The EMBO Journal vol.6 no.4 pp.1137–1142, 1987

Organization of multispecific DNA methyltransferases encoded by temperate *Bacillus subtilis* phages

B.Behrens, M.Noyer-Weidner, B.Pawlek, R.Lauster, T.S.Balganesh and T.A.Trautner

Max-Planck-Institut für molekulare Genetik, Ihnestrasse 73, D-1000 Berlin 33, FRG

Communicated by T.A.Trautner

B. subtilis **phage $\varrho 11_s$ codes for a multispecific DNA methyltransferase (Mtase) which methylates cytosine within the sequences GGCC and GAGCTC. The Mtase gene of $\varrho 11_s$ was isolated and sequenced. It has 1509 bp, corresponding to 503 amino acids (aa). The enzyme's M_r of 57.2 kd predicted from the nucleotide sequence was verified by direct M_r determinations of the Mtase. A comparison of the aa sequence of the $\varrho 11_s$ Mtase with those of related phages SPR and ϕ3T, which differ in their methylation potential, revealed generalities in the building plan of such enzymes. At least 70% of the aa of each enzyme are contained in two regions of 243 and 109 aa at the N and C termini respectively, which are highly conserved among the three enzymes. In each enzyme, variable sequences separate the conserved regions. Variability is generated through the single or multiple use of related and unrelated sequence motifs. We propose that the recognition of those DNA target sequences, which are unique for each of the three enzymes, is determined by these variable regions. Evolutionary relationships between the three enzymes are discussed.**

Key words: B. subtilis phages SPR, ϕ3T, $\varrho 11_s$/*Sac*I modification/enzyme domains/molecular evolution/DNA–protein interaction

Introduction

Sequence-specific DNA methylation is performed by DNA methyltransferases (Mtases). Such enzymes are DNA binding proteins with the capacity to recognize specific target sequences of a few nucleotides. Only in the presence of such sequences will the Mtase, at first non-specifically bound to DNA, mediate the transfer of a methyl group from S-adenosyl-methionine to a defined base within the recognition sequence. This property and the availability of a large number of Mtases with different specificities have made them, together with restriction enzymes, suitable proteins for a study of specificity recognition in DNA–protein interactions.

With the aim of describing domains of Mtases responsible for the various steps in DNA methylation, particularly for specificity recognition, two approaches have been initiated. A number of laboratories have cloned and sequenced genes of Mtases from different sources. Comparisons of the amino acid (aa) sequences derived showed the presence of some highly conserved regions within a large background of unrelated sequences (Hattman *et al.*, 1985; Kiss *et al.*, 1985; Mannarelli *et al.*, 1985). It is plausible to assume that conserved regions in such sequences represent domains, or parts thereof, which are involved in the same step of the methylation reaction. The inclusion in this comparison of enzymes which methylate the same or closely related sequences, as well as secondary structure predictions, have furthermore permitted tentative assignments of specificity-recognizing domains (Buhk *et al.*, 1984; Argos, 1985; R.Lauster *et al.*, in preparation).

Our laboratory's approach to the problem allows more specific predictions. For the last several years, we have been studying closely related multispecific Mtases from some temperate *Bacillus subtilis* bacteriophages. These enzymes recognize on the one hand the same sequence (GGCC), but differ in their capacity to recognize additional specificities (Buhk *et al.*, 1984; Günthert and Trautner, 1984; Günthert *et al.*, 1986a,b). In such cases we would anticipate domains responsible for general steps in the methylation reaction and for the common recognition of the sequence GGCC to reside in conserved regions. Domains involved in the recognition of the additional specificities must be identifiable through structural diversity. We have applied this rationale to a comparison of the aa sequences of two different Mtases derived from phages SPR and ϕ3T (Tran-Betcke *et al.*, 1986). In our comparison we could indeed identify, in addition to extensive regions of highly conserved aa sequences, several regions with sequence diversity, which would qualify as specificity-recognizing domains.

Here we describe the aa sequence of a related Mtase, derived from *B. subtilis* phage $\varrho 11_s$, with yet a different specificity. Expanding our previous comparison by the inclusion of the $\varrho 11_s$ sequence we can now report a considerably more reliable picture of the organization and evolution of these genes and their products.

Results

Cloning the Mtase gene of $\varrho 11_s$ and characterization of the gene product

The cytosine methylating Mtases of the temperate bacteriophages SPR, ϕ3T (SPβ, $\varrho 11_B$) were previously shown to be multispecific enzymes (Günthert and Trautner, 1984). These enzymes share the capacity to recognize the sequence GGCC (*Hae*III) as a methylating site. They differ in their ability to recognize additional target sequences which are CCGG (*Hpa*II and *Msp*I) and $CC^{A}_{T}GG$ (*Eco*RII) in SPR and GCNGC (*Fnu*4HI) in ϕ3T. The DNA of bacteriophage $\varrho 11_S$ was shown by Hemphill (personal communication) to be resistant to the GGCC- and GAGCTC-specific restriction endonucleases *Hae*III and *Sac*I, but sensitive to degradation by *Hpa*II/*Msp*I (CCGG) or *Fnu*4HI (GCNGC). The relatedness of $\varrho 11_S$ phage to the phages previously studied suggested that $\varrho 11_S$ also coded for a multispecific cytosine methylating Mtase, which generated the observed methylation pattern of its DNA. To identify the putative Mtase gene of $\varrho 11_S$, we cloned a partial *Hin*dIII digest of $\varrho 11_S$ DNA into pBR328. Selecting for *Hae*III-resistant recombinant plasmids, using the technique described by Szomolányi *et al.* (1980) we obtained plasmid pBN52 (Figure 1A), which contained a 3.15-kb fragment of $\varrho 11_S$ DNA. The inserted DNA carried

