1084–1091 *Nucleic Acids Research, 1998, Vol. 26, No. 4*

Cloning and analysis of the four genes coding for *Bpu*10I restriction–modification enzymes

Kornelijus Stankevicius, Arvydas Lubys, Albertas Timinskas, Donatas Vaitkevicius and Arvydas Janulaitis*

Institute of Biotechnology, Graiciuno 8, Vilnius 2028, Lithuania

Received September 19, 1997; Revised and Accepted December 22, 1997 DDBJ/EMBL/GenBank accession no. Y14683

ABSTRACT

The *Bpu*10I R–M system from *Bacillus pumilus* 10, which recognizes the asymmetric 5′-CCTNAGC sequence, has been cloned, sequenced and expressed in *Escherichia coli*. The system comprises four adjacent, similarly oriented genes encoding two m5C MTases and two subunits of *Bpu*10I ENase (34.5 and 34 kDa). Both *bpu10IR* genes either in *cis* or *trans* are needed for the manifestation of R.*Bpu*10I activity. Subunits of R.*Bpu*10I, purified to apparent homogeneity, are both required for cleavage activity. This heterosubunit structure distinguishes the *Bpu*10I restriction endonuclease from all other type II restriction enzymes described previously. The subunits reveal 25% amino acid identity. Significant similarity was also identified between a 43 amino acid region of R.*Dde*I and one of the regions of higher identity shared between the *Bpu*10I subunits, a region that could possibly include the catalytic/Mg^{2+} binding center. The similarity between *Bpu*10I and *Dde*I MTases is not limited to the conserved motifs (CM) typical for m5C MTases. It extends into the variable region that lies between CMs VIII and IX. Duplication of a progenitor gene, encoding an enzyme recognizing a symmetric nucleotide sequence, followed by concerted divergent evolution, may provide a possible scenario leading to the emergence of the *Bpu*10I ENase, which recognizes an overall asymmetric sequence and cleaves within it symmetrically.

INTRODUCTION

Bacterial restriction and modification (R–M) systems are traditionally divided into three classes, designated type I, II and III, on the basis of their enzyme subunit composition, cofactor requirements, substrate specificity and reaction products (1). Type II restriction endonucleases are the most prolific, encompassing over 2500 enzymes (2). By definition, these enzymes consist of homodimers of a single polypeptide subunit and they require only Mg^{2+} for DNA cleavage (1,3). Most type II enzymes now known recognize short sequences in DNA, which vary from 4 to 8 bp in length, maintaining a dyad axis of symmetry (palindromes), and they cleave within those sequences. However, there is a substantial number of type II restriction endonucleases, that recognize asymmetric continuous or interrupted sequences, which cleave at a short distance from those sequences on one or both sides. In contrast to palindromic recognition sequences the asymmetric recognition sequences are 5′–3′ different in each strand. The possibility that recognition is mediated by symmetrically organized homodimers, as is observed in type II ENases interacting with palindromic nucleotide sequences is therefore excluded. Further characterization of such enzymes revealed additional differences from the rest of the type II endonucleases not only as expected in their monomer or heteromer versus homodimer structure, but in some cases also in other features (cofactor requirement, bifunctionality) so substantial, that it is now suggested that they may constitute new types (type IIS, type IV, *Bcg*I-like) of restriction endonucleases (3–6). These observations indicate that restriction endonucleases recognizing asymmetric nucleotide sequences are likely candidates for discovery of even greater diversity.

Among type II restriction endonucleases recognizing asymmetric nucleotide sequences there is a subgroup of enzymes [tentatively designated type IIT ENases (7)] which cleave within the asymmetric recognition sequence. No one representative of this group has yet been characterized beyond the determination of its specificity. One such enzyme, R.*Bpu*10I recognizes the asymmetric nucleotide sequence 5′-CCTNAGC, whose central pentanucleotide is represented by an interrupted palindrome, and which cleaves within it symmetrically, at positions –5/–2 in different strands (8). Preliminary experiments indicate that R.*Bpu*10I is composed of two non-identical subunits (9). We now provide detailed genetic and biochemical evidence to demonstrate that the *Bpu*10I R–M system from *Bacillus pumilus* 10 comprises four genes encoding two m5C MTases and two non-identical subunits of the restriction endonuclease. Properties of the *Bpu*10I R–M system and *Bpu*10I ENase make them unique among the type II enzymes that have been characterized so far. A plausible evolutionary scenario leading to the emergence of *Bpu*10I type enzymes concomitant with the generation of restriction endonucleases with specificities different from that of the progenitor is discussed.

*To who correspondence should be addressed. Tel: +370 2 642 468; Fax: +370 2 642 624; e-mail: janulaitis@fermentas.lt

MATERIALS AND METHODS

Bacterial strains and plasmids

Bacillus pumilus 10, used as the source of DNA for the cloning of the *Bpu*10I R–M system, was obtained from MBI Fermentas. The *Escherichia coli* K-12 strain RR1, (10) was used as a host for generating the *B.pumilus* 10 DNA library and for the propagation of recombinant plasmids carrying the cloned genes for R.*Bpu*10I or its individual subunits and *Bpu*10I MTases. *Escherichia coli* strain ER2267 of genotype [e14$^-$(*mcrA*$^-$)] *endA1 supE44 thi-1* Δ*(mcrC-mrr) 114::IS10* Δ*(argF-lac)U169 recA1/F′ proA*$^+$ *B*$^+$ *lacI*^{q} Δ*(lacz)M15 zzf::*mini-Tn*10* (Km^R) was used as host for subcloning and deletion procedures. The positive selection cloning vector pBR-R, containing the intact *cfr9IR* gene with unique cloning sites (11) was used for cloning of the *Bpu*10I R–M system. Plasmids pUC19 (12), pACYC184 (13) and pAL4A (A.L., unpublished) were used as vectors for subcloning and deletion experiments.

Media and transformation

Escherichia coli cells were grown in LB medium containing ampicillin (Ap, 60 μg/ml), kanamycin (Km, 50 μg/ml) and chloramphenicol (Cm, 30 μg/ml) as required. Cells were transformed using the $CaCl_2$–heat shock method (14). Transformants were selected by plating onto LB agar supplemented with appropriate antibiotics.

Enzymes and chemicals

Restriction enzymes, T4 DNA ligase, DNA polymerase I large fragment (Klenow), nuclease *Bal*31, bacterial alkaline phosphatase (BAP), *Exo*III/S1 Deletion Kit, DNA Labeling Kit (version 2.0), *Sma*I-digested and dephosphorylated pUC19 and sequencing primers were products of MBI Fermentas. The DNA Sequencing Kit used was purchased from Pharmacia. All enzymes and kits were used according to their respective manufacturer's recommendations. [α-^{33}P]dATP was obtained from Izotop (St. Petersburg, Russia).

