Nucleic Acids Research, 1992, Vol. 20, No. 22 **6051–6056**

Cloning and sequence analysis of the genes coding for *Eco*57I type IV restriction-modification enzymes

Arvydas Janulaitis*, Romualdas Vaisvila, Albertas Timinskas, Saulius Klimasauskas+ and Viktoras Butkus
Institute of Biotechnology FERMENTAS, Graiciuno 8, 2028 Vilnius, Lithuania

Received July 6, 1992; Revised and Accepted October 26, 1992 EMBL accession no. X61122

ABSTRACT

A 6.3 kb fragment of *E.coli* RFL57 DNA coding for the type IV restriction-modification system *Eco*57I was cloned and expressed in *E.coli* RR1. A 5775 bp region of the cloned fragment was sequenced which contains three open reading frames (ORF). The methylase gene is 1623 bp long, corresponding to a protein of 543 amino acids (62 kDa); the endonuclease gene is 2991 bp in length (997 amino acids, 117 kDa). The two genes are transcribed convergently from different strands with their 3'-ends separated by 69 bp. The third short open reading frame (186 bp, 62 amino acids) has been identified, that precedes and overlaps by 7 nucleotides the ORF encoding the methylase. Comparison of the deduced *Eco*57I endonuclease and methylase amino acid sequences revealed three regions of significant similarity. Two of them resemble the conserved sequence motifs characteristic of the DNA[adenine-N6] methylases. The third one shares similarity with corresponding regions of the *Pae*R7I, *Taq*I, *Cvi*BIII, *Pst*I, *Bam*HI and *Hinc*II methylases. Homologs of this sequence are also found within the sequences of the *Pae*R7I, *Pst*I and *Bam*HI restriction endonucleases. This is the first example of a family of cognate restriction endonucleases and methylases sharing homologous regions. Analysis of the structural relationship suggests that the type IV enzymes represent an intermediate in the evolutionary pathway between the type III and type II enzymes.

INTRODUCTION

Restriction-modification systems are conventionally grouped into three different classes (type I, II and III) on the basis of their subunit structure, cofactor requirements and some other features (1). The *Eco*57I modification methylase (M_r of 63 kDa) and the cognate restriction endonuclease (M_r of 104–108 kDa) were purified to homogeniety and characterized (2). Unexpectedly, it was discovered that a single polypeptide of the endonuclease is a bifunctional enzyme, which can catalyze both DNA cleavage and methylation. No restriction enzyme with such properties has been previously known. Enzymatic properties of the *Eco*57I modification methylase are also different from those of the known type I–III DNA methyltransferases. The entirety of biochemical and structural data suggest that the *Eco*57I restriction-modification system should be regarded as a novel type of RM enzymes—type IV (2, 3, 4).

In this paper we present data on cloning and sequence analysis of the *eco*57IRM genes. The data show the existence of two separate genes coding for a methylase and an endonuclease. Sequence comparison revealed three regions of homology between the *Eco*57I protein molecules. One of the consensus motifs is shared among several other restriction endonucleases and methylases. This has not been shown before for a family of restriction-modification enzymes, and provides evidence that cognate restriction endonucleases and methylases may be related evolutionarily. Analysis of the structural relationship suggests that the type IV enzymes may be an intermediate form in the evolutionary pathway between the type III and type IIS enzymes.

MATERIALS AND METHODS

Bacterial strains and DNAs

The *Escherichia coli* strain RFL57 producing the *Eco*57I endonuclease and methylase was isolated in our laboratory (3). The *E.coli* strains HB101 (5) and RR1 (6) used as hosts in cloning experiments were grown in LB medium. Transformants were selected by plating onto LB-agar supplemented with 50 mg/l Ampicillin (Ap) or 30 mg/l Chloramphenicol (Cm). The pJRD184 (7), pACYC184 (8), pUC19 (9) vectors were used in cloning and subcloning experiments. Plasmids carrying the *eco*57I genes were named according to the previously described nomenclature (10). Phage stocks of λvir were prepared according to Maniatis et al. (11). Plasmid DNA was prepared by the alkaline lysis method (12) and further purified by binding to glasspowder (13). *E.coli* RFL57 chromosomal DNA was extracted and purified by the method of Marmur (14).

* To whom correspondence should be addressed

+ Present address: Cold Spring Harbor Laboratory, PO Box 100, Cold Spring Harbor, NY 11724, USA

Enzymes and chemicals

Restriction enzymes (except for *Spe*I), nuclease *Bal*31, T4 DNA ligase, DNA polymerase I large fragment (Klenow), T4 polynucleotide kinase and terminal deoxynucleotidyl transferase were products of MBI Fermentas. The *Spe*I restriction endonuclease was from New England Biolabs. Nuclease S1, DNA Sequencing Kit and calf intestinal alkaline phosphatase were purchased from Pharmacia. DNAseI was obtained from Serva. All enzymes were used according to the manufacturers' recommendations. Synthetic oligonucleotide primers were prepared on 'GENE ASSEMBLER' using methylphosphoramidite chemistry. [α-^{33}P]dATP was obtained from Izotop, Leningrad. All other chemicals were reagent grade commercial products.

