The EMBO Journal vol.15 no.6 pp.1443–1450, 1996

Identification of a subdomain within DNA-(cytosine-C5)-methyltransferases responsible for the recognition of the 5′ part of their DNA target

Christina Lange[1], Carola Wild and Thomas A.Trautner[2]

Max-Planck-Institut für molekulare Genetik, Ihnestrasse 73, D-14195 Berlin, Germany

[1]Present address: Istituto Richerche Immunobiologiche Siena, Via Fiorentina 1, I-53100 Siena, Italy

[2]Corresponding author

In previous work on DNA-(cytosine-C5)-methyltransferases (C5-MTases), domains had been identified which are responsible for the sequence specificity of the different enzymes (target-recognizing domains, TRDs). Here we have analyzed the DNA methylation patterns of two C5-MTases containing reciprocal chimeric TRDs, consisting of the N- and C-terminal parts derived from two different parental TRDs specifying the recognition of 5′-CC(A/T)GG-3′ and 5′-GCNGC-3′. Sequences recognized by these engineered MTases were non-symmetrical and degenerate, but contained at their 5′ part a consensus sequence which was very similar to the 5′ part of the target recognized by the parental TRD which contributed the N-terminal moiety of the chimeric TRD. The results are discussed in connection with the present understanding of the mechanism of DNA target recognition by C5-MTases. They demonstrate the possibility of designing C5-MTases with novel DNA methylation specificities.

Keywords: bisulfite modification of DNA/DNA target recognition/methylation specificity/protein–DNA interaction

Introduction

DNA-(cytosine-C5) methylation is found in organisms as diverse as bacteria, plants and mammals. C5-DNA methyltransferases (C5-MTases), which catalyze the transfer of a methyl group from the co-factor *S*-adenosyl-L-methionine to the C5 position of DNA cytosine, share a similar architecture, where an ordered set of well-conserved amino acid motifs alternates with non-conserved spans (Lauster *et al.*, 1989; Pósfai *et al.*, 1989; Figure 1A). By sequence comparisons (Behrens *et al.*, 1987; Lauster *et al.*, 1989) of the multispecific phage-encoded C5-MTases, several non-overlapping contiguous subdomains, located within the largest non-conserved or variable region ('V' in Figure 1A), could be identified. Genetic analyses showed that these subdomains are responsible for the different DNA recognition specificities of the enzymes [Balganesh *et al.*, 1987; Trautner *et al.*, 1988, 1996 (accompanying paper); Wilke *et al.*, 1988; Walter *et al.*, 1992], and they were termed target-recognizing domains (TRDs). By amino acid sequence comparisons (Lauster *et al.*, 1989) and by hybrid constructions (Klimasauskas *et al.*, 1991), some TRD-like motifs were also identified in the 'V' regions of the monospecific bacterial C5-MTases. Based on these studies, the concept of a common modular organization of all C5-MTases was developed: an enzyme 'core' of conserved amino acid motifs, which is responsible for the common catalytic functions of the methyl transfer reaction, is associated with different TRDs, located in the 'V' regions of the enzymes, that mediate specific DNA target recognition. This general concept was confirmed and refined significantly by the resolution of the atomic structures of two monospecific bacterial C5-MTases. While in the case of M.*Hha*I, crystals of the ternary co-factor–protein–DNA complex were analyzed (Cheng *et al.*, 1993; Klimasauskas *et al.*, 1994), the structure of M.*Hae*III, complexed with its DNA substrate site, had been analyzed in the absence of the co-factor (Reinisch *et al.*, 1995). In both structures, most of the invariant amino acid residues are located within the active sites of the enzymes, while in the protein–DNA interphase several amino acid residues located within the C-terminal part of their 'V' regions provide almost all the base-specific contacts to the recognized target sites. The residues involved in base-specific recognition are located within the putative TRDs of M.*Hha*I and M.*Hae*III, whose identity had been predicted by amino acid sequence comparisons (Lauster *et al.*, 1989). TRDs of the multispecific C5-MTases, specifying the recognition of different DNA targets, show only weak sequence similarity (Figure 1B). Point mutations causing the exchange of single amino acid residues within the TRDs affected their activity but never led to a change of specificity (Trautner *et al.*, 1988, 1996; Wilke *et al.*, 1988; Lange *et al.*, 1991). Hence, it was concluded that target recognition must be mediated by a network of several amino acid residues within the TRD. In this communication, we attempted to elucidate the contribution of different segments of a TRD to target recognition. The approach was to analyze the DNA recognition activity of two chimeric TRDs, in which the N- and C-terminal segments of two TRDs with different target specificities had been recombined. With these constructs, we addressed the following questions: (i) does the fusion of parts of two different TRDs result in a functional TRD which can endow the recipient C5-MTase with methylation activity? (ii) What would be the specificity of such methylation? (iii) Is it possible to identify subdomains of the TRDs involved in the recognition of separate parts of its DNA target sequence?

Results

Construction of C5-MTases carrying chimeric TRDs FE and EF

In earlier studies (Balganesh *et al.*, 1987; Figure 1C), reciprocal chimeras '3' and '4' between the multispecific

