The EMBO Journal vol.6 no.11 pp.3543–3549, 1987

Construction and use of chimeric SPR/ϕ3T DNA methyltransferases in the definition of sequence recognizing enzyme regions

Tanjore S.Balganesh[1], Luzia Reiners, Roland Lauster, Mario Noyer-Weidner, Klaus Wilke and Thomas A.Trautner

Max-Planck Institut für Molekulare Genetik, Ihnestrasse 73, D-1000 Berlin 33, Germany

[1]Present address: Astra Research Centre, Bangalore 560003, India

Communicated by T.A.Trautner

Multispecific DNA methyltransferases (Mtases) of temperate *Bacillus subtilis* phages SPR and ϕ3T methylate the internal cytosine of the sequence GGCC. They differ in their capacity to methylate additional sequences. These are CCGG and CC(A/T)GG in SPR and GCNGC in ϕ3T. Introducing unique restriction sites at equivalent locations within the two genes facilitated the construction of chimeric genes. These expressed Mtase activity at a level comparable to that of the parental genes. The methylation specificity of chimeric enzymes was correlated with the location of chimeric fusions. This analysis, which also included the use of mutant genes, showed that domains involved in the recognition of target sequences unique to each enzyme [CCGG, CC(A/T)GG or GCNGC] are represented by the central non-conserved parts of the proteins, whilst recognition of the sequence (GGCC), which is a target for both enzymes, is determined by an adjacent conserved region.

Key words: B. subtilis phages SPR, ϕ3T/DNA methylation/protein–DNA interaction/protein engineering/target recognition

Introduction

Sequence specific DNA methylation is mediated by DNA methyltransferases (Mtases). These enzymes attach to DNA, bind the donor of methyl groups, S-adenosyl methionine, and recognize characteristic DNA sequences, a base within which becomes methylated in the reaction. Together with a variety of other DNA binding proteins, the capacity of Mtases to recognize specific sequences in DNA and to alter them chemically make them model molecules for elucidating protein–DNA interaction.

Towards defining amino acid sequences of Mtases involved in sequence recognition, we have investigated three closely related Mtases from the *Bacillus subtilis* phages SPR, ϕ3T and $\varrho 11_s$. These enzymes are multispecific Mtases. They share the capacity to methylate the sequence GGCC, but differ from each other in their capacity to methylate additional sequences (Buhk *et al.*, 1984; Günthert and Trautner, 1984; Behrens *et al.*, 1987; Günthert and Reiners, 1987). The isolation of mutants of the SPR and the ϕ3T Mtases, which are deficient in either one of their methylation specificities showed that structures involved in the recognition of specific target sequences may represent independent domains of the enzymes (Trautner *et al.*, 1980; Noyer-Weidner *et al.*, 1983; Buhk *et al.*, 1984; Günthert *et al.*, 1986b). In a preceding communication (Behrens *et al.*, 1987) we have compared the amino acid sequences of the three enzymes. This revealed that they all contain two extensive and very similar regions of amino acids at their N and C termini respectively. These blocks of homology are separated in each enzyme by sequences of amino acids which are variable in size and which have only limited homology to each other (Behrens *et al.*, 1987). This building plan suggested that the conserved sequences may be involved in general steps of the methylation reaction and in the recognition of the sequence GGCC, which is methylated by either enzyme. In contrast, domains involved in the recognition of those target sequences which distinguish each enzyme would be represented by the variable region of each Mtase. Evidence for this interpretation is presented in this communication, where we report the construction of chimeric Mtases. These enzymes were generated through *in vitro* recombination at equivalent locations of the SPR and ϕ3T Mtase genes. All chimeric constructs were enzymatically active. Analyzing their methylation potential permitted the allocation of enzyme regions responsible for target recognition.

Results

General considerations

If Mtases consisted of functionally independent regions mediating the various steps in sequence specific DNA methylation, one should be able to localize their specificity determining regions by analyzing the methylation potential of chimeric enzymes generated by exchanging functionally equivalent segments of enzymes with different methylation specificities. Since the Mtases of phages SPR and ϕ3T show extensive amino acid sequence conservation and have at the same time different methylation specificities, they appeared to be particularly suited for such studies. However, wide nucleotide sequence divergence between

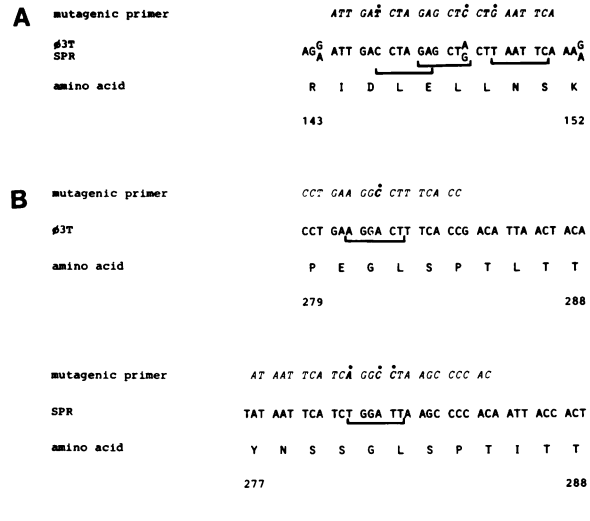


Fig. 1. Introduction of new restriction sites into ϕ3T and SPR DNAs. **A**: mutagenesis to introduce the *Xba*I (TCTAGA), *Sac*I (GAGCTC) and *Eco*RI (GAATTC) sites at the underlined regions. An identical primer could be used for both DNAs. **B**: mutagenesis by two different primers to introduce the *Stu*I (AGGCCT) site at the regions underlined. In all mutagenic primers, nucleotides to be exchanged are marked by dots.

