sensitivity to LB4023 restriction were presumed to have lost their targets for *Sty*SKI. The simple interpretation of the results is that λ*b*527 has two targets for *Sty*SKI; phages with an e.o.p. of 10^{-1} to 10^{-2} retain one target and those with an e.o.p. of 1 have lost both. A 1.8 kb *Sma*I–*Eco*RI fragment from the left-hand end of the *b*2 region of a λ*b*527 derivative resistant to restriction by *Sty*SKI was cloned in M13mp19 and the nucleotide sequence of the insert determined. A single base change of adenine to guanine was detected at position 20936 in the nucleotide sequence of λ (37,39; GenEMBL sequence data base accession no. J202495).

Oligonucleotide primers were used to amplify, by the polymerase chain reaction (PCR), a segment of the λ genome that includes the base-pair identified by mutation. The nucleotide sequences of the amplified DNA from each of five restriction-resistant

a) Mutations identified in targets within the λ genome

```
Target 1  5'cgCGATtgcagatGTTAtc
                ↓
              ↓ G        ↓
              T          C

Target 2  5'caCGATggaacagGTTAac
                ↓
              ↓ G          ↓
              T            G
```

b)

	Target sequences	Source
1	5' cgCGATtgcagatGTTAtc	λ bp. 20932 - 50
2	5' caCGATggaacagGTTAac	λ bp. 39821 - 39
3	5' acCGATactgagaGTTAcg	oligos T10012/3
4	5' gtCGATctgctccGTTAgg	oligos T6731/2

c) Non-target sequences

1	5' acCtATactgagaGTTAcg	oligos T10012/3
2	5' gtCGAgctgctccGTTAgg	oligos T6731/2
3	5' gtCGAcctgctccGTTAgg	oligos T6731/2
4	5' gtCGAactgctccGTTAgg	oligos T6731/2

Figure 2. The target sequence recognised by *Sty*SKI. (**a**) The two target sequences in phage λ and the mutations that lead to loss of the target. (**b**) Four sequences that make M13mp18 sensitive to *Sty*SKI. (**c**) Sequences differing from those in (b) by one degeneracy within the 5′ tetranucleotide. These sequences all fail to elicit restriction.

derivatives of λ*b*527 identified two additional base changes (Target 1 in Fig. 2a). A putative target sequence of 5′-CGAT$(N)_7$GTTA was deduced as the only sequence to occur twice in λ that is compatible with the three mutations identified in the first target sequence; the other occurrence of 5′-CGAT$(N)_7$GTTA is close to the immunity region in gene *O* (bp 39823–39837). The *Bgl*II–*Sma*I fragment that includes this second putative target sequence (Target 2 in Fig. 2a) was transferred from a restriction-resistant derivative of λ*b*527 to M13mp18. The nucleotide sequence indicated a single change within the predicted target sequence. Oligonucleotide primers were chosen to amplify the region of λ identified by this mutation, and the sequences of products derived from five restriction-resistant derivatives of λ*b*527 were determined. Three different mutations were detected (Fig. 2a). The predicted target sequence, CGAT$(N)_7$GTTA, was changed in all the restriction-resistant derivatives of λ*b*527. Degeneracies at all but the second and fourth positions within the bipartite, octameric sequence require more than two targets in the λ genome. The target sequence compatible with the data is 5′-CKAB$(N)_7$GTTA, where K is either G or T and B is T, C or G.

In order to determine whether there are degeneracies in the target sequence, two pairs of complementary oligonucleotides were made, and the annealed products cloned in M13mp18 to produce phage with putative target sequences. The phages were checked for their sensitivity to restriction by a strain encoding

*Sty*SKI (LB4023 F′*kan*), and the nucleotide sequence of each insert was determined. One oligonucleotide pair (T10012 and T10013) has a degeneracy at the second position. The test showed that a G (Fig. 2b, line 4), but not a T (Fig. 2c, line 1), at the second position resulted in an active target sequence. The second oligonucleotide pair (T6731 and T6732) introduced degeneracies at position four. Only a T at this position produced a restriction-sensitive phage (Fig. 2b, line 4, and c, lines 3 and 4). The bases immediately flanking the two components of the target sequences (1–4 in Fig. 2b) are completely degenerate, as are four of the seven bases in the spacer. The data identify 5′-CGAT(N)$_7$GTTA or its complementary sequence as the target for *Sty*SKI.

