Volume 17 Number 23 1989 Nucleic Acids Research

Cloning, characterization and heterologous expression of the *Sma*I restriction-modification system

Stefan Heidmann, Wolfgang Seifert, Christoph Kessler[1] and Horst Domdey

Laboratorium für Molekulare Biologie, Genzentrum, der Ludwig-Maximilians-Universität München, Am Klopferspitz 18a, D-8033 Martinsried and [1] Boehringer Mannheim GmbH, Biochemical Research Center, Department of Genetics and Molecular Biology, Nonnenwald 2, D-8122 Penzberg, FRG

Received August 8, 1989; Revised and Accepted October 27, 1989 EMBL accession no. X16458

ABSTRACT

The genes coding for the class-II *Serratia marcescens* restriction-modification system have been cloned and expressed in *E. coli*. Recombinant clones restricted incoming phage only poorly; the recombinant plasmids, however, became fully modified *in vivo*, i. e. completely resistant against digestion with R·*Sma*I. The determined nucleotide sequence of the cloned system revealed three open reading frames with lengths of 252 bp, 741 bp, and 876 bp. Through various deletion experiments and an insertion-mutation experiment the 876 bp open reading frame could be assigned to the *Sma*I DNA modification enzyme and the 741 bp open reading frame to the *Sma*I restriction endonuclease. Mapping of the transcription start sites of the genes revealed that the *Sma*I endonuclease is transcribed as a polycistronic mRNA together with a 252 bp long preceding open reading frame of unknown function. No homology was found when comparing the amino acid sequence of M·*Sma*I with the published sequences of m^5C-specific DNA modification methyltransferases. On the other hand, a stretch of 14 amino acids in the C-proximal region of M·*Sma*I shows a significant homology to the C-proximal amino acid sequences of the N^6A-methyltransferases M·*Hin*fI and M·*Dpn*IIA and the N^4C-methyltransferase M·*Pvu*II.

INTRODUCTION

Class-II restriction-modification systems are a wide-spread phenomenon in procaryotic organisms (1−3). Several possible functions have been correlated with these systems. The mainly discussed function is the observed protection of the microorganisms against infection by bacteriophages (4−6). This phenomenon is consistent with the observation that the G/C content of a recognition site for a specific system correlates well with the G/C content of the host genome and thus also of the bacteriophage of the bacterial host (7). Class-II restriction endonucleases which recognize GC-rich hexa- or octanucleotide recognition sites are almost exclusively found in organisms with a high G/C content. On the other hand there are also exceptions from this rule, e. g. *Sphaerotilus* species with a G/C content of 66 % and an *Ssp*I recognition site consisting only of A and T nucleotides (AATATT) (7). Exceptions like this and the existence of bacteriophage-harbouring bacteria with no detectable restriction-modification systems imply, however, that other functions than solely protection from phage infection are conceivable, too (2). The *E. coli* M·*Eco*dam I DNA modification methyltransferase, e. g., for which a cognate endonuclease has not yet been found seems to be involved in the repair of mismatches, during gene expression, replication, transposition and chromosome segregation (8).

In addition to class-I and class-III systems more than 1 000 class-II restriction-modification systems have been described until to-date, and more than 100 have been cloned (9−12). The restriction-modification system of *Serratia marcescens* is of special interest, since the

*Sma*I endonuclease is the only known class-II restriction endonuclease which in addition to magnesium ions requires potassium ions as cofactor (13). Furthermore R·*Sma*I is used together with the isoschizomer R·*Xma*I to analyze the methylation pattern of the CpG dinucleotide in the hexanucleotide CCCGGG of the DNA of higher eucaryotes (14, 15). In the present study we report the molecular cloning and characterization of the genes encoding the *Sma*I DNA modification enzyme (M·*Sma*I) and restriction endonuclease (R·*Sma*I).

MATERIALS AND METHODS

Bacterial strains and culture conditions

Serratia marcescens BMTU 1373 [*amp*^{R}] (16) was used as source for isolating genomic *Serratia marcescens* DNA. An *mcr*B$^-$ derivative of *E. coli* JM107 [*end*A1, *gyr*A96, *thi*, *hsd*R17, *sup*E44, *rel*A1, *mcr*A, λ$^-$, Δ(*lac-pro*AB), [F′, *tra*D36, *pro*AB, *lac*I^q ZΔM15]] (17) and *E. coli* K802 [F$^-$, *lac*Y1, *gal*K2, *gal*T22, *met*B1, *hsd*R2, *sup*E44, 1$^-$, *rfb*D1, *mcr*AB1] (18) were used for cloning and expression. Competent *E. coli* cells were produced by the procedure of Hanahan (19). *S. marcescens* was grown at 30°C and *E. coli* at 37°C.

Enzymes and chemicals

Restriction endonucleases were obtained from Boehringer Mannheim, *Eag*I from New England Biolabs, the Klenow enzyme (large fragment of *E. coli* polymerase I), exonuclease III, T4-DNA ligase, AMV reverse transcriptase, T4 polynucleotide-kinase and the *Sma*I-linker were from Boehringer Mannheim. Exonuclease VII was obtained from BRL. [γ-^{32}P]-ATP (185 TBec/mmol) and [α-^{35}S]-dATP (24 TBec/mmol) were from Amersham. The following oligonucleotides were synthesized with an automated DNA synthesizer from Applied Biosystems:

5′-GGTTCTGTTGAAAGAAAAG-3′ (O1)
5′-GTTCCATATGATATCAGTCC-3′ (O2)
5′-CTTTCCCCAAAGTGTAAAG-3′ (O3)
5′-GGATTTCTTTAATATCCATG-3′ (O4)
5′-CTGTTTGATCGAATAAGTC-3′ (O5)

Construction of a plasmid vector (pWS2) with three SmaI sites

To construct pWS2 (Fig. 1), two additional *Sma*I sites were introduced into pKK223 (20) which contains only one *Sma*I site in its polylinker region. One of them was obtained by digesting pKK223 with *Hin*dIII, filling up the protruding ends with Klenow enzyme and religation in the presence of *Sma*I linkers. The third *Sma*I recognition site was created by replacing the 1129 bp pKK223[*Eag*I/*Pvu*II]-fragment comprising parts of the *rop*-gene and of the non-functional *tet*-gene through a *Sma*I linker. Finally the remaining part of the *tet*-gene was deleted through a digestion with *Nae*I and *Nar*I, filling up the protruding ends and religation of the longest fragment.