DNA preparation and manipulation

Bacillus pumilus 10 genomic DNA was extracted and purified as described by Marmur (15). Plasmid DNAs were prepared by the alkaline lysis procedure (16) and were further purified by binding to glass powder (17). Restriction and deletion mapping, agarose gel electrophoresis, isolation of individual DNA restriction fragments from agarose gels, subcloning of DNA fragments and *Bal*31 deletions were carried out by standard procedures (14). *Exo*III/S1 deletions were obtained following the manufacturer's recommendations.

Library construction and selection of the clones harboring *Bpu*10I methyltransferase

An aliquot (50 μg) of *B.pumilus* 10 genomic DNA was partially digested with *Hin*1II and ligated with T4 DNA ligase at 16°C for 24 h in a 800 ml reaction volume to 7 mg *Pae*I-cleaved, BAP-dephosphorylated pBR-R vector DNA. The ligation mixture was used to transform competent *E.coli* RR1 cells. Total plasmid DNA (10 μg) isolated from the pooled 100.000 Cm^R transformants was digested with excess R.*Bpu*10I (30 U) for 3 h and the transformed back into RR1. Plasmid DNA of 24 randomly picked individual transformants obtained was then screened for resistance to R.*Bpu*10I digestion.

Mapping of the *B.pumilus* 10 genomic R–M locus

Southern-transfer of *B.pumilus* 10 genomic or pBpu10IM9.0 recombinant plasmid DNA, digested with various restriction enzymes (single and double digests), was performed as previously described (14). A DNA Labeling Kit was used for the preparation of a radioactive DNA probe containing a *Hin*dIII fragment of the cloned *Bpu*10I MTase gene.

DNA sequence determination

Overlapping deletions used in DNA sequencing were constructed using the nuclease *Bal*31 and *Exo*III/S1 Deletion Kit. Sequencing was performed by the dideoxynucleotide chain-termination method (18) using the DNA Sequencing Kit, [α-^{33}P]dATP, M13/pUC (direct, reverse) standard sequencing primers and double-stranded, supercoiled plasmid DNA as template. The reaction products were resolved by electrophoresis on wedge-shaped gels.

Analysis of *Bpu*10I endonuclease and methylase activity

In vitro endonuclease activity was tested by incubation of various amounts (1, 3, 5 μl) of cell-free extracts prepared as described (19) with 2 μg of λ DNA at 37°C for 40 min in 40 μl reaction mixture containing 10 mM Tris–HCl, pH 8.5, 10 mM $MgCl_2$, 100 mM KCl and 0.1 mg/ml BSA (standard reaction mixture), followed by electrophoresis in 0.8% agarose gels. Chromatographic fractions containing individual subunits were assayed by complementation using aliquots of the crude cell extract prepared from recombinant *E.coli* cultures expressing the alternate subunit, as a supplement to the standard reaction mixture. To determine the *Bpu*10I-specific modification generated *in vivo*, plasmid DNA isolated from transformants was challenged with an excess of R.*Bpu*10I followed by agarose gel electrophoresis.

Amino acid sequence comparison

The deduced amino acid sequences of all ORFs (50 and more amino acids residues) translated in six reading frames, were compared to all sequences deposited with the EMBL (Release 42) and the SWISS-PROT (Release 31) sequence data bases using the BLAST (20) procedure. Alignment of selected sequences was obtained using the MULTALIN (21) procedure applying the BLOSUM 62 (22) amino acid scoring matrix. Similarity of the *Bpu*10I subunit and MTase sequences to those of the available restriction endonucleases and m5C methyltransferases, respectively, was tested using the HR-SEARCH procedure (23).

Gel electrophoresis of proteins

Gel electrophoresis of proteins under denaturing conditions was performed as previously described (24). SDS–PAGE was carried out on 10% separating gels. Protein bands were visualized after Coomasie blue R250 staining.

Cell growth and initial fractionation

Escherichia coli RRI[pAL-Bpu10IRα] and *E.coli* RRI[pAL-Bpu10IRβ], expressing the individual genes for the subunits of R.*Bpu*10I (α or β, respectively) cloned in the pAL4A vector,

under the control of the thermoinducible λ P_L promoter, were used as the source of subunits. The cells were grown into late logarithmic phase at 37°C with aeration in L broth containing 10 g/l trypton, 5 g/l yeast extract and 5 g/l NaCl. The cells were harvested by centrifugation and stored at –20°C. All further steps were carried out at 40°C. Cells (15 g) were thawed in 60 ml of buffer A (10 mM K-phosphate pH 7.0, 7 mM ME, 1 mM Na_2EDTA) containing 0.25 M KCl. Then, following cell rupture by sonication, 10% Polymin P solution was gradually added to a final concentration of 0.7%, with stirring. The precipitate was removed by centrifugation at 12 000 *g* for 10 min. Soluble proteins were salted out with $(NH_4)_2SO_4$ at 80% saturation and recovered by centrifugation. The pellet was dissolved in 1/3 of the initial volume of buffer A and dialyzed against buffer A containing 0.15 M KCl (buffer B). Dialysates were used for chromatographic purification of the individual α or β subunits of R.*Bpu*10I.

Purification of the R.*Bpu*10I α subunit from *E.coli* RRI[pAL-Bpu10IRα]

Phosphocellulose chromatography. The dialysate was applied to a phosphocellulose P-11 column (1.5 × 20 cm) equilibrated with buffer B. The column was washed with 80 ml of the same buffer and eluted with a 400 ml linear gradient of 0.15–1 M KCl in buffer A. *Bpu*10Iα eluted around 0.65 M KCl. The peak fractions containing the α subunit were pooled and dialyzed against buffer B.

Bordo–Sepharose chromatography. The dialysate was applied to a Bordo–Sepharose column (1 × 10 cm) equilibrated with buffer B. After the application of the dialysate, the column was washed with 20 ml of buffer B and then eluted with a 100 ml linear gradient of 0.15–1 M KCl in buffer A. The α subunit eluted at ~0.43 M KCl. The peak fractions were pooled and dialyzed against buffer B.

Heparin–Sepharose chromatography. The dialysate was applied to a Heparin–Sepharose (1 × 10 cm) column, equilibrated with buffer B. The column was washed with 20 ml of the same buffer and eluted with a 100 ml linear gradient of 0.15–1 M KCl in buffer A. The subunit eluted around 0.5 M KCl. Fractions containing the α subunit were pooled, dialyzed against storage buffer (10 mM Tris–HCl, pH 7.5, 50 mM KCl, 1 mM DDT, 0.1 mM Na_2EDTA, 50% glycerol) and stored at –20°C.

