Cloning and nucleotide sequence of the genes coding for the *Sau*96I restriction and modification enzymes

László Szilák+, Pál Venetianer and Antal Kiss*
Institute of Biochemistry, Biological Research Center, Hungarian Academy of Sciences, PO Box 521, 6701 Szeged, Hungary

Received June 29, 1990; Revised and Accepted July 23, 1990 EMBL accession no: X53096

ABSTRACT

The genes coding for the GGNCC specific *Sau*96I restriction and modification enzymes were cloned and expressed in *E. coli*. The DNA sequence predicts a 430 amino acid protein (M_r: 49,252) for the methyltransferase and a 261 amino acid protein (M_r: 30,486) for the endonuclease. No protein sequence similarity was detected between the *Sau*96I methyltransferase and endonuclease. The methyltransferase contains the sequence elements characteristic for m^5C-methyltransferases. In addition to this, M.*Sau*96I shows similarity, also in the variable region, with one m^5C-methyltransferase (M.*Sin*I) which has closely related recognition specificity (GGA/TCC). M.*Sau*96I methylates the internal cytosine within the GGNCC recognition sequence. The *Sau*96I endonuclease appears to act as a monomer.

INTRODUCTION

To understand the molecular mechanism by which type II restriction endonucleases and methyltransferases (MTases) recognize their target sequence, the genes of several restriction enzymes and MTases were cloned and sequenced (ref. in 1) Comparison of the deduced amino acid sequences revealed that endonucleases do not share homology with their cognate methylase (2). Endonucleases seem to have very different primary structures, there is only one case where homology was detected between two enzymes (3).

MTases can be classified in three groups according to the methylated base they produce: m^5C-, m^4C- and m^6A-MTases. Analysis of the amino acid sequences revealed that MTases of each type show a varying degree of similarity among themselves (2,4,5,). m^5C-MTases are the most uniform, they share extensive similarities in their amino acid sequence (6–18). The homologies detected in m^5C-MTases suggest a common building plan and evolutionary relatedness for these enzymes. This holds also for prokaryotic m^5C-MTases that are not part of a restriction-modification system (19–24) and even for a eukaryotic MTase (25).

More is known about m^5C-MTases than about the other two types. A Pro-Cys dipeptide has been proposed to be part of the catalytic site (26) and in the multispecific m^5C-MTases the domain that is responsible for the sequence specific DNA recognition has been identified: it is in the so-called variable region (27). Although there is no experimental evidence for it, the common architecture suggests that also in monospecific m^5C-MTases sequence specificity is determined by the variable region (28).

Obviously, the detailed structural analysis and enzymological studies of more enzymes with different specificities is necessary to understand the structural basis of sequence specificity. We started a systematic study of restriction and modification enzymes recognizing GGCC and related sequences (6,7,14,20). One restriction system belonging to this group is *Sau*96I (recognition sequence: GGNCC, ref.29). This system is especially interesting because the amino acid sequence of the *Sin*I endonuclease and methyltransferase has become available (13). The specificity of the *Sin*I system (GGA/TCC) is closely related to that of *Sau*96I. We report here the cloning and characterization of the genes coding for the *Sau*96I endonuclease and methyltransferase. These are the first GGNCC specific restriction endonuclease and MTase for which the amino acid sequence has been determined.

MATERIALS AND METHODS

Strains and media

Staphylococcus aureus PS 96 (29) was obtained from J.S. Sussenbach. *E. coli* strains ER1398 (*mcrB*$^-$) and ER1382 (*mcrB*$^+$, ref.30) were from E. Raleigh. M13 mp18 and mp 19 phages were grown in *E. coli* JM107 (31) or JM107 MA2 (*mcrB*$^-$, ref.32). Bacteria were grown in LB medium at 37°C. $Lambda_{vir}$ and $phi80_c$ phages (obtained from. J.Brooks) were used to test phage restriction.