Construction of the Serratia marcescens genomic library

Genomic DNA of *S. marcescens* was isolated according to Rodriguez and Tait (21). The DNA was partially digested with *Sau*3AI and fractionated by sucrose gradient centrifugation; fragments of 6 to 9 kb were isolated. The *Sau*3AI-fragments were ligated to *Bam*HI-digested and dephosphorylated pWS2 and *E. coli* JM107 was transformed with the resulting plasmids.

Selection of clones expressing the SmaI restriction-modification system

Plasmid DNA was isolated from 2×10^4 transformants representing around 1.1×10^4 insert-containing clones and digested extensively with *Sma*I. By using 125 ng of the digested plasmid DNA for retransformation of *E.coli* JM107, 70 transformants were obtained.

Recombinant clones were analyzed for *in vivo* restriction of the phages T1 and λ_{cts}. Isolated plasmid DNA of recombinant clones was characterized by *in vitro* restriction digestion with *Hin*fI and *Pst*I/*Hin*dIII.

Plasmid manipulations

Subcloning of the *S. marcescens* DNA inserts via partial digestions and restriction site mapping was performed according to standard procedures (22); digestions with exonucleases III and VII were according to Yanisch-Perron et al. (17).

DNA sequencing

DNA was sequenced by the chain termination method of Sanger et al. (23) as modified for supercoiled plasmids (24). Inserts were shortened from both ends by the combined action of exonucleases III and VII (17) and sequenced with standard M13 primers. Three specific oligodeoxynucleotide primers (O1, O2, and O4) were constructed in order to verify ambiguous sequences.

Primer extension analysis

RNA was extracted from *S. marcescens* and from transformed *E. coli* by the hot phenol method (25). The 5′ ends of the modification enzyme and endonuclease mRNAs were mapped by primer extension with AMV reverse transcriptase (26). 25 μg of total bacterial RNA were used. cDNA synthesis of the endonuclease transcript was primed with the 5′ labeled oligodeoxynucleotide O3 which is complementary to the DNA sequence between bases 379 and 397 (see Fig. 3) close to the 5′ end of the coding region. The 5′ end of the M·*Sma*I transcript was mapped by primer extension with oligodeoxynucleotide O5 which corresponds to the DNA sequence between bases 1924 and 1942 (see Fig. 3) close to the 5′ end of the coding region of *sma*IM. The RNA was hydrolyzed by alkali treatment and, after neutralization and ethanol precipitation, the cDNAs were analyzed on a sequencing gel together with Sanger dideoxy sequencing reactions on plasmid DNA using the same primers.

In vitro assay of the SmaI restriction endonuclease

Cells were disrupted by sonication for 3×1 min in 20 mM KCl, 6 mM Tris-HCl (pH 8.0/25°C), 6 mM $MgCl_2$, 6 mM 2-mercaptoethanol and 0.1 mg/ml BSA. The samples were allowed to cool in between for one minute on ice. Intact bacteria and cell debris were removed by centrifugation for 30 min with 10 000 × g. The supernatant was loaded on a phosphocellulose column (Whatman P11) and the proteins were eluted with a solution containing 1.2 M KCl, 50 mM Tris-HCl (pH 8.0/25°C), 0.1 mM EDTA, 10% glycerol and 7 mM 2-mercaptoethanol. Eluted proteins were dialysed against a buffer containing 50 mM KCl , 20 mM Tris-HCl (pH 8.0/25°C), 0.1 mM EDTA, 55% glycerol, 0.01% Triton X-100 and 7 mM 2-mercaptoethanol. *Sma*I endonuclease activity was detected by incubating either the centrifugation supernatants or the column eluates with adenovirus 2 DNA in 50 μl reactions containing 20 mM KCl, 6 mM Tris-HCl (pH 8.0/25°C), 6 mM $MgCl_2$,
6 mM 2-mercaptoethanol and 0.1 mg/ml BSA, followed by electrophoresis in a 1% agarose gel.

SDS polyacrylamide gel electrophoresis

Phosphocellulose column eluted proteins were separated by electrophoresis on 15 % polyacrylamide gels (27). After electrophoresis, proteins were stained with the Amersham silver stain kit according to the instructions of the supplier.

Computer analysis

Analyses of DNA and protein sequences were carried out on a Micro VAX computer (Digital Equipment Corp.) with the SEQED, MAP, COMPARE, and GAP (28) programs

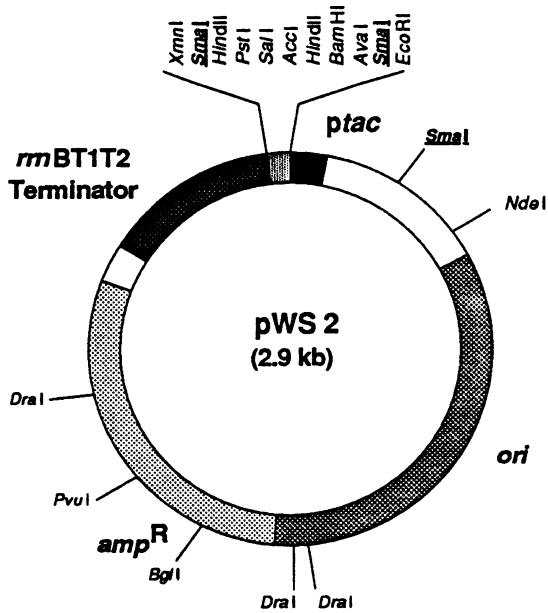


Figure 1: Structure of pWS2. pWS2 is a derivative of pKK223 in which the *tet*- and *rop*-genes have been deleted and two additional *Sma*I recognition sites (indicated by bold letters) have been introduced (for details of construction see *materials and methods*).

of the University of Wisconsin Genetics Computer Group (UWGCG) program package version 4 and the FastP (29) program of the Protein Identification Resource (PIR) (30) program package.