Does the amino TRD of *Sty*SKI specify 5′-CGAT?

Each natural type I R–M system studied prior to *Sty*SKI recognizes a target sequence that includes a trinucleotide component specified by the amino TRD of the HsdS subunit, and by convention the target sequence is written beginning with the component specified by the amino TRD. The target site of *Sty*SKI differs in that it comprises two tetranucleotide components and the data presented do not orient the TRDs with respect to the two components of the target sequence. Formally, the target sequence for *Sty*SKI could be either 5′-CGAT(N)$_7$GTTA or 5′-TAAC(N)$_7$ATCG. The component of the target sequence specified by the amino TRD of *Sty*SKI must be established. This was done by the analysis of a hybrid system.

A chimaeric HsdS subunit was made in which the amino TRD of the *Eco*AI system was replaced with that of the *Sty*SKI system. This was achieved in a λ*hsd* phage encoding the modification component of *Eco*AI, by replacing the *Bam*HI fragment encoding the amino TRD of *Eco*AI with that encoding the amino TRD of *Sty*SKI. The resulting phage no longer encoded the modification system of *Eco*AI, but conferred a new specificity. Lysogens of this phage in a strain (NM789) that retains *hsdR* and *hsdM* of *Eco*AI, but has a deletion within *hsdS*, were restriction-proficient. A lysogen was cured of its prophage and a strain (NM821) that had replaced the *hsdS*Δ with a hybrid *hsdS* gene was used to deduce the target recognised by the hybrid R–M system.

If the amino TRD of the new system recognizes the sequence 5′-CGAT and the spacing of N=7 is retained, the predicted specificity of the chimaeric subunit is 5′-CGAT(N)$_7$GTCA, where GTCA is the component specified by the carboxy TRD of *Eco*AI.

This target sequence occurs once within the sequence of M13mp18 or -19 and twice within that of phage λ, where one of the two targets is within the immunity region in DNA deleted by the mutation *c*IΔKH54. Consistent with the predictions, both wild-type λ and M13mp18 were restricted by the hybrid system. Furthermore, the e.o.p. of wild-type λ was less than that for either λ*c*IΔKH54 or λ*imm*^{434}, as expected if one target is within the immunity region. This distribution of targets in λ and M13mp18 is inconsistent with a spacing of N=6 rather than N=7, and with the alternative combinations of target sequences. Nevertheless, a derivative of M13mp18 was selected that was resistant to restriction by the hybrid system, and the sequence of this mutant M13 was checked in the region of the predicted target sequence. The three restriction-resistant derivatives of M13mp18 tested had a mutation in the 3′ component of the target sequence; 5′-GTCA was changed to 5′-GTTA. This mutation confirms the predicted target sequence for the hybrid system and consequently defines the orientation of the components of the target sequence for *Sty*SKI. Concomitantly, conversion of the target sequence to 5′-CGAT(N)$_7$GTTA creates a new target sequence for *Sty*SKI. As expected, the mutant phage has acquired sensitivity to restriction by *Sty*SKI.

DISCUSSION

*Sty*SKI *hsd* genes

The *Sty*SKI type I R–M system is the first representative of the IB family to be identified in *Salmonella*. Both *E.coli* and *Salmonella* are now shown to have members in each of the three families of allelic *hsd* genes. Hybrid *hsdS* genes have been isolated for both the IA (25,27) and IC (24) families of enzymes and have been shown to confer the predicted new specificies (24,26). Hybrid *hsdS* genes generated from members of the IB family have not been described previously; the one reported in this paper functions with the anticipated specificity.

