Nucleic Acids Research, 1992, Vol. 20, No. 23 6267–6273

Structure and evolution of the *Xcy*I restriction-modification system

Barbara E.Withers, Linda A.Ambroso and Joan C.Dunbar*
Wayne State University School of Medicine, Detroit, MI 48201, USA

Received September 4, 1992; Revised and Accepted October 27, 1992 GenBank accession no. M98768

ABSTRACT

The *Xcy*I restriction-modification system from *Xanthomonas cyanopsidis* recognizes the sequence, CCCGGG. The *Xcy*I endonuclease and methylase genes have been cloned and sequenced and were found to be aligned in a head to tail orientation with the methylase preceding and overlapping the endonuclease by one base pair. The nucleotide sequence codes for an N4 cytosine methyltransferase with a predicted molecular weight of 33,500 and an endonuclease comprised of 333 codons and a molecular weight of 36,600. Sequence comparisons revealed significant similarity between the *Xcy*I, *Cfr*I and *Sma*I methylisomers. In contrast, no similarity was detected between the primary structures of the *Xcy*I and *Sma*I endonucleases. The *Xcy*I restriction-modification system is highly homologous to the *Xma*I genes, although the DNA sequences flanking the genes rapidly diverge. The sequence of the *Xcy*I endonuclease contains two motifs which have recently been identified as essential to the activity of the *Eco*RV endonuclease.

INTRODUCTION

A large number of restriction-modification systems have been identified and the genes for more than 60 systems have been cloned to date(1). The genes for the endonuclease and methylase are generally linked although the relative orientation of the genes varies(1). Many of the enzymes comprising the restriction-modification systems are members of families of endonucleases and methylases which have similar recognition and catalytic specificity(2). Comparative analyses of the organization and structure of the genes provides an insight into the evolution of the enzymes and the potential to identify conserved domains which may be critical to the structure and function of the proteins.

The *Sma*I, *Xcy*I, *Xma*I and *Cfr*91 restriction-modification systems have a common recognition sequence, CCCGGG. The methylases from the *Xma*I(1), *Cfr*91(3), *Xcy*I(Ambroso and Dunbar, unpublished) and *Sma*I(4,5) systems are all N4 cytosine methyltransferases and have identical catalytic specificity(CmCCGGG). The *Xma*I, *Cfr*91 and *Xcy*I endonucleases are perfect isoschizomers and cleave between the external cytosines. In contrast, the *Sma*I endonuclease is an imperfect isoschizomer which cleaves between the internal CpG of the recognition site. Sequence comparisons have revealed significant primary structure similarity between the *Cfr*91 and *Sma*I methylases(4). The *Sma*I endonuclease, however, is the only restriction enzyme of this family for which the sequence has been reported(5). In an attempt to utilize sequence conservation as a method for defining potential functional domains of the endonucleases and methylases the *Xcy*I restriction-modification system has been cloned and sequenced.

Cloning of chromosomal endonuclease and methylase genes is generally based on selection for the restriction or modification phenotype(6). Early studies used lambda infection assays to screen for recombinants containing the endonuclease gene. However, *in vitro* selection for the methylation phenotype, by plasmid resistance to cognate endonuclease cleavage, is now more frequently utilized. Piekarowicz *et al.*(7) have recently described an alternative method for selecting subcloned methyltransferase genes based on the induction of the SOS response by Mcr and Mrr systems in the presence of methylated DNA. In only a few cases has screening by homology been attempted. Rodicio and Chater(8) detected hybridization between some but not all isoschizomers of the *Sal*GI genes. Similarly, Mullings *et al*(9) have screened, by Southern analyses, DNA from bacteria which produce *Pst*I isoschizomers. Hybridization was detected with approximately 30% of the strains when using a probe corresponding to the N-terminus of the endonuclease. Greater success has recently been achieved in detecting homologous genes in related bacteria: Dusterhoft *et al*(10) have successfully cloned the *Hgi*BI restriction-modification system from *H. giganteus* Hpg5 using a probe derived from the *Hgi*CII methylisomer.

The high degree of sequence similarity between the N4 methyltansferases, particularly those with the common recognition sequence, CCCGGG, and the close phylogenetic relationship between *X. malvacaerum* and *X.cyanopsidis* prompted the use of Southern blot analysis, using a probe derived from the *Xma*I methylase, to screen for homologous sequences in genomic DNA from *X. cyanopsidis*. The genes for the *Xcy*I methylase and endonuclease were consequently cloned and sequenced and found to be highly homologous to *Xma*I restriction-modification enzymes. Despite the identity of the endonuclease and methylase genes, the flanking sequences

* To whom correspondence should be addressed

diverged less than 2kb from the restriction- modification systems. In addition, there appears to be no system comparable to the *Xma*III restriction- modification present in *X. cyanopsidis*. Sequence comparisons have revealed that two motifs, which have recently been shown to be essential for the activity of *Eco*RV endonuclease(11,12) are also present in the sequence of the *Xcy*I endonuclease.