RESULTS

Isolation of the genes coding for the restriction-modification system of S. marcescens

The strategy which was used to clone the genes coding for the restriction-modification system of *S. marcescens* was based on the selection for self-modifying recombinant plasmids which should become resistant against digestion with the cognate restriction endonuclease *Sma*I. Since in most presently known cases the genes coding for the modifying and the restricting activity are located next to each other (1), a clone containing the methyltransferase gene should contain the gene for the endonuclease, too.

Two conditions had to be considered for the construction of the cloning vector: (i) The vector should include more than a single recognition site for the respective endonuclease (R·*Sma*I) in order to minimize the number of background transformants due to uncleaved plasmids which do not contain the M·*Sma*I gene. (ii) In case that the promoter of the methyltransferase was not active in *E. coli*, expression of the methyltransferase should be possible through an *E. coli* promoter present in the cloning vector. According to these conditions pWS2 (Fig. 1) containing three *Sma*I sites and the inducible *tac*-promoter was constructed. 6 to 9 kb fragments of *S. marcescens* DNA obtained through a partial digestion with *Sau*3AI were cloned into the unique *Bam*HI site of pWS2. Plasmid DNA was isolated from around 20 000 recombinant clones and digested extensively with *Sma*I.

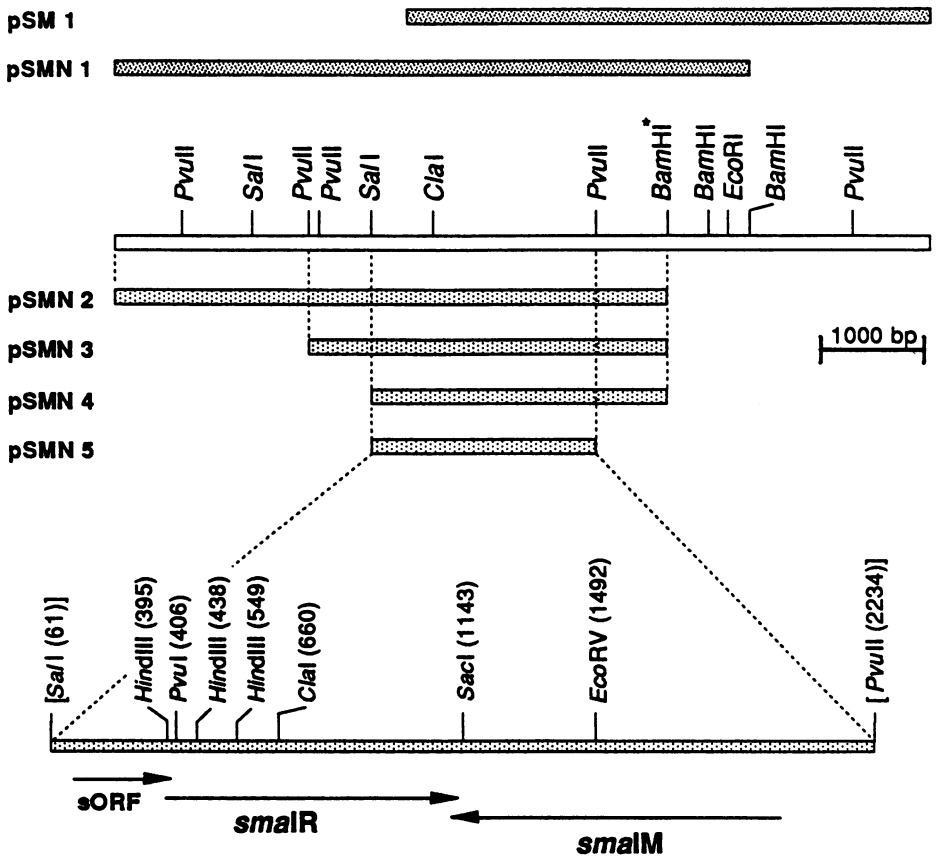
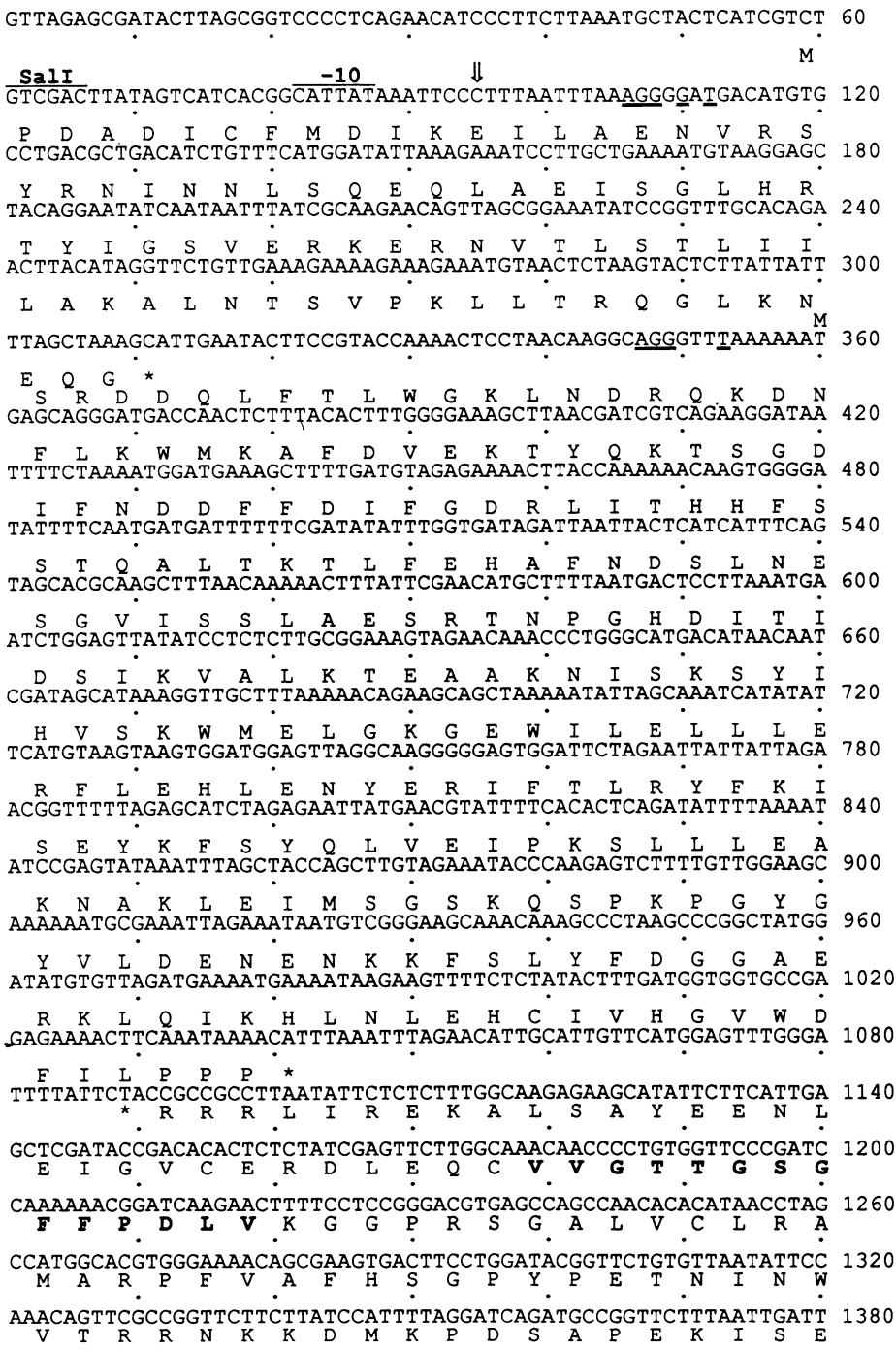


