Mol Gen Genet (1998) 260: 226–231
© Springer-Verlag 1998

ORIGINAL PAPER

S.-y. Xu · J.-p. Xiao · L. Ettwiller
M. Holden · J. Aliotta · C. L. Poh · M. Dalton
D. P. Robinson · T. R. Petronzio · L. Moran
M. Ganatra · J. Ware · B. Slatko · J. Benner

Cloning and expression of the *Apa*LI, *Nsp*I, *Nsp*HI, *Sac*I, *Sca*I, and *Sap*I restriction-modification systems in *Escherichia coli*

Received: 15 April 1998 / Accepted: 3 August 1998

Abstract The genes encoding the *Apa*LI (5′-G^TGCA-C-3′), *Nsp*I (5′-RCATG^Y-3′), *Nsp*HI (5′-RCATG^Y-3′), *Sac*I (5′-GAGCT^C-3′), *Sap*I (5′-GCTCTTCN1^-3′, 5′-^N4GAAGAGC-3′) and *Sca*I (5′-AGT^ACT-3′) restriction-modification systems have been cloned in *E. coli*. Amino acid sequence comparison of M.*Apa*LI, M.*Nsp*I, M.*Nsp*HI, and M.*Sac*I with known methylases indicated that they contain the ten conserved motifs characteristic of C5 cytosine methylases. *Nsp*I and *Nsp*HI restriction-modification systems are highly homologous in amino acid sequence. The C-termini of the *Nsp*I and *Nla*III (5′-CATG-3′) restriction endonucleases share significant similarity. 5mC modification of the internal C in a *Sac*I site renders it resistant to *Sac*I digestion. External 5mC modification of a *Sac*I site has no effect on *Sac*I digestion. N4mC modification of the second base in the sequence 5′-GCTCTTC-3′ blocks *Sap*I digestion. N4mC modification of the other cytosines in the *Sap*I site does not affect *Sap*I digestion. N4mC modification of *Sca*I site blocks *Sca*I digetion. A DNA invertase homolog was found adjacent to the *Apa*LI restriction-modification system. A DNA transposase subunit homolog was found upstream of the *Sap*I restriction endonuclease gene.

Key words Restriction endonuclease · Methylase selection · Gene expression · DNA methylation · Recombinant DNA

Communicated by W. Arber

S.-y. Xu (✉) · J.-p. Xiao · L. Ettwiller · M. Dalton
D. P. Robinson · T. R. Petronzio · L. Moran · M. Ganatra
J. Ware · B. Slatko · J. Benner
New England Biolabs, Inc., 32 Tozer Road, Beverly
MA 01915, USA
e-mail: xus@neb.com
Tel.: +1-978-927-5054; Fax: +1-978-921-1350

M. Holden
Brown University, Box 1258, Providence
RI 02912, USA

J. Aliotta
Tufts University Medical School
136 Harrison Ave. Boston, MA 02111, USA

C. L. Poh
Dept. of Microbiology, Faculty of Medicine
National University of Singapore
Lower Kent Ridge Rd. Singapore 0511

Introduction

Type II restriction endonucleases are indispensable tools in recombinant DNA technology. They usually recognize 3–8 bp of DNA sequence and cleave within or downstream of the recognition sequences (Heitman 1993; Roberts and Halford 1993; Wilson and Murray 1991). Over 100 restriction-modification (R-M) systems have been cloned so far (Roberts and Macelis 1998). Cloning and expression of restriction-modification genes have made it possible to carry out structural and mechanistic studies of these enzymes (reviewed in Heitman 1993). Comparative studies of cloned R-M systems show that the restriction endonuclease and methylase genes in a particular R-M system map in close proximity to each other. Little sequence homology is found amongst different restriction enzymes, except for some isoschizomers (Wilson and Murray 1991). However, significant sequence similarities have been found among methylases. There are usually ten conserved amino acid sequence motifs in C5 cytosine methylases (Lauster et al. 1989; Posfai et al. 1989). The amino-methylases (N4-cytosine and N6-adenine methylases) fall into three groups, α, β, and γ, based on the circular permutation of nine conserved motifs (Wilson and Murray 1991; Malone et al. 1995).

