The EMBO Journal vol.6 no.5 pp.1493-1497, 1987

DNA recognition by a new family of type I restriction enyzmes: a unique relationship between two different DNA specificities

Clive Price, John C.W.Shepherd and Thomas A.Bickle

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

Communicated by T.A.Bickle

The DNA sequences recognized by the allelic type I restriction enzymes *Eco*R124 and *Eco*R124/3 were determined. *Eco*R124 recognizes 5′-GAA(N_6)RTCG-3′ and *Eco*R124/3 recognizes 5′-GAA(N_7)RTCG-3′. These are typical of sequences recognized by type I recognition enzymes in that they consist of two specific domains separated by a non-specific spacer sequence. For these two enzymes, the specific sequences are identical but the length of the non-specific spacer is different. The specific domains of *Eco*R124/3 are thus 3.4 Å further apart than those of *Eco*R124 and rotated with respect to each other through a further 36°.

Key words: DNA recognition/type I restriction enzymes/ *Eco*R124/*Eco*R124/3

Introduction

R124 is a large conjugative plasmid belonging to the IncFIV incompatibility group and encoding a unique DNA restriction and modification system (Hedges and Datta, 1972; Bannister and Glover, 1968). R124/3 is a derivative of R124 that carries a restriction and modification system of altered specificity (Firman *et al.*, 1983). Both specificities are carried out on the same plasmid but only one of them is normally expressed (Firman *et al.*, 1983), and it is thought that one specificity has arisen from the other by a process of gene duplication and sequence divergence. Moreover under certain conditions the plasmids are able to switch specificity (Glover *et al.*, 1983).

We have purified and characterized the restriction enzymes coded by the R124 and R124/3 plasmids, *Eco*R124 and *Eco*R124/3 and have shown that they belong to the type I class of restriction systems (Price, 1984). Type I R-M systems all consist of three subunits which combine to form a complex enzyme exhibiting DNA cleavage, ATP hydrolysis, DNA methylation and topoisomerase activities (reviewed in Yuan, 1981; Bickle, 1982). Genetic analysis has defined three genes coding type I R-M systems; *hsdR*, coding the restriction subunit and *hsdM* and *hsdS* encoding the modification and specificity subunits respectively (Hubacek and Glover, 1970; Glover and Colson, 1969). Recently, it has been shown that the type I R-M systems form at least two distinct families, type IA and type IB (Murray *et al.*, 1982; Fuller-Pace *et al.*, 1985; Suri and Bickle, 1985). The members of each family are allelic, as shown by genetic complementation, DNA hybridization and immunological cross reactivity, but the two families are unrelated to each other by these criteria (Murray *et al.*, 1982; Fuller-Pace *et al.*, 1985; Suri and Bickle, 1985). We have shown that *Eco*R124 and *Eco*R124/3 form a third family of type I systems which we call type IC (Bickle, 1987).

The DNA sequences recognized by eight different type I R-M systems have been determined (Kan *et al.*, 1979; Lautenberger *et al.*, 1978; Ravetch *et al.*, 1978; Sommer and Schaller, 1979; Kröger and Hobom, 1984; Suri *et al.*, 1984; Nagaraja *et al.*, 1985a,b,c; Piekarowicz and Goguen, 1985). They all have a very characteristic structure in that they consist of two specific sequences separated by a non-specific spacer sequence of fixed length, between 6 and 8 bp long depending on the system. When the sequences are modified, one adenosyl residue in each of the specific domains is methylated but on opposite strands, so that a fully modified site has methyl groups in both strands of the DNA.

The study of the DNA sequences recognized by allelic families of enzymes is expected to aid in the understanding of how these enzymes recognize DNA. The most fruitful of these studies so far has been with the *Salmonella* type I systems, *Sty*SP, *Sty*SB and a new specificity, *Sty*SQ that arose by recombination between the *hsdS* genes of *Sty*SB and *Sty*SP (Bullas *et al.*, 1976; Fuller-Pace *et al.*, 1984). The DNA sequence recognized by *Sty*SQ is a hybrid of those recognized by *Sty*SB and *Sty*SP in that it consists of one specific domain from *Sty*SB and the other from *Sty*SP (Nagaraja *et al.*, 1985a,b), indicating that the recombination event that generated *Sty*SQ reassorted two domains of the protein molecule that are involved in DNA recognition. Despite these analyses, the DNA binding domains of any of the type I *hsdS* gene products remain ill defined. We are studying the *Eco*R124 and *Eco*R124/3 systems because they are the most closely related type I R-M systems so far discovered. In particular, the two *hsdS* genes are so closely related that they appear identical when analysed by DNA-DNA hybridization techniques (Price, 1984), and we hope that differences in the DNA recognition sequences can be correlated with differences in the *hsdS* gene product sequences. In this paper we describe the two DNA sequences recognized by the two systems and show that they have a unique relationship to each other.