Assay for *Eco*57I activities

The activity of the *Eco*57I restriction endonuclease in cell extracts was assayed as described previously (2). The restriction activity *in vivo* was analyzed by measuring the plating efficiency of the λvir bacteriophage in the transformed cells as compared to the control, *E.coli* RR1 cells harboring the pJRD184 vector plasmid (11). To determine the *Eco*57I specific modification *in vivo* plasmid or chromosomal DNA was isolated and challenged with an excess of the *Eco*57I restriction endonuclease *in vitro*.

DNA cloning

Transformation of *E.coli*, restriction mapping, agarose gel electrophoresis, *Bal*31 deletions and subcloning of DNA fragments for sequencing were carried out by standard procedures (11). Deletion mapping of the clones carrying the *eco*57I methylase and endonuclease genes was performed using DNaseI and nuclease S1 as described previously (15).

Construction and selection of the *E.coli* RFL57 libraries

pJRD184 (10 μg) DNA was linearized by *Bam*HI cleavage. Homopolymer (dG) extensions were added to the *Bam*HI-linearised pJRD184 by incubation in a reaction mixture containing 5 μg DNA, dGTP and 60 units of terminal transferase for 10 min at 37°C as described in (16). The poly(dG) extensions were estimated in preliminary experiments with [^{3}H]dGTP to average 20 nucleotides in length. *E.coli* RFL57 chromosomal DNA (50 μg) was fragmented by sonication. The sonication products were then fractionated on a low-melting agarose gel and 3–20 kb fragments were isolated. Poly(dC) extensions were added to the *E.coli* RFL57 DNA fragments in a similar fashion. Annealing of the extended DNAs was carried out in a 3 ml volume under standard conditions (16). The clone library was obtained by transforming competent *E.coli* RR1 cells with the annealed DNA and selecting for Ap resistant transformants. Selection on the basis of the acquired resistance to bacteriophage infection was carried out by replicating the transformants onto top-layer agar containing 10^6 phage λvir particles per plate (11).

DNA sequencing

The nucleotide sequence was determined by the dideoxynucleotide chain-termination procedure (17) using double-stranded, supercoiled plasmid DNA as templates. Fragments generated by serial *Bal*31 deletions were subcloned into a pUC19 vector and were sequenced using Sequencing Kit, [α-^{33}P]dATP and synthetic primers (complementary to the plasmid region upstream from the cloned fragments). Some regions of the cloned fragments were directly sequenced using specific primers. A small portion of the DNA sequencing was carried out using the chemical cleavage method of Maxam and Gilbert (18). The reaction products were resolved by electrophoresis on wedge-shaped gels.

Sequence comparisons

Published sequences of the following DNA-methylases: *Hha*II, *Hinf*I, *Dpn*A, *Dpn*B, *Eco*P1, *Eco*P15, T4 *dam*, *Eco dam*, *Eco*RV, *Eco*KI, *Taq*I, *Pae*R7I, *Cvi*BIII, *Pst*I, *Eco*RI, *Hha*I, *Eco*RII, *Hpa*II, *Dde*I, *Sin*I, *Phi*3T, SPR, *Bsu*RI and *Hae*III and restriction endonucleases: *Bsu*RI, *Dde*I, *Dpn*I, *Dpn*II, *Eco*PI, *Eco*RI, *Eco*RV, *Hha*II, *Pae*R7I, *Pst*I, *Sin*I and *Taq*I were used in sequence comparisons. The references for these sequences are contained in ref. 19. The sequences of other methylases used included: *Cfr*9I (20), *Mva*I (20), *Sma*I (21), *Pvu*II (22), *Bam*HI (23), *Bam*HII (24), *Eca*I (25), *Rsr*I (26), *Fok*I (27), *Hinc*II (28), *Dcm* (29), *Msp*I (30), *Bsu*FI (31), *Ban*I (32), *Bep*I (33), *Ngo*PII (34), *Aqu*I (35), *Cvi*JI (36), *Sau*3AI (37), *Sss*I (38) and *Sau*96I (39).

The sequences for restriction endonucleases used for comparison included: *Bam*HI (23), *Bsu*FI (40), *Eco*RII (41), *Fok*I (27), *Hinc*II (28), *Msp*I (30), *Ngo*PII (34), *Pvu*II (42), *Rsr*I (43), *Sau*3AI (37), *Sau*96I (39) and *Sma*I (21). The sequences for the restriction endonucleases and methylases *Bcn*I, *Eco*72I, *Mun*I, the methylases *Bsp*6I, *Cfr*10I and the restriction endonucleases *Cfr*9I and *Mva*I were determined in our laboratory (unpublished data). The sequences for the *Hgi*BI, *Hgi*CI, *Hgi*CII, *Hgi*EI, *Hgi*DI and *Hgi*GI restriction endonucleases and methylases were kindly provided by M. Kröger.

The procedure developed in our laboratory (20) was used to manage the sequence data.