*Apa*LI, *Nsp*I, *Nsp*HI, *Sac*I, *Sap*I, and *Sca*I were isolated from *Acetobacter pasteurianus* (ATCC 12875), *Nostoc* species (ATCC 29411), *Nostoc* species (ATCC 29106), *Streptomyces achromogenes*, *Saccharopolyspora* species, and *Streptomyces caespitosus*, respectively (Roberts and Macelis 1998). The *Nostoc* and *Saccharopolyspora* strains are difficult to culture on a large

scale and therefore sufficient amounts of the enzymes to meet the needs of molecular biology applications cannot be produced from the native strains. To produce these enzymes in large quantities, it is necessary to clone and overexpress the R-M genes. In this work we report a modified methylase selection method for the cloning of methylase genes in *E. coli* and DNA sequence analysis of these six R-M systems. We also tested the effect of cytosine modifications in *Sac*I, *Sap*I, and *Sca*I recognition sites.

Materials and methods

Media, reagents, and *E. coli* strains

Bacterial culture media were prepared as described (Catalogue of Bacteria and Phages; ATCC, Rockville, MD). Molecular biology reagents were from New England Biolabs. The genotype of *E. coli* K strain ER2206 is *endA1 thi-1 supE44* Δ(*mcr-hsd-mrr*)114::IS10 Δ*lacU169* /F′ *proAB lacI^q lacZ* Δ*M15* Tn*10*. T7 expression hosts were *E. coli* B strain ER2504 (λDE3) [*endA1 fhuA2* Δ (*mcr-hsd-mrr*)] and ER2566 [*endA1 fhuA2* Δ(*mcr-hsd-mrr*), with the T7 RNA polymerase gene inserted in the *lac* operon (E. Raleigh and W. E. Jack, unpublished)], *E. coli* K strain ER1821 (λDE3), e14$^-$ (McrA$^-$) *endA1 supE44 thi-1* Δ (*mcrC-mrr*)) (S.-y. Xu, unpublished), or RR1 (λDE3) (J. Menin and W. E. Jack, unpublished).

Purification of *Apa*LI, *Sac*I, and *Sca*I proteins from native sources

*Apa*LI restriction endonuclease (hereafter referred to as *Apa*LI) was purified to near homogeneity by chromatography on DEAE Sepharose, Heparin Sepharose, Mono Q and Sephacryl S-100 HR (Pharmacia Biotech, Piscataway, N.J.) columns. *Sac*I was purified to near homogeneity by chromatography on DEAE Sepharose, Heparin Sepharose, and Mono-Q columns. *Sca*I was purified by chromatography on Heparin Sepharose, Phosphocellulose P11 (Whatman, Maidstone, UK) and DEAE Sepharose columns. The purified proteins were subjected to electrophoresis and electroblotted as previously described (Matsudaira 1987). The membrane was stained and protein bands were excised and subjected to sequential degradation in an automated sequencer (ABI Model 470 A).

Preparation of genomic DNA and library construction

Genomic DNA was prepared from *Acetobacter pasteurianus* (ATCC 12875), *Nostoc* sp. (ATCC 29411, *Nsp*I-producing strain), *Nostoc* sp. (ATCC 29106, *Nsp*HI-producing strain), *Streptomyces achromogenes* (ATCC 12767), *Streptomyces caespitosus* and *Saccharopolyspora* sp. as previously described (Maniatis et al. 1982). *Apo*I, *Nla*III, and *Sau*3AI partial genomic DNA libraries were constructed as described (Fomenkov et al. 1994).

DNA sequence analysis, PCR and inverse PCR

Plasmid DNA was sequenced by the dideoxy termination method using the AmpliTaq dideoxy terminator sequencing kit and the ABI 373 A sequencing machine. Inverse PCR amplification of unknown DNA was carried out as previously described (Ochman et al. 1988). Inverse PCR products were gel-purified and sequenced directly or cloned in pUC19 and then sequenced.

Expression of R-M genes in *E. coli*

Methylase genes were cloned into the Tc^R gene of pACYC184 (Rose 1988), pSYX20 (Km^R and Tc^R, pSC101 origin) (Morgan et al. 1996) or pBR322 (Sutcliffe 1978) and the cloned genes were constitutively expressed under the control of the Tc promoter. Restriction endonuclease genes were expressed from the *lac* promoter of pRRS (Skoglund et al. 1990) or the T7 promoter of pAII17 (pET11 derived) (Kong et al. 1993), pET21at (Aliotta et al. 1996) or pACYC-T7 (S.-y. Xu, unpublished) in a premodified strain that had been transformed with a plasmid bearing the gene for the respective cognate methylase. Plasmid pSYX20-LysS was constructed and provided by J. Menin and W. E. Jack (New England Biolabs, unpublished). Plasmid pSYX33 is derived from pLG339 (Stoker et al. 1982) and carries a pSC101 origin, Tc^R and Km^R markers, and a *lacI^q* gene inserted in the *Eco*RI site (S.-y. Xu, unpublished).

