Nucleic Acids Research, 2002, Vol. 30, No. 7 **1547–1557**

Circular permutation of DNA cytosine-N4 methyltransferases: *in vivo* coexistence in the *Bcn*I system and *in vitro* probing by hybrid formation

Giedrius Vilkaitis, Arvydas Lubys*, Eglė Merkienė, Albertas Timinskas, Arvydas Janulaitis and Saulius Klimašauskas

Institute of Biotechnology, Graičiūno 8, LT-2028 Vilnius, Lithuania

Received December 18, 2001; Revised and Accepted February 4, 2002 DDBJ/EMBL/GenBank accession no. AF472611

ABSTRACT

Sequence analysis of the *Bcn*I restriction-modification system from *Bacillus centrosporus* revealed four open reading frames (*bcnIC*, *bcnIR*, *bcnIB* and *bcnIA*) that are arranged as two converging collinear pairs. One pair encodes a putative small regulatory protein, C.*Bcn*I, and the restriction endonuclease R.*Bcn*I. The other two gene products are the DNA cytosine-N4 methyltransferases M.*Bcn*IA and M.*Bcn*IB, which differ by circular permutation of conserved sequence motifs. The *Bcn*I methyltransferases are isospecific on double-stranded DNA [methylation specificity C<u>C</u>(C/G)GG], but M.*Bcn*IA can also methylate the target sites in single-stranded DNA. Functional analysis shows that *bcnIA* is dispensable (*bcnIB* is capable of protecting the DNA against the *in vivo* activity of *bcnIR*); in contrast, no stable clones were obtained if *bcnIB* alone was deleted from the system. By analogy with the *Dpn*II system, the second methylase M.*Bcn*IA may play a role in the transformation proficiency of its gram-positive host. The interchangeability of homologous elements in the β class of cytosine-N4 methylases was probed by hybrid formation between M.*Bcn*IB and its closest homolog M.*Cfr*9I (C<u>C</u>CGGG) employing a novel semi-random strategy combined with selection for catalytic activity. The fusion points in the active hybrids mapped in a narrow region located between sequence motifs X and I. Our data illustrate that recombination of two related sequences by circular permutation may serve as an evolutionary mechanism for creating new specificities of amino MTases.

INTRODUCTION

Restriction-modification (R-M) systems are widespread across the full spectrum of bacterial microorganisms (1). Among the major functional roles of R-M systems are the control of inter- and intraspecies exchange of genetic material ('immigration control') (2) as well as the mobility and maintenance of their own DNA ('selfish behavior') (3). R-M systems have been divided into three major types (I, II and III) on the basis of their biochemical properties and gene organization (for a review see 4). Type II R-M systems are the simplest and are usually composed of a single methyltransferase (MTase) and its counterpart restriction endonuclease (ENase). Both enzymes act on the same short recognition sequence (4–8 bp long). The ENase cleaves DNA if neither strand of the recognition site is methylated, while the MTase serves to protect the cellular DNA against this action by methylation of a specific nucleotide.

The *Bcn*I R-M system with two MTases (5) has been identified in *Bacillus centrosporus* strain RFL1. The *Bcn*I ENase cleaves the sequence 5′-C<u>C</u>↓(C/G)GG-3′ at the position indicated by the arrow (6), whereas a DNA modification activity isolated from either the wild-type strain or from *Escherichia coli* cells harboring the cloned MTase genes (7–9) was found to modify the second cytosine (underlined), yielding N4-methylcytosine (10). The complete *Bcn*I R-M system has been cloned and expressed in *E.coli* (11). In the present work we report sequence analysis of the complete *Bcn*I R-M system and demonstrate that both modification enzymes found in the *Bcn*I R-M system are isospecific (modify the same cytosine residue). Functional and evolutionary implications of the dual *Bcn*I MTases are discussed in the light of our recent observation that the enzymes exhibit different preferences for double-stranded or single-stranded DNA (5).

Each of the more than 100 known type II MTase sequences can be attributed to one of four main groups (m5C, α, β and γ), which differ in the linear arrangement of characteristic motifs (12). However, despite the altered linear order of the sequence motifs, DNA MTases share the same tertiary fold of the AdoMet-binding pocket and the catalytic site (13–17). The evolutionary origin of this circular permutation of DNA MTases is not yet understood. In contrast to cytosine-5 MTases, where the structural similarity of proteins was demonstrated directly by domain swapping (18,19), analogous experiments involving representatives of the other groups have not been reported. In this work we have probed the structural peculiarities of the β group of MTases by hybrid formation between M.*Bcn*IB and its closest homolog, M.*Cfr*9I. The results

*To whom correspondence should be addressed. Tel: +370 2 602105; Fax: +370 2 602116; Email: lubys@ibt.lt

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

of this analysis and their implications for the evolutionary mechanism of circular permutation of MTases are discussed.

MATERIALS AND METHODS

Bacterial strains and media

Escherichia coli K-12 strain ER1727 [F′ *proA*$^+$*B*$^+$ *lacI*^{q} Δ(*lacZ*)*M15*/*trp-31 his-1 rpsL104* (Str^r) Δ(*lacZ*)*r1 glnV44 xyl-7 mtl-2* Δ(*mcrBC-hsdRMS-mrr*)2::Tn*10 mcrA1272::*Tn*10* (Tet^r) *metB1 fhuA2*] was used as a host in DNA subcloning experiments (20). Semi-random *cfr9IM*'::'*bcnIB* fusions were constructed and analyzed using as host ER2267 [F′ *proA*$^+$*B*$^+$ *lacI*^{q} Δ(*lacZ*)*M15 zzf*::mini-Tn*10* (Km^R)/e14$^-$ (*mcrA*$^-$) *endA1 glnV44 thi-1* Δ(*mcrC-mrr*)*114*::*IS10* Δ(*argF-lacZ*)*U169 recA1*]. Both strains were obtained from New England Biolabs. *Escherichia coli* cells were grown in LB medium (21) at 37°C. Ampicillin, kanamycin and tetracycline were added to Luria broth or LB agar at 100, 50 and 10 μg/ml, respectively. *Escherichia coli* transformations were carried out by the $CaCl_2$–heat shock method (21).

Enzymes and chemicals

Oligonucleotides, sequencing primers, kits, phage λ DNA, enzymes and preparations of the non-commercial restriction endonucleases *Esp*1396I (recognition sequence CCANNNN↓NTGG, arrow indicates cleavage position) and *Ppu*21I (YAC↓GTR) were from and were used as recommended by Fermentas. Modification enzyme *Bcn*IA, used for the determination of N-terminal amino acids, was purified as described (5). [α-^{33}P]dATP was obtained from Izotop (St Petersburg, Russia).

Recombinant DNA techniques

Plasmid DNAs were prepared by the alkaline lysis procedure (21) and purified additionally as described previously (22). Restriction and deletion mapping, agarose gel electrophoresis, isolation and subcloning of individual DNA restriction fragments and nuclease *Bal*31 digestions were carried out by standard procedures (21). Construction of nested unidirectional deletions used for the sequencing of R-M genes was carried out using the *Exo*III/S1 Deletion Kit (Fermentas) according to the manufacturer's recommendations.

Plasmids

The recombinant plasmids pBcnRM-1, pBcnRM-2 (11), pBC1 and pBC-1 (8) were used as sources of *Bcn*I R-M genes. pBcnA-1 was constructed by inserting a 1.5 kb *Mun*I–*Bgl*II DNA fragment carrying the methyltransferase gene *bcnIA* from pBC1 into the *Eco*RI and *Bam*HI sites of pUC19. Cloning of the same DNA fragment in the opposite orientation (pBcnA-2) was achieved by end-filling with Klenow fragment and insertion into the *Sma*I site of pUC19. Subcloning of either the blunt-ended 1.33 kb *Bse*XI–*Bsu*RI fragment or the 1.13 kb *Bsu*RI fragment of pBcnRM-2 into the *Ecl*136II site of pUC19 yielded plasmids pDer1 and pDer2, respectively (Fig. 1). pBcnB was constructed in two steps. The DNA fragment coding for the N-terminal part of M.*Bcn*IB was PCR amplified using pBC1 as template and a pair of primers. Primer 1, 5′-TTAAGGAGGTAAAACAT**AT-G**AATAAACAATTAGAACAA-3′, corresponds to the 5′-terminal part of *bcnIB* and possesses an optimized Shine–Dalgarno sequence (underlined) and translation start codon ATG (bold),