Analysis of N-terminal oligopeptides

Sequencing of the amino termini of the restriction endonuclease and methylase was performed by sequential Edman degradation (44). The phenylthiohydantoin derivatives were identified by HPLC on Cyanopropyl Column.

RESULTS AND DISCUSSION

Cloning and localization of the *Eco*57I restriction-modification genes

To clone the *Eco*57I restriction-modification system a clone library of the host DNA consisting of 30,000 transformants was prepared in *E.coli* strain RR1. The transformants carrying the *eco*57IR gene were selected on the basis of their restriction phenotype—the acquired resistance to bacteriophage infection. A single clone out of 3,000 tested survived the infection by phage λvir *in vivo*. The plating efficiency of the unmodified phage was 10^{-5} fold lower in this clone than in control cells carrying the vector plasmid. A recombinant plasmid containing a 6.3 kb (p*Eco*57RM6.3) insert was recovered from this clone. Chromosomal and plasmid DNAs isolated from the clone were resistant when challenged with the *Eco*57I restriction endonuclease *in vitro* (data not shown) indicating that they were fully modified *in vivo*. A crude cell extract prepared from the cells carrying the recombinant plasmid was assayed for the R.*Eco*57I activity. The tests were positive, revealing a fragmentation pattern characteristic of R.*Eco*57I (data not shown). The results indicate that the recombinant plasmid carries both *eco*57IM and *eco*57IR.

To determine the location of the individual genes, the cloned fragment was subjected to deletion mapping. The functional

methylase in the transformants was assayed by recovering recombinant plasmids carrying different deletions and challenging them *in vitro* with R.*Eco*57I. The activity of the restriction endonuclease of the deletion mutants was tested *in vivo* by evaluating λvir restriction in the transformed cells and *in vitro* by its ability to produce characteristic DNA digestion patterns. The results of the mapping experiments are summarized in Fig. 1A. Plasmid p*Eco*57RM6.1, the smallest in the series, conferred the restriction and modification phenotypes. The shortest derivative still showing a modification phenotype was p*Eco*57M3.2, while p*Eco*572.9 conferred neither phenotype.

As demonstrated earlier, the *Eco*57I restriction endonuclease possesses both endonuclease and methylase activities *in vitro* (2). Its ability to protect chromosomal DNA *in vivo* was tested. A major part of the methylase gene was deleted by excising the *Mun*I-*Spe*I fragment from p*Eco*57RM6.3 (Fig. 1A). The resulting p*Eco*57R4.2 was used to transform *E.coli* HB101(pACYC184) or *E.coli* HB101(p*Eco*57M3.3). The latter plasmid is a derivative of pACYC184 carrying the *eco*57IM gene on the *Eco*RV-*Pvu*II 3.3 kb fragment (Fig. 1B). Equal amounts of p*Eco*57R4.2 were used to transform both types of recipient cells, although the yield of Ap resistant transformants was 100 fold lower in *E.coli* HB101(pACYC184) as compared to *E.coli* HB101(p*Eco*57-M3.3). Both types of clones were assayed for restriction activity and their plasmid structure was investigated. *E.coli* HB101(p*Eco*57M3.3) transformants revealed R^+M^+ phenotype indicating that the *eco*57IR gene was present in its entirety. Two types of plasmids were recovered from the transformants. One of these was identical to p*Eco*57M3.3, while the other was indistinguishable from p*Eco*57R4.2. The latter plasmid, however, could not be detected in any of the several clones obtained after transformation of *E.coli* HB101(pACYC184). Only intact pACYC184 and various derivatives of p*Eco*57R4.2 were observed. None of these clones showed the R^+ phenotype. This indicates that *eco*57IR can not function in M^- cells suggesting that probably the endonucleolytic activity of R.*Eco*57I prevails over its modification activity *in vivo* resulting in the destruction of the cellular DNA.

