Volume 12 Number 8 1984 Nucleic Acids Research

Characterization of the genes coding for the *Eco* RV restriction and modification system of *Escherichia coli*

L.Bougueleret, M.Schwarzstein*, A.Tsugita and M.Zabeau+

EMBL, Meyerhofstrasse 1, Postfach 102209, 6900 Heidelberg, FRG

Received 7 February 1984; Revised and Accepted 26 March 1984

ABSTRACT

A plasmid encoding the recently described Eco RV restriction and modification system has been isolated and characterized. This plasmid, pLB1, is 6.2 kb long and carries only the Eco RV genes. A subclone of 3 kb has been inserted in pBR322. The relative positions of the endonuclease and the methylase genes were determined by the construction of a set of overlapping deletions generated by Bal31 resection. The DNA sequence of a 2.2 kb fragment containing the two genes was determined. The two genes are transcribed divergently from a 310 bp region and the assignment of the coding region has been confirmed by direct aminoacid sequence analysis. Possible mechanisms of regulation of the endonuclease gene expression at the translational level are proposed and discussed.

INTRODUCTION

Interest in the type II restriction enzymes - apart from their broad use in genetic engineering - has increased lately, and there have been several reports of the cloning of such systems in Escherichia coli (1-6). Apart from the obvious advantage of overproducing these enzymes, their cloning in Escherichia coli provides the means to study more fundamental questions. First the potential lethality of these systems brings up the problem of their establishment and maintenance. It is known that plasmids coding for restriction and modification systems can be transformed in bacterial hosts deprived of any protecting methylase. This would imply at least some control - of the expression of the endonuclease. Until now no such control mechanism has been identified. The second most interesting feature of these enzymes is their ability to recognize and interact - either by cutting or methylating - with specific DNA sequences generally 4 to 6 bp long. They can thus

provide an interesting model for studying protein-DNA interactions. Until now only the Eco RI endonuclease has been crystallized (7), and little is known about the restriction endonucleases structure. It would then be of great interest to have available different restriction systems to allow their comparison at the structural level and determine whether, as in the case of other DNA binding proteins (8-9), some general features can be found for restriction enzymes.

The recent description of the Eco RV restriction and modification system opened a new possible field of investigation. The Eco RV endonuclease, like the Eco RI endonuclease recognizes a 6 bp long target. The sites of action of those two enzymes are similar - GATATC and GAATTC respectively, - but while Eco RI generates 5' overhanging cohesive ends, Eco RV generates blunt ends (R. Brown, L. Bougueleret; manuscript in preparation). It was therefore interesting to investigate the relationship between these two systems. We report here the cloning, the characterization and the DNA sequence of the genes encoding the Eco RV restriction and modification system.

MATERIAL AND METHODS

Strains:

Strains J62(pGL74) was a gift from Noreen Murray. Strain K514 (C600, r_K^- m_K^+) was used as an indicator strain for λ tests and as a recipient strain for transformations. The cells were grown in L. Broth or on L. Agar. When necessary ampicillin was added to the medium at a final concentration of 100 μg/ml. Phage tests were done on BBL Agar plates supplemented with MgSO4 0.1 M final concentration.

Restriction and Modification Tests:

Phage λ vir was used. Restriction spot test: dilutions of a non methylated λ vir.0 stock (10^{10} pfu/ml) were spotted on a lawn of the cells to be tested, as well as on an indicator non restricting strain (K514). The efficiencies of plating were compared and the ratio eop(K514)/eop(cell) gives the efficiency of restriction. Methylation test: 0.1 ml of an appropriate dilution of λ vir (≃ 10^5 to 10^9 pfu/ml for non restricting and

restricting clones respectively) were mixed with 0.1 ml of freshly grown cells. The mixture was incubated 15' at 37°C without agitation to allow adsorbtion of phage λ. 3ml of BBL top layer Agar containing MgSO4 was then added to the mixture and poured on a BBL Agar plate. Because of the extremely high restriction efficiency of the Eco RV system, the methylation could be tested by simply picking single plaques on a non restricting strain and on the appropriate restricting strain. The methylated phages are giving plaques on both plates, whereas the non methylated phages are unable to grow on the restricting strain. This allows the discrimination between methylating and non methylating clones. An average of 10 single plaques are tested for each clone.

λ Selection:

K514 competent cells were transformed with total DNA extracted from strain J62(pGL74) according to the standard procedure. The mixture was then diluted in fresh medium and infected with λ vir at m.o.i. of 1. Aliquots were plated on non-selective plates after 1, 2 ,3 and 4 hours. The surviving clones were assumed to be either λ resistant mutants or restricting clones and were tested for restricting properties.

Ligation and Transformation:

Ligation reactions were performed at 15°C in 50 mM tris pH 7.9, 50 mM NaCl, 10 mM MgCl2, 5 mM dithiotreitol, 0.5 mM ATP. T4 DNA ligase was a gift from E. Remaut. Prior to transformation the DNAs were phenol and chloroform extracted, and isopropanol precipitated. Transformations were done as described by Dagert and Ehrlich (1979) (11).

Plasmid DNA Preparation and Gel Electrophoresis:

Large scale plasmid DNA preparation was according to Clewell (1972) (12), and for small scale preparation a modification of the procedure described by Cough and al. (1980) (13) was used. Gel electrophoresis was done on vertical agarose gels ranging from 1.2 to 2% agarose according to the molecular weights to be detected (14). The electrophoresis was done in Tris-Acetate buffer. The gels were stained in a 0.5 mg/ml EtBr solution and photographed under short wave U.V. light.

