The EMBO Journal vol.10 no.4 pp.951 - 957, 1991

Increased protein flexibility leads to promiscuous protein - DNA interactions in type IC restriction - modification systems

Marcel Gubler and Thomas A.Bickle

Microbiology Department, Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland

Communicated by T.A.Bickle

We have investigated the role of a four amino acid element that is repeated twice and three times, respectively, in the specificity polypeptides of the two allelic restriction - modification systems *Eco*R124 and *Eco*R124/3. We had earlier shown that this difference in amino acid sequence between the two systems is solely responsible for the different DNA sequence specificities of the two systems. The effect of single amino acid substitutions and small insertion and deletion mutations on restriction activity and modification specificity was determined *in vivo* by phage infection assays and *in vitro* by methylation of DNA with purified modification methylases. Mutant restriction - modification systems with changes in the number and the length of the central amino acid repeats exhibited decreased restriction activity and in some cases relaxed substrate specificity. Our data strongly support the idea that the repetitive amino acid motif in the specificity polypeptides forms part of a flexible interdomain linker. It may be responsible for positioning on the DNA the two major specificity polypeptide domains which are thought to contact independently the half sites of the split recognition sequences typical for all type I restriction - modification systems.

Key words: DNA recognition/relaxed specificity/repetitive sequences/type I restriction systems

Introduction

Type I restriction - modification (R - M) systems have attracted relatively little attention compared with the well known type II restriction endonucleases, largely due to the impact on molecular biology of the latter as tools for recombinant DNA technology. Type I restriction enzymes, on the other hand are interesting objects for the investigation of structure - function relationships. These enzymes are the most complex restriction endonucleases with respect to subunit composition, cofactor requirements and the variety of different reactions that they catalyze: they are endonucleases, methylases and ATPases (see Bickle, 1982, for a review). Moreover, although they recognize and methylate only specific DNA target sites, actual cleavage of non-methylated DNA occurs randomly, from hundreds to thousands of base pairs (bp) distant from the target site (Horiuchi and Zinder, 1972; Adler and Nathans, 1973).

Three genes are necessary to account for all activities of a type I R - M system (Sain and Murray, 1980; Suri and Bickle, 1985; Price *et al.*, 1989). *HsdM* and *hsdS* genes encode the subunits required for a functional modification methylase and are transcribed as an operon from a single promoter in front of *hsdM*. *HsdR* is transcribed from its own promoter and codes for the R subunit which, in assembly with M and S subunits, forms a restriction endonuclease with potent ATPase activity upon cleavage of DNA (Eskin and Linn, 1972).

Structural homologies between different type I R - M systems have led to their division into three distinct families A, B and C (see Bickle, 1987, for a review). Members of the same family exhibit genetic complementation, antigenic crossreaction and, to a variable extent, DNA sequence homology. The C family of type I R - M systems includes *Eco*R124, *Eco*R124/3 (Firman *et al.*, 1985) and *Eco*DXXI (Skrzypek and Piekarowicz, 1989) which reside on large conjugative plasmids and, a fourth member recently localized to the *Escherichia coli* chromosomal *prr* locus (Linder *et al.*, 1991). The key role for DNA target site specificity of type I R - M systems has been assigned to the product of the *hsdS* gene (Boyer and Roulland-Dussoix, 1969; Fuller-Pace *et al.*, 1985; Skrzypek and Piekarowicz, 1989). The DNA sequences recognized by type I R - M systems consist of two non-symmetrical target half sites, each 3 - 5 bp long separated by a spacer of 6 - 8 non-specific nucleotides. For the type IA and IB systems it has been shown that each half site is recognized by a specific domain of the HsdS polypeptide. The regions of the genes coding these domains have no homology with each other (unless they recognize the same DNA sequence) and they are separated by regions of good homology which code amino acids that are thought to be important for interactions between the subunits of the

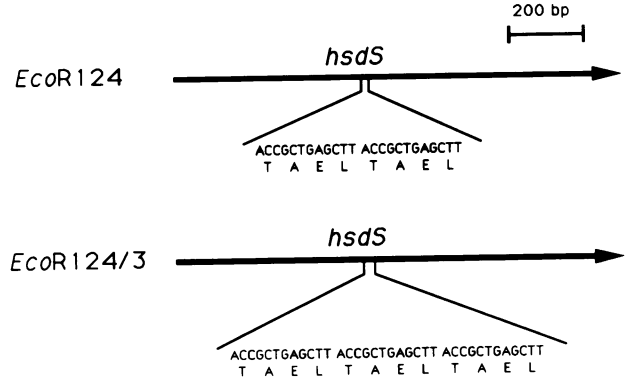

Fig. 1. Position of the repetitive elements in the allelic R - M systems *Eco*R124 and *Eco*R124/3. A 12 bp DNA sequence is repeated two times in the *Eco*R124 and three times in the *Eco*R124/3 *hsdS* gene thus, giving rise to the repeated amino acid motif T-A-E-L in the specificity polypeptides of either system.

enzymes (Nagaraja *et al.*, 1985; Fuller-Pace and Murray, 1986; Cowan *et al.*, 1989; Kannan *et al.*, 1989).

The recognition sites of *Eco*R124 and *Eco*R124/3 have been determined: 5′-GAA(N_6)RTCG-3′ and 5′-GAA(N_7)-RTCG-3′, respectively (Price *et al.*, 1987). They are identical but for the number of nucleotides in the non-specific spacer. In accordance with this striking similarity, the *hsdS* genes were found to be almost identical with the only difference being a 12 bp sequence in the centre of both genes which is repeated two times in the *Eco*R124 and three times in the *Eco*R124/3 *hsdS* genes (Price *et al.*, 1989) (Figure 1). In this work we have investigated the role of the repetitive elements in *Eco*R124 and *Eco*R124/3 HsdS polypeptides. We asked whether particular features of the primary structure of these repeats are important for the specific recognition of DNA target sequences. By site directed mutagenesis we generated mutant R−M systems with variations in the amino acid composition and alterations in the length of the central repeats in *Eco*R124 and *Eco*R124/3 HsdS polypeptides, respectively and determined the effects on the activity and specificity of the mutant enzymes.