DNA specificity of *Sty*SKI

The amino TRD sequences of the *Eco*AI and *Sty*LTIII share only 40% amino acid identity whilst they both recognise the same DNA target (5′-GAG), implying that many amino acids within the TRD are not directly involved in determining DNA specificity. On the assumption that both domains recognise their DNA target by the same mechanism, the residues critical for specific DNA recognition will be amongst the conserved residues. An analysis of point mutations within the 5′ variable region of the *Eco*KI *hsdS* gene has shown that many changes have no effect on *Eco*KI activity; currently those that abolish the restriction and modification activity are confined to a short segment of the TRD (M.O'Neill, personal communication). It is possible that the TRDs impart nucleotide specificity in a similar way to the type II endonucleases and MTases (9,44–47). For these enzymes, structural studies demonstrate that short loops of the polypeptide extend into the DNA helix to make base-specific contacts.

The amino TRD sequence of *Sty*SKI is 36% identical to the carboxy TRD of *Eco*R124I (Fig. 3), implying that these two TRDs are related despite their different positions within the HsdS subunit and their presence in enzymes from different families. The level of identity is similar to that found for the amino TRDs of *Eco*AI and *Sty*LTIII but is considerably greater than the 20% identity between the carboxy TRD sequences of *Eco*KI and *Sty*SPI, which also recognise a similar but degenerate target sequence. Although 20% identity seems a low figure for two TRDs with a similar target sequence, the carboxy TRDs of *Eco*KI and *Sty*SPI were the most similar of the 10 comparisons that could be made at that time (27).

It has been postulated (10,48–52) that an HsdS subunit has a symmetrical structure within which the two TRDs are separated by two spacer regions. This symmetrical structure is predicted to be common to all members of the known families of Type I enzymes (Fig. 4). If the HsdS subunit is functionally symmetrical then amino and carboxy TRDs, as well as the spacer regions, should be interchangeable within each family. Experiments in which the 3′ half of an *hsdS* gene was deleted have shown that amino TRDs can function as carboxy TRDs (50,51); in such cases the dimerized TRDs recognise symmetrical rather than asymmetrical target sequences. Similar experiments have recently been reported in which the 5′ half of an *hsdS* gene has been

```
          1     **       *  +**  **** * *+*   * ***+ * ***   *+
SKI         lpvgwEWvtf shlGhfFGGk TPSKmKdEYW g.GtIPWVtP KDMktnLIvd
R124I       kegevEWktl geiGkwYGGg TPSKnKiEFW enGsIPWIsP KDMgrtLVds
          196

         50*** +*  *+     **  *   **  * **  **  ++* *    +  * ***
SKI         SEDkVTslAI edglTKvsPg .SIlfVaRSg ILrrIFPvAi tsIecTvNQD
R124I       SEDyITeeAV lhssTKliPa nSIaiVvRSs ILdkVLPsAl ikVpaTlNQD
          246

         99+*   *       **  *        *      ***  * *       +
SKI         LKvlsPflse isyYIrlMmn gferyIvenl tKTGttVeSl lf.ddFishp
R124I       MKaviPheni lvkYIyhMig srgsdIlraa kKTGgsVaSi dskknYfhlk
          296

         148  +*   **
SKI         fmIPpfaEQ
R124I       ipVPninEQ
          346
```

Figure 3. An alignment of the amino TRD of *Sty*SKI (SKI) and the carboxy TRD of *Eco*R124I (R124I). *, amino acid identities; +, amino acid similarities (similar amino acids are grouped as follows: [isoleucine and valine], [phenylalanine, leucine, tyrosine, methionine and tryptophan] and [glutamic acid and aspartic acid]). All similar or identical amino acids are shown in capital letters. The numbers indicate the residues from the start of the relevant HsdS sequence; the two sequences are 36% identical.