© IRL Press Limited, Oxford, England

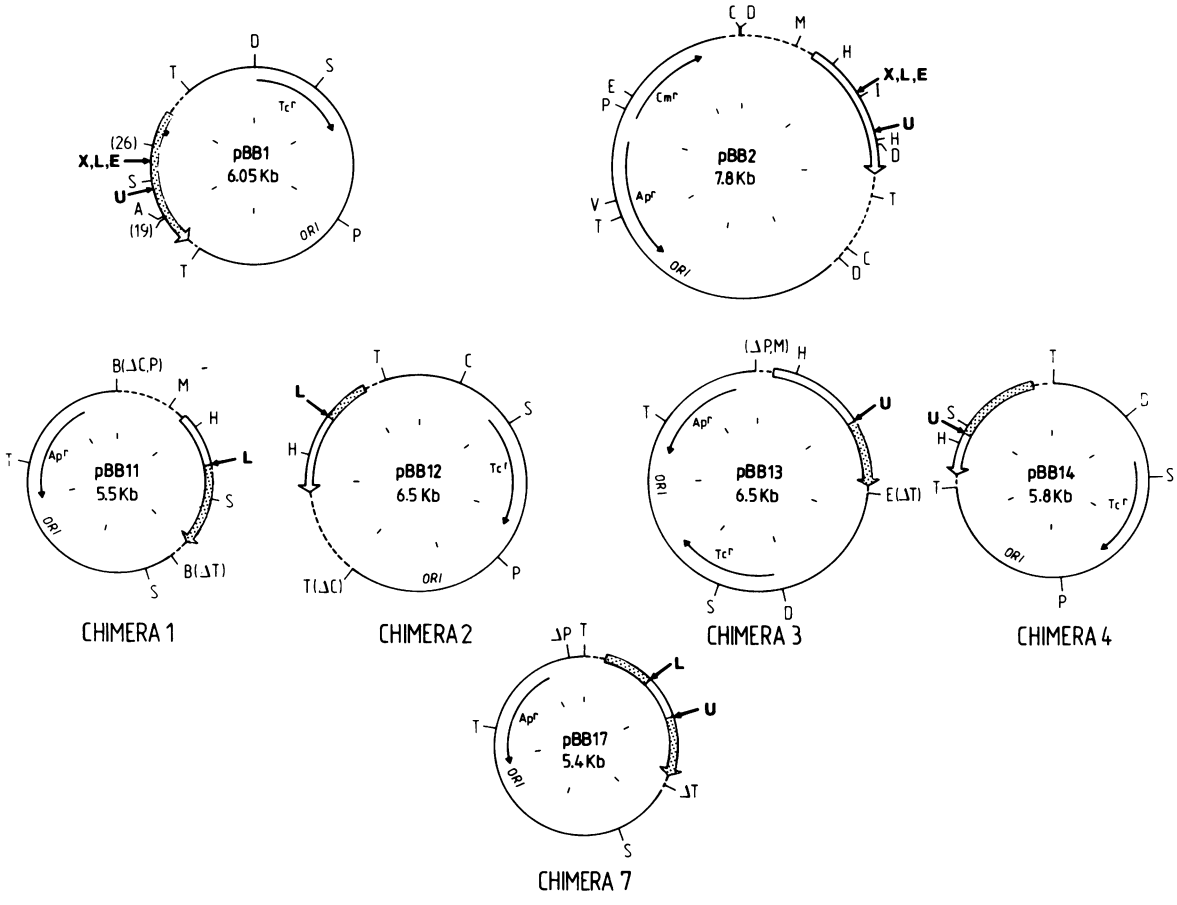


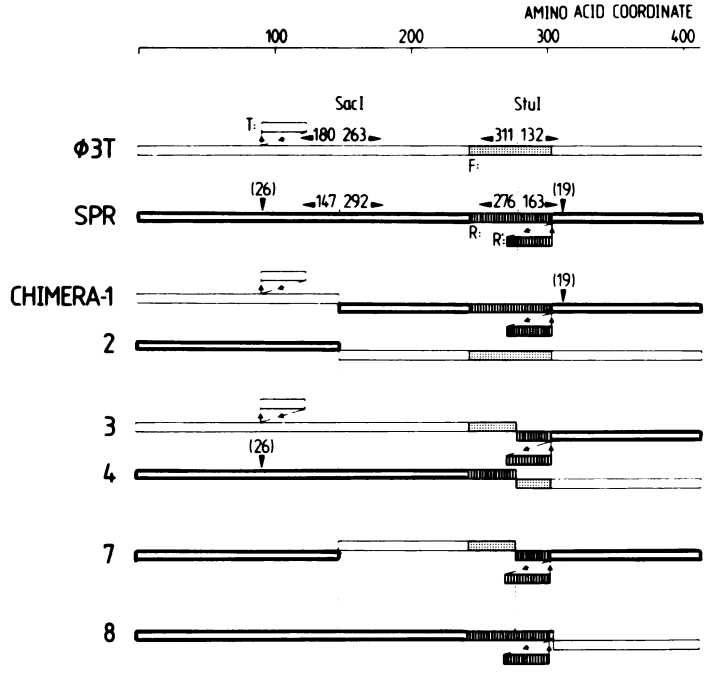
Fig. 2. Plasmids containing various Mtase genes. The cloned Mtase genes, (SPR: shaded, ϕ3T: blank), and their direction of transcription are given by double lined arrows. Vector DNA and insert DNA are shown as a continuous or broken line. Engineered restriction enzyme sites are shown in bold letters (**X**: *Xba*I, **L**: *Sac*I, **E**: *Eco*RI and **U**: *Stu*I), other restriction sites as plain letters (A: *Hae*III, B: *Bam*HI, C: *Cla*I, D: *Hin*dIII, H: *Hpa*I, M: *Xmn*I, P: *Pvu*II, S: *Sal*I, T: *Pst*I and V: *Pvu*I). Δ describes the deletion of a restriction site. The location and transcription of vector antibiotic genes are indicated by arrows.

these genes has prevented both *in vivo* and *in vitro* recombination (unpublished results). The recent establishment of the amino acid sequences of the phage Mtases has provided an alignment of these sequences which allowed us to describe possible equivalence points on the two proteins (Behrens *et al.*, 1987). Gene fusion by *in vitro* recombination could then be attempted at locations on the DNA corresponding to the equivalent positions defined. In the absence of identical restriction sites at equivalence points, such sites had to be introduced into the genes through oligonucleotide directed mutagenesis. The number of locations, however, where such manipulations could be performed, was limited by two constraints: (i) the site introduced had to be a (novel) single site in each gene; (ii) mutagenesis had to be neutral both with respect to the reading frames of the genes and the amino acid sequence of parental enzymes and chimeras.