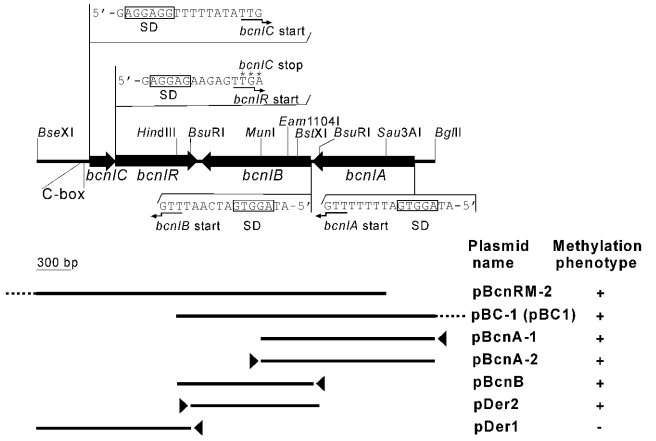

Figure 1. Gene organization and subcloning mapping of the *Bcn*I R-M system. Black solid arrows denote the locations and orientations of the *bcnIC* (putative regulatory protein C.*Bcn*I), *bcnIR* (restriction endonuclease), *bcnIA* and *bcnIB* (modification enzymes M.*Bcn*IA and M.*Bcn*IB, respectively) genes. C-box refers to the predicted operator sequence of C.*Bcn*I. SD marks the Shine–Dalgarno sequence (boxed). The direction of transcription from the P_{lac} promoter is indicated by arrowheads. Thin black lines represent sequenced regions; undetermined sequences are indicated by dotted lines (not to scale). The 'Methylation phenotype' column refers to the resistance of plasmids to R.*Bcn*I cleavage. +, complete protection; –, no protection. Plasmids pBC-1 and pBC1 differ in the orientation of the insert (8).

which replaced the TTG present in the wild-type gene. Primer 2, 5′-ATGCGTCCGGCGTAGA-3′, is complementary to the region of vector DNA just after the unique *Bam*HI target. The amplified fragment was digested with *Eam*1104I and blunt-ended. The resulting 170 bp fragment was gel purified and inserted into *Eco*32I-cleaved pUC57. In the next step, full-length *bcnIB* was reconstituted by replacing the *Bsp*68I–*Bst*XI fragment of the resulting plasmid with the 1.03 kb *Hin*dIII–*Bst*XI fragment from pBC1.

DNA sequence determination and analysis

The nucleotide sequence of the *Bcn*I R-M system was determined on both DNA strands by the dideoxy nucleotide chain termination approach using a Cycle Reader DNA Sequencing Kit (Fermentas) and standard M13/pUC sequencing primers. DNA of plasmids pDer1, pDer2, pBcnA-1 and pBcnA-2 or their *Exo*III/S1 deletion derivatives served as templates. Sequence data were compiled and analyzed with the MicroGenie sequence analysis software program (Beckman Instruments). The experimental nucleotide sequence and its translation products were used in searches against current DNA and protein databases with the PSI-BLAST (23) program network services available at the National Center for Biotechnology Information (Bethesda, MD; http://www.ncbi.nlm.nih.gov) or the FASTA3 (24) program network services available at the European Bioinformatic Institute (Cambridge, UK; http://www.ebi.ac.uk). The alignment of amino acid sequences was performed with the CLUSTALW (25) program available at the EBI. The regions of statistically significant similarities were tested using the HR-SEARCH program (26). The GeneDoc program (http://www.psc.edu/biomed/genedoc) was used to edit and annotate the alignments.

Functional analysis and determination of methylation specificity of *Bcn*I MTases

*Bcn*I-specific *in vivo* methylation was tested by incubating plasmid DNA with an excess of ENase *Bcn*I followed by agarose gel electrophoresis. Precise methylation patterns were visualized by bisulfite-mediated genomic sequencing (27). Briefly, *in vivo* modified plasmids pBcnB and pBcnA-1 as well as a control plasmid pUC19 were linearized with *Alw*44I, ethanol precipitated and dissolved in water. Each DNA sample was then treated with bisulfite as previously described (28), with the exception that five repetitive cycles of incubation with sodium bisulfite at 95°C for 3 min and 50°C for 60 min were used to modify DNA preparations. Individual strands of bisulfite-treated DNA were amplified using strand-specific pairs of primers (5′-GTTGTGTAGATAATTATGATATGGGAGGG-3′ and 5′-CATTTTCCAATAATAAACACTTTTAAAATTCT-3′, upper strand; 5′-TTTCATTCATCCATAATTACCTAACTCCCC-3′ and 5′-ATGATGAGTATTTTTAAAGTTTTGTTATGTGG-3′, lower strand). The PCR-amplified fragments correspond to positions 1702–2300 (upper strand) and 1672–2291 (lower strand) in pUC19 (GenBank/EMBL accession no. L09137) and include two *Bcn*I targets at positions 1881–1885 and 2232–2236. Both DNA fragments were agarose gel purified and inserted into *Sma*I-digested pUC19. Recombinant plasmids isolated from individual transformants were sequenced through the regions of interest.

cfr9IM'::'*bcnIB* fusions and their *in vivo* methylation activities

The plasmid pCfrBcn2 (Fig. 2), which served as a starting derivative for the generation of *cfr9IM*'::'*bcnIB* fusions, was constructed by a stepwise cloning procedure and verified by sequencing and ENase mapping (details available upon request). The construction of semi-random *cfr9IM*'::'*bcnIB* fusion libraries is outlined in Figure 2. Briefly, pCfrBcn2 was linearized with either R.*Kpn*2I or R.*Pst*I and then treated with nuclease *Bal*31. The *Bal*31 reaction probes were stopped at specified time points and the extent of DNA shortening of each probe was assessed by agarose gel electrophoresis. Probes that contained DNA molecules in the range 5000–7000 bp were combined, purified and digested in parallel reactions with nine ENases as listed in Figure 2. In the next step, the resulting DNA molecules were blunt-ended with Klenow fragment, circularized by ligation and used to transform competent ER2267 cells. Plasmid DNA isolated from the pooled ampicillin-resistant colonies (50–100 000 in each transformation) was digested to completion with an excess of R.*Cfr*9I and then re-transformed back into ER2267. Resulting individual transformants were analyzed for plasmids that possess a recombinant fragment of 1.15–1.25 kb, either directly by cleaving plasmids with R.*Bam*HI or by colony PCR (21) using M13/pUC universal primers. The selected plasmids were further analyzed for their resistance to cleavage with R.*Cfr*9I or R.*Bcn*I. *Cfr*9I-resistant plasmids were sequenced to determine the structure at their fusion points. DNA sequencing reactions were prepared using a CycleReader Auto DNA sequencing kit (Fermentas) with 5′-Cy5-labeled universal M13/pUC primers and analyzed using an ALFexpress II DNA sequencer (Amersham Pharmacia Biotech).

RESULTS

Gene organization of the *Bcn*I R-M system

The *B.centrosporus* RFL1 genomic DNA region flanked by R.*Bse*XI and R.*Bgl*II sites, encoding the complete *Bcn*I R-M system, was sequenced on both strands. Scanning the determined nucleotide sequence of 3475 bp revealed four open reading frames (ORFs) which were named *bcnIC*, *bcnIR*, *bcnIB* and *bcnIA* (Fig. 1). The ORFs are arranged as two collinear pairs that appear in convergent orientation to each other. The first two genes, *bcnIC* and *bcnIR*, overlap by 1 nt and encode proteins of 74 and 238 residues, respectively. Both ORFs are preceded by typical ribosome-binding sites (RBS) located just upstream of their putative TTG start codons (Fig. 1). The predicted product of the first gene has 40–50% identical amino acids with short regulatory proteins (not shown) observed in a few other R-M systems (1). In addition, a typical C-box, the candidate operator sequence for the mentioned regulatory proteins, was found upstream of the *bcnIC* gene (29). Altogether, these observations suggest that the product of *bcnIC* might be involved in regulation of the *Bcn*I R-M system. In accordance with the accepted nomenclature (30), the putative protein was denoted C.*Bcn*I.

The second gene, *bcnIR*, codes for the R.*Bcn*I ENase. This conclusion is based on earlier deletion mapping experiments (11) indicating that the DNA region surrounding the *Hin*dIII site in pBcnRM-2 (see Fig. 1) is essential for ENase activity. Moreover, comparative amino acid sequence analysis revealed substantial sequence similarities of R.*Bcn*I with the R.*Mva*I ENase (GenBank accession no. AF472612) from *Micrococcus varians* RFL19, which recognizes the related nucleotide sequence CC(A/T)GG (17% identical and additionally 16% similar amino acids). Another statistically significant similarity (18–25% identical amino acids) (Fig. 3) was identified between the two ENases and two ORFs (TVG1488832 and TVG1490552) present in the genome of the archaeal thermophile *Thermoplasma volcanium* (31). The functions of these two adjacent genes have not been experimentally established. Of note, the first ORF is preceded by TVG1487993, which is related to M.*Bcn*IA (see below), whereas the second one is followed by TVG1491781, which is most similar to cytosine-N4 MTase *Mva*I (data not shown). Taken together, all these observations point towards the likely role of TVG1488832 and TVG1490552 as the restriction components of an uncharacterized R-M system(s).

