The EMBO Journal vol.15 no.6 pp.1434–1442, 1996

Exact size and organization of DNA target-recognizing domains of multispecific DNA-(cytosine-C5)-methyltransferases

Thomas A.Trautner[1], Brigitte Pawlek, Bernhard Behrens and Jürgen Willert

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

[1]Corresponding author

A large portion of the sequences of type II DNA-(cytosine-C5)-methyltransferases (C5-MTases) represent highly conserved blocks of amino acids. General steps in the methylation reaction performed by C5-MTases have been found to be mediated by some of these domains. C5-MTases carry, in addition at the same relative location, a region variable in size and amino acid composition, part of which is associated with the capacity of each C5-MTase to recognize its characteristic target. Individual target-recognizing domains (TRDs) for the targets CCGG (M), CC(A/T)GG (E), GGCC (H), GCNGC (F) and G(G/A/T)GC(C/A/T)C (B) could be identified in the C-terminal part of the variable region of multispecific C5-MTases. With experiments reported here, we have established the organization of the variable regions of the multispecific MTases M.SPRI, M.φ3TI, M.H2I and M.ρ11$_S$I at the resolution of individual amino acids. These regions comprise 204, 175, 268 and 268 amino acids, respectively. All variable regions are bipartite. They contain at their N-terminal side a very similar sequence of 71 amino acids. The integrity of this sequence must be assured to provide enzyme activity. Bracketed by 6–10 'linker' amino acids, they have, depending on the enzyme studied, towards their C-terminal end ensembles of individual TRDs of 38 (M), 39 (E), 40 (H), 44 (F) and 54 (B) amino acids. TRDs of different enzymes with equal specificity have the same size. TRDs do not overlap but are either separated by linker amino acids or abut each other.

Keywords: DNA recognition by enzymes/molecular evolution/site-directed mutagenesis

Introduction

DNA methyltransferases (MTases) are ubiquitously distributed enzymes which catalyze the steps which eventually lead to post-replicative methylation of DNA (for reviews, see Wilson and Murray, 1991; Noyer-Weidner and Trautner, 1993). Reactions commonly found in DNA methylation include non-specific attachment of the enzyme to DNA, recognition of an enzyme-specific DNA target, binding of the methyl group donor *S*-adenosylmethionine (SAM) to the enzyme and the direct transfer of the methyl group to a base contained within the DNA target sequence with the release of methylated DNA, *S*-adenosylhomocysteine (SAH) and free enzyme. The specificity of methylation concerns the DNA base, which becomes methylated, and the atom within the base to which the methyl group will be transferred, leading to the distinction of 6-amino-A-, 4-amino-C- and C5-C methylation. These methylations are performed by three separate classes of MTases: A-N6-, C-N4- and C5-MTases. In addition to this, the specificity of methylation is exerted by the selective recognition of defined DNA target sequences, within which methylation occurs. The division of MTase function into mediating common steps of methylation and recognition of a specific DNA target is reflected in the organization of these enzymes. In particular, comparison of the amino acid sequences of some 50 different C5-MTases has revealed a separation into extensively conserved sequence motifs found in all enzymes, as opposed to a highly variable region, characterizing individual enzymes (Som *et al.*, 1987; Lauster *et al.*, 1989; Pósfai *et al.*, 1989; Noyer-Weidner and Trautner, 1993; Kumar *et al.*, 1994; Noyer-Weidner *et al.*, 1994). It was plausible to assign in these enzymes the conserved sequence motifs to an enzyme core, providing the structural requirements for enzyme activity and mediating general steps of the methylation reaction. One of the conserved motifs could indeed be shown to be involved in the covalent coupling between the enzyme and the cytosine to be methylated, which is an initial step in the methyl group transfer by C5-MTases (Wu and Santi, 1987; Chen *et al.*, 1991). Another conserved motif is involved in the binding of SAM, the universal methyl group donor (Ingrosso *et al.*, 1989). In contrast, the variable region of C5-MTases was held responsible for the individual capacity of each C5-MTase to recognize its specific DNA target(s). The elucidation of the three-dimensional structure of the C5-MTase M.*Hha*I (Cheng *et al.*, 1993) and of the same MTase complexed with its DNA target (Klimasauskas *et al.*, 1994) as well as of the M.*Hae*III–DNA complex (Reinisch *et al.*, 1995) has, as well as defining the structural parameters of MTase–DNA interaction, demonstrated the validity of this concept of a division of C5-MTases into catalytic and recognition domains.

Experimental proof for such a division into two types of basically different enzyme elements was first provided by genetic work performed with multispecific C5-MTases of temperate *Bacillus subtilis* phages (Balganesh *et al.*, 1987; Trautner *et al.*, 1988; Wilke *et al.*, 1988; Lange *et al.*, 1991a,b; Walter *et al.*, 1992) and also with the monospecific C5-MTases M.*Hpa*II and M.*Hha*I (Klimasauskas *et al.*, 1991). One of our experimental approaches towards this concept was based on the identification of two classes of mutations affecting the phenotypes of multispecific C5-MTases (Wilke *et al.*, 1988). Class I mutations largely eliminated the methylation capacity of C5-Mtases. These mutations caused amino

Table I. Methylation potential of multispecific C5-MTases

C5-MTase	Methylation potential and description of TRDs[a]				
	B[b]	F or (F)[c]	E or (E)	M	H
M.SPRI (204)[d]	–	–	+	+	+
M.φ3TI (175)	–	+	–	–	+
M.H2I (268)	+	+	(–)[c]	–	+
M.ρ11$_S$I (268)	+	(–)	(–)	–	+

[a]The TRDs are described in their N- to C-terminal order.
[b]The abbreviations correspond to the following specificities and their analogs in restriction–modification systems (C* = methylated base): B: *Bsp*1286I [G(G/A/T)GC*(C/A/T)C]; F: *Fnu*4HI (GC*NGC); E: *Eco*RII [C*C(A/T)GG]; M: *Msp*I (C*CGG); H: *Hae*III (GGC*C).
[c]Bracketed entries describe pseudodomains.
[d]The numbers in parentheses give the length of the variable region.

acid alterations exclusively in the conserved MTase regions. Class II mutations selectively affected only one of the multiple methylation capacities of such enzymes. They caused amino acid exchanges exclusively in the C-terminal part of the variable enzyme region and defined contiguous, apparently consecutively arranged sequence elements, which we termed target-recognizing domains (TRDs). The resolution of our previous analysis (Wilke *et al.*, 1988) was limited to only two C5-MTases and could not define all sequence parameters of the variable region of these enzymes and their TRDs. In particular, we could not determine the exact size of individual TRDs. We could not prove rigorously whether individual TRDs of a multispecific C5-MTase were overlapping, were separated by 'linker' amino acids not involved in target recognition, or immediately followed each other. Neither did we obtain information on how the ensemble of TRDs within the variable region was related to those amino acid sequences of the variable region apparently not responsible for target recognition.