Enzymes:

All enzymes were purchased from Boehringer Mannheim, Biolabs or B.R.L. and used according to the suppliers recommendations.

DNA Labelling and Sequencing:

Digested DNAs were either labelled in the 3' end by incorporating the appropriate α^{32}PdNTP with the klenow fragment of DNA polymeraseI or labelled in the 5' end by incorporating γ^{32}PdATP with T4 polynucleotide kinase on dephosphorylated DNA ends. The labelled DNA was then redigested and the fragment to be sequenced was purified from acrylamide gels or in case a combination of BamHI and EcoRI was used the small (15bp) undesirable labelled fragment was simply eliminated by isopropanol precipitation. The DNA was then submitted to chemical cleavage as described by Maxam and Gilbert (15). Samples of the four reactions (A, A+G, C, C+T) were loaded on ultra thin thermostated 7% acrylamide-urea gels.

Amino-acid Sequencing:

The amino-acid composition of the proteins was determined as described in (17). The sequences of the amino-terminal and the carboxy-terminal ends of the proteins were determined as described in (18) and (19).

Computer Analysis:

A combination of the programs from Staden (16), the Wisconsin university and the Intelligenetics Inc. was used for sequence analysis. The programs were run on a Vax 11/780 computer.

RESULTS

Isolation of a Plasmid Encoding the Eco RV Genes

Agarose gel electrophoresis of plasmid DNA isolated from the Eco RV strain J62(pGL74) (10) revealed several different species of plasmid DNA. This mixture was transformed in strain K514, and restricting transformants were isolated after phage selection (see Material and Methods). The restricting bacterial clones contained only a plasmid of 6.2 kb. These clones unlike the parental strain J62(pGL74) do not produce ColE1 type colicin nor are they immune to it, demonstrating that these properties

do not reside on the 6.2 kb plasmid. Earlier studies on the strain J62 (pGL74) suggest that the plasmid pGL74 encodes the colicin system, the resistance to Kanamycin and the Eco RV restriction and modification system. Therefore we chose to refer to the 6.2 kb plasmid described here as pLB1. This plasmid was shown to encode the Eco RV restriction and modification system according to the following criteria. i) The strain K514(pLB1) did not restrict phage λ grown on J62 (pGL74), which is therefore RV methylated, but was restricting unmodified λ vir.0 with an efficiency of 10^{-7}, thus showing Eco RV specific restriction. ii) λ phage grown on K514(pLB1) was able to grow with an e.o.p. of 1 on J62(pGL74), thus showing Eco RV specific methylation. iii) In vitro assays revealed that crude extracts from either K514(pLB1) or J62(pLG74) exhibited the EcoRV site specific cleavage activity on different substrate DNAs.

Mapping of the Eco RV Genes on pLB1

A restriction enzyme analysis of pLB1 showed that only a few of the most commonly used hexanucleotide recognizing enzymes, namely the enzymes Pst I, Bgl II, Hind II, and Cla I cleaved once, while there were no sites for Hind III, Sal I, Pvu I, Eco RI, Bam HI, Kpn I and Sph I. Pvu II was found to cut at least four times. Map positions of these sites are shown on Figure 1. In view of the few restriction sites available on pLB1 for subcloning we chose to construct pLB1::pBR322 cointegrates and to subsequently isolate deletion derivatives. For these constructions we took advantage of the unique sites present on both of these plasmids. The recombinant clones isolated from Cla I-Cla I, or Bgl II-Bam HI ligations exhibited normal restriction and modification properties (r+,m+) while the recombinant clones isolated from a Pst I-Pst I ligation had lost their capacity to restrict phage λ but retained the modifying activity (r-,m+). The Pst I site is thus located in the region coding for the Eco RV endonuclease. Because of the convenient arrangement of Pvu II sites (see Fig. 1) the Bgl II-Bam HI cointegrate was chosen to produce Pvu II deletion derivatives after Pvu II partial digestion. Such a strategy should yield smaller plasmids retaining the pBR322 origin of replication and the β-lactamase gene. All transformants exhibited a r+m+

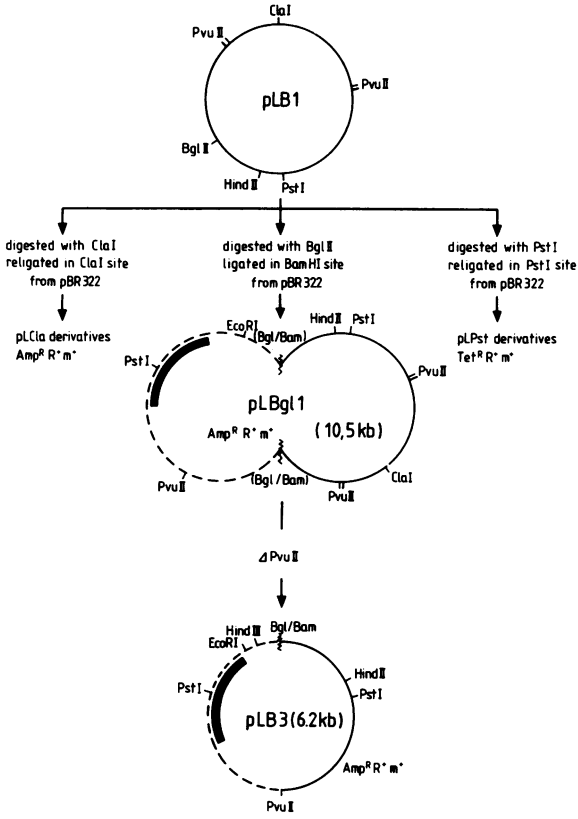

Figure 1: Construction of pBR322 derivatives carrying the Eco RV genes.
The DNA fragment coming from pBR322 is represented by a dot-line.

phenotype. The smallest of these plasmids, pLB3, 6 Kb long, retained 2.9 Kb of pBR322 and a 3 Kb Bgl II-Pvu II segment of pLB1 (see Fig. 1). Its restriction and modification properties were undistinguishable from the parental plasmid, thus showing that the 3 Kb Bgl II-Pvu II fragment contained all the necessary information for the expression of the Eco RV restriction and modification system.