Identification of modified bases within the recognition sequence

Deoxyoligonucleotides (oligos) were synthesized to test for resistance to digestion by *Sac*I or *Sap*I. The oligos used had the following sequences:

*Sac*I oligo 1: 5′-GCGGA**GAGCTC**TCCG-3′
*Sac*I oligo 2: 5′-GCGGA**GAGCTC**TCCG-3′; **C** = 5mC
*Sac*I oligo 3: 5-GCGGA**GAGCTC**TCCG-3′; **C** = 5mC
*Sap*I oligo 1: 5-CGTA**GCTCTTC**TTAGCT-3′
*Sap*I oligo 2: 5-CGTA**GCTCTTC**TTAGCT-3′; **C** = N4mC
*Sap*I oligo 3: 5-CGTA**GCTCTTC**TTAGCT-3′; **C** = N4mC
*Sap*I oligo 4: 5-CGTA**GCTCTTC**TTAGCT-3′; **C** = N4mC
*Sap*I oligo 5: 5-AGCTAA**GAAGAGC**TACG-3′.

*Sac*I oligos 1, 2, and 3 were labeled at their 5′ ends with [γ-^{33}P]ATP and T4 polynucleotide kinase. These three oligos are self-complementary and form duplexes at 37° C. *Sap*I oligos 1, 2, 3, and 4 were labeled at their 5′ ends with [γ-^{33}P]ATP and T4 polynucleotide kinase. After labeling, they were annealed to the complementary strand (*Sap*I oligo 5) and used for *Sap*I digestion. Restriction digestions were carried out at 37° C for 1 h. The resulting products were resolved in 20% DNA sequencing gels and X-ray films were exposed 24 to 48 h to obtain autoradiograms.

The fluorophore-labeled oligos used to test N4mC modification of *Sca*I site had the following sequences:

*Sca*I oligo 1: 5′-**X**CGATCGCTG**AGTACT**GGCTGAGC-3′
*Sca*I oligo 2: 5′-**X**CGATCGCTG**AGTACT**GGCTGAGCCC-3′;
C = N4mC; **X** = fluorescein
*Sca*I oligo 3: 5′-**Y**GTGGCTCAGCC**AGTACT**CAGCGATCG-CAT-3′
*Sca*I oligo 4: 5′-**Y**GTGGCTCAGCC**AGTACT**CAGCGATCG-CATG 3′;
C = N4mC; **Y** = tetrachlorofluorescein.

The non-methylated (oligos 1+3), hemi-methylated (oligos 1+4; oligos 2+3), and fully methylated (oligos 2+4) *Sca*I oligos were annealed and digested with *Sca*I at 37° C for 2 h in the *Sca*I buffer. The digested oligos were extracted with phenol-chloroform and chloroform, and diluted to 0.05 ng/μl. Aliquots (0.5 ng) of oligos were resolved in a 15% DNA sequencing gel and the cleaved products were detected with the ABI PRISM GeneScan 2.1.

Results and discussion

Cloning of the *Apa*LI R-M system

The *apaLIM* gene (1290 bp) was isolated from a *Sau*3AI partial DNA library constructed in a modified pUC19 vector (containing three *Apa*LI sites) by the methylase selection method (Szomolanyi et al. 1980; Kiss and Baldauf 1983). Most of the *apaLIR* gene (lacking the last