Here we present, at the resolution of individual amino acids, the organization of the variable regions of the multispecific enzymes M.SPRI (Buhk *et al.*, 1984), M.φ3TI (Tran-Betcke *et al.*, 1986), M.H2I (Lange *et al.*, 1991b) and M.ρ11$_S$I (Behrens *et al.*, 1987). This information was derived from experiments involving site-directed mutagenesis specifically concerning those amino acids which were inferred to be at the amino or carboxy ends of each TRD. Further information was obtained following exchanges of individual TRDs and their derivatives between various MTases. It was conjectured, furthermore, from comparisons between the various variable regions of different enzymes, where TRDs with identical specificity are located in different sequence contexts. The phenotypes of mutations introduced into the region N-terminal to the TRDs were analyzed to determine the significance of this part of the variable region for enzyme activity.

Results

All experiments reported concern the multispecific C5-MTases M.SPRI, M.φ3TI, M.H2I and M.ρ11$_S$I. The methylation potential of these enzymes, the sizes of their variable regions and the order of TRDs are given in Table I. Sequence comparisons of multispecific TRDs, together with mutagenesis experiments, had previously assigned the TRDs of multispecific enzymes to the C-terminal part of the variable region. Sequences N-terminal to this area are similar among all multispecific enzymes but distinct from variable sequences of monospecific MTases at equivalent positions. The TRDs of different MTases which recognize the same target only differ from each other by a few amino acids (Figure 4).

Two approaches were used here to establish the organization of the enzymes' variable regions, which include their TRDs. (i) Point mutations, in-frame deletions or insertions, and substitutions of sense by missense sequences were introduced at various defined locations in the variable regions by site-directed mutagenesis or by otherwise manipulating the sequences. As reported previously (Wilke *et al.*, 1988), three kinds of mutations could be identified as a consequence of mutagenesis: neutral mutations without an altered phenotype (class 0); those which destroyed the methylation capacity of the enzyme altogether (class I) and those which affected one or occasionally two of the multiple methylation potentials [class II, with an index (E, M, etc.) indicating the TRD concerned]. Such analyses were performed with M.SPRI to characterize its TRDs E, M and H and that part of the variable region preceding these TRDs, with M.φ3TI (TRDs F and H) and, to a lesser extent, with M.H2I to identify TRD B. (ii) We have reported previously (Walter *et al.*, 1992) the possibility of studying the properties of individual TRDs of multispecific MTases either by deletion of accompanying TRDs or by transferring the genomic equivalent of a putative TRD, following PCR amplifiction, into a gene of a multispecific MTase whose TRDs had been eliminated. Here we have used overlapping primers with staggered termini in nested PCR amplification to generate sets of DNA fragments corresponding to a TRD region, in which successive amino acids at either the N- or C-terminus of presumed TRDs were systematically eliminated. Determining the methylation proficiency of a recipient MTase into whose gene such a set of DNA fragments had been transferred allowed the identification of the terminal amino acids of the TRD. Using this technique, we have established the confines of TRDs F of M.φ3TI and B of M.H2I or M.ρ11$_S$I.

The confines of individual TRDs

In M.SPRI, the TRDs are arranged in the N- to C-terminal order E, M and H (Wilke *et al.*, 1988). Figure 1 shows the amino acid sequence of the variable region of M.SPRI with all mutational changes observed and their phenotypes. Figure 2 gives a selection of gel electrophoresis patterns of plasmids carrying the M.SPRI gene in its wild-type and various mutant forms following cleavage by the restriction endonucleases R.*Eco*RII, R.*Msp*I and R.*Hae*III. This analysis defined the phenotype of the mutants concerned.

TRD E. The most N-terminal class II mutation providing a phenotype in which the activity of TRD E is affected (class II$_E$) is represented by V254→E (Figures 1 and 2). A series of mutations of the 10 amino acids preceding this location (Figures 1 and 2) were class 0 mutations. Therefore, V254 must be the N-terminal amino acid of TRD E. Of several other class II$_E$ mutations downstream