Construction and characterization of chimera 1 and 2

A pair of reciprocal chimeras (1 and 2) between the ϕ3T and SPR Mtases were obtained by a chimeric fusion at a newly introduced *Sac*I site at amino acid coordinate 147. This *Sac*I site is located within a DNA sequence coding for 30 identical amino acids between coordinates 141 and 172. In Figure 1A we show the nucleotide sequences corresponding to ten of these amino acids and the mutagenic primer, which was used to introduce the *Sac*I site (in addition to unique *Eco*RI and *Xba*I sites) into the two Mtase genes. The presence of these sites in plasmids pBB1 (SPR) and pBB2 (ϕ3T) (Figure 2) and the expression of Mtase activity was verified by digestion with the relevant restriction enzymes (data not shown). By appropriate restriction enzyme digestions of these plasmids, and ligation at the *Sac*I sites of the individual fragments to effect the gene fusion, we obtained plasmids pBB11 and pBB12 (Figure 2), which contain the reciprocal chimeric genes 1 and 2.

The amino acid sequences of the ϕ3T, and SPR Mtases as well as those predicted for the chimeric Mtases 1 (472 amino acids) and 2 (410 amino acids) are schematically represented in Figure 3. Both chimeras express levels of methylation activity similar to the parental plasmids (Figure 3), although neither of the *Sac*I terminated subfragments of the genes alone have enzymatic activity (data not shown). This result demonstrates the interchangeability of the ϕ3T and SPR N- and C-terminal enzyme subfragments in providing Mtase activity when fused at the equivalent amino acid 147. With respect to methylation specificity, the chimeric Mtase 1 methylates target sequences identical to those methylated by the SPR Mtase, while the reciprocal chimeric Mtase 2 methylates sequences modified by the ϕ3T enzyme. Therefore enzyme regions responsible for the recognition of sequences characteristic for each Mtase [ϕ3T: GCNGC; SPR: CCGG and CC(A/T)GG] must be located C terminal to the fusion point corresponding to amino acid 147. Also the unique 33 amino acid region T of the ϕ3T Mtase (Figure 3) is obviously dispensable for both the general activity of the enzyme and the recognition of the target sequences.



LENGTH OF PROTEIN (AMINO-ACIDS)	SEQUENCE MODIFIED: GGCC (HaeIII)	CCGG (MspI)	GCNGC (FnuDII)	CCAGG (EcoRII)	RELATIVE SPECIFIC ACTIVITY
443	+	−	+	−	0.68
439	+	+	−	+	1.00
472	+	+	−	+	0.38
410	+	−	+	−	1.00
474	+	+	+/−	−	1.60
408	+	−	−	−	0.46
441	+	+	+*	−	4.60
439	−	+	−	+	n.d

Fig. 3. Schematic representation of structure and enzymatic activity of the φ3T, SPR and chimeric Mtases. The amino acid alignment coordinates and the description of the non-homologous blocks T, F, R and R′ are those of Figure 4B of Behrens *et al.* (1987). φ3T sequences are represented by boxes with light outlines while SPR sequences are indicated by boxes with heavy lines. The central variable regions are depicted with dotted (φ3T) or vertically striped (SPR) bars. The location of the restriction sites *Sac*I and *Stu*I introduced by oligonucleotide mutagenesis on the corresponding DNA sequences are indicated by vertical lines. The numbers left or right of these locations indicate the numbers of amino acids extending to the N or C termini of the proteins. 19 and 26 describe the location of the SPR mutations 19 and 26. The signs +, +/−, − describe complete, partial or no methylation of the respective sequences, as determined by restriction enzyme digestion of the plasmid containing the gene concerned. +* indicates the partial methylation of the GCNGC sequence by chimeric Mtase 7 as a reflection of the 'relaxed' specificity of the protein. The specific activities of the different Mtases were determined from crude extracts of cells as described by Günthert *et al.* (1986a), and related to the specific activity of an extract of cells with plasmid pRB121 (SPR) which had an absolute specific activity of 370 u/mg of total protein. The characterization of gene products synthesized is shown in Figure 4.

Construction and characterization of chimeric genes 3 and 4

Following the results obtained with chimeras 1 and 2, we constructed and analyzed a second set of chimeras, which had their fusion points further from the N termini, within the variable regions of the parental proteins (Figure 3). These regions, composed of 57 amino acids (motif F) in the φ3T Mtase and 86 amino acids (motifs R plus r′) in the SPR Mtase (Behrens *et al.*, 1987; see Figure 3) contain at different relative positions a short homologous stretch of seven amino acids at coordinates 281−288 (Figure 1B). A unique *Stu*I site could be generated within the corresponding DNA sequences through the use of two different primers as shown in Figure 1B. Genes fused at this site, represented by the conserved G residue at coordinate 281, would code for chimeric Mtases, which derive segments of the variable region from each protein in addition to the integral N- and C-terminal regions.

Chimera 3, present in plasmid pBB13 (Figure 2), derives its N terminal 311 amino acids from the φ3T Mtase, and the C terminal 163 amino acids from the SPR Mtase (Figure 3). Its variable region contains 35 amino acids of the φ3T and 53 amino acids of the SPR enzymes. The latter include the complete R′ motif and a part of the R motif. The reciprocal recombinant construct, chimera 4, is represented on plasmid pBB14 (Figure 2). This chimeric Mtase has the SPR Mtase sequence up to amino acid 276, fused at the conserved G residue to 132 C terminal amino acids of the φ3T Mtase (Figure 3). The recomposed variable region of the chimeric Mtase 4 is represented by 33 amino acids derived from the sequence motif R of the SPR protein joined with 22 amino acids from the motif F of the φ3T Mtase.