Results

Effects of single amino acid exchanges in the central repeats of the EcoR124 specificity polypeptide

Analysis of the *Eco*R124 and *Eco*R124/3 HsdS polypeptide regions in which the amino acid sequence threonine-alanine-glutamate-leucine (T-A-E-L) is repeated two and three times, respectively, strongly predicts an α-helical structure (Chou and Fasman, 1974; Garnier *et al.*, 1978). By site directed mutagenesis of the *Eco*R124 gene, we substituted the leucine residue of the first repeat with either proline, arginine or histidine, and the threonine residue of the second repeat with either proline, alanine or serine. Proline was expected to induce the most drastic alteration due to its inherent property of breaking α-helical structures. The mutant plasmid clones are all derivatives of pMG1 (Figure 2) i. e. *hsdM* and the mutated *hsdS* genes are cloned in the vector pBluescript SK(+). For phage infection assays, the *hsdR* gene was provided in *trans* on the plasmid pMG3 (Figure 2) whose p15A replicon makes it compatible with the ColE1 replicon of the pBluescript clones. Double transformants of *E.coli* WA921 harboring pMG3 plus either the wild-type methylase plasmid pMG1, or one of the mutant methylase clones or the vector pBluescript SK(+) were infected with non-modified λ*vir* phage. The level of restriction expressed as efficiency of plating (e.o.p.) is shown in Table I. All the mutants still restrict phage λ, indeed, they are somewhat more active than the wild type. Even the proline substitutions had no marked effect on enzymatic activity. The specificity conferred by the mutant HsdS polypeptides was analyzed by plating phage λ*vir* that had been modified by the mutant R−M systems on bacteria harboring either the *Eco*R124 or the *Eco*R124/3 wild-type systems on the plasmids pUNG20 and pUNG30, respectively. All phages were restricted by *Eco*R124/3 to an e.o.p. below 10^{-3} but were not restricted by *Eco*R124. This shows that the mutant R−M systems exhibit *Eco*R124 but not *Eco*R124/3 specificity. When λ*vir* modified by *Eco*R124 was plated on cells expressing the mutant R−M systems, no restriction was observed at all, indicating that no additional specificity had arisen in addition to the original *Eco*R124.

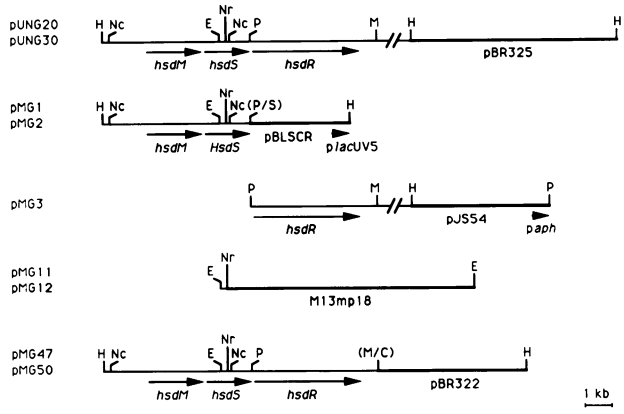

Fig. 2. Structure of the plasmids containing *hsd* DNA used in this work. pUNG20 and pUNG30 express the wild-type R−M systems *Eco*R124 and *Eco*R124/3, respectively (Firman *et al.*, 1985). In pMG1 and pMG2 expression of the respective modification methylase genes, *hsdM* and *hsdS*, is driven by the natural *hsdM* promoter and by the *lac*UV5 promoter of the vector pBluescript SK(+). In pMG3 the *hsdR* gene is transcriptionally fused to the promoter of the kanamycin resistance gene (*aph*) of the vector pJS54. pMG11 and pMG12 are derivatives of phage M13mp18. They contain 197 bp and 209 bp *Eco*RI−*Nru*I inserts spanning the repetitive elements of the *Eco*R124 and *Eco*R124/3 *hsdS* genes, respectively. pMG47 and pMG50 are smaller versions of pUNG20 and pUNG30 and they were used to transfer *hsdS* mutations into the original *hsd* operon. *hsd* DNA is drawn with thin lines, vector DNA with thick lines. The position of the promoters p*lac*UV5 and p*aph* are indicated by arrow heads. Restriction sites: C, *Cla*I; E, *Eco*RI; H, *Hin*dIII; M, *Mlu*I; Nc, *Nco*I; Nr, *Nru*I; P, *Pvu*I; S, *Sma*I.

Table I. Restriction of λ*vir* by plasmid-borne *Eco*R124 HsdS point mutants in *E.coli* WA921[a]

Plasmid	Type of HsdS[b]	e.o.p.[c]
pMG1	*Eco*R124 (wt)	6.0×10^{3}
pMG19	124LP	1.5×10^{-3}
pMG20	124LR	0.8×10^{-3}
pMG21	124LH	2.4×10^{-3}
pMG22	124TP	1.7×10^{-3}
pMG23	124TA	2.1×10^{-3}
pMG24	124TS	2.6×10^{-3}
pBluescript SK(+)	−	1.0

[a]*hsdR* provided in *trans* on pMG3; [b]LP, LR, LH = mutation of leucine to proline, arginine and histidine, respectively in the first HsdS repeat; TP, TA, TS = mutation of threonine to proline, alanine and serine, respectively in the second HsdS repeat; [c]mean of three independent experiments.