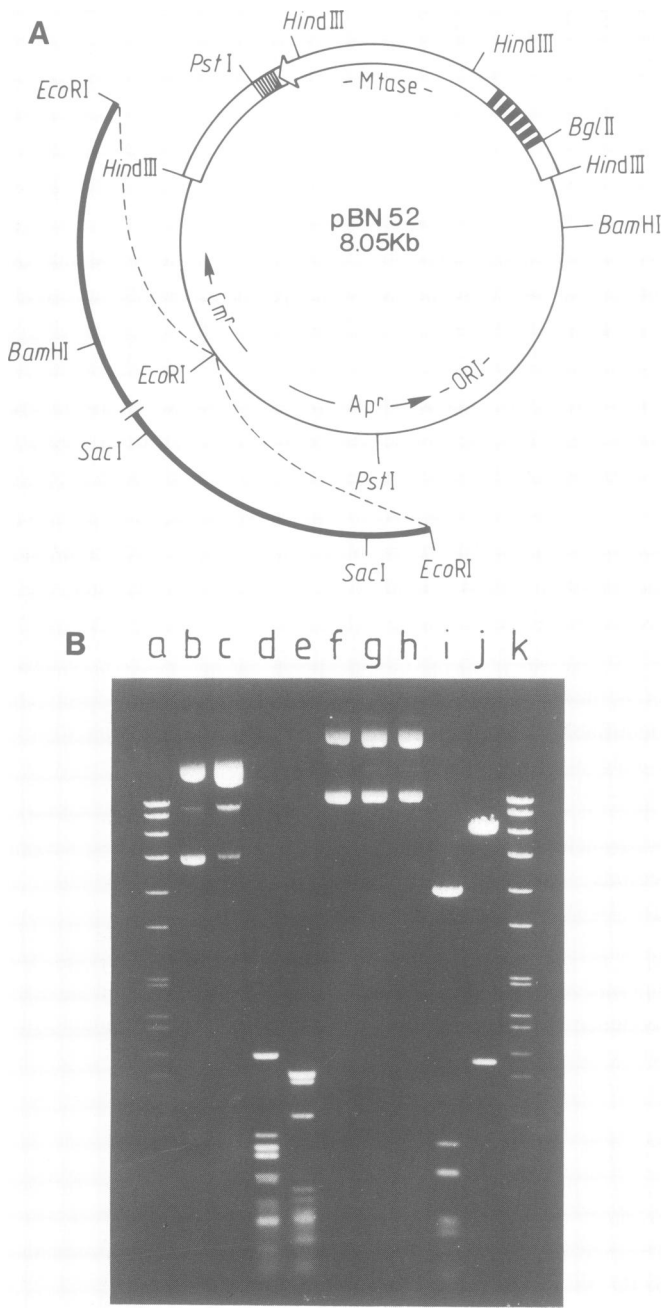

Fig. 1. (**A**) Schematic presentation of plasmid pBN52 and its derivatives pBK53 and pBK54. $\varrho 11_S$ DNA is shown as a double line, pBR328 as a single line, inserted λ DNA is symbolized by a heavy line. The $\varrho 11_S$ Mtase gene with its direction of transcription is shown as an arrow. Shadowing of $\varrho 11_S$ DNA surrounding the Mtase gene is the same as in Figure 7 of Tran-Betcke *et al.* (1986). Only restriction sites necessary to facilitate orientation are shown. The 4.9-kb of λ DNA inserted into the *Eco*RI site of pBN52 is shown (heavy line). The two *Bam*HI sites used in constructing the deleted plasmid pBK53 are represented. (**B**) Electropherogram in 1% agarose of pBN52, pBK53 and pBK54 with various restriction enzymes. **Lanes a** and **k**: SPP1 DNA digested with *Eco*RI (M_r standard, Ratcliff *et al.*, 1979); **lane b**: pBN52 uncut; **lanes c**, **g** and **i**: pBN52, pBK53 and pBK54 digested with *Hae*III; **lanes d** and *e*: pBN52 digested with *Msp*I and *Fnu*4HI respectively; **lane f**: pBK53 uncut; **lanes h** and **j**: pBK53 and pBK54 digested with *Sac*I.

the anticipated $\varrho 11_S$ Mtase gene, since plasmid pBN52 was resistant against degradation by *Hae*III and degradable with *Hpa*II/*Msp*I or *Fnu4*HI, whose recognition sites are methylated in corresponding clones from SPR or φ3T respectively (Figure