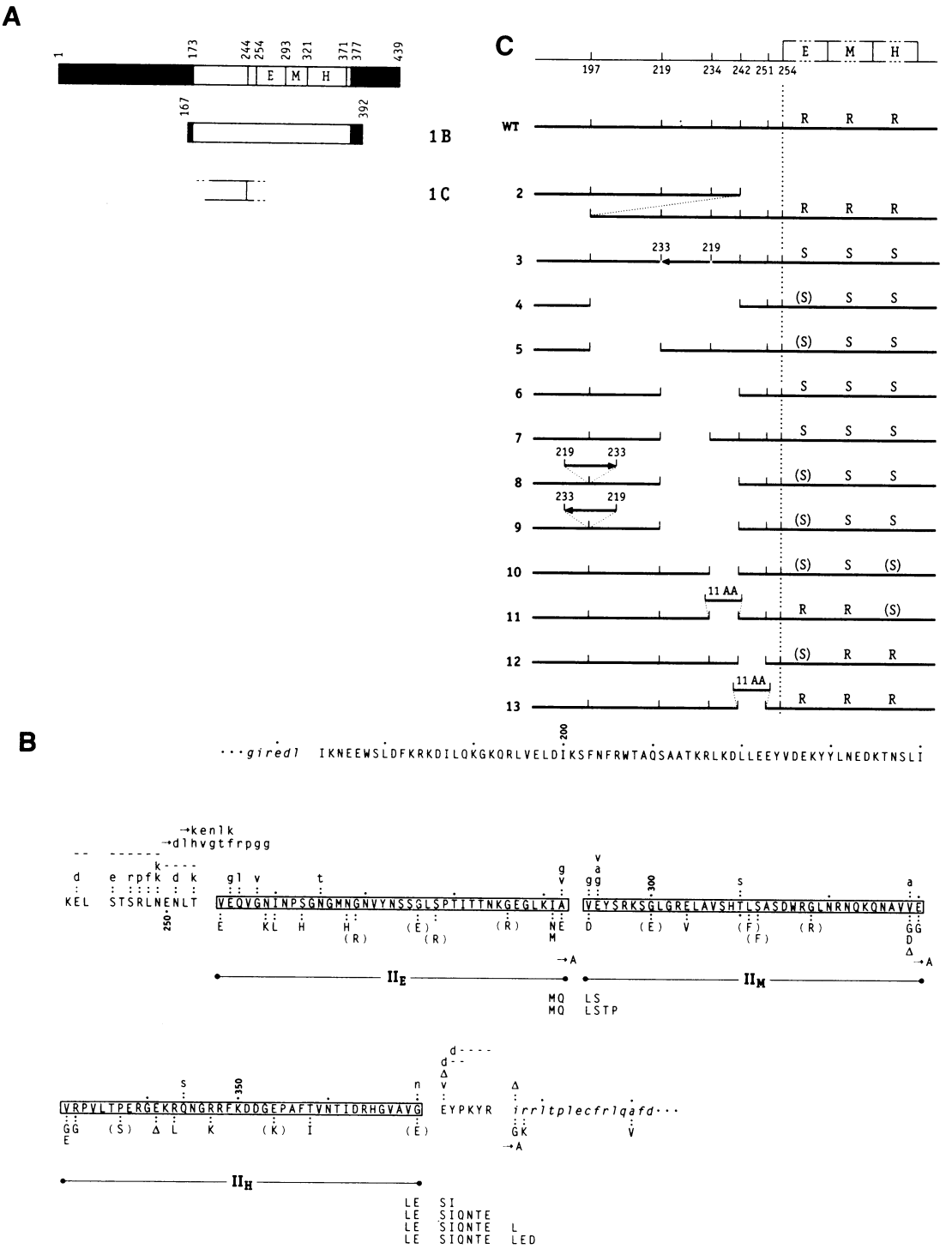

Fig. 1. Distribution of mutational amino acid changes in the variable region of M.SPRI. (**A**) Schematic drawing of M.SPRI. The numbers are the coordinates of amino acids. The filled bars are the conserved regions; the open bar is the variable region V. E, M and H describe the TRDs. The regions described in (B) and (C) are indicated. (**B**) The amino acid sequence (coordinates: Tran-Betcke *et al.*, 1986) of the variable region of the wild-type enzyme. Amino acids of the conserved domains of the protein, bracketing the variable region, are shown in italics, amino acids of the variable region in upper case letters. Amino acid changes, which generate an altered phenotype (class I or II), are shown as upper case letters below the wild-type sequence, neutral amino acid changes without a change of phenotype (class 0) in lower case letters above. Deletions are indicated by △ or – . Positions of insertions are marked by a horizontal arrow (→). These comprise substitutions of N251 and T253 by 11 and five amino acids, respectively, and the insertions of single As between amino acids at coordinates 292–293, 330–331 and 377–378. Class II mutations falling within the confines of the three horizontal lines had the phenotypes II_E, II_M and II_H respectively. The corresponding TRDs E, M and H (in this order) are boxed. Multiple amino acid exchanges affecting one or more TRDs are described below the horizontal lines. For their phenotypes see text. Mutations at coordinates 377 and 378 were class I. Some relevant mutations described by Wilke *et al.* (1988) (bracketed entries) are included in the figure. (**C**) Schematic representation of mutations 2–13 introduced into the upstream region of the TRDs together with their phenotypes. The numbers are the coordinates of (B), where neutral mutations generating *Bgl*I sites were introduced. Manipulations involving these sites generated the in-frame mutations 2 –13. Gaps indicate deletions; arrows between the coordinates an inversion (mutation 3) or insertion (mutations 8 and 9). In mutations 11 and 13, the amino acids between coordinates 234–242 and 242–251 were replaced by a *Bgl*I fragment encoding 11 amino acids unrelated to the M.SPRI gene. R, S and (S) indicate resistance, sensitivity or partial sensitivity against degradation by either R.*Eco*RII, R.*Msp*I or R.*Hae*III.

