M.ϕ3TII: a new monospecific DNA (cytosine-C5) methyltransferase with pronounced amino acid sequence similarity to a family of adenine-N6-DNA-methyltransferases

Mario Noyer-Weidner¹, Jörn Walter, Peter-Andreas Terschüren, Sunghee Chai and Thomas A. Trautner*

Max-Planck-Institut für Molekulare Genetik, Ihnestr. 73, D-14195 Berlin and ¹Walter de Gruyter & Co., Scientific Publishers, Genthiner Str. 13, D-10728 Berlin, Germany

Received July 26, 1994; Revised and Accepted September 5, 1994

EMBL accession no. X80202

ABSTRACT

The temperate *B. subtilis* phages ϕ3T and ϕ11_s code, in addition to the multispecific DNA (cytosine-C5) methyltransferases (C5-MTases) M.ϕ3TI and M.ϕ11_sI, which were previously characterized, for the identical monospecific C5-MTases M.ϕ3TII and M.ϕ11_sII. These enzymes modify the C of TCGA sites, a novel target specificity among C5-MTases. The primary sequence of M.ϕ3TII (326 amino acids) shows all conserved motifs typical of the building plan of C5-MTases. The degree of relatedness between M.ϕ3TII and all other mono- or multispecific C5-MTases ranges from 30 – 40% amino acid identity. Particularly M.ϕ3TII does not show pronounced similarity to M.ϕ3TI indicating that both MTase genes were not generated from one another but were acquired independently by the phage. The amino terminal part of the M.ϕ3TII (preceding the variable region 'V'), which predominantly constitutes the catalytic domain of the enzyme, exhibits pronounced sequence similarity to the amino termini of a family of A-N6-MTases, which — like M. *TaqI* — recognize the general sequence T_NNA. This suggests that recently described similarities in the general three dimensional organization of C5- and A-N6-MTases imply divergent evolution of these enzymes originating from a common molecular ancestor.

INTRODUCTION

Type II-DNA methyltransferases (MTases) transfer methyl groups from S-adenosylmethionine (AdoMet) to adenine or cytosine residues within specific DNA sequences (for a review see (1)). Based on the modification type generated, typeII-MTases can be divided into C5-MTases, C-N4-MTases and A-N6-MTases. C5-MTases methylate the cyclic carbon in position 5 of cytosines, C-N4- and A-N6-MTases the exocyclic nitrogen in position 4 or 6 of cytosine or adenine residues, respectively. The difference in the position and atom to be methylated in C5-

or A-N6- and C-N4-methylation correlates with differences in amino acid (aa) sequences of C5-MTases and A-N6/C-N4-MTases.

The primary structures of C5-MTases share a common building plan: Several conserved elements (CEs) forming the amino and carboxy terminal enzyme segments are separated by a variable region 'V', which includes the target recognizing domain(s) TRD(s) (1, 2). The analysis of the three dimensional structure of a typical monospecific C5-MTase, M. *HhaI*, either free or complexed with DNA, revealed that the enzyme consists of two globular domains (3). One, consisting of most of the conserved elements (CEs) of C5-MTases, is essentially involved in the catalytic steps of the reaction. The other domain, formed predominantly by the carboxy terminal part of the variable region 'V', is responsible for target recognition.

In contrast to the uniform building plan of C5-MTases, A-N6- and C-N4-MTases are much more heterogeneous in the organization of their primary structures. Still, the recent determination of the three dimensional structure of the A-N6-MTase M. *TaqI* complexed with AdoMet (4) revealed suggestive similarity to M. *HhaI* in the general organization of this enzyme. Also M. *TaqI* consists of two globular domains, in which the amino terminal domain comprises the AdoMet binding pocket and probably forms the catalytic subunit. The C terminal domain is less structured and its function is as yet unclear. A cleft formed between these two domains most likely serves to accommodate the DNA substrate.