Figure 2: Subclones and restriction map of the gene locus containing the *Sma*I restriction-modification system. The two overlapping inserts of pSM1 and pSMN1 are shown at the top of the figure. pSMN2, pSMN3, pSMN4, and pSMN5 are defined subclones of pSMN1 in the vector pWS2. The arrows indicate the orientation of the three identified reading frames. sORF designates a small open reading frame upstream of the reading frame encoding the endonuclease. The numbers on top of the restriction sites correlate to the nucleotide positions of the complete DNA-sequence shown in Figure 3. The small asterisk indicates a star activity restriction site of *Bam*HI. The two restriction sites shown in brackets at the ends of the enlarged insert of pSMN5 have been destroyed after the processes of filling up protruding ends generated by partial restriction digestions and following religation.

Retransformation with 125 ng of *Sma*I treated plasmid DNA yielded 70 transformants. 42 transformants were tested for restriction of phage. 11 transformants were sensitive, 31 transformants showed partial resistance (efficiency of plating of 2×10^{-2}) against incoming phage. A complete digest with *Hin*fI displayed a similar restriction fragment pattern in the 31 partially resistant clones and in three of the 11 sensitive clones. The remaining eight sensitive clones showed a pattern characteristic for the cloning vector pWS2 (data not shown).

Two out of the 42 partially characterized plasmids were analyzed in further detail. pSM1 was isolated from a phage-sensitive transformant but exhibited a *Hin*fI pattern similar to


SalI
-10
M
M

```
CCCAATCATAGTAATATTTTTTCCCTTTAGATAAGAGAAATATATATTCATGTGATCTAG 1440
  W  D  Y  Y  Y  K  K  G  K  S  L  L  F  I  Y  E  H  S  R  T
TTGGTCTATCTCTGACGCTTTCGGGTTGACAGTTCGGTTTGTTCCATATGATATCAGTCC 1500
  P  R  D  R  V  S  E  P  Q  C  N  P  K  N  W  I  I  D  T  R
TTAAATACCAACCATCATTTTGAAGTGCGAATGCCAGCCGCCATGGAACACCGATCAAAT 1560
  L  Y  W  G  D  N  Q  L  A  F  A  L  R  W  P  V  G  I  L  D
CTTTTGGTTTTAAACCTTCCGGCGTAGGCGGTCTATAGCTCATTGCTCGCCCTTTGTTTT 1620
  K  P  K  L  G  E  P  T  P  P  R  Y  S  M  A  R  G  K  N  K
TATCATCTTTATCACGCCAAGTTCTTCCTCCAGATGTATAGCTGTCACCTATATTAAGCC 1680
  D  D  K  D  R  W  T  R  G  G  S  T  Y  S  D  G  I  N  L  W
ATAGCGTTCCATCATCTTTAAGAGTTCTTCTTACATCTCTAAATAGATCAACTAAATCTT 1740
  L  T  G  D  D  K  L  T  R  R  V  D  R  F  L  D  V  L  D  K
TTATGTAATCATTTATGTTATCTTCAGCGCCTATTTGACCGCCATTACCGTAATCTCTTA 1800
  I  Y  D  N  I  N  D  E  A  G  I  Q  G  G  N  G  Y  D  R  L
ACCCCCAATAAGGTGGGGATGTAACAACACAGTCAAATATTTCAGAATCTAACCCCTGAA 1860
  G  W  Y  P  P  S  T  V  V  C  D  F  I  E  S  D  L  G  Q  V
CTGCTTCCCTAGCATCGCCTACAATAATTTTGCAAGAACGCAGGTTGCTCTCCAAACACT 1920
  A  E  R  A  D  G  V  I  I  K  C  S  R  L  N  S  E  L  C  E
CTTCTGTTTGATCGAATAAGTCTAAATTGTTACTGTGTTTTTTCATAGAACTCAAAGAAA 1980
  E  T  Q  D  F  L  D  L  N  N  S  H  K  K  M
          ⇓        -10
CAAAGAAATAATGCTACTTATGGTATTCATTTGAGAGCGTAAAAACAACATTTTTAGACT 2040
CAATTCCTTTACGTTGTTCAGTTAGCTCAACAGCCGCATTTTCGTCCGTCGAAGAAGCTA 2100
TAGTGTTGTTAGCATCTGCTTTTTGTAGCCGGTGGTCTGCCAGCACTCAGAATCTCCCTG 2160
TATGCAACACCCGCCTAGCTCACCCGGTAGGTATCTCCACACTTCCGGCAGCTACTTTTA 2220
CGCAGCTCGGCCAGCTGTTTATGCAACAGCTTGTTGTCTTGTCGTATGAGCCCGATCAGG 2280
          PvuII
```