Deletion Mapping of the Eco RV Genes

To locate the two genes more precisely we constructed a series of overlapping deletions (see Fig. 2). For that purpose we used the exonuclease Bal 31 to produce truncated DNA

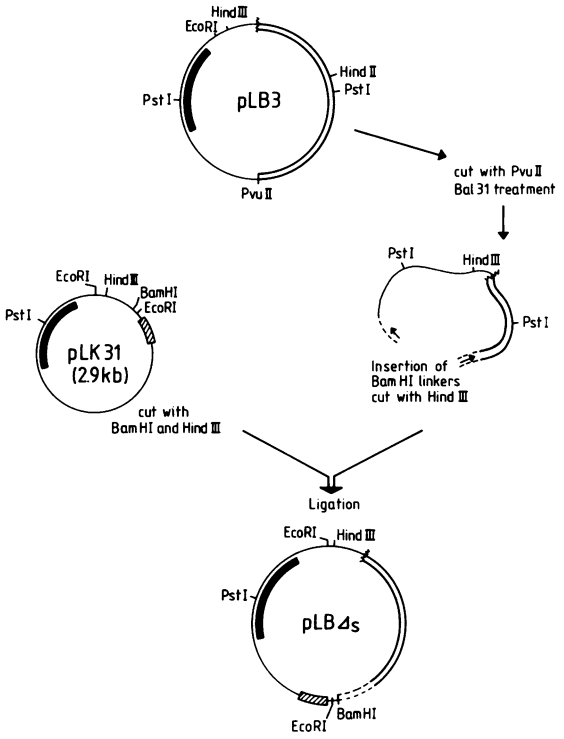

Figure 2: Construction of plasmids containing deletions entering the Eco RV genes.
The wavy bar indicates the fusion point between pBR322 and pBL1 in pLB3. The black box indicates the position of the Ampicillin resistance gene. The dashed box shows the position of the Fd terminators. The open box represents the DNA segment carrying the Eco RV genes.

fragments starting from the Pvu II site in the Pvu II-Bgl II fragment. After ligation of Bam HI linkers to the resected ends to facilitate subsequent DNA sequencing (21), the fragments were recloned in the polylinker of the plasmid pLK31 (M. Zabeau, manuscript in preparation). After transformation in strain K514, ampicillin resistant clones were isolated and analysed.

Depending on the relative location of the two genes two types of results could be expected. Either the resection first reaches the endonuclease gene in which case we could expect three classes of recombinant clones: r+,m+; r-,m+; r-,m-, or the deletions first enter the methylase gene in which case only

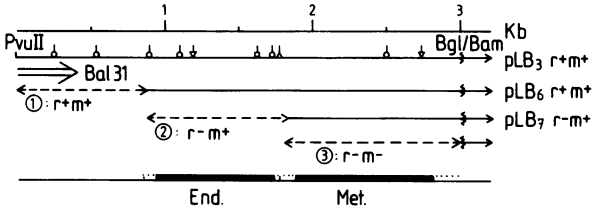

Figure 3: Deletion mapping of the Eco RV genes
The restriction map of the 3 Kb fragment of pLB3 is presented. The sites for the enzymes used for the mapping are shown (, HinFI; , AccI; , PstI; HindII). The three classes of deletion are indicated with the range of their corresponding end-points. The length of the shortest r^+m^+ (pLB6) and r^-m^+ (pLB7) plasmids are drawn. The approximate location of the endonuclease and methylase genes is indicated by black boxes.

two classes of recombinants are expected: r+,m+ or r-,m-, since the third class r+,m- would not be expected to be viable (5). The restricting and modifying properties of the ampicillin resistant recombinants were tested and correlated with the deletion map of the 3 Kb fragment (see Fig. 3). Three classes of recombinants were found in the expected order (r+,m+; r-,m+ and r-,m-) thus indicating that the endonuclease gene is the closest to the Pvu II site. The result of this deletion mapping experiment can be summarised as follows: (see Fig. 3) 880 bp of the Pvu II-Bgl II fragment could be deleted without affecting the restricting properties of the clones (r+,m+). Any deletion in the following 1100 bp abolished the restriction activity but the clones retained their modifying properties (r-,m+). All further deletions gave rise to clones having lost both restriction and modification phenotype (r-,m-). The relative position of the genes in pLB3 could therefore be established (Fig. 3).