Enzymes and chemicals

Restriction endonucleases were prepared in this institute or were purchased from New England Biolabs. DNA polymeraseI large fragment was from Vepex (Szeged), BAL31 nuclease from New

+Permanent address: Agricultural Biotechnology Center, POB 170, 2100 Gödöllő, Hungary
*To whom correspondence should be addressed

England Biolabs and modified T7 DNA polymerase (Sequenase) from United States Biochemical Corp. Deoxyadenosine 5'-alpha-(^{35}S)thiotriphosphate was purchased from Amersham.

Cloning methods

S. aureus PS 96 DNA was prepared by the SDS/proteinase K lysis method (33) with the modification that the cells were pretreated with lysostaphin (29). Isolation of plasmids, transformation of *E. coli* and agarose gel electrophoresis were done by standard procedures (34).

Determination of the nucleotide sequence

DNA fragments were cloned in mp18 and mp19 phage vectors (35) or in Bluescript plus vector (Stratagene) and sequenced by the chain termination method (36) using deoxyadenosine 5'-alpha-(^{35}S)thiotriphosphate and either DNA polymerase I large fragment or Sequenase. To sequence regions that were not readily accessible for sequencing with the universal sequencing primer, unidirectional deletions were prepared by the combined action of exonuclease III and mung bean nuclease (37).

Determination of the methylated base

The 246 bp *Hin*fI$_{3363}$–*Pst*I$_{3609}$ fragment of pBR322 was isolated from unmethylated (pBR322) and methylated (pSau3) plasmid DNA. The fragments labeled at the *Hin*fI end were subjected to pyrimidine specific chemical sequencing reactions (38,39) and run on a 6% sequencing gel.

Cloning of the *Sau*96I methyltransferase gene

S. aureus PS 96 DNA was partially digested with *Mbo*I and the digested DNA was ligated to pBR322 that had been cleaved by *Bam*HI and dephosphorylated by bacterial alkaline phosphatase. The ligated DNA was transformed into *E. coli* ER1398. Plasmid DNA isolated from 27,000 ampicillin resistant transformants was digested to completion with *Sau*96I. The digested DNA was used to transform ER1398. Ampicillin resistant colonies were selected.

Purification of *Sau*96I endonuclease

*Sau*96I endonuclease was purified from *S. aureus* PS 96 by a published procedure (29). This procedure, however, did not give satisfactory results when the *E. coli* clone containig the cloned *Sau*96I genes was used as starting material. To purify *Sau*96I endonuclease from ER1398(pSau5), cells from a 2 l dense culture were sedimented by centrifugation, then resuspended in PC buffer (10 mM K-phosphate pH 7.5, 1o mM 2-mercaptoethanol, 1 mM EDTA) and disrupted by French press. After centrifugation at 100,000×g for 1 h, 1 M NaCl was added to the supernatant and it was loaded on a Bio-Gel A0.5m column. Active fractions were dialysed against PC buffer and loaded on a Mono S FPLC column. *Sau*96I endonuclease was eluted with a 0–1 M NaCl gradient in PC buffer.

Determination of molecular weight by gel filtration

The molecular weight of *Sau*96I endonuclease under native conditions was determined by gel filtration on a Superose 12 HR10/30 (Pharmacia) FPLC column in a buffer containing 50 mM K-phosphate pH 7.0, 100 mM NaCl, 15 mM $MgCl_2$ and 10 mM 2-mercaptoethanol. *Sau*96I is active in this buffer (unpublished). Bovine serum albumin, ovalbumin and cytochrome c were used as molecular weight standards.

RESULTS AND DISCUSSION

Cloning of the *sau96IRM* genes

Transformation of *E. coli* by the *Sau*96I-digested *S. aureus* PS 96 plasmid library yielded several ampicillin resistant clones. Plasmids isolated from these clones were resistant to *Sau*96I digestion indicating that they carried the *Sau*96I methyltransferase gene. All plasmids analysed seemed to be identical by restriction enzyme digestions. One plasmid (pSau1, Fig.1) was chosen for further analysis. pSau1 contained an approximately 3.4 kb insert. The *E. coli* clone containing pSau1 did not restrict unmodified phage, suggesting that pSau1 did not code for the *Sau*96I endonuclease.