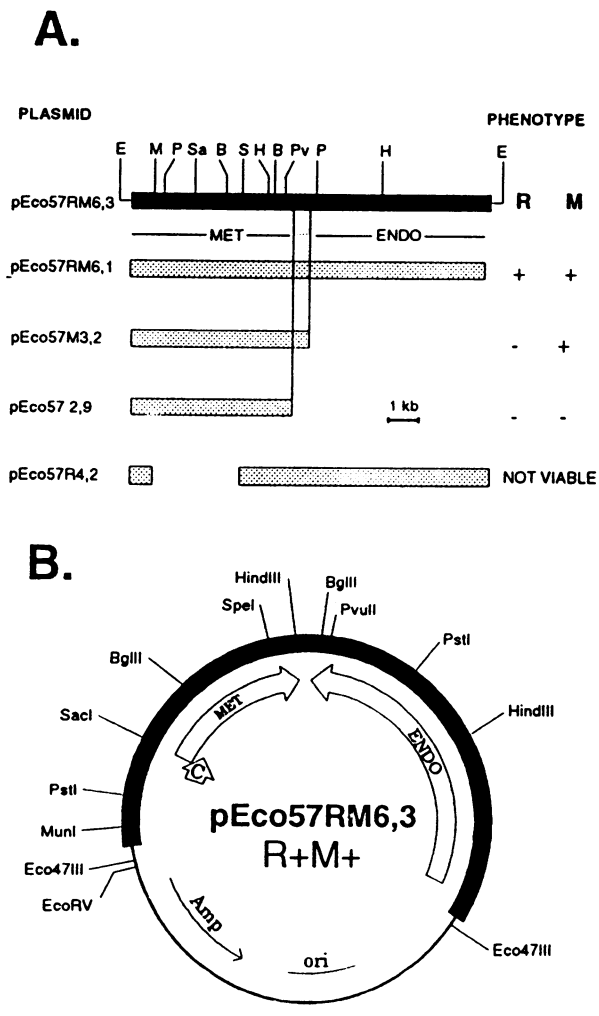

Figure 1. Schematic map of the DNA region coding for the *Eco*57I enzymes. **A)** Deletion mapping of the *eco*57I genes. Restriction endonuclease sites are as follows: E-*Eco*47III, M-*Mun*I, P-*Pst*I, Sa-*Sac*I, B-*Bgl*II, S-*Spe*I, H-*Hind*III, Pv-*Pvu*II. The deletion plasmids are indicated by dotted boxes. The inferred locations of the endonuclease and the methylase genes are indicated by thin lines and dots. **B)** Circular map of p*Eco*57IRM6.3. Vector sequences are shown as thin lines, the genes are shown as open arrows (C—the SORF (*eco*57IC), MET—*eco*57IM, ENDO—*eco*57IR).

Gene organization and DNA sequence analysis

A 5775 base pair region of the cloned 6.3 kb DNA fragment encompassing the *Eco*57I restriction-modification system was sequenced in both strands. Scanning the sequence for open reading frames revealed that two of them have a potential to code for proteins of the expected molecular weight.

The two open reading frames were in perfect agreement with the results of the deletion mapping experiment (Fig. 1). The first ORF codes for the methylase, while the second one codes for the restriction endonuclease. The strand assignment suggests that the two genes are transcribed from different strands convergently, with their tails separated by 69 base pairs. The *eco*57IM sequence specifies a protein of 544 residues with molecular weight of 62,020 Da (calculated from amino acid composition). The predicted molecular weight agrees well with the value (63 kDa) determined by SDS-polyacrylamide gel electrophoresis (2). The second ORF codes for a protein of 997 amino acids with the expected molecular weight of 116,724 Da. This number is just slightly larger than that of the purified restriction endonuclease (104−108 kDa) as estimated by SDS-polyacrylamide electrophoresis or by gel exclusion chromatography (2).

The assignment of the open reading frames and initiation codons was further confirmed by amino terminal sequence analysis of the purified methylase and endonuclease proteins (data not shown). In both cases the experimentally determined N-pentapeptides matched those deduced from the nucleotide sequence.

The methylase gene is preceded by a short open reading frame (SORF) that extends over 186 nucleotides, and is capable of coding for a peptide of 7,282 Da (Fig. 1B). The SORF and *eco*57IM overlap by seven nucleotides and are staggered by one frame.

The ATG codons of the three ORFs are immediately preceded by sequences showing complementarity to the 3'-end of *E.coli* 16S RNA. However, promoter-like sequences (67−80% similarity to the consensus) can be found only upstream from the sequences encoding SORF and the endonuclease and no such sequence appears upstream of the methylase gene itself. Most likely *eco*57M is transcribed from the promoter located upstream of the SORF. Due to the overlap of the coding regions its translation from this polycistronic messenger may be coupled with

the translation of the preceding SORF (45). The occurrence of SORFs tightly linked with one of the major genes has been demonstrated for some other type II RM systems, and in two cases (*Pvu*II and *Bam*HI) their regulatory role was proved (46, 47). Bearing in mind these precedents, we believe that the SORF is associated with the RM.*Eco*57I system and probably plays a regulatory role.

The average A+T content of the *Eco*57I genes is 65.1% (67.3% for *eco*57IR, 63.2% for *eco*57IM and 65.1% for the SORF), which differs considerably from that of *E.coli*—49%. The *Eco*RI and *Eco*RV RM genes (65% A+T) also exceed the *E.coli* average (48, 49). It has been suggested that these genes may have originated from species with a higher A+T content in their DNA (49). Alternatively, that may reflect an inherent feature of particular RM systems.