In this communication we present the sequences and properties of new C5-MTases, M.ϕ3TII and M.ϕ11_sII, encoded by the *Bacillus subtilis* phages ϕ3T and ϕ11_s, respectively. Both structurally identical MTases methylate the cytosine of the DNA target sequence TCGA. While this specificity is novel among C5-MTases, A-N6-MTases (M. *TaqI*, M. *TthHB8I*) which modify the adenine of this target sequence have been described before. The identification of these new C5-MTases has motivated us to compare their amino acid sequences with those of other C5-MTases and A-N6-MTases including the isospecific enzymes

*To whom correspondence should be addressed

M. *TaqI* and M. *ThiHB8I*. These comparisons provided evidence for an evolutionary link between C5- and A-N6-MTases.

MATERIALS AND METHODS

Bacterial strains, phages and media

E. coli NM675 (e14⁻(mcrA⁻), Δ (*mrr-hsdRMS-mcrBC*)202, *gal*, *met*; (5) served as host to establish and propagate MTase encoding plasmids. JM103 (*endA*, *hsdR*-4, *sbcB*-15, *strA*, *supE*, *thi*-1, Δ (*lac-pro*), (F', *traD*-36, *proAB*, *lacI*^q, ZDM15) (6) was used in cloning experiments employing M13mp18/19 vectors. *E. coli* strains were grown in Luria broth supplemented with 50 μ g ampicillin/ml, if required.

Enzymes

Restriction endonucleases, T4 DNA ligase, T4 polynucleotide kinase, Klenow fragment of DNA polymerase I, and calf intestinal phosphatase were purchased from Boehringer, Mannheim, Germany, and used as recommended by the manufacturer.

Molecular cloning procedures

Preparation of plasmid and chromosomal DNA, DNA restriction, isolation of restriction fragments, DNA ligation and transformation of *E. coli* cells followed standard procedures (7).

Construction of plasmids

MTase encoding recombinant plasmids were all derived by inserting ϕ 3T or ϕ 11_s DNA fragments into the *Hind*III site of pBR328 (8). Construction of plasmids pBA680 and pBA681, serving in the determination of the M. ϕ 3TII specificity, was based on pMS119EH, a derivative of pJF118EH (9) with an altered multicloning site (obtained from E.Lanka). pBA680 was generated by inserting a 36 bp oligonucleotide sequence, containing TCGA sites in various sequence contexts, into the unique *Bam*HI site of pMS119E/H. pBA681 was constructed by integration of a 1.2 kb *Ssp*I segment, carrying the M. ϕ 3TII gene, into the *Sma*I site of pBA680.

Nucleotide sequencing

The nucleotide sequence of the M. ϕ 3TII gene was determined for both strands by the dideoxy chain termination method (10) following cloning of appropriate pBA67 derived restriction fragments (see text) in M13mp18 and M13mp19 (11). The sequence information obtained was combined with previously established sequencing data (12). The site of methylation in M. ϕ 3TII modified DNA was determined by the chemical sequencing method (13) of the following DNA fragments: part of a 539bp *Eco*RV–*Hind*III fragment comprising pMS119EH vector sequences, a 36bp oligonucleotide insertion into the *Bam*HI site of pBA680 and pBA681 and a 117 bp *Mlu*I/*Hind*III segment of pBA681, representing a part of the M. ϕ 3TII gene (nucleotide positions 651 to 768, this paper).

Primer extension analyses

Primer extension experiments were carried out with m-RNA isolated by a procedure described before (14) from ϕ 3T lysogenic cells of *B. subtilis* strain YB886 (15) 90 or 120 min after phage induction with mitomycin. Synthetic oligonucleotides, whose sequences corresponded to the sense strand of the M. ϕ 3TII gene within nucleotide positions 361 to 383, were 5' end labeled with

(γ -³²P) ATP and hybridized with the RNA probes under appropriate conditions. The primer extension reactions were carried out for 1 h at 52°C using 4 units of AMV reverse transcriptase (Promega, USA). The products were analyzed on 8% PAA sequencing gels.

Amino acid sequence comparisons

Computer aided sequence comparisons were performed either on a VAX using the UWGCG software (16) or a PC (IBM-compatible) using the program 'Motif' (17). Amino acids in brackets were considered as related: (E,D), (Q,N), (I,L,V), (H,R,K), (F,W,Y), (S,T). Published sequences of the following MTases were included in the comparisons: M.*Accl* (18), M.*Bsu*BI (19), M.*Cvi*BIII (20), M.*Cvi*RI (21), M.*Eco*57I (22), M.*Hinc*II (23), M.*Pae*RV II (24), M. ϕ 3TI (12), M.*Pst*I (25), M.*Taq*I (26), M.*ThiHB8I* (26).