The other two collinear genes, *bcnIA* and *bcnIB* (Fig. 1), encode the cytosine-N4 MTases M.*Bcn*IA and M.*Bcn*IB, respectively (see below). The translation initiation codon of *bcnIA* cannot be immediately predicted from the gene sequence alone due to the presence of several potential initiation sites. The N-terminal amino acid sequence analysis of the M.*Bcn*IA protein (not shown) yielded the sequence XIKSNV, which coincides with residues 2–6 of the M.*Bcn*IA protein translated from TTG at positions 3288–3290. Furthermore, it is preceded by a RBS-like sequence (Fig. 1), suggesting that this triplet is the actual start codon. The downstream *bcnIB* gene has two potential start codons. The first one, GTG at positions 2395–2397, is not preceded by a sequence that might serve as a RBS, while the second in-frame start codon TTG at positions 2386–2388 is preceded by the RBS-like sequence 5′-AGGTG. If TTG is used, *bcnIB* codes for a *Bcn*IB methylase of 317 amino acids.

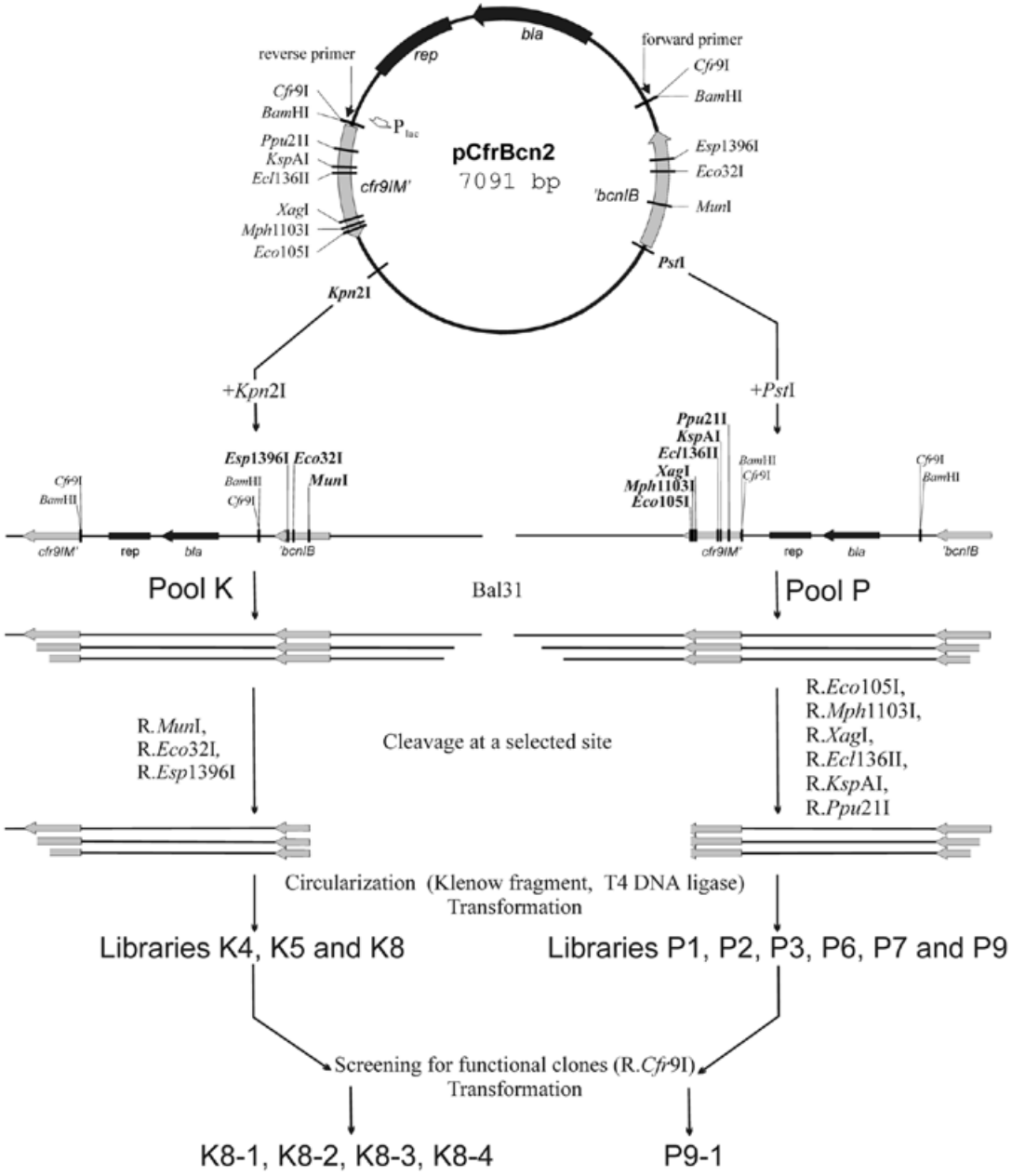

Figure 2. Construction of semi-random libraries of *cfr9IM*'::'*bcnIB* fusions. A detailed description is given in Materials and Methods. Solid black arrows indicate the position and orientation of the *bla* gene (ampicillin resistance). Shaded arrows denote the truncated MTase genes (*cfr9IM*' and '*bcnIB*), the open arrow indicates the direction of transcription from the *lac* promoter. Rep denotes the replication region.

The deduced amino acid sequences of both M.*Bcn*IA and M.*Bcn*IB contain all conserved motifs characteristic of amino DNA MTases (12,26). The linear order of motifs, however, is different and, therefore, the proteins are assigned to distinct structural classes. In M.*Bcn*IA they are arranged in the order X-I-II-III-IV-V-VI-VII-VIII, with the longest (56 amino acids) variable region (the putative target recognition domain, or TRD) at the N-terminus of the protein. This suggests that M.*Bcn*IA belongs to the ζ group of MTases (12) and, in fact, is the first characterized representative. A GenBank search revealed a remarkable similarity (62% identity) with the ORF TVG1487993 from *T.volcanium* (31). Considering such a high homology level it is likely that the specificity of TVG1487993 protein and M.*Bcn*IA is similar or identical. This conclusion is further supported by its proximity to the ENase-like genes TVG1490552 and TVG1488832 (see Discussion).

M.*Bcn*IB belongs to the β group of amino MTases (motif order IV-V-VI-VII-VIII-X-I-II-III) (12) and shows high resemblance to other representatives of the β group. The most similar among them is M.*Cfr*9I (34% identity) (32), which recognizes the related sequence C<u>C</u>CGGG. Comparison of M.*Bcn*IB with a circularly permuted version of M.*Bcn*IA identified sequence similarity that is largely limited to the conserved motifs.

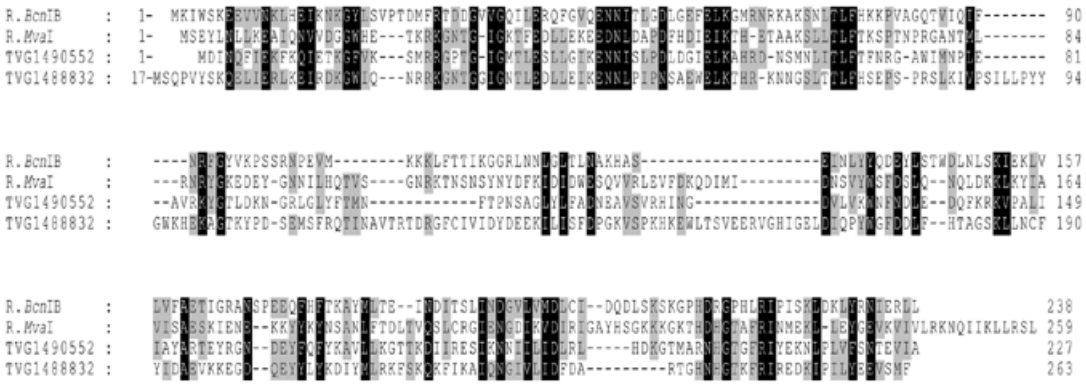

Figure 3. Sequence alignment of R.*Bcn*I, R.*Mva*I (GenBank accession no. AF472612) and two unidentified ORFs from *T.volcanium* (31). White letters on a black background and black letters on a shaded background indicate residues that are common or similar in all four or in three aligned sequences, respectively. Similar amino acids are grouped as follows: DENQ, KR, TS, VILM, FWY.

The two convergent gene pairs, *bcnIC bcnIR* and *bcnIA bcnIB*, are separated by a central intergenic region of 29 bp. The latter contains two inverted repeats at positions 1410–1419 and 1425–1434 which could potentially form a stem–loop structure and might serve as a common transcription terminator for both cistrons.