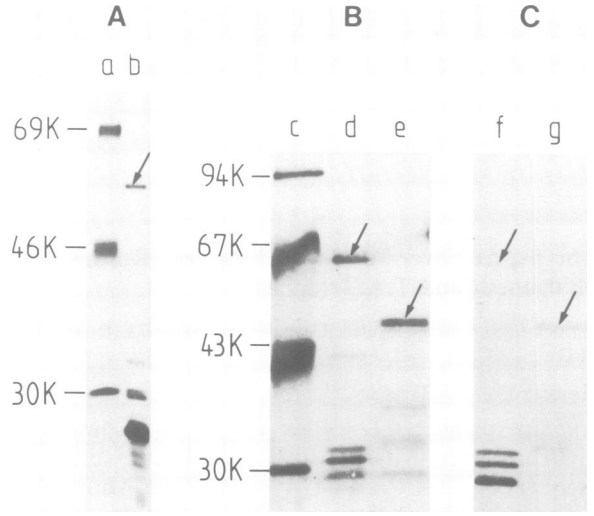

Fig. 2. (**A**) Analysis by SDS−PAGE (12.5%, 16 h, 110 V) of gene products synthesized in *E. coli* minicells. **Lane a**: M_r standard; **lane b**: autoradiogram of [^{35}S]methionine-labelled proteins specified by plasmid pBN52. (**B**) Silver-stained SDS−PAGE (12.5%, 3 h, 170 V) of the $\varrho 11_S$ and SPR Mtases: purified $\varrho 11_S$ (**lane d**) and SPR (**lane e**) Mtases were applied. **Lane c** is a M_r standard. (**C**) Western blot analysis of $\varrho 11_S$ and SPR Mtase with anti-SPR-Mtase antiserum. **Lanes f** and **g** contain the same fractions that were applied to **lanes d** and **e**. Arrows indicate the positions of the $\varrho 11_S$ (58 kd) and SPR (49 kd) Mtases.

1B). pBN52 has no *Sac*I site. To verify also the presence of a *Sac*I-recognizing Mtase activity, we have inserted a 4.9-kb *Eco*RI fragment of λ DNA containing two *Sac*I sites into pBN52. This plasmid, pBK53, was resistant to *Sac*I degradation, whereas a derivative plasmid, pBK54, with a deletion extending between the two *Bam*HI sites, causing loss of the $\varrho 11_S$ Mtase gene, was *Sac*I sensitive and showed the anticipated *Sac*I restriction pattern (Figure 1B).