RESULTS AND DISCUSSION

Cloning of the M. ϕ 3TII gene and determination of the M. ϕ 3TII specificity

We previously reported the cloning and sequence determination of genes encoding the multispecific C5-MTases, M. ϕ 3TI and M. ϕ 11_sI, of *B. subtilis* phages ϕ 3T and ϕ 11_s (27, 28). In the 5' region of the M. ϕ 3TI or M. ϕ 11_sI genes the sequenced DNA segments contained what seemed to be the 3' part of another C5-MTase gene, identical in both phage genomes (29, Figure 1). Extending the sequenced ϕ 3T and ϕ 11_s DNA segments in

Figure 1. Restriction maps of pBR328 derivatives carrying ϕ 3T or ϕ 11_s MTase genes and modification activities specified by these constructs. All plasmids were selected from a pBR328 based gene library containing partially *Hind*III digested ϕ 3T or ϕ 11_s DNA by resistance to *R. Hae*III restriction, a consequence of M. ϕ 3TI or M. ϕ 11_sI gene expression (27). Plasmids pBN16, and pBN52 had already been described before (27, 28). Segments of ϕ 3T DNA (in pBN16 and pBA66/68) or ϕ 11_s DNA (in pBN52 and pBA70) are represented by double lines, parts of flanking pBR328 sequences by single lines. The gap indicates an engineered deletion. The position and orientation of complete or incomplete MTase genes are shown by black (M. ϕ 3TI and M. ϕ 11_sI genes) or hatched (M. ϕ 3TII and M. ϕ 11_sI genes) arrows. Restriction sites shown are: B (*Bam*HI), Hi (*Hind*III) and Hp (*Hpa*I). The right panel indicates resistance or sensitivity of plasmid DNA to restriction endonucleases. Resistance to *R. Hae*III restriction reflects M. ϕ 3TI or M. ϕ 11_sI gene expression, resistance to *Sal*I restriction indicates M. ϕ 3TII or M. ϕ 11_sII gene expression (see text).

previously generated recombinant plasmids (pBN16 and pBN52, Figure 1) by 0.65 kb *Hind*III fragments representing DNA at the 5' flank of the *M.φ3TI* and *M.φ11_sI* genes (plasmids pBA66 and pBA70, Figure 1) indeed revealed the existence of additional phage MTase genes, whose products were termed *M.φ3TII* and *M.φ11_sII*. Maxam Gilbert sequencing of *M.φ3TII* modified DNA revealed that the enzyme methylates the C residue in TCGA targets (Figure 2). This activity is also expressed in the absence of *M.φ3TI* (plasmid pBA68, Figure 1). A C5-MTase with this specificity has not been described before. We did not detect any additional specificity suggesting that *M.φ3TII* and *M.φ11_sII* are monospecific MTases (see below).

C methylation within TCGA sites of plasmids carrying the *M.φ3TII* and *M.φ11_sII* genes or of chromosomal DNA of *E. coli* cells carrying such plasmids did not inhibit cleavage by *R. TaqI* but provided resistance to several restriction endonucleases (ENases) like *R. ClaI* (ATCGAG), *R. Csp45I* (TTCGAA), *R. SalI* (GTCGAC) and *R. XhoI* (CTCGAG). In *B. subtilis*, however, a hierarchy in target recognition, determined by bases surrounding the target, became apparent when the enzymes were expressed after induction of the temperate *B. subtilis* phages. DNAs of the phages *φ3T* or *φ11_s*, grown on hosts deficient in the PyTCGAPu specific *Bsu*MI R/M system (30), were only resistant to *R. ClaI* and *R. XhoI* cleavage but not to *R. SalI* or *R. Csp45I* restriction. The specificity of *M.φ3TII*

Figure 2. Autoradiogram of unmodified and *M.φ3TII* methylated DNAs sequenced according to (13). Parts of the following pMS119E/H derivatives were sequenced (the numbering of nucleotide positions is given according to that described for the pMS119EH precursor pJF118EH (9) or to that described for the *M.φ3TII* gene sequence, EMBL Acc.No. X80202): Lanes 1, 2, 3: pBA680, a pMS119E/H derivative containing a 36 bp oligonucleotide insert at the *Bam*HI site, providing TCGA sites in various sequence contexts; Lanes 4, 5: pBA681, a derivative of pBA680 encoding *M.φ3TII*; Lanes 6, 7, 8: An internal 117 bp *MluI/Hind*III fragment of the *M.φ3TII* gene, isolated from pBA681. Lanes 1, 4, 7 C+T reactions; Lanes 2, 5, 8: C reactions; Lanes 3, 6: G+A reactions. TCGA sites within the sequenced fragments are indicated. One of these sites forms the center of a *SalI* site, shown in brackets. Gaps in the C and C+T tracks, caused by methylation of C-residues, are indicated by arrows.