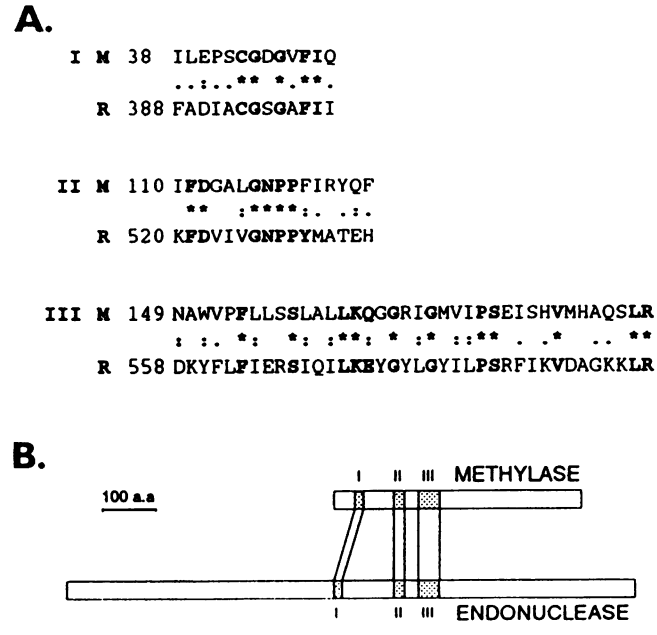

Figure 2. Regions of similarity between the *Eco*57I methylase (M) and restriction endonuclease (R). **A)** Sequence alignments within conserved blocks **I, II** and **III**. The numbers at the left indicate the position of the first amino acid in the alignments. '*' indicates identity, ':' similarity, '.' compatibility between amino acid residues. **B)** Diagram of conserved blocks I, II and III (connected by vertical lines) within the RM.*Eco*57I sequences (open boxes).

Comparison of amino acid sequences

The amino acid sequences of the RM.*Eco*57I proteins have been examined for intermolecular homology. Comparison revealed homology between R.*Eco*57I and the cognate methylase, limited to three conserved blocks I, II and III (Fig. 2). The conserved regions I and II resemble the two motifs which are common to all adenine and cytosine methylases analyzed to date (20) and are presumably involved in general steps of DNA methylation catalysis (AdoMet binding and methyl group transfer). This is not surprising given that the *Eco*57I restriction endonuclease is a bifunctional enzyme catalyzing both DNA cleavage and methylation (2).

We searched all available sequences of DNA-methylases for regions homologous with the conserved motif III. We found that this motif is present within the methylases *Pae*R7I, *Cvi*BIII, *Taq*I, *Pst*I, *Hinc*II and *Bam*HI (Fig. 3). In each case the conserved block III occurs within 12−27 amino acids downstream of the conserved motif II. In the case of M.*Eco*57I and R.*Eco*57I the interval is practically identical (24 a.a. and 23 a.a.). It should be noted that, as we found recently, this block partially overlaps with the pattern of sequence similarity previously described for five adenine methylases mentioned above: *Pae*R71, *Cvi*BIII, *Taq*I, *Pst*I and *Hinc*II (28, 50,).

The consensus motif (Fig. 3) was used as a search pattern against the database of the available restriction endonuclease sequences. At a slightly lower stringency R.*Pae*R7I, R.*Pst*I and R.*Bam*HI showed homology to the consensus pattern. No such motif has been found in R.*Taq*I, and R.*Hinc*II. Pairwise alignments of RM sequences in which patterns resembling the consensus motif has been found is shown in Fig. 4. Only in the cases of R.*Bam*HI and R.*Eco*57I is the N-terminus of the motif $(N,D)(X)_1(Y,W)(X)_2(F,Y,W)$ strictly conserved. Nonconservative substitutions in this part of the motif are within R.*Pae*R7I and R.*Pst*I. The same is true for R.*Bam*HI where an aliphatic uncharged amino acid located downstream of the conserved Y is substituted by R. For all enzymes under consideration the interval between the first (N,D) amino acid of the conserved motif and K is the same. Also remarkable is the conservation of the G residue, which is three amino acids downstream of the K

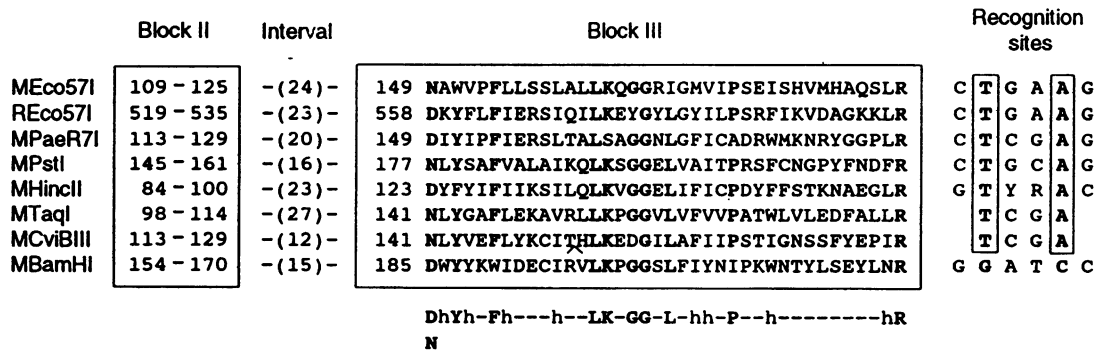

Figure 3. Schematic diagram and alignment of the methylase sequences within conserved Block III. The numbers in the box for Block II indicate the corresponding amino acid positions in each of the sequences, numbers in brackets indicate intervals between Blocks II and III. The numbers on the left in the box for Block III indicate the position of the first amino acid in the alignment: '.' stands for the E residue excluded from the alignment in the M.*Cvi*BIII sequence. A symbol of a single amino acid appears in the consensus if the residue is conserved in at least 6 sequences in the alignment; two symbols appear if two functionally similar amino acids, (e.g. **N** and **D**) are conserved; 'h' if a minimum 7 out of 8 residues are hydrophobic and '−' stands for poorly conserved positions. T and A residues common in the recognition sequences of the adenine methylases are boxed.

residue in R.*Pae*R7I and R.*Eco*57I, as well. This does not hold for R.*Pst*I and R.*Bam*HI. The conserved R (22 amino acid downstream of invariant K) in the restriction endonucleases is also retained.