Purification of the R.*Bpu*10I β subunit from *E.coli* RRI[pAL-Bpu10IRβ]

Phosphocellulose and Bordo–Sepharose chromatography. Fractionation on phosphocellulose and Bordo–Sepharose was performed as described above for the isolation of the α subunit. In this case though a (1.5 × 20 cm) Bordo-Sepharose column was used and the limits of the KCl linear gradient used were different (0.15–0.6 M phosphocellulose and 0.2–1 M Bordo–Sepharose). R.*Bpu*10Iβ eluted around 0.35 M KCl from the phosphocellulose column and at ~0.6 M KCl from the Bordo–Sepharose column. After Bordo–Sepharose chromatography the fractions containing the β subunit were pooled and dialyzed against buffer B.

AH–Sepharose chromatography. The dialysate was applied to an AH–Sepharose column (1.5 × 20 cm) equilibrated with buffer B. The column was washed with 80 ml of the same buffer and then eluted with a 400 ml linear gradient of 0.15–0.8 M KCl in buffer A. The subunit eluted at around 0.3 M KCl. Fractions containing R.*Bpu*10Iβ subunit were pooled and dialyzed against buffer B.

Heparin–Sepharose chromatography. The dialysate was applied to a Heparin–Sepharose column (1.5 × 20 cm) equilibrated with buffer B and then washed with the same buffer. The column was eluted with a 200 ml linear gradient of 0.15–0.8 M KCl in buffer A. The R.*Bpu*10Iβ subunit eluted at ~0.32 M KCl. Fractions containinng the β subunit were pooled, dialyzed against storage buffer and stored at –20°C.

RESULTS AND DISCUSSION

Cloning of the *Bpu*10I methyltransferase gene

Selection of the gene coding for *Bpu*10I MTase was based on the resistance of self modifying recombinant plasmids to digestion by *Bpu*10I ENase. Twenty four randomly picked transformants obtained after the selection procedure, described under Materials and Methods, were further screened for the presence of *Bpu*10I specific modification. Plasmid DNA of 22 out of 24 clones was found to be protected against *Bpu*10I ENase challenge *in vitro*. None of these clones showed any *Bpu*10I ENase activity when assayed *in vitro*. Restriction mapping of the plasmids resistant to *Bpu*10I cleavage revealed that they contained a 1.65 kb *Hin*dIII fragment in common (data not shown). The recombinant plasmids containing the shortest (4 kb, pBpu10IM4.0) and the longest (9 kb, pBpu10IM9.0) of the cloned *B.pumilus* 10 DNA fragments were selected for further experiments.

Deletion mapping of the pBpu10IM4.0 plasmid showed that the 2.2 kb region on the right side of the *Mph*1103I site contained a functionally active *Bpu*10I MTase gene (Fig. 1A). Southern hybridization, using the radioactive 1.65 kb *Hin*dIII DNA fragment from the cloned MTase gene as a probe, and *B.pumilus* 10 or pBpu10IM9.0 DNA, digested with various restriction endonucleases, indicated that the pBpu10M9.0 plasmid contains the intact genomic DNA fragment of *B.pumilus* 10 (data not shown). The DNA flanking the 1.65 kb *Hin*dIII fragment is large enough to include the ENase gene, usually closely linked to the genes of the cognate MTase. However, clones harboring the pBpu10IM9.0 plasmid did not reveal R.*bpu10I* activity, indicating that the corresponding gene was either not cloned or not expressed in *E.coli* cells. Sequencing of the *bpu10IM* gene and the flanking regions, revealed genes encoding the complete *Bpu*10I R–M system (see below).

Gene structure of the *Bpu*10I R–M system

A 4410 bp region of the cloned 9.0 kb fragment encompassing the gene for *Bpu*10I MTase was sequenced on both strands (EMBL accession no. Y14683). Four ORFs, all oriented in the same direction, were identified (Fig. 1B and C). They encode the *Bpu*10I R–M enzymes: two separate m5C MTases (*Bpu*10IC1 and *Bpu*10IC2) and two ENase heterosubunits (*Bpu*10Iα and *Bpu*10Iβ) (see below). The first and the second ORFs corresponded to the genes *bpu*10IMC1 and *bpu10IMC2* of two m5C MTases. ORF1 was 1194 bp long (termination codon not included), extended from nt 110 to 1303, and encoded a protein of 398 amino acids. It overlapped by 1 nt (or by 4 nt if the stop codon TGA of ORF1 was included) with the downstream ORF2 (nt 1303–2478 = 1176 bp, 392 amino acids; Fig. 1C). Putative RBS sequences AGAGG (nt 95–99) and GGGGG (nt 1288–1292; not shown in Fig. 1C) were detected upstream of both ORFs, respectively. A