Figure 3. (A) 5' part of the nucleotide sequence of the *M.φ3TII* gene and deduced amino acid sequence of part of the *M.φ3TII* protein. The *M.φ3TII* encoding ORF starts with ATG at nucleotide positions 286 to 288. A potential Shine-Dalgarno (S/D) sequence preceding the translational starting point is underlined. Transcriptional start sites at nucleotides 225 and 249, as revealed in primer extension analyses (Figure 3B), are indicated by vertical arrows. Regions resembling -10 promoter sequences, located 5' to the transcriptional starting points identified, are boxed. Segments of *M.φ3TII* corresponding to conserved elements (CEs) I to III, as defined by Posfai *et al.* (29), are indicated by underlining and bracketing the constituent amino acids. (B) Primer extension analyses of the *M.φ3TII* and *M.φ3TI* gene transcripts. The reactions were carried out with m-RNA isolated from mitomycin induced *φ3T*-lysogenic YB886 cells using oligonucleotides whose sequence corresponded to the sense strands of the *M.φ3TII* gene (nucleotide positions 361 to 383, this paper) or *M.φ3TI* gene (nucleotide positions 783-806 (12)). Lanes 1, 2, 3, 4, A, C, G and T tracks of a sequencing ladder obtained in sequencing a fragment of the *sus7* gene of *B. subtilis* phage SPP1 (J. Alonso *et al.*, unpublished), cloned into M13mp19, according to Sanger *et al.* (10) using the universal -40 sequencing primer. Lanes 5, 6, extension of m-RNA prepared from YB886(*φ3T*) 90 or 120 min after phage induction using the *M.φ3TII* gene specific primer. Lanes 7, 8, as lanes 5, 6 except for the use of the *M.φ3TI* gene specific primer.

(and M.ϕ11,II) therefore protects the phage genome, at least partially, against host restriction by the PyTCGAPu specific *R.Bsu*MI activity, which is naturally found in the *B.subtilis* Marburg strains (31). Together with the multispecific MTases the monospecific MTases contribute to a repertoire of methylation activities which provide protection against different host restriction systems. Coexistence of multispecific and monospecific MTases counteracting host restriction has also been described for the *B.amyloliquefaciens* phage H2 (32, 33).

Nucleotide sequence and transcription of the M.ϕ3TII gene

In view of the same target specificity and methylation hierarchy of the M.ϕ3TII and M.ϕ11sII MTases and the identity of the 373 nucleotides at the 3' end of their genes (12, 28), we considered M.ϕ3TII and M.ϕ11sII to be identical. Hence the complete nucleotide sequence was only determined for the M.ϕ3TII gene. This was achieved by sequencing the 0.65 kb *Hind*III fragment (Figure 1) containing the unidentified part of the M.ϕ3TII gene and a 402 bp *Eco*RV-*Bgl*II fragment overlapping the join of this fragment with the previously established sequence. An ORF of 978 nucleotides, which included the previously identified 3' end of the putative MTase gene was detected. Figure 3A shows the 5' part of the newly determined sequence. The entire sequence is available under the EMBL

Acc.No X80202. Assuming that the ATG site at the very beginning of the ORF (nucleotides 286–288), which is preceded by an appropriately spaced Shine–Dalgarno (SD) sequence, serves as the translational start codon, the frame encodes a protein of 326 amino acids with a calculated M_r of 36374 Da.

The M.ϕ3TII gene is separated by 87 bp from the identically oriented M.ϕ3TI gene (not shown). A spacer identical in size and conserved in its nucleotide sequence but for 3 bp, separates the M.ϕ11sII and M.ϕ11sI genes, whose orientations are also identical (28).