MATERIALS AND METHODS

Bacterial strains and plasmids

Xanthomonas cyanopsidis 13D5 was provided by C.Kado, University of California, Davis. *E.coli* K802(Hsd R_K^- Hsd M_K^-, McrA$^-$ McrB$^-$) and ER1451(Hsd R_K^-, Hsd M_k^-, Mcr A$^-$, Mcr B$^-$) were obtained from New England Biolabs. XL1-B(13) cells were obtained from Stratagene.

Enzymes and chemicals

Restriction enzymes were obtained from New England Biolabs or BRL and were used according to the manufacturers' instructions. T4 DNA ligase was also obtained from New England Biolabs. [alpha- ^{32}P]-ATP(>3,000Ci/mmmole), [gamma-^{32}P]-dCTP(>3,000 Ci/mmole), Hybond H nitrocellulose, Klenow and Hot Tub DNA polymerase were from Amersham. 'Gene Clean' was purchased from Bio101. Sequenase DNA sequencing kit and calf intestine alkaline phosphatase were obtained from United States Biochemical. Oligonucleotide primers used for sequencing or for PCR analysis were prepared by solid phase phosphoamidate synthesis at the Macromolecular Structure Facility, Michigan State University. Proteinase K was purchased from Boehringer Mannheim Biochemicals. DE-45 paper was from Schleicher and Schull.

DNA preparation

Genomic DNA from X. *cyanopsidis* and S. *marcescens* was isolated by detergent lysis. Bacterial cultures were grown overnight at 30°C in LB media. Cells were pelleted by centrifugation and resuspended in 5 mls of phosphate buffered saline. Proteinase K was added to a final concentration of 0.4mg/ml followed by the addition of a 1/10 volume of lysis buffer(100mM Tris-Cl, pH 8.0, 1% SDS and 0.1mM EDTA). The suspension was heated at 55°C until the solution cleared. The DNA was purified by phenol and chloroform/isoamyl alcohol extraction and recovered by ethanol precipitation.

Identification of genes and DNA sequence analysis

DNA fragments containing the *Xcy*I endonuclease and methylase genes were identified by hybridization to probes derived from the *Xma*I restriction and modification genes(as described below). The *Xma*I endonuclease and methylase probes were isolated from low melting temperature agarose gels and were random primer labelled using [gamma-^{32}P]-dCTP(14). Genomic DNA from *X. cyanopsidis* was digested with *Eco*RI, *Pst*I and *Hind*III and the products electrophoretically separated on a 1% agarose gel. The fragments were transferred to nitrocellulose according to the method of Southern(15). Hybridizations were carried out at 39°C overnight. Filters were washed four times at room temperature(2×SSC, 15min; 2×SSC, 15min; 1×SSC, 15min; 1×SSC, 0.1% SDS, 15min) and 1×SSC, 0.1% SDS for 15min at 55°C.

Fragments used for subcloning were isolated from agarose gels by either electroelution onto DE-45 paper or purification using Gene Clean according to manufacturer's instructions. Fragments were subcloned into M13mp19 and single stranded DNA was used as template for dideoxy sequence analysis with either universal primer or synthetic oligonucleotide primers complementary to predetermined sequences.

PCR amplification of sequences

Genomic and plasmid DNA sequences were amplified using Hot Tub DNA polymerase. Reactions containing 300ng plasmid DNA and 450ng each primer were carried out in 25mM Tris-Cl, pH 9.5, 50mM KCl, 10mM $MgCl_2$, 1mg/ml BSA, 200μM of each dNTP and 4% formamide. The DNA was subjected to 35 cycles of amplification with a cycle profile of 94°C, 2min; 52°C, 3min and 72°C for 4min. 500ng of template and 400ng of primers were used for amplification of genomic DNA. Reaction products were subsequently analysed on 1% agarose gels. The sequence of the primers used for amplification was as follows:

A′ CGTTCCATTGCTCCAATTCGGCG
A ATCGAAGTCAAGAGC
B′ AGATACCAGCCGTCA
B GCCGCGAGCGTCAGCGTCAGG
C′ CCTGACGCTGACGCTCGCGGC
C TCGTGCAGATCATGACGCAGCT

RESULTS

Isolation of the *Xcy*I restriction and modification genes

The strategy for cloning the *Xcy*I restriction and modification genes utilized the potential homology between the two *X. cyanopsidis* and *X. malvacaerum* species as well as the possible sequence conservation between the restriction-modification genes. Fig.1(A) shows the restriction map for the *Xma*I restriction-modification system. The genes were initially cloned into pUC19 on a 10kb *Hind*III fragment and were subsequently localized to the 3.1kb *Nru*I/*Eco*RI fragment(Wilson and Lunnen, unpublished). The latter fragment was digested with *Xho*I to yield a probe(Xh550) specific for the *Xma*I methylase gene. The Xh550 probe was hybridized to Southern blots of restriction digests of genomic DNA from *X. cyanopsidis*. The probe specifically hybridized to a 4.3kb *Pst*I fragment and a >10kb *Eco*R1 fragment(Fig.2). DNA fragments which comigrated with the hybridizing band from the *Pst*I digest were subcloned into M13 to generate a size fractionated library. The library was rescreened with the Xh550 probe and the positive clones isolated. A partial restriction map for the insert of the *Pst*I clone(pMP43) is shown in Fig.1(B). Further Southern hybridizations revealed that the Xh550 probe localized to a 1.1kb *Sma*I-*Pst*I fragment(Fig.2).