Structural similarity of proteins suggest that they may be functionally related. Although restriction endonucleases and methylases catalyze very different reactions, they share some common functions such as DNA binding and specific target recognition. A strong argument supporting the involvement of block III in sequence recognition is its occurrence within a family of RM enzymes, that recognize related sequences. All these enzymes share the TNNA element as a subset of their recognition sequences (Fig. 3) except for RM.*Bam*HI which target C at the same relative position of the recognition sequence. Recent analysis of R.*Bam*HI mutants with reduced cleavage and altered DNA binding properties has shown that such mutations map within a region (residues 77−135) (51) that almost completely covers the conserved block III (residues 70−108).

An abundance of conserved motifs was found within m^5C methylases (52, 53). The experimental evidence indicates that the variable region is responsible for sequence specificity in these enzymes (54, 55), while some of the conserved motifs most likely define structural elements important for proper folding (53). It could not be excluded that block III fulfills the same function.

The proteins encoded by controlling elements (SORF) in the *Pvu*II, *Bam*HI, *Eco*RV and *Sma*I RM systems share homologous region with each other and some DNA-binding proteins containing the helix-turn-helix motif (46). No clear similarity can be detected between them and the deduced amino acid sequence of the putative polypeptide encoded by the *Eco*57I SORF.

EVOLUTIONARY IMPLICATIONS

One of the most intriguing questions is the evolution of RM enzymes, where the restriction endonuclease and the cognate methylase usually recognize the same DNA sequence. It is straightforward to expect that such enzymes would share at least some elements of the target recognition machinery. Previous attempts to find evidence for such inheritance by pairwise comparison of corresponding restriction enzymes and methyltransferases were unsuccessful. Moreover, common themes of sequence recognition so far have not been found among restriction endonucleases themselves (56). One possibility is that each component evolved independently from the other and later combined to form RM systems. Alternatively, due to strong divergence the original domains are barely detectable at the primary structure level. The discovery of conserved patterns within the restriction endonucleases and methylases RM.*Eco*57I, RM.*Pst*I, RM.*Pae*R7I and RM.*Bam*HI indicates that at least in some cases, restriction endonucleases are evolutionarily related to methylases.

```
MEco57I  149 NAWVPFLLSSLALLKQGGRIGMVIPSEISHVMHAQSLR
             : :. *: .*: :**: * :*.::**.. .*  ...**
REco57I  558 DKYFLFIERSIQILKEYGYLGYILPSRFIKVDAGKKLR

MPaer7I  149 DIYIPFIERSLTALSAGGNLGFICADRWMKNRYGGPLR
             . . .*:   *..:.*.* . .*  :*.* . .*  .*
RPaer7I   48 KNMDGFLALVLDVIKANGLAHAIHQNRAMLTLPGY-FR
                                     E

MPstI    177 NLYSAFVALAIKQLKSGGELVAITPRSFCNGPYFNDFR
             *  *.:*.**: ::*. ... *..*  ........ .*
RPstI     24 NDRSGWVLLALANIKPEDSWKAA-P-LLPTVSIMEFIR
                                  EK

MBamHI   185 DWYYKWIDECIRVLKPGGSLFIYNIPKWNTYLSEYLNR
             * *  : :..:.:** .  .:::.*    **   *
RBamHI    70 DTYNWYREKPLDILKLEKKGGPIDVYKEFIENSE-LRR
                                   K
             D-Y--Fh---h--LK--G-h-hh----h--------hR
             N
```

Figure 4. Pairwise alignment and the consensus for the cognate restriction endonucleases and methylases. A symbol of a single amino acid appears in the consensus if the residue is conserved in at least 5 sequences in the alignment; two symbols appear if two functionally similar amino acids, (e.g. N and D) are conserved in at least 7 sequences out of 8 aligned; 'h' if a minimum 6 out of 8 residues are hydrophobic and '−' stands for poorly conserved positions. Residues excluded from the alignments are shown below the sequences.

The type III RM systems consist of two subunits : **mod**, which is a functional DNA-methylase, able to recognize and modify the target sequences, and **res**, which is responsible for the restriction phenotype, but functions only in a complex with the **mod** subunit (57). Therefore, the type III restriction endonuclease is a heterodimer which uses the **mod** subunit for sequence recognition, and the **res** subunit−for DNA cleavage. The cleavage always occurs in the vicinity of rather than within the asymmetric target sequence, so that both subunits can be accomodated.

The type II RM systems are composed of two structurally independent enzymes responsible for either modification (methylase) or cleavage (restriction endonuclease) of the specific DNA sequence. This is also true for a subclass of the type II, the type IIS enzymes, that recognize asymmetric DNA sequences and cleave both strands at a defined distance from it.