Primer extension experiments with m-RNA from induced $\phi 3T$ lysogenic *B. subtilis* cells, were carried out with primers corresponding to 5' parts of either $\phi 3T$ MTase gene. Experiments with the M. $\phi 3TII$ specific primer identified two m-RNA species (Figure 3B), whose 5' ends corresponded to transcription initiation sites around nucleotides 225 and 249, respectively (Figure 3A). Both sites are preceded by potential -10 promoter regions with no canonical -35 regions identifiable. At present it cannot be decided whether the two m-RNAs result from the use of alternative transcriptional start sites or from partial truncation of the larger m-RNA species. No reverse transcript was obtained in experiments with the M. $\phi 3TI$ gene specific primer (Figure 3B), suggesting that both $\phi 3T$ MTase genes form an operon with transcription initiating 5' of the M. $\phi 3TII$ gene.

Figure 4. Comparison of the amino acids sequences of M.ϕ3TII and TNNA recognizing A-N6-MTases. Alignments A, B, C and D show the conserved blocks (CBs) I, I', II and III and surrounding amino acids of the TNNA recognizing A-N6-MTases. Alignment E shows a sequence element found to be similar in pairwise comparisons between M.ϕ3TII and *M. taqI* or *M. TthHB8I*. The extension of CBs is indicated by the line below the aligned sequences. Matching segments of M.ϕ3TII are shown above the CB regions. Their coincidence with (parts of) CEs or Region 'V' in M.ϕ3TII is indicated by horizontal lines above the sequence. Numbers at the beginning or end of the sequence elements aligned refer to the positions occupied in the MTase primary sequences by the first and last amino acids shown. Amino acids identical or similar in M.ϕ3TII and the A-N6-MTases are boxed.

Amino acid sequence comparison of *M.φ3TII* with other C5-MTases

The general building plan of C5-MTases (1, 2, 29, 34, 35) is also realized in the case of *M.φ3TII* (Figure 5). Its primary sequence shows the characteristic arrangement of conserved elements CE I to X (29), which define the enzyme's core structure.

Amino acid identities in the C5-MTase core regions, extending from the amino terminus of CE I to the carboxy terminus of CE X with omission of the variable region 'V', showed that the degree of relatedness between *M.φ3TII* and other C5-MTases ranges from some 30 to 40% amino acids identity. No correlation between the degree of similarity and the basic functional subdivision of C5-MTases into mono- and multispecific enzymes or the phylogenetic distances of the species encoding these enzymes was observed. Particularly no exceptional similarity was found between *M.φ3TII* and *M.φ3TI*. This suggests that the two MTase genes, in spite of their genomic neighborhood, were not generated from one another but were acquired independently by the phage.

Amino acid comparison of *M.φ3TII* with A-N6-MTases

Comparing the *M.φ3TII* amino acids sequence with those of unselected C-N4- and A-N6 MTases showed similarities which did not exceed those previously reported by Klimasauskas *et al.* (36). Their analyses and most of the comparisons performed here yielded no evidence for a relatedness of global structures of C5- and C-N4-/A-N6-MTases. Similarities were confined to the short 'catalytic' sequences FXGXG and (G,S)XPCXXXSXX(G,S)/(D,N,S)PP(Y,F) whose order is fixed in C5-MTases but varies in different families of C-N4- and A-N6-MTases (not shown).

Similarities beyond these, however, were detected in the studies reported here, when *M.φ3TII* was compared with the members of the *TaqI*-family of A-N6-MTases. Members of this family recognize DNA targets which contain the general sequence TNNA and are, therefore, identical with or related to the

M.φ3TII DNA target TCGA (see Figure 4). Among the four identically arranged blocks of conserved amino acids (CBI, CBI', CBII and CBIII) contained within the approximately 200 amino terminal amino acids of these enzymes (22, 23, this study; Figure 4 and 5), CB III has been suggested to act in target recognition, as this element most prominently distinguishes these A-N6-MTases from those with other specificities.

The alignment of *M.φ3TII* with A-N6-MTases of the *TaqI*-family are shown in Figure 4 and 5, from which the following conclusions can be drawn:

(1) Sequence elements similar in *M.φ3TII* and the *TaqI*-MTase family are only contained within the amino terminal halves of these MTases and are arranged in identical order in all enzymes.