In a similar manner, a specific probe was isolated corresponding to the *Xma*1 endonuclease. A 670bp probe(Dd7) was generated by *Dde*I digestion of the 3.1kb*Nru*I/*Eco*RI fragment. Dd7 hybridized to the same 4.3kb *Pst*I and >10kb *Eco*RI fragments as did the methylase gene probe(Fig.2) Within the *Pst*I fragment, the Dd7 probe hybridized to the 1.6kb *Sma*I fragment, which is adjacent to that detected by the methylase gene probe.

Very little background hybridization was detected using the *Xma*I methylase and endonuclease specific probes. In addition, probes generated from regions flanking the 3.1kb *Nru*I/*Eco*RI fragment failed to hybridize to the genomic DNA from *X. cyanopsidis*. Similar experiments were also carried out using probes derived from the *Sma*I restriction and modification genes.

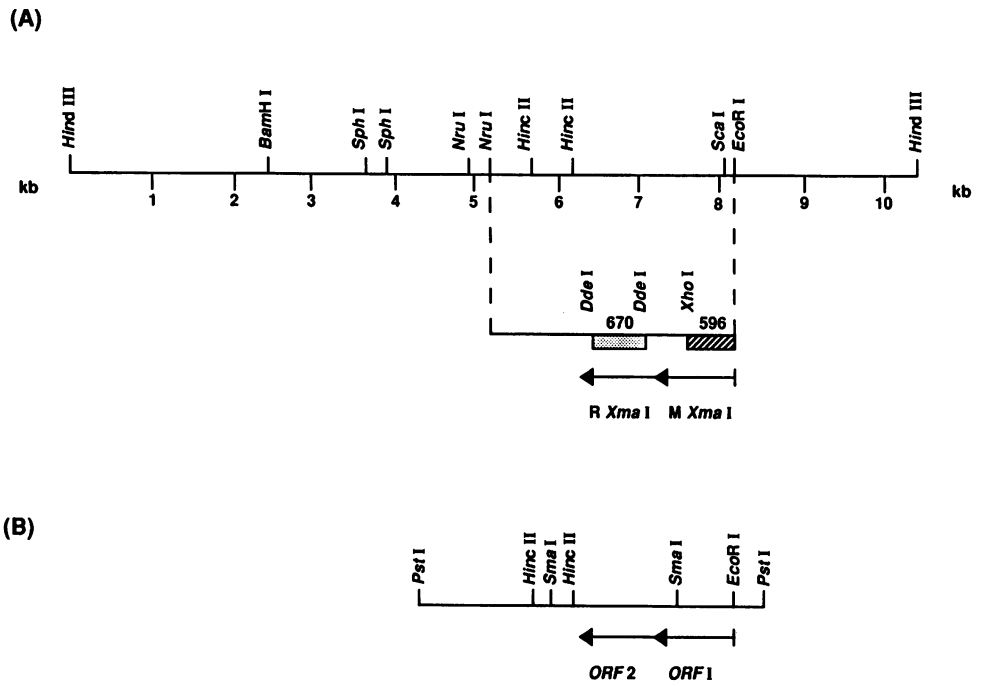

Figure 1. Comparison of the restriction maps of clones containing the **(A)** *Xma*I and **(B)** *Xcy*I restriction-modification systems. Probes specific for the *Xma*I methylase(Xh550) and endonuclease(Dd7) genes were isolated from the *Xma*I clone and used in Southern analyses of *X. cyanopsidis* genomic DNA to identify fragments containing homologous sequences. The location is shown for the two open reading frames detected by sequence analysis of the *Pst*I fragment containing the *Xcy*I restriction-modification system.(The restriction map of the *Xma*I clone was provided by Keith Lunnen and Geoff Wilson).

No hybridization was detected between the *Sma*I and *Xcy*I genes, even under conditions of low stringency.

Sequence analysis

DNA encompassing the two *Sma*I fragments from the pMP43 clone was further subcloned and sequenced. Two open reading frames were detected in a head-to-tail orientation(Fig 1B). ORF1 coded for a protein of 300 amino acids and a calculated molecular weight of 33,580. Initial examination of the sequence of ORF1 revealed the presence of two sequence motifs, FGGSG and TSSPY. While the former motif contains the GXG tripeptide which is present in all Adomet dependent enzymes(16), the TSPPY sequence frequently occurs in the N4 cytosine methyltransferases(17). The *Pst*I/*Hinc*II fragment was subcloned into pUC 19(pXcyI), propagated in *E.coli* K802 cells and analysed for *in vivo* methylase activity: Neither *Xma*I nor *Sma*I endonuclease cleaved the purified pXcyI plasmid.(In contrast to pMXcy, pMP43 was digested by both *Sma*I and *Xcy*I endonucleases. Presumably the higher copy number of the pUC plasmid enables sufficient expression of the methylase gene to confer resistance to cognate endonuclease digestion of the plasmid). ORF1 was consequently identified as the *Xcy*I methylase. In addition to the proposed initiator ATG codon at position 361, there is also a second potential initiator codon(albeit TTG) 32 amino acids upstream at nucleotide position 265. N-terminal sequencing of the purified protein will be necessary to determine which of the codons is utilized. However, sequence comparisons with the *Cfr*91 methylase(see below) would suggest the ATG codon as the correct initiation site.