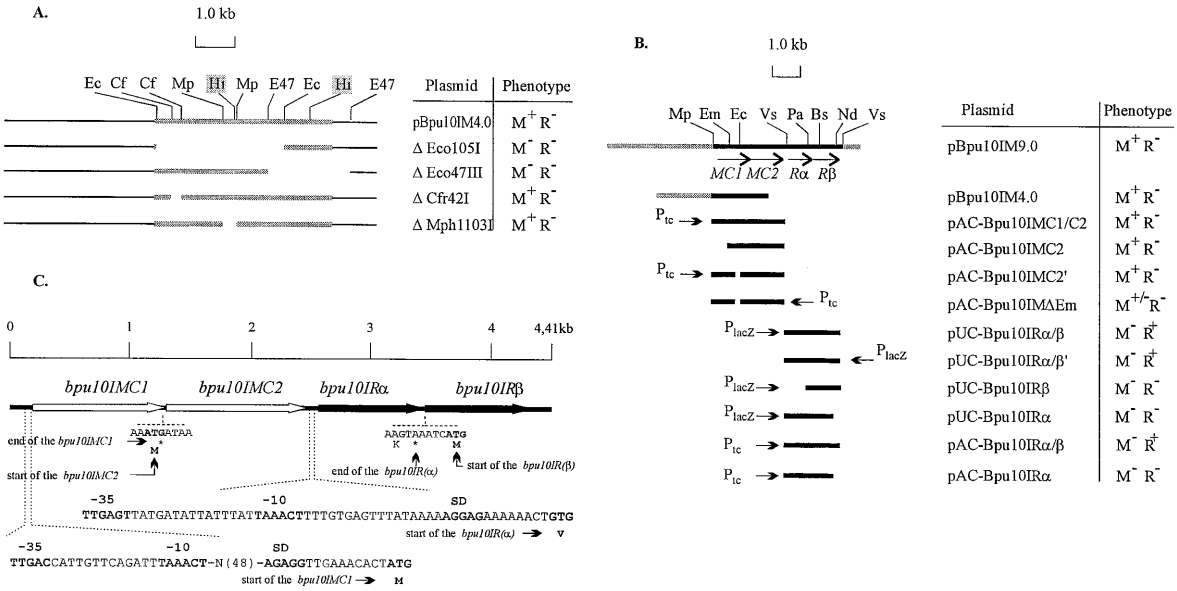

Figure 1. (**A**) Restriction endonuclease and deletion mapping experiments using the plasmid pBpu10IM4.0. Thin lines indicate the cloning vector pBR-R, hatched lines, the 4 kb cloned *B.pumilus* 10 DNA fragment. Restriction endonuclease sites are abbreviated as follows: Cf, *Cfr*42I; Ec, *Eco*105I; E47, *Eco*47III; Hi, *Hin*dIII; Mp, *Mph*1103I; Δ, deletion between restriction sites as indicated. (**B**) Subcloning and deletion mapping of the genes encoding the *Bpu*10I R–M system. Only *B.pumilus* 10 DNA fragments cloned or subcloned into pBR-R, pUC19 or pACYC184 vectors are represented. The black lines represent the sequenced DNA fragment. Directions of transcription and positions of the *Bpu*10I R–M genes are indicated by large arrows: *MC1* corresponds to *bpu10IMC1*; *MC2*, *bpu10IMC2*; *R*α, *bpu10IR*α; *R*β, *bpu10IR*β. Direction of transcription from the known vector promoters adjacent to the cloned fragments is indicated by small arrows. Recognition sites of the restriction endonucleases used in subcloning and deletion mapping experiments are abbreviated as follows: Mp, *Mph*1103I; Em, *Eam*1104I; Ec, *Eco*105I; Vs, *Vsp*I; Pa, *Pae*I; Nd, *Nde*I. Plasmid pAC-Bpu10IMC1/C2 was obtained after subcloning the 2.51 kb *Mph*1103I–*Vsp*I DNA fragment, containing the intact genes of *bpu10IMC1* and *bpu10IMC2*, from pBpu10IM9.0 into *Eco*32I-digested pACYC184. To obtain pAC-Bpu10IMC2ΔEm the 5′ part of the *bpu10IMC1* and upstream DNA region including the heterogenous P_{tet} promoter from the plasmid pAC-Bpu10IMC1/C2 was deleted (the second *Eam*1104I site being located within the vector). pAC-Bpu10IMC2 is a derivative of the pAC-Bpu10IMC1/C2, in which a part (~0.4 kb) of the *bpu10IMC1* gene was deleted by restriction with *Eco*105I, then being treated with *Bal*31 exonuclease. pAC-Bpu10IMC2′ was obtained after subcloning the 2.2 kb *Bfm*I–*Nhe*I DNA fragment from the pAC-Bpu10IMC2 (restriction sites reside within the vector; not shown) into *Eco*32I-cleaved pACYC184. Plasmids pUC-Bpu10IRα/β, pUC-Bpu10IRα/β′ and pAC-Bpu10IRα/β were obtained after insertion of the 1.9 kb *Vsp*I fragment from pBpu10IM9.0 into *Sma*I-digested pUC19 and *Eco*32I-cleaved pACYC184, respectively. Plasmids pUC-Bpu10IRβ and pUC-Bpu10IRα were constructed by deletion of the 1 kb *Pae*I or the 0.4 kb *Nde*I DNA fragments respectively from the plasmid pUC-Bpu10IRα/β (*Pae*I and *Nde*I sites located within the vector are not shown). The plasmid pAC-Bpu10IRα was formed by deletion of the 2 kb *Nde*I–*Eco*81I DNA fragment from pAC-Bpu10IRα/β (*Eco*81I site resides within the vector). R–M phenotypes were determined as described under Materials and Methods. M^+, MTase activity; M^-, no activity; $M^{+/-}$, partial activity; R^+, ENase activity; R^-, no activity. (**C**) Gene organization of the *Bpu*10I R–M system. White and black arrows indicate direction of transcription and relative position of the *bpu10I* R-M genes in the sequenced DNA fragment. Nucleotide sequences of the putative promoters and RBS are presented below. The putative start codons, RBS and –35, –10 promoter sequences are indicated in bold. Asterisks denote the stop codons.

putative promoter region was found only upstream of the *bpu10IMC1* gene. The third ORF, which corresponded to the *bpu10IR*α gene, started 95 bp downstream from the last nucleotide of *bpu10IMC2* and encoded 294 amino acid residues (34.5 kDa) of the *Bpu*10I ENase α subunit (start codon GTG; nt 2573–3454 = 882 bp). It was separated by the stop codon TAA and the ATC triplet (nt 3455–3460), from the structural part of the gene *bpu10IR*β encoding R.*Bpu*10Iβ (ORF4; start codon ATG; nt 3461–4324 = 864 bp; 288 amino acids = 34 kDa). The putative Shine–Dalgarno sequences AGGAGA (nt 2560–2565) and AGGA (nt 3448–3451; not shown in Fig. 1C) were detected for ORF3 and ORF4, respectively. Putative promoter sequences were found only upstream from ORF3. These findings suggest that genes of the *Bpu*10I R–M system are organized into two separate operons: one for the genes of the two *Bpu*10I MTases and another for the two subunits of *Bpu*10I ENase.

Gene assignment

Deletion and subcloning experiments were carried out to investigate the expression and function of the four ORFs identified.

Bpu10I MTases. The intact ORF1 and 5′ terminal part of the ORF2 in the 4.0 kb fragment were cloned (Fig. 1A and B). Thus, ORF1 is clearly expressed in *E.coli* as the *Bpu*10I specific MTase (M.*Bpu*10IC1). Incomplete plasmid DNA protection (~80%) from R.*Bpu*10I cleavage was found when the 0.69 kb DNA fragment containing the N-terminal part and the putative promoter of the M.*Bpu*10IC1, together with the exogenous promoter of the tetracycline resistance gene, was deleted from the plasmid pAC-Bpu10IMC1/C2 (pAC-Bpu10IMΔEm; Fig. 1B). Nevertheless, these results indicate that the ORF2 encodes another *Bpu*10I MTase (M.*Bpu*10IC2). These conclusions were supported by the sequence analysis of the ORF1 and ORF2, which revealed conserved amino acid motifs typical of DNA m5C MTases (25). Low level *bpu10IMC2* gene expression is consistent with the lack of promoter sequences upstream of the gene. As expected, given the operon like organization of *Bpu*10I MTase genes, the level of *bpu10IMC2* expression was increased under the control of the promoter located in front of *bpu10IMC1* (Fig. 1C) to a level which proved high enough to protect DNA against R.*Bpu*10I challenge *in vitro* (pAC-Bpu10IMC2 and pAC-Bpu10IMC2′; Fig. 1B).