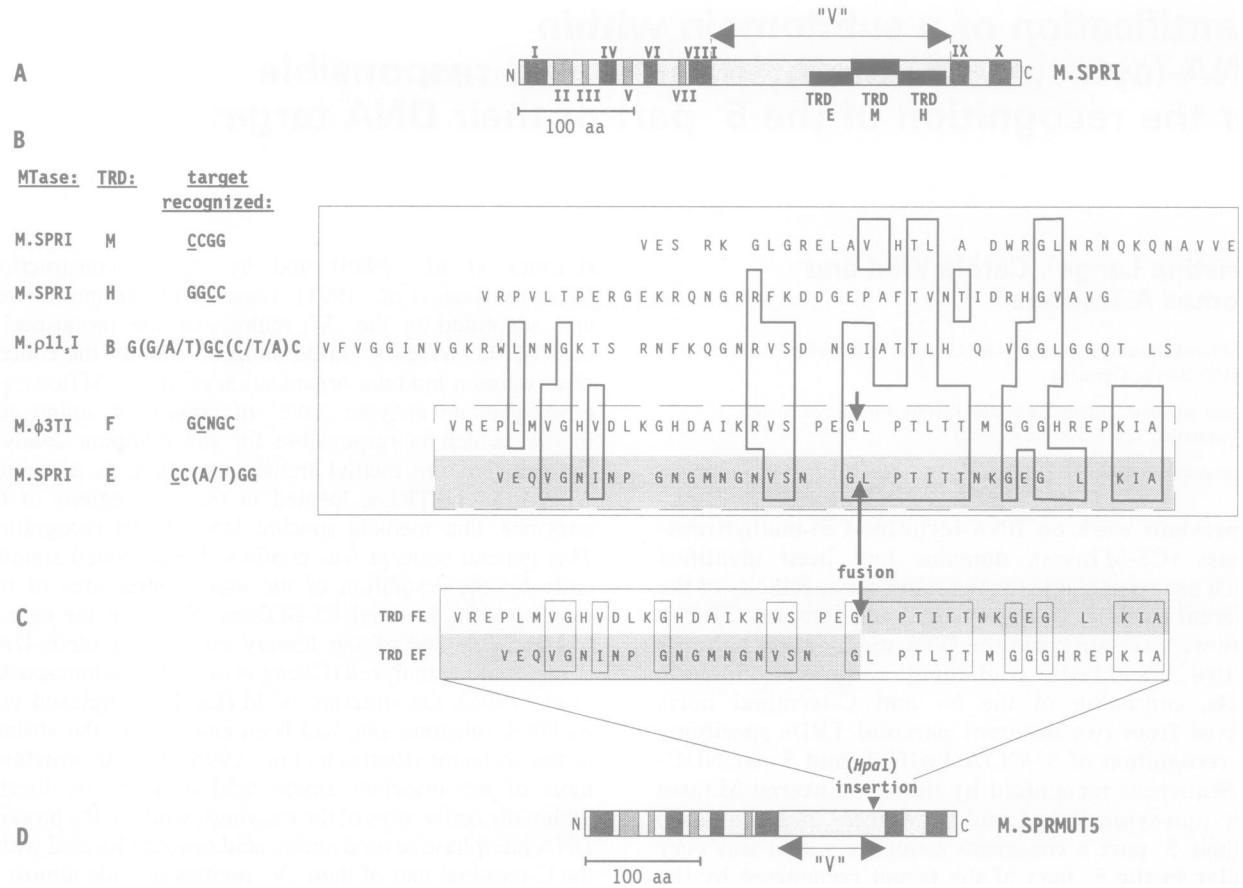

Fig. 1. Structure of multispecific C5-MTases and their target-recognizing domains (TRDs). (**A**) Arrangement of the six highly conserved sequence motifs I, IV, VI, VIII, IX and X and four less-well conserved motifs II, III, V and VII (gray boxes; Pósfai, *et al.*, 1989; Cheng *et al.*, 1993; Kumar *et al.*, 1994) within the multispecific MTase M.SPRI (Buhk *et al.*, 1984). In the C-terminal part of the largest variable region ('V'), the location of TRDs E, M and H, each responsible for recognizing a different target [as shown in (B)], is indicated by black rectangles (Trautner *et al.*, 1996). (**B**) Alignment of the amino acid sequences of TRDs with different specificities of various multispecific MTases. Methylated target Cs are underlined. Data obtained by bisulfite analysis of the methylation pattern of M.SPR-E, a derivative of the M.SPRI MTase carrying only TRD E (Walter *et al.*, 1992), revealed that the outer and not the inner C residue (as previously reported; Trautner and Noyer-Weidner, 1992) of the CC(A/T)GG target site is methylated. The sequences are aligned at the highly conserved T(L/I/V) motif that is found in most TRDs (Lauster *et al.*, 1989); similar amino acid residues are boxed. (**C**) Amino acid composition of the chimeric TRDs FE and EF, resulting from the fusion of parts of the M.ϕ3TI (Tran-Betcke *et al.*, 1986) and M.SPRI encoding genes at a conserved position [indicated by arrows at the corresponding amino acid positions in (B)], as described elsewhere (Balganesh *et al.*, 1987). (**D**) Schematic structure of M.SPRMUT5, a mutant derivative of M.SPRI lacking all TRDs (Walter *et al.*, 1992), into which chimeric TRDs were introduced.

C5-MTases M.ϕ3TI and M.SPRI had been constructed. These had their fusion point within TRD F and TRD E, recognizing GCNGC and CC(A/T)GG respectively (all DNA sequences are written 5′–3′ from left to right, methylated Cs are underlined). As a consequence, two reciprocal chimeric TRDs were generated, which had the N-terminal part of TRD F fused to the C-terminal part of TRD E (TRD FE; Figure 1B and C) or the N-terminal part of TRD E fused to the C-terminal part of TRD F (TRD EF).

To analyze the DNA recognition capacities of chimeric TRDs FE and EF, they were singularized by introducing them into the 'V' region of M.SPRMUT5, a mutant derivative of the multispecific C5-MTase M.SPRI lacking all TRDs (Walter *et al.*, 1992; Figure 1D). Earlier experiments had demonstrated that TRDs could be introduced into this construct with the concomitant acquisition of a methylation potential as specified by the incorporated TRD. For the insertion of the chimeric TRDs into M.SPRMUT5, the TRD FE- and EF-encoding gene segments were amplified by PCR and inserted into the M.SPRMUT5 gene (see Figure 1C and D). The resulting derivatives were termed M.SPR-FE and -EF. Analysis of the DNA methylation activity of the engineered MTases by an *in vitro* DNA methylation assay revealed that only M.SPR-FE was methylation proficient without further manipulation. Therefore, our studies concentrated first on an analysis of the methylation specificity of this engineered MTase.

Analysis of the DNA methylation activity of M.SPR-FE

Previously, we had attempted to determine the specificity of the MTase with TRD FE by measurement of the susceptibility of DNA methylated by this enzyme to various restriction enzymes (Balganesh *et al.*, 1987). From these studies, we concluded that the chimeric TRD had a 'relaxed' specificity, in that the enzyme introduced partial resistance to some endonucleases, to which DNA methylated by either parent of the chimera was sensitive. For a closer study of the sequence specificity of M.SPR-FE, we chemically sequenced 250 nucleotides of one strand of an