Gene expression of plasmid pBN52 in minicells showed, together with the bands found with pBR328 alone, a single additional protein band corresponding to a M_r of 58 000 (Figure 2A). To verify the identity of the 58-kd protein with the $\varrho 11_S$ Mtase, we have purified this enzyme from IPTG-induced *Escherichia coli* containing plasmid pBL1. pBL1 is a derivative of pUC18, containing a 2.8-kb *Bgl*II−*Pst*I fragment with the $\varrho 11_S$ Mtase gene brought under the control of the *lac*Z promoter. Enzyme purification followed essentially the procedure previously described for the SPR-Mtase (Günthert *et al.*, 1986b). In contrast, however, to the latter enzyme, the $\varrho 11_S$ Mtase proved to be extremely unstable during purification and storage. Purified enzyme obtained after the gel filtration step of the purification procedure was subjected to SDS−PAGE, where a prominent band corresponding to a protein of 58 kd was identifiable (Figure 2B). A Western blot with this material, using anti-(SPR-Mtase) antibodies also revealed staining of this band and several smaller M_r bands, which probably represent degradation products of the enzyme. The minicell results, together with the Western blot, demonstrate that the $\varrho 11_S$ Mtase has a M_r of 58 kd. The enzyme is hence considerably larger than those of SPR and φ3T, which have M_rs of 49 kd. The fact that the 58-kd enzyme band can also be recognized after immunoblotting using anti-(SPR-Mtase) antibodies, which also react with φ3T-Mtase, indicates a close structural relationship between these enzymes. Properties of the $\varrho 11_S$ Mtase — in particular the extreme instability of the enzyme — are presently under investigation.

```
   1 AGATCTGCCA GAGGTCGGAG AATACTTCTA CGTTAATCGG GGGGGTTGGG GTAAGTGTTT GGTGAAAAGT GTTTTCGGAC TCGAAGATGT ATCGCCGACT

 101 ATAAAAAGTC AAAACGCAAA TCTTCCTCCT GGTTATCCCG GACACCCTAA AGACGCCGAT CAAATCTGCA ACCGACCGAT TAGCCCAACA AAAGCGCCCC

 201 GCCGTTTCAC CGTTCGCGAA TGTCTCCGAA TCCAATCTGC GCCAGACACT TACGTCCTGC CAGACGATAT TTCCTTAACG GCGCAGTATC GGATTGTCGG
                                                                                                          -35
 301 TAACGGAATT GCTTCGCGTG TTGCGTGGTA TATCGGGCGG GCGCTTGCGG ATCAACTTTC ATCTAACTGT AAGTAATTAT GATAGATAAG ATGGTTGAAA
                          -10                                S.D.
 401 TATTACATAA AAGATCAATT TTATTTAGAA TAAAAATAAA ATAAATGGAG GTTGTTTA TTG AGT AAA CTA CGA GTA ATG AGT CTT TTT AGT
   1                                                                  M   S   K   L   R   V   M   S   L   F   S

 492 GGA ATC GGT GCG TTT GAA GCT GCA CTA AGA AAC ATT GGG GTT GAA TAT GAA CTG GTT GGC TTT AGT GAA ATT GAT AAA TAT
  12 G   I   G   A   F   E   A   A   L   R   N   I   G   V   E   Y   E   L   V   G   F   S   E   I   D   K   Y

 573 GCT ATT AAG TCA TAT TGT GCG ATT CAC AAC GCT GAT GAG CAA TTA AAT TTT GGT GAT GTA AGC AAG ATT GAT AAG AAA AAA
  39 A   I   K   S   Y   C   A   I   H   N   A   D   E   Q   L   N   F   G   D   V   S   K   I   D   K   K   K

 654 CTG CCT GAA TTT GAT CTT TTA GTA GGA GGA TCT CCT TGT CAA AGC TTT AGT GTA GCC GGT TAT CGG AAA GGG TTC GAA GAT
  66 L   P   E   F   D   L   L   V   G   G   S   P   C   Q   S   F   S   V   A   G   Y   R   K   G   F   E   D

 735 ACT AGA GGG ACA TTG TTT TTT CAG TAT ATA GAT ACC TTA AAA GAG AAG CAG CCG CGG TAT TTT GTT TTT GAA AAT GTT AAA
  93 T   R   G   T   L   F   F   Q   Y   I   D   T   L   K   E   K   Q   P   R   Y   F   V   F   E   N   V   K

 816 GGA TTG ATC AAT CAT GAT AAA GGA AAT ACA TTA AAT ATT ATG GCT GAA TCT TTT AGT GAA GTT GGG TAC AGA ATT GAT TTA
 120 G   L   I   N   H   D   K   G   N   T   L   N   I   M   A   E   S   F   S   E   V   G   Y   R   I   D   L

 897 GAG CTA CTT AAT TCA AAA TTT TTC AAT GTT CCT CAG AAT CGT GAG CGT ATA TAC ATA ATT GGA GTT CGT GAG GAT TTA ATT
 147 E   L   L   N   S   K   F   F   N   V   P   Q   N   R   E   R   I   Y   I   I   G   V   R   E   D   L   I

 978 GAA AAT GAC GAA TGG GTT GTC GAA AAG GGA AGG AAC GAT GTT TTA AGT AAG GGA AAA AAG AGA TTA AAA GAA TTA AAT ATA
 174 E   N   D   E   W   V   V   E   K   G   R   N   D   V   L   S   K   G   K   K   R   L   K   E   L   N   I

1059 AAA AGT TTC AAT TTT AAA TGG TCT GCA CAA GAT ATT GTT GGA AGG AGA TTG CGA GAA ATT CTT GAG GAA TAT GTA GAT GAG
 201 K   S   F   N   F   K   W   S   A   Q   D   I   V   G   R   R   L   R   E   I   L   E   E   Y   V   D   E

1140 AAG TAT TAC TTA AGC GAA GAA AAA ACA TCT AAA CTG ATT GAA CAA ATT GAA AAA CCA AAA GAA AAA GAT GTG GTG TTT GTT
 228 K   Y   Y   L   S   E   E   K   T   S   K   L   I   E   Q   I   E   K   P   K   E   K   D   V   V   F   V

1221 GGT GGC ATT AAC GTA GGA AAG AGG TGG CTG AAT AAC GGA AAA ACA TAT TCC AGA AAC TTT AAA CAA GGC AAT AGA GTC TAT
 254 G   G   I   N   V   G   K   R   W   L   N   N   G   K   T   Y   S   R   N   F   K   Q   G   N   R   V   Y

1302 GAT TCA AAT GGC ATT GCA ACA ACT TTG ACA TCC CAA TCG GTC GGT GGT CTT GGA GGG CAG ACT TCG CTA TAC AAA GTG GAA
 282 D   S   N   G   I   A   T   T   L   T   S   Q   S   V   G   G   L   G   G   Q   T   S   L   Y   K   V   E

1383 GAC CCG ATC ATG ATT GGT CAC ATT GAT CTA AAA GGA CAT GAT GCA ATT AAA AGA GTG TAC TCG CCT GAT GGG GTG TCA CCA
 309 D   P   I   M   I   G   H   I   D   L   K   G   H   D   A   I   K   R   V   Y   S   P   D   G   V   S   P

1464 ACA TTA ACA ACT ATG GGA GAA GGT CAT AGA GAA CCT AAA ATT GCT GTT GAG TAT GTT GGT AAT ATT AAT CCT TCA GGA AAA
 336 T   L   T   T   M   G   E   G   H   R   E   P   K   I   A   V   E   Y   V   G   N   I   N   P   S   G   K

1545 GGA ATG AAT GAC CAG GTT TAC AAT TCA AAT GGG CTA AGT CCA ACC TTA ACA ACA AAT AAG GGT GAA GGG GTG AAA ATT TCT
 363 G   M   N   D   Q   V   Y   N   S   N   G   L   S   P   T   L   T   T   N   K   G   E   G   V   K   I   S

1626 GTG CCA AAC CCT GAA ATA AGA CCC GTC TTA ACT CCA GAA AGG GAA GAA AAA AGA CAA AAT GGC AGA CGC TTT AAA GAA GAC
 390 V   P   N   P   E   I   R   P   V   L   T   P   E   R   E   E   K   R   Q   N   G   R   R   F   K   E   D

1707 GAT GAA CCA GCC TTC ACT GTT AAT ACA ATT GAT CGT CAT GGT GTA GCA ATT GGC GAA TAT CCA AAA TAC AGA ATT AGA AAG
 417 D   E   P   A   F   T   V   N   T   I   D   R   H   G   V   A   I   G   E   Y   P   K   Y   R   I   R   K

1788 CTC ACT CCA TTA GAA TGT TGG AGG CTG CAA GCA TTT GAT GAA GAA GAT TTT GAG AAA GCT TTA TCA GTG GGA ATT AGT AAT
 445 L   T   P   L   E   C   W   R   L   Q   A   F   D   E   E   D   F   E   K   A   L   S   V   G   I   S   N

1869 TCG CAG TTG TAC AAG GAA GCC GGC AAT TCA ATT ACT GTA ACT GTA CTT GAG TCA ATA TTC AAG GAA TTA ATA CAT ACA TAC
 471 S   Q   L   Y   K   E   A   G   N   S   I   T   V   T   V   L   E   S   I   F   K   E   L   I   H   T   Y

1950 GTT AAT GAA GAA TCT GAA TAA AATTTGTCTT TTAAACAAAT GCAAAATAAG AGGGATTGAG AGGTGAGGAA GATGGATAGT TACCCAGAGT
 498 V   N   E   E   S   E   *

2041 CTTTAAAAAG AGAGACAGAG GAGATTAAAG AGCGTGTTAG GAATGGAAAT ATCAAAGAAG ACAGGATTAA AGAAATTGCA GAAACGACAG TTGAGTTTTT

2141 GAAATCAGAG GAGAAAAGAC ATAAATACTT TTCTGAAGTT GCTGCAG
```

