

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

proper distance between the two DNA binding domains, and thus determine the distance between the two target half sites (Price et al., 1989; Gubler and Bickle, 1991; Gubler et al., 1992).

In this paper we focus on a novel type IC R-M system mutant resulting from a Tn5 insertion in the middle of the *EcoDXXI hsdS* gene. Surprisingly, this mutant still encodes a functional methylase and restriction enzyme albeit with a different specificity than the wild type parent. This mutant is a rare example of a transposition event which results in a new, functional enzyme rather than just abolishing activity and demonstrates that transposition can indeed play a role in the evolution of a bacterium.

Results

Location of the Tn5 element in plasmid pES14

While mapping the genes encoding the *EcoDXXI* enzyme encoded on plasmid pES14, a Tn5 insertion into the *hsdS* gene was recovered (Skrzypek and Piekarowicz, 1989). Unexpectedly, instead of disrupting enzymatic function, the mutant was fully active in both restriction and modification but appeared to have a different DNA specificity from the wild type. DNA sequencing of the wild type and mutant *hsdS* genes showed that the Tn5 insert was located such that it divided the *hsdS* gene into a 5' portion of 673 bp (*hsdS* 5')

and a 3' portion of 545 bp (*hsdS* 3') (Figure 1). The wild type gene codes for a protein of 406 amino acids; the mutant protein is predicted to contain 233 amino acids with the nine carboxyl-terminal ones being encoded by Tn5 DNA. The sequence of the wild type gene has been deposited with the EMBL DNA sequence database under accession number X73984. Except for the Tn5 disruption and a 9 bp duplication of the flanking sequence normally seen upon Tn5 insertion, no other sequence differences were found between the wild type and mutant *hsdS* genes.

At 91 bp from the 3' end of the *hsdS* gene the sequence was found to read, GAAGGTCTTC, which differs from the published *hsdS* sequence of the related *EcoR124/3I* enzyme, which has GAAGTCTTC at the same position (Price et al., 1989). As the additional G is located in one of three regions of high DNA homology between the *EcoR124/3I* and *EcoDXXI hsdS* genes, we reinvestigated the *EcoR124/3I* sequence in this region. Fortunately, the discrepancy coincided with an *XmnI* restriction site and so it was easy to confirm the existence of this site in the *EcoR124/3I hsdS* gene. After correction of the *EcoR124/3I* DNA sequence, the 35 amino acids at the carboxyl-terminus of these two *hsdS* proteins are identical.

Properties of the mutant enzyme

To demonstrate that the enzyme encoded by pES14::Tn5 methylates DNA with a specificity different from that of