Assuming that the *Sau*96I methyltransferase and endonuclease genes are linked but the insert cloned in pSau1 did not contain both genes, we tried to reclone the *sau96IM* gene on an overlapping fragment. The pSau1 insert contains a unique *Sal*I site close to one end. Subcloning experiments revealed that this *Sal*I site lies outside of the methyltransferase gene. *S. aureus* PS 96 DNA was digested with *Sal*I and several other enzymes and hybridized, in Southern experiments, with a probe representing the central part of the pSau1 insert. In the *Sal*I-*Sph*I digest a 3.1 kb hybridizing fragment was identified. This fragment was cloned in *E. coli* ER1398 using pBR322 as vector and the same type of selection which was used to obtain pSau1. The plasmid (pSau3) isolated from this clone was resistant to *Sau*96I and the clone restricted unmodified phage (restr. ratio: 10^{-1}–10^{-2}). Since this clone did not restrict phage grown on ER1398(pSau1), we concluded that pSau3 coded, in addition to *Sau*96I

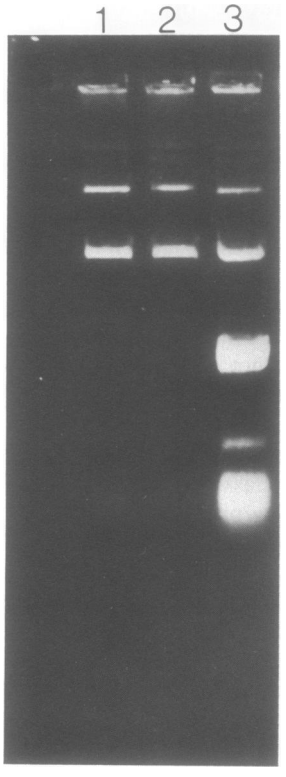

Fig. 1. Plasmid carrying the *Sau*96I methyltransferase gene is resistant to *Sau*96I digestion.
1. pSau1 undigested
2. pSau1 digested with *Sau*96I
3. pSau1 and pBR322 digested with *Sau*96I