Figure 3: DNA sequence of the restriction-modification system of *S. marcescens*. The predicted amino acid sequences of R·*Sma*I (nt. 359−1099) and the preceding sORF (nt. 118−369) are shown above and of M·*Sma*I (nt. 1966−1091) below the nucleic acid sequence. The putative *Pribnow*-boxes in the −10 regions are indicated. Arrows define mapped transcription start sites. Potential *Shine-Dalgarno* sequences are underlined. The peptide sequence which is highly homologous to sequences in M·*Dpn*IIA, M·*Pvu*II, and M·*Hin*fI is written in bold letters.

those plasmids which were able to confer partial resistance to phage infection. pSMN1 was isolated from a phage-resistant transformant. Both plasmids with insert sizes of 5.1 kb for pSM1 and 6.2 kb for pSMN1 (Fig. 2) were completely resistant against cleavage with *Sma*I. From both transformants extracts were prepared and tested for R·*Sma*I activity. Only the extract obtained from the pSMN1 clone exhibited endonuclease activity (data not shown).

DNA sequence analysis

By stepwise reductions of the insert in pSMN1 the complete restriction-modification system could be narrowed down to a pSMN1[2173 bp *Sal*I/*Pvu*II]-fragment cloned in pWS2 (Fig. 2). The nucleotide sequence of the entire insert was determined and revealed three open reading frames of significant lengths (Fig. 3). These three reading frames could be translated into three polypeptide sequences of 84, 247 and 292 amino acids, i.e. polypeptides of calculated molecular weights of 9442, 28749 and 33428 daltons, respectively.

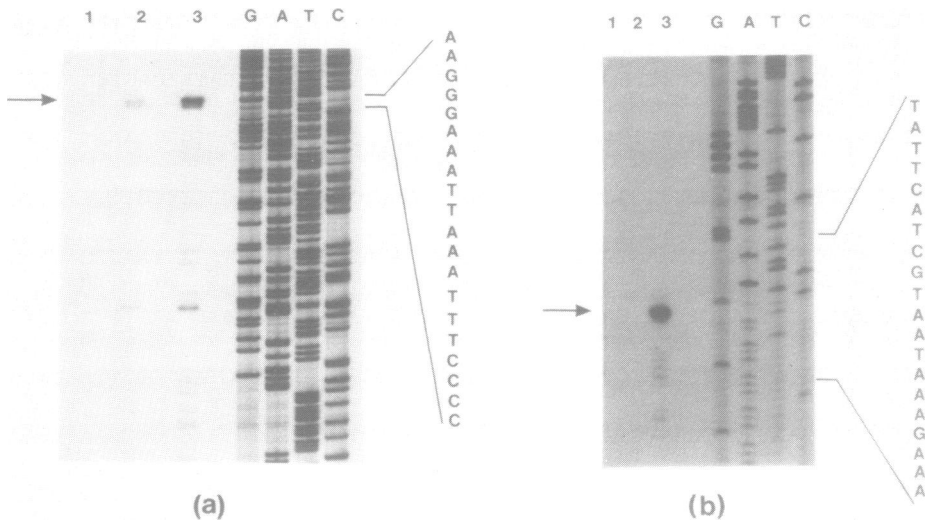


Figure 4: Primer extension analyses of the R·*Sma*I (a) and M·*Sma*I (b) transcripts using the oligodeoxynucleotides O3 and O5, respectively, as primers. The analyzed total RNA was isolated from *E. coli* JM107 containing pWS2 (1), *E. coli* JM107 containing pSMN3 (2), and *S. marcescens* (3). Arrows indicate the transcription start sites. The adjacent DNA sequence analysis was carried out with the same primers and pSMN3 as DNA template.

Assignment of two of the three reading frames to the DNA modification enzyme and the restriction endonuclease

Since extracts from clone pSM1 which lacked the 45 5′-terminal nucleotides of the 741 bp reading frame and the entire 252 bp reading frame did not show any endonuclease activity, the 876 bp reading frame could easily be assigned to the modification gene. Therefore the endonuclease had to be encoded by any of the two other reading frames. An insertion of 2 bp was introduced in pSMN4 at the singular *Cla*I site (position 660, see Fig. 2) in the 741 bp reading frame to yield plasmid pSMN4I. This mutation caused an R·*Sma*I-negative phenotype. A deletion comprising the sequence from nucleotide 662 to 1358 additionally destroyed the modification phenotype (data not shown). Exonucleolytic digestions starting from the *Pvu*II site at position 2234 generated only deletion mutants which extended into the 741 bp reading frame (data not shown). No mutants were created in which only parts of or the total 876 bp open reading frame were deleted and the 741 bp open reading frame was left intact. Since it seems reasonable that an active endonuclease gene cannot be stably maintained without the corresponding modification gene, this finding strengthens the conclusion that the 876 bp reading frame encodes M·*Sma*I and the 741 bp reading frame R·*Sma*I.

