

Generation of new DNA binding specificity by truncation of the type IC *EcoDXXI hsdS* gene

Maria P. MacWilliams and Thomas A. Bickle¹

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

¹Corresponding author

The *hsdS* subunit of a type IC restriction-modification enzyme is responsible for the enzyme's DNA binding specificity. Type I recognition sites are characterized by two defined half-sites separated by a non-specific spacer of defined length. The *hsdS* subunit contains two independent DNA binding domains, each targeted towards one DNA half-site. We have shown previously that the 5' half of *hsdS* can code for a functional substitute of the full-length *hsdS*. Here we demonstrate that the 3' half of the gene, when fused to the appropriate transcriptional and translational start signals, also codes for a peptide which imparts DNA binding specificity to the enzyme. About half the natural *hsdS* size, the mutant peptide contains a single DNA recognition domain flanked by one copy of each internal repeat found in the full-length *hsdS*. Deletion of either repeat sequence results in loss of activity. Like the 5' *hsdS* mutant, the 3' mutant recognizes an interrupted palindrome, GAAYN₂RTTC, suggesting that two truncated subunits participate in DNA recognition. Co-expression of the 5' *hsdS* mutant and the 3' *hsdS* mutant along with *hsdM* regenerates the wild-type methylation specificity. Thus, there is a free assortment of subunits in the cell.

Keywords: DNA modification/DNA recognition/DNA restriction/gene duplication/type I restriction systems

Introduction

Classical DNA restriction and modification (R-M) systems can be broken down into three groups based on subunit composition and co-factor requirements. Type I R-M systems have been found in several species of the eubacterial kingdom, including Gram-negative enterics such as *Escherichia coli*, *Salmonella typhimurium*, *Citrobacter freundii* and *Klebsiella pneumoniae* (Daniel *et al.*, 1988; Barcus *et al.*, 1995; Valinluck *et al.*, 1995) as well as in such diverse species as *Mycoplasma pneumoniae* and *Bacillus subtilis* (Dybvig and Yu, 1994; Xu *et al.*, 1995). The better characterized type II systems are more widespread, being found throughout the Eu- and Archaeobacterial kingdoms. Type III systems, like the type Is, are, so far, more restricted in their distribution, with only four characterized systems to date (for reviews, see Wilson and Murray, 1991; Bickle and Krüger, 1993).

The two main (competing) activities of all R-M systems are methylation of the bacterial DNA at a specific sequence

and the endonucleic cleavage of DNA which is not methylated at that sequence. Type II R-M systems accomplish these tasks with two independent enzymes, a DNA methyltransferase and a restriction endonuclease. Type I systems differ from their type II cousins in that a single, multisubunit enzyme accomplishes both DNA modification and restriction. This enzyme is composed of three subunits, *hsdM*, *hsdS* and *hsdR*. A complex of *hsdM* and *hsdS* alone can catalyse methylation, but all three subunits are required for restriction. The type I group can be subdivided further into three families, IA, IB and IC, based on sequence similarities, antigenic cross-reactivity and genetic complementation (Bickle, 1987).

The mode of DNA recognition by the type I enzymes is interesting in that they generally recognize non-palindromic sequences composed of two defined half-sites, 3-5 bp in length, separated by a non-specific spacer of defined length. For example, the *EcoDXXI* system, the subject of this study, recognizes TCAN₇RTTC (where N = any base and R = G or A). The spacer arranges the half-sites on the DNA such that there is approximately one helix turn length between the two adenines to be methylated.

A single peptide, *hsdS*, is responsible for recognition of both DNA half-sites as well as for spacer length determination. In addition, *hsdS* must also form inter-subunit contacts with the other peptides of the enzyme complex. It has been shown previously that *hsdS* is composed of two independent DNA binding domains (Fuller-Pace *et al.*, 1984; Nagaraja *et al.*, 1985; Fuller-Pace and Murray, 1986; Cowan, 1989; Gubler *et al.*, 1992). The amino-proximal domain (ARD; Chen, *et al.*, 1995) recognizes the 5' half-site, and the carboxy-proximal domain (CRD) the 3' half site. These variable regions which constitute the DNA binding domains are separated by amino acid segments that are conserved between members of a particular type I family (Figure 1; reviewed in Kneale, 1994). Since the *hsdS* subunits within a family are interchangeable, the conserved regions are thought to mediate interactions with the other subunits of the enzyme complex (Fuller-Pace and Murray, 1986; Abadjieva *et al.*, 1994; Cooper and Dryden, 1994). Sequence comparisons amongst the IC enzymes reveal three conserved regions, one at either end of the peptide and a third, centrally located region (Meister *et al.*, 1993; Tyndall *et al.*, 1994). Within the central conserved region, there is a tandem repeat of a tetraamino acid sequence, TAEL. It has been shown that the number of repeats governs the length of the recognition site spacer: two repeats result in a 6 bp spacer while three repeats result in a 7 bp spacer (Price *et al.*, 1989; Gubler and Bickle, 1991; Gubler *et al.*, 1992).

When the amino acid sequence of a type IC *hsdS* subunit is compared with itself, another level of sequence conservation is revealed (Meister *et al.*, 1993; Abadjieva