Results

The experimental approach used to determine the *Eco*R124 and *Eco*R124/3 recognition sequences was essentially that employed by Nagaraja *et al.* (1985b). It involves the *in vitro* methylation by the enzymes of DNA molecules of known sequence using [methyl-^{3}H] S-adenosyl methionine (AdoMet) as the methyl donor. The methylated sites in the DNA are then mapped and a computer is used to find sequences present in all methylated regions that are absent from non-methylated regions. Such sequences are candidates for being the recognition sequence(s) and they must be tested in further experiments using other sequenced DNA molecules.

Determination of the EcoR124 recognition sequence

Initially, various sequenced plasmid and phage DNA molecules were screened for the presence of methyl accepting sites. This resulted in the identification of two plasmids, pBR325::IS*5* and pBR325::IS*1*, both of which had a single site within the insertion sequences. Restriction enzyme mapping of these sites revealed one of them to be between coordinates 71 and 158 of the IS*5*

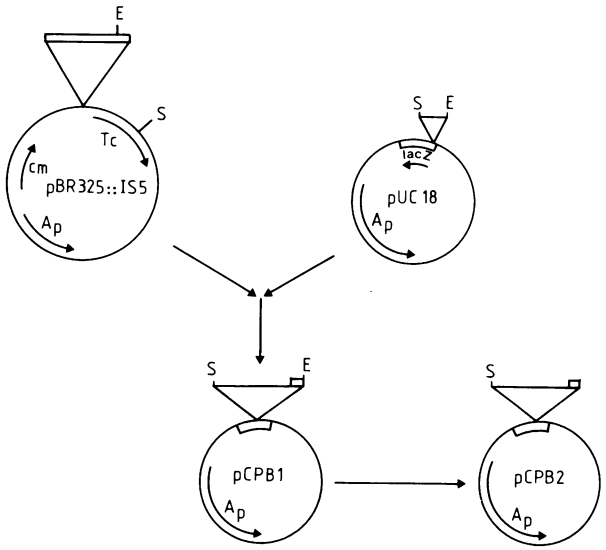

Fig. 1. Creation of an *Eco*R124 site. The smallest *Eco*RI−*Sal*I fragment of pBR325::IS*5* was cloned into *Eco*RI- and *Sal*I-cut pUC18 to yield the plasmid pCPB1 which now contains an *Eco*R124 site. The plasmid pCPB1 was cleaved with *Eco*RI and the ends were made blunt by treatment with DNA polymerase I in the presence of dATP and TTP. Ligation yielded the plasmid pCPB2 which no longer has an *Eco*R124/3 site.