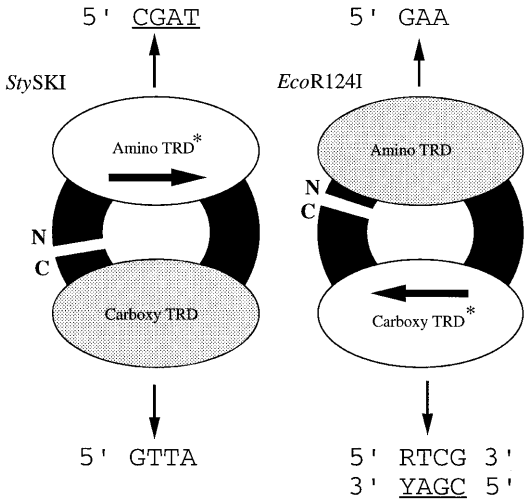

Figure 4. The HsdS subunits of *Sty*SKI and *Eco*R124I along with the DNA targets specified by the TRDs are shown. The amino TRD of *Sty*SKI and the carboxy TRD of *Eco*R124I, indicated with an asterisk, have 36% amino acid identity. The DNA targets recognised by these domains are indicated. The complement of the *Eco*R124I carboxy TRD target (underlined) a degenerate version of the *Sty*SKI amino TRD target (also underlined). The predicted 2-fold rotational symmetry of the HsdS subunit is emphasised in this diagram.

deleted for a member of the IC family and the target sequence in this case comprises two symmetrically arranged tetranucleotide sequences flanking the spacer (52). If the amino and carboxy TRD have opposing orientations within a symmetrically organised subunit (Fig. 4) the target sequence of the amino TRD of *Sty*SKI is predicted to be the complement of that of the carboxy TRD of *Eco*R124I. Our data meet this prediction; 5′-CGAT is the complement of 5′-ATCG.

Recently, Taylor *et al.* (53) chemically modified the surface lysine residues in the Mtase component of *Eco*R124I in both the presence and absence of DNA. Lysine residues 261, 297 and 327 within the carboxy TRD of *Eco*R124I were protected from modification by the binding of a DNA target sequence. Interestingly, these three lysine residues align with lysines in the *Sty*SKI amino TRD sequence (Fig. 3), whilst lysine residue 326 within the *Eco*R124I carboxy TRD is neither conserved nor protected from modification by DNA. Hence, residues which are conserved in these two TRDs have been implicated in the interaction of *Eco*R124I with its target sequence. The data are consistent with these two TRDs recognising their DNA targets using the same mechanism, despite their predicted inverted orientation within HsdS (Fig. 4).

Evolution of new DNA specificities

The chromosomally encoded *hsd* genes can be transferred between strains by conjugation and by transduction. The type I R–M systems also have plasmid-borne members which are readily transferred between bacterial hosts. Type I R–M systems with new specificities can arise spontaneously *in vivo* (19,24,54–56). Experiments have shown that the sequences encoding TRDs can be switched between *hsd* genes of the same family to create new combinations of TRDs, and consequently enzymes with new specificities (24–27 and the data presented here for the IB family). Homology-dependent recombination can generate new combinations of TRDs within a family of enzymes, but it seems less likely that a sequence encoding a TRD can be transferred between different families of *hsd* genes in this way, since there is little DNA sequence identity. Similar TRD sequences in different families of enzymes may reflect their existence prior to divergence of the *hsd* genes into separate families. Alternatively short conserved sequences that flank the variable regions in some *hsdS* genes may identify the remnants of sites associated with the homology-dependent recombination a considerable time ago, or illegitimate recombination events could be involved in the transfer of segments of DNA coding for TRDs. If the variable regions encoding TRDs can be transferred not only between *hsdS* genes in different families, but also from the 5′ to 3′ end of a gene and vice versa, clearly the scope for varying specificity is increased.

Repeated nucleotide sequences, particularly within the *hsdS* genes of the IB and IC families, support the idea that present *hsdS* genes have been derived by a gene duplication event (22–24,52), in which case the carboxy TRD sequence of *Eco*R124I may result from the duplication of an ancestral *hsdS* gene common to the type IB and IC families of R–M systems. Regardless of the evolutionary relationship, our data from natural enzymes indicate that the variable regions have the potential to encode functional TRDs irrespective of their orientation within the subunit. Such an adaptable DNA binding domain extends the repertoire of new specificities that can be generated from the available TRDs.

ACKNOWLEDGEMENTS

We thank our colleagues for their helpful discussions and advice, particularly Lynn Powell and David Dryden for constructive comments on the manuscript and Annette Titheradge for the genomic library. We are indebted to Karen Witherspoon for preparing the manuscript and the Medical Research Council for support of the research and a studentship (P.H.T.).