Fig. 3. Nucleotide sequence and derived aa sequence of the *Bgl*II-*Pst*I DNA fragment of $\varrho 11_S$, containing the Mtase gene (see Figure 1). Nucleotides 1–429 of the $\varrho 11_S$ sequence were identical to nucleotides 257–686 of the published ϕ3T sequence (Tran-Betcke *et al.*, 1986). This comparison also eliminated three errors in our previous communication of the ϕ3T sequence: one C each had to be added to the ϕ3T sequence at locations 441 and 599; the A at location 551 had to be eliminated from the sequence. S.D. describes the ribosomal binding site. '−10' and '−35' are sequences which might represent sequences involved in transcriptional regulation.

The nucleotide and derived amino acid sequence of the $\varrho 11_S$ Mtase gene

Appropriate overlapping restriction fragments from the $\varrho 11_S$ DNA of plasmid pBN52 were cloned into M13mp8 or M13mp9 and sequenced by the Sanger dideoxy method (Sanger *et al.*, 1977), using commercially available universal primers. The nucleotide sequence of the Mtase containing the *Bgl*II–*Pst*I fragment of plasmid pBN52 and the deduced aa sequence is shown in Figure 3. The largest open reading frame, extending 5′ to 3′ from left to right in Figure 3, contains the translational initiation codon TTG at position 459 and ends at the termination codon TAA at position 1968. It has 1509 bp corresponding to 503 aa with a calculated M_r of the protein of 57 166.31. This value