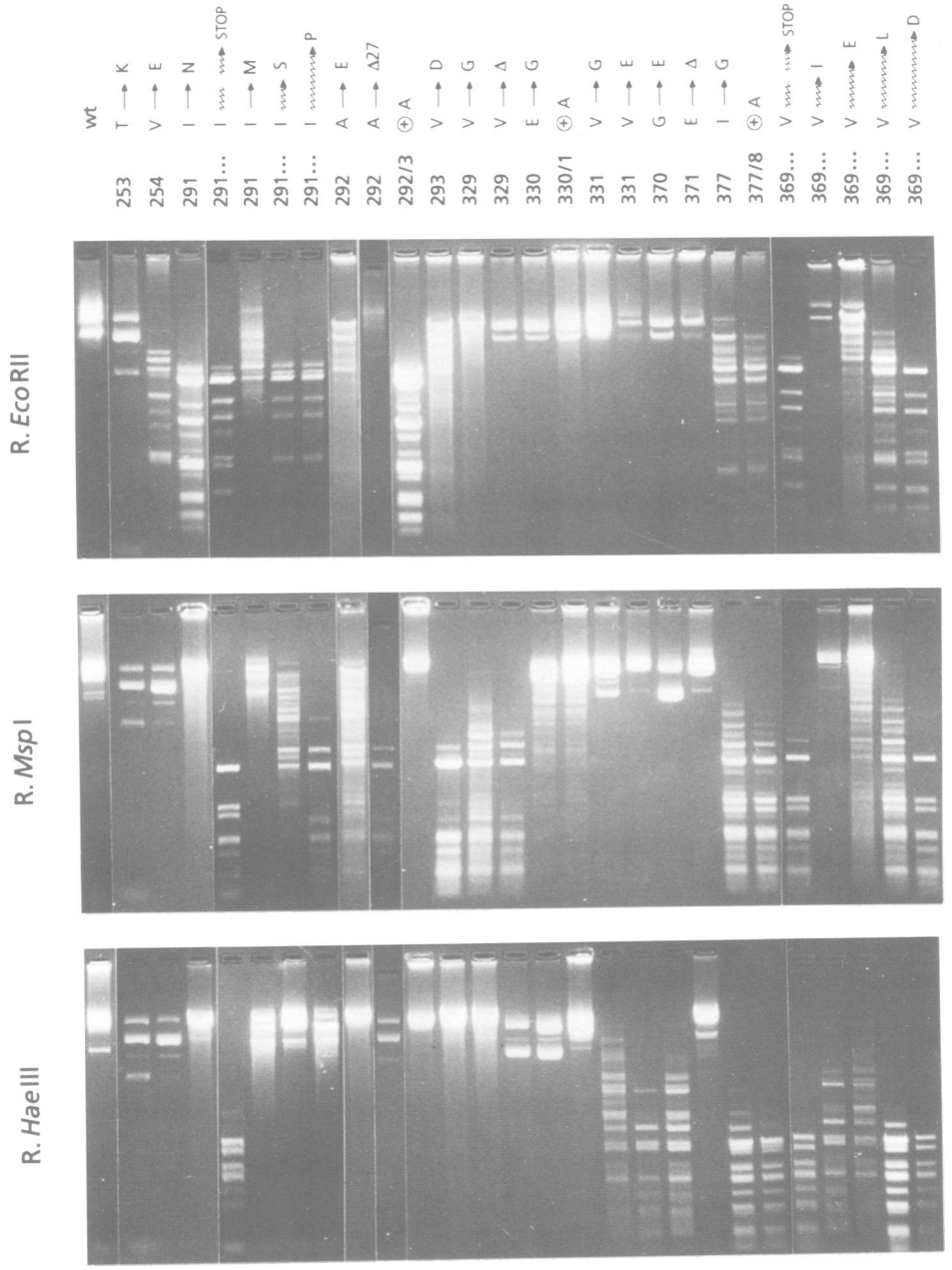

Fig. 2. Gel electrophoresis patterns of selected plasmids with various alleles of the M.SPRI gene following degradation by R.*Eco*RII, R.*Msp*I and R.*Hae*III. Individual tracks are designated by the coordinate at which a single amino acid change has occurred or where multiple amino acid exchanges (deletions, missense sequences) initiated (see Figure 1B). Arrows describe single amino acid exchanges or deletions. Wavy lines show the beginning and end of missense sequences introduced by single or double frameshift mutations (see Figure 1B).

of V254, the most C-terminal ones are at coordinate 291. Mutations of A292 were either class 0 or produced a phenotype (A292→E) which affected the activity of TRD E, but also of the neighboring TRD M. An insertion of a single amino acid, A, between A292 and V293 produced a mutant enzyme which caused inactivation of TRD E but left TRD M virtually unaffected. Since the mutation V293→D generated a class II_M phenotype, we assign the C-terminal end of TRD E to A292, with a total of 39 amino acids making up this TRD and, by inference, the inactive pseudodomains (E) of M.H2I and M.ρ11$_S$I (Lange *et al.*, 1991a). This assignment is supported by a TRD transfer experiment (Figure 3B) where the transfer of a PCR-amplified DNA equivalent of the sequence V254 to A292 was accompanied by a gain of E activity.

TRD M. A class II_M point mutation, V293→D, identified the N-terminal amino acid of TRD M. The assignment of E330 as the C-terminal amino acid follows from the class II_M phenotype of the E330→G exchange, and the identification of class II_H mutations at V331, representing the N-terminal amino acid of the neighboring TRD H. TRD M hence has 38 amino acids. This size determination was supported by a TRD transfer experiment (Figure 3B), where a fragment extending between V293 and V329 gave M activity. In this construct, the vector-derived N