DNA Sequencing

A subset of the deletion plasmids was used to sequence the 2.2 Kb DNA fragment in an ordered fashion (Fig. 4). Since all plasmids carry the same polylinker flanking the end-points of the deletions, they could be sequenced from the Bam HI site. Sequencing an average of 200 bp on each deletion fragment would give an overlap of 30-50 bp with the next fragment. In one case only was the distance between the end-points of adjacent

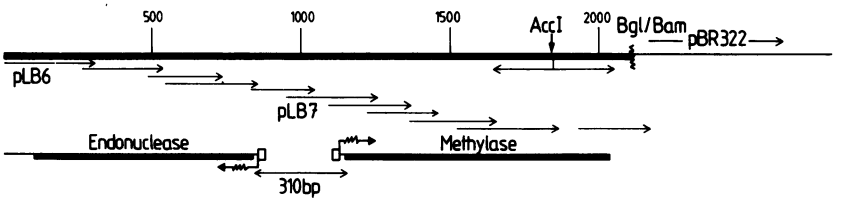

Figure 4: Sequencing strategy
The arrows represent the length of the region sequenced from each deletion plasmid. The organization of the genes and their potential promotors is represented schematically.

deletions too long an an Acc I site was used to sequence across the gap on both directions on both strands (see Fig. 4). For each deletion plasmid both strands were sequenced using both 5' and 3' labelled DNA as described in Material and Methods. The resulting sequence is presented in Fig. 5 and starts with the shortest r^+m^+ plasmid pLB6. The 54 bp beyond the BglII/BamHI fusion point belong to the tetracycline resistance gene of pBR322.

Gene Assignment

Scanning the sequence for open reading frames showed that only two had the potential to code for proteins of the expected molecular weight (about 30 Kd) (see Fig. 7). The two open reading frames were in perfect agreement with the results of the deletion mapping experiment. The strand assignment of the open reading frames suggests that the two genes are transcribed divergently from a 310 bp intergenic region.

The endonuclease gene is probably coded from position 834 to 97 with the deduced aminoacid sequence shown in Fig. 6. The following experiments established the deduced protein sequence: i) The aminoacid composition agrees with the deduced sequence (Table 1); ii) The molecular weight of this polypeptide would be 29 Kd which is in agreement with the value of 30 Kd observed on SDS-PAGE (A. D'Arcy and R. Brown; personal communication). iii) Amino-acid sequencing provided the final confirmation: the sequence of the first 40 N-terminal amino-acids and of the 11 C-terminal amino-acids was established by protein sequencing. The formylmethionine is obviously processed.