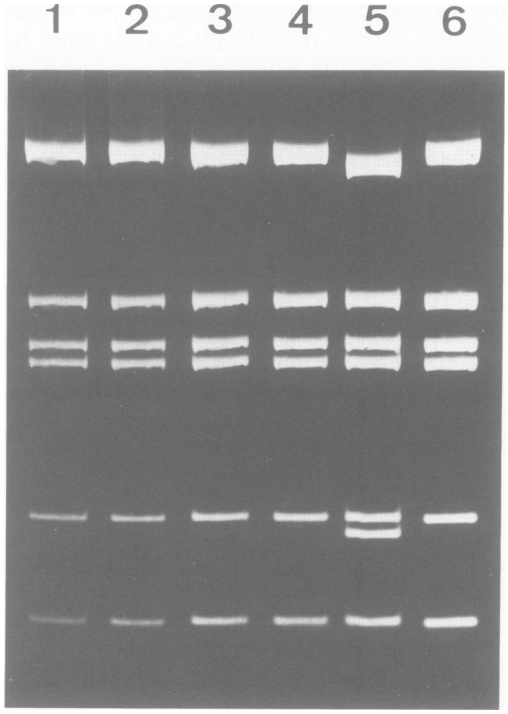

Fig. 2. The coincidence of the *Eco*R124 site in IS*5* with the *Eco*RI site. The figure shows an ethidium bromide-stained 4% polyacrylamide gel. **Lane 1**, pCPB2 DNA cleaved with *Eco*RI and *Dde*I; **lane 2**, pCPB2 DNA cleaved with *Dde*I alone; **lane 3**, pCPB1 DNA modified *in vivo* with *Eco*R124 and cleaved with *Eco*RI and *Dde*I; **lane 4**, the same DNA cleaved with *Dde*I alone; **lane 5**, pCPB1 DNA unmodified by *Eco*R124 and cleaved with *Eco*RI and *Dde*I; **lane 6**, the same DNA cleaved with *Dde*I alone.

sequence (Schoner and Kahn, 1981) and the other to be between coordinates 71 and 158 of IS*1* (Ohtsubo and Ohtsubo, 1978). Computer analysis identified only one sequence common to these two regions that was also present in the phage λ genome sufficiently frequently to account for the methylation data with λ DNA: GAA(N)$_6$GTCG. This sequence precisely overlaps the single *Eco*RI site at position 99 in IS*5* in the sequence *GAATTC*-ACTGTCG where the *Eco*RI recognition sequence is in italics. Thus, cloning of the smallest *Sal*I−*Eco*RI fragment of pBR325:: IS*5* into pUC8 (which contains no *Eco*R124 site) should create an *Eco*R124 site at the *Eco*RI point of insertion. This cloning strategy is outlined in Figure 1, and the resulting plasmid is called pCPB1 which we showed, by methylation, had acquired an *Eco*R124 site. If the data were correctly interpreted, two predictions can be made. Firstly, pCPB1 DNA modified *in vivo* by the *Eco*R124 R-M system should be refractory to cleavage by *Eco*RI and, secondly, destruction of the *Eco*RI site should result in the concomitant loss of the *Eco*R124 site. Both predictions are borne out by the data presented in Figure 2. Lane 3 of Figure 2 shows that *Eco*R124-modified pCPB1 DNA is resistant to cleavage by *Eco*RI despite a many-fold over-digestion. Lanes 1 and 2 show the plasmid pCPB2 (Figure 1) in which the *Eco*RI site of pCPB1 has been destroyed by filling in the *Eco*RI sticky ends with DNA polymerase followed by religation. pCPB2 is not a substrate for methylation by *Eco*R124. We believe that these data adequately confirm that *Eco*R124 recognizes the sequence GAA(N)$_6$GTCG.

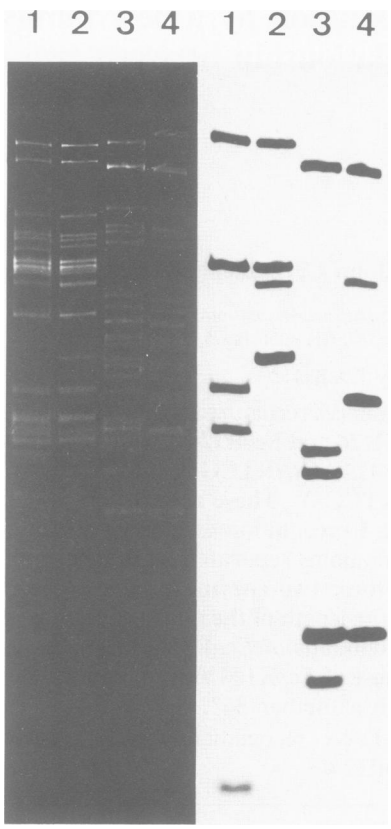

Fig. 3. *In vitro* labelling of *Eco*R124 sites on pIL254 DNA. The DNA was labelled with [methyl-^{3}H] AdoMet and *Eco*R124 as described in Materials and methods. **Lane 1**, cut with *Hin*fI and *Fsp*I; **lane 2**, cut with *Hin*fI alone; **lane 3**, cut with *Hae*III and *Fsp*I; **lane 4**, cut with *Hae*III alone. The **left panel** shows an ethidium bromide-stained 4% polyacrylamide gel, the **right panel** shows the fluorogram of the same gel.