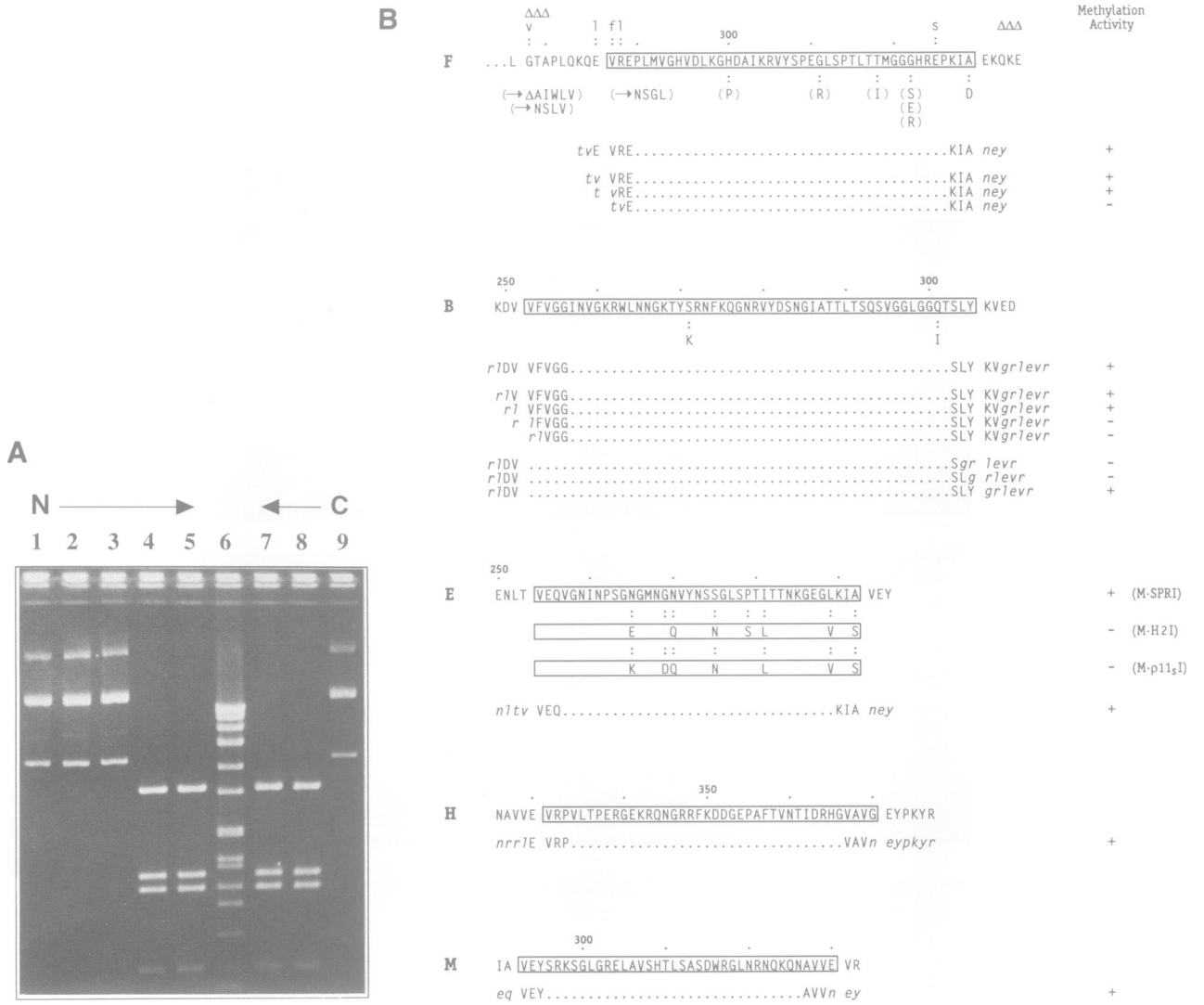

Fig. 3. Confines of TRDs and their derivatives following TRD singularization. (**A**) The DNA of the putative TRD B of M.H2I and derivatives in which consecutive N-terminal or C-terminal amino acids were systematically excluded (see B, line B) was introduced into a plasmid with the inactive C5-MTase M.SPRMUT 5 (Walter *et al.*, 1992). The degradation patterns following digestion with R.*Bsp*1286I of the plasmids containing the various TRDs, as shown in (B) line B, are presented. The first track shows the pattern with the largest fragment transferred. Tracks 2–5 have TRDs with decreasing numbers of amino acids at the N-terminus. Tracks 9–7 show the three mutant enzymes, which have decreasing numbers of amino acids on the C-terminal side. Track 6 contains *Eco*RI-degraded SPP1 DNA as a molecular weight reference (Ratcliff *et al.*, 1979). The N-terminal confine of TRD F was established by a similar approach. (**B**) Schematic representation of TRDs F, B, E, H and M (boxed sequences, top line) in their normal context. Lines below show the methylation activities of the inactive C5-MTase-containing plasmids into which the PCR-generated TRD derivatives shown (upper case letters) had been introduced. Amino acid symbols in italics represent vector sequences. Dots (....) indicate amino acids identical to those of the active TRDs. Mutations in TRDs F, B and E are shown, following the convention of Figure 1B. The description of TRD E includes the amino acid exchanges in which the inactive pseudodomains (E) of M.H2I and M.ρ11$_S$I differ from domain E of M.SPRI. + and – indicate the presence or absence of methylation activity.

apparently substitutes for E330. Similarly to an insertion of an A between TRDs E and M, the insertion of an A between E330 and V331 only affected the TRD M immediately upstream of the insert (Figure 2B)

TRD H. Following the identification of V331 as the N-terminus of TRD H, the analysis of class II$_H$ mutations further downstream assigns the C-terminal amino acid to G370 (Figures 2 and 3). Mutations introduced between coordinates 371 and 376 were neutral. Mutations affecting amino acids beyond 376 were class I mutations. Thus, the TRD H of M.SPRI, and also of the other enzymes studied here, has 39 amino acids (Figure 4). The confines of TRD H, as defined, are compatible with singularization of the TRD (Figure 3B).

TRD F. From previous work (Wilke *et al.*, 1988), the N-terminus of TRD F of M.ϕ3TI (Figure 3B) was anticipated to be N-terminal to a four amino acid insertion (NSGL) between R287 and E288, which gave a class II$_F$ phenotype. Five new mutations further upstream, which included a deletion of three amino acids (T278 to P280), were neutral. Since such mutagenesis did not provide the precise location of the N-terminal amino acid of this TRD, a domain transfer experiment using staggered primers for TRD amplification by PCR (Figure 3B) was performed.

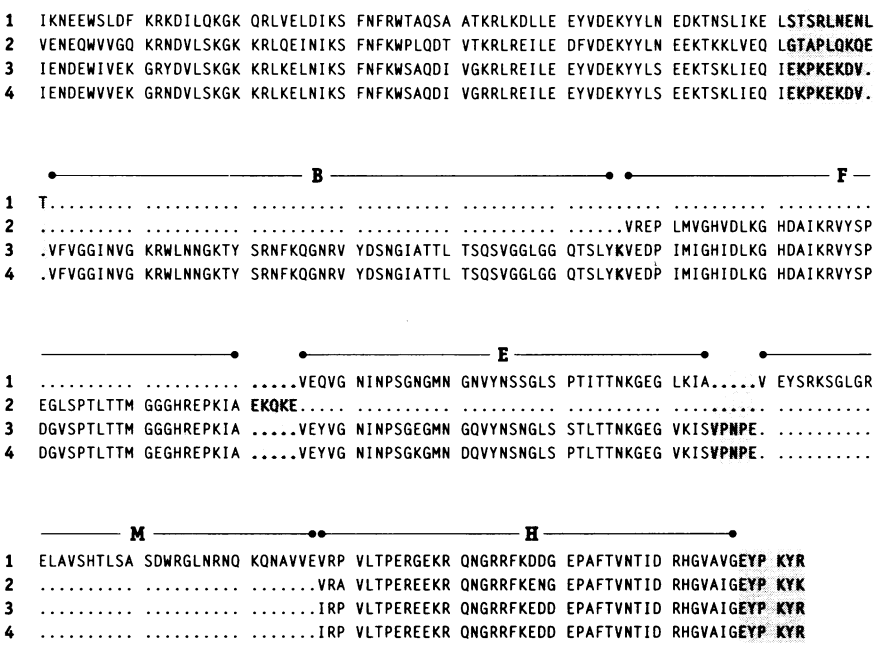

Fig. 4. Alignment of the variable regions of M.SPRI (1), M.ϕ3TI (2), M.H2I (3) and M.ρ11$_S$I (4). The alignment extends between amino acids 173 and 376 of M.SPRI and the equivalent amino acids of the other enzymes. Gaps (. . .) are introduced within the sequences to allow alignment of TRDs or pseudodomains with equal specificity. Confines of the TRDs and 'linker' sequences are marked.