The methylase gene appears to be coded by the 1141-2037

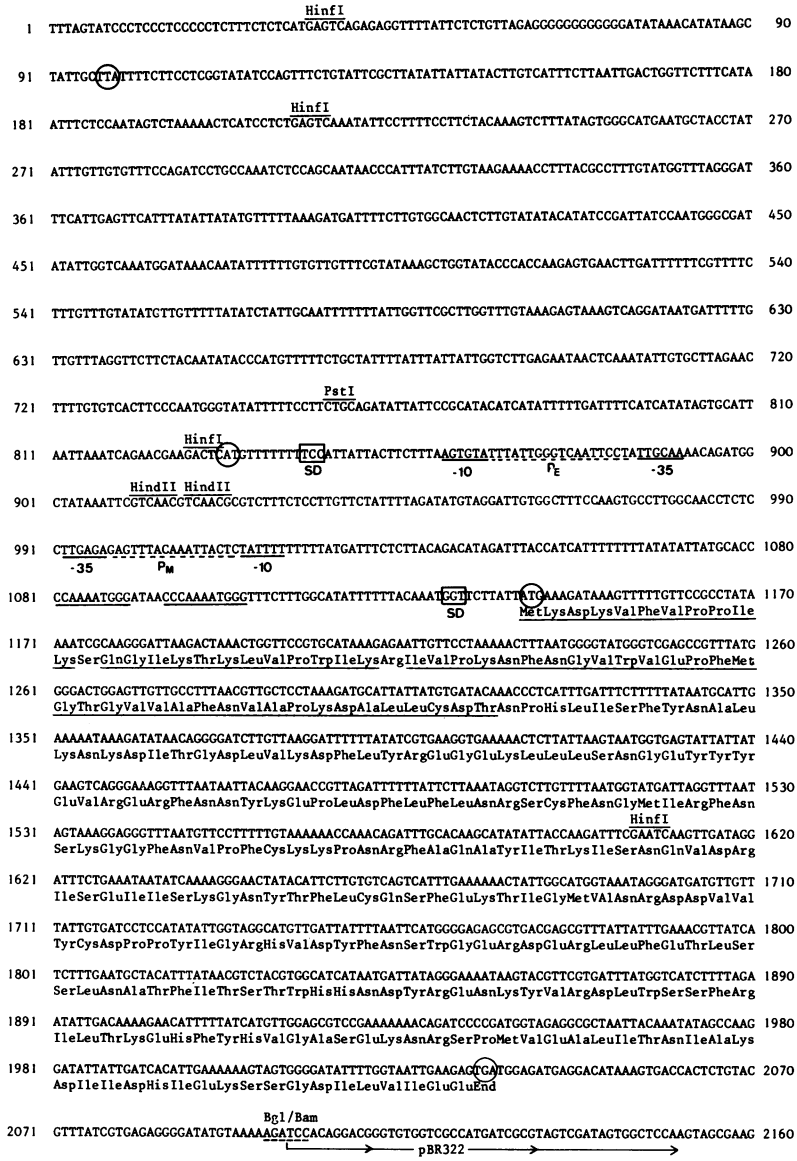

Figure 5: Complete DNA sequence of the Eco RV genes, present in pLB6 and aminoacid translation of the methylase gene.
The circles indicate the positions of start and stop codons for the endonuclease (on the reverse strand) and the methylase gene. The putative Shine and Dalgarno sequences are boxed. The potential promotors are underlined. Underlined aminoacids have been confirmed by aminoacid sequencing.

```
   1 ATGAGTCTTCGTTCTGATTTAATTAATGCACTATATGATGAAAATCAAAAATATGATGTATGCGGAATAATATCTGCAGAAGGAAAAATA  90
     MetSerLeuArgSerAspLeuIleAsnAlaLeuTyrAspGluAsnGlnLysTyrAspValCysGlyIleIleSerAlaGluGlyLysIle

  91 TACCCATTGGGAAGTGACACAAAAGTTCTAAGCACAATATTTGAGTTATTCTCAAGACCAATAATAAATAAAATAGCAGAAAAACATGGG 180
     TyrProLeuGlySerAspThrLysValLeuSerThrIlePheGluLeuPheSerARgProIleIleAsnLysIleAlaGluLysHisGly

 181 TATATTGTAGAAGAACCTAAACAACAAAATCATTATCCTGACTTTACTCTTTACAAACCAAGCGAACCAAATAAAAAAATTGCAATAGAT 270
     TyrIleValGluGluProLysGlnGlnAsnHisTyrProAspPheThrLeuTyrLysProSerGluProAsnLysLysIleAlaIleAsn

 271 ATAAAAACAACATATACAAACAAAGAAAACGAAAAAATCAAGTTCACTCTTGGTGGGTATACCAGCTTTATACGAAACAACACAAAAAAT 360
     IleLysThrThrTyeThrAsnLysGluAsnGluLysIleLysPheThrLeuGlyGlyTyrThrSerPheIleArgAsnAsnThrLysAsn

 361 ATTGTTTATCCATTTGACCAATATATCGCCCATTGGATAATCGGATATGTATATACAAGAGTTGCCACAAGAAAATCATCTTTAAAAACA 450
     IleValTyrProPheAspGlnTyrIleAlaHisTrpIleIleGlyTyrValTyrThrArgValAlaThrArgLysSerSerLeuLysThr

 451 TATAATATAAATGAACTCAATGAAATCCCTAAACCATACAAAGGCGTAAAGGTTTTCTTACAAGATAAATGGGTTATTGCTGGAGATTTG 540
     TyrAsnIleAsnGluLeuAsnGluIleProLysProTyrLysGlyValLysValPheLeuGlnAspLysTrpValIleAlaGlyAspLeu

 541 GCAGGATCTGGAAACACAACAAATATAGGTAGCATTCATGCCCACTATAAAGACTTTGTAGAAGGAAAAGGAATATTTGACTCAGAGGAT 630
     AlaGlySerGlyAsnThrThrAsnIleGlySerIleHisAlaHisTyrLysAspPheValGluGlyLysGlyIlePheAspSerGluAsp

 631 GAGTTTTTAGACTATTGGAGAAATTATGAAAGAACCAGTCAATTAAGAAATGACAAGTATAATAATATAAGCGAATACAGAAACTGGATA 720
     GluPheLeuAspTyrTrpArgAsnTyrGluArgThrSerGlnLeuArgAsnAspLysTyrAsnAsnIleSerGluTyrArgAsnTrpIle

 721 TACCGAGGAAGAAAATAA 738
     TyrArgGlyArgLysEnd
```

Figure 6: DNA sequence and aminoacid translation of the endonuclease gene. Position 1 corresponds to nucleotide 834 on the complete sequence (Fig. 5). Underlined aminoacids have been confirmed by aminoacid sequencing.

fragment. As for the endonuclease, the following data support the deduced protein sequence: i) The aminoacid composition fits the deduced sequence (Table 1). ii) The molecular weight of the polypeptide would be 35 Kd which is in agreement with the value of 33 Kd observed on SDS-PAGE (A. D'Arcy and R. Brown, personal communication). iii) The NH2 terminal end was sequenced till Thr(59) and exactly confirms the presented translation (see Fig.

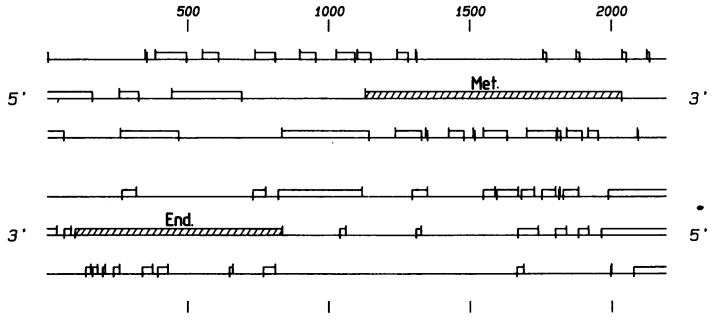

Figure 7: Reading frames in the sequence DNA fragment. Boxes show the open reading frames. Stop codons (—|—) are shown only when preceeded by a methionine (—|——) and internal methionines are not shown.

Table 1: Amino-acid composition of the endonuclease and methylase.
Numbers in columns 25' and 50' both represent the average of two analyses performed for 25 and 50 minutes respectively. The number of cysteines was separately analysed from 25' hydrolysis of performic acid oxidated samples. The number of tryptophans was separately analysed by a new micromethod (Maeda, Schefler & Tsugita, unpublished results). Column Int. gives the integer values calculated after protein analysis: the values for threonine and serine are 0 time extrapolations; the values for valine, isolencene and phenylalanine are taken from extreme times; values for tyrosine take the highest values because of chlorination. Column DNA give the aminoacid composition as predicted from the DNA sequence.