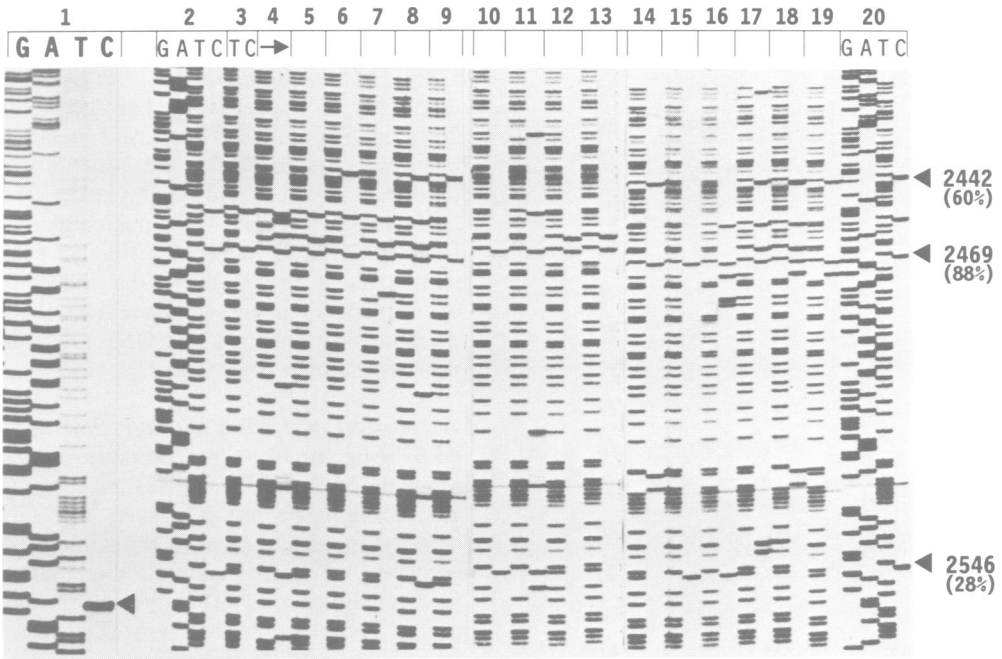

Fig. 2. Analysis of the target specificity of M.SPR-FE after bisulfite treatment and PCR amplification of *in vivo* methylated DNA (lower strand of region A). Lane 1 shows a control sequence of a plasmid (pMS-SPR-F) *in vivo* methylated by a GCNGC (*Fnu*4HI)-specific MTase (C.Lange, unpublished). All four base-specific sequencing tracks are shown. Only one C-specific band, identifying a 5mC residue, is observed within the single *Fnu*4HI site present within the section shown (indicated by an arrow; nucleotide position 2549). Hence all non-methylated C residues of the control fragment have been converted into T residues following bisulfite treatment and PCR amplification under the conditions used. Lanes 2–20 show representative sequence examples derived from individual clones of the same region after amplification from a methylated and bisulfite-treated plasmid population (pMS-SPR-FE). Only lanes 2 and 20 show all four reaction tracks, lanes 3–19 show only the pyrimidine (T and C) reaction tracks. The arrows indicate the Cs at positions 2442, 2469 and 2546 that are located within three frequently methylated GCTGGC sites contained within the section shown. The percentages that are given in parentheses describe the percentages of clones that were observed to have non-converted cytosines at the indicated positions (a total of 32 clones was analyzed).

in vivo M.SPR-FE-modified DNA fragment with the Maxam–Gilbert (M/G) procedure (Maxam and Gilbert, 1980), which identifies 5mC locations (Ohmori *et al.*, 1978). Two protected cytosine positions could be identified, which were located within the identical asymmetrical sequence GCTGGC, while no methylation of either parental target (GCNGC and CC(A/T)GG) was detectable (results not shown). However, the analyzed DNA strand contained an additional GCTGGC site which was apparently not protected against cleavage. Since the sequence obtained with M/G sequencing represents a population average for many DNA molecules, we decided to study the methylation activity of M.SPR-FE more accurately by using a more sensitive method based on the PCR amplification of bisulfite-modified DNA templates (Frommer *et al.*, 1992, in the modification of Feil *et al.*, 1994). This method provides a positive identification of 5mC residues and can determine the strand-specific methylation status of individual DNA molecules. We selected two different regions (A and B) of 435 and 551 bp, respectively, of a plasmid expressing M.SPR-FE for the analysis of the methylation patterns of both DNA strands. The choice of these regions was based on the fact that they contain both the parental GCNGC and CC(A/T)GG sites and the asymmetrical GCTGGC site identified by M/G sequencing. To facilitate a comparison between the results obtained with the two techniques, the 250 bp fragment sequenced by the M/G technique is contained in region A. In initial analyses, we observed that parallel clones of the PCR products obtained after bisulfite treatment of plasmids modified *in vivo* by M.SPR-FE differed from each other with respect to both the position and frequency of 5mC residues. Such variation was only rarely observed if, in a parallel reaction under the same conditions, methylation was mediated by one of the parental TRDs (Figures 1 and 2). Hence, we attribute the variation of the methylation patterns by M.SPR-FE mostly to its apparently inherent relaxed methylation specificity, rather than to deficiencies of the method. Given the circumstance of a relaxed specificity, we determined the methylation specificity of the chimeric MTase by a statistical analysis of the methylation patterns of 30–40 independent clones for each strand of the two analyzed DNA segments. An example of such an analysis is given in Figure 2, which shows the distribution in 19 (of a total of 32) clones of the methylation pattern observed in a segment of DNA from the lower strand of region A. A compilation of all data is shown in Figure 3. The results obtained indicate the following. (i) The methylation activity of M.SPR-FE is not directed against a unique DNA target but is characterized by a preference of the C5-MTase for methylating the 5′ C residue within the non-symmetrical hexameric sequence GCTGGC. (ii) Methylation is detectable at lower frequencies within several other non-symmetrical hexameric sequences of the general form GC(T/C)NNC/G. (iii) Methylation of GCCAGC sequences, complementary to the preferred GCTGGC target, was observed only with very low fre-