This identified V286 as the N-terminal amino acid of TRD F. Using the same approach, A329 was identified as the C-terminal amino acid of TRD F. In line with this assignment, the mutation A329→D gave a class II$_F$ phenotype. Hence, TRD F and, by inference, the pseudodomain (F) of M.ρ11$_S$I have 44 amino acids.

TRD B. The confines of TRD B of M.H2I and M.ρ11$_S$I were determined by TRD transfer experiments following the synthesis of the TRD B DNA equivalent with staggered primers. The data (Figure 3A and B) identify V252 as the N- and Y305 as the C-terminal amino acid of this TRD. The two class II$_B$ point mutations at coordinates 272 and 302, identified by Rausch (1989), fall within this region. Hence TRD B has 54 amino acids.

The arrangement of TRDs

In Figure 4, we compare the variable regions of all MTases studied here with an alignment of all TRDs with identical specificity. Within the variable region, the ensembles of all TRDs are bracketed by different sequences of 8–10 amino acids at their N-terminal side and by essentially identical hexapeptides at their C-terminal end. With the exception of two insertions at the N-terminus of this region in M.ϕ3TI, which were class I mutations (Figure 3B), mutations within either set of these sequences have not produced novel phenotypes. These regions are hence dispensable for the activity of the TRDs and the MTases generally, and may represent 'linker' sequences.

The distribution of phenotypes of the mutations studied here and the confines of TRDs determined in the transfer of singularized TRDs confirm our previous assumption that the TRDs of multispecific MTases represent contiguous domains. The identification in each TRD (or pseudo TRD) of its N- and C-terminal amino acid and their location (Figure 4) show that none of the neighboring TRDs overlap. In relation to each other, they may either abut with each other, e.g. TRDs E–M–H of M.SPRI or B–F–E of M.H2I or M.ρ11$_S$I, or they may be separated by pentapeptides (VPNPE), as observed between E and H in M.H2I or M.ρ11$_S$I or between F and H in M.ϕ3TI (EKQKE). Dispensability of at least part of this sequence for TRD F activity was demonstrated. A single K separates TRDs B and F of M.H2I and M.ρ11$_S$I.

Mutations affecting more than one amino acid at the termini of TRDs support the conclusions concerning the domain order. Two missense sequences in M.SPRI spanning the junction of TRDs E and M between coordinates 290 and 295 and 290 and 297, respectively, affect both TRD E and TRD M activity (Figures 1B and 2). At the C-terminal end of TRD H of M.SPRI, we have introduced short segments of missense amino acids between A368 and various positions further C-terminal by pairs of oppositely oriented frameshift mutations. Among four mutations of this kind (Figures 1B and 2), the one providing the shortest run of missense amino acids was a class II$_H$ mutation. The two longest ones, extending into the conserved region were, as expected, class I mutations. They differed, however, from each other in that the shorter mutation retained some methylation specificity for all three targets (Figure 2). The mutation terminating with E376 provided sensitivity to R.*Hae*III digestion, but also partial sensitivity to degradation by the two other restriction enzymes. These data indicate that amino acids beyond the C-terminus of the ensemble of TRDs, although not necessary for individual TRD activity, may very well modulate the overall methylation potential of these enzymes.

The region preceding the TRDs

As can be seen from the alignment of the variable regions of the four multispecific C5-MTases discussed here (Figure 4), 71 amino acids preceding the N-terminal linker

sequences are very similar among these enzymes. Previous mutagenesis studies had demonstrated (Wilke *et al.*, 1988) that, both in M.SPRI and M.ϕ3TI, short insertions at various locations of this region had produced class I phenotypes, pointing to the need for preservation of this region for enzyme activity. We have analyzed this requirement by first introducing neutral mutations providing several new *Bgl*I sites into the region between E197 and E250, which in turn facilitated the construction of a number of duplications, deletions, insertions and inversions, whose structure and phenotypes are summarized in Figure 1C. From these data, it is apparent that: (i) if a mutated phenotype is observed, it is mostly of class I; in some mutations, partial sensitivity to degradation by R.*Eco*RII was observed; (ii) a duplication of the region between coordinates 197 and 242 (mutation 2) does not destroy enzyme activity; (iii) an inversion of 15 amino acids between coordinates 218 and 233 leading to incorporation of 15 missense amino acids (mutation 3) is mutagenic; (iv) deletion of the regions between 197 and 242, 197 and 218, 218 and 242 and 218 and 233 is detrimental to enzyme activity; (v) the enzymatic activity of deletions cannot be restored by partially compensating the loss of sequences by reintroducing deleted sequences at other locations; and (vi) deletions and insertions beyond coordinate 242 are tolerated in terms of enzyme activity.

Discussion

Comparison of the amino acid sequences of C5-MTases has led to the distinction of amino acid domains which are highly conserved among these enzymes and of an extended variable domain (Noyer-Weidner and Trautner, 1993; Kumar *et al.*, 1994). This variable region, previously identified as being the essential component of target recognition by multispecific C5-MTases, was shown here to contain three sequence elements: (i) a region of 71 amino acids at the N-terminus of the variable region, which is very similar among the multispecific MTases; (ii) the mostly dissimilar TRDs; and (iii) minor 'linker' sequences.

The function of the 71 amino acids at the N-terminus of the variable region in DNA methylation is not obvious. In the monospecific C5-MTase M.*Hha*I, the equivalent region of some 60 amino acids represents part of the large catalytic domain and the connector region to the small domain involved in target recognition (Klimasauskas *et al.*, 1994). While the sequence of this region is variable amongst the monospecific C5-MTases, it is highly conserved among the multispecific MTases. A number of extended mutations, which we have generated in the M.SPRI gene within this region, have led mostly to the generation of class I phenotypes. Only a large duplication or smaller deletions towards the (dispensable) linker region preceding TRD E have not caused a change in phenotype. A class I phenotype was also the consequence of the conversion of the sequence of 15 amino acids to a missense configuration by inverting the corresponding sense DNA fragment of the MTase gene. The class I phenotype of deletions could not be reverted by partially compensating the loss of genetic material by introduction of other DNAs at other locations. These data suggest that the 71 N-terminal acids do not represent an inert 'spacer' region, but that their composition and sequence are crucial for the activity of the multispecific C5-MTases, irrespective of which of the multiple targets is methylated. It is conceivable that the high degree of conservation of this region amongst these enzymes reflects such a stringent functional/structural requirement.