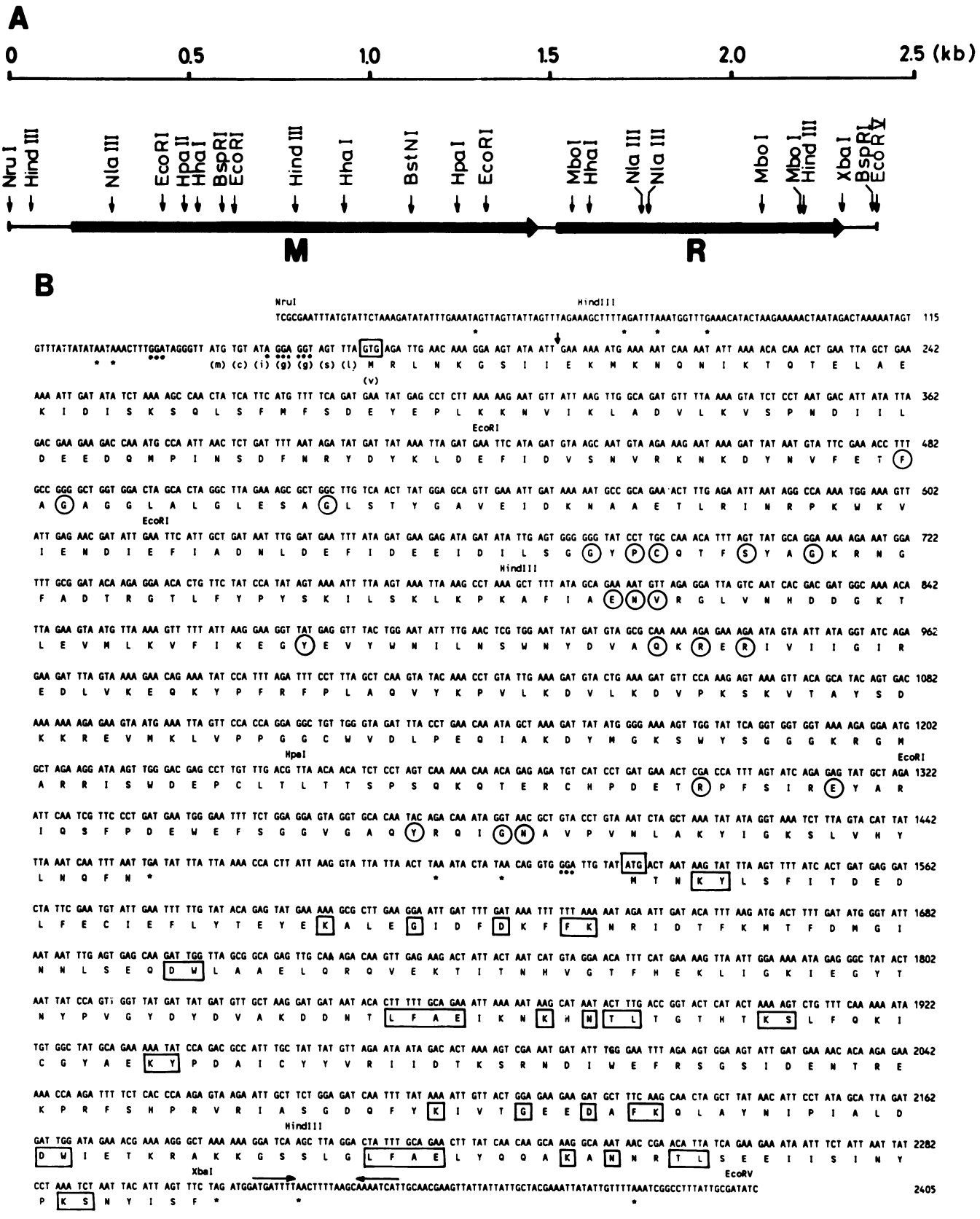

Fig. 2. A) Schematic map of the *Nru*I-*Eco*RV fragment containing the *sau96IRM* genes. Horizontal arrows indicate the coding regions. B) Nucleotide sequence of the *sau96IRM* genes and the deduced amino acid sequence of the *Sau*96I MTase and endonuclease. The putative translational initiation codons are boxed, Shine-Dalgarno sequences are indicated by dots and translational stop codons by asterisk. The sequence of a seven amino acid N-terminal peptide that would be translated if the ATG triplet at position 147 were used as initiation codon is shown in brackets, with lower case letters. Amino acid residues in the MTase which are most highly conserved in m^5C-methyltransferases (4) are circled. Amino acid residues in the nuclease that are part of an intramolecular repeat are boxed. Horizontal arrows mark a sequence which is likely to form a hairpin structure in the putative mRNA. A vertical arrow at nucleotide 194 indicates the endpoint of a deletion cloned in pSau6.

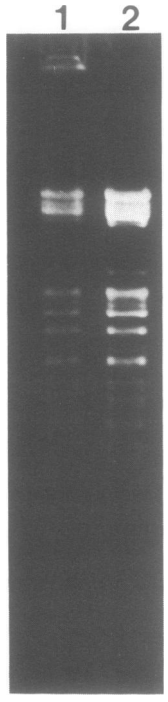

Fig. 3. Digestion of lambda phage DNA with *Sau*96I purified from (lane 1) ER1398(pSau5) and *S. aureus* PS 96 (lane 2).

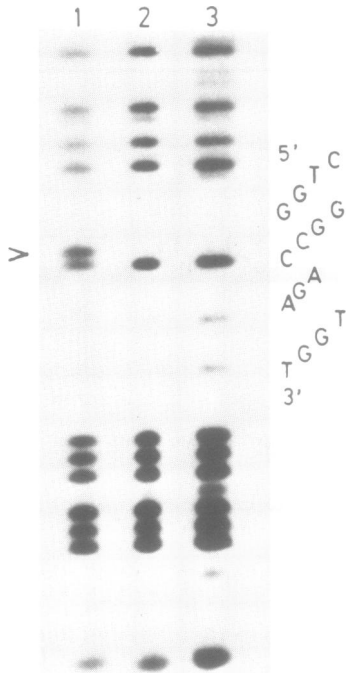

Fig. 4. M.*Sau*96I methylates the internal C within the GGNCC recognition sequence. .
1. Unmethylated DNA−C reaction
2. Methylated DNA−C reaction
3. Methylated DNA−C + T reaction
The absence of a band in the methylated DNA indicates the methylated base within the *Sau*96I site at 3410 in the pBR322 sequence. Part of the sequence is shown at right.