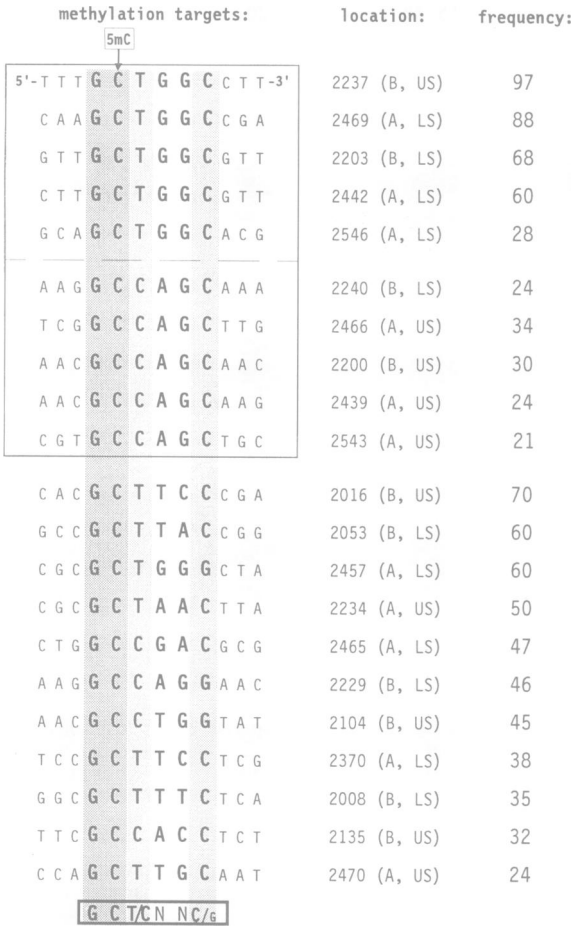

Fig. 3. Methylation profile of individual molecules within both DNA strands of regions A and B of an *in vivo* M.SPR-FE-specific methylated plasmid pMS-SPR-FE. The sequencing data (Figure 2) from the following numbers of independent clones were compiled: region A, upper DNA strand (US), 38; lower DNA strand (LS), 32; region B, US, 40; LS, 35. The average methylation frequencies are given as calculated from the percentages of clones that contained non-converted cytosines at the indicated positions. Only those positions where 20% or more of all analyzed clones contained 5mC residues were considered as significant and were included in the compilation. For each identified 5mC position, the analyzed regions (A or B) and strands (US or LS) are specified in parentheses (see also Figure 2). The upper part of the figure shows a comparison of the methylation frequencies observed within the hexameric sequence GCTGGC and its complementary GCCAGC sequences. The consensus that was inferred from all identified methylation sites is shown at the bottom of the figure. Altogether, comparisons of the nucleotide sequences surrounding all identified 5mC residues did not reveal any obvious similarities in addition to the indicated consensus positions.

quencies (Figure 3, upper part). This is also true for other pairs of complementary targets. Hence the identified targets are, for the most part, hemimethylated. (iv) No methylation was detected in either strand of a total of nine *Fnu*4HI sites (the targets of the parental TRD F) present within the analyzed DNA segments A and B. In two of the four *Eco*RII sites (the targets of the parental TRD E) contained in the segments, methylation was detected only within one of the two strands. In these targets, G residues precede the methylated C residues at their 5′ flanks (nucleotide positions 2229 and 2104 in Figure 3). (v) Quantitatively different methylation frequencies are observed at C residues within identical hexameric targets, which are surrounded by different 5′- and 3′-flanking sequences. Methylation was also detectable, albeit to a small extent (28%), within the GCTGGC site that was not protected detectably from chemical cleavage (nucleotide position 2546 in Figure 3). This result demonstrates that the chemical sequencing method is not sufficiently sensitive to detect low levels of methylation. A gel retardation analysis with short oligonucleotides containing the GCTGGC target within different sequence contexts demonstrated that, indeed, the nucleotide composition of sequences surrounding the identified targets has a pronounced influence on the DNA recognition and binding process (data not shown).

Activation of a DNA methylation activity of M.SPR-EF by site-directed mutagenesis

Since the M.SPRMUT5 derivative carrying chimeric TRD EF (M.SPR-EF) did not show any DNA methylation activity in an *in vitro* methylation assay, we tried to activate this construct by site-directed mutagenesis. The C-terminal part of TRD F differs from the corresponding part of TRD E in three regions (I, II and III in Figure 4). In order to activate the chimeric enzyme, we introduced amino acid exchanges, insertions and deletions into TRD EF, to assimilate its sequence stepwise to the corresponding sequence of TRD E (Figure 4). For an analysis of the DNA methylation capacities of the resulting mutant derivatives of M.SPR-EF (M.SPR-EF-1 to -EF-8) we tested the *in vitro* methylation activity of *Escherichia coli* cells overexpressing the different mutant proteins as described previously (Günthert *et al.*, 1986). These analyses revealed that alterations of amino acids within regions I and II or in both regions (M.SPR-EF-1–4) had no activating effect. Only the substitution of the four amino acids of region III of TRD EF by a single amino acid (L or P in mutants EF-7 and EF-8, respectively) caused an activation.

Analysis of the DNA methylation activity of M.SPR-EF-7 and -8

A study of the DNA methylation activities of both activated M.SPR-EF mutant derivatives was again carried out by sequencing PCR-amplified regions of bisulfite-treated plasmids expressing M.SPR-EF-7 or -EF-8. The same regions (A and B) that had been selected for the analysis of M.SPR-FE were analyzed in a total of 80 independent clones. These studies revealed that the methylation patterns observed with both constructs were identical. The targets recognized by M.SPR-EF-7 and -8 were much more degenerate than those observed with M.SPR-FE. A detailed compilation of the sequencing data obtained with M.SPR-EF-7 within region A is shown in Figure 5.

Analysis of the sequences surrounding the identified 5mC residues showed that all 5mCs have a C residue on their 3′ side [as in one of the parental sequences CC(A/T)GG], and A, T or C but not G 3′ to this C (5′..CC(A/T/C)..3′). In contrast to the methylation patterns observed with M.SPR-FE, no further consensus positions can be deduced from the sequencing results obtained with M.SPR-EF-7 or -8. As with M.SPR-FE, no methylated C residues are present within any of the *Fnu*4HI sites located within the analyzed DNA segments. Of the three *Eco*RII sites in fragment B, two were hemimethylated and one was weakly

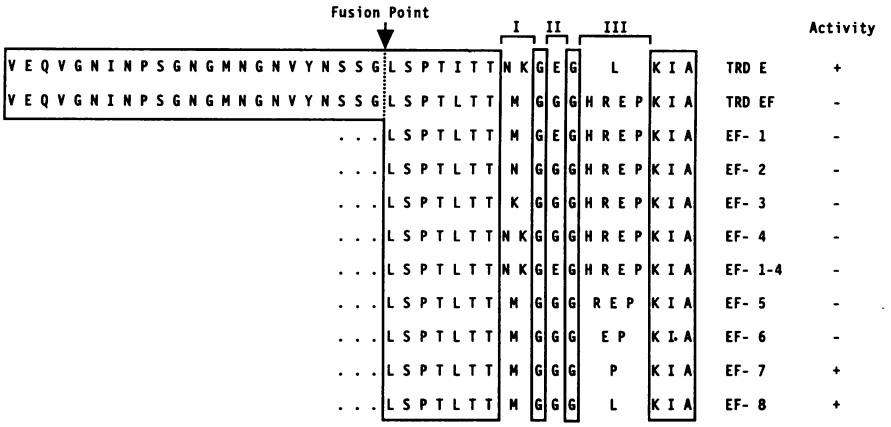

Fig. 4. Construction of mutant derivatives of chimeric TRD EF. The amino acid sequence of TRD E of the multispecific C5-MTase M.SPRI (Buhk *et al.*, 1984) is shown in comparison with the sequence of chimeric TRD EF of M.SPR-EF. The C-terminal part of TRD EF was gradually assimilated to the corresponding sequence of TRD E by a stepwise deletion, insertion or exchange of amino acid residues, using the site-directed mutagenesis procedure of Kunkel (1985). Regions I, II and III indicate the three positions where the sequences of TRD E and TRD EF diverge.

methylated in both strands, as was the only *Eco*RII site of fragment A (coordinate 2604/2600 in Figure 5). With the exception of these four (partially) methylated *Eco*RII sites, the majority of the methylated 'CC' dinucleotides are located within non-palindromic sequence contexts.