The start codon for the second open reading frame overlaps the stop codon of the methylase gene by one nucleotide and codes

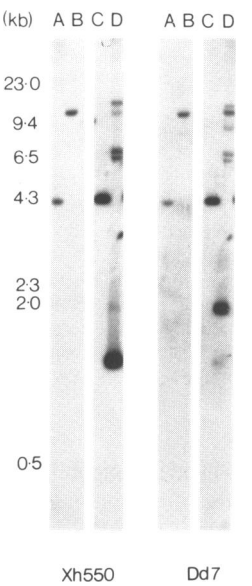

Figure 2. Hybridization of the (1) Xh550 and (2) Dd7 probes with **(A)** Genomic DNA from *X. cyanopsidis* digested with *Pst*I, **(B)** *X.cyanopsidis* genomic DNA digested with *Eco*RI, **(C)** pMP43 digested with *Pst*I and **(D)** pMP43 digested with *Sma*I.

for a protein of 333 amino acids with a predicted molecular weight of 36,644. The *Xcy*I endonuclease has been purified to near homogeneity and has a molecular weight, calculated from electrophoretic mobility on denaturing polyacrylamide gels, of

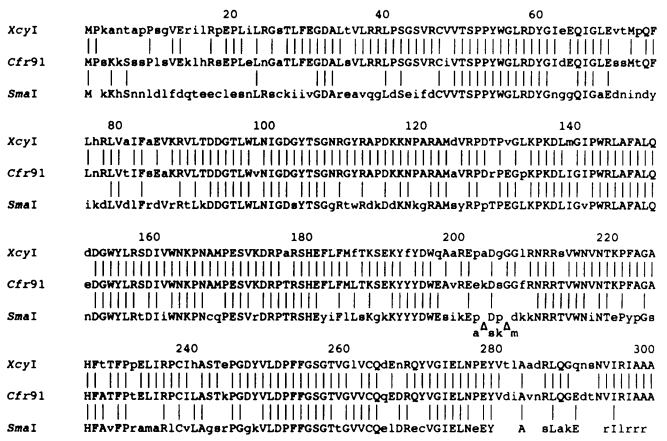

Figure 3. Alignment of the amino acid sequences of the *Xcy*I, *Cfr*91 and *Sma*I methylases. Amino acid residues that are identical between two or more sequences are shown in capital letters. Arrowheads indicates the position of additional amino acids within the sequence of the *Sma*I methylase. The numbering system is for the *Xcy*I methylase.

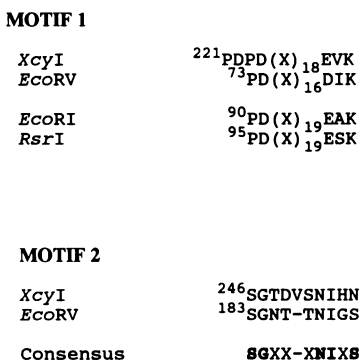

Figure 4. Alignment of the two motifs common to the *Xcy*I and *Eco*RV endonucleases. Motif 1 corresponds to the proposed magnesium binding domain and the corresponding sequences are also shown for the *Eco*RI and *Rsr*I endonucleases. The sequences for Motif 2 have been aligned for the *Xcy*I and *Eco*RV endonucleases. A gap has been inserted into the sequence of the *Eco*RV sequence to maximize the alignment. The consensus sequence is derived from Thielking *et al.*(11) with the inclusion of a gap to accommodate the additional amino acid present in the sequence of the *Xcy*I endonuclease.

approximately 35,000(Dunbar, unpublished). ORF2 is, therefore, proposed to encode the *Xcy*I endonuclease.

Approximately 2.4kb of total DNA was sequenced including 350bp upstream of the methylase gene and 150bp distal of the endonuclease. No other full length open reading frames were detected within these flanking regions, although the available sequence upstream of the methylase does contain an open reading frame of nearly 70 amino acids. A search of the PIR data bank did not reveal any significant similarity of this sequence with other known proteins.

The methylase and endonuclease start codons are preceded by potential ribosome binding sites, GAGAGA and AAGGA, centered at positions 339 and 1237, respectively. The 5′ region was also screened for sequences similar to the *E.coli* major promoter. A potential promoter at '−10'(TAAGATT) and '−35'(GTGACCATTGCAG) occurs upstream of the methylase gene as shown. The *in vivo* activity which was detected in the cells transformed with pMXcy may reflect the activity of this putative promoter in *E.coli*. The alternative TTG start codon also contains potential Shine−Dalgarno and promoter regions. No obvious rho independent termination signals were detected downstream of the endonuclease gene.