methyltransferase, also for *Sau*96I endonuclease. A 2.4 kb *Nru*I-*Eco*RV fragment of the pSau3 insert was cloned in pBR322 to give pSau5 (Fig.2.). pSau5 was methylated, ER1398(pSau5) exhibited phage restriction and was shown to produce *Sau*96I endonuclease (Fig.3), thus the *Nru*I-*Eco*RV fragment codes for the complete *Sau*96I system (Fig.2). The level of phage restriction shown by ER1398 containing pSau3 or pSau5 is lower than usually found with cloned restriction-modification systems (e.g. *Bsu*RI, ref.7.). A possible explanation for this is the low level of *Sau*96I endonuclease found in the clone. It was probably due to this low activity, that in order to get a usable enzyme preparation from ER1398(pSau3 or pSau5) we had to modify the original purification procedure described for *Sau*96I (29).

DNA sequence

The nucleotide sequence (2405 nucleotides) of the *Nru*I-*Eco*RV fragment was determined. Two large open reading frames were found (Fig.2): a longer (ATG_{147}-TGA_{1458}) coding for 437 amino acids, and a shorter (ATG_{1524}-TAG_{2307}) coding for 261 amino acids. Since pSau1 contained only the longer ORF, this was assigned to the *Sau*96I methyltransferase and the shorter one to *Sau*96I endonuclease. The two genes are tandemly arranged with 63 nucleotides between the coding regions.

The ORF coding for M.*Sau*96I contains, in addition to ATG at 147, another potential start codon: GTG at 168 and several less likely initiation codons further downstream. Starting from the *Nru*I site, BAL31 deletions were prepared to identify the functioning initiation codon. The shortened fragments were cloned in pUC18 in the orientation which would enable the *sau96IM* gene to be transcribed from the *lacZ* promoter. The fusion points were sequenced. One of the clones (pSau6), in which the first 194 nucleotides were deleted (Fig.2) and the *lacZ* and *sau96IM* genes were fused in different reading frames, did not express *Sau*96I methyltransferase, suggesting that translation starts upstream from this deletion end point. Of the two potential initiation codons located upstream of position 194, GTG_{168} is preceded by a much stronger Shine-Dalgarno sequence than ATG_{147} (16.9 kcal/mol vs. 7.7 kcal/mol, Fig.2, ref. 40). Gram positives are characterized by strong ribosomal binding sites (41) thus we think that GTG_{168} is the real translational start codon of the *Sau*96I methyltransferase gene. The reading frame between GTG_{168} and TGA_{1458} determines a protein of 430 amino acids (M_r: 49,252). The calculated molecular weight of the endonuclease is 30,486.

The *sau96IRM* genes are highly AT-rich. The coding regions are 67.6% and 70.1% A+T for the MTase and the endonuclease, respectively. There is no *Sau*96I recognition site in the sequenced region (2405 bp).