Both chimeras 3 and 4 were enzymatically active and expressed the universal GGCC specific Mtase activity, showing that also the fusion site chosen within the variable regions was an equivalence point as assumed in our alignment. Chimeric Mtase 3 modifies also the sequence CCGG, associated with the SPR Mtase. In contrast neither the other SPR specific sequence CC(A/T)GG nor the φ3T characteristic sequence GCNGC are appreciably methylated. Chimera 4 does not express any of the three Mtase activities distinguishing the parental Mtases. Relating these methylation specificities to the configurations of the two chimeras, we propose that variable regions interrupted by the *Stu*I fusion site are determinants of these methylation specificities which have been affected in chimera formation. Thus motifs F and R would be involved in the recognition of GCNGC and CC(A/T)GG respectively. Obviously partial representation of these motifs in chimeras 3 and 4 is inadequate for the full expression of this function. This suggests that amino acids on either side of the conserved G at coordinate 281 contribute to the functioning of these specificity determinants. On the other hand, from the preservation of CCGG methylating activity in chimera 3, we associate the domain responsible for CCGG recognition with the C terminal 163 amino acids derived from SPR, which includes the integral motif R′.

To further delineate this region, 106 amino acids of the C terminus of the SPR Mtase were replaced with the equivalent 106 amino acids of the C terminal region of the φ3T Mtase (chimera

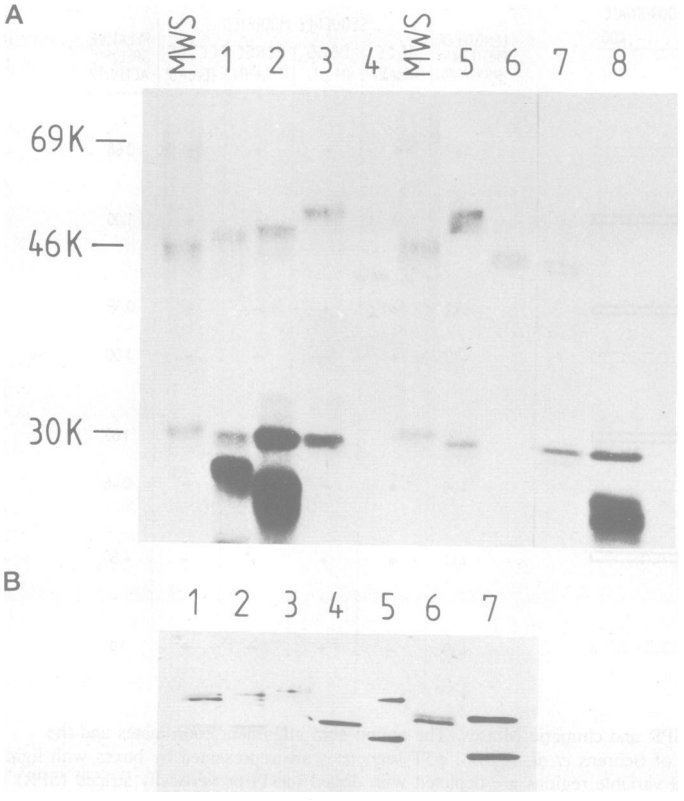


Fig. 4. Identification of products of various Mtase genes. **A**: Autoradiogram of [^{35}S]methionine labelled plasmid coded proteins from *E. coli* minicells (15% SDS−PAGE). **Lanes 1** and **2** show extracts of minicells with parental plasmids pRB121 (SPR) and pBN16 (ϕ3T). Products from plasmids with chimeric genes are shown in **lane 3** (pBB11, chimera 1), **4** (pBB12, chimera 2), **5** (pBB13, chimera 3), **6** (pBB14, chimera 4) and **7** (pBB17, chimera 7). In **lane 8** is shown an extract from minicells containing pBR328. Molecular weight markers (MWS) were from Amersham. The differences of product patterns seen between the tracks in the lower M_r range can be attributed to different cloning sites utilized in these plasmids and the use of either pBR322 or pBR328 as vectors. **B**: Western blot analysis of partially purified fractions of the various Mtases using anti-SPR serum. The patterns presented were obtained with partially purified enzymes from *E. coli* cells containing plasmids with parental (1,2) and chimeric (3−7) Mtase genes. **Lanes 1** through **7** correspond to extracts from cells with the same plasmids as used in the minicell analysis.

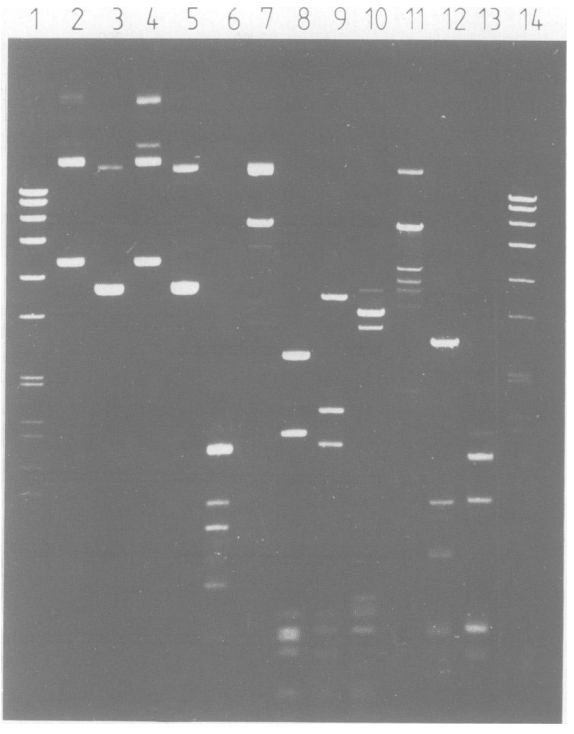


Fig. 5. Agarose gel electrophoresis of plasmids restricted with various restriction enzymes. **Lanes 1** and **14** show *Eco*RI digested SPP1 DNA, serving as a molecular weight standard (Ratcliffe *et al.*, 1979). The pairs of **lanes 2/5, 4/5, 6/7, 10/11, 12/13** shows plasmid DNA with the chimeric genes 3 and 7 digested with *Hae*III, *Msp*I, *Fnu*4HI, *Hha*I and *Sau*3A, respectively. **Lanes 8** and **9** show the restriction pattern obtained on digesting plasmids pBB2 and pBB1 with *Hha*I.