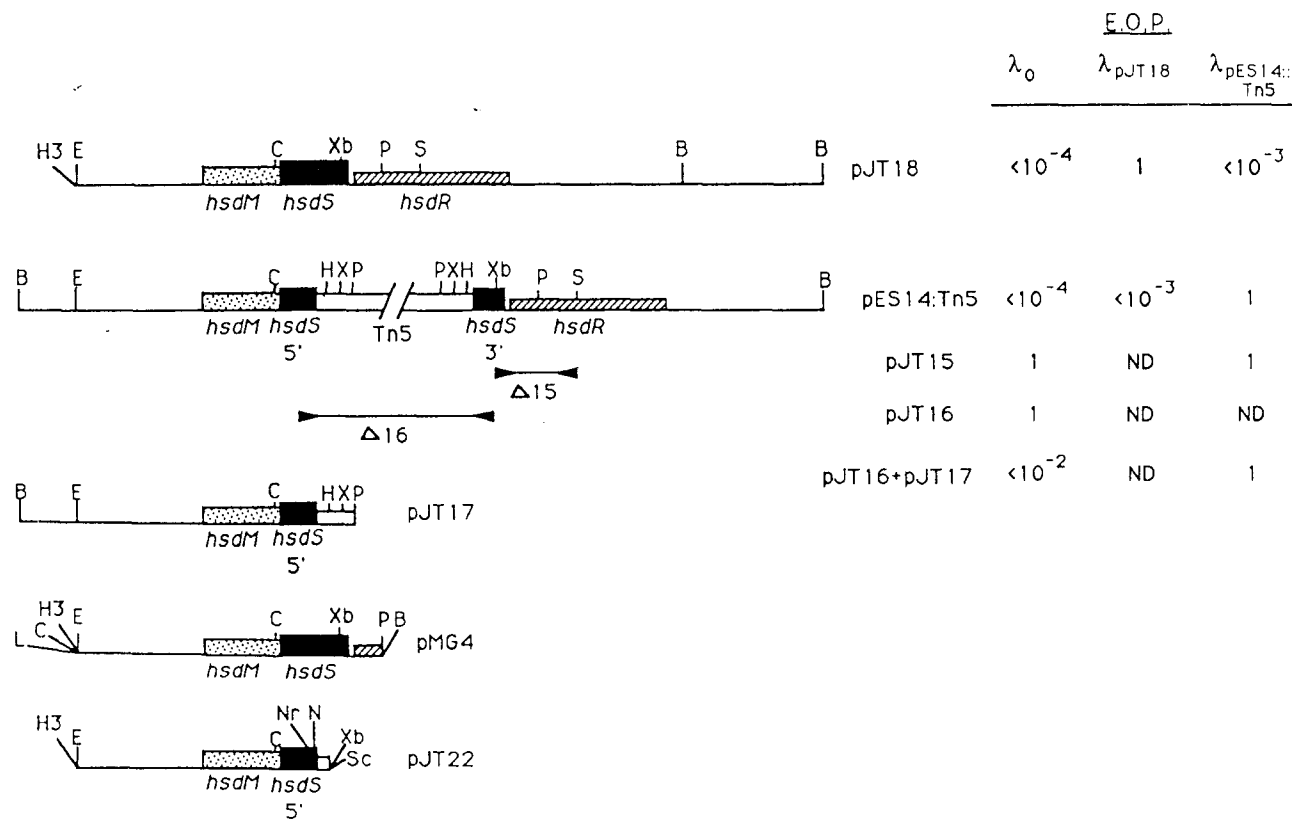


Fig. 1. Plasmid constructs and e.o.p. values. Plasmids pJT18 and pMG4 code for the wild type *EcoDXXI* restriction enzyme and methylase, respectively. The *EcoDXXI* locus in pJT18 is identical to that of pES14. pES14::Tn5 codes for the *EcoDXXI* mutant R-M system. The mutant phenotype is a result of the Tn5 insertion within the *hsdS* gene. Beneath pES14::Tn5 are indicated the regions deleted in plasmids pJT15 and pJT16 (see text for further description). Also indicated are the plating efficiencies (E.O.P.) of bacteriophage λ lysates on HB101 cells containing the plasmids diagrammed on the left. The subscripts refer to the modification that the phage had prior to infection (ND = not determined). Plasmids pJT15, 17 and 22 all express mutant methylase activity but lack restriction activity. Gene sequences are indicated by boxes shaded as follows: stippled, *hsdM*; black, *hsdS*; white, Tn5; striped, *hsdR*. The pertinent restriction sites are abbreviated as follows: B, *Bam*HI; C, *Cla*I; E, *Eco*RI; H, *Hpa*I; H3, *Hind*III; L, *Sal*I; N, *Nco*I; Nr, *Nru*I; P, *Pst*I; S, *Sna*I; Sc, *Sac*I; X, *Xho*I; and Xb, *Xba*I.

EcoDXXI, 'cross-plaquing' assays were performed between cells harboring either a plasmid coding for the wild type *EcoDXXI* restriction enzyme (pJT18) or the mutant enzyme-coding plasmid (pES14:Tn5) (Figure 1). Cells carrying either pES14:Tn5 or pJT18 restrict unmodified λ vir with an efficiency of plating (e.o.p.; see Materials and methods for definition) of between 10^{-4} and 10^{-8} . λ vir modified with the wild type *EcoDXXI* specificity by passage through HB101/pMG4 cells (Figure 1) was protected against restriction when plated onto HB101/pJT18 cells (e.o.p. = 1), but was restricted by HB101/pES14:Tn5 cells (e.o.p. = 10^{-3}). Conversely, λ vir modified by passage through HB101/pES14:Tn5 was restricted by HB101/pJT18 (e.o.p. < 10^{-3}), but not by HB101/pES14:Tn5 cells (e.o.p. = 1). These results clearly show that pES14:Tn5 encodes an R-M enzyme with a specificity different from that of the wild type *EcoDXXI*.