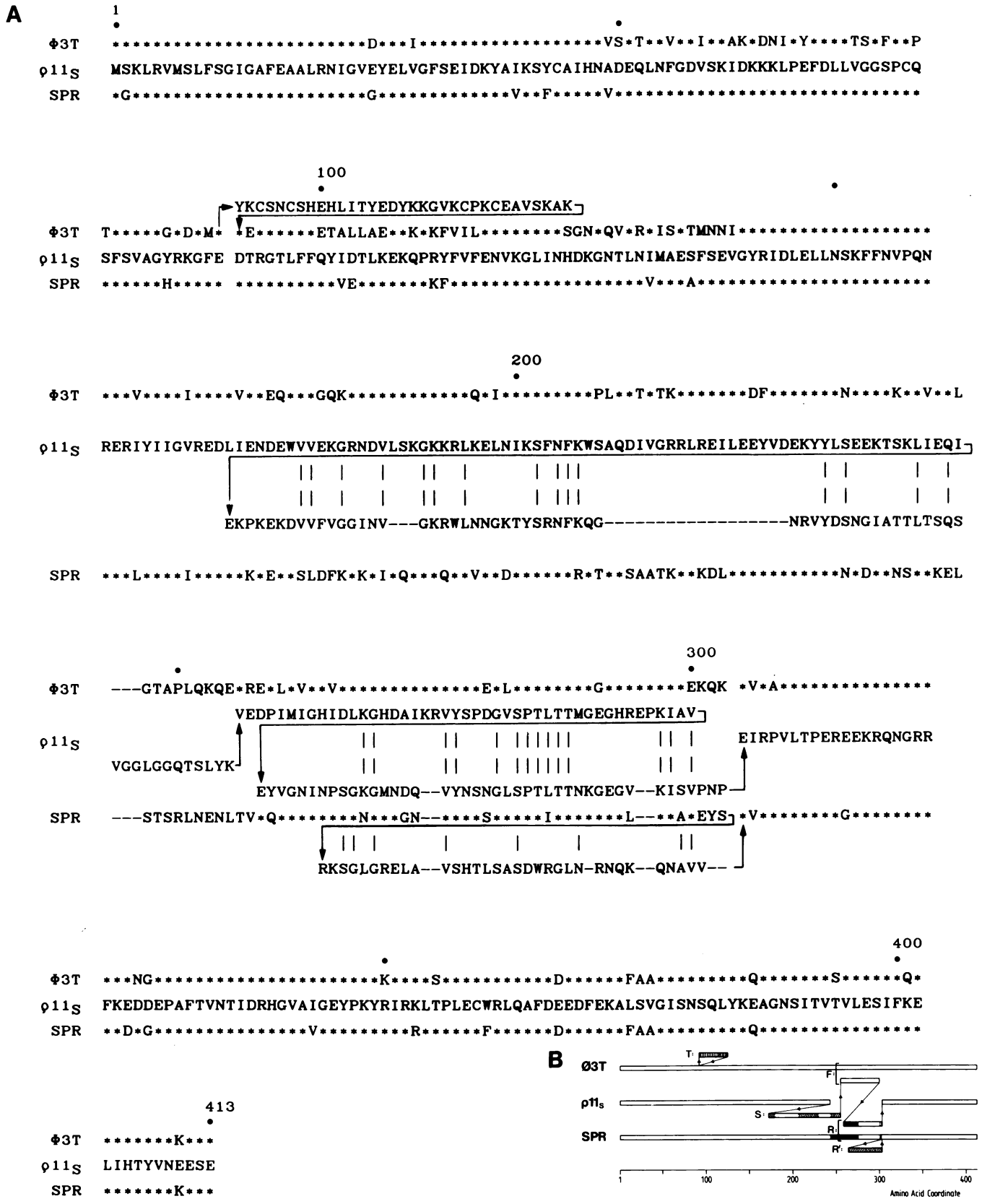

Fig. 4. Maximal alignment of aa sequences of the $\varrho 11_S$ aa sequence with those of ϕ3T and SPR. (**A**) Juxtaposition of aa sequences. The sequences of SPR and ϕ3T are from Buhk *et al.* (1984) and from Tran-Betcke *et al.* (1986). aa in ϕ3T and SPR identical to those of $\varrho 11_S$ are entered as asterisks (*). Hyphens (−) describe gaps introduced to maximize alignments. Vertical lines connecting identical aa show intramolecular homology. The numbers above the sequences describe aa alignment coordinates. (**B**) Schematic presentation of the aa alignment of Figure 4A. Regions symbolized by the same shadowing represent homologous regions where aa sequences do not differ by >5 consecutive, non-conservative exchanges. T, F, S, R and R′ indicate aa sequence blocks which lead to variations characteristic for each enzyme.

agrees well with those obtained of the purified enzyme and by minicell-analysis (Figure 2).

The nucleotide sequence of DNA flanking the $\varrho 11_S$ Mtase gene and the corresponding sequences of ϕ3T (Tran-Betcke *et al.*, 1986) are virtually identical (Figure 3). The upstream region includes a sequence of 52 bp immediately preceding the start codon TTG, which is highly conserved in all phage DNAs at this location. Here the $\varrho 11_S$ sequence differs only by three exchanges from those of ϕ3T/SPβ or SPR. The sequence includes the ribosomal binding site GGAGG and a tentatively assigned −10 region (TTTATT) of a potential promoter. Extending further 5′ from the 52 bp region, where the DNA sequences of ϕ3T, SPβ and SPR were previously shown to be unrelated (Tran-Betcke *et al.*, 1986), the $\varrho 11_S$ sequence follows that of ϕ3T. Also beyond the 3′ end of the gene, where the ϕ3T, SPR and Z sequences were distinguishable through only slight differences, the sequenced portion of the $\varrho 11_S$ DNA was identical to that of ϕ3T. Within this sequence of 219 bp neither phage DNA contained a conventional transcriptional terminator signal. Direct measurements of *in vivo* generated phage mRNA to define transcriptional start and terminus are pending to determine whether the Mtase transcripts are part of a polycistronic mRNA.

Discussion

To compare the organization of the phage Mtase genes, we have aligned in Figure 4 the derived aa sequence of $\varrho 11_S$ with those of ϕ3T and SPR, which we have determined before (Buhk *et al.*, 1984, Tran-Betcke *et al.*, 1986). The aa sequences have been placed in register to give maximal inter- and intramolecular aa homology. We can identify two large contiguous regions of aa homology among the three enzymes extending between the N terminus and the aa at coordinate 243 and between the aa at position 304 and the C terminus. More than three-quarters of the enzymes' aa fall into these regions of homology. In addition we recognize two areas where the aa sequences cannot be readily aligned. One is a sequence of 33 aa (T of Figure 4B), only present in the ϕ3T gene at coordinate 92. The other zone is represented by sequences F, S, R and R′ (Figure 4B), which extend between the N- and C-terminated blocks of homology. The difference in the number of aa in this region of the three enzymes indicates that the spacing of the N- and C-terminated homology blocks cannot be a critical parameter in providing Mtase activity. The organization of aa separating the conserved regions is such that nine (ϕ3T and SPR) or 63 ($\varrho 11_S$) nonmatched aa extending from coordinate 244, and three nonmatched aa preceding aa 304, bracket an arrangement of aa which is characterized by the use of two related sequence motifs. One of these (F) is present only once in the ϕ3T Mtase and at the equivalent position in $\varrho 11_S$. At the corresponding position the SPR Mtase has a different, though related motif (R). In $\varrho 11_S$ the ϕ3T-type motif (F) is followed by the SPR-type motif (R), whereas the SPR Mtase contains at this location an aa sequence (R′), which only shows marginal relationship to the SPR-type motif.