Functional analysis of the *Bcn*I MTases

The discovery of two closely linked genes encoding a pair of cytosine-N4 methylases raised the question of whether both of them are required for protection of host DNA against R.*Bcn*I *in vivo*. Current sequence analysis revealed that the plasmid pBcnRM-2 (see Fig. 1), which rendered the host cells R-M proficient (11), encodes R.*Bcn*I and only one full-length methylase, M.*Bcn*IB. This finding implies that *bcnIA* is dispensable and *bcnIB* alone is sufficient for protection of the host DNA against *in vivo* action of the cognate ENase. Since both MTases modify the same target sites, the functional role of *bcnIA* remains unclear. To elucidate the function of the *Bcn*IA MTases, a series of plasmids (pBcnA-1, pBcnA-2 and pBcnB) carrying the individual *bcnIA* and *bcnIB* genes were constructed and their modification phenotypes examined (see Fig. 1). The three plasmids, which contained multiple *Bcn*I sites, were resistant to R.*Bcn*I cleavage *in vitro* (data not shown). Therefore, the presence of either *bcnIA* or *bcnIB* renders efficient *in vivo* modification of the CC(C/G)GG sites. The final experiment was to check if M.*Bcn*IA can perform the *in vivo* protection function, as does M.*Bcn*IB. The plasmid pBcnRM-1 possessing all four *bcnI* genes was deleted for *bcnIB* or, as a control, for both *bcnIB* and *bcnIR* (not shown). Attempts to inactivate *bcnIB* alone were unsuccessful, resulting in rearranged derivatives only. In contrast, simultaneous deletion of both *bcnIB* and *bcnIR* was readily achieved and yielded R.*Bcn*I-resistant plasmids (data not shown). These observations indirectly suggest that M.*Bcn*IA alone cannot protect all its target sites in the presence of the active ENase gene, at least in the heterologous model system employed.

Specificity of the *Bcn*I MTases

Previous experiments (7–10) established that a DNA MTase activity (called M.*Bcn*I) purified from *B.centrosporus* RFL1 or from *E.coli* HB101 [pBC1] modifies the second cytosine residue in the target sequence C<u>C</u>(C/G)GG yielding N4-methylcytosine. In the light of the discovery of two *Bcn*I MTases, it became unclear which of the two enzymes, M.*Bcn*IA or M.*Bcn*IB, or a mixture of both, was analyzed in those early studies. Subsequent specificity studies of individual MTases demonstrated that both enzymes methylate CC(C/G)GG sites in double-stranded DNA and that M.*Bcn*IA can also, with comparable efficiency, modify these sites in single-stranded DNA (5). However, the precise specificities, i.e. which nucleotide in the target site is modified by the individual MTases, remained unclear. To determine the position of methylation, we applied a bisulfite sequencing protocol adopted recently for mapping N4-methylcytosine residues in DNA (28,33). The DNAs of pBcnA-1 and pBcnB, which express individual MTases, were analyzed to map the positions of methylated cytosines in the complementary strands CCCGG and CCGGG of two neighboring *Bcn*I targets (Table 1). In all cases only the second cytosine residue on each strand of the *Bcn*I site CC(C/G)GG was refractory to bisulfite attack. Consistent with the previous studies, N4-methylcytosines were displayed on the sequencing ladders with frequencies ranging from 14 to 50%. Taken together, our results demonstrate that both *Bcn*I methylases modify the second C base on both strands of the target sequence C<u>C</u>(C/G)GG yielding N4-methylcytosine. These results are in full agreement with data obtained in the previous studies (7–10).

Table 1. Bisulfite sequencing analysis of DNA methylated by M.*Bcn*IA and M.*Bcn*IB

MTase	Frequency of display of modified cytosines at the underlined positions in the individual strands	
	5′-C<u>C</u>GGG-3′	5′-C<u>C</u>CGG-3′
M.*Bcn*IA	6/13 (46%)[a]	2/14 (14%)
M.*Bcn*IB	2/13 (15%)	7/14 (50%)

[a]Frequency of display is determined as the ratio of positive displays over the total number of clones analyzed.

Construction of hybrid cytosine-N4 MTases

Structural probing by hybrid formation has become an essential tool for studying relationships between the evolutionary, structural and functional organization of proteins (reviewed in 34,35). Although the formation of functionally active hybrids has been demonstrated for monospecific 5-methylcytosine MTases (19,36), no such analysis in the amino MTase family has been reported to date. In the context of the natural occurrence of circularly permuted MTases in the *Bcn*I system, we have undertaken such an analysis by constructing a library of single-crossover hybrids between M.*Bcn*IB and its closest homolog, M.*Cfr*9I. The enzymes recognize overlapping targets [C<u>C</u>(C/G)GG and C<u>C</u>CGGG, respectively], belong to the same structural β group and share 34% sequence identity. MTase hybrids were constructed employing a novel strategy that generates libraries of semi-random fusions as outlined in Figure 2. The starting step was the construction of a 'tandem' plasmid containing the two parent genes separated by a large intervening region. Both MTase genes were inactivated by deleting short terminal regions (*cfr9IM*' lacks 93 nt at the 3′-end and '*bcnIB* lacks 118 nt at the 5′-end) to avoid a high background of wild-type variants at the selection step (see below). The tandem plasmid pCfrBcn2 was linearized in parallel with either *Pst*I, which cleaves near the 5′-terminus of '*bcnIB*, or with *Kpn*2I, which cleaves near the 3′-terminus of *cfr9IM*', and then treated with nuclease *Bal*31 to yield two random size pools of linear molecules. In the first pool (named P) '*bcnIB* was further progressively shortened at the 5′-end, whereas *cfr9IM*' was protected from exonuclease digestion by the intervening region. In the reciprocal experiment (pool K) a similar procedure led to variable digestion of *cfr9IM*' from the 3′-end with full retention of '*bcnIB*. The *Bal*31 reaction products were fractionated to select fragments that were 0–2000 nt shorter than the initial linearized plasmid. In the next step, pool P was divided into parallel fractions and treated with ENases that produce unique cuts in *cfr9IM*' only (*Ppu*21I, *Ksp*AI, *Ecl*136II, *Xag*I, *Mph*1103I or *Eco*105I). The resulting DNA molecules were blunt-ended with Klenow fragment, self-ligated and transformed into *E.coli* to yield six independent libraries (named P1, P2, P3, P6, P7 and P9, respectively). Pool K was fragmented in separate fractions with *Mun*I, *Eco*32I or *Esp*1396I, which uniquely cut within '*bcnIB*. The same work-up procedure resulted in the K4, K5 and K8 libraries, respectively. Inspection of the junction points showed that coding triplets should occur upon in-frame fusion of the MTase genes and, therefore, full-length hybrid MTases were expected in all nine cases. The sizes of the individual clone libraries (50 000–100 000) ensured that all variants within a tolerance range of 1000 bp were present with at least 50–100-fold statistical redundancy.

The nine hybrid libraries were individually screened for plasmids that express active MTases using a positive selection procedure. The method is based on the resistance of self-modifying recombinant plasmids to digestion with a cognate ENase (37). R.*Cfr*9I, which has two target CCCGGG sites on pCfrBcn2, does not cut DNA if any of the cytosine residues in the site is replaced with N4-methylcytosine (38), and thus is inhibited by either M.*Cfr*9I or M.*Bcn*IB. Digestion of the clone libraries with *Cfr*9I *in vitro* followed by re-transformation into *E.coli* allowed selective removal of all variants devoid of MTase activity.

Analysis of hybrid cytosine-N4 MTases

Individual transformants obtained after the screening step (typically 50–200) were analyzed with respect to their plasmid structure and methylation phenotype. Several modification-positive hybrids were obtained from two libraries, K8 and P9, but no such hybrids were obtained in the case of P1, P2, P3, K4, K5, P6 and P7. Analysis of the P9 library revealed 16 active hybrids, all of which proved identical (designated P9-1). As expected, the junction point appeared just downstream of the remnant *Eco*105I site (Fig. 4) such that the C-terminal part of M.*Bcn*IB replaced the homologous region of M.*Cfr*9I without any amino acid insertions or deletions. In the K8 library four different chimeric MTases were identified among 19 modification-positive clones sequenced. The most abundant (12 representatives) was the K8-1 group, which turned out to be identical to P9-1. Other hybrids possessed three new structures and were denoted K8-2 (three clones), K8-3 (two clones) and K8-4 (two clones; see Fig. 4). The K8 hybrids were expected to contain a fixed C-terminal segment of *bcnIB* extending up to the *Esp*1396I site fused to a variable length fragment from *cfr9IM*. However, the *bcnIB* sequence is shorter than expected by 8 nt in K8-1, by 17 nt in K8-2 and by 23 nt in K8-3. This heterogeneity of the fusion point position might have resulted from excessive exonucleolytic activity of the Klenow fragment, which removed not only cohesive DNA ends but also adjacent nucleotides. However, the 'lost' terminal nucleotides in the *bcnIB* gene were faithfully replaced with appropriate regions from *cfr9IM*. The K8-4 variant contains the expected fragment from *bcnIB*, but also an extra 69 nt derived from *cfr9IM*, which leads to an insertion of 23 amino acids in the hybrid. As is evident from Figure 4, the amino acid sequences of K8-2 and K8-3 are identical despite two mismatches at the nucleotide level. The amino acid sequence of K8-1/P9-1 differs from those of K8-2 and K8-3 by a single amino acid residue (Phe250→Leu replacement).