To ascertain whether the observed DNA restriction activity of pES14:Tn5 was due to the *EcoDXXI* mutant locus or to another restriction system occurring elsewhere on the plasmid, the DNA deletion construct pJT15 was made (Figure 1). Plasmid pJT15 has the 5' half of the *EcoDXXI* *hsdR* gene deleted, presumably inactivating it. In addition, the last 85 bp of the *hsdS* 3' segment are also deleted. HB101/pJT15 cells plated unmodified λ vir with an e.o.p. of 1 indicating that the deletion in the *EcoDXXI* *hsdR* gene resulted in loss of restriction activity. Thus no other DNA restriction system was supplied by the plasmid. On the other hand, phage harvested from (and thus modified by) cells harboring pJT15 plated onto HB101/pES14:Tn5 cells with an e.o.p. of 1. Thus pJT15 still codes for a methylase with pES14:Tn5 specificity in spite of the carboxyl-terminal *hsdS* gene deletion.

Plasmid pJT16 was constructed to prove that the truncated *hsdS* subunit functions as the specificity subunit for the mutant enzyme. pJT16 has most of both *hsdS* 5' and *hsdS* 3' deleted while leaving *hsdM* and *hsdR* intact. Unmodified λ vir plated on HB101/pJT16 cells with an e.o.p. of 1, which shows that no other active *hsdS* gene is located on the plasmid.

To demonstrate that the mutant methylase activity relied solely on the truncated *hsdS* 5' subunit without any contribution from the 3'-terminal portion of the Tn5-divided *hsdS* gene, the plasmid pJT17 (Figure 1) was made. pJT17 carries DNA sequences derived from the extreme 5' *Bam*HI site of pES14:Tn5, including *hsdM*, *hsdS* 5' and 680 bp of Tn5 to the *Pst*I site in Tn5. The entire *hsdS* 3' fragment is missing. λ vir passaged through HB101/pJT17 cells plated on HB101/pES14:Tn5 cells with an e.o.p. of 1, indicating that *hsdS* 5' fragment confers the same specificity on the methylase as that encoded by pES14:Tn5. Thus, a truncated *hsdS* gene (*hsdS* 5') plus a portion of Tn5 is sufficient to confer specificity.

In a complementation experiment, HB101 cells harboring both pJT16 and pJT17 were challenged by λ vir. The resulting e.o.p. of 10^{-2} indicates that restriction activity was conferred by supplying the *hsdR* subunit *in trans* from pJT16 to the truncated *hsdS* subunit from pJT17 (both plasmids code for *hsdM*). Neither HB101/pJT17, which has no *hsdR* gene, nor HB101/pJT16, lacking a functional *hsdS* gene, individually can restrict λ . Thus the truncated *hsdS* subunit encoded by pJT17 can participate in the formation of a functional R-M enzyme.

The minimum *hsdS* sequences required for activity

Two approaches were taken to determine the minimum-sized *hsdS* peptide required for a functional methylase. To map out the region of interest, fortuitously located restriction sites were utilized to remove discrete portions of the truncated *hsdS* of pJT22. The resulting *hsdS* deletion variants were tested *in vivo* for methylase activity.

Removal of the *Nco*I–*Xba*I fragment of pJT22 (Figure 1) deleted the remaining Tn5 sequences as well as 10 bp from the 3' end of the truncated *hsdS* gene (Figure 2). This mutant, pHJ1, still encoded a functional methylase. Removal of an additional 94 bp in the *Nru*I–*Xba*I deletion resulted in the loss of methylase activity.

To narrow down further the minimum *hsdS* fragment required for function, a series of exonuclease III deletion mutants was constructed using the *Nco*I site as the starting point for digestion. The resulting mutant plasmids were screened for the size of the remaining *hsdS* region and for *in vivo* methylase activity. Five representative clones were sequenced. The results are shown in Figure 2. The shortest *hsdS* derivative (609 bp) isolated which still encoded a functional subunit is predicted to produce a peptide of 205 amino acids with the last two being contributed by 3' flanking sequence. The endpoint for the longest methylase negative clone was at bp 572 of the *hsdS* gene. Thus the endpoint of the area critical for *hsdS* function must lie between nucleotides 572 and 609.