	AA composition of endonuclease				AA composition of methylase			
	25'	50'	Int.	DNA	25'	50'	Int.	DNA
Asx	33.9	33.7	34	34	37.0	36.7	37	41
Thr	13.1	13.2	15	15	11.1	11.4	12-13	13
Ser	14.2	12.8	15	15	19.1	19.2	20-21	19
Glx	22.0	22.0	22	22	20.7	22.2	21-22	23
Pro	8.8	8.4	9	9	10.7	10.6	11	13
Gly	16.2	16.0	16	15	18.9	19.1	19	16
Ala	10.4	10.0	10	9	10.5	10.5	10-11	10
Val	7.8	9.9	10	10	15.4	17.8	18	21
Met	0	0	0	1	2.9	2.8	3	5
Ile	17.7	20.5	21	26	15.3	17.2	17	23
Leu	13.9	14.1	14	14	20.6	21.3	21	22
Tyr	16.8	17.6	18	20	9.8	10.0	10	14
Phe	8.9	9.3	9-10	10	22.3	21.3	21-22	21
His	4.5	4.8	5	5	5.6	3.9	6	7
Lys	20.8	22.9	23	24	21.4	21.3	21-22	24
Arg	10.7	10.8	11	11	16.7	16.5	16-17	16
Cys	0.7		1	1	(4.8)		5	5
Trp	3.9		4	4	(5.0)		5	5

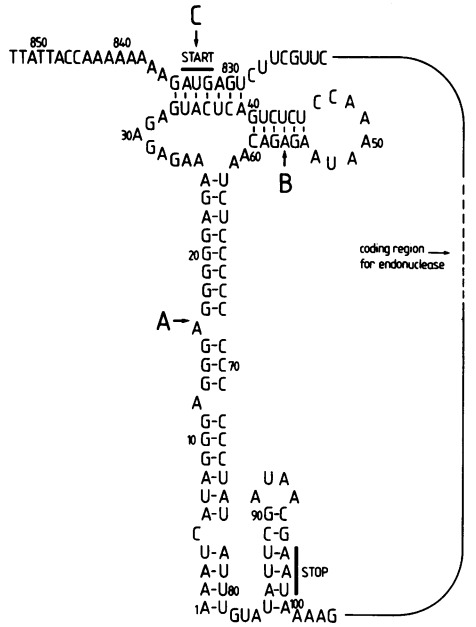

Figure 8: Secondary structure of the Eco RV endonuclease mRNA showing the stem loop structure downstream of the endonuclease coding region (stems A and B). The stem C would result from an interaction between the region surrounding the AUG start codon and the bulge loop between stems A and B.

5). The COOH terminal end is not readily accessible to carboxypeptidation and seems to be buried.

Transcriptional and Translational Signals

Promoter-like sequences can be found in both directions in the 310 bp intergenic region. The ones presenting the most homology with the plasmids and transposons promoter sequences (21) are indicated on Fig. 5. Two weak putative Shine and Dalgarno sequences are indicated on Fig. 5. Downstream of the endonuclease gene we find a GC rich region containing several inverted repeats. As shown on Fig. 8, these could form a very stable stem and loop structure (ΔG=-64 Kcal.) typical for transcription termination signals.

DISCUSSION

The characterisation of the Eco RV restriction and modification system can be summarized as follows: i) The

restriction efficiency of phage λ is estremely high (> 10^7 as compared to 10^4 for Eco RI). ii) This system involves two proteins, an endonuclease and a methylase of molecular weight 29 Kd and 32 Kd respectively. The DNA sequence of a 2.2 Kb subclone containing the two genes has been determined, and the genes located on that fragment. The A+T content of this sequence is 65.3% which differs greatly from the average E. coli A+T content (49%). The codon usage in the two genes reflects this A+T richness, and preference is given to codons ending with A or T, and when possible even to codons starting with A or T (see Table 2). The two genes are most probably transcribed in

Table 2: Codon usage in the Eco RV genes
The numbers indicate the fraction of one codon for each amino acid. The values for low expressed genes are from Grantham (31, 32) (E. coli L). The values for highly expressed genes are from The Wisconsin University (E. coli H).

AmAcid		FRACTION Endo RV	Meth RV	E.Coli H	E.Coli L
Arg	AGG	0.00	0.50	0.00	0.05
	AGA	0.73	0.25	0.00	0.12
	CGG	0.00	0.00	0.00	0.10
	CGA	0.18	0.00	0.00	0.10
	CGU	0.09	0.25	0.75	0.30
	CGC	0.00	0.00	0.25	0.32
Leu	UUG	0.14	0.27	0.00	0.10
	UUA	0.43	0.55	0.00	0.15
	CUG	0.00	0.05	1.00	0.43
	CUA	0.14	0.05	0.00	0.05
	CUU	0.21	0.05	0.00	0.15
	CUC	0.07	0.05	0.00	0.11
Ser	AGU	0.20	0.21	0.00	0.18
	AGC	0.33	0.00	0.22	0.14
	UCG	0.00	0.11	0.00	0.23
	UCA	0.20	0.26	0.00	0.16
	UCU	0.27	0.32	0.39	0.15
	UCC	0.00	0.11	0.39	0.14
Ala	GCG	0.00	0.20	0.33	0.28
	GCA	0.56	0.40	0.29	0.27
	GCU	0.11	0.20	0.38	0.18
	GCC	0.33	0.20	0.00	0.27
Gly	GGG	0.13	0.44	0.00	0.15
	GGA	0.67	0.25	0.00	0.11
	GGU	0.13	0.25	0.61	0.26
	GGC	0.07	0.06	0.39	0.48
Pro	CCG	0.00	0.38	0.87	0.38
	CCA	0.67	0.15	0.13	0.23
	CCU	0.33	0.46	0.00	0.20
	CCC	0.00	0.00	0.00	0.18
Thr	ACG	0.00	0.23	0.00	0.28
	ACA	0.73	0.46	0.00	0.11
	ACU	0.13	0.23	0.36	0.20
	ACC	0.13	0.08	0.64	0.40
Val	GUG	0.00	0.00	0.19	0.29
	GUA	0.50	0.19	0.29	0.21
	GUU	0.50	0.71	0.52	0.29
	GUC	0.00	0.10	0.00	0.22