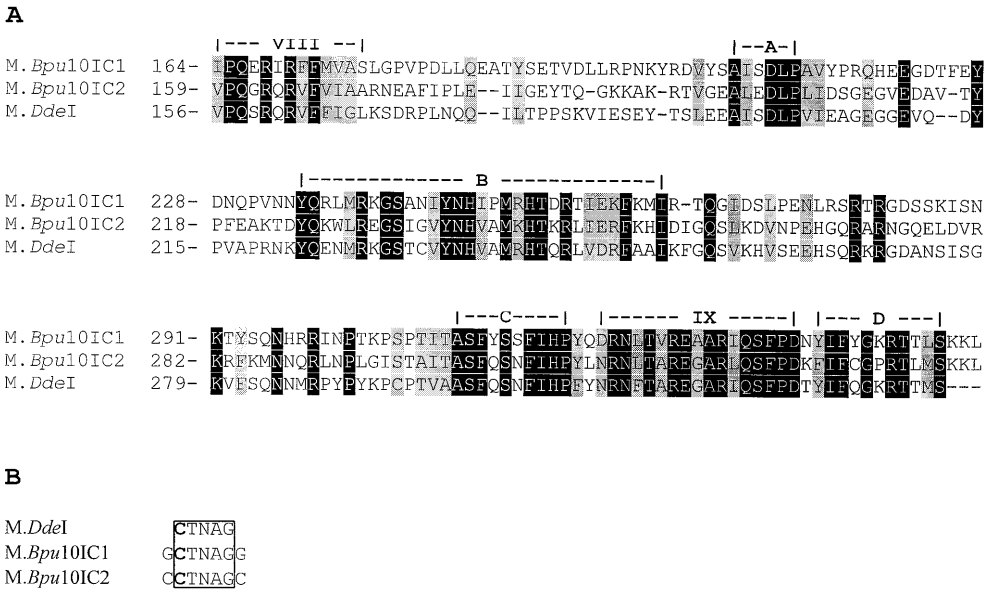

Figure 2. Alignment of M.*Bpu*10IC1, M.*Bpu*10IC2 and M.*Dde*I amino acid sequences (**A**) and their DNA target sequences (**B**). (A) Only the protein sequences encompassing CM VIII and CM IX, which are common to m5C MTases, the variable region between them, and conserved region D downstream of CM IX, are shown. The numbers on the left margin denote the amino acid position relative to the N-terminus. White letters on the black background and the black letters on the shaded background indicate amino acid residues that are common and similar in aligned sequences, respectively. Similar amino acids are grouped as follows: (K, R), (T, S), (D, E, N, Q), (A, V, I, L, M, F, W, Y), (A, G). Uppercase letters (A, B, C, D) above aligned sequences correspond to the conserved regions common for all three aligned MTases. (B) The 5′-CTNAG sequences common to the recognition sequences of *Bpu*10I MTases and M.*Dde*I are boxed. The target cytosines are boldfaced.

Bpu10I ENase. No activity of the *Bpu*10I ENase was detected in crude extracts prepared from the cells carrying the pBpu10I9.0 plasmid (Fig. 1B). The 1.9 kb DNA fragment encompassing both the intact ORF3 and ORF4, when subcloned from the plasmid pBpu10IM9.0 into the pUC19 vector in both directions (pUC-Bpu10IRα/β and pUC-Bpu10IRα/β′; Fig. 1B), manifest *Bpu*10I ENase activity *in vitro*. This indicated that ORF3 and/or ORF4 encoded a restriction endonuclease and that the endogenous promoter(s) operates in *E.coli*. Interestingly, host cell DNA protection by specific methylation was not required to generate viable transformants expressing *Bpu*10I ENase activity, as has been observed in some other R–M systems [for references see (26)]. In cells harboring the intact *Bpu*10I R–M system (plasmid pBpu10IM9.0) no activity of the *Bpu*10I ENase was detected. It is unlikely, however, that some increase of plasmid copy number, after the subcloning of the *Bpu*10I ENase gene(s) from pBR-R vector (pBpu10IM9.0) into the pUC19 (pUC-Bpu10IRα/β; Fig. 1B) could result in so dramatic an increase in enzyme activity. The ORF3 and ORF4 were precisely excised from the pBpu10IM9.0 plasmid before subcloning, leaving a 4 kb DNA region located upstream of the MTase genes and a 95 bp intergenic region separating ORF2 and ORF3. The existence of some regulatory elements of R.*Bpu*10I expression within these regions cannot be excluded.

No *Bpu*10I ENase activity was observed in crude cell extracts of the clones harboring deletion derivatives, where either the ORF3 (pUC-Bpu10IRβ) or ORF4 (pUC-Bpu10IRα) were inactivated (Fig. 1B). When plasmids pAC-Bpu10IRα and pUC-Bpu10IRβ were transformed into the same cell, on the other hand, *Bpu*10I ENase activity was detected in cell extracts (data not shown). This suggests that the proteins encoded by ORF3 and ORF4 function in a mutually dependent fashion, such that both genes, either in *cis* or *trans*, are needed for the manifestation of R.*Bpu*10I activity. In parallel with these observations, crude cell extracts prepared from the individual clones carrying the cloned intact ORF3 or ORF4 (pUC-Bpu10IRα or pUC-Bpu10IRβ) complemented when added to the reaction mixture (data not shown).

Comparison and analysis of deduced amino acid sequences

Bpu10I MTases. A 39% identity between M.*Bpu*10IC1 and M.*Bpu*10IC2 was observed with an additional 24% similarity derived from conservative substitutions. Both possess 10 conserved amino acid motifs (CM; not shown), characteristic of DNA m5C MTases (25). The greatest similarity amongst all m5C MTases was observed in comparison with *Dde*I. The level of similarity between *Dde*I and the *Bpu*10I MTases was approximately equal to that between the *Bpu*10I MTases themselves. This similarity is not limited to the conserved motifs but also extend into the region that lies between the CMs VIII and IX and beyond CMs IX where four conserved regions (CR) of higher identity are found (CRs A, B, C and D; Fig. 2). Genetic and biochemical evidence has been obtained which indicates that a variable region located between CM VIII and CM IX of both phage multi-specific and bacterial mono-specific MTases contains a target recognition domain (TRD) which defines specificity of these enzymes as well as the base to be methylated within the target sequence (27–30). This general concept was confirmed and refined significantly by the resolution of the atomic structures of two m5C MTases, *Hha*I (31) and *Hae*III (32) complexed with substrate DNA.