8, Figure 3). The use of a linker to effect this fusion resulted in the change of amino acids V and L at coordinates 308 and 309 to A and G. The chimeric Mtase 8, which contains the entire variable region of the SPR Mtase, methylates CCGG and CC(A/T)GG sequences indicating that the 106 C terminal amino acids of the SPR protein do not contribute to CCGG or CC(A/T)GG specific interaction. Hence the determinants of CCGG methylation are confined within amino acids 276 and 333 of the SPR enzyme, which is essentially composed of the sequence motif R′ of the SPR Mtase. The failure of chimeric Mtase 8 to methylate GGCC will be discussed later.

Characterization of chimera 7, which codes for a Mtase with a 'relaxed' specificity

The N terminal amino acids of SPR and ϕ3T Mtases contained within chimeras 1 and 2 are distinguished by the presence in the ϕ3T sequence of the motif T and a small block of non-conservative amino acid exchanges (alignment coordinate 101−107) (Behrens *et al.*, 1987). In spite of these differences, these fragments were apparently fully interchangeable in the construction of chimeras 1 and 2. To analyze whether this was also the case in combination with the specificity region as defined by chimera 3, we have constructed a derivative — chimera 7 — in which the amino acids extending from amino acid coordinate 1 through 147 were derived from SPR, rather than from ϕ3T. Chimera 7, like chimera 3, fully expresses both GGCC and CCGG specific methylation. Specificity recognition differs, however, from that of the chimeric Mtase 3 in that the GCNGC sequence specific methylation is clearly enhanced, albeit not to a level to give complete protection against *Fnu*4HI restriction (Figure 5, lane 7). The enhancement is apparently not a mere restoration of the GCNGC methylating activity, as found in the ϕ3T Mtase, but rather reflects the generation of a novel specificity of the Mtase. This followed from our observation that the plasmid DNA with the chimeric Mtase gene 7 was not only partially resistant against *Fnu*4HI cleavage but also against cleavage by other restriction enzymes, to which either parental DNAs were completely sensitive. Such enzymes which like *Fnu*4HI recognize a 'GC' containing target sequence, including *Hha*I (Figure 5, lane 11), *Alu*I and *Hin*dIII (not shown). No protection against *Sau*3A cleavage was detectable (Figure 5, lane 13) and hence the specificity of the chimeric Mtase 7 is not relaxed to an extent which would allow random methylation of cytosines alone. Work is in progress to define the consensus target sequence of this chimeric Mtase and also to understand the observation that the 'GC' containing sequences are only partially protected in contrast to the GGCC and CCGG target sequences. Beyond this analysis, it remains to be answered whether the introduction of the N terminal 147 amino acids from the SPR Mtase in chimera 7 only enhanced a specificity recognizing pattern, which was not

detectably expressed in chimera 3, or whether the specificity of chimera 7 was indeed generated *de novo*.

Characterization of the mutant chimeric gene 1−19

Like the parental enzymes all chimeras analyzed (with the exception of chimera 8) were proficient in expressing GGCC methylation, indicating that restructuring the variable regions does not affect this methylation potential which is shared by both enzymes. Hence, the domain required for recognition of the sequence GGCC must reside within homologous regions of the enzymes, which cannot be resolved using chimera analysis. To characterize also this domain, we have probed the effect of defined specificity mutations within the conserved regions on the methylation specificity of chimeric enzymes. In addition, we have tested our expectation that mutations in the SPR Mtase gene known to impede Mtase activity at large would have the same effect on the expression of chimeric Mtases.

Analyzing the methylation specificity of the chimeric Mtase 1 showed that the C terminal 292 amino acids of the SPR Mtase are involved in CCGG and CC(A/T)GG interaction. Within the conserved 110 C terminal amino acids of this region at coordinate 311 we had previously localized the SPR mutation 19, which selectively abolishes GGCC methylation (Buhk *et al.*, 1984). When mutation 19 was introduced into the chimeric gene 1, it also rendered the chimeric Mtase 1−19 GGCC methylation deficient. This indicates that the determinator of the recognition of GGCC is included within the 110 C terminal conserved amino acids of both enzymes. This result as well as the loss of GGCC specific interaction in the construction of chimera 8, as a consequence of the amino acid changes at amino acids 308 and 309 (see previous section) suggest furthermore that the residues mediating GGCC interaction must be immediately C terminal to the variable regions.

Characterization of the mutant chimera 4−26

Mutation 26 of the SPR gene leads to a complete loss of specific methylating activity (Günthert *et al.*, 1986b) and had been previously identified as a D to N change at amino acid coordinate 92 (Buhk *et al.*, 1984). The mutation being located within the conserved N terminal region of the protein, it was of interest to determine if the loss of the methylation potential of the mutant was limited to the SPR protein or would also affect the activity of an enzyme containing a chimeric variable region. To this end the mutation 26 was introduced into the chimeric gene 4. This construction resulted in the loss of the ability of the chimeric Mtase 4 to methylate the sequence GGCC, which is the only target recognized, suggesting the functional equivalence of amino acid 92 in both the SPR and the chimera 4 Mtase.

Identification of the chimeric gene products

To estimate the level of expression and the stability of the chimeric enzymes we have compared the methylation activities of crude extracts prepared from cells harbouring plasmids encoding the different parental and chimeric Mtases. The values of relative specific activity presented in Figure 3 show that the methylation activity of crude extracts of cells containing the chimeric Mtases 1−4 is in a comparable range to that of SPR or ϕ3T Mtase containing extracts. The extract containing the chimeric Mtase 7 however has a significantly higher activity. These data demonstrate the stability of the chimeric Mtases. The higher methylation activity of the crude extract containing the chimeric Mtase 7 may reflect the relaxed specificity of this enzyme.