To determine their DNA modification specificities and relative catalytic efficiencies *in vivo*, the recombinant plasmids were challenged with ENases *Cfr*9I and *Bcn*I. Positive controls were carried out to ensure that the restriction enzymes performed as expected. All hybrids proved weaker methylases as compared to their wild-type versions since both progenitor genes rendered complete modification of the cognate sites (data not shown). R.*Cfr*9I analysis (Fig. 5) demonstrated similar DNA methylation levels to K8-2 and K8-3, consistent with their identical protein sequences. Both plasmids are nearly completely *Cfr*9I resistant. P9-1, which differs by a single amino acid, renders a slightly lower level of modification of the CCCGGG sites. In contrast, very weak DNA modification was observed in the case of K8-4. Analysis of the R.*Bcn*I fragmentation patterns showed that in all cases only the two *Bcn*I sites that overlap with the *Cfr*9I sites were resistant to cleavage (data not shown). Altogether our results clearly show that the hybrids exhibit the specificity of M.*Cfr*9I. This observation is fully consistent with the primary structure of the hybrids, >75% of which (including the putative TRD) is derived from M.*Cfr*9I.

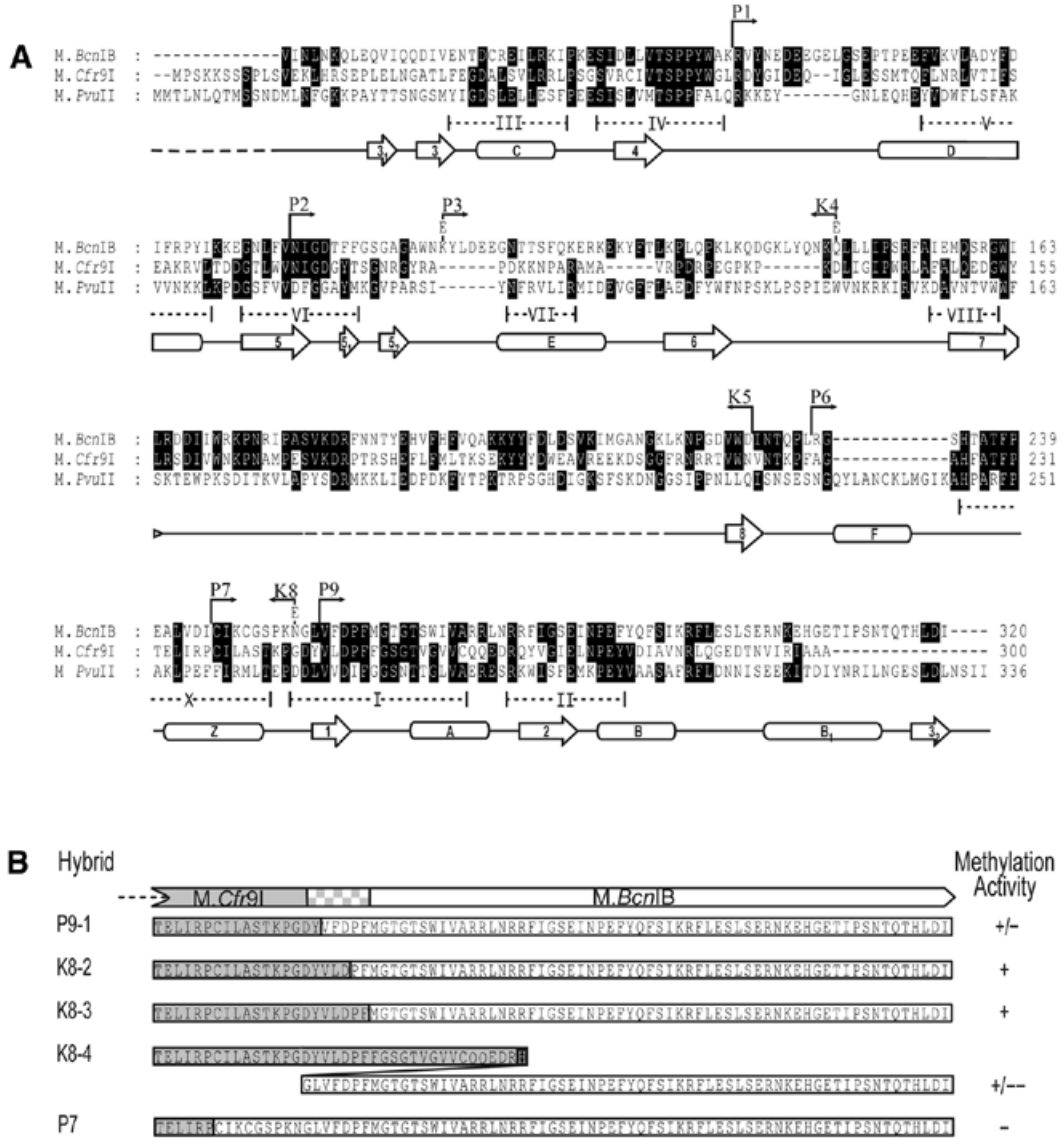

Figure 4. Structure and functional activity of hybrid MTases. (**A**) Alignment of amino acid sequences of M.*Bcn*IB and M.*Cfr*9I (GenBank accession no. X17022) along with M.*Pvu*II (GenBank accession no. M77223). Residues invariant in more than two sequences are shown as white letters on a black background. The secondary structure of M.*Pvu*II (from 15) is shown below the alignment. Cylinders stand for α-helixes, arrows for β-strands. Conserved motifs I–VIII and X (12) are marked below the alignment. The crossover points P1–P9 are marked above the sequences with arrows pointing in the direction of the variable length fragment. In-frame fusion of the two genes in P3, K4 and K8, produced E codons (shown above the corresponding residues). (**B**) Schematic representation of primary structures of hybrid proteins [aligned with the bottom section of (A)]. Sequences derived from M.*Cfr*9I and M.*Bcn*IB are gray and white, respectively. Mutant residues at the fusion point are shown as white letters on a black background. The derived crossover region is shown in mixed gray-white. The 'Methylation activity' column indicates the resistance of plasmids to *in vitro* cleavage with R.*Cfr*9I. +, complete resistance; +/–, partial resistance; +/—, weak resistance; –, complete fragmentation.

DISCUSSION

Structure and function of the *Bcn*I system

The four genes of the type II R-M system *Bcn*I are arranged as two converging collinear pairs. One pair encodes the putative regulatory protein C (gene *bcnIC*) and ENase *Bcn*I (*bcnIR*) and the other codes for the two cytosine-N4 MTases M.*Bcn*IA and M.*Bcn*IB (genes *bcnIA* and *bcnIB*, respectively). In most R-M systems possessing regulatory genes a methylase gene is transcribed in the opposite direction from the *CR* operon. The gene arrangement of *Bcn*I best resembles that of the *Sma*I and *Nme*SI systems, in which *M* and *CR* converge (39). The main difference between them is the occurrence of two MTase genes in *Bcn*I. In addition, the location of putative regulatory sequences, so called C-boxes, appears to be different. These were identified upstream of both the *M* and *C* genes in the *Sma*I and *Nme*SI systems (39), whereas no such putative operators were found to precede *bcnIA* or *bcnIB*. These facts suggest that the overall mechanisms of gene regulation in *Sma*I/*Nme*SI and *Bcn*I may be different.