AmAcid		FRACTION Endo RV	Meth RV	E.Coli H	E.Coli L
Ile	AUA	0.62	0.35	0.00	0.13
	AUU	0.23	0.65	0.17	0.50
	AUC	0.15	0.00	0.83	0.37
Asn	AAU	0.70	0.74	0.00	0.50
	AAC	0.30	0.26	1.00	0.50
Asp	GAU	0.50	0.94	0.36	0.59
	GAC	0.50	0.06	0.64	0.41
Gln	CAG	0.00	0.25	0.86	0.64
	CAA	1.00	0.75	0.14	0.36
Glu	GAG	0.19	0.32	0.22	0.42
	GAA	0.81	0.68	0.78	0.58
Cys	UGU	0.00	1.00	0.43	0.52
	GUC	1.00	0.00	0.57	0.48
His	CAU	0.80	0.86	0.16	0.69
	CAC	0.20	0.14	0.84	0.31
Lys	AAG	0.13	0.29	0.26	0.37
	AAA	0.88	0.71	0.74	0.63
Phe	UUU	0.70	0.90	0.25	0.62
	UUC	0.30	0.10	0.75	0.38
Tyr	UAU	0.75	0.86	0.23	0.64
	UAC	0.25	0.14	0.77	0.36

opposite directions from a 310 bp intergenic region. In the other E. coli restriction and modification system Eco RI the gene organization is different. The two genes are transcribed from the same DNA strand, starting from the endonuclease. Both the Eco RV and the Eco RI DNA sequences being available it is therefore possible to compare these two systems.

Comparison of the Eco RI and the Eco RV Systems

Heteroduplex analysis (Data not shown) revealed that pLB1 and pMB4 (a ColE1 type plasmid carrying the Eco RI genes (24)) are closely related but that the DNA homology does not extend in the genes coding for the restriction and modification systems. Further computer-aided DNA sequence comparison were unfruitful and no significant homology could be identified between the two systems. The same kind of search was performed at the aminoacid sequence level and gave no further result. The secondary structure of the Eco RV proteins has been predicted both by the Chou and Fassman method (23) and by the Robson method (24). It shows that like the Eco RI enzymes, they belong to the α-β alternated type but no region of higher similarity is found. The tertiary structure of those proteins is currently under investigation (7) and might provide more clues about their possible structural relationship. In the Eco RI system, as well as in the Eco RV system, the base composition (65% A+T) (25) (26) and the codon usage are different from the E. coli average. The argument was made that the Eco RI genes might have been translocated on the pMB1 plasmid where they were first found, and that those genes might originate from a species whose DNA has a high A+T content. The results of the heteroduplex experiment, as mentioned previously, indeed show a strong homology between the carrier plasmids. That would corroborate the hypothesis of a possible translocation of those restriction systems in a Col E1 type plasmid. Although these findings do not support a genetic relationship between those two systems, it shows that they are equally different from E. coli. Either that reflects a foreign origin of these genes or it is an inherent feature of this type of systems. It had been mentioned (25) that codons corresponding to minor tRNA species were commonly used in the Eco RI genes, as they are in the Eco RV case, and

that that could represent a very simple mechanism to modulate the rate of translation. However, since those enzymes (endonuclease and methylase) can be overproduced [Bougueleret manuscript in prep.], such a mechanism would not be sufficient alone to guarantee a low translational level of the protein.

mRNA Secondary Structure and Possible Regulation Mechanism

It is clear that the expression of the endonuclease gene must be tightly controlled. We have already shown that the rate of translation of the Eco RV endonuclease could be modulated at the levels of translation initiation (poor ribosome binding site) and elongation (codon usage). We now propose an additional mechanism of regulation suggested by the endonuclease mRNA secondary structure, which is analogous to that proposed for the translational regulation of the A-protein of phage MS2 (27) (28). The secondary structure of the mRNA potentially transcribed from the endonuclease gene has been predicted, using the Zucker and Stiegel algorithm (29). The main body of this RNA can form a series of low energy stem loop structures. The most stable structure (ΔG=-64 Kcal) is found in the region 1-110 (see Fig. 5), located immediately downstream of the endonuclease gene. The main stem involves the GC rich region which as suggested earlier could act as a transcription terminator. This stem exposes in a bulge loop a 7 bases sequence which can be perfectly paired with the sequence surrounding the endonuclease start codon (see Fig. 8). This new stem alone (stem C on Fig. 8) would have a free energy of 9 Kcal as calculated from Tinoco et al. (32). Moreover this stem could form a single helical structure together with stem B and stabilize even more the presented structure. If such an interaction really occurs it would, as soon as a full length mRNA is synthesized, impede further translation initiation on this mRNA molecule. Thus only nascent RNA could be translated and that would ensure a low level of endonuclease production. Since no such structure can be identified in the methylase gene, upon transformation of the restriction and modification system, the methylase could quickly reach a higher level in the cell and protect the endogenous DNA. Since such a mechanism is based on the intrinsic structure of the mRNA it would override all other control system operational

in the cell, in this way the endonuclease production would be tightly controlled in any type of cell. If the hypothesis of translocability of the Eco RV system is true, it would help the spreading of such a system. It would be interesting to check other restriction endonuclease DNA sequences for that type of long distance interaction.