Determination of the methylated base

Plasmids containing the *sau96IM* gene transform the *mcrB*$^+$ *E. coli* strain ER1382 several orders of magnitude poorer than the isogenic *mcrB*$^-$ ER1398. Since the *mcrB* system recognizes GmC or PumC (30), the restriction shown by the mcrB$^+$ strain suggests that M.*Sau*96I methylates the internal cytosine (or in a less likely case both). The nature of the *Sau*96I methylation reaction was further investigated by the chemical sequencing method. A fragment of pBR322 DNA isolated from modified and unmodified plasmid was subjected to pyrimidine specific chemical sequencing reactions. The missing band in the modified

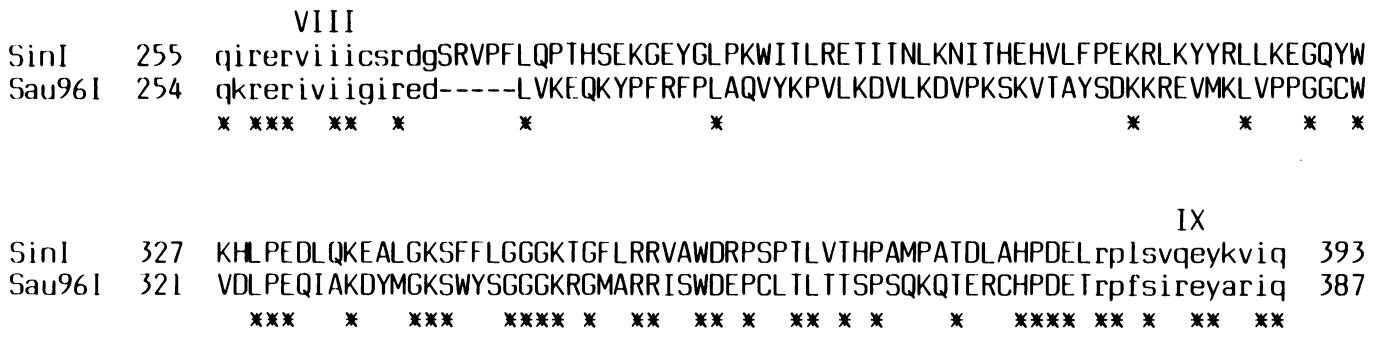

Fig. 5. Amino acid sequence in the variable region of the *Sau*96I and *Sin*I MTases. Residues in the variable region are typed in upper case and conserved blocks VIII and IX forming the bounderies of the variable regions (16) are typed in lower case. Identical amino acids are marked with asterisk.

DNA indicates that M.*Sau*96I methylates the inner C in the recognition sequence (Fig.4) Since N^4m-cytosine, unlike C^5m-cytosine, is known to react with hydrazine (43) this observation also suggests that M.*Sau*96I produces C^5m-cytosine.

Comparison of amino acid sequences

No amino acid sequence similarity was found between the *Sau*96I endonuclease and methyltransferase. This seems to be a common feature of type II restriction and modification enzymes and probably reflects the different molecular mechanism by which they recognize their target sequence. M.*Sau*96I contains all sequence motifs characteristic for m^5C-methyltransferases (4,Fig.2). In addition to this, its variable region exhibits striking similarity to the variable region of the *Sin*I methyltransferase (Fig.5). The variable region of multispecific m^5C-methyltransferases is responsible for the sequence specific DNA recognition (27). The common general architecture shared by the multispecific and monospecific enzymes and the identification of a conserved 'core' structure in the variable region of multispecific and monospecific enzymes led to the suggestion that the variable region determines sequence specificity also in monospecific enzymes (28). The similarity shown by the variable regions of M.*Sau*96I and M.*Sin*I is even more significant if we consider that these enzymes are in taxonomically unrelated bacteria: *Staphylococcus*, a Gram positive and *Salmonella*, a Gram negative. It seems likely that the similarity of the putative target recognizing domains of M.*Sau*96I and M.*Sin*I is not a chance similarity, rather it reflects the use of common structural elements for the recognition of closely related targets.

In contrast to the MTases, there is no similarity between R.*Sau*96I and R.*Sin*I. A search for intramolecular homologies found a pattern of a few amino acids in the N-terminal half of the *Sau*96I endonuclease which are repeated in the C-terminal half (Fig.2.). The observation of similar twofold and fourfold repeats in a few other type II restriction enzymes led to the assumption that these enzymes evolved by gene duplication (42). The intramolecular repeat detected in R.*Sau*96I gives further support to this idea. Such intramolecular duplication, however,is not evident at the level of primary structure in all type II restriction enzymes.