The sequence of the *Xcy*I genes demonstrated an unusual base composition. Many of the restriction- modification genes exhibit a lower G+C content compared to the genomic DNA of the host cell and also display a preference for A/U bases at the third position of the codon(18). A slight, approximate 10%, decrease is detected in the G+C content for the *Xcy*I system; however, *Xcy*I endonuclease and methylase genes display a C/G(65%) rather than A/U bias at the wobble position. These differences may reflect the relative GC richness of *Xanthomonas* spp (approximately 65%(19)). A similar difference in the content and base utilization at the third position has also been detected for the *Rsr*I endonuclease which exhibits a GC bias consistent with the relative high GC content *R. sphaeroides*(20).

Sequence comparisons

Comparison of the sequence of the *Xcy*I methylase with published sequences of other type II methyltransferases revealed the greatest similarity with the *Cfr*91 methylase(Fig.3). The proteins shared approximately 80% amino acid sequence identity and an additional 5% similarity arising from conservative substitutions. The *Xcy*I methylase also revealed greater than 50% identity with the *Sma*I methylase. Fig.3 also reveals that the non-conserved amino acids in the *Sma*I methylase are clustered at the amino terminus of the protein. However, the most significant alignment of the *Xcy*I methylase occurred with the *Xma*I methylase(Lunnen and Wilson, unpublished) for which there was 100% homology for both the protein and the DNA sequences.

In addition to the sequence conservation between the methylases, the *Xcy*I and *Xma*I endonucleases also revealed 100% identity. In contrast, the *Xcy*I endonuclease exhibited no significant similarity with other endonucleases which have been sequenced to date, including the *Sma*I endonuclease. The sequence of the *Xcy*I perfect isoschizomer, *Cfr*91 endonuclease, has not yet been published: however, it is reported(1) to exhibit greater than 70% similarity to the *Xma*I(and hence, *Xcy*I) endonuclease.

Despite the lack of global sequence conservation, the *Xcy*I endonuclease contains two amino acid motifs which have recently been identified and shown to be essential to the activity of the *Eco*RV endonuclease(11,12). Fig.4 shows the alignment of the conserved motifs between the *Eco*RV and the *Xcy*I endonucleases. Motif 1 is comprised of two conserved sequence elements separated by 16−20 amino acids. The second motif, identified by Thielking *et al.*(11) has the consensus sequence, SG---NI-S. A similar sequence occurs in the *Xcy*I endonuclease, albeit it with a conservative substitution(S-N) at the C-terminus and an amino acid insertion between the conserved glycine and asparagine.

Flanking sequence conservation

The high degree of sequence similarity between the genes for the *Xma*I and *Xcy*I restriction and modification enzymes prompted further investigation of the extent to which the DNA sequences of *X. cyanopsidis* and *X. malvacaerum* were conserved. Comparison of the restriction map of the *Xma*I clone with the pMP43 subclone revealed that the *Pst*I site upstream of the methylase gene was conserved as were the two *Hin*cII sites distal

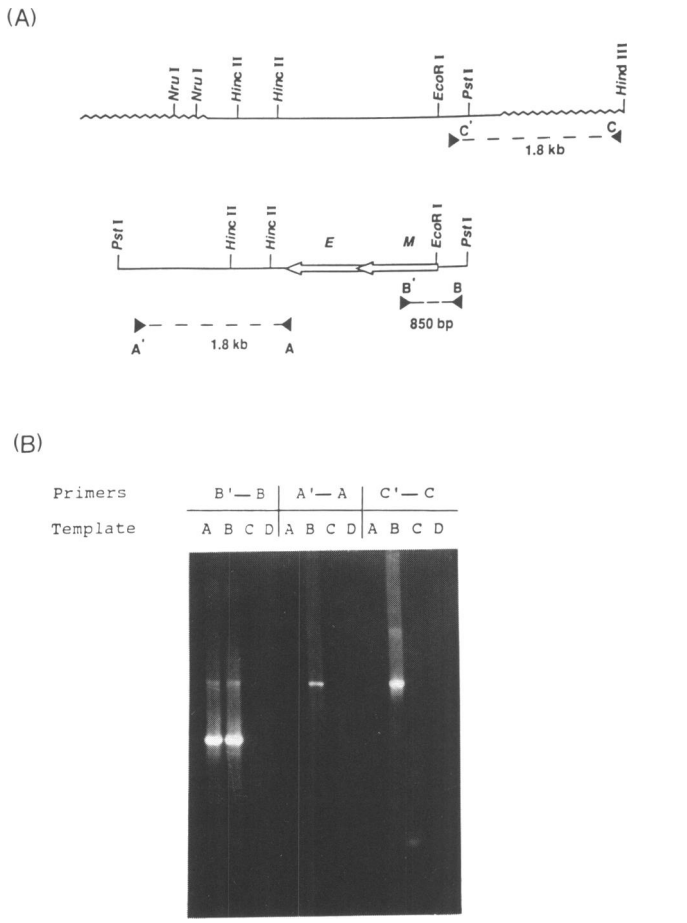

Figure 5. DNA sequence conservation between *X. malvacaerum* and *X. cyanopsidis*. **(A)** Restriction maps of the clones containing the *Xma*I and *Xcy*I restriction-modification genes indicating the location of the primers used for PCR amplification of the DNA sequences flanking the genes. **(B)** Gel electrophoresis of the PCR products obtained from amplification of (A) pMP4.3, containing the *Xcy*I genes subcloned in M13mp19, (B) pXmaI, containing the *Xma*I genes subcloned into pUC 19, **(C)** M13mp19 and **(D)** pUC 19.

of the endonuclease. However, additional sites downstream of the endonuclease appeared to diverge as evidenced by the lack of the *Pst*I site in the *Xma*I clone and and the failure to detect the *Nru*I site in the restriction map of the pXcyI clone(Fig.1).