The properties of R.*Eco*57I (2) indicate that it may be regarded as an intermediate type of RM reflecting the evolutionary link between the enzymes of types III and type IIS. This hypothesis is based on the following arguments. R.*Eco*57I possessing both DNA methylation and cleavage functions very closely resembles a hypothetical **mod-res** fusion protein (see Fig. 2B). It recognizes an asymmetric sequence and cleaves at 16/14 nucleotides away from it. Its stimulation by AdoMet resembles that of enzymes of type III. Interestingly, R.*Eco*57I is able to methylate only one of the two strands of the target sequence, which might be due to steric restrictions imposed by the asymmetric location of the cleavage site. This methylation most likely is a relic of the progenitor activity since it is not by itself sufficient to protect the host DNA *in vivo* and therefore an additional copy of the 'free' methylase is needed.

The *Gsu*I restriction endonuclease, which recognizes 5'CTGGAG sequence and cleaves DNA 14/16 nucleotides away from it (3,4) is another example of a plausible evolutionary linking member. Like R.*Eco*57I it is stimulated by AdoMet but differs from it by the absence of methylation activity and by its ability to perform DNA cleavage to completion. Consequently, R.*Gsu*I has greater similarity to the type II enzymes than has R.*Eco*57I. It may be speculated that the R.*Gsu*I is a further intermediate in the evolutionary pathway towards type IIS enzymes.

ACKNOWLEDGEMENTS

The authors wish to thank Dr. M.Kröger for making the sequences for the *Hgi* enzymes available to us prior to publication and Dr. R.J.Roberts for helpful comments on the manuscript.

We also thank A.Neniskis for N-terminal amino acid sequence analysis and R.Bityte and T.Tarvainis for their help in manuscript preparation.