There are few data available for the subunit structure of type II endonucleases. Some enzymes, e.g. *Bsu*RI (44) act as monomers, others, e.g. *Eco*RI (45) as dimers. It was interesting to know what subunit structure *Sau*96I has. We found that, when tested by gel filtration under native conditions, it eluted from the column between ovalbumin (M_r: 43,000) and cytochrome c (M_r: 12,500) indicating that it consists of a single subunit (not shown). It was suggested (42) that it is the symmetrical structure of *Bsu*RI, indicated by the fourfold repeat in the amino acid sequence, that enables this enzyme to act as a monomer. Based on the available data, we think it is possible that type II endonucleases that act as monomers have retained more of the original intramolecular symmetry and, unlike enzymes that evolved as dimers, are characterized by intramolecular repeats evident also at the level of the primary structure. Further data will be needed to see whether such correlation really exists.

ACKNOWLEDGEMENTS

We thank Dr.J.S.Sussenbach, E.A.Raleigh, J.Brooks and R.Blumenthal for strains and E.Csorba and E.Komócsin for the technical assistance.

REFERENCES

1. Lunnen,K.D., Barsomian,J.M., Camp,R.R., Card,C.O., Chen,S.-Z., Croft,R., Looney,M.C., Meda,M.M., Moran,L.S., Nwankwo,D.O., Slatko,B.E., Van Cott,E.M. and Wilson,G.G. (1988) Gene, **74**, 25–32.
2. Chandrasegaran,S. and Smith,H.O. (1988) in Structure and Expression. Vol.1: From Proteins to Ribosomes. Eds. Sarma,R.H. and Sarma,M.H, Adenine Press, New York, pp. 149–156.
3. Stephenson,F.H., Ballard,B.T., Boyer,H.W., Rosenberg,J.M. and Greene,P.J. (1989) Gene, **85**, 1–13.
4. Pósfai,J., Bhagwat,A.S., Pósfai,G. and Roberts,R.J. (1989) Nucl. Acids Res., **17**, 2421–2435.
5. Klimasauskas,S., Timinskas,A., Menkevicius,S., Butkiene,D., Butkus,V. and Janulaitis,A. (1989) Nucl. Acids Res., **17**, 9823–9832.
6. Pósfai,G., Kiss,A., Erdei,S., Pósfai,J. and Venetianer,P. (1983) J. Mol. Biol., **170**, 597–610.
7. Kiss,A., Pósfai,G., Keller,C.C., Venetianer,P. and Roberts,R.J. (1985) Nucl. Acids Res., **13**, 6403–6421.
8. Sznyter,L.A., Slatko,B., Moran,L., O'Donnell,K.H. and Brooks,J.E. (1987) Nucl. Acids Res., **15**, 8249–8266.
9. Caserta,M., Zacharias,W., Nwankwo,D., Wilson,G.G. and Wells,R.D. (1987) J. Biol. Chem., **262**, 4770–4777.
10. Som,S., Bhagwat,A.S., and Friedman,S. (1987) Nucl. Acids Res., **15**, 313–331.
11. Slatko,B.E., Croft,R., Moran,L.S. and Wilson, G.G. (1988) Gene, **74**, 45–50.
12. Sullivan,K.M. and Saunders,J.R. (1988) Nucl. Acids Res., **16**, 4369–4387.
13. Karreman,C. and de Waard,A. (1988) J. Bacteriol., **170**, 2527–2532.
14. Kupper,D., Zhou,J.-G., Venetianer,P. and Kiss,A. (1989) Nucl. Acids Res., **17**, 1077–1088
15. Lin,P.M., Lee,C.H. and Roberts,R.J. (1989) Nucl. Acids Res., **17**, 3001–3011.
16. Card,C.O., Wilson,G.G., Weule,K., Hasapes,J., Kiss,A. and Roberts,R.J. (1990) Nucl. Acids Res., **18**, 1377–1383.