12 codons) was also present in the original M^+ clone. The entire insert of 3923 bp was sequenced. Three ORFs were found: ORF1 encodes a DNA invertase homolog, whose predicted amino acid sequence shows 68% similarity (54% identity) to the known DNA invertase Cin of phage P1 (Hiestand-Nauer and Iida 1983) and 84% similarity (71% identity) to a putative DNA invertase found adjacent to the *Pae*R7I R-M system (Vaisvila et al. 1995). An inverted repeat (5′-GTTTTTN$_{5-6}$GTCCA-GAAAAC-3′) is also present upstream of the putative invertase. We do not know if this invertase is active in the native strain. We did not identify a second copy of the inverted repeat in the insert of 3923 bp. In *Mycoplasma pulmonis*, a type I restriction-modification system (*hsd1* locus) has been found to be regulated by DNA inversion (Dybvig and Yu 1994). ORF2 was identified as the *apaLIM* gene because the predicted amino acid sequence contains conserved C5 methylase motifs and a deletion in ORF2 abolished *Apa*LI site modification (data not shown). ORF2 (*apaLIM* gene) was amplified by PCR and inserted into the *Bam*HI site in pSYX33 to generate plasmid pSYX33-ApaLIM. ORF3 is a partial sequence ending at a *Sau*3AI site. Amplification of the adjacent DNA by inverse PCR indicated that ORF3 continued in frame for an additional 12 codons. ORF3 was amplified by PCR and ligated to *Hin*dIII and *Sph*I sites in pRRS, and the ligated DNA was transformed into an M.*Apa*LI premodified host. A cell extract prepared from IPTG-induced cells displayed *Apa*LI activity. Thus, ORF3 was confirmed as the *apaLIR* gene (1128 bp, Fig. 1). The DNA sequence of the *Apa*LI R-M system has been determined independently (Suzuki et al. 1996). Comparison of our sequence with the published sequence indicates that the published sequence contains a frame shift between codons 284 and 314 in the *apaLIR* gene. There is a deletion (C) at nucleotide 849 and an insertion (T) at position 945 of the *apaLIR* gene. A third discrepancy was found at position 1024, which is a C as compared to a T in the published sequence. Since this part of our sequence was cloned directly from genomic DNA and the published sequence was derived from inverse PCR product, we believe that our sequence is likely to be the correct one.

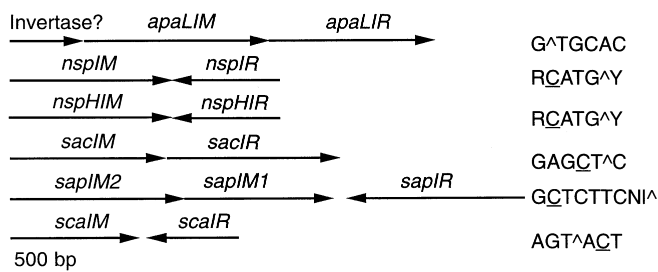

Fig. 1 Genetic organization of *Apa*LI, *Nsp*I, *Nsp*HI, *Sac*I, *Sap*I, and *Sca*I R-M systems. The GenBank accession numbers are: *Apa*LI (AF044847); *Nsp*I (AF056036), *Nsp*HI (AF056037), *Sac*I (AF027867), *Sap*I (AF045021), *Sca*I (AF044681)

In all C5 methylases that recognize sites containing 5′-GC-3′ as part of the recognition sequence, there is a conserved Arg residue within the variable region (target-recognizing domain; TRD) that recognizes the guanine 5′ to the modified cytosine. In the crystal structure of *Hha*I complexed with DNA, this conserved Arg_{240} residue contacts the 5′ G and the Thr_{250} residue makes a water-mediated contact with the phosphate backbone (Klimasauskas et al. 1994). The recognition sequence for M.*Apa*LI also contains a 5′-GC-3′ dinucleotide and the sequence of M.*Apa*LI possesses the conserved motif R-9(Xaa)-T (Fig. 2C). It is likely that M.*Apa*LI modifies the internal C.

Cloning of the *Nsp*I R-M system

The *nspIM* gene (1194 bp) was isolated from an *Apo*I partial DNA library using a modified pRRS vector (containing two *Nsp*I sites). The entire insert of 2212 bp in the original clone was sequenced. ORF1 was identified as the *nspIM* gene, because the predicted amino acid sequence shows homology to C5 methylases. The *nspIM* gene was amplified by PCR and cloned into the *Bam*HI site of pBR322 to generate a premodified host RR1(λDE3) [pBR322-NspIM]. One partial ORF oriented in the opposite direction to the *nspIM* gene was found. Amplification of the adjacent DNA by inverse