REFERENCES

1 Wilson,G.G. and Murray,N.E. (1991) *Annu. Rev. Genet.*, **25**, 585–627.
2 Redaschi,N. and Bickle,T.A. (1996) In Neidhardt,F.C. (ed.), *Escherichia coli and Salmonella*. Edition 2. ASM Press. Washington, DC. pp. 773–781.

3 Bickle,T.A. and Kruger,D.H. (1993) *Microbiol. Rev.*, **57**, 434–450.
4 Halford,S.E., Taylor,J.D., Vermote,C.L.M. and Vipond,I.B. (1993) In Eckstein,F. and Lilley,D.M.J. (eds), *Nucleic Acids and Molecular Biology.* Springer-Verlag, Berlin, **7**, pp. 47–69.
5 Heitman,J. (1993) In Setlow,J.K. (ed.), *Genetic Engineering.* Plenum Press, New York, **15**, pp. 57–108.
6 King,G. and Murray,N.E. (1994) *Trends Microbiol.*, **2**, 465–469.
7 Barcus,V.A. and Murray,N.E. (1995) In Baumberg,S., Young,J.P.W., Saunders,S.R. and Wellington,E.M.H. (eds) *Population Genetics of Bacteria*. Society for General Microbiology, Symposium 52, Cambridge University Press, pp. 31–58.
8 Price,C. and Bickle,T.A. (1986) *Microbiol. Sci.*, **3**, 296–299.
9 Cheng,X., Kumar,S., Posfai,J., Pflugrath,J.W. and Roberts,R. (1993) *Cell*, **74**, 299–307.
10 Dryden,D.T.F., Sturrock,S.S. and Winter,M. (1995) *Nature Struct. Biol.*, **2**, 632–635.
11 Labahn,J., Granzin,J., Schluckebier,G., Robinson,D.P., Jack,W.E., Schildkraut,I. and Saenger,W. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 10957–10961.
12 Malone,T., Blumenthal,R.M. and Cheng.X. (1995) *J. Mol. Biol.*, **253**, 618–632.
13 Yuan,R., Hamilton,D.L. and Burckhardt,J. (1980) *Cell*, **20**, 237–244.
14 Endlich,B. and Linn,S. (1985) *J. Biol. Chem.*, **260**, 5729–5738.
15 Studier,F.W. and Bandyopadhyay,P.K. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 4677–4681.
16 Szczelkun,M.D., Dillingham,M.S., Janscak,P., Firman,K. and Halford,S.E. (1996) *EMBO J.*, **15**, 6335–6347.
17 Dreier,J., MacWilliams,M.P. and Bickle,T.A. (1996) *J. Mol. Biol.*, **264**, 722–733.
18 Barcus,V.A., Titheradge,A.J.B. and Murray,N.E. (1993) *Genetics*, **140**, 1187–1197.
19 Bullas,L.R., Colson,C. and Neufeld,B. (1980) *J. Bacteriol.*, **141**, 275–292.
20 Titheradge,A.J.B., Ternent,D. and Murray,N.E. (1996) *Mol. Microbiol.*, **22**, 437–448.
21 Daniel,A.S., Fuller-Pace,F.V., Legge,D.M. and Murray,N.E. (1988) *J. Bacteriol.*, **170**, 1775–1782.
22 Kannan,P., Cowan,G.M., Daniel,A.S., Gann,A.A.F, and Murray,N.E. (1989) *J. Mol. Biol.*, **209**, 335–344.
23 Gough,J.A. and Murray,N.E. (1983) *J. Mol. Biol.*, **166**, 1–19.
24 Gubler,M., Braguglia,D., Meyer,J., Piekarowicz,A. and Bickle,T.A. (1992) *EMBO J.*, **11**, 233–240.
25 Fuller-Pace,F.V., Bullas,L.R., Delius,H. and Murray,N.E. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 6095–6099.
26 Nagaraja,V., Shepherd,J.C.W. and Bickle,T.A., (1985) *Nature*, **316**, 371–372.
27 Gann,A.A.F., Campbell,A.J.B., Collins,J.F., Coulson,A.F.W. and Murray,N.E. (1987) *Mol. Microbiol.*, **1**, 13–22.