R.*Bcn*I is the first sequenced prototype among over 40 ENases that recognize the CC(C/G)GG DNA sequence (1). In

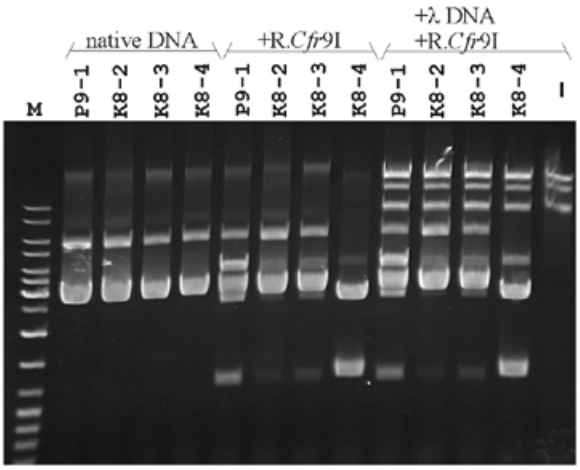

Figure 5. The efficiency of *in vivo* DNA modification by hybrid MTases. The plasmids encoding the respective hybrids (left) were challenged with R.*Cfr*9I (center). The control reaction (right) includes R.*Cfr*9I cleavage of hybrid plasmids along with λ phage DNA. M, DNA size marker.

general, the primary sequences of type II ENases show little resemblance to each other (except for isoschizomers) despite their similar functions and common tertiary structure. The degree of similarity detected between R.*Bcn*I, R.*Mva*I and the two ORFs TVG1490552 and TVG1488832 from *T.volcanium* is marginal, but comparable with that observed between other ENases recognizing related targets. For example, *Mun*I and *Eco*RI recognize the related hexanucleotide sequences CAATTG and GAATTC, respectively, and share 18% sequence identity (40). Subsequent X-ray structures revealed that these enzymes possess high similarity in their tertiary folds and in the structural elements interacting with the inner tetranucleotide sequence AATT (41). Taking into account that R.*Bcn*I and R.*Mva*I recognize related DNA targets [CC(C/G)GG and CC(A/T)GG, respectively] and cleave DNA at identical positions (after the second C), one could expect that the two homologs identified in *T.volcanium* are also ENases of similar or identical specificity. This notion is reinforced by the fact that these two genes are flanked by two ORFs (TVG1487993 and TVG1491781) that are highly homologous to the cytosine-N4 MTases *Bcn*IA and *Mva*I, respectively. Therefore, it is likely that the aforementioned locus in *T.volcanium* encodes two adjacent R-M systems with CC(C/G)GG and CC(A/T)GG specificity, or a single complex system consisting of two MTases and two ENases that recognize related/overlapping targets.

The biological function of the two MTases in the *Bcn*I R-M system is not completely clear. While M.*Bcn*IA and M.*Bcn*IB are absolutely isospecific on double-stranded DNA, they show clear differences in that the first MTase is also active on single-stranded substrates. Moreover, subcloning experiments indicated that M.*Bcn*IB alone is sufficient to support the growth of *E.coli* cells carrying the *Bcn*I ENase gene; this could not be demonstrated for M.*Bcn*IA. Taken together with their dissimilar primary structures, this disparity indirectly suggests that the MTases may have distinct functional roles. A few other known examples of type II R-M systems in which a gene for ENase is accompanied by tandem MTase genes can be divided into two major groups. The first group is represented by R-M systems such as *Bfi*I and *Bpu*10I that recognize asymmetrical (non-palindromic) DNA targets (33,42). The modification of target bases on both DNA strands involves two separate binding and recognition events, which in the case of a non-symmetrical site requires two distinct sets of interactions. This problem is usually overcome by employing two distinct MTase activities each responsible for the modification of a specific DNA strand and which may be encoded either by two separate (33,42) or a single polypeptide chain (43). The other group consists of R-M systems that recognize palindromic DNA sequences. The best studied among them is *Dpn*II from *Streptococcus pneumoniae*, which recognizes the sequence GATC (44,45). The two DNA-adenine methylases encoded by the *Dpn*II R-M system, namely *Dpn*M and *Dpn*A, methylate double-stranded DNA substrates, but *Dpn*A can also modify single-stranded DNA (46). Few other cloned and sequenced GATC-specific R-M systems (*Lla*DCHI, *Lla*AI, *Mbo*I and *Ssu*DAT1I) are highly homologous with *Dpn*II. Tandem 5-methylcytosine MTases are exemplified by the *Scr*FI system from *Lactococcus lactis*, which recognizes the sequence CCNGG (47). However, the specific roles of dual MTases have been studied in detail only in the case of *Dpn*II. The primary biological role of *Dpn*A was suggested to be facilitation of the natural transfer of plasmid DNA between strains by the transformation pathway in which the incoming plasmid DNA is in a single-stranded form. Since a linear single-stranded plasmid cannot circularize and replicate itself, two complementary strands entering separately must interact to establish a plasmid. *Dpn*A is expected to perform its function through the methylation of incoming single strands, resulting in resistance of the subsequently reconstituted double-stranded plasmid DNA to R.*Dpn*II cleavage (46,48). Clearly, the modification component of the *Bcn*I system resembles that of *Dpn*II: M.*Bcn*IB, like *Dpn*M, is active on double-stranded DNA targets and M.*Bcn*IA methylates both single-stranded and double-stranded substrates, like *Dpn*A (5). In all likelihood, M.*Bcn*IA is a dedicated single-stranded MTase required for plasmid transfer, as demonstrated for *Dpn*M (46), and which is dispensable (at least when located on a multicopy plasmid in *E.coli*) under normal cultivation conditions, as demonstrated above (Fig. 1).

Exchangeability of structural elements and circular permutation in cytosine-N4 MTases

The formation of functionally active MTase hybrids between two N4-methylcytosine MTases of the β class, *Bcn*IB and *Cfr*9I, was probed at nine different crossover positions along the MTase sequence. Two such positions yielded functionally active hybrids in which four different hybrid sequences were identified among 35 clones examined. These derivatives represent the first reported functional hybrids obtained from two different MTases of the cytosine-N4 family. Their analysis demonstrated that the crossover points cluster in a narrow window of several residues between sequence motifs X and I located near the C-terminus of the protein (see Fig. 4). One of the variants, K8-4, showed that an insertion of 23 amino acid residues in that region can also be tolerated. According to the prototype structure of the β class N4-methylcytosine MTase *Pvu*II (15), the core of the protein possesses a structural fold consisting of a seven-stranded β-sheet flanked by six parallel α-helices. Of note, in K8-4 the crossover point lies in the loop connecting helix αZ and strand β1 (Fig. 6), which is a

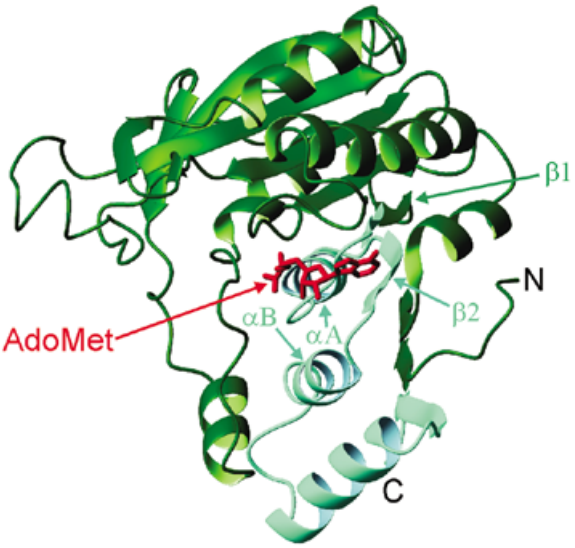

Figure 6. Presumed tertiary structure of the catalytically active M.*Cfr*9I–M.*Bcn*IB hybrids K8-1–K8-3. The region of M.*Pvu*II (PDB accession no. 1BOO) corresponding to the N-terminal section (M.*Cfr*9I) is green and the C-terminal section (M.*Bcn*IB) is cyan. The cofactor AdoMet is colored red. Helices stand for α-helixes, arrows for β-strands (from 15). The crossover region (residues 263–269) lies in the β1-strand. Prepared with MOLMOL (55).

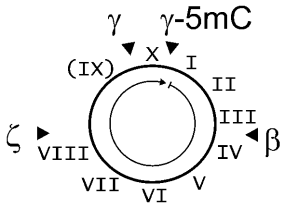

Figure 7. Evolution of DNA MTases by a circular permutation mechanism. Conserved sequence motifs I–X represented in a circular manner. Breaking the circle at defined points (shown by arrowheads) leads to different classes of circularly permuted linear sequences. The inner arrow shows the direction of the polypeptide chain.

favorable factor for non-disruptive accommodation of the inserted 23 residue sequence. The other three hybrids recombined within strand β1. All the hybrids received their C-terminal part from M.*Bcn*IB, which included helix αA, strand β2, helix αB and their connecting loops. These elements of secondary structure form one wall of the AdoMet-binding pocket, the so-called Rossman fold. This fold is very well conserved not only in DNA MTases but also in most known AdoMet-dependent methylases that modify various substrates and come from distinct origins (49), as well as many other adenine nucleotide-binding proteins. It is thus not so surprising that such a conserved fold exhibits certain modularity and can be exchanged between related proteins.