Mapping of the transcription start sites

Primer extension analyses were carried out to map the transcription start sites of the R·*Sma*I and M·*Sma*I genes in *S. marcescens* and in pSMN3 transformed *E. coli*. For mapping the 5′ terminal nucleotide of the R·*Sma*I mRNA oligodeoxynucleotide O3 which starts 20 nucleotides downstream of the ATG translation initiation codon was used as primer. Surprisingly a cDNA of around 300 nucleotides in length was obtained which comprised the entire small open reading frame (sORF) upstream of the R·*Sma*I coding region (Fig.

4a). Thus sORF and the R·*Sma*I coding region are transcribed as a polycistronic mRNA which starts with one of the cytosines at positions 93 to 95 in Fig. 3. The exact start site (C95) was obtained by primer extension analysis with oligonucleotide O4 which is complementary to the sORF sequence between positions 81 and 100 (data not shown). The transcription start site of the M·*Sma*I mRNA was mapped to position 1992 (Fig. 4b). A reverse cDNA transcript was obtained, however, only from RNA isolated from *S. marcescens*, but not from RNA isolated from *E. coli* transformed with the plasmid pSMN3.

Relationship of the S. marcescens restriction modification system to other systems

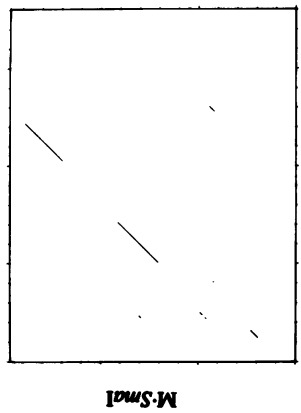
The amino acid sequences from sORF, R·*Sma*I and M·*Sma*I were compared with the amino acid sequences contained in the MIPSX protein data bank using the FastP program. The only significant homology was found between M·*Sma*I and the *Dpn*IIA methyltransferase of *Streptococcus pneumoniae* (31). A dot plot comparison of the two sequences is shown in Fig. 5a. No similarity was found between M·*Sma*I and any m^5C-methyltransferase for which sequence information was available (data not shown). Furthermore, a comparison of the M·*Sma*I protein sequence with the conserved sequence motifs of m^5C-methyltransferases described by Pósfai et al. (32) did not reveal any homology, too. Further dot plot comparisons (Fig. 5b and 5c) were carried out between the amino acid sequences of M·*Sma*I and the recently published sequences of M·*Pvu*II (33) and M·*Hin*fI (34). As shown in Fig. 5d, all four DNA modification enzymes (M·*Sma*I, M·*Dpn*IIA, M·*Pvu*II, and M·*Hin*fI) share a box of high homology in their C-proximal regions.

SDS-polyacrylamide gel-electrophoresis

Extracts from *S. marcescens* and *E. coli* JM107 transformed with pWS2, pSMN4 and pSMN4I were partially purified by phosphocellulose column chromatography and the column eluates were analyzed by SDS-polyacrylamide gel-electrophoresis (Fig. 6). The protein pattern displayed a unique polypeptide of around 29 kD in the extract prepared from *E. coli* transformed with pSMN4. This protein was also faintly visible in the extract from *S. marcescens*, but absent in the extracts prepared from *E. coli* transformed with pWS2 and pSMN4I. Therefore this polypeptide which comigrates with purified R·*Sma*I could be assigned to the *Sma*I endonuclease. One of the three extra bands in lanes 3 and 4 with sizes of approximately 28.5, 34.5 and 37 kD which are not present in lane 2 of Fig. 6 could possibly correspond to M·*Sma*I.

DISCUSSION

The genes encoding the class-II restriction-modification system of *Serratia marcescens* comprise three open reading frames two of which could be assigned to the endonuclease and to the DNA modification enzyme. The gene encoding the endonuclease is transcribed as a bicistronic mRNA together with a preceding small open reading frame (sORF) of 252 bp which starts with a GUG translation initiation codon. The last 10 nucleotides of this sORF overlap with the first 10 coding nucleotides of the endonuclease gene. A protein deriving from sORF with a calculated MW of 9442 daltons could, however, not be detected in phosphocellulose column eluates of extracts prepared from *S. marcescens* or *E. coli* transformed with the corresponding plasmids. A function of this sORF including its hypothetical gene product remains uncertain; our experiments only exclude the possibility that this open reading frame encodes a separate endonuclease or methyltransferase like the *Dpn*IIMA gene of the *Dpn*II restriction-modification system (35). An auxiliary or regulatory role of the open reading frame is nevertheless conceivable.


M·SmaI
M·HinfI
(c)
M·SmaI
M·PvuII
(b)
M·SmaI
M·DpnIIA
(a)

M·*Sma*I	250	V	L	D	P	F	F	G	S	G	T	T	G	V	V	263
M·*Dpn*IIA	223	I	L	D	P	F	V	G	S	G	T	T	G	V	V	236
M·*Pvu*II	269	V	V	D	I	F	G	G	S	N	T	T	G	L	V	282
M·*Hin*fI	215	V	L	D	P	F	F	G	T	G	T	T	G	A	V	228

(d)

Figure 5: Dot matrix plot comparisons between M·*Dpn*IIA and M·*Sma*I (a), M·*Pvu*II and M·*Sma*I (b), M·*Hin*fI and M·*Sma*I (c). The dot matrix plots were generated by the programs COMPARE and DotPlot of the UWGCG program package. Regions of partial homology are represented by diagonal stretches. For all comparisons the window size and the stringency value were 30 and 15, respectively. (d) Alignment of the highly homologous amino acid sequences in the C-proximal regions of the four analyzed DNA modification enzymes.

The *Sma*I restriction-modification system described here shares some common features with other cloned and characterized class-II systems. The two genes encoding the modification enzyme and the restriction endonuclease are located very close to each other in a tail-to-tail orientation, i. e., here the two mRNAs actually do overlap in their 3'-terminal coding regions. Since no classical terminators were found within this region when using the UWGCG TERMINATOR program, one might speculate that the found overlap of the two genes could play a role in transcriptional termination.