This sequence occurs four times in the λ genome (Sanger *et al.*, 1982) and one of these sequences is found in the 7.55-kb *Eco*RI fragment that includes the immunity region. The plasmid pIL254 contains this fragment (derived from λ *c*I857*s*7) cloned

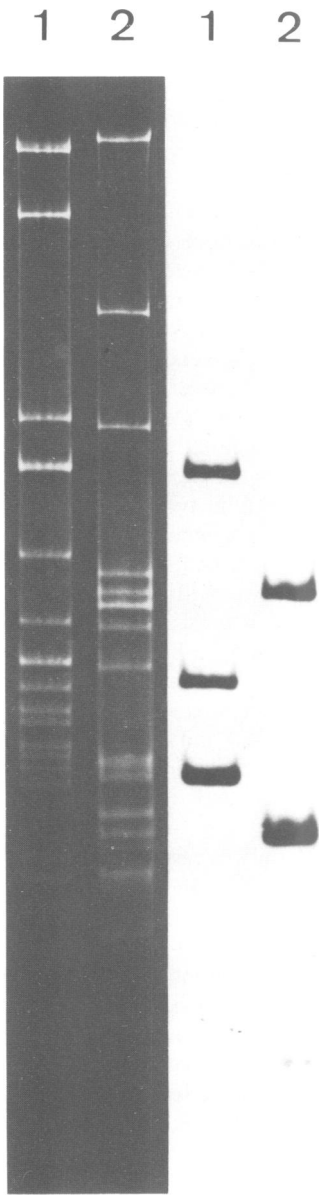

Fig. 4. Identification of the *Eco*R124/3 sites in pSC101 DNA. The DNA was methylated *in vitro* with *Eco*R124/3 and then cut with *Hae*III **(lane 1)** or *Fnu*DII **(lane 2)**. The **left panel** shows the ethidium bromide-stained gel and the **right panel** shows the fluorogram of the same gel.

into pBR322. Figure 3 demonstrates that when this plasmid is methylated *in vitro* by *Eco*R124 and then cleaved with *Hin*fI or *Hae*III four labelled DNA fragments, all of them derived from the phage λ insert, can be detected. The *Eco*R124 recognition sequence is thus degenerate, as are the recognition sequences of the other type I systems *Eco*D, *Sty*SB, *Sty*SP and *Sty*SQ (Nagaraja *et al.*, 1985a,b,c,). When the regions labelled in the experiment shown in Figure 3 were compared, a second sequence was found which occurred nowhere else in the λ insert or in the vector: $GAA(N)_6ATCG$. Fortuitously one of these sites, located at position 32679 in the λ DNA sequence (Sanger *et al.*, 1982), contains an *Fsp*I site within the non-specific spacer. Thus cleavage of the methylated plasmid with *Hin*fI together with *Fsp*I (or *Hae*III and *Fsp*I) should produce five rather than four labelled fragments. Figure 3 confirms that this is indeed the case and the *Eco*R124 recognition sequence is thus $GAA(N)_6RTCG$ where R can be