These conclusions are confirmed by the analysis of proteins synthesized by *Escherichia coli* minicells harbouring the plasmids with the different chimeric genes. As seen in Figure 4A, the chimeric Mtases are synthesized with an efficiency comparable to that of the parental SPR and ϕ3T enzymes. The apparent mol. wt of the chimeric Mtases 1, 3 and 4 which are 52.4, 52.5 and 45 kd, respectively, are in good agreement with the predicted values which are 53.1, 53.8 and 46.3 kd. The chimeric Mtases 2 and 7, however, with their anticipated mol. wt of 46.9 and 50.2 kd, respectively, had a mobility corresponding to mol. wts of 42.7 and 45.7 kd. Several experiments were performed through which this discrepancy could be attributed to an aberrant mobility of these proteins. Thus a detailed restriction analysis of the plasmids with the chimeric genes 2 and 7 confirmed the expected structure of the recombinant genes and argued against the possible occurrence of deletions in the construction. Analysis of the encoded products of independently constructed chimera 2 revealed an identical mobility for the chimeric protein. Furthermore the products encoded by the SPR genes with and without the introduction of the engineered restriction sites have the same electrophoretic mobility. In addition the methylation specificity of the chimeric Mtase 2 includes both the targets methylated by the ϕ3T Mtase which again argues against the presence of a deletion affecting a functional component of the protein. The authenticity of the chimeric products 2 and 7 to be the Mtase proteins was confirmed by reaction with the anti-SPR antibodies (Figure 4B, lane 4 and 7).

The SPR Mtase has been previously shown to purify as a doublet band in *E. coli* (Günthert *et al.*, 1986b) and the smaller of the doublet has been shown to be a processed product (Lauster, unpublished). Also the chimeric Mtases seen in Figure 4A are apparently processed in the same manner as the parental enzymes.

The chimeric proteins were also identified by immunoblotting of partially purified fractions of crude extracts from cells harbouring the plasmids coding for the different Mtases, using anti-SPR Mtases serum (Figure 4B). The anti-SPR Mtase antibodies had been previously shown to cross-react with the ϕ3T Mtase (Günthert *et al.*, 1986b). As seen in Figure 4B the relative mobilities of the different chimeric Mtases identified by immunoblotting correspond to those seen with the minicell analysis. In addition cross-reacting material is seen with the chimeric Mtases 5 and 7 (Figure 4B, lane 5 and 7). Since these degradation products are not detected among the proteins synthesized in minicells they might probably have accumulated during the purification procedure applied.

Discussion

A comparison of the amino acid sequences of the related Mtases of SPR, ϕ3T and $\varrho 11_s$ has revealed a similar organization of these enzymes (Behrens *et al.*, 1987). Long N- and C-terminal blocks of amino acids, highly conserved between the three enzymes, bracket a shorter amino acid sequence, which is variable between the enzymes in size and configuration. Specificity recognizing domains were assumed to be associated with the variable regions, while domains responsible for enzymatic activities shared by all enzymes were allocated to the conserved areas. Here we have presented experimental evidence for the validity of this concept. Chimeric Mtases in which conserved regions of one enzyme became associated to the variable region of another and also chimeras, where fusions were performed within the variable regions of two enzymes, proved to be active Mtases. Since the junctions chosen to build chimeras were based on the amino acid alignment map of the two proteins described by Behrens *et al.* (1987) this finding verified the fundamental assumption that the alignment performed did not only represent

a convenient formalism but reflected the juxtapositioning of equivalent amino acids of the proteins.

The hybrid Mtases vary in length from 408 amino acids (chimeric Mtase 4) to 474 (chimeric Mtase 3). These differences are mainly brought about by the presence or absence of the unique 33 amino acid block (T of the ϕ3T MTase) at coordinate 92 of the protein. The functional capacity of the hybrid proteins 2 and 4 shows that this block is entirely dispensable with reference to the methylation of the GGCC or the GCNGC sequences. Another factor contributing to the length variations is the difference in the size of the variable regions of the individual proteins. Though it cannot be generalized that the size of these proteins is not a critical parameter for their function, the experiments described here indicate that moderate increases or decreases of length at specific points on the protein sequence are indeed feasible. This may reflect an inherent flexibility of the proteins to accommodate and evolve sequences conferring additional specificities.

In agreement with our expectation, the construction of chimeras 1 and 2 placed the determinator of sequence specificity in a region C terminal of the chimeric fusion coordinate 147. Between the reciprocal chimeras 1 and 2 the entirety of the methylation potential of the parental input [GGCC; CCGG, CC(A/T)GG; GCNGC] was preserved. The situation was more complex with the reciprocal chimeras 3 and 4, which had the SPR/ϕ3T junction at coordinate 281 within the variable sequence. Here neither the SPR specific methylation potential affecting CC(A/T)GG nor that of ϕ3T leading to GCNGC methylation was retrieved between the chimeras. The *Stu*I scission mediating fusion of SPR and ϕ3T DNAs hence defines a specificity determining domain of both SPR and ϕ3T. On the other hand, the CCGG methylating capacity characterizing SPR is observed in chimera 3 and could be correlated with the presence of the motif R′ in chimera 3. This result is fully compatible with the identification of SPR mutant 83, which is deficient only in CCGG methylation due to a G to E change at coordiante 267 within the R′ motif (Buhk *et al.*, 1984).

All chimeras derived from recombination between unaltered wild type sequences methylate the sequence GGCC, strongly suggesting that in the SPR and ϕ3T Mtases the same sequence motif located within a conserved region is mediating this methylation specificity. Obviously the method of localizing specificity determining domains through chimera formation is inadequate for the identification of the domain responsible for GGCC methylation. However, through the persistence of the phenotype of SPR mutant 19, which is deficient solely in GGCC methylation, in chimera 1−19, and through the generation of a GGCC deficiency concomitant with a double amino acid exchange at locations 308 and 309 incurred in the course of constructing chimera 8, we conclude that the GGCC specific methylation is determined by a conserved region adjacent to the C terminal section of motifs F or R′. The positioning of the double amino acid exchange causing only GGCC methylation deficiency does also give a maximal size of the domain of the SPR Mtase responsible for CCGG methylation. The region is represented by the 57 amino acids between alignment coordinates 281 and 307. Similarly, the region in ϕ3T responsible for GCNGC methylation would extend for 57 amino acids between the end of the N terminal region of homology at coordinate 244 and coordinate 304. No estimates can be given on the size of the GGCC recognizing domains. Such information and also a more precise description of the confines of other specificity determining domains, like the region in SPR specifying CC(A/T)GG methylation, will be provided in a subsequent communication through a mutational analysis of the Mtase genes.