The ensemble of TRDs represents the second major sequence component of the variable region. The lengths of the individual TRDs are between 38 (TRD M) and 54 (TRD B) amino acids. No correlation between the lengths of the individual targets and the sizes of the corresponding TRDs can be found from the data presented. TRDs with identical specificities from different multispecific, but not monospecific, enzymes show remarkable sequence similarity and have the same length. Minimal amino acid differences between such TRDs point to a common origin of such TRDs followed by a different evolutionary history of the enzymes concerned. Together with several neutral mutations which we have generated during these studies, these data also indicate some flexibility in the usage of amino acids of isospecific TRDs. Other amino acid exchanges are not tolerated for TRD activity. These include, in addition to those generated by site-directed mutagenesis, the natural exchanges, which have led to the generation of inactive pseudodomains of TRDs E and F of M.H2I and M.ρ11$_S$I, described previously (Lange *et al.*, 1991a). Comparing TRDs with different specificities has provided an extension and a refinement of our previous comparisons (Lauster *et al.*, 1989) of TRD sequences. As had been realized previously, all TRDs contain a centrally located dipeptide T (L, I, V) which is also ubiquitously found in the monospecific C5-MTases (Lauster *et al.*, 1989). They also share another common dipeptide, G (L, V, H), in some cases in a G-rich region, six or seven (TRD F) amino acids downstream of the TL motif. In analogy to the defined TRD recognition mechanisms of M.*Hha*I (Klimasauskas *et al.*, 1994) and M.*Hae*III (Reinisch *et al.*, 1995), these amino acids would fall into the C-terminal part of the bipartite set of target-recognizing amino acids. A further consensus element of these TRDs is an N-terminal hydrophobic amino acid, I in the TRDs H of M.H2I and M.ρ11$_S$I, and V in all others. Furthermore, TRDs E and F have the C-terminal tripeptides KIA or KIS. The significance of these consensus amino acids must await structural elucidation of multispecific MTases interacting with their target substrates.

Comparing the TRDs E, F and B, which have the greatest sequence similarity, we realized that the characteristics of a TRD are determined by its particular amino acid composition, the number of amino acids between conserved amino acid motifs and the termini of TRDs. We have investigated the contribution of individual amino acids and of their spacing to the specificity of a TRD using experiments involving enzymes with chimeric TRDs. These experiments indeed have shown that both parameters contribute to the specificity [Lange, 1993; see Lange *et al.*, 1996 (accompanying paper)].

TRDs are fully active following singularization, defining them as autonomous functional elements of multispecific enzymes. In their natural context with other TRDs, however, some mutations of amino acids, located outside the TRD itself but in its vicinity, have been found to modulate the activity of the TRD (Figure 2) in cases where TRDs

abut with each other. As was observed with the insertion of A residues C-terminal to TRDs E and M, such insertions had a prominent effect on the TRD preceding the insert. Again structural analyses with multispecific MTases must be performed to understand the molecular basis of this modulation. DNA target recognition by M.*Hha*I (Klimasauskas *et al.*, 1994) and by M.*Hae*III (Reinisch *et al.*, 1995) has been shown to involve several amino acids, most of which were heavily clustered within some 35 amino acids of the variable region, including the T of the T (L, I, V) motif. These amino acids are clustered in two separate regions which interact with the bases of the double-stranded DNA target and with selected phosphate residues of the sugar phosphate backbone within and beyond the target sequence. It is plausible to assume that target recognition by the multispecific MTases follows the same principle. We cannot, however, assign structural parameters to the sequences identified here as TRDs. Their similarity in size to the cluster of amino acids in *Hha*I and *Hae*III target recognition, however, and the presence of the T (L, I, V) motif and other conserved amino acids would easily accommodate a bipartite recognition mechanism also in multispecific MTases. Genetic evidence for such recognition is provided in the accompanying paper (Lange *et al.*, 1996).

Within the variable region, the ensemble of the TRDs of each enzyme is bracketed by short regions which are dispensable for enzymatic activity. At the C-terminus this is invariably the hexapeptide EYPKYR/(K), which is conserved in all multispecific MTases. At the N-terminus either an octapeptide EKPKEKDV (M.H2I and M.ρ11$_S$I), a nonapeptide GTAPLQKQE (M.ϕ3TI) or the decapeptide STSRLNENLT (M.SPRI) serve as such structures. The role of these sequences and of those separating TRDs in the activity of MTases is not obvious in the absence of structural information.

Materials and methods

MTase genes

The genes of the multispecific MTases (Table I) in their wild-type and mutant alleles were present in *E.coli* plasmids pBR328 (Soberon *et al.*, 1980), and were maintained and replicated in the methylation-tolerant *E.coli* strains GM1499 or GM271 (Raleigh *et al.*, 1988; Krüger *et al.*, 1992; Marinus, unpublished). The nucleotide and derived amino acid sequences of MTase genes and the amino acid coordinates of these enzymes used here are described in the following references: M.SPRI and M.ϕ3TI: Tran-Betcke *et al.* (1986); M.H2I: Lange *et al.* (1991b); M.ρ11$_S$I: Behrens *et al.* (1987).

Mutations

Point mutations, deletions, insertions, frameshift mutations and their intragenic suppressors were generated by site-directed mutagenesis as described previously (Wilke *et al.*, 1988), or by recombinant DNA technology, when closely neighboring restriction sites, either natural or introduced by mutagenesis, could be used to generate deletions, inversions, insertions or duplications. Some relevant mutations of M.SPRI and M.ϕ3TI, described in our previous work (Wilke *et al.*, 1988), were included in these analyses. Some mutations, kindly provided by C.Lange, arose in connection with an unrelated project, which involved amplification of the TRD H of M.SPRI by PCR technology. Mutations of TRD B in M.ρ11$_S$I were described by Rausch (1989). All induced mutational changes, including those whose generation was accompanied by the loss or gain of a restriction site, were verified by DNA sequencing using the Sanger dideoxy method (Sanger, 1977). All mutational amino acid changes used in these analyses are shown in Figures 1 and 3.