What is the significance of the various regions identified for the Mtase activity, particularly for their capacity to recognize different sequence specificities? The identity in the three enzymes of the aa sequence in the C-terminated homology region excludes the possibility that this region is decisive in the recognition of target sequences, which are unique to each enzyme. Similarly the finding of identity of aa in the N-terminal portion of the $\varrho 11_S$ and SPR Mtase genes excludes the possibility that this region is critical in the recognition of the $\varrho 11_S$- and SPR-characteristic methylation specificities, GAGCTC and CCGG. Recent experiments from our laboratory, in which ϕ3T/SPR chimeric Mtases were constructed (T.S.Balganesh *et al.*, in preparation), have also excluded the possibility that determinants for the ϕ3T-specific recognition of GCNGC reside in the equivalent portion of the ϕ3T sequences by the presence of the 33-bp insert. These observations and considerations assign domains responsible for general steps of the methylation reaction to the conserved areas and at least to the essential parts, if not the whole domain responsible for recognition of unique specificities to the aa sequences extending between coordinates 244 and 304 in Figure 4. On the other hand, the determinants for the recognition of GGCC, which is common to all enzymes, is apparently located in the conserved C-terminated homology region. This follows from the previously reported location of specificity mutations in the SPR Mtase gene (Buhk *et al.*, 1984) and more recent mutagenesis experiments (Wilke *et al.*, in preparation). The structural basis for the variable regions between the two homologies serving in specificity recognition is difficult to assess on the basis of only the primary sequence, which is presented here.

With respect to the unique insert T (Figure 4B) of 33 aa of the ϕ3T Mtase, we have noted that it consists essentially of the general aa sequence $CX_2CX_{16}CX_2C$, which is strikingly similar to metal binding sequences which have recently been identified in a large variety of DNA binding proteins (Berg, 1986; Giedroc *et al.*, 1986). In one case it was shown that such an aa sequence can form, in the presence of Zn^{2+} ions, a protein 'finger', which is apparently involved in DNA binding (Miller *et al.*, 1985). The ϕ3T Mtase is the first example where such a sequence has been identified in a DNA binding protein which is a Mtase. In the absence of experiments directed at specifically modifying this domain we cannot establish whether or at what level of Mtase/DNA interaction this domain is effective.

The establishment of the nucleotide and derived aa sequences of the three phage Mtases invites speculation about their evolution. With the great similarity of the three enzymes it is plausible to assume that their genes have evolved from the same ancestral gene. Such a gene could have encoded a GGCC recognizing Mtase, of which essential sequences, including those which determined domains responsible for the methylation at large, have remained conserved in the present enzymes. Possibly the pressure on the phages to expand their host range into bacterial strains with other restriction systems and at the same time the pressure to economize with genome size caused divergence of evolution by the acquisition of additional, though different, methylation potentials. Determinants for such specificities, characterized by intramolecular homology, were apparently generated in the SPR and $\varrho 11_S$ Mtases through sequential duplications and rearrangements of related sequence motifs (Figure 4). The fact that this sequence variability occurs in all enzymes at the same location would indicate that specificity-determining sequence diversity, both in quality and number of aa, can be tolerated in this part of the enzyme.

The comparison of the nucleotide and aa sequences in the conserved and variable regions gives some indication of the evolutionary relationship between the three enzymes. Looking at the homology regions, we can identify a few, mostly conservative, aa exchanges. These and some diversities of codon usage for one and the same aa are responsible for the differences in the nucleotide sequence of the conserved areas, pointing to

divergence in their evolution. Such divergence, however, has affected these regions differently: among the 110 C-terminal aa, conservation is close to 90%. Within the N-terminal region both the nucleotide and aa sequences of $\varrho 11_S$ and SPR are more closely related to each other than to those of ϕ3T, which is in addition distinguished by the presence of the potential DNA binding finger T of 33 aa. Analyzing the DNA sequences described before (Buhk *et al.*, 1984; Tran-Betcke *et al.*, 1986) of the variable region between the blocks of homology, we have observed considerable degeneracy in codon usage between the sequence motifs F and R shared by the three Mtases: out of 37 identical aa in corresponding regions F of ϕ3T and $\varrho 11_S$ 17 used different codons. Similarly the same codons were used for only 12 out of 28 aa which are identical in motifs R shared by SPR and $\varrho 11_S$. Five out of a total of eight aa which are identical between the SPR motifs R and R′ used different codons. Among 14 aa shared between motifs F and R within $\varrho 11_S$ only six were identical. Of 15 identical aa between segment S and the corresponding sequence $\varrho 11_S$ only four aa had identical codons. These results, like those obtained with the N- and C-terminal aa sequences, suggest also that the central arrangements of the Mtase genes were established at a time which allowed subsequent divergent evolution.

In interspecific comparisons of related proteins, codon degeneracy could often be correlated with species-specific codon preferences. This explanation is not applicable to this comparison of related proteins of phages which have the same host organism. In organisms like *B. subtilis*, which have very potent recombination systems, codon degeneracy could be an instrument to maintain repetitious aa sequences.

Our results provide the first, albeit rather coarse, definition of the region involved in specificity recognition of a Mtase. We shall provide a refinement of our analysis from the application of genetic analyses, which are presently under way in our laboratory. These approaches, combined with three-dimensional analysis through X-ray crystallography of enzymes interacting with DNA substrate should provide structural understanding of specificity recognition by these Mtases.

Materials and methods

Bacterial strains and phages

Bacteriophage $\varrho 11_S$ was kindly provided by E.Hemphill (Syracuse, NY). We have called the phage $\varrho 11_S$ to distinguish it from another $\varrho 11$ phage now called $\varrho 11_B$, which had previously been shown in our laboratory to have the same methylation potential as phage ϕ3T and SPβ (Noyer-Weidner *et al.*, 1983). *B. subtilis* strain SB1207 (Trautner *et al.*, 1980) was used for the plating of phages and in transformation experiments. The *rglB*$^-$ *E. coli* strains HB101 (Boyer and Roulland-Dussoix, 1969) and TC600 (Noyer-Weidner *et al.*, 1986) were used in cloning experiments. The minicell-producing strain, TC410 (*rglB*$^-$), derived from DS410 (Dougan and Sherrat, 1977) served for the analysis of the $\varrho 11_S$ Mtase gene product.