(2) In *M.φ3TII* three (four) of the elements discussed are part of CEs, i.e. structures generally found in C5-MTases. One segment including part of the region 'V', however, is unique to *M.φ3TII*. This suggests a particularly close relationship between *M.φ3TII* and this group of A-N6-MTases and strongly supports the notion of a general structural and evolutionary relatedness of the catalytic domains of C5-MTases and the A-N6-MTase as observed in case of the three dimensional structures of *M.HhaI* and *M.TaqI*.

(3) Looking at individual matches between CEs of *M.φ3TII* and sequences in the A-N6-MTase group discussed here, the most amino terminal conserved element, CE I of C5-MTases, contains the sequence FXGXG. This motif is degenerate in the TNNA recognizing A-N6-MTase family. It is readily identifiable, though, by its spacing to the highly conserved tripetide motif VLE shared between *M.φ3TII* and the A-N6-MTases discussed here.

The amino acid sequence ALDID of CEII of *M.φ3TII* is the next sequence element with similarity to the motif of CBI' of TNNA MTases. In both the C5-MTases and the A-N6-MTases the distance of these motifs from the carboxy termini of CEI or CBI is highly conserved. In the case of *M.HhaI* and *M.TaqI* the D,E residues of these motifs (residues E-40 in *M.HhaI* and D-71 in *M.TaqI*, respectively) have been shown to be involved in AdoMet binding (3,4).

Figure 5. Schematic representation of the position of regions similar between *M.φ3TII* and TNNA-recognizing A-N6-MTases. The primary sequences of the MTases indicated are represented by bars. Matching regions in *M.φ3TII* and the A-N6-MTases are shown as differently labeled boxes. The position of CEs and CBs in *M.φ3TII* and *M.Eco57I* is indicated above and below the corresponding MTase schemes.

The amino terminus of CEIV contains the characteristic C5-MTase motif (G,S)XPCXXSXX(G,R), the C residue of which initiates methyl group transfer by covalently binding to the cytosine to be methylated (37,38). CBII of the TNNA recognizing A-N6-MTases includes the sequence NPPY, a representative of the motif (D,N,S)PP(Y,F), generally found in A-N6- and C-N4-MTases. In the alignment presented in Figure 4, the Y residue of the NPPY motif is juxtaposed to the C-residue of the C5-MTase motif. This deviates from a previous report of Klimasauskas *et al.* (36) who recognized the general similarity of the motifs discussed here, but juxtaposed the Y residue of the NPPY motif to an amino acids located carboxy terminal to the C-residue in the C5-MTase motif. The direct correspondence of the C- and Y-residues reported here and the identical position of the motifs discussed relative to other sequence elements of C5- and TNNA recognizing A-N6-MTases (Figure 4 and 5) provide additional support for a role of NPPY and related motifs in methyl group transfer by A-N6- and C-N4-MTases. The alignment furthermore shows a strong conservation of a D/E residue in the TNNA recognizing A-N-6 MTase family (D-89 in *M.TaqI*). This residue (D-63 in *M.φ3TII*) is also highly conserved among C5-MTases (2). Similar as discussed for the D/E residues of the motifs CEII and CBI' (see above) also the D/E residues considered here are involved in complexing the AdoMet cofactor (3,4).

A fourth region of sequence similarity involves the amino acids of CEVI of *M.φ3TII*. In this case similarity is, however, limited to *M.TaqI* and *M.TthHB8I* MTases.

(4) Outside of the CEs, a segment of *M.φ3TII* representing part of the 'V' region reveals some similarities to the previously defined element CBIII of all A-N6-MTases described here. In the case of *M.TaqI* and *M.TthHB8I* these similarities extend significantly also into sequences immediately joining the carboxy terminus of CBIII (Figure 4, Alignment D and Figure 5): within the 42 carboxy terminal amino acids (extending from V₁₉₆), which can be contiguously aligned, 18 and 15 amino acids of *M.TaqI* or *M.TthHB8I*, respectively, are identical or similar to the *M.φ3TII* sequence. This 42 amino acids segment of the *M.φ3TII* sequence would include the amino terminal segment of the enzyme's TRD. The similarity, however, is not sufficient to postulate a conserved mechanism of target recognition between *M.φ3TII* and *M.TaqI* and *M.TthHB8I*, respectively. Rather this similarity and that observed with CEVI (Figure 5, Alignment E) suggest a common evolutionary history of these C5- and A-N6-MTases.