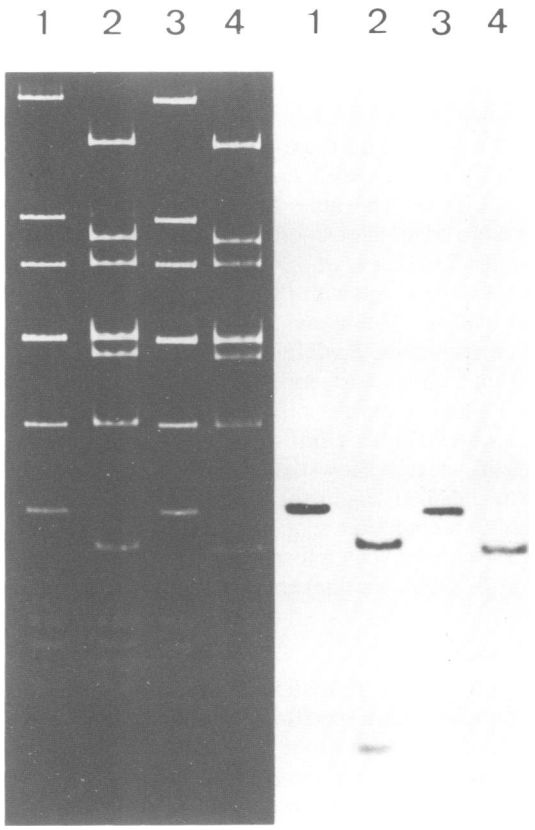

Fig. 5. *Eco*R124/3 methylation of pUC4K DNA. **Lanes 1** and **2** show pUC4K DNA isolated from a dam^- host, methylated *in vitro* with *Eco*R124/3, and cleaved with *Hin*fI **(lane 1)** or *Hin*fI plus *Pvu*I. **Lanes 3** and **4** show pUC4K DNA isolated from a dam^+ host that was methylated *in vitro* with *Eco*R124/3 and cleaved with *Hin*fI **(lane 3)** or *Hin*fI plus *Pvu*I **(lane 4)**. The **left panel** shows the ethidium bromide-stained gel, the **right panel** shows the fluorogram of the same gel.

either purine. Phage λ DNA contains 14 *Eco*R124 sites, and this was confirmed by methylation, and phage T7 DNA (Dunn and Studier, 1983) contains 10 sites. An examination of these 24 sites showed that the positions 5′ and 3′ to the sites as well as all positions in the non-specific spacer can be occupied by any nucleotide.

Determination of the EcoR124/3 recognition sequence

Screening of sequenced plasmid DNAs revealed the presence of recognition sites for *Eco*R124/3 in pBR322 and in pSC101. These sites were mapped by methylation and restriction analysis. The single site in pBR322 mapped to the coordinates 1303-1424 (coordinates from Sutcliffe, 1979) and the three sites in pSC101 (Figure 4) mapped to 1261-1445, 2831-3362 and 8943-8677 (Bernardi and Bernardi, 1984). A comparison of all of these regions failed to reveal a single sequence as a candiate for the *Eco*R124/3 recognition site. However, a comparison of the pBR322 region with the pSC101 region between 1261 and 1445 revealed a sequence $GAA(N)_7ATCG$ that was found in these two regions and which occurred nowhere else in pBR322 or in non-methylatable regions of pSC101. A comparison of this sequence with the two remaining methylatable regions of pSC101 revealed a related sequence, $GAA(N)_7GTCG$, which is absent from non-methylatable regions.

We tested both of these putative *Eco*R124/3 recognition sequences separately. The sequence at PSC101 coordinate 8905

contains an *Fnu*DII site within the non-specific spacer in the sequence GAAGCTGCGCGTCG (*Fnu*DII site underlined) and cleavage by *Fnu*DII of *Eco*R124/3 methylated pSC101 DNA should produce four labelled fragments rather than three. Figure 4 shows that this is the case. Furthermore, the size of the two smallest fragments are those predicted by cleavage within the *Eco*R124/3 site. This result confirms that *Eco*R124/3 recognizes GAA(N)$_7$GTCG. The alternative sequence, GAA(N)$_7$ATCG, was found to occur at position 1590 in the kanamycin resistance gene of Tn*903* (Oka *et al.*, 1981) and this site contains a *Pvu*I site such that cleavage with *Pvu*I should cut the site in two. The plasmid pUC4K (Vieira and Messing, 1982) contains this site and no other putative *Eco*R124/3 site. Plasmid pUC4K was prepared from a *dam*$^-$ host, for reasons to be described later, and Figure 5 shows cleavage of this DNA after methylation with *Eco*R124/3 with either *Hin*fI alone or *Hin*fI together with *Pvu*I. The results clearly show that the *Pvu*I digestion has created two extra labelled fragments of the expected size. Thus, the *Eco*R124/3 recognition sequence is GAA(N)$_7$RTCG. As for *Eco*R124 we have shown that the 15 *Eco*R124/3 sites in λ and the 12 in T7 DNA are flanked by all possible nucleotide combinations.