Primers complementary to the sequences of the *Xma*I and *Xcy*I clones were used to amplify homologous sequences in the pMP43(*Xcy*I) and *Xma*I clones. Amplification of sequences downstream of the *Xcy*I endonuclease using primers A′ and A yielded a 1.8kb fragment for pMP43 (Fig.5). No amplification of the *Xma*I clone was detected. The lack of an amplified product is consistent with the *Pst*I and *Nru*I polymorphisms which were previously detected in this region. The DNA sequences, therefore, appear to diverge less than 2kb from the endonuclease gene. Sequences upstream of the methylase genes are conserved for at least 400bp as evidenced by (i) conservation of the *Pst*I site and (ii) amplification of the *Xma*I clone using the *Xcy*I specific primers B′ and B(Fig.5). The extent of conservation at the 5′ end of the genes was further defined. The sequence of the oligonucleotide (C′) corresponded to the non-coding strand of the amino terminus of the *Xma*I methylase. The sequence of the second primer(C) was derived from subcloning and sequencing of the *Pst*I/*Hin*dIII fragment from the *Xma*I clone. Primers C′ and C, complementary to the *Xma*I sequence, were used to amplify genomic DNA from *X. cyanopsidis*. A 1.8kb fragment was amplified from the *Xma*I clone. However, no amplification of the *X. cyanopsidis* genomic DNA was detected.(In control experiments, the same template DNA was readily amplified using the *Xcy*I specific primers B′ and B). The sequences of *X. cyanopsidis* and *X. malvacaerum* therefore appear to diverge less than 2kb on either side of the conserved restriction-modification system.

Conservation of the *Xma*III restriction-modification system

X. malvacearum has been reported to contain at least three distinct restriction-modification systems(21,22). De Feyter and Gabriel(23) have localized two of the methylation activities within a 13.8kb *Eco*RI fragment and have identified one of the genes as the the *Xma*III modification gene. Comparison of the restriction map of clones containing the *Xma*I and *Xma*III genes identifies the second component as the *Xma*I restriction-modification system. The *Xma*I and *Xma*III restriction- modification systems, therefore, reside within approximately 5kb on a single *Eco*RI fragment. RFLP and PCR analyses suggested that sequences greater than 2kb beyond the *Xcy*I restriction-modification system were not conserved in *X. malvacaerum*. Nonetheless, in view of the similarity between the *Xcy*I and *Xma*I systems, the possibility was explored that a restriction-modification system, homologous to *Xma*III, may also occur in *X. cyanopsidis*.

The 5.1kb *Eco*RI/*Hin*dIII fragment from pUFR059(a plasmid containing the *Xma*III restriction- modification genes and kindly provided by D. Gabriel) was used as a probe in Southern analysis of *X. cyanopsidis* genomic DNA. Very weak, non-specific, hybridization was detected, suggesting that a system homologous to *Xma*III is not present in *X. cyanopsidis*. The lack of hybridization with the *Xma*III probe is consistent with the detection of only a single endonuclease(*Xcy*I) activity in crude cellular extracts from *X. cyanopsidis*(24, Dunbar, unpublished).

DISCUSSION

The genes for the *Xcy*I restriction-modification system have been identified by hybridization to probes specific for the *Xma*I methylase and endonuclease genes. Subsequent sequence analyses revealed that the organization and primary structure of the *Xcy*I restriction-modification system is identical *Xma*I.

Analyses of isoschizomeric restriction-modification systems have revealed that while some systems, for example, *Hha* and *Hin*fI(25) appear to have evolved independently, others such as the *Taq*I and *Tth*HB8I(26) enzymes, are clearly related. However, no other restriction-modification systems have as yet demonstrated the absolute homology which was detected between the genes of the *Xma*I and *Xcy*I systems. The phylogenetic relationship of the host organisms and the identical catalytic activities of both the endonucleases and methylases were a predisposition towards potential sequence similarities. Indeed, it has been suggested that *X. cyanopsidis* and *X. malvacaerum* are different isolates of the same species(19). However, whereas the restriction-modification systems have been well conserved, the flanking sequences diverge less than 2kb beyond the endonuclease and methylase genes. Furthermore, despite the proximity(<5kb) of the *Xma*I and the *Xma*III systems and the potential for co-inheritance, genes homologous to *Xma*III have not been conserved in *X. cyanopsidis*. *X. cyanopsidis* and *X. malvacaerum*, therefore, appear to be distinct subspecies.

