

Macroevolution by transposition: drastic modification of DNA recognition by a type I restriction enzyme following Tn5 transposition

Jennifer Meister, Maria MacWilliams, Philipp Hübner, Helga Jütte, Elzbieta Skrzypek^{1,2}, Andrzej Piekarowicz¹ and Thomas A. Bickle³

Department of Microbiology, Biozentrum, Basel University, Klingelbergstrasse 70, CH-4056 Basel, Switzerland and ¹Institute of Microbiology, Warsaw University, Nowy Swiat 67, PL-00-046 Warsaw, Poland

²Present address: Department of Microbiology and Immunology, University of Kentucky College of Medicine, Lexington, KY 4036-0084, USA

³Corresponding author

Communicated by T.A. Bickle

We have characterized a novel mutant of *EcoDXXI*, a type IC DNA restriction and modification (R-M) system, in which the specificity has been altered due to a Tn5 insertion into the middle of *hsdS*, the gene which encodes the polypeptide that confers DNA sequence specificity to both the restriction and the modification reactions. Like other type I enzymes, the wild type *EcoDXXI* recognizes a sequence composed of two asymmetrical half sites separated by a spacer region: TCA(N₇)RTTC. Purification of the *EcoDXXI* mutant methylase and subsequent *in vitro* DNA methylation assays identified the mutant recognition sequence as an interrupted palindrome, TCA(N₈)TGA, in which the 5' half site of the wild type site is repeated in inverse orientation. The additional base pair in the non-specific spacer of the mutant recognition sequence maintains the proper spacing between the two methylatable adenine groups. Sequencing of both the wild type and mutant *EcoDXXI hsdS* genes showed that the Tn5 insertion occurred at nucleotide 673 of the 1221 bp gene. This effectively deletes the entire carboxyl-terminal DNA binding domain which recognizes the 3' half of the *EcoDXXI* binding site. The truncated *hsdS* gene still encodes both the amino-terminal DNA binding domain and the conserved repeated sequence that defines the length of the recognition site spacer region. We propose that the *EcoDXXI* mutant methylase utilizes two truncated *hsdS* subunits to recognize its binding site. The implications of this finding in terms of subunit interactions and the malleability of the type I R-M systems will be discussed.

Key words: DNA methylation/evolution/restriction enzymes/sequence specificity/transposition mutagenesis/type I restriction

DNA at a specifically recognized DNA sequence (this distinctive pattern of methylation defines 'self' DNA) and double-stranded cleavage of 'foreign' DNA that is not methylated at that sequence. There are presently four distinct types or classes of R-M systems known based on the subunit composition of the enzymes and cofactor requirements. With the type I R-M systems, both the DNA restriction and modification functions are carried out by a single enzyme composed of three heterologous subunits: *hsdS* (DNA binding specificity), *hsdM* (modification/methylation) and *hsdR* (restriction). A complex of only *hsdS* and *hsdM* can catalyse methylation but not restriction. The type I enzymes are further divided into three families, A, B and C, based on genetic complementation, sequence homology and antigenic cross-reactivity amongst family members. In fact, the individual subunits are interchangeable within each family group (reviewed in Wilson and Murray, 1991; Bickle and Krüger, 1993).

All type I methylase activities characterized so far are adenine-specific and the sequenced type I *hsdM* genes share some homology with other adenine methylases (Sharp *et al.*, 1992; Wilson, 1992). *S*-adenosyl methionine (AdoMet) is required for the methylation reaction. In contrast to the more familiar type II restriction endonucleases, which cleave at defined positions within or close to their recognition sites, DNA restriction by type I enzymes occurs at random sites a great distance away from the recognition site (reviewed in Bickle, 1982). Mg²⁺ ion, AdoMet and ATP are required for type I restriction activity. Type I restriction of unmodified DNA is accompanied by large amounts of ATP hydrolysis which continues long after cleavage. It is thought that the ATP hydrolysis fuels the 'pumping' of DNA past the bound enzyme to reach the cleavage site (Yuan *et al.*, 1980; Endlich and Linn, 1985; Studier and Bandyopadhyay, 1988).

The *hsdS* subunit is responsible for binding to the enzyme's DNA recognition site. A typical type I recognition site is composed of two asymmetric half sites separated by 6-8 bp of nonspecific spacer DNA. The spacer arranges the half sites such that the adenine groups to be methylated are 10 or 11 bp apart on the same face of the DNA helix. It has previously been shown that the two halves of the DNA binding site are recognized by two discrete domains in the *hsdS* protein, an amino-terminal domain for the 5' and a carboxyl-terminal domain for the 3' sequence (Fuller-Pace *et al.*, 1984; Nagaraja *et al.*, 1985; Fuller-Pace and Murray, 1986; Cowan *et al.*, 1989). A third region of the *hsdS* protein has been assigned a specific role in DNA recognition for the type IC family of proteins: comparison of two type IC systems, *EcoR124I* and *EcoR124/3I*, whose recognition sequences are identical except for the length of the unspecified spacer DNA, showed that the spacer length of the DNA binding site is correlated with the number of repeats of a conserved tetraamino acid sequence (TAEL) present in the middle of *hsdS*. Located between the 5' and 3' DNA binding domains, the TAEL repeats appear to provide the

Introduction

Restriction-modification (R-M) systems are found throughout the bacterial world and serve as a means to protect the host bacterium from foreign DNA. The two main functions of all R-M systems are methylation of the host