Purification and properties of the mutant methylase

The high copy number plasmid, pJT22, was used to obtain a partially pure preparation of mutant methylase as described in Materials and methods. Wild type *EcoDXXI* methylase consists of the *hsdM* and *hsdS* subunits, whose gel-estimated molecular weights correspond to the predictions from the DNA sequence of 58 and 46 kDa (Figure 3, lanes 1 and 14). The band running slightly ahead of the *hsdM* subunit is most likely a proteolytic breakdown product. Fractions from a Sephacryl-200 column of a mutant enzyme preparation are also shown (lanes 2–9). Active fractions were identified by *in vitro* DNA methylation assays with the highest activity being found in lanes 3–5. Fractions represented by lanes 3 and 4 were pooled for *in vitro* methylation studies.

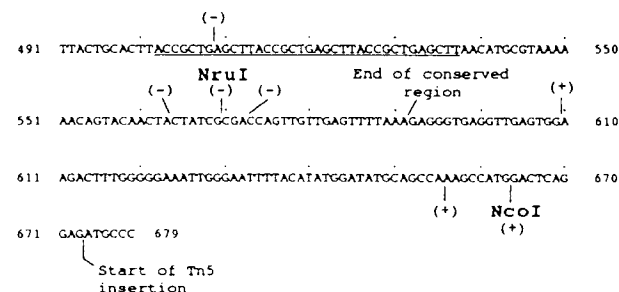


Fig. 2. Deletion analysis of the *EcoDXXI* mutant *hsdS* gene. To map the minimum *hsdS* fragment required for a functional methylase, deletion derivatives of the truncated *hsdS* gene in pJT22 were constructed. Illustrated are the deletion endpoints of several derivatives in the critical region along with their *in vivo* methylation phenotypes [(+), active methylase; (-), inactive]. The Tn5 sequence begins after nucleotide 673 of the *hsdS* gene. The underlined bases code for the TAEI repeats which govern the spacing between the recognition half sites (see Introduction). Also indicated is the 3' end of the internally conserved region which may play a role in subunit interactions (see Discussion).

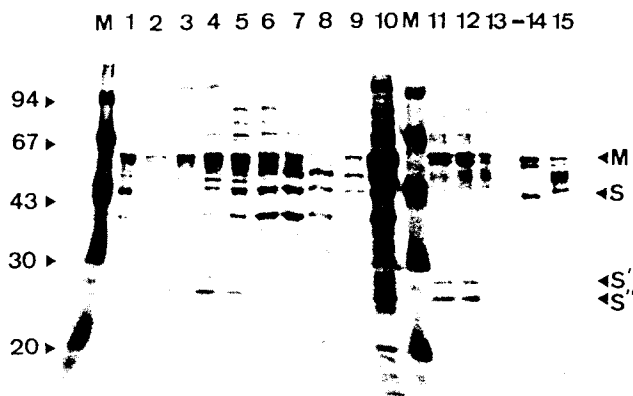


Fig. 3. Purification and immunoprecipitation of wild type and mutant methylases. The 12% SDS-polyacrylamide gel was stained with Coomassie Blue. Lanes 1 and 14, *EcoDXXI* methylase. Lane 15, immunoprecipitated *EcoDXXI* methylase. Lanes 2-9, fractions from Sephacryl-200 chromatography of the mutant enzyme. Lanes 11-13, immunoprecipitates of material in the fractions shown in lanes 3, 4 and 6, respectively. Lanes marked M contain molecular weight markers whose sizes are indicated to the left of the figure. M and S show the positions of the wild type hsdM and hsdS proteins, S' and S'' the positions of the presumed mutant truncated hsdS protein.

Immune serum raised against the *EcoR124I* enzyme cross-reacts with the hsdM and hsdR subunits of the *EcoDXXI* enzyme. Although the serum has no affinity for the hsdS polypeptides from either *EcoR124I* or *EcoDXXI*, it can complex with the holoenzyme so that the hsdS subunits of the methylases will be captured and coimmunoprecipitated. In an effort to identify the mutant methylase components, fractions from the protein purification were analyzed by immunoprecipitation with anti-*EcoR124I* antiserum. The precipitated proteins were solubilized and run on a 12% polyacrylamide gel (Figure 3). Lane 15 shows the immunoprecipitated products from the purified *EcoDXXI* methylase preparation, and lanes 11, 12 and 13 those from fractions represented in lanes 3, 4 and 6, respectively, of the mutant enzyme preparation. In lanes 11, 12 and 13 the hsdM subunit has the same mobility as the wild type in lanes 14 and 15. In contrast to the 46 kDa hsdS subunit found in the wild type preparation, peptides of 28 and 25 kDa, designated S' and S'', were precipitated from the active mutant methylase fractions (lanes 11 and 12). No S' or S'' proteins were immunoprecipitated from the low-activity fraction in lane 6 (lane 13), showing that the *in vitro* methylase activity correlates with the presence of S' and S''. The S' and S'' peptides were electroeluted and subjected to amino-terminal amino acid analysis. The 25 kDa protein has an amino-terminal stretch of 9 amino acids which correspond exactly to the amino-terminus of the wild type hsdS protein. The 28 kDa protein was blocked at the amino-terminus. From the DNA sequence, the truncated hsdS peptide is predicted to be 233 amino acids long which fits in well with a molecular weight of 25 kDa. The 28 kDa peptide may be a slightly longer derivative of the 25 kDa peptide, or the amino-terminal modification may account for the slower mobility.