Length variation in the repetitive HsdS T-A-E-L motif differentially affects activity and specificity of restriction and modification in vivo

A series of *hsdS* mutants were constructed with variations in the length of the central repetitive T-A-E-L motif. The range of mutations extends from four repeats in the mutant 124/4 to no repeats in the mutant 124/0 as shown in Table II. All the mutations generated in pMG12 by site directed mutagenesis were transferred to pMG1 (Figure 2) to obtain clones encoding mutant modification methylases and also to pMG47 in order to construct non-disrupted mutant R−M systems identical to the original operon structure. The restriction activity exerted by these systems was monitored by phage infection assays in the *recA* host HB101 to prevent recombination events that could occur between the repetitive

Table II. Plasmid-borne R−M systems with length variations in the central repetitive HsdS elements

Plasmid	R−M system	Amino acid sequence in the central repetitive HsdS elements
pMG48	124/4	T A E L T A E L T A E L T A E L
pMG49	124/4ΔTAE	L T A E L T A E L T A E L
pMG50	*Eco*R124/3 (wt)	T A E L T A E L T A E L
pMG51	124/3ΔT	A E L T A E L T A E L
pMG52	124/3ΔTA	E L T A E L T A E L
pMG53	124/3ΔTAE	L T A E L T A E L
pMG54	124/3ΔL	T A E L T A E L T AE
pMG55	124/3ΔEL	T A E L T A E L TA
pMG56	124/3ΔAEL	T A E L T A E LT
pMG47	*Eco*R124 (wt)	T A E L T A EL
pMG57	124/2ΔEL	T A E L TA
pMG58	124/2ΔAEL	T A E LT
pMG59	124/1	T A EL
pMG60	124/1ΔL	T AE
pMG61	124/1ΔEL	TA
pMG62	124/0	–

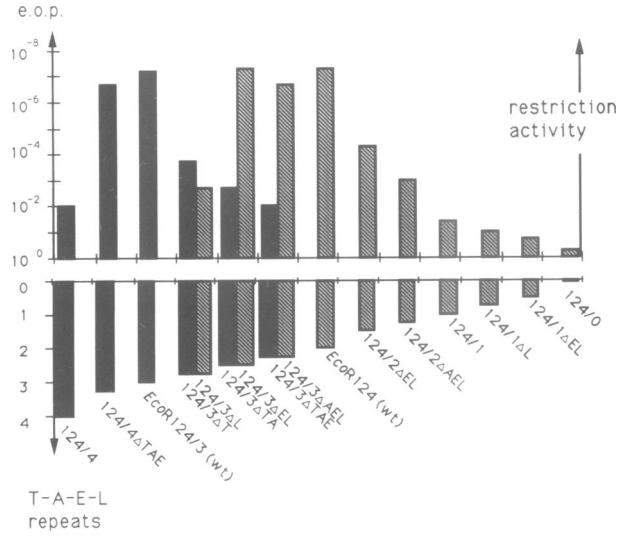

Fig. 3. Efficiency of plating (e.o.p.) of λ*vir* on *E.coli* HB101 expressing the wild-type and mutant R−M systems listed in Table II. The e.o.p. in the upper histogram is a measure of restriction activity. The bars in the lower histogram represent the length of the repetitive T-A-E-L motif in the HsdS polypeptides of the R−M systems indicated. Black and hatched bars differentiate between deletions (Δ) at the amino and carboxy termini of the T-A-E-L motif, respectively.

hsdS sequences. The e.o.p. of non-modified λ*vir* on strains expressing the mutant R−M systems is shown in Figure 3. Three major facts are apparent: (i) reduction in the number of the repetitive elements to less than the two found in *Eco*R124 led to a drastic decrease in restriction activity; (ii) an increase in the number of elements to more than the three found in *Eco*R124/3 also reduced restriction activity; (iii) strains with repeat lengths intermediate between *Eco*R124 and *Eco*R124/3, such as124/3ΔTA and 124/3ΔEL, which are identical in repeat length but permutated in amino acid sequence, exhibit differences in restriction activity of up to four orders of magnitude. This stands in clear contrast to the results obtained with *Eco*R124 point mutants, suggesting that restriction activity is relatively insensitive to changes of amino acid composition in the central HsdS repeats.

A more complex picture of target site specificity emerged

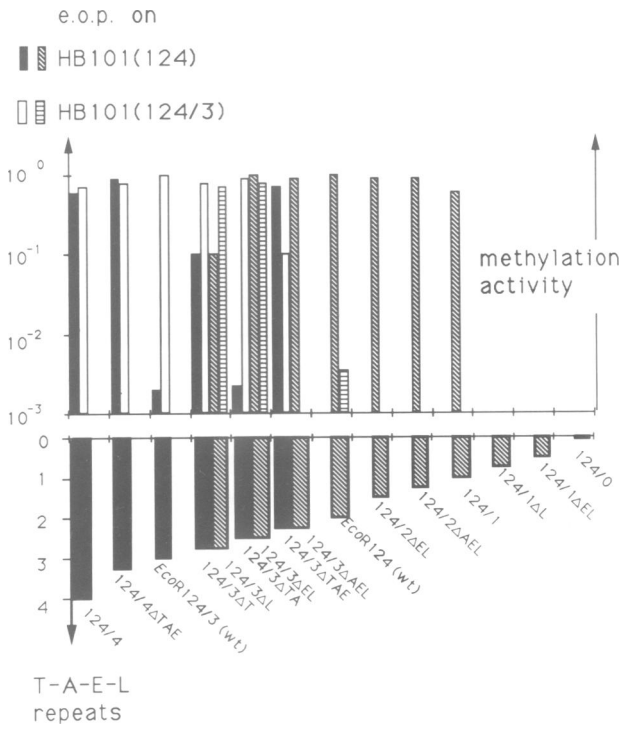

Fig. 4. Efficiency of plating (e.o.p.) of λ*vir* modified by the wild-type and mutant R−M systems on *E.coli* expressing *Eco*R124 or *Eco*R124/3. In the upper histogram an e.o.p. of 10^0 on HB101(124) indicates full modification of GAA(N_6)RTCG and similarly, an e.o.p. of 10^0 on HB101(124/3) indicates full modification of GAA(N_7)RTCG. Black and white e.o.p. bars in the upper panel correlate with black T-A-E-L bars in the lower panel as do hatched e. o. p. bars with hatched T-A-E-L bars. e.o.p. values below 10^{-3} could not be reproducibly measured due to the low titers of the modified λ*vir* phages and are not shown in the histogram.