In the light of a quite even distribution of homologous regions along the MTase sequence (Fig. 4), the reason for the heavily biased distribution of crossover points appears somewhat puzzling. One possibility might be the peculiarities of our new semi-random hybrid construction strategy used to create multiple collections of *cfr9IM*'::'*bcnIB* fusions. Since the position of a crossover point is confined on one side by the cleavage site of the ENase employed, that particular junction point may appear incompatible with the overall structure leading to a misfolded or inactive protein. However, as exemplified by the K8 hybrids, a junction point does not necessarily lie exactly at the corresponding cleavage site. Due to some inadvertent action of the enzymes during the experimental manipulations, the fixed fragment could be shortened, leading to a slight shift of the junction position, such as in K8-1, K8-2 and K8-3. Alternatively, in-frame fusion with a longer fragment on the variable side would lead to an inserted peptide sequence, which would either loop out itself or replace the equivalent sequence from the other MTase, forcing it to loop out (as in K8-4). This additional limited flexibility at crossover points leads to an increased diversity of hybrid variants. On the variable side the 50–100-fold redundancy of the library ensured the statistical presence of fragments of any reasonable size. Therefore, technical problems were unlikely to be the major reason that no active hybrid MTases were observed at seven of nine junction points examined. This was further confirmed by constructing the P6 and P7 variants by an exact fusion of the two genes using PCR (data not shown). In both cases the constructs expressed full-length chimeric proteins, which, however, showed no detectable methylation activity at CCCGGG or CC(C/G)GG sites.

Altogether, these facts lead to the conclusion that the observed limited exchangeability of sequence elements may reflect certain inherent structural features of the N4-methylcytosine MTases or protein organization in general. The MTases differ in the arrangement and order of sequence motifs giving rise to the subclassifications α, β, γ, γ-5mC and ζ. The β, γ and γ-5mC arrangements of motifs can be obtained by direct circular permutation of sequences, which may have occurred during evolution, very much in the same manner that was used to create the current set of MTase hybrids. The junction point in the active hybrids lies between the conserved sequence motifs X and I. This turns out to almost exactly match the sequence break point in γ-5mC MTases (Fig. 7). Two other known arrangements, α and γ, have sequence break points in the region preceding motif X. The K5 and P6 junction points are exactly in this region, but no active derivatives were obtained using our semi-random hybrid approach. Consistent with this, the P6 hybrid, which has been constructed in a separate experiment by specific joining of sequences, showed no MTase activity. Therefore, we conclude with a high degree of confidence that no active hybrids of this type can be obtained in this system. The sequences of the remaining ζ class possess discontinuity in the region subsequent to motif VIII, which could not be constructed using available ENase sites on either MTase gene.

Our study experimentally validated one region that could potentially serve as a junction/breakage point during evolutionary circular permutation of MTase sequences. The other two known sequence junction points, which are much more centrally located in the employed MTases, did not yield active derivatives and thus could not be confirmed in these *in vitro* experiments. It should be noted that studies of other systems showed that active single-crossover hybrid proteins were obtained only with crossovers close to the N- or C-terminus (50,51). Single junctions in the central region may be too disruptive, simply because larger regions would be exchanged leading to a larger number of disturbed interactions. This

limitation can probably be overcome by generating additional crossovers, or by mutational fine tuning to counter such disruptions.

Recombination is an important mechanism for the acquisition of novel functions in proteins. One evolutionary model of amino MTases envisions the following steps: duplication of a predecessor gene, joining of the original N- and C-termini of the two proteins, followed by generation of new termini within the tandem protein (52). This simple model explains how different permuted versions of MTases could have arisen, but it does not explain what the purpose of it is (other than creation of circularly permuted versions themselves). Indeed, after a duplicate copy of a gene is created, there are at least two other alternatives to evolve higher diversity: (i) the second copy can evolve independently to achieve a certain desired phenotype; (ii) molecular shuffling can be achieved by homologous recombination without circular permutation if the duplicate copy is highly homologous with the original. Our results shed new light by demonstrating that new MTases can be created not only from two copies of the same gene but also from two distinct parental genes. Clearly, such a circular permutation mechanism would be well suited for molecular breeding of more distantly related sequences. The available structural and genetic data on adenine-N6 and cytosine-N4 specific MTases (15,53,54) indicate that the recognition elements are not clustered in a single small domain, but rather come from several regions separated in the primary structure. Since the resulting hybrids would carry sequence elements from both predecessors, this mechanism could serve to create new recognition specificities of MTases. In combination with mutational fine tuning, the repertoire of useful crossover points *in vivo* should not be limited to a single variant, as observed in the present *in vitro* study.

ACKNOWLEDGEMENTS

We thank D. Šikšnienė and A. Rukšėnaitė for the sequencing of hybrid genes. We are also grateful to V. Naktinis and J. Zareckaja for N-terminal amino acid sequence analysis of *Bcn*IA methyltransferase and Fermentas for continued support. This work was in part supported by a Howard Hughes Medical Institute International Research Scholarship and State Science and Study Foundation of Lithuania (project no. 96-170/2B) grants to S.K.

REFERENCES

1. Roberts,R.J. and Macelis,D. (2001) REBASE—restriction enzymes and methylases. *Nucleic Acids Res.*, **29**, 268–269.
2. Bickle,T.A. and Kruger,D.H. (1993) Biology of DNA restriction. *Microbiol. Rev.*, **57**, 434–450.
3. Kobayashi,I. (2001) Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. *Nucleic Acids Res.*, **29**, 3742–3756.
4. Wilson,G.G. and Murray,N.E. (1991) Restriction and modification systems. *Annu. Rev. Genet.*, **25**, 585–627.
5. Merkiene,E., Vilkaitis,G. and Klimasauskas,S. (1998) A pair of single-strand and double-strand DNA cytosine-N4 methyltransferases from *Bacillus centrosporus*. *Biol. Chem.*, **379**, 569–571.
6. Janulaitis,A.A., Petrusyte,M.A., Jaskelevicene,B.P., Krayev,A.S., Skryabin,K.G. and Bayev,A.A. (1982) A new restriction endonuclease *Bcn*I from *Bacillus centrosporus* RFL1. *FEBS Lett.*, **137**, 178–180.
7. Petrusyte,M.P. and Janulaitis,A.A. (1981) Specific methylase from *Bacillus centrosporus*. *Bioorg. Khim.*, **7**, 1885–1887.
8. Janulaitis,A., Povilionis,P. and Sasnauskas,K. (1982) Cloning of the modification methylase gene of *Bacillus centrosporus* in *Escherichia coli*. *Gene*, **20**, 197–204.
9. Janulaitis,A., Klimasauskas,S., Petrusyte,M. and Butkus,V. (1983) Cytosine modification in DNA by *Bcn*I methylase yields N4-methylcytosine. *FEBS Lett.*, **161**, 131–134.
10. Klimasauskas,S., Butkus,V. and Janulaitis,A.A. (1987) Investigation of specificity of DNA-methylases *Bcn*I, *Cfr*I and *Cfr*10I. *Mol. Biol. (Mosk.)*, **21**, 87–92.
11. Povilionis,P.I., Lubys,A.A. and Janulaitis,A.A. (1988) Cloning of the restriction-modification genes of *Bacillus centrosporus* in *Escherichia coli*. *Genetika*, **24**, 210–215.
12. Malone,T., Blumenthal,R.M. and Cheng,X. (1995) Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases and suggests a catalytic mechanism for these enzymes. *J. Mol. Biol.*, **253**, 618–632.
13. Cheng,X., Kumar,S., Klimasauskas,S. and Roberts,R.J. (1993) Crystal structure of the *Hha*I DNA methyltransferase. *Cold Spring Harbor Symp. Quant. Biol.*, **58**, 331–338.
14. Tran,P.H., Korszun,Z.R., Cerritelli,S., Springhorn,S.S. and Lacks,S.A. (1998) Crystal structure of the *Dpn*M DNA adenine methyltransferase from the *Dpn*II restriction system of *Streptococcus pneumoniae* bound to S-adenosylmethionine. *Structure*, **6**, 1563–1575.
15. Gong,W., O'Gara,M., Blumenthal,R.M. and Cheng,X. (1997) Structure of *Pvu*II DNA-(cytosine N4) methyltransferase, an example of domain permutation and protein fold assignment. *Nucleic Acids Res.*, **25**, 2702–2715.
16. Schluckebier,G., Kozak,M., Bleimling,N., Weinhold,E. and Saenger,W. (1997) Differential binding of S-adenosylmethionine, S-adenosylhomocysteine and Sinefungin to the adenine-specific DNA methyltransferase M.*Taq*I. *J. Mol. Biol.*, **265**, 56–67.
17. Scavetta,R.D., Thomas,C.B., Walsh,M.A., Szegedi,S., Joachimiak,A., Gumport,R.I. and Churchill,M.E. (2000) Structure of *Rsr*I methyltransferase, a member of the N6-adenine beta class of DNA methyltransferases. *Nucleic Acids Res.*, **28**, 3950–3961.
18. Balganesh,T.S., Reiners,L., Lauster,R., Noyer-Weidner,M., Wilke,K., Trautner,T.A. (1987) Construction and use of chimeric SPR/Phi3T DNA methyltransferases in the definition of sequence recognizing enzyme regions. *EMBO J.*, **6**, 3543–3549.
19. Klimasauskas,S., Nelson,J.L. and Roberts,R.J. (1991) The sequence specificity domain of cytosine-C5 methylases. *Nucleic Acids Res.*, **19**, 6183–6190.
20. Dila,D., Sutherland,E., Moran,L., Slatko,B. and Raleigh,E.A. (1990) Genetic and sequence organization of the *mcrBC* locus of *Escherichia coli* K-12. *J. Bacteriol.*, **172**, 4888–4900.
21. Sambrook,J. and Russell,D.W. (2001) *Molecular Cloning: A Laboratory Manual*, 3rd Edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
22. Marko,M.A., Chipperfield,R. and Birnboim,H.C. (1982) A procedure for the large scale isolation of highly purified plasmid DNA using alkaline extraction and binding to glass powder. *Anal. Biochem.*, **121**, 382–387.
23. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
24. Pearson,W.R. and Lipman,D.J. (1988) Improved tools for biological sequence comparison. *Proc. Natl Acad. Sci. USA*, **85**, 2444–2448.
25. Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.
26. Timinskas,A., Butkus,V. and Janulaitis,A. (1995) Sequence motifs characteristic for DNA [cytosine-N4] and DNA [adenine-N6] methyltransferases. Classification of all DNA methyltransferases. *Gene*, **157**, 3–11.
27. Frommer,M., McDonald,L.E., Millar,D.S., Collis,C.M., Watt,F., Grigg,G.W., Molloy,P.L. and Paul,C.L. (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl Acad. Sci. USA*, **89**, 1827–1831.
28. Vilkaitis,G. and Klimasauskas,S. (1999) Bisulfite sequencing protocol displays both 5-methylcytosine and N4-methylcytosine. *Anal. Biochem.*, **271**, 116–119.
29. Rimseliene,R., Vaisvila,R. and Janulaitis,A. (1995) The *eco72IC* gene specifies a trans-acting factor which influences expression of both DNA