It is interesting to note that the carboxy termini of all enzymes discussed here (with exception of the nearly identical *M.TaqI* and *M.TthHB8I* MTases) massively diverge. This holds even for members of the *TaqI*-MTase family itself. It therefore appears that the 42 amino acids segment of *M.φ3TII* and matching sequences in the TNNA-recognizing A-N6-MTases are located at the borderline of a related amino terminal enzyme moiety linked to a divergently evolved (or independently acquired) carboxy terminal segment. The evolutionary concept proposed here, is supported by three observations. Firstly, evidence has been provided that the amino terminal sequences preceding the TRD region of C5-MTases and the carboxy terminal sequences following it represent independent enzyme entities, which must not necessarily be covalently joined to provide MTase activity (39,40). Furthermore, recently C5-MTases of type II have been reported in which the sequence of conserved elements is permuted in such a way that the variable region and the conserved motifs

IX and X represent the amino terminal part of the enzyme, whereas the other conserved elements I to VIII are at its carboxy terminus (J.Bitinaite, pers. Comm.). Finally, recent three dimensional analyses of *M.TaqI* (4) and *M.HhaI* (3, 41) have shown that both A-N6- and C5-MTases consist of two globular domains, which are joined by one (*M.TaqI*) or two (*M.HhaI*) rather flexible polypeptide sequences. In the case of *M.HhaI* this flexible 'hinge' includes the variable part of the enzyme, located some 200 amino acids away from its amino end. In *M.TaqI*, the major globular structure observed in the enzyme is similarly composed of some two hundred amino terminal amino acids, making up the highly conserved elements described before.

The high conservation of amino acids within the carboxy terminal side of the variable region of C5-MTases, represented by CEs IX and X, might reflect the constraints put on these enzymes in which the carboxy terminal end folds back into the large domain defined by CEs I through VIII. In the A-N6-MTases no folding back of the carboxy terminus is observed allowing for a more relaxed evolution of their carboxy termini and hence a greater sequence variation.

ACKNOWLEDGEMENTS

We are most grateful to Hal Zabin for his advice and help with the computational analysis and like to thank W.Saenger for providing us with the data on the *M.TaqI* crystal structure prior to publication. Part of this work was supported by the Deutsche Forschungsgemeinschaft (SFB 344).

REFERENCES

1. Noyer-Weidner, M. and Trautner, T.A. (1992) In Jost, J.P. and Saluz, H.P. (eds.), *DNA Methylation: Molecular Biology and Biological Significance*. Birkhäuser Verlag, Basel-Boston-Berlin, pp. 39–108.
2. Kumar, S., Cheng, X., Klimasauskas, S., Mi, S., Posfai, J., Roberts, R.J. and Wilson, G.G. (1994) *Nucleic Acids Res.* 22, 1–10.
3. Cheng, X., Kumar, S., Posfai, J., Pflugrath, J.W. and Roberts, R.J. (1993) *Cell* 74, 299–307.
4. Labahn, J., Granzin, J., Schluckebier, G., Robinson, D.P., Jack, W.E., Schidkraut, I. and Saenger, W. (1994) *Proc. Natl. Acad. Sci.*, in press.
5. Woodcock, D.M., Crowther, P.J., Doherty, J., Jefferson, S., DeCruz, E., Noyer-Weidner, M., Smith, S.S., Michael, M.Z. and Graham, M.W. (1989) *Nucleic Acids Res.* 17, 3469–3478.
6. Messing, J., Crea, R. and Seeburg, P.H. (1981) *Nucleic Acids Res.* 9, 309–321.
7. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
8. Soberon, X., Covarrubias, L. and Bolivar, F. (1980) *Gene* 9, 287–305.
9. Fürste, J.P., Pansegrau, W., Frank, R., Blöcker, H., Scholz, P., Bagdasarian, M. and Lanka, E. (1986) *Gene* 48, 119–131.
10. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463–5467.
11. Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) *Gene* 33, 103–119.
12. Tran-Betcke, A., Behrens, B., Noyer-Weidner, M. and Trautner, T.A. (1986) *Gene* 42, 89–96.
13. Maxam, A.M. and Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* 74, 560–564.
14. Chai, S., Bravo, A., Lüder, G., Nedlin, A., Trautner, T.A. and Alonso, S. (1992) *J. Mol. Biol.* 224, 87–102.
15. Yasbin, R. F., Fields, P. I. and Anderson, B. J. (1980) *Gene* 12, 155–159.
16. Devereux, J., Haerberli, P. and Smithies, O. (1984) *Nucleic Acids Res.* 12, 387–395.
17. Smith, H.O., Annau, T.M. and Chandrasegaran, S. (1990) *Proc. Natl. Acad. Sci. USA* 87, 826–830.
18. Kawakami, B., Christophe, H., Nagatomo, M. and Oka, M. (1991) *Agric. Biol. Chem.* 55, 1553–1559.