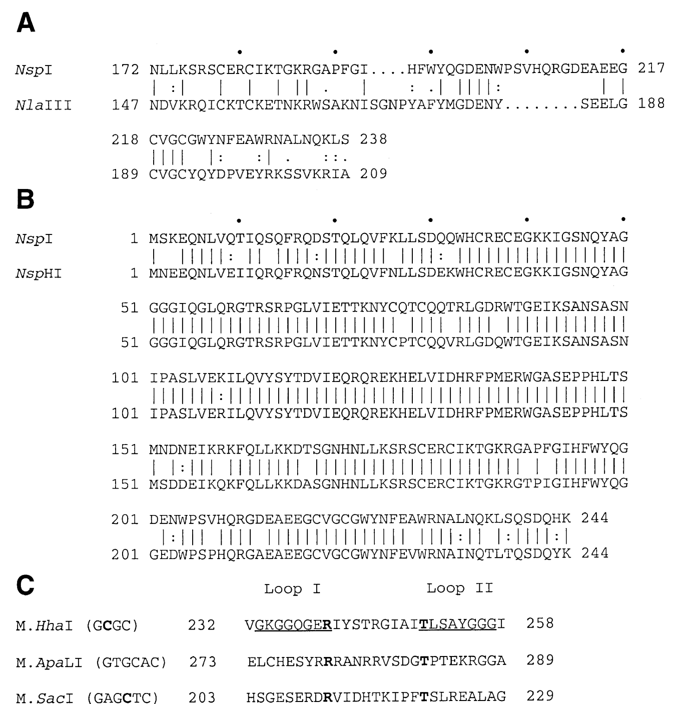

Fig. 2A Amino acid sequence alignment of C-termini of *Nsp*I and *Nla*III generated by the GCG program Bestfit (Devereux et al. 1984). **B** Amino acid sequence alignment of *Nsp*I and *Nsp*HI. The two endonucleases share 91% similarity (89% identity). Identical residues are indicated by *vertical lines* and residues with similar properties are indicated by *colons*. **C** Part of the TRDs of M.*Hha*I, M.*Apa*LI and M.*Sac*I. The conserved Arg and Thr residues are highlighted in *bold*

PCR indicated that the partial ORF continues for an additional five codons. The entire ORF was amplified by PCR and inserted into the *Bam*HI site in pACYC-T7 and the ligated DNA was transformed into an M.*Nsp*I premodified host RR1(λDE3) [pBR322-NspIM]. Cell extracts prepared from IPTG-induced cells showed *Nsp*I activity (data not shown). Thus, the second ORF was identified as the *nspIR* gene (735 bp). The *nspIM* and *nspIR* genes run into each other, overlapping by 38 bp. Both *Nsp*I and *Nla*III generate DNA fragments with a 5′-CATG-3′ extension. The BESTFIT and DOTPLOT programs (Devereux et al. 1984) were used to assess sequence similarity between the two endonucleases. The comparison revealed that the C-terminal 66 residues of *Nsp*I share 44% similarity (35% identity) with *Nla*III (recognition sequence 5′-CATG-3′) (Fig. 2A). It is possible that this is part of the specificity domain that recognizes the CATG sequence. Weak amino acid sequence similarity (28% identity) has also been observed between *Nla*III and *Sph*I (recognition sequence 5′-GCATGC-3′) (Morgan et al. 1996). Another pair of restriction enzymes whose recognition sequences are related comprises *Bss*HII (5′-GCGCGC-3′) and *Asc*I (5′-GGCGCGCC-3′); these two endonucleases are 40% similar (28% identical) in sequence (K. Lunnen, S.-y. Xu and G.G. Wilson, unpublished).

Originally we tried to express the *nspIM* gene in *E. coli* B strain ER2566, but the transformants were unstable, probably due to partial restriction of the methylated DNA. The *E. coli* K strains ER1821 (λDE3) and RR1 (λDE3) were more tolerant of M.*Nsp*I than *E. coli* B.

The native *Nostoc* sp. (ATCC 29411) was isolated from a moderately hot spring and produces five restriction endonucleases, *Nsp*I–*Nsp* V (Reaston et al. 1982). The cloning and expression of the *Nsp*I *R-M* genes in *E. coli* should facilitate production and purification of *Nsp*I. This would also avoid any possibility of contamination by the other four restriction endonucleases produced by the native strain.

Cloning of *Nsp*HI R-M system

The *nspHIM* gene (1194 bp) was isolated from an *Apo*I partial DNA library using a modified pRRS as a cloning vector. The *nspHIM* gene was amplified by PCR and cloned into the *Bam*HI site of pACYC184 to produce plasmid pACYC-NspHIM. The original M^+ isolate carries a partial *nspHIR* gene. After two rounds of inverse PCR amplification and DNA sequencing, the entire *nspHIR* gene (735 bp) was amplified by PCR and cloned into the *Nde*I and *Bam*HI sites of the T7 expression vector pET21at. To reduce the basal level of expression, a third compatible plasmid pSX20-LysS was also transformed into the host cell. The *Nsp*HI expression strain was ER1821 (λDE3) [pACYC-NspHIM, pSX20-LysS, pET21at-NspHIR].