Discussion

New types of TRDs of C5-MTases could be generated by the fusion of the N- and C-terminal parts of the parental TRDs F and E. Whereas one of the TRDs (FE) was active as constructed, the reciprocal TRD EF could only be converted to methylation proficiency following changes to the amino acid spacing in its C-terminal part by site-directed mutagenesis. The targets recognized by these chimeric TRDs, particularly those of TRD EF, are degenerate, in contrast to those of the parental TRDs. The consensus sequence inferred from target site comparisons of M.SPR-FE (Figure 3) reveals a strong conservation of the nucleotides surrounding the 5mC residues in the 5′ part of the identified targets (5′..GC(T/C)..3′), while the nucleotides located more 3′ are mostly variable. Amongst the targets described by the consensus, methylation occurs with high preference in the hexameric sequence GCTGGC. For the highly degenerate targets recognized by the mutagenized TRD EF, we could only define a three nucleotide consensus including the methylated C, which is also located at its 5′ end.

In both chimeras, the 5′ ends of their targets (GC in the case of TRD FE and CC in the case of TRD EF) closely match the 5′ end of the target which is recognized by the segment of the parental TRD which is represented at the N-terminal part of the chimeric TRD. Despite the poor sequence identity found between the variable regions of the phage-encoded and the bacterial C5-MTases, this finding suggests a similar mechanism of target recognition by TRDs from multispecific and monospecific MTases. In the co-crystal structure of M.*Hha*I with its DNA substrate (Klimasauskas *et al.*, 1994), the interaction between the enzyme and the GCGC target is mediated by two successive glycine-rich recognition loops, an N-terminal loop (loop I) and a C-terminal loop (loop II). Assuming a similar tertiary structure for multispecific MTases, the fusion point of the chimeric TRDs studied here would fall between the two recognition loops. Aligning the amino acid sequences of the chimeric TRDs FE and EF with the region of M.*Hha*I, building these recognition loops, we realize that R240 within loop I, which makes two contacts with the 5′ G of the target C in the co-crystal, is also present at the same position within TRD FE and in the majority of other C5-MTases with 5′..GC..3′ methylation activity (Figure 6). By analogy with the co-crystal structure of M.*Hha*I, we attribute the methylation of the 5′ GC dinucleotide by TRD FE to the presence of the N-terminal subdomain of TRD F containing the R residue critical for contacting the 5′ G (Figure 6). None of the putative TRDs of C5-MTases having other nucleotides 5′ to their target cytosines (e.g. the TRD of M.*Eco*RII or TRD EF) contain this R at the same spacing from the conserved T(L/I/V) motif. This provides a consensus primary sequence [R(N)$_9$ T(L/I/V)] to be found in the majority of 5′..GC..3′-recognizing TRDs which we have studied. Further support for this concept came from the more recent publication of the crystal structure of another monospecific bacterial C5-MTase, M.*Hae*III, complexed with its GGCC target site (Reinisch *et al.*, 1995). Although the structure of the DNA-contacting region of M.*Hae*III clearly differs from the M.*Hha*I structure, R227, which corresponds to R240 of M.*Hha*I, is making a bidentate contact to the G located immediately 5′ to the target cytosine. In M.*Hae*III, the residues that interact directly with the bases lie on either side of a short helix (residues 220–223 in Figure 6) or they are located in a short hairpin turn (residues 241–244). These structural elements fall exactly within the region that corresponds to the two recognition loops of M.*Hha*I. Despite the structural differences of the enzymes and the different configurations of DNA targets following the interaction with these C5-MTases, the basic concept of target recognition deduced from the M.*Hha*I structure appears to hold also for M.*Hae*III: residues that are located in the N-terminal part of the DNA-contacting protein region do mainly interact with the 5′ part of the recognition sequence (see Reinisch *et al.*, 1995).

The methylation patterns that were observed with the activated chimeric TRD EF support the concept of a correspondence between the N-terminal segment of a TRD

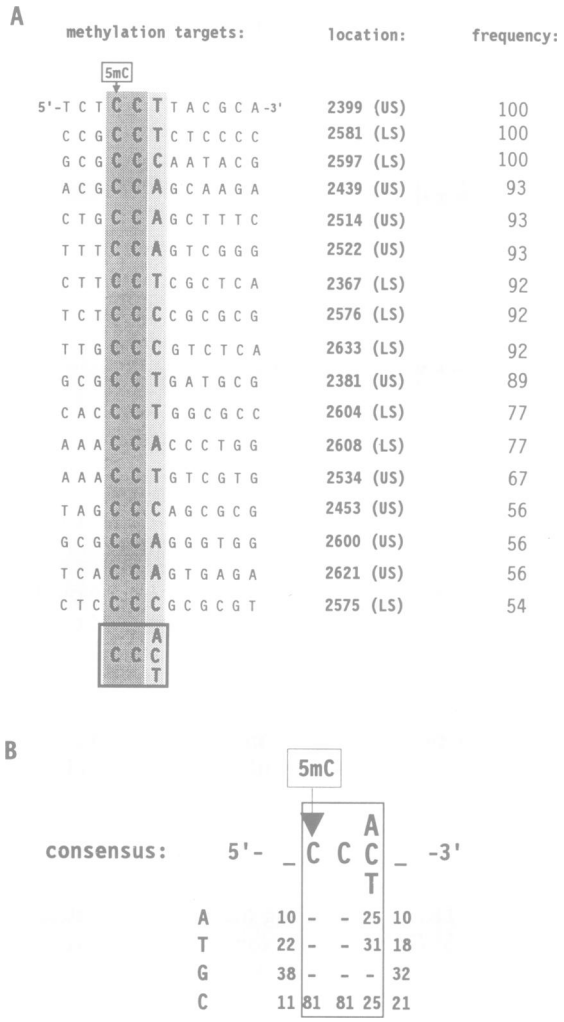

Fig. 5. Compilation of the data obtained after sequencing of bisulfite-treated plasmids overexpressing M.SPR-EF-7 and -EF-8 (see Figure 4). (**A**) A total number of 80 independent clones of fragments A and B within plasmids pMS-SPR-EF-7 and -8 were analyzed. Within both analyzed regions, 81 sites were found to be frequently methylated. Representation of the distribution and sequence context of 5mC residues within both strands of region A of the *in vivo* M.SPR-EF-7-specific methylated plasmid pMS-SPR-EF-7. (**B**) Distribution of bases within and outside the consensus amongst 81 frequently methylated Cs in regions A and B.

and the recognition of the 5′ half of the methylated target site. This TRD, in which the N-terminal part (corresponding to recognition loop I) originates from the CC(A/)TGG-recognizing TRD E, mediates the methylation of the 5′ C residue within the dinucleotide 5′..CC..3′. In the M.*Hha*I structure, recognition loop II interacts extensively with the phosphate backbone of the methylated DNA strand and primarily makes contacts with bases in the 3′ half of the target site (Klimasauskas *et al.*, 1994). These interactions are largely conserved in M.*Hae*III and M.*Hha*I (Reinisch *et al.*, 1995). With the exception of the—generally—conserved T residue, none of the DNA-contacting residues of M.*Hha*I are found at an identical position within the corresponding regions of TRDs FE and EF.