The similarity between variable regions of m5C MTases is observed only in some of the enzymes which recognize identical

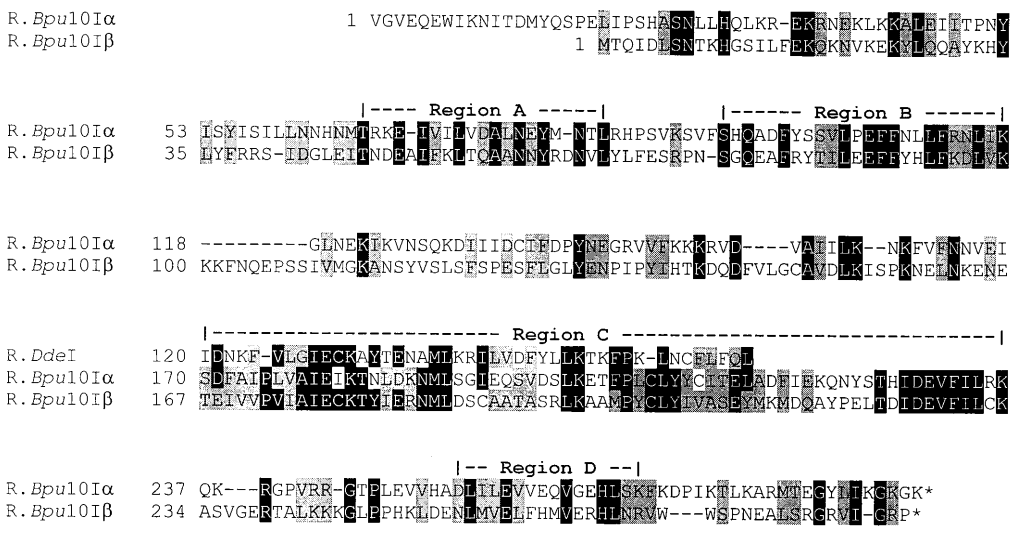

Figure 3. The amino acid alignment of R.*Bpu*10Iα, *Bpu*10Iβ and the 43 amino acid region of R.*Dde*I discussed. The conventions used are the same as those in Figure 2. In the alignment of R.*Dde*I with R.*Bpu*10I subunit sequences aa residues common to R.*Dde*I and at least one of the subunits are shown. Uppercase letters (A, B, C, D) above aligned sequences correspond to the regions of higher identity between the subunits of R.*Bpu*10I.

or related nucleotide sequences which modify the same target base (33). M.*Dde*I recognizes the 5′-CTNAG pentanucleotide and modifies the C base (34). The same pentanucleotide represents the central symmetrical part of the nucleotide sequences recognized by *Bpu*10I MTases. *Bpu*10IC1 MTase modifies the 5′-GCTNAGG sequence (target base underlined), while M.*Bpu*10IC2–5′-CCTNAGC (Z.Maneliene, personal communication) i.e. modifies the C base which belongs to the 5′-CTNAG pentanucleotide. The sequence similarities between the variable regions of M.*Dde*I and *Bpu*10I MTases may therefore reflect a similarity of structural organisation between the M.*Dde*I specificity domain and the component of the *Bpu*10I MTases which recognises the same target sequence 5′-CTNAG.

One monomeric MTase is sufficient for the recognition and modification of a palindromic nucleotide sequence (33). For type IIS MTases it has been demonstrated that a range of variations in enzyme structural organisation are used in the modification of asymmetric recognition sequences. These include a single monomeric bifunctional enzyme modifying adenine residues in both strands of target DNA (35); DNA MTases yielding m6A or m5C on complementary strands in the reaction catalysed by a single tandemly arranged bifunctional MTase or two separate m6A and m5C MTases (36; J.Bitinaite, personal communication); and two separate m5C MTases each responsible for the methylation of different DNA strands (11,37). The *Bpu*10I MTase is likely to belong to the last group.

Bpu10I ENase. Comparison of the R.*Bpu*10Iα and R.*Bpu*10Iβ deduced amino acid sequences revealed quite a low degree of similarity (25% of amino acids are identical and 17% are functionally similar). Analysis using HR-SEARCH (23), however, resulted in the extraction of four regions of higher identity (43% for A region; 46%, B; 37%, C and 31%, D; Fig. 3). The subunits of R.*Bpu*10I resembled none of the protein sequences in the EMBL and SWISS-PROT databases, nor did they resemble any of those in our personal restriction enzyme data base. The most significant regional similarity was identified between the 43 amino acid region in the *Dde*I ENase (which extends from amino acid 120 to 162 of the R.*Dde*I), and the N-terminal part of the conserved region C (28% identical amino acids for R.*Dde*I versus R.*Bpu*10Iα or R.*Bpu*10Iβ). The consensus motif P(E/D)$X_{9\text{-}18}$(E/D)XK, identified as essential for catalysis and Mg^{2+} binding in some restriction endonucleases (38), could not be found in R.*Bpu*10I subunits. However, the EXK motif preceded 8 amino acids upstream by a negatively charged amino acid residue (D or E) is conserved, as indicated by alignment of R.*Bpu*10I subunit sequences with that of R.*Dde*I (Fig. 3; region C), elements reminiscent of the P(E/D)$X_{9\text{-}18}$(E/D)XK consensus motif (except for P). If experimental evidence is obtained to support the suggestion that the catalytic/Mg^{2+} binding motif (or part of it) of R.*Bpu*10I and R.*Dde*I is of the structure (D/E)X_8EXK, then it could represent a new variant.

Subunit structure of the *Bpu*10I ENase

Initial experiments, using crude cell extracts prepared from *B.pumilus* 10, revealed a total loss of enzyme activity (measured using the standard reaction mixture) after chromatography on phosphocellulose P11. The separation of the two R.*Bpu*10I heterosubunits, whose existence was demonstrated on the basis of genetic evidence, explained these findings. Indeed *Bpu*10I ENase activity was recovered in two peaks after fractionation on phosphocellulose P11, using the standard reaction mixture supplemented with aliquots of crude cell extracts prepared from the recombinant clones *E.coli* RRI[pAL-Bpu10IRα] or *E.coli* RRI[pAL-Bpu10IRβ] expressing the individual subunits α and β, respectively. *Bpu*10I ENase activity recovered eluted at around 0.33 M (β subunit) and 0.66 M KCl (α subunit). *Bpu*10I subunits were purified from the recombinant clones to apparent homogeneity. The molecular masses of these proteins (Fig. 4A) were determined to be close to those predicted for the α and β subunits on the basis of their nucleotide sequences (34.5 kDa α subunit and

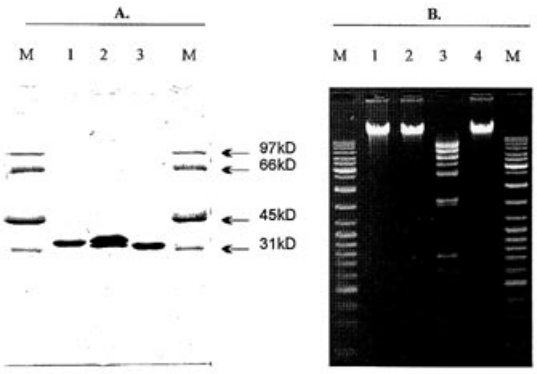

Figure 4. (**A**) SDS–PAGE electrophoresis of R.*Bpu*10Iα and R.*Bpu*10Iβ. An aliquot (0.8 µg) of each subunit was prepared and electrophoresed on a 10% SDS–polyacrylamide gel as described under Materials and Methods. M, molecular weight marker (Boehringer Ingelheim); lane 1, R.*Bpu*10Iα; lane 2, R.*Bpu*10Iα + R.*Bpu*10Iβ; lane 3, R.*Bpu*10Iβ. (**B**) Restriction activity of the R.*Bpu*10I. For each reaction 1.5 µg of λ DNA were incubated with purified α and β subunits of R.*Bpu*10I. Products were then applied to a 0.8% agarose gel, using the same conditions as described under Materials and Methods. Lane 1, λ DNA; lane 2, λ DNA + R.*Bpu*10Iα; lane 3, λ DNA + R.*Bpu*10Iα and β; lane 4, λ DNA + R.*Bpu*10Iβ.