when the methylation activity of the mutant enzymes was assayed. Phage λ*vir* modified by the mutant R−M systems was used to infect HB101 cells expressing the wild-type *Eco*R124 or *Eco*R124/3 systems. The e.o.p. values shown in Figure 4 are a measure for methylation activity at *Eco*R124 and *Eco*R124/3 target sites. Surprisingly, most of the active enzymes exhibit promiscuous substrate specificity in that they recognize both target sites, though with different efficiency. Even the wild-type enzymes show very low levels of 'allelic' target site specificity. Again, pairs of mutants with repeats of identical length but different sequence not only exhibit different restriction activity but also differ in selectivity for *Eco*R124 and *Eco*R124/3 sites as substrates for methylation. It is also noteworthy that some mutant R−M systems such as 124/1 or 124/4 have very high methylation activity, although they act very poorly as restriction enzymes.

In vitro confirmation of the site specificity of purified mutant modification methylases

Five modification methylases were purified as described in Materials and methods: *Eco*R124, 124/3ΔTA, 124/3ΔEL, *Eco*R124/3 and 124/4. Figure 5 shows the proteins in enzyme containing fractions during the course of purification of all five enzymes. A sixth methylase (124/1) did not bind to the first column (heparin Sepharose) for reasons that are not clear and was therefore not investigated further. All enzymes eluted at approximately the same NaCl concentration (250 mM) from the heparin Sepharose and

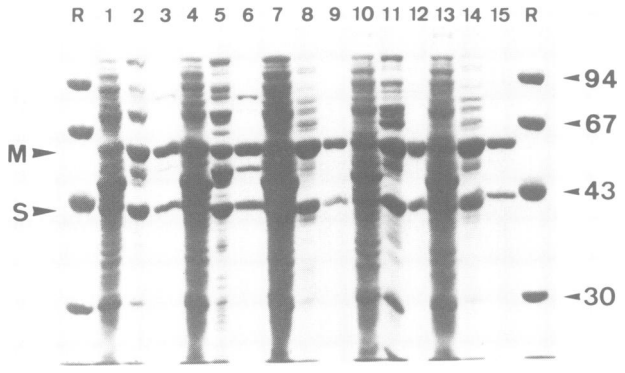

Fig. 5. Purification of modification methylases. The 10% SDS−polyacrylamide gel stained with Coomassie blue shows methylase preparations of *Eco*R124 (lanes 1-3), 124/3ΔTA (lanes 4-6), 124/3ΔEL (lanes 7-9), *Eco*R124/3 (lanes 10-12) and 124/4 (lanes 13-15). Lanes 1, 4, 7, 10, and 13 show whole cell extracts, lanes 2, 5, 8, 11, and 14 pooled fractions after heparin Sepharose chromatography and lanes 3, 6, 9, 12, and 15 pooled fractions after Affi-Gel Blue chromatography. R lanes contain reference proteins whose M_r in kd are indicated to the right of the figure. The positions of HsdM and HsdS subunits are indicated to the left of the figure by M and S.

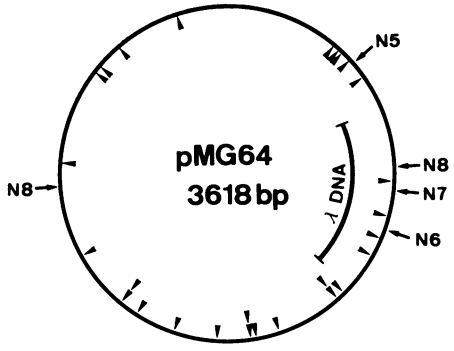

Fig. 6. Methylation target sites and *Fnu*4HI cleavage sites on the pMG64 map. The construction of pMG64 is described in Materials and methods. Arrow heads inside the circle indicate cleavage sites for *Fnu*4HI. Arrows on the outside of the circle point to the positions of possible methylation sites: N5 specifies the sequence GAA(N_5)RTCG in the 73 bp *Fnu*4HI fragment, N6 the sequence GAA(N_6)RTCG in one of the two 81 bp fragments, N7 the sequence GAA(N_7)RTCG in the 122 bp fragment and N8 the sequence GAA(N_8)RTCG in the 328 bp and 380 bp fragments.

Affi-Gel Blue columns. Addition of glycerol to 10% for chromatography on Affi-Gel Blue was essential for the 124/3ΔEL and 124/4 methylases since without glycerol the HsdM and HsdS subunits dissociated on the column. HsdM eluted at 50 mM NaCl whereas HsdS bound very tightly and was eluted at 600 mM NaCl. Methylation activity could not be restored by pooling column fractions containing the dissociated HsdM and HsdS subunits.

In order to have a suitable DNA substrate for *in vitro* methylation experiments, we constructed plasmid pMG64 (Figure 6) which contains the unique sites GAA(N_5)RTCG, GAA(N_6)RTCG, GAA(N_7)RTCG and GAA(N_8)RTCG at locations that can be separated from each other by cleavage with the type II restriction enzyme *Fnu*4HI. pMG64 DNA was methylated with the five purified modification enzymes using [^{3}H-methyl] S-adenosyl methionine as methyl donor