Sites of methylation

As mentioned in the Introduction, type I R-M systems modify DNA by methylating specific adenosyl residues to N^6-methyladenosine, one in each of the specific domains of the sequence and in opposite strands of the DNA. It is evident from inspection of the sequences that both the *Eco*R124 and *Eco*R124/3 recognition sequences can sometimes overlap with the recognition site of the *Escherichia coli* DNA adenine methylase, the *dam* gene product which methylates GATC. The recognition sequence of *Eco*R124/3 in pUC4K is one such site: GAAATACGC*GATC*G, the *dam* site being italicized. Thus, *dam* methylated pUC4K DNA will contain two methylated adenosyl residues in the ATCG domain of the *Eco*R124/3 recognition site, one of them in the first A shown and the other in the A complementary to the T residue. It is this second A that must be methylated in *Eco*R124/3-modified DNA for two reasons. Firstly, the GAA domain of the recognition sequence contains no A residues in the other strand and must therefore be methylated on one of these two As, predicting that the other specific domain must be methylated in the other strand. Secondly, only one of the two possible RTCG domains contains an A residue in the strand as written, again predicting that methylation in this domain has to be in the opposite strand. The experiment shown in Figure 5 gives the result predicted from the above considerations. It shows that when pUC4K DNA is purified from *dam*$^+$ *E. coli* cells, *Eco*R124/3 can only transfer [^{3}H]methyl groups to the GAA domain of the recognition sequence. This is also true of the *Eco*R124 site at position 34776 in the λ DNA sequence, although in this case there is no type II restriction site in the non-specific spacer and the evidence is that the restriction fragment containing the site (see the *Hin*fI digest in Figure 3) has only about half of the radioactivity of the other fragments containing sites.

Despite several attempts, we have been unable to determine which of the A residues in the GAA domain of *Eco*R124 or *Eco*R124/3 is methylated in modified DNA. In the case of *Eco*R124 we favour the second A residue, because of the fact that when the sequence overlaps with an *Eco*RI site, *Eco*R124-modified DNA is resistant to *Eco*RI cleavage and the second A in the sequence is the one methylated by the *Eco*RI methylase (Dugaigczyk *et al.*, 1974). However, it has not been formally demonstrated that methylation of the other A would not also block *Eco*RI cleavage. These results also prove that, as with other type I systems, it is adenosyl residues that are methylated in the modified recognition sites.

Discussion

The first conclusion to be drawn is that the structure of the recognition sequences of both *Eco*R124 and *Eco*R124/3 is typical of that of other type I R-M systems in that they consist of two specific domains separated by non-specific spacers. This characteristic structure has now been found for 10 different type I R-M systems. These results also lend weight to the argument that the length of the non-specific spacer is such that the two methylatable positions are separated by one helical turn (Nagaraja *et al.*, 1985b). It was also suggested that type I enzyme methylation involves interactions of the enzyme along one face of the helix making specific contacts in two successive major groups with the non-specific spacer sequence tucked away in the intervening minor groove. Again, the *Eco*R124 and *Eco*R124/3 recognition sequences are compatible with this model.

The point of major interest in this paper is the relationship between the *Eco*R124 and the *Eco*R124/3 recognition sequences. They have identical specific domains and differ only in the length of the non-specific spacer sequence separating two specific domains, 6 bp for *Eco*R124 and 7 bp for *Eco*R124/3. As a consequence of this, the specific domains in the *Eco*R124/3 sequence are 3.4 Å further apart and rotated 36° with respect to those of *Eco*R124.

For the type IA family of restriction enzymes, the *hsdS* genes of different systems are largely non-homologous with the exception of a short region towards the middle of the gene and another region at the carboxyl terminus (Gough and Murray, 1983). In one case, recombination within the central region of homology resulted in the generation of new sequence specificity (Fuller-Pace *et al.*, 1984), the new sequence recognized by the recombinant enzyme being a hybrid with one specific domain from one parent and the other from the second (Nagaraja *et al.*, 1985a,b). One exceptional pair of *hsdS* genes was found that showed considerable homology towards the amino terminus of the genes (Fuller-Pace and Murray, 1986). This pair of enzymes, *Eco*K and *Sty*SP, both have AAC as the sequence of one of the specific domains of their recognition sequence. The conclusion from these studies is that the amino terminus of the proteins recognize the

Table I. Plasmids used in this work

Plasmid	Source	Reference
pSC101	A.Bernardi	Cohen and Chang (1977); Bernardi and Bernardi (1984)
pUC4K	Pharmacia	Vieira and Messing (1982)
pBR325::IS*5*	C.Nager	Personal communication
pBR325::IS*1*	C.Nager	Personal communication
pUC8	D.Pridmore	Vieira and Messing (1982)
pBR322	Laboratory collection	Bolivar *et al.* (1977); Sutcliffe (1979)
pIL254	A.Solonin	Personal communication
R124	Laboratory collection	Hedges and Datta (1972)
R124/3	Laboratory collection	Firman *et al.* (1983)
pCPB1	This work	
pCPB2	This work	

trinucleotide domain of the recognition sequences while the other domain is recognized by the carboxyl portion of the proteins. If the specific domains of the recognition sequences are different from each other, the genes are non-homologous in this region, if they are the same, they show homology (Fuller-Pace and Murray, 1986).