Domain exchange experiments

The technique followed here was essentially that described previously by Walter *et al.* (1992). To determine the terminal amino acids of TRDs F and B, DNA fragments coding for TRDs or sets of derivatives of TRDs systematically shortened at their putative N- or C-terminal ends were amplified using PCR technology with appropriate primers. These were then introduced by ligation into the *Hpa*I site of M.SPRMUT5 (Walter *et al.*, 1992) or into the *Aat*II site of M.SPRMUT1 (Walter *et al.*, 1992). The sense orientation of the PCR-generated inserts was verified in all constructs.

Identification of phenotypes

The basis for the determination of phenotypes of mutations in the MTase gene was the susceptibility of *in vivo* methylated plasmids, carrying mutant alleles of the MTase gene, to restriction by the restriction endonucleases R.*Eco*RII, R.*Msp*I, R.*Hae*III (M.SPRI); R.*Fnu*4HI, R.*Hae*III (M.ϕ3TI) or R.*Fnu*4HI, R.*Bsp*1286, R.*Eco*RII, R.*Hae*III (M.ρ11$_S$I, M.H2I). Plasmids were prepared either from lysates of *E.coli* cells according to the method of Birnboim and Doly (1979) or after further purification following conventional methods. Restriction enzymes were used in excess, under conditions recommended by the suppliers. Class II mutations are characterized by susceptibility of the mutant plasmid to only one of the restriction enzymes tested and resistance of the plasmid to the activity of the other restriction enzymes used. This latter analysis also served as an internal control, indicating synthesis of wild-type levels of mutant MTase. Some class II mutations caused susceptibility of the corresponding plasmid to more than one of the enzymes tested. The gain of a novel methylation specificity following insertion of PCR-amplified TRDs into plasmids pSPRMUT5 or pSPRMUT8 was taken as evidence of the addition of a complete TRD in these constructs.

Acknowledgements

Discussions with M.Noyer-Weidner and J.Walter are greatly appreciated, as is computer assistance given by R.Lurz and H.Zabin. This work was partially supported by EEC Grant SC1*-CT90-0472 (TSTS) and by the DFG within Sonderforschungsbereich 344.

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**, 3543–3549.

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.

Birnboim,H.C. and Doly,J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.*, **7**, 1513–1523.

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.

Chen,L., MacMillan,A.M., Chang,W., Ezaz-Nikpay,K., Lane,W.S. and Verdine,G.L. (1991) Direct identification of the active-site nucleophile in a DNA (cytosine-5)-methyltransferase. *Biochemistry*, **30**, 11018–11025.

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.

Ingrosso,D., Fowler,A.V., Bleibaum,J. and Clarke,S. (1989) Sequence of the D-aspartyl/L-isoaspartyl protein methyltransferase from humam erythrocytes. *J. Biol. Chem.*, **264**, 20131–20139.

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.

Krüger,T., Grund,C., Wild,C. and Noyer-Weidner,M. (1992) Characterization of the mcr BC region of *Escherichia coli* K-12 wild-type and mutant strains. *Gene*, **114**, 1–12.

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

Lange,C. (1993) Charakterisierung potentieller Sequenzerkennungsdomänen in multispezifischen DNA-Methyltransferasen. Dissertation, FU Berlin.

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

Lange,C., Noyer-Weidner,M., Trautner,T.A., Weiner,M. and Zahler,S.A. (1991b) M.H2I, a multispecific 5C-DNA methyltransferase encoded by *Bacillus amyloliquefaciens* phage H2. *Gene*, **100**, 213–218.

Lange,C., Wild,C. and Trautner,T.A. (1996) Identification of a subdomain within DNA-(cytosine-C5)-methyltransferases responsible for the recognition of the 5′ part of their DNA target. *EMBO J.*, **15**, 1443–1450.

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.

Noyer-Weidner,M. and Trautner,T.A. (1993) Methylation of DNA in prokaryotes. In Jost,J.P. and Saluz,H.P. (eds), *DNA Methylation, Molecular Biology and Biological Significance*. Birkhäuser Verlag, Basel, Switzerland, pp. 39–108.

Noyer-Weidner,M., Walter,J., Terschüren,P.-A., Chai,S. and Trautner,T.A. (1994) M.phi3TII: 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.

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

Raleigh,E.A., Murray,N.E., Revel,H., Blumenthal,R.M., Westaway,D., Reith,A.D., Rigby,P.W.J., Elhai,J. and Hanahan,D. (1988) McrA and McrB restriction phenotypes of some *E.coli* strains and implications for gene cloning. *Nucleic Acids Res.*, **16**, 1563–1575.

Ratcliff,S.W., Luh,J., Ganesan,A.T., Behrens,B., Thompson,R., Montenegro,M.A., Morelli,G. and Trautner,T.A. (1979) The genome of *Bacillus subtilis* phage SPP1: the arrangement of restriction endonuclease generated fragments. *Mol. Gen. Genet.*, **168**, 165–172.

Rausch,S. (1989) Mutationen im Gen der $\rho 11_S$-DNA-Methyltransferase, die zum Verlust der *Bsp*1286-Methylierungsaktivität führen. Diploma Thesis, FU Berlin.

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.

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

Soberon,X., Covarrubias,L. and Bolivar,F. (1980) Construction and characterization of new cloning vehicles. IV. Deletion derivatives of pBR322 and pBR325. *Gene*, **9**, 287–305.

Som,S., Bhagwat,A.S. and Friedman,S. (1987) Nucleotide sequence and expression of the gene encoding the *Eco*RII modification enzyme. *Nucleic Acids Res.*, **15**, 313–332.

Tran-Betcke,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., Balganesh,T.S. and Pawlek,B. (1988) Chimeric multispecific DNA methyltransferases with novel combinations of target recognition. *Nucleic Acids Res.*, **16**, 6649–6657.

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**, 505–627.

Wu,J.C. and Santi,D.V. (1987) Kinetic and catalytic mechanism of *Hha*I methyltransferase. *J. Biol. Chem.*, **262**, 4778–4786.

Received on July 3, 1995; revised on November 20, 1995