In spite of the extensive overall homology between the amino acid sequences of the SPR and ϕ3T fragments extending from the N terminus to the *Sac*I site corresponding to amino acid 147, these sequences differ from each other by an insertion of 33 amino acids present only in the ϕ3T Mtase and by a few short runs of nonconservative amino acid changes. To determine whether such differences affect the specificity we have constructed chimera 7. In addition to the most promiment Mtase activities found also in chimera 3, chimera 7 has an additional, obviously relaxed specificity towards recognition of 'GC' containing sequences, which had neither been detected in SPR nor in ϕ3T. Although we cannot provide a molecular interpretation of this observation, these results indicate that sequences outside the specificity determining variable region can modulate its activity quantitatively and/or qualitatively.

Altogether, the results presented provide further support for a modular organization of Mtase genes and enzymes. Particularly, domains involved in recognizing specific DNA sequences are apparently arranged as contiguous successive units of ~60 amino acids. These include the domains determining specificities characteristic for each enzyme, which represent the previously acknowledged variable regions as well as part of a successive conserved region responsible for methylation of the sequence GGCC. In spite of these data we realize that the analysis and description of the organization needs further refinement in terms of the possible overlap of the domains involved in the recognition of the different target sequences and the role of individual amino acids within each domain.

Bearing on the generality of our observation, a modular organization of the methylating activity as described here for the type II Mtases of SPR and ϕ3T phages is also apparent in type I restriction/modification (R/M) systems of *E. coli* and *Salmonella*. Here the modules are represented on the one hand by the interchangeable proteins of the type I R/M enzyme complex. But also within the protein S, the subunit determining the sequence specificity of the complex, two separated variable regions could be identified, each of which is responsible for the recognition of one of the two composite elements of the enzymes' target sequences (Gough and Murray, 1983). Reassortment of the two variable regions within proteins from different genes through *in vivo* recombination generated a novel recognition specificity (Nagaraja *et al.*, 1985; Fuller-Pace and Murray, 1986).

Materials and methods

All *E. coli* strains used as hosts were *rgl*B (Noyer-Weidner *et al.*, 1986). *E. coli* strain HB101 (Boyer and Roulland-Dussoix, 1969) was used for cloning experiments. The *E. coli* strain GM1499, deficient in 5-methyl cytosine modification (kind gift of M.G.Marinus) was used as the host to detect CC(A/T)GG modification of plasmids and for the *in vitro* methylation assays. *E. coli* strain TC410 (Noyer-Weidner *et al.*, 1986) was used for the analysis of the plasmid encoded gene products in mini cells.

Plasmids

Plasmid pBR328 (Soberon *et al.*, 1980) and pBR322 (Bolivar *et al.*, 1977) were used as cloning vectors. Plasmid pBB2 is a derivative of pBN16 with the ϕ3T Mtase gene (Noyer-Weidner *et al.*, 1985). pBB1 was derived from pRB121 (Günthert *et al.*, 1986a) and contains the SPR Mtase gene. Plasmids pRB421 and pRB321 which contain the SPR Mtase gene with the 19 and the 26 mutation, respectively, have been described previously (Günthert *et al.*, 1986a).

Reagents and general techniques

Restriction enzymes, synthetic linkers, *Pol*I (Klenow) and other relevant enzymes were purchased from Boehringer (Mannheim, FRG). 8-mer phosphorylated linker *Pvu*II d(pCCAGCTGG) was purchased from New England Biolabs, Beverly, MA, USA. [γ-^{32}P]ATP, [^{35}S]methionine and S-adenosyl-L methyl-[^{3}H]methionine was purchased from Amersham-Buchler (UK). Standard transformation and cloning techniques were used as described by Maniatis *et al.* (1982).

Oligonucleotide site directed mutagenesis

The primers shown in Figures 1A,B, were synthesized in a Applied Biosystems DNA synthesizer. The gapped duplex method for site directed mutagenesis described by Kramer *et al.* (1984a) was followed. The 0.8 kb *Hpa*I(H) fragment of the φ3T Mtase gene of plasmid pBB2 (Figure 2) was cloned in M13 vectors and used as template for the mutagenesis with both primers. The 5′ *Pst*I(T)−*Sal*I(S) (0.93 kb) fragment and the 0.78 kb, 3′ *Sal*I−*Pst*I fragment of the SPR Mtase gene of plasmid pBB1 (Figure 2) cloned in separate M13 vectors were used as templates for mutagenesis with the primers in Figure 1A and B, respectively. The DNA mixture after repair synthesis was transformed into *E. coli* strain BMH-71-18, *mut*S (Kramer *et al.*, 1984b). Individual phage isolates were screened for the mutation by dot blot analysis (Zoller and Smith, 1983). The base changes made were then confirmed by restriction enzyme analysis. The mutagenized fragments were recloned into the φ3T or the SPR Mtase gene. The methylation activity of the mutagenized Mtases was confirmed to be identical to that of the parental Mtases.

Plasmid construction

Like plasmids with the parental Mtase genes, recombinant plasmids harbouring the chimeric genes could be transformed into and stably maintained in *E. coli* strains, provided the host was deficient in *rgl*B nuclease activity (Noyer-Weidner *et al.*, 1986). Unlike all other plasmid carrying cells, those with the plasmid pBB17, containing the chimeric gene 7 coding for the Mtase with a relaxed specificity, were found to grow very slowly. Possibly the extensive methylation of host DNA through chimera 7 Mtase is deleterious to the cells. As the expression of the SPR and the φ3T Mtase genes had previously been shown to be strongly influenced by read-through transcription from external promotor sequences (Günthert *et al.*, 1986a) the chimeric genes described here were cloned at sites which would permit such read-through transcription in addition to transcription originating from the indigenous promotors of the genes. All chimeric genes were preceded by at least 150 nucleotides of phage DNA 5′ to the translation initiation codon of the gene on the cloned fragments and thus included the putative −35, −10 and the ribosomal binding site of the genes (Tran-Betcke *et al.*, 1986). The plasmids pBB11 (a pBR328 derivative) and pBB12 (pBR322 derived recombinant) were constructed by fusing appropriate restriction fragments derived from plasmids pBB2 and pBB1. Except for the fusion internal to the gene all manipulations of the DNA sequence leading to the construction of the plasmids with the chimeric genes were performed at locations external to the coding regions of the genes. The plasmid pBB13 (a pBR328 derivative) and pBB14 (a pBR322 recombinant) were, however, to be constructed from subfragments of the φ3T and SPR Mtase genes because of the inhibition of *Stu*I cleavage on the methylation of the internal C within its recognition sequence. A subset (GGCC) of the *Stu*I recognition sequence (AGGCCT) is the target of the φ3T and the SPR Mtases. pBB17 harboring the chimeric gene 7 was constructed from plasmids pBB12 and pBB13. Chimeras 1−19 and 4−26 were constructed by exchanging appropriate restriction fragments from the chimeric genes 1 and 4, with the corresponding fragments derived from the mutant genes. Chimeric gene 8 (plasmid pBB18) was constructed by replacing the *Hae*III(A)−*Pst*I fragment of the SPR gene from plasmid pBB1 with a *Hpa*I−*Pst*I of the φ3T gene from the plasmid pBB2. The *Hae*III and *Hpa*I sites are, however, separated by 8 nucleotides corresponding to amino acids 307(P), 308(V) and 309(L) on the alignment maps. The *Pvu*II linker used to fuse the fragments introduced the amino acids P, A and G at the respective positions.