The *Eco*R124 and *Eco*R124/3 systems are not in the same family as those described above and are thought to have arisen as the result of a recent gene duplication (Firman *et al.*, 1983). Nevertheless, assuming that the functional organization of the *hsdS* genes is the same, the high degree of homology between the genes (Price, 1984) predicts that the two specific domains of the recognition sequence ought to be the same. It is difficult to predict the nature of the mutational event that would lead to the change in specificity that we see. This question will be answered when the DNA sequences of the two *hsdS* genes have been determined.

Finally, we believe that this report represents the first evidence for an alteration of DNA specificity of this kind that arose by spontaneous mutation, thus extending the possibilities to be considered when thinking about the evolution of DNA binding specificities. Other changes in the specificity of DNA binding proteins have been reported recently (e.g. Youderian *et al.*, 1983; Ebright *et al.*, 1984) but these have all resulted from strong genetic selection applied in the laboratory. The type I R-M systems still provide the only documented examples of changes in specificity arising through the natural processes of spontaneous mutation (this communication) and homologous recombination (Fuller-Pace *et al.*, 1984).

Materials and methods

Biological materials

Table I lists the plasmids used in the course of this work. The *E. coli* K12 strains AB2463 (R124) and AB2463 (R124/3) were used for *in vivo* methylation (Firman *et al.*, 1983). GM99 (Marinus and Morris, 1973) was the *dam*$^-$ host used to prepare unmethylated plasmid DNA. Phage λ*c*I857*s*7 DNA was prepared from a lysogen of AB2463 (Price, 1984). Plasmid DNA was purified from cleared lysate by CsCl–ethidium bromide equilibrium centrifugation (Clewell and Helinski, 1969). The purification of the *Eco*R124 and *Eco*R124/3 restriction enzymes will be described elsewhere.

Nucleic acid procedures

Type II restriction enzymes and DNA ligase were from New England Biolabs. DNA fragments were separated on 4% polyacrylamide gels which were stained with ethidium bromide and photographed under u.v. light illumination. Tritium-labelled fragments were detected by fluorography according to the method of Chamberlain (1979).

In vitro DNA methylation with *Eco*R124 and *Eco*R124/3 was carried out exactly as described by Nagaraja *et al.* (1985b). Most of the nucleotide sequences used in this study were taken from the EMBL Nucleotide Sequence Data Bank.

Acknowledgements

We would like to thank A.Bernardi and N.Murray for communicating results prior to publication and B.Suri and V.Nagaraja for their advice and interest in this work. We would also like to thank A.Bernardi, D.Pridmore, C.Nager and A.Solonin for gifts of materials used in this work, which was supported by grants from the Swiss National Science Foundation. C.P. was the recipient of a Royal Society European Fellowship.