Three of the four restriction-modification systems which recognize the CCCGGG sequence are very similar. The genomic organization is identical for the *Xcy*I, *Xma*I and *Cfr*91 genes and the methylases exhibit greater than 80% amino acid sequence identity. The primary structure of the *Cfr*91 endonuclease is also reported to be very similar to the *Xma*I(and *Xcy*I) endonuclease(1). These properties suggest a common ancestor for the *Xcy*I, *Xma*I and *Cfr*91 restriction-modification systems.

A different evolutionary pathway appears to have occurred for the *Sma*I endonuclease and methylase genes. The genes of the *Sma*I restriction-modification system are convergent and there is an additional open reading frame, encoding a proposed regulatory protein(27), immediately upstream of the endonuclease. Furthermore, there is no apparent sequence similarity between the *Sma*I restriction enzyme and the other endonucleases of this family. In contrast, the *Sma*I methylase does exhibit significant similarity to the *Xcy*I and *Cfr*91 methylases. The greatest divergence in the sequence of the *Sma*I methylase occurs at the amino terminus and calculations from which the N-terminal 50 amino acids have been excluded yield an estimate of greater than 65% amino acid identity between the *Sma*I and *Xcy*I methylases and which increases to greater than 75% similarity if conservative substitutions are included. These observations suggest that the *Sma*I endonuclease and methylase genes were recruited independently with the latter derived from the same ancestor genes as the *Xcy*I and *Cfr*91 methylases. Similarly, it has been proposed that the *Rsr*I and *Eco*RI endonucleases have evolved from a common ancestor, while the corresponding methylases were derived from unrelated progenitor genes(20).

The functional domains of the N4 cytosine methyltransferases have yet to be defined. The FXGSG and TSPPY motifs have been assigned possible roles in AdoMet binding and methyl transfer(17) and Klimasauskas *et al.*(17) have further predicted that the recognition domain lies between these two conserved motifs. A DDG tripeptide occurs twice (at positions 92 and 151) within the sequence of the *Xcy*I methylase and in the region encompassing the potential recognition domain. A role for the DDG motif in DNA recognition has been suggested by Card *et al.*(28) who observed the presence of the tripeptide within the target recognition domains of several C5 methyltransferases which recognize the sequence, CCGG. In addition to the presence of the DDG motifs within the sequence of the *Xcy*I methylase, an homologous DDG sequence(at position 92) and, conservative substitutions of this sequence, EDG and NDG(at position 151) occur within the primary structures of the *Cfr*91 and *Sma*I methylases, respectively.(The sequence is not present in the *Mva*I(CCA/TGG) or *Pvu*II(CAGCTG) N4 cytosine methyltransferases). The significance of this observation and the relationship between the binding domains of the methylases which recognize the CCGG and the CCCGGG remains to be determined.

A lack of primary structure similarity has, in the past, been a characteristic of the type II endonucleases. However, recent studies have revealed that some isoschizomers, such as *Eco*RI and *Rsr*I(30), share significant sequence similarities. No similarities have previously been detected between endonucleases which have different catalytic specificities. However, two motifs which have recently been assigned functional roles in the *Eco*RV endonuclease(11) were detected within the primary structure of the *Xcy*I endonuclease. The $PD(X)_{16-20}EXK$ motif has been identified as a magnesium binding domain(12) and the crystal structure of *Eco*RV positions these amino acids close to the scissile bond. The motif is also present in the *Dpn*II, *Hinc*II and *Hpa*I endonucleases, as well as the *Eco*RI and *Rsr*RI endonucleases(11). This motif may represent a structural element utilized by a number of different endonucleases. A similar metal binding domain, however, is not apparent in the sequence of the *Sma*I endonuclease.

The second conserved domain has the consensus sequence, SG---NI-S. The motif present in the *Xcy*I endonuclease contains an additional amino acid between the conserved serine and asparagine. The X-ray crystal structure of the *Eco*RV enzyme suggests that the amino acids occur within a loop(29). If a similar structure exists for the *Xcy*I endonuclease, the insertion of one amino acid may well be accomodated.

Site directed mutagenesis studies have indicated that the SG---NI-S domain is essential for the activity of the *Eco*RV endonuclease(11). The conserved motif was also found by Thielking *et al.*(11) to be present only in the *Dpn*II, *Hinc*II and *Hpa*I endonucleases which have recognition (GW(WW)WC, where W is A or T) and cleavage specificities(blunt end scissions) similar to *Eco*RV. The *Dpn*II methylase also contains the consensus sequence. These observations led to the proposal that the motif may participate in DNA recognition. It is perhaps unexpected that a similar motif was detected in *Xcy*I which not only recognizes a different sequence(CCCGGG) but produces cohesive termini. However, a significant difference between the *Xcy*I and *Eco*RV enzymes is the relative location of the two domains which are immediately adjacent in the *Xcy*I endonuclease but separated by approximately 100 amino acids in *Eco*RV enzyme. The different location of the domains may be related to the different positions at which these enzymes cleave within their recognition sequences. Nonetheless, the identification of the two conserved motifs will facilitate site-directed mutagenesis studies to determine whether there is a comparable role for these amino acids in the structure and function the *Xcy*I endonuclease.