Plasmid constructions

E. coli plasmid cloning vectors were pBR328 (Soberon *et al.*, 1980) and pUC18 (Yanisch-Perron *et al.*, 1985). A pBR328 derivative (pBN52) containing the $\varrho 11_S$ Mtase gene was obtained using the technique described by Szomoloányi *et al.* (1980). pBR328 was digested with *Hin*dIII and ligated in the presence of a partial *Hin*dIII digest of $\varrho 11_S$ DNA. After transformation of HB101 with the ligation mixture, plasmid DNA was isolated from Ap^R transformants and incubated with an excess of *Hae*III. *Hae*III-resistant plasmids, one of which, pBN52, was used in all further studies, contained a 3.15-kb fragment of $\varrho 11_S$ DNA with the $\varrho 11_S$ Mtase (Figure 1A). A 4.87-kb *Eco*RI fragment of λ DNA with two *Sac*I sites was inserted at the *Eco*RI site of pBN52 (shown as insert in Figure 1A). The plasmid was designated pBK53. pBK54 was constructed by deleting the DNA between the two *Bam*HI sites of pBK53. The $\varrho 11_S$ Mtase gene on the 2.8-kb *Bgl*II–*Pst*I fragment was cloned behind the *lacZ* promoter of pUC18. The plasmid was designated pBL1.

Enzyme purification and immunoblotting

E. coli strain TC600 harboring the plasmid pBL1 was grown to an OD_{560} of 0.4 followed by induction with 1 mM IPTG. The protein purification and immunoblotting were performed essentially following the procedure described for the purification of the SPR Mtase (Günthert *et al.*, 1986b).

Reagents and general techniques

Restriction enzymes, T4 DNA ligase, *Pol*I (Klenow) and alkaline phosphatase were purchased from Boehringer (Mannheim, FRG) and New England Biolabs (Beverly, MA). [^{35}S]Methionine and [^{35}S]dATP were obtained from Amersham-Buchler (UK). *E. coli* plasmids were isolated by the procedure of Birnboim and Doly (1979). *E. coli* plasmid transformation and other techniques were following standard procedures (Maniatis *et al.*, 1982). Characterization of labelled proteins from minicells was done as described by Noyer-Weidner *et al.* (1985). DNA sequencing was carried out following the Sanger dideoxy chain termination method (Sanger *et al.*, 1977). Sequence evaluation was based on computer programs of Staden (1978) and Isono (1982).

Acknowledgements

We thank Dr E.Hemphill for bacteriophage $\varrho 11_S$ and the communication of unpublished results, K.Wilke for constructing plasmids pBK53 and pBK54, and P.A.Terschüren for Southern blot analyses. The technical assistance of Luzia Reiners-Schramm is also gratefully acknowledged. M.Noyer-Weidner was supported by grant Tr 25/10-1 of Deutsche Forschungsgemeinschaft.

References

Argos,P. (1985) *EMBO J.*, **4**, 1351–1355.
Berg,J.M. (1986) *Science*, **232**, 485–487.
Birnboim,H.C. and Doly,J. (1979) *Nucleic Acids Res.*, **7**, 1513–1523.
Boyer,H.W. and Roullard-Dussoix,D. (1969) *J. Mol. Biol.*, **41**, 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.
Dougan,G. and Sherratt,D. (1977) *Mol. Gen. Genet.*, **151**, 151–160.
Giedroc,D.P., Keating,K.M., Williams,K.R., Konigsberg,W.H. and Coleman,J.E. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 8452–8456.
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.
Hattman,S., Wilkinson,J., Swinton,D., Schlagman,S., McDonald,P.M. and Mosig,G. (1985) *J. Bacteriol.*, **164**, 932–937.
Isono,K. (1982) *Nucleic Acids Res.*, **10**, 85–89.
Kiss,A., Posfai,G., Keller,C.C., Venetianer,P. and Roberts,R.J. (1985) *Nucleic Acids Res.*, **13**, 6403–6421.
Maniatis,T., Fritsch,D.F. and Sambrook,J. (eds) (1982) *Molecular Cloning. A Laboratory Manual*. Cold Spring Habor Laboratory Press, NY.
Mannarelli,B.M., Balganesh,T.S., Greenberg,B., Springhorn,S.S. and Lacks,S.A. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 4468–4472.
Miller,J., McLacklan,A.D. and Klug,A. (1985) *EMBO J.*, **4**, 1609–1614.
Noyer-Weidner,M., Jentsch,S., Pawlek,B., Günthert,U. and Trautner,T.A. (1983) *J. Virol.*, **46**, 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-Schramm,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.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Soberon,X., Covarrubias,L. and Bolivar,F. (1980) *Gene*, **9**, 287–305.
Staden,R. (1978) *Nucleic Acids Res.*, **5**, 1013–1015.
Szomolányi,E., Kiss,A. and Venetianer,P. (1980) *Gene*, **10**, 219–225.
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.
Yanisch-Perron,C., Vieira,J. and Messing,J. (1985) *Gene*, **33**, 103–119.

Received on December 29, 1986; revised on February 4, 1987

Note added in proof

These sequence data have been submitted to the EMBL/GenBank Data Libraries under the accession number Y00053.