17. Walter,J., Noyer-Weidner,M. and Trautner,T.A. (1990) EMBO J., **9**, 1007–1013.
18. Karreman,C. and De Waard,A. (1990) J. Bacteriol., **172**, 266–272.
19. Buhk,H.-J., Behrens,B., Tailor,R., Wilke,K., Prada,J.J., Gunthert,U., Noyer-Weidner,M., Jentsch,S. and Trautner,T.A. (1984) Gene, **29**, 51–61.
20. Pósfai,G., Baldauf,F., Erdei,S., Pósfai,J., Venetianer,P. and Kiss,A. (1984) Nucl. Acids Res., **12**, 9039–9049.
21. Tran-Betcke,A., Behrens,B., Noyer-Weidner,M. and Trautner,T.A. (1986) Gene, **42**, 89–96.
22. Behrens,B., Noyer-Weidner,M., Pawlek,B., Lauster,R., Balganesh,T.S. and Trautner,T.A. (1987) EMBO J., **6**, 1137–1142.
23. Hanck,T., Gerwin,N. and Fritz,H.-J. (1989) Nucl. Acids Res., **17**, 5844.
24. Renbaum,P., Abrahamove,D., Fainsod,A., Wilson,G.G., Rottem,S. and Razin,A. (1990) Nucl. Acids Res., **18**, 1145–1152.
25. Bestor,T., Laudano,A., Mattaliano,R. and Ingram,V. (1988) J. Mol. Biol., **203**, 971–983.
26. Wu,J.C. and Santi,D.V. (1987) J.Biol.Chem., **262**, 4778–4786.
27. Balganesh,T.S., Reiners,L., Lauster,R., Noyer-Weidner,M., Wilke,K. and Trautner,T.A. (1987) EMBO J., **6**, 3543–3549.
28. Lauster,R., Trautner,T.A. and Noyer-Weidner,M. (1989) J. Mol. Biol., **206**, 305–312.
29. Sussenbach,J.S., Steenbergh,P.H., Rost,J.A., van Leeuwen,W.J. and van Embden,J.D.A. (1978) Nucl. Acids Res., **5**, 1153–1163.
30. Raleigh,E.A. and Wilson,G. (1986) Proc. Natl. Acad. Sci. USA **83**, 9070–9074.
31. Yanisch-Perron,C., Vieira,J. and Messing,J. (1985) Gene, **33**, 103–119.
32. Blumenthal,R.M., Gregory,S.A. and Cooperider,J.S. (1985) J. Bacteriol., **164**, 501–509.
33. Silhavy,T.J., Berman,M.L. and Enquist,L.W. (1984) Experiments with Gene Fusions. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
34. Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
35. Norrander,J., Kempe,T. and Messing,J. (1983) Gene, **26**, 101–106.
36. Sanger,F., Nicklen,S. and Coulson,A.R. (1977) Proc. Natl. Acad. Sci. USA, **74**, 5463–5467.
37. Henikoff,S. (1984) Gene, **28**, 351–358.
38. Maxam,A.M. and Gilbert,W. (1980) Methods Enzymol., **65**, 499–560.
39. Ohmori,H., Tomizawa,J. and Maxam,A.M. (1978) Nucl. Acids Res., **5**, 1479–1485.
40. Tinoco,I., Borer,P.N., Dengler,B., Levine,M.D., Uhlenbeck,O.C., Crothers,D.M. and Gralla,J. (1973) Nature New Biol., **246**, 40–41.
41. McLaughlin,J.R., Murray,C.L. and Rabinowitz,J.C. (1981) J. Biol. Chem., **256**, 11283–11291.
42. Lauster,R. (1989) J. Mol. Biol., **206**, 313–321.
43. Butkus,V., Klimasauskas,S., Kersulyte,D., Vaitkevicius,D., Lebionka,A. and Janulaitis,A. (1985) Nucl. Acids Res., **13**, 5727–5746.
44. Bron,S. and Horz,W. (1980) Methods Enzymol., **65**, 112–132.
45. McClarin,J.A., Frederick,C.A., Wang,B.-C., Greene,P., Boyer,H.W., Grable,J. and Rosenberg,J.M. (1986) Science, **234**, 1526–1541.