Our results show that the M·*Sma*I and R·*Sma*I promoters are functionally active in *E. coli*. The fact that no M·*Sma*I specific mRNA was detectable in *E. coli* transformed with pSMN3 could indicate that the M·*Sma*I promoter works less efficiently.

The cloned system conferred only partial resistance against incoming phage. Similar reduced phage-resistances have already been described for other class-II restriction-modification systems like *Taq*I (36) or *Msp*I (37). These findings might imply that also other functions than solely phage restriction could be the natural function of a restriction-modification system (2).

All m^5C-DNA modification methyltransferases investigated so far exhibit homologies in distinct sequence motifs arranged in 5 boxes and weaker similarities in another 5 boxes as recently reviewed by Pósfai et al. (32). Although these homologies are fairly weak, they can be well detected by dot-plot comparisons (38). The protein sequence deduced from the DNA sequence coding for M·*Sma*I did not exhibit any similarity with any of the sequence motifs which are characteristic for m^5C-methyltransferases. These results could suggest that M·*Sma*I is an N^4-methyltransferase, but this assumption would disagree with results that exclude the presence of N^4-methylcytosine in the DNA of *S. marcescens* (39, 40).

Another striking result is the extensive homology found in the C-terminal regions of M·*Sma*I, M·*Dpn*IIA, M·*Pvu*II, and M·*Hin*fI. Whereas M·*Pvu*II is an N^4C-methyltransferase, M·*Dpn*IIA and M·*Hin*fI are N^6A-methyltransferases. This similarity

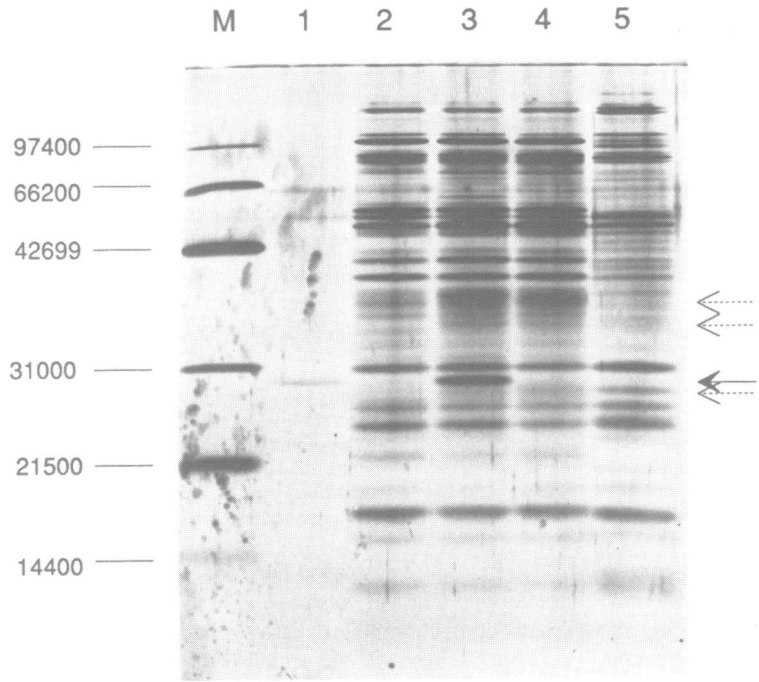


Figure 6: SDS-polyacrylamide gel electrophoresis of phosphocellulose column eluted proteins. M, protein molecular weight marker; 1, 180 U commercially available R·*Sma*I; 2, proteins from *E. coli* JM107 containing pWS2; 3, proteins from *E. coli* JM107 containing pSMN4; 4, proteins from *E. coli* JM107 containing pSMN4I which differs from pSMN4 by a two nucleotide insertion in the *Cla*I site (nucleotide 660 in Fig. 3) of the R·*Sma*I gene (*sma*IR); 5, proteins from *S. marcescens*. The bold arrow indicates the *Sma*I endonuclease, the thin dotted arrows indicate bands which could represent the *Sma*I modification enzyme.

between adenine-specific methyltransferases and a cytosine-specific DNA modification enzyme correlates well with a recent finding by Tao et al. (33), who detected substantial homologies between the cytosine-specific methyltransferase M·*Pvu*II and the adenine-specific methyltransferase M·*Eco*RI. The significance of these homologies remains uncertain, but could imply a different mechanism of modification in the *Sma*I restriction-modification system. This assumption is strengthened by the lack of a proline-cysteine pair in the amino acid sequence of M·*Sma*I, which is discussed to be a crucial constituent of the catalytic center of modification methyltransferases (41). This unusual feature has also been described for M·*Pvu*II, for which an alternative mechanism of methyltransfer has been proposed (33).

In spite of several attempts using different assay conditions, it was not possible to detect a *Sma*I-specific methylating activity neither in cell extracts of *S. marcescens* nor in extracts of the recombinant clones (data not shown), although the used assay conditions for analysis of *in vitro* activity of M·*Sma*I were varied broadly according to various described assay procedures (42–45). This failure to detect an M·*Sma*I activity *in vitro* might be due to a general instability of M·*Sma*I. Another possibility might be that M·*Sma*I acts via a different mechanism of DNA modification. One well-known example of a different modification is the generation of N^6-(1-acetamido)-adenine mediated by the *mom*-gene

product of the bacteriophage mu (46). Such a type of modification would not have been covered by the applied assay systems referring to N^6A-, 5C- or N^4C-methyltransfer (12). It should also be mentioned that no *Sma*I methyltransferase including the respective reaction conditions has been described yet in the literature. Possibly M·*Sma*I might need different cofactors as compared to other known methyltransferases which are part of a restriction-modification system. As it is well known that R·*Sma*I is unique in requiring potassium ions (13), it might also be that other specific requirements are characteristic for the *Sma*I modifying activity.

ACKNOWLEDGEMENTS

We thank H.-J. Höltke, K. Köhrer, G. G. Schmitz and K. Vogel for helpful discussions and advice. This work was supported by donations from Boehringer Mannheim GmbH.