The strand symmetry and the stringency of target recognition observed with the parental TRDs and their loss following the formation of chimeric TRDs indicate a co-evolution and coordinated activity of the two recognition loops, leading to a finely tuned ensemble of peptide motifs. The amino acid composition of the recognition loops and their spacing in relation to each other and within the variable region of C5-MTases are elements determining this interdependence and hence the specificity of target recognition. We also attribute the absence of activity in the case of TRD EF to the incompatibility between its composite subdomains. It was only after mutagenesis within the TRD F-derived part of this chimera that methylation proficiency was observed. In these experiments, a hierarchy became detectable in the significance of alterations causing activity: whereas assimilations to TRD E in regions I and II and in both regions had no effect, activation was possible by reducing the number of amino acids within region III from four to one (Figure 4). The significance of the spacing of amino acids in this region also became apparent in mutagenesis studies with TRD FE, where the addition of amino acids to region III destroyed the methylation capacity of the resulting mutant enzymes (data not shown). Further mutagenesis experiments involving the spacing between the putative recognition subdomains and the introduction of amino acid exchanges within the N-terminal moieties of the TRDs could lead to enzymes with further, and possibly more defined, target specificities.

In the specificity-determining proteins, S, of type I restriction–modification systems, separate subdomains responsible for target recognition have been identified (Nagaraja *et al.*, 1985; Gann, 1987). The targets of type I enzymes are bipartite, consisting of defined sequences of 3 bp at their 5′ ends and 4 or 5 bp at their 3′ ends, separated by 6–8 variable base pairs (Wilson and Murray, 1991). Within the S proteins of these enzymes, separated subdomains recognize the 5′ and 3′ segments of the target. Such subdomains derived from enzymes with different specificities can be reassorted (Fuller-Pace *et al.*, 1984; Nagaraja *et al.*, 1985; Gann, 1987). The recombinant S proteins resulting from such reassortment now recognize DNA targets whose conserved 5′ sequence is that of one parental target, its 3′ sequence that of the other. Hence the target-recognizing subdomains of type I S proteins are fully independent recognition elements, in contrast to those of type II C5-MTases which we have described here.

Both engineered enzymes, M.SPR-FE and -EF, can recognize and methylate non-symmetrical DNA targets. These results illustrate that the palindromic nature of the targets usually recognized by the C5-MTases is not a prerequisite for the DNA recognition and methylation process. They also demonstrate the possibility of generating C5-MTases with new DNA methylation specificities, which would not necessarily be directed against symmetrical DNA targets.

Materials and methods

Bacterial strains, phages and media

The *E.coli* K12 strain GM271 (*mcrA*$^-$, *mcrB*-1, *hsdR*-2, *dcm*-6; obtained from M.G.Marinus) was used for propagating plasmids encoding M.SPR wild-type and mutant derivatives. XL1Blue [*endA*1, *gyrA*46, *hsdR*17, *lac*, *recA*1, *relA*1, *supE*44, *thi*, F′*lac*: *lac*Iq, Δ(*lacZ*)M15, Tn*10*, *proA*+, *proB*+; Sambrook *et al.*, 1989] and RZ1032 {HfrKL16 PO/45 [*lys*A(61–62)], *dut*-1, *ung*-1, *thi*-1, *rel*A-1, *sup*E44, Zbd-279::Tn*10*; Kunkel *et al.*, 1987} were used for the cloning with M13mp18/19 vectors. OneShot