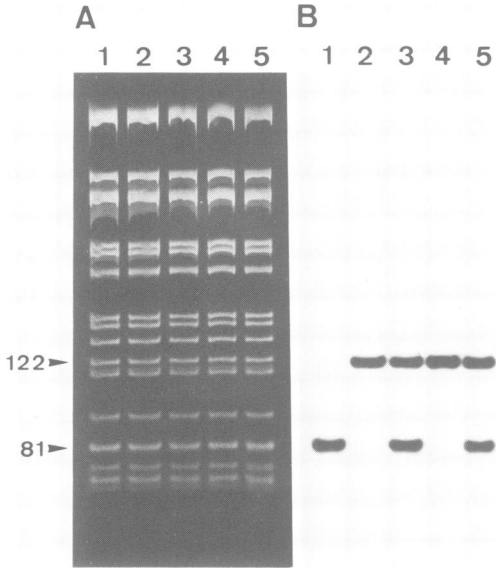

Fig. 7. *In vitro* methylation of pMG64 DNA with purified modification methylases. The DNA was labeled with [^{3}H-methyl] S-adenosyl methionine and subsequently cleaved with *Fnu*4HI as described in Materials and methods. The modification methylases were from the R−M systems *Eco*R124 (lane 1), 124/3ΔTA (lane 2), 124/3ΔEL (lane 3), *Eco*R124/3 (lane 4) and 124/4 (lane 5). **Panel A** shows an ethidium bromide stained 6% polyacrylamide gel, **panel B** shows the fluorogram of the same gel. Arrow heads point to the positions of the 122 bp and the 81 bp fragments.

as described in Materials and methods. Prolonged incubation was found to be necessary as the methylation was very slow, as reported for other type I modification methylases (Suri and Bickle, 1985). The modified plasmid DNA was subsequently cleaved with *Fnu*4HI and the resulting fragments were separated on the polyacrylamide gel shown in Figure 7A. The fluorogram of the same gel is depicted in Figure 7B; it shows the DNA fragments which had been labeled by the purified modification enzymes. The distribution of the [^{3}H-methyl] groups among the DNA fragments and the intensity of the signals represent the substrate specificity of the modification methylases with respect to the four target sites offered. [^{3}H]-label was incorporated in two DNA fragments only, namely in one of the two 81 bp fragments containing the *Eco*R124 site GAA(N_6)RTCG and in the 122 bp fragment containing the *Eco*R124/3 site GAA(N_7)RTCG (Figure 7A and B). No methylation occurred in other DNA fragments, such as for example in the 328 bp and 380 bp fragments which harbor the sequence GAA(N_8)RTCG, a potential target site for 124/4. Thus, we have observed no specificity other than the original ones of *Eco*R124 and *Eco*R124/3. However, the signals in lanes 3 and 5 of Figure 7B clearly confirm the double specificity of the enzymes 124/3ΔEL and 124/4 as was also seen in the phage infection assays. These two enzymes are the first R−M systems which tolerate length variation in their target sequence. Close inspection of the fluorogram also reveals promiscuous target site specificity for *Eco*R124 and 124/3ΔTA (Figure 7B, lanes 1 and 2) because weak signals can be seen at the positions of the 122 bp and 81 bp fragments, respectively. The highest specificity seems to reside in *Eco*R124/3, since even prolonged exposure of the fluorogram did not give rise to a signal from

the *Eco*R124 site in the 81 bp DNA fragment (Figure 7B, lane 4).

Discussion

While the double helical structure of DNA is relatively uniform, proteins which bind to DNA exhibit a variety of structures of which only a few have been identified so far. The DNA binding motifs known to date belong to one of the following categories: helix−turn−helix (Pabo and Sauer, 1984), zinc finger (Evans and Hollenberg, 1988), leucine zipper (Landschulz *et al.*, 1988; Vinson *et al.*, 1989) and helix−loop−helix (Murre *et al.*, 1989). Although these structures are typical for some DNA binding proteins, it is not clear what particular features mediate the specificity in the protein−DNA interaction. In spite of the large selection of sequence specific DNA binding proteins represented by the type II restriction endonucleases, the understanding of how these proteins recognize DNA target sequences is still very poor. Their highly divergent primary structures make it difficult to define structural and functional units. However, the relatively small group of type I R−M systems has offered new possibilities to gain insight into protein structures interacting with specific DNA target sequences. This is largely due to the existence of families of closely related enzymes recognizing different DNA sequences. Here, we have investigated the function of a repetitive four amino acid unit in the HsdS subunits of the two allelic type I R−M systems, *Eco*R124 and *Eco*R124/3. The only difference between these proteins concerns the number of times that the sequence is repeated.

Single amino acid exchanges in the central repeats of the *Eco*R124 HsdS polypeptide did not significantly alter either the specificity or the restriction activity of the mutant R−M systems. Thus, we conclude, that conservation of the amino acid residues in the *Eco*R124 HsdS repeats is not an absolute requirement, for either sequence specificity or for restriction and modification. Even substitutions of proline for leucine or threonine had no effect which makes it very unlikely that the predicted α-helical structure in this region of the protein, if it exists, is of importance for enzymatic function and specificity. It should be noted, that the restriction activities observed in phage infection assays with *Eco*R124 wild-type and point mutants were rather low (e.o.p. 10^{-3}) compared with those observed with spacer length mutants in subsequent experiments (e.o.p. 10^{-8}). This is most certainly due to unbalanced gene expression: the methylase operon was on the high copy number vector pBluescript SK(+) and the *hsdR* gene was on the medium copy number vector pJS54.

Except for extreme reductions to one or no repetitive elements (e. g. in 124/1, 124/1ΔL, etc.), length variations in the *Eco*R124/3 HsdS central repeats did not lead to complete loss of either restriction or modification although none of the mutants restricted as well as the two wild-type R−M systems. This may indicate that this region of the protein, besides its spacer function, also interacts with the HsdR polypeptide. Spacer length variation did not lead to any entirely new specificities, but in some cases (notably 124/3ΔEL and 124/4), sequence specificity was relaxed so that both GAA(N_6)RTCG and GAA(N_7)RTCG were very efficiently recognized. Obviously, length conservation in the central repeats is not an all or none prerequisite for a functional HsdS polypeptide. Therefore the question arises as to why all four type IC R−M systems so far discovered have either two or three repeats (see below). What are the criteria for maintaining an integral number of central repeats in HsdS polypeptides and why is their sequence conserved even at the DNA level? Comparing the data for both restriction and modification specificity of all the R−M systems investigated in this work, it looks as though only *Eco*R124 and *Eco*R124/3 satisfy the criteria for biologically effective R−M systems: they are at the same time (i) highly active in restriction and (ii) efficient and very specific in modification. We assume that the regions of the proteins that include the repeats act as structural elements which correctly position the two HsdS polypeptide domains on either side so that they independently interact with two target half sites on the DNA. The repeats form an 'interdomain linker' with some flexibility that allows for example the 124/3ΔEL enzyme to stretch or contract in order to be able to recognize target sites with six or seven non-specific nucleotides between both target half sites.