REFERENCES

1. Wilson, G.G. (1988) Trends in Genetics **4**, 314–318.
2. Modrich, P. and Roberts, R.J. (1982) In Linn, S.M. and Roberts, R.J. (eds.), Nucleases. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp. 109–154.
3. Kessler, C. (1987) In Obe, G. and Basler A. (eds.), Cytogenetics. Springer-Verlag, New York Berlin Heidelberg Tokyo, pp. 225–279.
4. Bertani, G. and Weigle, J.J. (1953) J. Bacteriol. **65**, 113–121.
5. Luria, S.E. and Human, M.L. (1952) J. Bacteriol. **64**, 557–569.
6. Linn, S. and Arber, W. (1968) Proc. Natl. Acad. Sci. USA **59**, 1300–1306.
7. McClelland, M. (1988) Nucleic Acids Res. **16**, 2283–2294.
8. Messner, W. and Noyer-Weidner, M. (1988) Cell **54**, 735–737.
9. Kessler, C. and Höltke, H.-J. (1986) Gene **47**, 1–153.
10. Wilson, G.G. (1988) Gene **74**, 281–289.
11. Kessler, C. and Manta, V. (1989) Gene, in press.
12. McClelland, M. and Nelson, M. (1988) Gene **74**, 291–304.
13. Endow, S.A. and Roberts, R.J. (1977) J. Mol. Biol. **112**, 521–529.
14. Gautier, F., Bünemann, H. and Grotjahn, L. (1977) Eur. J. Biochem. **80**, 175–183.
15. Youssoufian, H. and Mulder, C. (1981) J. Mol. Biol. **150**, 133–136.
16. Grimont, P.A.D. and Grimont, F. (1978) Ann. Rev. Microbiol. **32**, 221–248.
17. Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene **33**, 103–119.
18. Bachmann, B.J. (1987) In Neidhardt, F.C. et al. (eds.), *Escherichia coli* and *Salmonella typhimurium*, Cellular and molecular biology. American Society for Microbiology, Washington DC, Vol. 2, pp. 1190–1220.
19. Hanahan, D. (1983) J. Mol. Biol. **166**, 557–580.
20. Brosius, J. and Holy, A. (1984) Proc. Natl. Acad. Sci. USA **81**, 6929–6933.
21. Rodriguez, R.L. and Tait, R.C. (1983) In Rodriguez, R.L. and Tait, R.C. (eds.), Recombinant DNA techniques. Addison-Wesley, Reading, Mass., pp. 45–46.
22. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
23. Sanger, F., Nicklen, S. and Coulson, A.R. (1977). Proc. Natl. Acad. Sci. USA **74**, 5463–5467.
24. Chen, E.Y. and Seeburg, P. (1985) DNA **4**, 165–170.
25. Aiba, H., Adhya, S. and de Crombrugghe, B. (1981) J. Biol. Chem. **256**, 11905–11910.
26. Domdey, H., Apostol, B., Lin, R.-J., Newman, A., Brody, E. and Abelson, J. (1984) Cell **39**, 611–621.
27. Fling, S.P. and Gregerson, D.S. (1986) Anal. Biochem. **155**, 83–88.
28. Needleman, S.B. and Wunsch, C.D. (1970) J. Mol. Biol. **48**, 443–453.
29. Lipman, D.J. and Pearson, W.R. (1985) Science **227**, 1435–1441.
30. George, D.G., Barker, W.C. and Hunt, L.T. (1986) Nucleic Acids Res. **14**, 11–15.
31. Lacks, S.A., Mannarelli, B.M., Springhorn, S.S., and Greenberg, B. (1986) Cell **46**, 993–1000.
32. Pósfai, J., Bhagwat, A.S., Pósfai, G. and Roberts, R.J. (1989) Nucleic Acids Res. **17**, 2421–2435.
33. Tao, T., Walter, J., Brennan, K.J., Cottermann, M.M. and Blumenthal, R.M. (1989) Nucleic Acids Res. **17**, 4161–4175.
34. Chandrasegaran, S., Lunnen, K.D., Smith, H.O. and Wilson, G.G. (1988) Gene **70**, 387–392.
35. de la Campa, A.G., Kale, P., Springhorn, S.S. and Lacks, S.A. (1987) J. Mol. Biol. **196**, 457–469.

36. Slatko, B.E., Benner, J.S., Jager-Quinton,T., Moran, L.S., Simcox, T.G., Van Cott, E.M. and Wilson, G.G. (1987) Nucleic Acids Res. **15**, 9781–9796.
37. Nwankwo, D.O. and Wilson, G.G. (1988) Gene **64**, 1–8.
38. Som, S., Bhagwat, A.S. and Friedman, S. (1987) Nucleic Acids Res. **15**, 313–331.
39. Butkus, V., Petrauskiene, L., Maneliene, Z., Klimasauskas, S., Laucys, V. and Janulaitis, A. (1987) Nucleic Acids Res. **15**, 7091–7102.
40. Ehrlich, M., Wilson, G.G., Kuo, K.C., and Gehrke, C.W. (1987) J. Bacteriol. **169**, 939–943.
41. Wu, J.C. and Santi, D.V. (1987) J. Biol. Chem. **262**, 4778–4786.
42. Mann, M.B. and Smith, H.O. (1977) Nucleic Acids Res. **4**, 4211–4221.
43. McClelland, M., Kessler, L. and Bittner, M. (1984) Proc. Natl. Acad. Sci. USA **81**, 983–987.
44. Nelson, M., Christ, C. and Schildkraut, I. (1984) Nucleic Acids Res. **12**, 5165–5173.
45. Nelson, M. and Schildkraut, I. (1988) In Wu, R. (ed.), Methods in Enzymology. Academic Press, New York, pp. 41–48.
46. Swinton, D., Hattman, S., Crain, P.F., Cheng, C.-S., Smith, D.L. and McCloskey J.A. (1983) Proc. Natl. Acad. Sci. USA **80**, 7400–7404.

This article, submitted on disc, has been automatically converted into this typeset format by the publisher.