Determination of the methylation specificity encoded by the chimeric genes

Purified preparations of plasmids harbouring the chimeric genes were digested with restriction enzymes *Hae*III, *Msp*I, *Fnu*4HI, *Eco*RII, *Hha*I and *Alu*I, all of which have 15−30 cleavable sites on the respective plasmids. These enzymes are inhibited by the methylation of the C within their recognition sequence (Kessler and Höltke, 1986). Unmodified pBR328 DNA was added as an internal control to the restriction digest. The resistance of the recombinant plasmids against the individual restriction enzymes indicated the modification specificity of the encoded chimeric enzymes.

In vitro methylation assays

The determination of Mtase activity in crude extracts prepared from cells harbouring the recombinant plasmids was performed as described previously (Günthert *et al.*, 1986a). *Micrococcus lysodeikticus* DNA was used as substrate. 1 unit of specific activity for the Mtase is defined as that amount of enzyme which incorporates 1 pmol methyl groups from S-adenosyl-L methyl-[^{3}H]methionine into DNA within 1 h at 37°C.

Analysis of plasmid coded gene products in E. coli minicells

The procedure described by Noyer-Weidner *et al.* (1985) was followed.

Immunoblotting

Crude extracts of cells harbouring plasmids coding for the various Mtases were prepared and subjected to the purification steps I (DEAE cellulose) and III (DNA cellulose) described by Günthert *et al.* (1986b) for the purification of the SPR Mtase. The analysis of the Mtases by Western blotting with anti SPR-serum was carried out as described by the same authors.

Acknowledgement

M.N.-W. was supported by grant Tr 25/10-1 of Deutsche Forschungsgemeinschaft.

References

Behrens,B., Noyer-Weidner,M., Pawlek,B., Lauster,R., Balganesh,T.S. and Trautner,T.A. (1987) *EMBO J.*, **6**, 1137−1142

Bolivar,F., Rodriguez,R.L., Greene,P.J., Betlach,M.C., Heyneker,H.L. and Boyer,H.W. (1977) *Gene*, **2**, 95−113.

Boyer,H.W. and Roulland-Dussoix,D. (1969) *J. Mol. Biol.*, **4**, 459−472.

Buhk,H.-J., Behrens,B., Tailor,R., Wilke,K., Prada,J.J., Günthert,U., Noyer-Weidner,M., Jentsch,S. and Trautner,T.A. (1984) *Gene*, **29**, 51−61.

Fuller-Pace,F.V. and Murray,N.E. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 9368−9372.

Gough,J.A. and Murray,N.E. (1983) *J. Mol. Biol.*, **66**, 1−19.

Günthert,U. and Reiners,L. (1987) *Nucleic Acids Res.*, **9**, 3689−3702.

Günthert,U. and Trautner,T.A. (1984) In Trautner,T.A. (ed.), *Methylation of DNA, Current Topics in Microbiology and Immunology*. Springer, Berlin, Vol. 108, pp. 11−22.

Günthert,U., Reiners,L. and Lauster,R. (1986a) *Gene*, **41**, 261−270.

Günthert,U., Lauster,R. and Reiners,L. (1986b) *Eur. J. Biochem.*, **159**, 485−492.

Kessler,C. and Höltke,H.-J. (1986) *Gene*, **47**, 1−153.

Kramer,W., Drutsa,V., Jansen,H.W., Kramer,B., Pflugfelder,M. and Fritz,H.-J. (1984a) *Nucleic Acids Res.*, **12**, 9441−9456.

Kramer,B., Kramer,W. and Fritz,H.-J. (1984b) *Cell*, **38**, 879−887.

Maniatis,T., Fritsch,E. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Nagaraja,V., Shepherd,J.C.W. and Bickle,T.A. (1985) *Nature*, **316**, 371−372.

Noyer-Weidner,M., Jentsch,S., Pawlek,B., Günthert,U. and Trautner,T.A. (1983) *J. Virol.*, **42**, 446−453.

Noyer-Weidner,M., Jentsch,S., Kupsch,J., Bergbauer,M. and Trautner,T.A. (1985) *Gene*, **35**, 143−150.

Noyer-Weidner,M., Diaz,R. and Reiners,L. (1986) *Mol. Gen. Genet.*, **205**, 469−475.

Ratcliff,S.W., Luh,J., Ganesan,A.T., Behrens,B., Thompson,R., Montenegro,M.A., Morelli,G. and Trautner,T.A. (1979) *Mol. Gen. Genet.*, **168**, 165−172.

Soberon,X., Covarrubias,L. and Bolivar,F. (1980) *Gene*, **9**, 287−305.

Tran-Betcke,A., Behrens,B., Noyer-Weidner,M. and Trautner,T.A. (1986) *Gene*, **42**, 89−96.

Trautner,T.A., Pawlek,B., Günthert,U., Canosi,U., Jentsch,S. and Freund,M. (1980) *Mol. Gen. Genet.*, **180**, 361−367.

Zoller,M.J. and Smith,M. (1983) *Methods Enzymol.*, **100**, 468−500.

Received on July 9, 1987; revised on August 7, 1987