The *Nsp*I and *Nsp*HI R-M systems are very similar. The two systems are 88% identical at the DNA level. *Nsp*I and *Nsp*HI endonucleases are 91% similar (89% identical) in amino acid sequence (Fig. 2B). M.*Nsp*I and M.*Nsp*HI are 92% similar (90% identical) in sequence. The 148 bp of sequence upstream of the two methylase genes are also 79% identical. The gene organization of the *Nsp*I and *Nsp*HI R-M systems is also the same (Fig. 1) and it is likely that the two R-M systems also share the same gene regulation mechanism.

Cloning of the *Sac*I R-M system

The M.*Sac*I gene (*sacIM*, 1173 bp) was isolated from a *Sau*3AI partial DNA library using the expression vector pRRS. The vector contains a stem-loop structure at the end of the modified pUC19 polylinker which increases gene expression (Skoglund et al. 1990). For instance, 12-fold more *Hin*fI activity is expressed from pRRS than from a pUC19 vector containing a lac_{UV5} promoter (Skoglund et al. 1990). The *sacIM* gene cloned in pRRS ended within a *Sau*3AI site and lacked only the stop codon. Thus, M.*Sac*I was produced as a fusion protein by addition of a Leu residue from the polylinker (a stop codon follows the Leu codon in the polylinker). The fusion protein is apparently active in *E. coli*, because it was selected by virtue of expression of the methylase and modification of *Sac*I sites in vivo. Plasmid pRRS-SacIM is partially resistant to *Sac*I digestion. Approximately 50% of the supercoiled DNA was converted to nicked circular DNA by *Sac*I digestion (data not shown). This result probably reflects the poor ribosome binding site (5′-TTAGGG-3′) and GC-rich spacer (5′-TGGCGCG-3′) upstream of the start codon (5′-GTG-3′). When pUC19 was used as the vector to clone the *sacIM* gene, no M^+ clones were found in either *Nla*III, *Sau*3AI, or *Sma*I partial DNA libraries after the methylase selection procedure (S.-y. Xu, J.-p. Xiao, unpublished). To ensure sufficient methylase gene expression in *E. coli*, it is desirable to use a high-copy-number vector such as pRRS for the construction of genomic DNA libraries.

By modifying the ribosome binding site and spacer to 5′-GGAGGTAAATAA-3′, the *sacIM* gene was expressed in pLG339 under the control of a Tc promoter in *E. coli* strain ER2206. The *sacIR* gene (1077 bp) was amplified by PCR and cloned into the *Bam*HI and *Hin*dIII sites of pRRS and transformed into M.*Sac*I premodified cells. *Sac*I activity was detected in cell extracts made from IPTG induced cells. The *sacIM* and *sacIR* genes are separated by 17 bp and transcribed in the same orientation. The gene organization of the *Sac*I R-M system is shown in Fig. 1.

The recognition sequence for M.*Sac*I also contains a 5′-GC-3′ dinucleotide and M.*Sac*I possesses the conserved motif R-9(Xaa)-T (Fig. 2C). Based on this conserved motif, it was predicted that M.*Sac*I modifies the internal C. To test this prediction, deoxyoligonucleotides were synthesized with 5 mC within the recognition sequence. When the internal C is modified (*Sac*I oligo 2, see Materials and methods), the oligo is resistant to *Sac*I

digestion. Oligos with an unmodified *Sac*I site or in which the external C was methylated were cleaved by *Sac*I. The efficiency of cleavage was approximately 50%, probably due to the short length of the oligos (data not shown).

Cloning of the *Sap*I R-M system

The *Sap*I M1 gene (*sapIM1*, 1131 bp) was isolated from a *Sau*3AI partial DNA library and the *Sap*I M2 gene (*sapIM2*, 1302 bp) was isolated from a *Kpn*I DNA library by the methylase selection method. There are two *Sap*I methylases in the *Sap*I R-M system (Fig. 1), reflecting the asymmetry of the recognition sequence. Their predicted sequences indicate that both are amino-methyltransferases. The *sapIM1* and *sapIM2* genes overlap by 1 bp. To premodify the *E. coli* chromosome, *sapIM1* and *sapIM2* genes were amplified separately, cloned into the *Bam*HI and *Sph*I sites of pBR322 and transformed into the T7 expression host ER2566. The *sapIM1* and *sapIR* genes are convergent and separated by 106 bp. The *sapIR* gene (1299 bp) was amplified by PCR and cloned into the *Bam*HI site of the T7 expression vector pACYC-T7. The final expression strain was ER2566 [pBR322-SapIM1 & SapIM2, pACYC-T7-SapIR]. There is an ORF adjacent to the *sapIR* gene that has 66% similarity (55% identity) to a transposase subunit of IS*1372* found in *Streptomyces lividans* (Fischer et al. 1996). There are other cases of restriction enzyme or methylase genes associated with mobile genetic elements. For example, the *Acc*I R-M system was found adjacent to a transposon-like sequence (Brassard et al. 1995) and a Dam methylase homolog is associated with a retron (Hsu et al. 1990). The *Eco*T22I R-M system (*Nsi*I isoschizomer) was found associated with a Tn*3*-like transposon, Tn*5396*, in a plasmid (Elhai et al. 1994). It is possible that association of R-M genes with mobile genetic elements facilitates transfer of R-M genes between microorganisms.

