Volume 15 Number 1 1987 Nucleic Acids Research

Nucleotide sequence and expression of the gene encoding the *Eco*RII modification enzyme

Subhendu Som, Ashok S.Bhagwat+ and Stanley Friedman

Department of Pharmacology, SUNY-Health Science Center, 450 Clarkson Avenue, Brooklyn, NY 11203 and +Cold Spring Harbor Laboratory, PO Box 100, Cold Spring Harbor, NY 11724, USA

Received August 20, 1986; Revised October 27, 1986; Accepted November 17, 1986

ABSTRACT

The gene coding for the EcoRII modification enzyme has been cloned and the nucleotide sequence of 1933 base pairs containing the gene has been determined. The gene codes for a protein of 477 amino acids. Two transcriptional start sites have been mapped by S_1 mapping. One deletion that removes 34 N-terminal amino acids was found to have partial enzyme activity. Comparison of the EcoRII methylase sequence with other cytosine methylases revealed several domains of partial homology among all cytosine methylases. Cloning the gene in multicopy pUC vectors increased the expression by 6-18 fold. A 40 fold overproduction of the EcoRII methylase was obtained by cloning the gene in the expression vector carrying the lambda P_L promoter.

INTRODUCTION

The sequence CC(A/T)GG is the common recognition site for the E. coli DNA cytosine methylase (dcm) and the EcoRII restriction-modification system (1). Over 35 enzymes have been described that recognize this DNA sequence (2), but none has been sequenced as yet. Both dcm and EcoRII enzymes methylate the internal cytosine residue in this sequence. Methylation of this base protects the DNA from restriction by the EcoRII endonuclease.

Both these enzymes, as well as all other DNA(cytosine-5)methyltransferases tested are inhibited by DNA containing 5-azacytosine in the sequence they recognize (3,4). Tight DNA-enzyme complexes are formed with azacytosine containing DNA that can be identified by gel electrophoresis (5). In the case of the EcoRII methylase (EcoRII.M) it has been shown that the enzyme binds to DNA fragments that contain 5-azacytosine in the position that would normally be methylated (6). As a step towards understanding the basis of this binding reaction we have determined the sequence of the gene coding for the EcoRII methylase and prepared plasmids that overproduce the enzyme.

We have compared the primary sequence of the EcoRII methylase with the sequence of other adenine and cytosine methylases. This comparison has shed light on the evolution of bacterial methylases.

MATERIALS AND METHODS

Strains and Plasmids

E. coli HB101 (7) was used as host in all cloning experiments involving pBR322 (8) and pUC (9,10) vectors. *E. coli* F1100 (Su^+ *endI thi*) was kindly provided by Dr. S. Hattman (11). *E. coli* N4830 (12), a temperature sensitive mutant carrying the λ *cI857* and N genes was obtained from PL biochemicals and was used as host in cloning pPL-λ recombinants. Plasmid pPL-λ , a pKC30 derivative carrying the P_L-λ promoter and N gene was purchased from PL Biochemicals. *E. coli* JM107 (10) was used as host for M13 cloning. Plasmid N3(Tc^r Su^r Sm^r Res^+ Mod^+ (RII)) (13) was also provided by Dr. Hattman. Plasmid vectors pUC18 and pUC19 were from Bethesda Research Laboratories.

Media and Growth Condition

Unless specifically mentioned, bacteria were grown in LB medium (14) at 37°C. Antibiotics were used in the following concentrations: ampicillin, 50 μg/ml; streptomycin, 25 μg/ml; sulfathiazole, 200 μg/ml and chloramphenicol 170 μg/ml. Sulfathiazole was used only in M9 medium (14).

Enzymes and Chemicals

Restriction enzymes (except *XhoII*), nuclease Bal-31 and T_4 DNA ligase were from New England Biolabs. Synthetic linkers *BamHI*, *HindIII*, *PstI* and *SalI* were also from the same vendor and were phosphorylated before use. Restriction endonuclease *XhoII* and nuclease S_1 were purchased from Boehringer Mannheim. DNA polymerase I large fragment was obtained from BRL. Calf intestinal alkaline phosphatase and proteinase K were from Sigma Chemical Co. T_4 polynucleotide kinase and deoxyribonuclease were the products of PL Biochemicals and Worthington respectively. Enzymes were used according to manufacturers recommendations. ^{35}S- α -dATP(>1000 Ci/mmole) was purchased from New England Nuclear. S-adenosyl-L-methionine[methyl-^{14}C](48 Ci/mmole) was from Research Products. L-[^{35}S]methionine(>800 Ci/mmole) and gamma ^{32}P-ATP(4500 Ci/mmole) were obtained from Amersham and ICN respectively. All other chemicals were of reagent grade quality.

Cloning Techniques

Plasmid DNA was purified by the alkaline lysis method followed by CsCl centrifugation (14). DNA fragments used for subcloning were prepared by Bal-31 digestion. The ends were made flush with Klenow fragment of DNA polymerase I, synthetic linkers attached, digested with the appropriate restriction enzyme(s) and ligated into a pUC vector. DNA was purified either from low melting point agarose or by electroelution. Ligation of fragments to vector DNA was performed at 16°C overnight. The ligation mix was heated

to 65°C prior to use in transformation. Transformation was done by Hanahan's procedure (15). Transformed colonies were screened after restriction digestion of the plasmid minipreps by agarose gel electrophoresis.