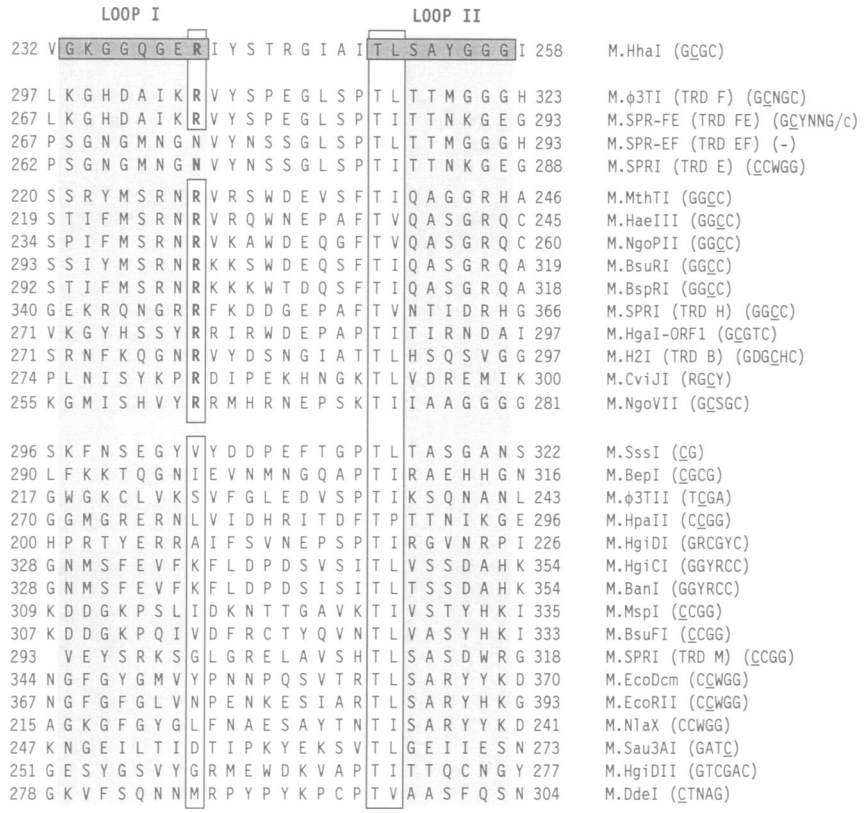

Fig. 6. Amino acid sequence alignment of the (putative) TRDs of various multi- and monospecific C5-DNA MTases. The amino acid sequences of the 'V' regions of the different MTases were aligned at the conserved T(L/I/V) motif (Lauster *et al.*, 1989), which is boxed, as well as the position 10 amino acids N-terminal to this motif, which is frequently occupied by R. While the TRDs E, F, M, H and B of the multispecific phage-encoded MTases and the TRD of the monospecific bacterial MTases M.*Hha*I, M.*Hae*III have been defined experimentally (Balganesh *et al.*, 1987; Trautner *et al.*, 1988; 1996; Wilke *et al.*, 1988; Walter *et al.*, 1992; Klimasauskas *et al.*, 1994; Reinisch *et al.*, 1995), the TRDs of all other bacterial enzymes were only predicted by comparative analyses (Lauster *et al.*, 1989 and unpublished results). The regions of the different TRDs corresponding to the two recognition loops of M.*Hha*I contacting the GCGC target site in the co-crystal structure (Klimasauskas *et al.*, 1994), are highlighted. The numbers indicate the amino acid coordinates of the respective enzymes. Only those segments of the TRDs corresponding to the region of M.*Hha*I building the two recognition loops are shown. The sequences recognized by the different MTases are given in parentheses; the target C residues are underlined. References for the different C5-MTases included in this alignment are summarized elsewhere (Kumar *et al.*, 1994). The amino acid sequence of M.*Ngo*VII was kindly provided by D.Stein, prior to publication; the sequence of M.φ3TII was taken from Noyer-Weidner *et al.* (1994).

INVαF′-competent *E.coli* cells included in the TA cloning kit (Invitrogen) were used for cloning the PCR amplification products from bisulfite-modified plasmids.

Reagents and general techniques

Restriction endonucleases, Klenow fragment of DNA polymerase I, T4 DNA ligase and *Taq* DNA polymerase were purchased from Boehringer Mannheim; shrimp alkaline phosphatase and [^{35}S]dATP were obtained from USB/Amersham. All enzymes were used as recommended by the manufacturer. All molecular biological manipulations followed standard procedures (Sambrook *et al.*, 1989). DNA sequencing was carried out by the dideoxy chain termination method (Sanger *et al.*, 1977) using a T7 sequencing kit from Pharmacia. All oligonucleotides were synthesized on an Applied Biosystems Synthesizer.

Construction of M.SPRI derivatives containing chimeric TRDs

For the insertion of the chimeric TRDs into M.SPRMUT5 (Walter *et al.*, 1992), the TRD FE- and EF-encoding gene segments were amplified by PCR from plasmids pBB13 and pBB14, encoding the chimeric MTases (chimeras 3 and 4; Balganesh *et al.*, 1987), using oligonucleotides complementary to the precisely defined boundaries of TRDs F and E (Trautner *et al.*, 1996). The following pairs of oligonucleotides served as primers for the PCR-based amplification: TRD FE: 5′-GAAGTAA-GGGAGCCGTTA-3′/5′-AACTGCAATTTTCAGTCC-3′; TRD EF: 5′-GTTGAGCAAGTAGGTAAC-3′/5′-TGCGATTTTAGGTTCTCTGTG-3′. The amplified DNA fragments were inserted in the sense orientation into a single *Hpa*I site within the 'V' region of the M.SPRMUT5 gene (see Figure 1) through blunt end ligations.

Construction of mutant M.SPR-EF derivatives

Mutant derivatives of M.SPR-EF were constructed by site-directed mutagenesis according to Kunkel (1985), by using appropriate 24–35mer oligonucleotides. The genes encoding M.SPR-EF-5, -6 and -7 result from the stepwise deletion of the nucleotide triplets corresponding to the amino acid positions 293 (ΔH); 293, 294 (ΔH, R); and 293, 294, 295 (ΔH, R, E) of M.SPR-EF. In all cases, mutagenesis was verified by DNA sequencing of the regions concerned. For expression of the mutant MTases, the mutated genes were cloned into the expression vector pMS119EH (Fürste *et al.*, 1986).

Analysis of methylation patterns after bisulfite treatment

For the expression of the MTases, the genes encoding M.SPRI and derivatives of M.SPRI containing single or chimeric TRDs were cloned into the expression vector pMS119EH, a derivative of pJF118 (Fürste *et al.*, 1986) and transformed into *E.coli* strain GM271. Bisulfite treatment of *in vivo* methylated plasmids expressing M.SPRI or its derivatives, that had been isolated from non-induced transformants, was performed for 18 h at 50°C following the protocol of Frommer *et al.* (1992) with the following modification. For the recovery of the reacted DNA from the unreacted bisulfite, it was bound to 25 μl of JETSORB suspension (Genomed kit) and washed twice with buffer A2 (Genomed). Regions A and B, located within the pMS119EH moiety of the bisulfite-reacted plasmids, were amplified by PCR from the upper and lower DNA strands in separate reaction mixtures. The following pairs of strand-specific PCR primers were used. For region A (upper strand) 5′-CAAACAATAAA-AAACAATCAACTATTACCCA-3′/5′-TGATTTTGTGGAGAATTCT-ATTATTGTTTTT-3′ and (lower strand) 5′-TTATCCCCGAATTCTATA-AATAACCATATTAC-3′/5′-GGGTAAATTAGTGTGGATTGTTTG-3′

were used to generate a 384 and a 434 bp DNA fragment, respectively; for region B (upper strand) 5′-CAATTATCCACAAAATCAAAAAATAACACAAA-3′/5′-TAAGAATTTTGTAGTATTGTTTATATATTTTG-3′ and (lower strand) 5′-CACTCTACTAATCCTATTACCAATAAC-3′/5′-GTTTGGTTGTGGTGAGTGGTATTAGTTTATTT-3′ were used to generate a 507 and a 520 bp DNA fragment, respectively. Standard PCR conditions, with the annealing temperatures varying between 55 and 60°C, were used. The amplified DNA fragments were cloned directly into pre-cut plasmid pCRII contained in a TA cloning kit (Invitrogen). Plasmid preparations from individual clones were sequenced directly.

Acknowledgements

We are grateful to J.Walter for his advice with the genomic sequencing method. We thank him and M.Noyer-Weidner for helpful comments on the manuscript and R.Lurz for his cheerful assistance with all computer graphics. This work was partially supported by SFB344 of the Deutsche Forschungsgemeinschaft.

References

Balganesh,T.S., Reiners,L., Lauster,R., Noyer-Weidner,M., Wilke,K. and Trautner,T.A. (1987) Construction and use of chimeric SPR/ϕ3T DNA methyltransferases in the definition of sequence recognizing enzyme regions. *EMBO J.*, **6**, 3543–3539.

Behrens,B., Noyer-Weidner,M., Pawlek,B., Lauster,R., Balganesh,T.S. and Trautner,T.A. (1987) Organization of multispecific DNA methyltransferases encoded by temperate *Bacillus subtilis* phages. *EMBO J.*, **6**, 1137–1142.