34 kDa β subunit). Both subunits were required for R.*Bpu*10I activity (Fig. 4B). Investigation of the *Bpu*10I ENase oligomeric organization, however, under non-denaturing conditions, was hampered due to the irreversible adsorption (or inactivation) of at least of one of the subunits, to the gel filtration media used. During PAGE electrophoresis under non-denaturing conditions the mixture of α and β subunits formed heavy aggregates which failed to enter the gel further than a few mm (data not shown). The same was observed when the individual subunits were subjected to electrophoresis.

The two ORFs encoding *Bpu*10I ENase are separated by the translation stop codon TAA followed by the ATC codon which precedes the putative ATG start codon for *bpu10IR*β (Fig. 1C). The assignment of the initiation codon was further confirmed by the N-terminal sequence of the purified β subunit. The experimentally determined N-terminal pentapeptide matched the Thr-Gln-Ile-Asp-Leu sequence deduced from the nucleotide sequence except for the absence of the N-terminal Met. It could not be excluded though that the R.*Bpu*10I produced in *B.pumilus* 10 is composed of one polypeptide (e.g. due to suppression of stop codon). Fractionation of the crude cell extract prepared from *B.pumilus* 10, the natural host of the *Bpu*10I R–M system, on the phosphocellulose P11 column, revealed a separation of the subunits identical to that for the subunits expressed in the recombinant clones. These findings indicate, that the molecular organization of R.*Bpu*10I, generated from the recombinant clone, is the same as that in *B.pumilus* 10, which demonstrates that R.*Bpu*10I is composed of two heterosubunits. The association between subunits seems to be particularly weak, such that numerous attempts to purify *Bpu*10I ENase proved unsuccessful. Instead, α and β subunits can be purified and when mixed, reconstitute restriction activity. This indicates that subunit interaction necessary for target recognition and (or) catalysis, is mediated by substrate. Therefore, it can be concluded that the active form of *Bpu*10I corresponds to the heteromer.

*Bpu*10I ENase's structure establishes it as different from other restriction endonucleases, recognizing asymmetric nucleotide sequences (4–6), all of which originally were classified as type II enzymes. The diversity in structure, cofactor requirement, type of DNA cleavage and functional properties (monofunctionality verses bifunctionality) of this group of enzymes, confirms this group as distinct from the group recognizing palindromes, which are homodimers activated by Mg^{2+}. Among the enzymes recognizing asymmetric nucleotide sequences the properties of R.*Bpu*10I are most close to the type IIS ENases (4,39). Both are monofunctional enzymes revealing only DNA cleavage activity and requiring only Mg^{2+} for activity. They differ though, in position of cut point relative to the asymmetric recognition sequence (R.*Bpu*10I within the sequence, type IIS – outside) and protein structure (heteromer versus monomer). The modification component of some type IIS R–M systems is represented by two MTases (11,37) as in case of *Bpu*10I R–M system. The *Bpu*10I R–M system consists of four genes (and proteins), which is unique among other representatives of the type II R–M systems. Thus, consistent with the subdivision of the class II ENases to establish the group of type IIS enzymes (4), an additional group designated type IIT might be introduced for enzymes like R.*Bpu*10I as has been previously proposed by Kessler and Manta (7).

The possibility of heteromeric structure of type II restriction endonucleases recognizing asymmetric nucleotide sequences has previously been addressed (40). Co-purification of two polypeptides of slightly different size during isolation of *Cau*II ENase, which recognizes the 5′-CCSGG nucleotide sequence was used as an indication of heteromeric composition. Biochemical and genetic data on R.*Bcn*I, an isoschizomer of R.*Cau*II, unambiguously shows that *Bcn*I ENase operates in the form of a single protein (41; A.J., unpublished). However, it cannot been excluded that R.*Cau*II recognizes the 5′-CCSGG sequence as asymmetric, while R.*Bcn*I recognizes the same sequence as symmetric.

Evolutionary implications

It is generally assumed that restriction endonucleases have evolved numerous times (42). This assumption, however, does not exclude the possibility that ENases emerged independently to evolve further, generating new specificities i.e. recognizing nucleotide sequences possibly related to that of a predecessor (43,44). Although the overall primary sequence similarity between subunits of the *Bpu*10I ENase is quite low, it is high enough to indicate the similarity of their 3-dimensional structures (45) and their evolutionary relatedness (46). This notion is supported by the presence of several regions of higher identity which are distributed over the aligned sequences of R.*Bpu*10I subunits in the same relative positions. R.*Bpu*10I recognizes an overall asymmetric sequence whose central part encompasses the interrupted palindrome (5′-CTNAG) and cleaves symmetrically within this part (8). It is possible that a restriction endonuclease recognizing a symmetric sequence (e.g. CTNAG or CCTNAGG) may have been a predecessor of R.*Bpu*10I. Such a progenitor gene would encode one polypeptide, which after forming a homodimer would be sufficient for recognition of the palindromic nucleotide sequence. Divergent concerted evolution of duplicated genes may have generated an enzyme of a new specificity, each subunit of which might recognize different non-symmetrical parts of the target (outside bases) in a different way. The similarity between subunits of the R.*Bpu*10I and R.*Dde*I is restricted to a

stretch of 43 amino acids that plausibly includes the catalytic/Mg^{2+} binding center. Therefore, R.*Dde*I can not be assumed to be a recent predecessor of R.*Bpu*10I, although this would not exclude the possibility that the catalytic/Mg^{2+} binding center and recognition of the 5′-CTNAG sequence might reside within the 43 amino acid stretch for both *Dde*I and *Bpu*10I ENases. A scenario similar to that described above for *Bpu*10I ENase might well be deduced for *Bpu*10I MTases.

