

Recombination of constant and variable modules alters DNA sequence recognition by type IC restriction – modification enzymes

Marcel Gubler², Diego Braguglia³,
Jürg Meyer⁴, Andrzej Piekarowicz¹ and
Thomas A. Bickle

Department of Microbiology, Biozentrum, Basel University, CH-4056 Basel, Switzerland and ¹Institute of Microbiology, Warsaw University, 00-046 Warsaw, Poland

²Present address: Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

³Present address: Institut Suisse de Recherches Experimentales sur le Cancer, CH-1066 Epalinges s/Lausanne, Switzerland

⁴Present address: Zahnärztliches Institut, CH-4056 Basel, Switzerland

Communicated by T.A. Bickle

***EcoR124* and *EcoDXXI* are allelic type I restriction–modification (R–M) systems whose specificity genes consist of common structural elements: two variable regions are separated by a constant, homologous region containing a number of repetitive sequence elements. *In vitro* recombination of variable and constant elements has led to fully active, hybrid R–M systems exhibiting new and predictable target site specificities. Methylation of synthetic DNA sequences with purified, hybrid modification methylases was used to confirm the proposed recognition sequences. The results clearly demonstrate the correlation between protein domains and target site specificity. Our data suggest that a bacterial population may switch the recognition sequences of its type I R–M system by single recombination events and thus is able to maintain a prokaryotic analogue of the immune system of variable specificity.**

Key words: DNA restriction and modification/DNA sequence restriction/evolution of sequence specificity/type IC restriction enzymes

Introduction

The physical interaction between bacterial restriction–modification (R–M) systems and DNA is basically threefold: R–M systems recognize specific sequences on DNA; they cleave DNA which contains non-protected target sequences; and they prevent cleavage of DNA by methylation of adenosine or cytosine residues within target sequences. R–M systems have been classified as type I, II and III based on subunit composition, cofactor requirements and the enzymatic reactions that they catalyse (Bickle, 1987). Most of the R–M systems known to date are type II. They are extensively used in recombinant DNA technology primarily because restriction and modification reactions are carried out by two different enzymes and cleavage of DNA occurs precisely within, or close to, the sequences that they recognize. In addition, type II restriction enzymes have very simple reaction conditions: they require no cofactor other than Mg^{2+} to cleave DNA. Type III R–M systems are

somewhat more complex in that the restriction enzyme is composed of two different subunits, requires ATP and Mg^{2+} to cleave DNA and, in the presence of *S*-adenosyl methionine (AdoMet), can also act as a modification methylase (Hadi *et al.*, 1983). Type I R–M systems consist of three different subunits, they require Mg^{2+} , ATP and AdoMet to digest DNA, and they exhibit high ATPase activity upon incubation with non-modified substrate DNA (see Bickle, 1982 for a review). The process leading to cleavage of DNA by type I restriction enzymes at seemingly random sites up to several thousand base pairs (bp) distant from their recognition sequence has been explained by ATP-stimulated translocation of the enzymes along the DNA (Studier and Bandyopadhyay, 1988).

All the genetic loci coding for type I R–M systems identified so far are organized in two transcriptional units, one containing the genes *hsdM* and *hsdS*, the other containing *hsdR* only (Sain and Murray, 1980; Suri and Bickle, 1985; Price *et al.*, 1989). Genetic analysis showed that all three gene products are required for restriction, whereas *hsdM* and *hsdS* gene products are sufficient for modification methylation. The specificity of type I R–M systems is determined by the *hsdS* gene: mutations in *hsdS* abolish restriction and methylation activity while complementation with an allelic *hsdS* gene restores both activities and confers the specificity of the complementing allele (Boyer and Roulland-Dussoix, 1969; Hubacek and Glover, 1970; Fuller-Pace *et al.*, 1985; Skrzypek and Piekarowicz, 1989). In fact, detection of genetic complementation, and later of DNA homologies and antigenic cross-reactivity (Murray *et al.*, 1982) led to the recognition of three distinct families of type I R–M systems: A, B and C (see Bickle, 1987 for a review). In this report, we focus on members of the C family which are encoded by large conjugative plasmids: *EcoR124*, *EcoR124/3* (Firman *et al.*, 1985) and *EcoDXXI* (Skrzypek and Piekarowicz, 1989). However, localization on plasmids seems not to be a necessary criterion for type IC R–M systems. The recent finding of extensive homology between chromosomal DNA of the *Escherichia coli* *prf* locus and the *EcoR124/3 hsd* genes (Linder *et al.*, 1990) indicates a more widespread occurrence of type IC R–M systems than originally anticipated.

All type I R–M systems whose target sites have been determined recognize bipartite sequences consisting of two half-sites, each of three to five nucleotides, separated by a non-specific spacer region of six to eight nucleotides. Recombination of the *hsdS* genes of the *Salmonella* type I R–M systems *StySBI* and *StySPI* gave rise to the hybrid systems *StySQI* and *StySJI* which recognize hybrid target sites that are combinations of the original SB and SP half-sites (Bullas *et al.*, 1976; Fuller-Pace *et al.*, 1985; Nagaraja *et al.*, 1985; Gann *et al.*, 1987). It was concluded that in the wild-type and the recombinant HsdS polypeptides, amino-terminal domains specify the 5' half and carboxy-terminal domains specify the 3' half of bipartite recognition