REFERENCES

1. Yuan, R. (1981) *Ann. Rev. Biochem.*, **50**, 285–315.
2. Janulaitis, A., Petrusyte, M., Maneliene, Z., Klimasauskas, S., Butkus, V. (1992) *Nucl. Acids Res.*, **20**, 6043–6049.
3. Petrusyte, M.P., Bitinaite, J.B., Kersulyte, D.R., Menkevicius, S.J., Butkus, V.V., Janulaitis, A.A. (1987) *Dokl. Akad. Nauk USSR*, **295**, 1250–1253.
4. Petrusyte, M., Bitinaite, J., Menkevicius, S., Klimasauskas, S., Butkus, V., Janulaitis, A. A. (1988) *Gene*, **72**, 89–91.
5. Boyer, H.W. and Roulland-Dussoix, D. (1969) *J. Mol. Biol.*, **41**, 459–472.
6. Bolivar, F., Rodriquez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L., Boyer, H.W., Crosa, J.H. and Falkow, S. (1977) *Gene*, **2**, 95–113.
7. Heusterspreute, M., Ha Thi, V., Emery, S., Tournis-Gamble, S., Kennedy, N. and Davidson, J. (1982) *Gene*, **39**, 299–304.
8. Chang, A. and Cohen, S. (1978) *J. Bacteriol.*, **134**, 1141–1156.
9. Viera, J. and Messing, J. (1982), *Gene*, **19**, 259–268.
10. Gingeras, T.R. and Brooks, J.E. (1983) *Proc. Natl. Acad. Sci.* USA, **80**, 402–406.
11. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
12. Birnboim, H.C. and Doly, J. (1979) *J. Bacteriol.*, **110**, 1513–1523.
13. Marko, M.A., Chipperfield, R., Birnboim, H.C. (1982) *Ann. Biochem.*, **121**, 382–387.
14. Marmur, J.A. (1961) *J. Mol. Biol.*, **3**, 208–218.
15. Povilionis, P.J., Lubys, A.A., Janulaitis, A.A. (1988) *Genetika* (Moscow), **24**, 210–215.
16. Mann, M.B., Rao, R.N. and Smith, H.O. (1978) *Gene*, **3**, 97–112.
17. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci.* USA, **74**, 5463–5467.
18. Maxam, A., Gilbert, W.(1977) *Proc. Natl. Acad. Sci. USA*, **74**, 560–564.
19. Wilson, G.G. (1988) *Gene*, **74**, 281–289.
20. Klimasauskas, S., Timinskas, A., Menkevicius, S., Butkiene, D., Butkus, V. and Janulaitis, A. (1989) *Nucl. Acids Res.*, **17**, 9823–9832.
21. Heidmann, S., Seifert, W., Kessler, C. and Domdey, H. (1989) *Nucl. Acids Res.*, **17**, 9783–9796.
22. Tao, T., Walter, J., Brennan, K.J., Cotterman, M.M. and Blumenthal, R.M. (1989) *Nucl. Acids Res.*, **17**, 4161–4175.
23. Brooks, J.E., Nathan, P.D., Landry, D., Sznyter, L.A., Waite-Rees, P., Ives, C.L., Moran, L.S., Slatko, B.E. and Benner, J.S. (1991) *Nucl. Acids Res.*, **19**, 841–850.
24. Connaughton, J.F., Kaloss, W.D., Vanek, P.G., Nardone, G.A. and Chirikjian, J.G. (1990) *Nucl. Acids Res.*, **18**, 4002.
25. Brenner, B., Venetianer, P. and Kiss, A. (1989) *Nucl. Acids Res.*, **18**, 355–359.
26. Kaszubska, W., Aiken, C., O'Connor, C.D. and Gumport, R.I. (1989) *Nucl. Acids Res.*, **17**, 10403–10425.
27. Kita, K., Kotani, H., Sugisaki, H. and Takanami, M. (1989) *J. Biol. Chem.*, 264, 5751–5756.
28. Ito, H., Sadaoka, A., Kotani, H., Hiroaka, N. and Nakamura, T. (1990) *Nucl. Acids Res.*, **18**, 3903–3911.
29. Hanck, T., Gerwin, N. and Fritz, H.J. (1989) *Nucl. Acids Res.*, **17**, 5844.
30. Lin, P.M., Lee, C.H. and Roberts, R.J. (1989) *Nucl. Acids Res.*, **17**, 3001–3011.
31. Walter, J., Noyer-Weidner, M. and Trautner, T.A. (1990) *EMBO J.*, **9**, 1007–1013.
32. Maekawa, Y., Yasukawa, H. and Kawakami, B. (1990) *J. Biochem.*, **107**, 645–649.
33. Kupper, D., Zhou, J.G., Venetianer, P. and Kiss, A. (1989) *Nucl. Acids Res.*, **17**, 1077–1088.
34. Sallivan, K.M., and Saunders, J.R. (1988) *Nucl. Acids Res.* **16**, 4369–4387.
35. Karreman, C. and Waard, A. (1990) *J. Bacteriol.*, **172**, 266–272.
36. Shields, S.L., Burbank, D.E., Grabherr, R. and van Etten, J.L. (1990) *Virology*, **176**, 16–24.
37. Sceber, S., Kessler, C. and Gotz, F. (1990) *Gene*, **94**, 37–43.
38. Renbaum, P., Abrahamove, D., Fainsod, A., Wilson, G.G., Rottem, S. and Razin, A. (1990) *Nucl. Acids Res.*, **18**, 1145–1152.
39. Szilak, L., Venetianer, P. and Kiss, A. (1990) *Nucl. Acids Res.*, **18**, 4659–4664.
40. Walter, J., Noyer-Weidner, M.and Trautner, T.A. (1990) *EMBO J.*, **9**, 1007–1013.
41. Bhagwat, A.S., Johnson, B., Weule, K. and Roberts, R.J. (1990) *J. Biol. Chem.*, **265**, 767–773.
42. Athanasiadis, A., Gregoriu, M., Thanos, D., Kokkinidis, M. and Papametheakis, J. (1990) *Nucl. Acids Res.*, **18**, 6434.
43. Stephenson, F.H., Ballard, B.T., Boyer, H.W., Rosenberg, J.M. and Greene, P.J. (1989) *Gene*, **85**, 1–13
44. Tarr, G.E. (1986) *Methods of Protein microcharacterization* (Shively, J.E. ed), pp.185–187, Humana, Clifton NJ.
45. Spanjaard, R.A., van Duin, J. (1989) *Nucl. Acids Res.* **17**, 5501–5507.
46. Tao, T., Bourne, J.C. and Blumenthal, R.M. (1990) *J.Bacteriol.*, **173**, 1367–1375.
47. Nathan, P.D. and Brooks, J.E. (1988) *Gene*, **74**, 35–36.
48. Bougueleret, L., Schwarzstein, M., Tsugita, A. and Zabeau, M. (1984) *Nucl. Acids Res.*, **12**, 3659–3677.
49. Greene, P.J., Gupta, M. and Boyer, H.W. (1981) *J. Biol. Chem.*, **256**, 2143–2153.
50. Mullings, R., Bennett, S.P, Brown, N.L. (1988) *Gene*, **74**, 245–251.
51. Xu, S.-Y. and Schildkraut, I.(1991) *J. Biol. Chem.*, **299**, 4425–4429.
52. Lauster, R., Trautner, T.A., Noyer-Weidner, M. (1989) *J. Mol. Biol.*, **206**, 305–312.
53. Posfai, J., Bhagwat, A.S., Posfai, G., Roberts, R.J. (1989) *Nucl. Acids Res.*, **17**, 2421–2435.
54. Wilke, K., Rauhut, E., Noyer-Weidner, M., Lauster, R., Pawlek, B., Behrens, B. and Trautner, T.A. (1988) *EMBO J.*, **7**, 2601–2609.
55. Klimasauskas, S., Nelson, J. and Roberts, R.J. (1991) *Nucl. Acids Res.*, **19**, 6183–6190.
56. Wilson, G.G. and Murray, N.E. (1991) *Annu. Rev. Genet.*, **25**, 585–627.
57. Rao, D.N., Page, M.G.P., Bickle, T.A. (1989) *J. Mol. Biol.*, **209**, 599–606.