19. Xu, G., Kapfer, W., Walter, J. and Trautner, T.A. (1992) *Nucleic Acids Res.* 20, 6517–6523.
20. Narva, K.E., Wendell, D.L., Skrdla, M.P. and Van Etten, J.L. (1987) *Nucleic Acids Res.* 15, 9807–9823.
21. Stefan, C., Xia, Y. and Van Etten, J.L. (1991) *Nucleic Acids Res.* 19, 307–311.
22. Janulaitis, A., Vaisvila, R., Timinskas, A., Klimasauskas, S. and Butkus, V. (1992) *Nucleic Acids Res.* 20, 6051–6056.
23. Itoh, H., Sadaoka, A., Kotani, H., Hiraoka, N. and Nakamura, T. (1990) *Nucleic Acids Res.* 18, 3903–3911.
24. Theriault, G., Roy, P. H., Howard, K. A., Benner, J. S., Brooks, J. E., Waters, A. F. and Gingeras, T. R. (1985) *Nucl. Acids Res.* 13, 8441–8461.
25. Walder, R.Y., Walder, J.A. and Donelson, J.E. (1984) *J.Biol.Chem.* 259, 8015–8026.
26. Barany, F., Danzitz, M., Zebala, J. and Mayer, A. (1992) *Gene* 112, 3–12.
27. Noyer-Weidner, M., Jentsch, S., Kupsch, J., Bergbauer, M. and Trautner, T. A. (1985) *Gene* 35, 143–150.
28. Behrens, B., Noyer-Weidner, M., Pawlek, B., Lauster, R., Balganesch, T. S. and Trautner, T. A. (1987) *EMBO J.* 6, 1137–1142.
29. Posfai, J., Bhagwat, A.S., Posfai, G. and Roberts, R.J. (1989) *Nucleic Acids Res.* 17, 2421–2435.
30. Uozumi, T., Hoshino, T., Miwa, K., Horinouchi, T., Beppu, T and Arima, K. (1977) *Mol. Gen. Genet.* 152, 65–69.
31. Guha, S. (1985) *J. Bacteriol.* 163, 573–579.
32. Connaughton, J.F., Kaloss, W.D., Vanek, G.P., Nardone, G.A. and Chirikjian, J.G. (1990) *Nucleic Acids Res.* 18, 4002.
33. Lange, C., Noyer-Weidner, M., Trautner, T.A., Weiner, M. and Zahler, S. (1991) *Gene* 100, 213–218.
34. Som, S., Bhagwat, A.S. and Friedman, S. (1987) *Nucleic Acids Res.* 15, 313–323.
35. Lauster, R., Trautner, T.A. and Noyer-Weidner, M. (1989) *J. Mol. Biol.* 206, 305–312.
36. Klimasauskas, S., Timinskas, A., Menkevicius, S., Butkiene, D., Butkus, V. and Janulaitis, A. (1989) *Nucleic Acids Res.* 17, 9823–9832.
37. Wu, J.C. and Santi, D.V. (1987) *J. Biol. Chem.* 262, 4776–4786.
38. Chen, L., MacMillan, A.M., Chang, W., Ezaz-Nikpay, K., Lane, W.S. and Verdine, G.L. (1991) *Biochemistry* 30, 11018–11025.
39. Karreman, C. and de Waard, A. (1990) *J. Bacteriol.* 172, 266–272.
40. Posfai, G., Kim, S., Szilak, L., Kovacs, A. and Venetianer, P. (1991) *Nucleic Acids Res.*, 19, 4843–4847.
41. Klimasauskas, S., Kumar, S., Roberts, R.J. and Cheng, X. (1994) *Cell*, 76, 357–369.