methyltransferase and endonuclease from the *Eco*72I restriction-modification system. *Gene*, **157**, 217–219.
30. Tao,T., Bourne,J.C. and Blumenthal,R.M. (1991) A family of regulatory genes associated with type II restriction-modification systems. *J. Bacteriol.*, **173**, 1367–1375.
31. Kawashima,T., Amano,N., Koike,H., Makino,S., Higuchi,S., Kawashima-Ohya,Y., Watanabe,K., Yamazaki,M., Kanehori,K., Kawamoto,T. *et al.* (2000) Archaeal adaptation to higher temperatures revealed by genomic sequence of *Thermoplasma volcanium*. *Proc. Natl Acad. Sci. USA*, **97**, 14257–14262.
32. Lubys,A., Menkevicius,S., Timinskas,A., Butkus,V. and Janulaitis,A. (1994) Cloning and analysis of translational control for genes encoding the *Cfr*9I restriction-modification system. *Gene*, **141**, 85–89.
33. Sapranauskas,R., Sasnauskas,G., Lagunavicius,A., Vilkaitis,G., Lubys,A. and Siksnys,V. (2000) Novel subtype of type IIs restriction enzymes. *Bfi*I endonuclease exhibits similarities to the EDTA-resistant nuclease Nuc of *Salmonella typhimurium*. *J. Biol. Chem.*, **275**, 30878–30885.
34. Nixon,A.E., Ostermeier,M. and Benkovic,S.J. (1998) Hybrid enzymes: manipulating enzyme design. *Trends Biotechnol.*, **16**, 258–264.
35. Beguin,P. (1999) Hybrid enzymes. *Curr. Opin. Biotechnol.*, **10**, 336–340.
36. Mi,S. and Roberts,R.J. (1992) How M.*Msp*I and M.*Hpa*II decide which base to methylate. *Nucleic Acids Res.*, **20**, 4811–4816.
37. Szomolanyi,E., Kiss,A. and Venetianer,P. (1980) Cloning the modification methylase gene of *Bacillus sphaericus* R in *Escherichia coli*. *Gene*, **10**, 219–225.
38. Butkus,V., Petrauskiene,L., Maneliene,Z., Klimasauskas,S., Laucys,V. and Janulaitis,A. (1987) Cleavage of methylated CCCGGG sequences containing either N4-methylcytosine or 5-methylcytosine with *Msp*I, *Hpa*II, *Sma*I, *Xma*I and *Cfr*9I restriction endonucleases. *Nucleic Acids Res.*, **15**, 7091–7102.
39. Bart,A., Dankert,J. and van der Ende,A. (1999) Operator sequences for the regulatory proteins of restriction modification systems. *Mol. Microbiol.*, **31**, 1277–1278.
40. Siksnys,V., Zareckaja,N., Vaisvila,R., Timinskas,A., Stakenas,P., Butkus,V. and Janulaitis,A. (1994) CAATTG-specific restriction-modification *munI* genes from *Mycoplasma*: sequence similarities between R.*Mun*I and R.*Eco*RI. *Gene*, **142**, 1–8.
41. Deibert,M., Grazulis,S., Janulaitis,A., Siksnys,V. and Huber,R. (1999) Crystal structure of *Mun*I restriction endonuclease in complex with cognate DNA at 1.7 Å resolution. *EMBO J.*, **18**, 5805–5816.
42. Stankevicius,K., Lubys,A., Timinskas,A., Vaitkevicius,D. and Janulaitis,A. (1998) Cloning and analysis of the four genes coding for *Bpu*10I restriction-modification enzymes. *Nucleic Acids Res.*, **26**, 1084–1091.
43. Looney,M.C., Moran,L.S., Jack,W.E., Feehery,G.R., Benner,J.S., Slatko,B.E. and Wilson,G.G. (1989) Nucleotide sequence of the *Fok*I restriction-modification system: separate strand-specificity domains in the methyltransferase. *Gene*, **80**, 193–208.
44. Lacks,S.A., Mannarelli,B.M., Springhorn,S.S. and Greenberg,B. (1986) Genetic basis of the complementary *Dpn*I and *Dpn*II restriction systems of *S. pneumoniae*: an intercellular cassette mechanism. *Cell*, **46**, 993–1000.
45. de la Campa,A.G., Kale,P., Springhorn,S.S. and Lacks,S.A. (1987) Proteins encoded by the *Dpn*II restriction gene cassette. Two methylases and an endonuclease. *J. Mol. Biol.*, **196**, 457–469.
46. Cerritelli,S., Springhorn,S.S. and Lacks,S.A. (1989) *Dpn*A, a methylase for single-strand DNA in the *Dpn*II restriction system and its biological function. *Proc. Natl Acad. Sci. USA*, **86**, 9223–9227.
47. Davis,R., van der Lelie,D., Mercenier,A., Daly,C. and Fitzgerald,G.F. (1993) *Scr*FI restriction-modification system of *Lactococcus lactis* subsp. cremoris UC503: cloning and characterization of two *Scr*FI methylase genes. *Appl. Environ. Microbiol.*, **59**, 777–785.
48. Lacks,S.A., Ayalew,S., de la Campa,A.G. and Greenberg,B. (2000) Regulation of competence for genetic transformation in *Streptococcus pneumoniae*: expression of *dpnA*, a late competence gene encoding a DNA methyltransferase of the *Dpn*II restriction system. *Mol. Microbiol.*, **35**, 1089–1098.
49. Fauman,E.B., Blumenthal,R.M. and Cheng,X. (1999) In Blumenthal,R.M. and Cheng,X. (eds), *S-Adenosylmethionine-dependent Methyltransferases: Structures and Functions*. World Scientific Publishing, Singapore, pp. 1–38.
50. Bellamine,A., Gautier,J.C., Urban,P. and Pompon,D. (1994) Chimeras of the human cytochrome P450 1A family produced in yeast. Accumulation in microsomal membranes, enzyme kinetics and stability. *Eur. J. Biochem.*, **225**, 1005–1013.
51. Sieber,V., Martinez,C.A. and Arnold,F.H. (2001) Libraries of hybrid proteins from distantly related sequences. *Nat. Biotechnol.*, **19**, 456–460.
52. Jeltsch,A. (1999) Circular permutations in the molecular evolution of DNA methyltransferases. *J. Mol. Evol.*, **49**, 161–164.
53. Goedecke,K., Pignot,M., Goody,R.S., Scheidig,A.J. and Weinhold,E. (2001) Structure of the N6-adenine DNA methyltransferase M.*Taq*I in complex with DNA and a cofactor analog. *Nature Struct. Biol.*, **8**, 121–125.
54. Beck,C., Cranz,S., Solmaz,M., Roth,M. and Jeltsch,A. (2001) How does a DNA interacting enzyme change its specificity during molecular evolution? A site-directed mutagenesis study at the DNA binding site of the DNA-(adenine-N(6))-methyltransferase *Eco*RV. *Biochemistry*, **40**, 10956–10965.
55. Koradi,R., Billeter,M. and Wüthrich,K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics*, **14**, 51–55.