28 Cowan,G.M., Gann,A.A.F. and Murray,N.E. (1989) *Cell*, **56**, 103–109.
29 Cooper,L.P. and Dryden,D.T.F. (1994) *J. Mol. Biol.*, **236**, 1011–1021.
30 Webb,M., Taylor,I.A., Firmage,K. and Kneale,G.G. (1995) *J. Mol. Biol.*, **250**, 181–190.
31 Fuller-Pace,F.V. and Murray,N.E. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 9368–9372.
32 Ochman,H. and Selander,R.K. (1984) *J. Bacteriol.*, **157**, 690–693.
33 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.
34 Hansen,E.B., Atlung,T., Atlung,F.G., Skovgaard,O. and Van Mayenberg,K. (1984) *Mol. Gen. Genet.*, **196**, 387–396.
35 Borck,K., Beggs,J.D., Brammar,W.J., Hopkins,A.S. and Murray,N.E. (1976) *Mol. Gen. Genet.*, **146**, 199–207.
36 Whittaker,P.A., Campbell,A.J.B., Southern,E.M. and Murray,N.E. (1988) *Nucleic Acids Res.*, **16**, 6725–6736.
37 Fuller-Pace,F.V., Cowan,G.M. and Murray,N.E. (1985) *J. Mol. Biol.*, **186**, 65–75.
38 Daniels,D.L., Schroeder,J.L., Blather,F.R., Szybalski,W. and Sanger,F. (1983) In Hendrix,R.W., Roberts,J.W., Stahl,F.W. and Weisberg,R.A. (eds), *Lambda II*. Cold Spring Harbour Laboratory Press, NY. pp. 469–517.
39 Sanger,F., Nicklem,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5465.
40 Sanger,F., Coulson,A.R., Hong,G.F., Hill,D.F. and Peterson,G.B. (1982) *J. Mol. Biol.*, **162**, 729–773.
41 Ryu,J.I., Rajadas,P.T. and Bullas,L.R. (1988) *J. Bacteriol.*, **170**, 5785–5788.
42 Murray,N.E., Daniel,A.S., Cowan,G.M. and Sharp,P.M. (1993) *Mol. Microbiol.*, **9**, 133–143.
43 Devereux,J., Haeberli,P. and Smithies,O. (1984) *Nucleic Acids Res.*, **12**, 387–395.
44 Kim,Y., Grable,J.C., Love,R., Greene,P.J. and Rosenberg,J.M. (1990) *Science*, **249**, 1307–1309.
45 Rosenberg,J.M. (1991) *Curr. Opin. Struct. Biol.*, **1**, 104–113.
46 Winkler,F.K., Banner,D.W., Oefner,C., Tsernoglou,D., Brown,R.S., Heathman, S.P., Bryan,R.K., Martin,P.D., Petratos,K. and Wilson,K.S. (1993) *EMBO J.*, **12**, 1781–1795.
47 Klimasaukas,S., Kumar,S., Roberts,R.J. and Cheng,X. (1994) *Cell*, **76**, 357–369.
48 Willcock,D.F., Dryden,D.T.F. and Murray,N.E. (1994) *EMBO J.*, **13**, 3902–3908.
49 Kneale,G.G. (1994) *J. Mol. Biol.*, **243**, 1–5.
50 Abadjieva,A., Patel,J., Webb,M., Zinkevich,V. and Firman,K. (1993) *Nucleic Acids Res.*, **21,** 4435–4443.
51 Meister,J., MacWilliams,M., Hübner,P., Jütte,H., Skrypek,E., Piekarowicz,A. and Bickle,T.A. (1993) *EMBO J.*, **12**, 4585–4591.
52 MacWilliams,M.P. and Bickle,T.A. (1996) *EMBO J.*, **15**, 4775–4783.
53 Taylor,I.A., Webb,M. and Kneale,G.G. (1996) *J. Mol. Biol.*, **258**, 62–73.
54 Hughes,S.G. (1977) Studies of plasmid encoded restriction and modification systems. PhD thesis, University of Edinburgh, Scotland.
55 Glover,S.W., Firman,K., Watson,G., Price,C. and Donaldson,S. (1983) *Mol. Gen. Genet.*, **190**, 65–69.
56 Skrzpek,E. and Piekarowicz,A. (1989) *Plasmid*, **21**, 195–204.