ACKNOWLEDGEMENTS

We thank Keith Lunnen and Geoff Wilson for the clone containing the *Xma*I restriction-modification genes and for providing the sequence of the genes prior to publication. We also thank Dean Gabriel for providing the *Xma*III clone. We are grateful to Bob Bumenthal for helpful suggestions and critical review of the manuscript. We also appreciate the technical asistance of Kim Zoski. This work was supported by grant DMB-900461 from the National Science Foundation and BW was supported by a Rumble Fellowship from Wayne State University.

REFERENCES

1. Wilson, G.G and Murray, N.E. (1991) Annu. Rev. Genet. **25**, 585–627.
2. Kessler, C. and Mantra, V. (1990) Gene **92**, 1–248.
3. Butkus, V., Petrauskiene, L., Maneliene, Z., Klimasauskas, S., Laucys, V. and Janulaitis, A.A. (1985) Nucl. Acids Res. **15**, 7091–7102.
4. Klimasauskas, S., Steponaviciene, D., Maneliene, Z., Petrusyte, M., Butkus, V. and Janulaitis, A. (1990) Nucl. Acids Res. **18**, 6607–6609.
5. Heidemann, S., Siefert, W., Kessler, C. and Domday, H. (1989) Nucl. Acids Res. **17**, 9783–9796.
6. Lunnen, K., Barsomian, J., Camp, R., Card, C., Chen, S., Croft, R., Looney, M., Meda, M., Moran, l., Nwankwo, D., Slatko, B., VanCott, E. and Wilson, G. (1988) Gene **74**, 25–32.
7. Piekarowicz, A., Yuan, R. and Stein, D.C. (1991) Nucl. Acids Res. **19**, 1831–1835.
8. Rodicio, M.R. and Chater, K.F. (1988) Gene **74**, 39–42.

9. Mullings, R., Bennett, S.P. and Brown, N.L. (1988) Gene **74**, 245–251.
10. Dusterhoft, A., Erdmann, D. and Kroger, M. (1991) Nucl. Acids Res. **19**, 3207–3211.
11. Thielking, V., Selent, U,M., Kohler, E., Wolfes, H., Pieper, U., Geiger, R., Urbanke, C., Winkler, F. and Pingoud, A.(1991). Biochemistry **30**, 6416–6422.
12. Selent, U., Ruter, T., Kohler, E., Liedtke, M., Thielking, V., Alves, J., Oelgeschlager, T., Wolfes, H., Peters, F. and Pingoud. A. (1992) Biochemistry **31**, 4804–4815.
13. Bullock, W.O., Fernandez, J.M. and Short, J.M. (1987) Biotechniques **5**, 376–378.
14. Sambrook, J., Fritsh, E.F. and Maniatis, T. (1989). In Molecular Cloning. A Laboratory Manual. 2nd Edition. Cold Spring Harbor University Press. Cold Spring Harbor.
15. Southern, E. M. (1975) J. Mol. Biol **98**, 503–517.
16. Ingrosso, D., Fowler, A., Bleibaum, J. and Clarke, S. (1989) J. Biol. Chem. **264**, 20131–20139.
17. Klimasauskas, S., Timinskas, A., Menkevicius, S., Butkiene, D., Butkus, V. and Janulaitis, A. (1989) Nucl. Acids. Res. **17**, 9823- 9832.
18. Brooks, J., Nathan, P., Landry, D., Sznyter, L., Waite-Rees, Ives, C., Moran, L., Slatko, B. and Benner, J.(1991) Nucl. Acids. Res. **19**, 841–850.
19. Bergey, A. In Krieg, N and Holt, J.(eds.), Bergey's Manual of Systematic Bacteriology. Williams and Wilkins, London. Vol. 1, pp. 409–498. 1984.
20. Kaszubska, W., Aiken, C., O'Connor, C. and Gumport, R. (1989) Nucl. Acids Res. **17**, 10403–10425.
21. Endow, S., and Roberts, R. (1977) J. Mol. Biol. **112**, 521–529
22. Kunkel, L., Silberklang, M and McCarthy, B. (1979) J. Mol. Biol. **132**, 133–139.
23. DeFeyter, R. and Gabriel, D. (1991) J. Bacteriol. **173**, 6421–6427.
24. Froman, B., Tait, R., Kado, C. and Rodriguez, R. (1984) Gene **28**, 331–335.
25. Chandrasegran, S., Lunnen, K., Smith, H. and Wilson, G. (1988) Gene **70**, 387–392.
26. Barany, F., Dantzitz, M., Zebala, J. and Mayer, A. (1992) Gene **112**, 3–12.
27. Tao, T., Bourne, J. and Blumenthal, R. (1991) J. Bacteriol. **173**, 1367–1375.
28. Card, C., Wilson, G., Weule, K., Hasapes, J., Kiss, A. and Roberts, R. (1990) Nucl. Acids. Res. **18**, 1377–1383.
29. Winkler, F. (1992) Curr. Opin. Struct. Biol. **2**, 93–99.
30. Stephenson, F.H., Ballard, B.T., Boyer, H.W., Rosenberg, J.M. and Greene, P.T. (1989) Gene **89**, 1–13.