It is somewhat surprising that the 124/3ΔEL and 124/3ΔTA enzymes are not identical in their specificity because their HsdS interdomain linkers are of identical length. It is clear that in some cases the amino acid composition of the linker does indeed affect the target site specificity. This is not in contradiction to the results with the *Eco*R124 point mutants since they are all derivatives of the *Eco*R124 wild-type enzyme. 124/3ΔEL and 124/3ΔTA however, are intermediate structures with an obvious tendency to promiscuous substrate interactions. Because the amino acid sequence of their central repeats is permuted, the differences in sequence may influence the flexibility of the interdomain linker, which then explains why the target site specificity of the two systems is not exactly identical.

It is not immediately obvious why the 124/4 R−M system efficiently recognizes GAA(N_6)RTCG in addition to GAA(N_7)RTCG. This seems to stand in contrast to the specificity of *Eco*R124/3 which recognizes only the latter target site. It may be due to extra flexibility in the 124/4 interdomain linker due to its unusual length.

In conclusion, there are several reasons for the conservation of the central repeats in *hsdS* specificity genes: i) they code for structural elements of an interdomain linker which positions two major specificity polypeptide domains on the half sites of split DNA recognition sequences; ii) from the genetic point of view, they may be the fusion point of two ancient genes, both encoding DNA binding proteins recognizing one half of a typical split target site; iii) they may be the site for recombination between *hsdS* genes of other members of the C family of type I R−M systems. In this context it is noteworthy that the *E.coli prr* locus contains a *hsdS* gene with three central repeats 100% identical with those of the *Eco*R124/3 *hsdS* gene (Linder *et al.*, 1991). In addition, preliminary DNA sequence analysis of *Eco*DXXI *hsdS* revealed the same repetitive elements within 180 bp of homology to the corresponding region of the *Eco*R124 *hsdS* gene (M.Gubler, unpublished observation). The association of the repetitive T-A-E-L interdomain linker unit with specificity polypeptides may be considered diagnostic of type IC R−M systems. Based on our results, we propose a modular composure of type IC specificity polypeptides that allows bacterial cells to vary the specificity of R−M systems with mininal effort by recombination of structural elements residing in *hsdS* genes. Among these elements are the

repetitive sequences with their dual function as possible sites for genetic recombination and as structural components determining enzymatic substrate specificity.

Materials and methods

E.coli strains and bacteriophages

WA921 (r_K^- m_K^- *thr leu met*) (Wood, 1966) was used to propagate phage λ*vir* (Jacob and Wollmann, 1954) and for phage infection assays with *hsdS* point mutants. HB101 (r_B^- m_B^- *recA*) (Boyer and Roulland-Dussoix, 1969) was used for the construction and purification of plasmids, for phage infection assays with *hsdS* deletion and insertion mutants and for the purification of mutant modification methylases. ER1451 (*mcrA mcrB*) (New England Biolabs), a derivative of JM107 (Yanisch-Perron *et al.*, 1985) was the host for phage M13mp18 (Yanisch-Perron *et al.*, 1985) and its derivatives constructed in this work.

Plasmid constructions

Plasmids pUNG20 and pUNG30 (Firman *et al.*, 1985) were the source for the R−M systems *Eco*R124 and *Eco*R124/3, respectively. The high copy number plasmids pMG1 and pMG2 expressing the *Eco*R124 and *Eco*R124/3 modification methylases were constructed by cloning the 4.2 kb *Hin*dIII−*Pvu*I fragments from pUNG20 and pUNG30, respectively into the vector pBluescript SK(+) (Stratagene Cloning Systems) cut with *Hin*dIII and *Sma*I. The *Pvu*I sites were made blunt with T4 DNA polymerase to join them to the *Sma*I site of the vector. For site directed mutagenesis of the central 12 bp repeats in each *hsdS* gene, the 197 bp and the 209 bp *Eco*RI−*Nru*I fragments from pMG1 and pMG2, respectively, were subcloned into pBluescript SK(+) digested with *Eco*RI and *Cla*I. A filling in reaction at the 5′ protruding *Cla*I end of the vector DNA allowed its ligation to the blunt *Nru*I end of the insert DNA and regenerated the *Nru*I site in the resulting plasmids pMG7 and pMG8. Both *hsdS* segments were subsequently retrieved from pMG7 and pMG8 as *Eco*RI−*Sal*I cartridges and cloned into M13mp18 cleaved with *Eco*RI and *Sal*I. Single stranded DNA of the resulting clones pMG11 and pMG12 was then used as target for site directed mutagenesis. The mutated *hsdS* DNA in pMG11 and pMG12 progeny phages was used to replace the wild-type *Eco*RI−*Nru*I fragment in pMG1 and resulted in clones expressing mutant modification methylases. Reconstitution of complete *hsd* operons for the coordinate expression of mutant modification and restriction activities was achieved by first subcloning the wild-type *Eco*R124 *hsd* operon as an 8.0 kb *Hin*dIII−*Mlu*I fragment from pUNG20 into pBR322 (Bolivar *et al.*, 1977) linearized with *Hin*dIII and *Cla*I giving rise to pMG47. In a second step, the *hsdS* mutations were excised from the various modification methylase clones as 3.5 kb *Nco*I fragments and used to replace the corresponding *Eco*R124 *Nco*I fragment in pMG47. The clones thus obtained (pMG48 to pMG62) all express R−M systems differing only in the length of the central 12 bp repeats of *hsdS*.