References

Bannister,D. and Glover,S.W. (1968) *Biochem. Biophys. Res. Commun.*, **30**, 735–738.
Bernardi,A. and Bernardi,F. (1984) *Nucleic Acids Res.*, **12**, 9415–9426.
Bickle,T.A. (1982) In Linn,S.M. and Roberts,R.J. (eds), *Nucleases*. Cold Spring Harbor Laboratory Press, New York, pp. 85–108.
Bickle,T.A. (1987) In Ingraham,J., Low,K.B., Magasanik,B., Neidhardt,F.C., Schaechter,M. and Umbarger,H.E. (eds), *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*. American Society for Microbiology, Washington, D.C., in press.
Bolivar,R., Rodriguez,R.L., Greene,P.J., Betlach,M.L., Heyneker,A.L., Boyer, H.W., Crosa,J.H. and Falkow,S. (1977) *Gene*, **2**, 95–113.
Bullas,L.R., Colson,C. and Van Pel,A. (1976) *J. Gen. Microbiol.*, **95**, 166–172.
Chamberlain,J.P. (1979) *Analyt. Biochem.*, **98**, 132–135.
Clewell,D.B. and Helinski,D.R. (1969) *Proc. Natl. Acad. Sci. USA*, **62**, 1159–1166.
Cohen,S.N. and Chang,A.C.Y. (1977) *J. Bacteriol.*, **132**, 724–734.
Dugaigczyk,A., Hedgpeth,J.M., Boyer,H.W. and Goodman,H.M. (1974) *Biochemistry*, **13**, 503–512.
Dunn,J.J. and Studier,F.W. (1983) *J. Mol. Biol.*, **166**, 477–535.
Ebright,R.H., Cossart,P., Gicquel-Sanzey,B. and Beckwith,J. (1984) *Nature*, **311**, 232–235.
Firman,K., Creasey,W.A., Watson,G., Price,C. and Glover,S.W. (1983) *Mol. Gen. Genet.*, **191**, 145–153.
Fuller-Pace,F.V. and Murray,N.E. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 9368–9372.
Fuller-Pace,F.V., Bullas,L.R., Delius,H. and Murray,N.E. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 6095–6099.
Fuller-Pace,F.V., Cowan,G.M. and Murray,N.E. (1985) *J. Mol. Biol.*, **186**, 65–75.
Glover,S.W., Firman,K., Watson,G., Price,C. and Donaldson,S. (1983) *Mol. Gen. Genet.*, **190**, 65–69.
Glover,S.W. and Colson,C. (1969) *Genet. Res.*, **13**, 227–240.
Gough,J.A. and Murray,N.E. (1983) *J. Mol. Biol.*, **166**, 1–19.
Hedges,R.W. and Datta,N. (1972) *J. Gen. Microbiol.*, **71**, 403–405.
Hubacek,J. and Glover,S.W. (1970) *J. Mol. Biol.*, **5**, 111–127.
Kan,K.C., Lautenberger,J.A., Edgell,M.H. and Hutchison,C.A., III (1979) *J. Mol. Biol.*, **130**, 191–210.
Kröger,M. and Hobom,G. (1984) *Nucleic Acids Res.*, **12**, 887–899.
Lautenberger,J.A., Kan,N.C., Lackey,D., Linn,S., Edgell,M.H. and Hutchison, C.A., III (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 2271–2275.
Marinus,M.G. and Morris,N.R. (1973) *J. Bacteriol.*, **114**, 1143–1150.
Murray,N.E., Gough,J.A., Suri,B. and Bickle,T.A. (1982) *EMBO J.*, **1**, 535–539.
Nagaraja,V., Shepherd,J.C.W., Pripfl,T. and Bickle,T.A. (1985a) *J. Mol. Biol.*, **182**, 579–587.
Nagaraja,V., Shepherd,J.C.W. and Bickle,T.A. (1985b) *Nature*, **316**, 371–372.
Nagaraja,V., Stieger,M., Nager,C., Hadi,S.M. and Bickle,T.A. (1985c) *Nucleic Acids Res.*, **13**, 389–399.
Ohtsubo,H. and Ohtsubo,E. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 615–619.
Oka,A., Sugisaki,H. and Takanami,M. (1981) *J. Mol. Biol.*, **247**, 217–226.
Piekarowicz,A. and Goguen,J.D. (1985) *Eur. J. Biochem.*, **154**, 295–298.
Price,C. (1984) *Ph.D. Thesis*, University of Newcastle-upon-Tyne.
Ravetch,J.V., Horiuchi,K. and Zinder,N.D. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 2266–2270.
Sanger,F., Coulson,A.R., Hong,G.F., Hill,D.F. and Petersen,G.B. (1982) *J. Mol. Biol.*, **162**, 729–773.
Schoner,B. and Kahn,M. (1981) *Gene*, **14**, 165–174.
Sommer,R. and Schaller,H. (1979) *Mol. Gen. Genet.*, **168**, 331–335.
Suri,B. and Bickle,T.A. (1985) *J. Mol. Biol.*, **186**, 77–85.
Suri,B., Shepherd,J.C.W. and Bickle,T.A. (1984) *EMBO J.*, **3**, 575–579.
Sutcliffe,J.G. (1979) *Cold Spring Harbor Symp. Quant. Biol.*, **43**, 77–90.
Vieira,J. and Messing,J. (1982) *Gene*, **19**, 259–268.
Youderian,P., Vershon,A., Bouvier,S., Sauer,R.T. and Susskind,M.M. (1983) *Cell*, **35**, 777–783.
Yuan,R. (1981) *Annu. Rev. Biochem.*, **50**, 285–315.

Received on February 2, 1987; revised on March 2, 1987