To test the effects of cytosine modification of the *Sap*I sequence, deoxyoligonucleotides were synthesized that contain N4mC bases in the second, fourth, or seventh position within the recognition sequence (5′-GCTCTTC-3′). *Sap*I oligo 2 contains N4mC in the second position. When *Sap*I oligo 2 was annealed to oligo 5, the duplex was resistant to *Sap*I digestion. When *Sap*I oligos 3 and 4 containing N4mC modification at other positions were annealed to oligo 5, they were cleaved by *Sap*I and the cleavage efficiency was about 70% (data not shown). Our preliminary results indicate that N4mC modification of the other strand (5′-GAAGAGC-3′) does not block *Sap*I digestion (S.-y. Xu. unpublished). The effect of N6mA modification of the 5′-GAAGAGC-3′ sequence has yet to be determined. *Sap*IM1 and *Sap*IM2 contain SPPF and SPPY motifs, respectively. *Sap*IM2 is most probably an N4 cytosine methylase since the SPPY motif is found in all N4 cytosine methylases discovered so far (G. G. Wilson, personal communication).

Cloning of the *Sca*I R-M system

The *scaIM* gene (915 bp) was isolated from a *Sau*3AI partial DNA library constructed in the expression vector pRRS using the methylase selection procedure. No M^+ clones were found in DNA libraries using a pUC19 vector (S.-y. Xu and J.-p. Xiao, unpublished). The original M^+ clone also contains the majority of the *scaIR* gene and only lacks the first three codons. The *scaIM* gene was amplified by PCR and inserted into the BamHI site of pACYC184. The entire *scaIR* gene (684 bp) was amplified by PCR, ligated to the *Sph*I site of pRRS and transformed into M.*Sca*I premodified cells. *Sca*I activity was detected in cell extracts of an IPTG-induced culture. The *scaIR* and *scaIM* genes are 45 bp apart, and convergently oriented. The gene organization of the *Sca*I R-M system is shown in Fig. 1. M.*Sca*I contains the amino-methyltransferase motif SPPY and is most likely to be an N4 cytosine methylase.

The recognition sequence of M.*Sca*I is 5′-AGTACT-3′, in which base 5 is the only candidate for N4mC modification. To test the effects of modification of the *Sca*I site, fluorophore-labeled oligos were synthesized that contained N4mC (see Materials and methods for oligo sequences). Hemi-methylated (*Sca*I oligos 1+4; oligos 2+3) or fully methylated duplexes (*Sca*I oligos 2+4) were resistant to *Sca*I digestion; oligos without the N4mC modification (*Sca*I oligos 1+3) were cleaved by *Sca*I (data not shown). This experiment demonstrates that N4mC modification at the *Sca*I site blocks *Sca*I digestion in vitro.

Acknowledgements We thank Ira Schildkraut and Richard Roberts for critical comments, Linda Coe for technical assistance, Julia Menin for providing pSYX20-LysS, Elisabeth Raleigh for *E. coli* strains, Roger Knott for primer synthesis, Jay Wayne and Pei-chung Hsieh for discussion, and Donald Comb for support.

References

Aliotta JM, Pelletier JJ, Ware JL, Moran LS, Benner JS, Kong HM (1996) Thermostable *Bst* DNA polymerase I lacks a 3′ to 5′ proofreading exonuclease activity. Genet Anal Biomol Engin 12:185–195

Brassard S, Paquet H, Roy PH (1995) A transposon-like sequence adjacent to the *Acc*I restriction-modification operon. Gene 157:69–72

Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12:387–395

Dybvig K, Yu H (1994) Regulation of a restriciton and modification system via DNA inversion in *Mycoplasma pulmonis*. Mol Microbiol 12:547–560