Plasmid pMG3 is a derivative of pACYC177 (Chang and Cohen, 1978) which expresses the *Eco*R124 *hsdR* gene and confers chloramphenicol resistance. Prior to insertion of *hsdR*, the original vector had been modified as follows: the ampicillin resistance gene of pACYC177 had been inactivated by deletion of the small internal *Hin*cII−*Pvu*I fragment to yield pAW329 (R.Stalder, personal communication). The *Bam*HI chloramphenicol resistance cartridge from pAW302 (Stalder and Arber, 1989) was then inserted at the unique *Bam*HI site resulting in pJS54 (Spring, 1986). The 120 bp *Pvu*I−*Hin*dIII fragment of the kanamycin resistance gene of pJS54 was finally replaced by the 9.8 kb *Pvu*I−*Hin*dIII fragment from pUNG20.

Plasmid pMG64 contains a 654 bp *Msp*I fragment of phage λ DNA cloned into the *Cla*I site of pBluescript II SK(+) (Stratagene Cloning Systems). λcI857Sam7 DNA (New England Biolabs) was first cleaved with *Hin*dIII and *Eco*RI and the isolated 831 bp fragment was then cut with *Msp*I. The resulting 654 bp *Msp*I fragment extends from nucleotide positions 44225 to 44879 in the λ DNA sequence (Daniels *et al.*, 1983).

Site-directed mutagenesis and DNA sequence analysis

The procedure for site-directed mutagenesis was that described by Vandeyar *et al.* (1988) based on the incorporation of 5-methyl-dCTP in the mutant strand and subsequent treatment of the hemimethylated DNA with *Msp*I and exonuclease III. The two oligonucleotides used to introduce point mutations into the *Eco*R124 *hsdS* gene were 17mers degenerate at position 9, a position corresponding to the second nucleotide of the leucine codon and the first nucleotide of the threonine codon in the centre of the 12 bp repeats, respectively. The oligonucleotides used to generate *hsdS* length mutants were 20mers lacking one, two or three contiguous triplets on the 5′ or 3′ side of the 12 bp repeats of the *Eco*R124/3 *hsdS* gene. As the mutagenic oligonucleotides could prime at any one of the 12 bp repeats, they gave rise to sets of mutant *hsdS* genes retaining identical fractions of a single repeat but varying in the integral number of repeats; e. g. 124/3ΔEL, 124/2ΔEL, 124/1ΔEL were all produced with a single oligonucleotide. The mutagenic oligonucleotides were designed such that no codons other than those occurring in the *Eco*R124 and *Eco*R124/3 wild-type repeats were introduced in the mutant *hsdS* genes except for that of 124/4. The *hsdS* mutation in 124/4 was generated using a 47mer oligonucleotide in which the thymine residue at position 12 of the fourth repeat is replaced by cytosine. It generated a *Sac*I restriction site and resulted in the change of the original leucine codon CTT to the analogue CTC. The structure of all of the mutants was verified by DNA sequencing.

DNA sequence analysis was performed using T7 DNA polymerase from PL-Pharmacia and the conditions suggested by the supplier. Sequencing templates were either single stranded DNA from M13mp18 derivatives or denatured double stranded plasmid DNA in conjunction with a M13 sequencing primer from New England Biolabs or with a 17mer oligonucleotide hybridizing ~70 bp downstream from the various *hsdS* mutations.

Phage infection assays

E.coli cells were grown in 2× YT containing 0.2% maltose, 200 μg/ml ampicillin and, where appropriate 20 μg/ml chloramphenicol to select for plasmids harboring wild-type or mutant *hsd* genes. At an OD_{600} of 1.0 the cells were infected with serial dilutions of λ*vir*, kept at 37°C for 30 min for adsorption and were then mixed with molten top agar prior to pouring on 2× YT agar without antibiotics followed by over night incubation at 37°C.

Protein purification

The wild-type and mutant modification methylases were purified using the same protocol. Purification of the enzymes was monitored by Western blot analysis (Towbin *et al.*, 1979) of column fractions using antibodies raised against the purified *Eco*R124 enzyme. *E.coli* HB101 harboring high copy number modification methylase plasmids were grown in 2× YT supplemented with 200 μg/ml ampicillin at 37°C overnight to stationary phase. All the following steps were carried out at 4°C. The cells from a two liter culture were resuspended in 20 ml of buffer A (20 mM Tris−HCl, pH 8.0; 50 mM NaCl; 5 mM $MgCl_2$; 7 mM 2-mercaptoethanol) and broken by two passages through a French pressure cell at 20 000 psi. The extract was clarified by high speed centrifugation (100 000 *g*) for 2 h. The NaCl concentration in the supernatant was increased to 400 mM and the nucleic acids were precipitated by addition of neutralized polyethyleneimine to 0.4%. The proteins in the supernatant were precipitated in 70% ammonium sulphate, redissolved in 4 ml of buffer B (20 mM Tris−HCl, pH 8.0; 50 mM NaCl; 0.5 mM EDTA; 7 mM 2-mercaptoethanol) and dialyzed against the same buffer. The proteins were applied to a 3 ml column of heparin Sepharose CL-6B (PL-Pharmacia), washed extensively with buffer B and eluted with a 50-1000 mM NaCl gradient. Fractions containing enzyme were pooled, dialyzed against buffer B containing 10% glycerol and adsorbed to a 3 ml Affi-Gel Blue (Bio Rad) column. Elution was done with a 50-1000 mM NaCl gradient and active fractions were pooled and supplemented with glycerol to 44% for storage at −20°C.