Buhk,H.J., Behrens,B., Tailor,R., Wilke,K., Prada,J.J., Guenthert,U., Noyer-Weidner,M., Jeubier,S. and Trautner,T.A. (1984) Restriction and modification in *Bacillus subtilis*: nucleotide sequence, functional organization and product of the DNA methyltransferase gene of bacteriophage SPR. *Gene*, **29**, 51–61.

Cheng,X., Kumar,S., Posfai,J., Pflugrath,J.W. and Roberts,R.J. (1993) Crystal structure of the *Hha*I DNA methyltransferase complexed with S-adenosyl-L-methionine. *Cell*, **74**, 299–307.

Feil,R., Charlton,J., Bird,A.P., Walter,J. and Reik,W. (1994) Methylation analysis on individual chromosomes: improved protocol for bisulphite genomic sequencing. *Nucleic Acids Res.*, **22**, 695–696.

Frommer,M., McDonald,L.E., Millar,D.S., Collis,C.M., Watt,F., Grigg,G.W., Molloy,P.L. and Paul,C.L. (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl Acad. Sci. USA*, **89**, 1827–1831.

Fuller-Pace,F.V. (1984) Genetic recombination can generate altered restriction specificity. *Proc. Natl Acad. Sci. USA*, **81**, 6095–6099.

Fürste,J.P., Pansegrau,W., Frank,R., Bloecker,H., Scholz,P., Bagdasarian,M. and Lanka,E. (1986) Molecular cloning of the plasmid RP4 primase region in a multi-hostrange TacP expression vector. *Gene*, **48**, 119–131.

Gann,A.C. (1987) Reassortment of DNA recognition domains and the evolution of new specificities. *Mol. Microbiol.*, **1**, 13–22.

Günthert,U., Lauster,R. and Reiners,L. (1986) Multispecific DNA methyltransferases from *Bacillus subtilis* phages. Properties of wild-type and various mutant enzymes with altered DNA affinity. *Eur. J. Biochem.*, **159**, 485–492.

Klimasauskas,S., Nelson,J.L. and Roberts,R.J. (1991) The sequence specificity domain of cytosine-C5 methylases. *Nucleic Acids Res.*, **19**, 6183–6190.

Klimasauskas,S., Kumar,S., Roberts,R.J. and Cheng,X. (1994) *Hha*I methyltransferase flips its target base out of the DNA helix. *Cell*, **76**, 357–369.

Kumar,S., Cheng,X., Klimasauskas,S., Mi,S., Pósfai,J., Roberts,R.J. and Wilson,G.G. (1994) The DNA (cytosine-5) methyltransferases. *Nucleic Acids Res.*, **22**, 1–10.

Kunkel,T.A. (1985) Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl Acad. Sci. USA*, **82**, 488–492.

Kunkel,T.A., Roberts,J.D. and Zakour,R.A. (1987) Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods Enzymol.*, **154**, 367–382.

Lange,C., Jugel,A., Walter,J., Noyer-Weidner,M. and Trautner,T.A. (1991) 'Pseudo' domains in phage-encoded DNA methyltransferases. *Nature*, **352**, 645–648.

Lauster,R., Trautner,T.A. and Noyer-Weidner,M. (1989) Cytosine-specific type II DNA methyltransferases—a conserved enzyme core with variable target-recognizing domains. *J. Mol. Biol.*, **206**, 305–312.

Maxam,A.M. and Gilbert,W. (1980) Sequencing with base-specific chemical cleavages. *Methods Enzymol.*, **65**, 499–559.

Nagaraja,V., Shepherd,J.C.W. and Bickle,T. (1985) A hybrid recognition sequence in a recombinant restriction enzyme and the evolution of DNA sequence specificity. *Nature*, **316**, 371–372.

Noyer-Weidner,M., Walter,J., Terschüren,P.-A., Chai,S. and Trautner,T.A. (1994) M.ϕ3TII: a new monospecific DNA (cytosine-C5) methyltransferase with pronounced amino acid sequence similarity to a family of adenine-N6-DNA-methyltransferases. *Nucleic Acids Res.*, **22**, 4066–4072.

Ohmori,H., Tomizawa,J. and Maxam,A.M. (1978) Detection of 5-methylcytosine in DNA sequences. *Nucleic Acids Res.*, **5**, 1479–1485.

Pósfai,J., Bhagwat,A.S., Pósfai,G. and Roberts,R.J. (1989) Predictive motifs derived from cytosine methyltransferases. *Nucleic Acids Res.*, **17**, 2421–2435.

Reinisch,K.M., Chen,L., Verdine,G.L. and Lipscomb,W.N. (1995) The crystal structure of *Hae*III methyltransferase covalently complexed to DNA: an extrahelical cytosine and rearranged base pairing. *Cell*, **82**, 143–153.

Sambrook,F., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sanger,F., Nicklen,S. and Coulson,A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl Acad. Sci. USA*, **74**, 5463–5467.

Tran-Betcke,B.A., Behrens,B., Noyer-Weidner,M. and Trautner,T.A. (1986) DNA methyltransferase genes of *Bacillus subtilis* phages: comparison of their nucleotide sequences. *Gene*, **42**, 89–96.

Trautner,T.A. and Noyer-Weidner,M. (1992) Restriction/modification – and methylation systems in *Bacillus subtilis*, related species and their phages. In Hoch,J.A., Losick,R. and Sonenshein,A.L. (eds), *Bacillus subtilis, the Model Gram-Positive Bacterium: Physiology, Biochemistry and Molecular Genetics*. ASM, Washington, DC, pp. 539–552.

Trautner,T.A., Balganesh,T.S. and Pawlek,B. (1988) Chimeric multispecific DNA methyltransferases with novel combinations of target recognition. *Nucleic Acids Res.*, **16**, 6649–6656.

Trautner,T.A., Pawlek,B., Behrens,B. and Willert,J. (1996) Exact size and organization of DNA target recognizing domains of multispecific DNA-(cytosine-C5)-methyltransferases. *EMBO J.*, **15**, 1434–1442.

Walter,J., Trautner,T.A. and Noyer-Weidner,M. (1992) High plasticity of multispecific DNA methyltransferases in the region carrying DNA target recognizing enzyme modules. *EMBO J.*, **11**, 4445–4450.

Wilke,K., Rauhut,E., Noyer-Weidner,M., Lauster,R., Pawlek,B., Behrens,B. and Trautner,T.A. (1988) Sequential order of target-recognizing domains in multispecific DNA-methyltransferases. *EMBO J.*, **7**, 2601–2609.

Wilson,G.G. and Murray,N.E. (1991) Restriction and modification systems. *Annu. Rev. Genet.*, **25**, 585–627.

Received on July 30, 1995; revised on November 20, 1995