Elhai J, Cai Yp, Wolk CP (1994) Conduction of pEC22, a plasmid coding for MR.*Eco*T22I, mediated by a resident Tn3-like transposon, Tn5396. J Bacteriol 176:5059–5067

Fischer J, Maier H, Viell P, Altenbuchner J (1996) The use of an improved transposon mutagenesis system for DNA sequencing leads to the characterization of a new insertion sequence of *Streptomyces lividans* 66. Gene 180:81–89

Fomenkov A, Xiao J-p, Dila D, Raleigh E, Xu S-y (1994) The ′endo-blue method′ for direct cloning of restriction endonuclease genes in *E. coli*. Nucleic Acids Res 22:2399–2403

Heitman J (1993) On the origins, structures and functions of restriction-modification enzymes. Plenum Press, New York

Hiestand-Nauer R, Iida S (1983) Sequence of the site-specific recombinase gene cin and of its substrates serving in the inversion of the C segment of bacteriophage P1. EMBL J 2:1733–1740

Hsu MY, Inouye M, Inouye S (1990) Retron for the 67-base multicopy single-stranded DNA from *Escherichia coli*: a potential transposable element encoding both reverse transcriptase and Dam methylase functions. Proc Natl Acad Sci USA 87:9454–9458

Kiss A, Baldauf F (1983) Molecular cloning and expression in *Escherichia coli* of two modification methylase genes of *Bacillus subtilis*. Gene 21:111–119

Klimasauskas S, Kumar S, Roberts RJ, Cheng X (1994) *Hha*I methyltransferase flips its targe base out of the DNA helix. Cell 76:357–369

Kong H, Kucera RB, Jack WE (1993) Characterization of a DNA polymerase from the hyperthermophile archaea *Thermococcus litoralis*. J Biol Chem 268:1965–1975

Lauster R, Trautner TA, Noyer-Weidner M (1989) Cytosine-specific type II DNA methyltransferases: a conserved enzyme core with variable target-recognizing domains. J Mol Biol 206:305–312

Malone T, Blumenthal RM, Cheng X (1995) Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechnism for these enzymes. J Mol Biol 253:618–632

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Matsudaira P (1987) Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. J Biol Chem 262:10035–10038

Morgan RD, Camp RR, Wilson GG, Xu S-y (1996) Molecular cloning and expression of *Nla*III restriction-modification system in *E. coli*. Gene 183:215–218

Ochman H, Gerber AS, Hartl DL (1988) Genetic applications of an inverse polymerase chain reaction. Genetics 120:621–623

Posfai J, Bhagwat AS, Posfai G, Roberts RJ (1989) Predictive motifs derived from cytosine methyltransferases. Nucleic Acids Res 17:2421–2435

Reaston J, Duyvesteyn MGC, Waard Ad (1982) *Nostoc* PCC7524, a cyanobacterium which contains five sequence-specific deoxyribonucleases. Gene 20:103–110

Roberts RJ, Halford SE (1993) Type II restriction endonucleases (2nd edn). Cold Spring Harbor Press, Cold Spring Harbor, New York

Roberts RJ, Macelis D (1998) REBASE-restriction enzymes and methylases. Nucleic Acids Res 26:338–350

Rose RE (1988) The nucleotide sequence of pACYC184. Nucleic Acids Res 16:355

Skoglund CM, Smith HO, Chandrasegeran S (1990) Construction of an efficient overproducer clone of *Hin*fI restriction endonuclease using the polymerase chain reaction. Gene 88:1–5

Stoker NG, Fairweather NF, Spratt BG (1982) Versatile low-copy-number plasmid vectors for cloning in *Escherichia coli*. Gene 18:335–341

Sutcliffe JG (1978) Nucleotide sequence of the ampicillin resistance gene of *Escherichia coli* plasmid pBR322. Proc Natl Acad Sci USA 78:3737–3741

Suzuki T, Sugimoto E, Tahara Y, Yamada Y (1996) Cloning and nucleotide sequence of *Apa*LI restriction-modification system from *Acetobacter pasteurianus* IFO 13753. Biosci Biotech Biochem 60:1401–1405

Szomolanyi E, Kiss A, Venetianer P (1980) Cloning the modification methylase gene of *Bacillus sphaericus* R in *Escherichia coli*. Gene 10:219–225

Vaisvila R, Vilkaitis G, Janulaitis A (1995) Identification of a gene encoding a DNA invertase-like enzyme adjacent to the *Pae*R7I restriction-modification system. Gene 157:81–84

Wilson GG, Murray NE (1991) Restriction and modification systems. Annu. Rev. Genet. 25:585–627