ACKNOWLEDGEMENTS

We thank MBI Fermentas for providing enzymes and chemicals, Elisabeth A. Raleigh for *E.coli* strain ER2267. We are also grateful to Elena Zareckaja for sequencing N-terminal part of the R.*Bpu*10Iβ subunit, Saulius Klimasuskas for discussion and to Barbara Richmond-Smith and Geoff Werstuck for linguistic help.

REFERENCES

1 Yuan,R. (1981) *Annu. Rev. Biochem.*, **50**, 285–315.
2 Roberts,R.J. and Macelis,D. (1997) *Nucleic Acids Res.*, **25**, 248–262.
3 Pingoud,A. and Jeltsch,A. (1997) *Eur. J. Biochem.*, **246**, 1–22.
4 Szybalski,W., Kim,S.C., Hasan,N. and Podhajska,A.J. (1991) *Gene*, **100**, 13–26.
5 Janulaitis,A., Petrusyte,M., Maneliene,Z., Klimasauskas,S. and Butkus,V. (1992) *Nucleic Acids Res.*, **20**, 6043–6049.
6 Kong,H., Roemer,S.E., Waite-Rees,P.A., Benner, J.S., Wilson,G.G. and Nwankwo,D.O. (1994) *J. Biol. Chem.*, **269**, 683–690.
7 Kessler,C. and Manta,V. (1990) *Gene*, **92**, 1–248.
8 Degtyarev,S.Kh., Zilkin,P.A., Prihodko,G.G., Repin,V.E. and Rechkunova,N.I. (1989) *Molekularnaja Biologija*, **23**, 1051–1056.
9 Stankevicius,K., Lubys,A., Timinskas,A. and Janulaitis,A. (1996) *Biologija*, **2**, 51–53.
10 Maniatis,T., Fritsch,E. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
11 Lubys,A., Lubiene,J., Kulakauskas,S., Stankevicius,K., Timinskas,A. and Janulaitis,A. (1996) *Nucleic Acids Res.*, **24**, 2760–2766.
12 Yanisch-Perron,C., Vieira,J. and Messing,J. (1985) *Gene*, **33**, 103–119.
13 Chang,A.C.Y. and Cohen,S.N. (1978) *J. Bacteriol.*, **134**, 1141–1156.
14 Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edn. Cold Spring Harbor Laboratory press, Cold Spring Harbor, NY.
15 Marmur,J. (1961) *J. Mol. Biol.*, **3**, 208–218.
16 Birnboim,H.C. and Doly,J. (1979) *Nucleic Acids Res.*, **7**, 1513–1523.
17 Marko,M.A., Chipperfield,R. and Birnboim,H.C. (1982) *Anal. Biochem.*, **121**, 382–387.
18 Chen,E.Y. and Seeburg,P.H. (1985) *DNA*, **4**, 165–170.
19 Whitehead,P.R. and Brown,N.L. (1985) *Arch. Microbiol.*, **141**, 70–74.
20 Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) *J. Mol. Biol.*, **215**, 403–410.
21 Corpet,F. (1988) *Nucleic Acids Res.*, **16**, 10881–10890.
22 Henikoff,S. and Henikoff,J.G. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 10915–10919.
23 Timinskas,A., Butkus,V. and Janulaitis,A. (1995) *Gene*, **157**, 3–11.
24 Ausubel,F.M., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl,K. (1992) In Ausubel,F.M (ed.), *Short Protocols In Molecular Biology* (2nd edition). Green Publishing Associates and John Wiley & Sons.
25 Posfai,J., Bhagwat,A.S., Posfai,G. and Roberts,R.J. (1989) *Nucleic Acids Res.*, **17**, 2421–2435.
26 Wilson,G.G. (1988) *Gene*, **74**, 281–289.
27 Balganesh,T.S., Reiners,L., Lauster,R., Noyer-Weidner,M., Wilke,K. and Trautner,T.A. (1987) *EMBO J.*, **6**, 3543–3549.
28 Trautner,T.A., Balganesh,T.S. and Pawlek,B. (1988) *Nucleic Acids Res.*, **16**, 6649–6657.
29 Wilke,K., Rauhut,E., Noyer-Weidner,M., Lauster,R., Pawlek,B., Behrens,B. and Trautner,T.A. (1988) *EMBO J.*, **7**, 2601–2609.
30 Klimasauskas,S., Nelson,J.L. and Roberts,R.J. (1991) *Nucleic Acids Res.*, **19**, 6183–6190.
31 Klimasauskas,S., Kumar,S., Roberts,R.J. and Cheng,X. (1994) *Cell*, **76**, 357–369.
32 Reinisch,K.M., Chen,L., Verdine,G.L. and Lipscomb,W.N. (1995) *Cell*, **82**, 143–153.
33 Noyer-Weidner,M. and Trautner,T.A. (1993) In Jost,J.P. and Saluz,H.P. (eds) *DNA Methylation: Molecular Biology and Biological Significance*. Birkhäuser Verlag, Basel/Switzerland pp. 39–107.
34 Sznyter,L.A., Slatko,B., Moran,L., O'Donnell,K.H. and Brooks,J.E. (1987) *Nucleic Acids Res.*, **15**, 8249–8266.
35 Looney,M.C., Moran,L.S., Jack,W.E., Feehery,G.R., Benner,J.S., Slatko,B.E. and Wilson,G.G. (1989) *Gene*, **80**, 193–208.
36 Bitinaite,J., Maneliene,Z., Menkevicius,S., Klimasauskas,S., Butkus,V. and Janulaitis,A. (1992) *Nucleic Acids Res.*, **20**, 4981–4985.
37 Sugisaki,H., Yamamoto,K. and Takanami,M. (1991) *J. Biol. Chem.*, **266**, 13952–13957.
38 Anderson,J.E. (1993) *Curr. Opin. Struct. Biol.*, **3**, 24–30.
39 Wah,D.A., Hirsch,J.A., Dorner,L.F., Schildkraut,I. and Aggarwal,A.K. (1997) *Nature*, **388**, 97–100.
40 Bennet,S.P. and Halford,S.E. (1987) In Guschlbauet and Saenger,W. (eds), *DNA-ligand Interactions from Drugs to Proteins.* Plenum, New York pp. 239–250.
41 Petrusyte,M.P. and Janulaitis,A.A. (1982) *Eur. J. Biochem.*, **121**, 377–381.
42 Wilson,G.G. and Murray,N.E. (1991) *Annu. Rev. Genet.*, **25**, 585–627.
43 Siksnys,V., Zareckaja,N., Vaisvila,R., Timinskas,A., Stakenas,P., Butkus,V. and Janulaitis,A. (1994) *Gene*, **142**, 1–8.
44 Janulaitis,A. (1996) *Biologija*, **2**, 28–31.
45 Sanders,C. and Schneider,R. (1991) *Proteins*, **9**, 56–68
46 Doolittle,R.F. (1981) *Science*, **214**, 149–159.