The DNA sequence recognized by the mutant methylase

As a further demonstration that the wild type and mutant methylases recognize different DNA sequences, *in vitro*

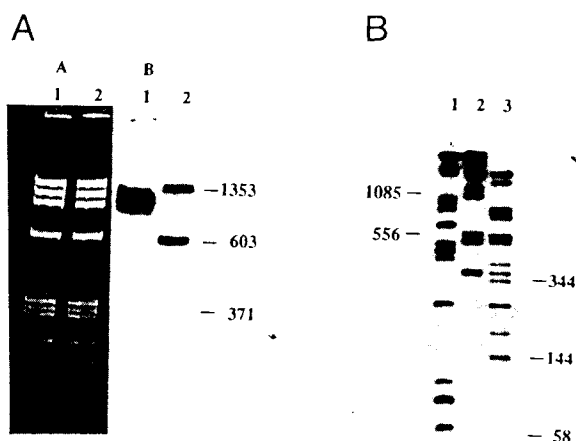


Fig. 4. *In vitro* methylation of DNA sequences by *EcoDXXI* wild type and mutant methylases. DNA aliquots were labelled with S-adenosyl [methyl-³H]methionine, isolated and cleaved with type II endonucleases as described in Materials and methods. (A) *HaeIII*-restricted ϕ X174 DNA which has been modified by either *EcoDXXI* wild type (lane 1) or mutant (lane 2) methylase. Panel A: photograph of ethidium bromide stained 5% polyacrylamide gel. Panel B: fluorograph of same gel. (B) *EcoDXXI* mutant methylase modification of λ DNA. After modification, DNA aliquots were digested with *HaeIII* (lane 1), *HincII* (lane 2) or *HincII* and *DdeI* (lane 3).

DNA methylation assays were performed. In these assays, S-adenosyl [methyl-³H]methionine is used as the methyl donor so that the DNA sites become radioactively labelled upon modification by the purified methylase. DNA was labelled by the wild type or mutant methylases, digested with a type II restriction endonuclease and the fragments separated on a polyacrylamide gel. Ethidium bromide staining of the gel reveals all the DNA fragments from the digest, while fluorography shows which of the fragments have been radioactively labelled by the enzyme. Figure 4A shows a photograph of the stained gel and the corresponding fluorogram from an experiment in which bacteriophage ϕ X174 DNA was labelled using either the *EcoDXXI* (lane 1) or the mutant methylase (lane 2). After labelling, the DNA was cut with *HaeIII*. Both lanes have an identical pattern of stained bands (panel A); however, the fluorogram (panel B) reveals that the methylases label different regions of the phage DNA. Since the wild type and mutant methylases do not methylate the same *HaeIII* fragments, it can be inferred that they have different sequence specificities.

In vitro methylation assays were used to identify the recognition sequence of the mutant enzyme using previously established methods (Suri *et al.*, 1984; Nagaraja *et al.*, 1985; Price *et al.*, 1987). Briefly, methylatable sites are mapped in sequenced DNA molecules and a computer is then used to sort out all sequences common to the labelled regions. Various tests using other sequenced DNA molecules are then made to distinguish between the various possibilities. In these experiments, bacteriophage λ DNA was the first substrate used. The methylation pattern of *HaeIII*-restricted λ DNA (Figure 4B, lane 1) shows at least 14 labelled bands and, in fact, 20 methylatable sites were eventually mapped in λ DNA using a variety of type II endonucleases to cleave the methylated DNA (data not shown). The analysis of the data from λ DNA was consistent with the mutant enzyme methylating the unique sequence, TCA(N₈)TGA.

Other DNA molecules were searched for occurrences of