Protein methods and enzyme assays

SDS−polyacrylamide gel electrophoresis was carried out as described by Laemmli (1970), Western blotting was done according to Towbin *et al.* (1979) and immunologic detection of Hsd proteins has been described by Murray *et al.* (1982). *In vitro* DNA methylation assays were carried out in buffer containing 100 mM Tris−HCl, pH 8.0; 0.25 mM EDTA; 6.5 mM $MgCl_2$; 7 mM 2-mercaptoethanol. pMG64 plasmid DNA was present at 0.5 μg/ml and [^{3}H-methyl] S-adenosyl methionine (85 Ci/mmol) at 1 μM. For every μg of DNA approximately 0.5 μg purified methylase enzyme was added and the reactions were incubated at 37°C during 12−14 h. Following phenol extraction and removal of non-incorporated radioactive material by gel filtration, the DNA samples were cut with *Fnu*4HI and electrophoretically separated on a 6% polyacrylamide gel. DNA fragments were visualized by staining with ethidium bromide and [^{3}H]-labeled fragments were detected by fluorography (Chamberlain, 1979).

Acknowledgements

We thank Clive Price for a series of *hsd* plasmids and help to get this project started; Sheikh Hadi for his support in protein purification and for many interesting discussions; and Gabi Rebmann for critically reading the manuscript. This work was supported by the Swiss National Science Foundation.

References

Adler,S.P. and Nathans,D. (1973) *Biochim. Biophys. Acta*, **299**, 177–188.

Bickle,T.A. (1982) In Linn,S.M. and Roberts,R.J. (eds), *Nucleases*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor NY, pp. 85–108.

Bickle,T.A. (1987) In Neidhardt,F.C., Ingraham,J.L., Low,K.B., Magasanik,B., Schaechter,M. and Umbarger,H.E. (eds), *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*. American Society for Microbiology, Washington DC, pp. 692–696.

Bolivar,F., Rodriguez,R.L., Greene,P.J., Betlach,M.C., Heyneker,H.L. and Boyer,H.W. (1977) *Gene*, **2**, 95–113.

Boyer,H.W. and Roulland-Dussoix,D.(1969) *J. Mol. Biol.*, **41**, 459–472.

Chamberlain,J.P. (1979) *Anal. Biochem.*, **98**, 132–135.

Chang,A.C.Y. and Cohen,S.N. (1978) *J. Bacteriol.*, **134**, 1141–1156.

Chou,P.Y. and Fasman,G.D. (1974) *Biochemistry*, **13**, 222–245.

Cowan,G.M., Gann,A.A.F. and Murray,N.E. (1989) *Cell*, **56**, 103–109.

Daniels,D.L., Schroeder,J.L., Blattner,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 Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 519–676.

Eskin,B.and Linn,S. (1972) *J. Biol. Chem.*, **247**, 6192–6196.

Evans,R.M. and Hollenberg,S.M. (1988) *Cell*, **52**, 1–3.

Firman,K., Price,C. and Glover,S.W. (1985) *Plasmid*, **14**, 224–234.

Fuller-Pace,F.V. and Murray,N.E. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 9368–9372.

Fuller-Pace,F.V., Cowan,G.M. and Murray,N.E. (1985) *J. Mol. Biol.*, **186**, 65–75.

Garnier,J., Osguthorpe,D.J. and Robson,B. (1978) *J. Mol. Biol.*, **120**, 97–120.

Horiuchi,K. and Zinder,N.D. (1972) *Proc. Natl. Acad. Sci.USA*, **69**, 3220–3224.

Jacob,F. and Wollman,E.L. (1954) *Ann. Inst. Pasteur*, **87**, 653–673.

Kannan,P., Cowan,G.M., Daniel,A.S., Gann,A.A.F. and Murray,N.E. (1989) *J. Mol. Biol.*, **209**, 335–344.

Laemmli,U.K. (1970) *Nature*, **227**, 680–685.

Landschulz,W.H., Johnson,P.F. and McKnight,S.L. (1988) *Science*, **240**, 1759–1764.

Linder,P., Doelz,R., Gubler,M. and Bickle,T. A. (1991) *Nucleic Acids Res.*, **18**, 7170.

Murray,N.E., Gough,J.A., Suri,B. and Bickle,T.A. (1982) *EMBO J.*, **1**, 535–539.

Murre,C., McCaw,P.S. and Baltimore, D. (1989) *Cell*, **56**, 777–783.

Nagaraja,V., Shepherd,J.C.W. and Bickle,T.A. (1985) *Nature*, **316**, 371–372.

Pabo,C.O. and Sauer,R.T. (1984) *Annu. Rev. Biochem.*, **53**, 293–321.

Price,C., Lingner,J., Bickle,T.A., Firman,K. and Glover,S.W. (1989) *J. Mol. Biol.*, **205**, 115–125.

Price,C., Shepherd,J.C.W. and Bickle,T.A. (1987) *EMBO J.*, **6**, 1493–1497.

Sain,B. and Murray,N.E. (1980) *Mol. Gen. Genet.*, **180**, 35–46.

Skrzypek,E. and Piekarowicz,A. (1989) *Plasmid*, **21**, 195–204.

Spring,J. (1986) Diploma Thesis, *Investigation of the Regulation of the Restriction Genes of R124 and R124/3*. Biozentrum, University of Basel, Basel.

Stalder,R. and Arber,W. (1989) *Gene*, **76**, 187–193.

Suri,B. and Bickle,T.A. (1985) *J. Mol. Biol.*, **186**, 77–85.

Towbin,H., Staehelin,T. and Gordon,J. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 4350–4353.

Vandeyar,M.A., Weiner,M.P., Hutton,C.J. and Batt,C.A. (1988) *Gene*, **65**, 129–133.

Vinson,C.R., Sigler,P.B. and McKnight,S.L. (1989) *Science*, **246**, 911–916.

Wood,W.B. (1966) *J. Mol. Biol.*, **16**, 118–133.

Yanisch-Perron,C., Vieira,J. and Messing, J. (1985) *Gene*, **33**, 103–119.

Received on November 15, 1990; revised on January 4, 1991