ACKNOWLEDGMENT

We would like to greatly thank Prof. R. Cortese for helpful discussions and Heide Seifert for typing the manuscript. Thanks are due to Dr. J. Devereux from the Genetics Computer Group of the Wisconsin University for providing his programs.

*Present address: EMBRAPA-CENARGEN S.A.I.N., Parque Rural CP 102372, CEP 70770, Brasilia DF, Brazil

+Present address: P.G.S., Plateaustraat, 22 Gent, Belgium

REFERENCES

1. Mann M.B. (1981) in Chirikian J.G. (Ed). Gene amplification and analysis Vol. 1 Restriction endonucleases, Elsevier/North Holland, New York pp 229-237.
2. Mann M., R. Nagaraga Rao and H.O. Smith (1978) Gene 3 pp 97-112
3. Schoner B., S. Kelly and H.O. Smith (1983) in press.
4. Walder R.Y., J.L. Hartley, J.E. Donelson and J.A. Walder (1981) P.N.A.S. 78 pp 1503-1507.
5. Kosykh V.G., Y.I. Buryanov and A.A. Bayev (1980) M.G.G. 178 pp 717-718.
6. O'Connor C.D. and G.O. Humpheys (1982) Gene 20 pp 219-229.
7. Young T.S., S.H. Kim, P. Modrich, A. Beth and E. Jay (1981) J. Mol. Biol. 145 pp 607-610.
8. Sauer R.T., R.R. Yocum, R.F. Doolittle, M. Lewis and C.O. Pabo (1982) Nature 298 pp 447-451.
9. Gicquel-Sanzey B. and P. Cossart (1982) E.M.B.O. Journal 1 pp 591-595.
10. Kholmina G.V., B.A. Rebentish, Y.S. Koblov, A.A. Miromov, N.K. Yankovskii, Y.I. Kozlov, L.I. Glatman, A.F. Moroz and V.G. Debabov (1980) Dokl. Acad. Nauk. SSSR, 253 pp 495-497.
11. Dagert H. and S.D. Ehrlich (1979) Gene 6 pp 23-28.
12. Clewell D. (1972) J. Bact. 110 pp 667-676.
13. Cough N.M., E.A. Webb, S. Cory and J.M. Adams (1980) Biochemistry 19 pp 2702-2710.
14. Sudgen B., B. De Troy, R.J. Roberts and J. Sambrook (1975) Anal. Biochem. 68 pp 36-46
15. Maxam A. and W. Gilbert (1980) in Grossman L. and Moldane K. (eds) Methods in enzymology 65 Academic press N.Y. pp 499-560.
16. Staden R. (1980) N.A.R. 8 pp 3673-3694.
17. Tsugita A. and J.J. Scheffer (1982) Eur. J. Biochem. 124 pp 585-588.

18. Tsugita A., I. Gregor, I. Kubota and R. v.d. Broek (1979) in Cytochrome oxydase (King, T.E. et al., eds), pp 67-77. Elsevier Biomedical Press, Amsterdam.
19. Tsugita A., B. Blazy, M. Takahashi and B. Baudras (1982) FEBS Lett. 144 pp 304-307.
20. Ruther K., M. Koenen, K. Otto and B. Muller-Hill (1981) N.A.R. 9 pp 4087-4098.
21. Hawley D.K. and W.R. McClure (1983) N.A.R. 11 pp 2237-2255.
22. Chou P.Y. and G.D. Fassman (1978) Adv. Enzymol. 47 pp 45-148.
23. Garnier J., D.J. Osguthorpe and B. Robson (1978) J. Mol. Biol. 88 pp 97-120.
24. Betlach M., V. Hershfield, L. Chow, W. Brown, H.M. Goodman and H.W. Boyer (1976) Federation proceedings 35 pp 2037.
25. Greene P.J., M. Gupta and H.W. Boyer (1981) J. Biol. Chem. 256 pp 2143-2153.
26. Newman A.K., R.A. Rubin, S.H. Kim and P. Modrich (1981) J. Biol. Chem. 256 pp 2131-2139.
27. Fiers, W., R. Contreras, F. Duerinck, G. Haegmean, J. Merregaert, W. MinJou, A. Raeymakers, G. Volckaert, M. Ysebaert, J. v.d. Kerckhove, F. Nolf and M. Van Montagu (1975) Nature 256 pp 273-278.
28. Weissman C., M.A. Billeter, H.M. Goodman, J. Hindley and H. Weber (1983) A. Rev. Biochem. 42 pp 303-328.
29. Zucker M. and Stiegler D. (1981) N.A.R. 9 pp 133-148.
30. Grantham R., C. Gautier, M. Gouy, R. Mercier and A. Pave (1980) N.A.R. 8 r49-r62.
31. Grantham R., C. Gautier and M. Gouy (1980) N.A.R. 8 pp 1893-1912.
32. Tinoco Jun. I., P.N. Borer, B. Dengler, M.P. Levine, O.C. Uhlenbeck, D.M. Mc Crothers and J. Gralla (1973) Nature 246 pp 40-41.