DNA Sequencing

Nucleotide sequences were determined by Sanger's dideoxy chain termination method (16) using M13 mp18 and mp19 recombinants as templates. Both strands were sequenced (strategy described in Figure 3). Fragments generated by restriction digestion and serial Bal-31 deletion were cloned in M13mp18/mp19 RF DNA with the help of linkers in the desired orientation for sequencing and transfected into E. coli JM107. M13 recombinants were isolated from transfected cells and screened by restriction digestion of the RF DNAs. Template DNAs for sequencing were purified from clones having inserts differing from one another by 150 to 200 bp and sequenced using a 17-mer synthetic oligonucleotide primer and α-^{35}S-dATP as label (17).

Computer Analysis

Data handling and sequence analysis were performed on an IBM personal computer with a program developed by A.D. Delaney (18). DNA and protein sequences were analyzed for homology using DIAGON (19) and SEQHP (20) programs on a SUN computer.

S_1 Mapping of the Transcript

Total bacterial RNA was isolated by the method of Brosius et al (21) from E. coli HB101 carrying pSS18 (see Figure 1) and digested with proteinase K treated DNase. The total RNA was precipitated with ethanol after phenol extraction and dissolved in water to a concentration of approximately 4 µg/µl. A 362 bp TthIIII-XhoII fragment covering the methylase promoter region was isolated from plasmid pSS18, purified by gel electrophoresis, and labeled at the 5' end with ^{32}P. The strands were separated on a 5% polyacrylamide gel and extracted by diffusion (22). Approximately $3X10^3$ CPM of DNA was mixed with 20 µg of total bacterial RNA in 20 µl hybridization buffer containing 40 mM PIPES, 1 mM EDTA, 0.4 M NaCl, 50% formamide, pH 6.4, and incubated at 37°C for 15 hrs. The slow moving strand was found to be the minus strand when a small aliquot from each hybridization mix was tested by gel electrophoresis on a 5% polyacrylamide gel. The rest of the mixture (19 µl) was added to 200 µl of S_1 nuclease buffer containing 0.28 M NaCl, 50 mM CH_3COONa, pH 4.6, 4.5 mM $ZnSO_4$, 20 µg/ml denatured salmon sperm DNA (14) and incubated with 500 U/ml S_1 nuclease at 30°C for 75 minutes. The reaction was stopped by adding ammonium acetate and EDTA to final concentrations of 75 and 10 mM respectively. The mixture was extracted with

phenol-chloroform and the S_1 resistant hybrid was precipitated with ethanol. The pellet was suspended in 10 µl 95% formamide, 0.5% each of bromophenol blue and xylene cyanol, 1 mM EDTA, heated to 90°C for 2 min and loaded onto a 7% polyacrylamide gel containing 7.5 M urea. A sequencing ladder was used as molecular weight markers to identify the exact length of the S_1 protected fragment.

Assay of Methylase Activity

Bacteria were grown in LB medium at 37°C with the appropriate antibiotic to A_{550} of about 2.0 to 2.25. Cells from 20 ml culture were harvested, washed with 10 mM tris-HCl, pH 8.0, containing 1 mM EDTA and 5 mM β-mercaptoethanol and suspended in 1.0 ml buffer containing 10 mM potassium phosphate, pH 7.0, 1 mM EDTA, 7 mM β-mercaptoethanol, 1 mM sodium azide, 0.4 M NaCl, 25 µg/ml phenylmethylsulfonylfluoride and lysozyme (50 µg/ml). Cells were sonicated and centrifuged at 100,000 X g as described (23) and the supernatant was used as the crude source of enzyme. Protein was determined by Lowry's method (24). In the case of E. coli N4830 carrying plasmids containing the thermo-inducible P_L-λ promoter, cells were grown initially at 30°C to A_{550} of 2.0, the temperature was quickly elevated to 42°C by addition of an equal volume of medium preheated to 65°C and incubation was continued for another hour before harvesting. The assay procedure measured the amount of ^{14}C-methyl incorporated into DNA from S-adenosyl-L-methionine[methyl-^{14}C] (23).

^{35}S Labeling of Protein

E. coli N4830, carrying plasmid pSS113 or pSS114 (see Figure 7) was grown at 30°C in LB medium containing 50 µg/ml ampicillin to A_{550} of 0.5. One ml of the culture was centrifuged. The pellet was washed with ten volumes of M9 medium, resuspended in 10 ml M9 supplemented with 0.1% of all amino acids except methionine, 0.2% glucose, 1 mM $MgSO_4$, 1 µg/ml vitamin B_1, 40 µg/ml ampicillin and reincubated at 30°C. When the A_{550} reached 0.25 to 0.3, the temperature was elevated to 42°C. Ninety minutes after thermal induction 200 µl of culture was pulse labeled for one minute with 20 µCi of ^{35}S-methionine (>800 Ci/mmole). An equivalent pulse was also given to the cells growing at 30°C. The reaction was stopped by adding β-mercaptoethanol and SDS to final concentrations of 5 and 1% respectively and heating the samples at 90°C for 1 minute. The total protein was precipitated with TCA and resuspended in gel-loading buffer (25). Equivalent aliquots were loaded after heating at 95°C onto a 5-15% gradient polyacrylamide gel containing 0.1% SDS and electrophoresed at 100 volts for 6 to 7 hrs. The gel was fixed in 7% acetic acid, 20% methanol, dried, and exposed for autoradiography.

RESULTS

Cloning of the EcoRII Methylase Gene

Plasmid N3 carries the EcoRII restriction-modification genes (26). Both restriction and modification genes have been cloned together as an EcoRI fragment of approximately 9 kb from N3 (27). This fragment also contains the region for sulphonamide resistance (Su^r) (27). We cloned this 9 kb EcoRI fragment from N3 in the EcoRI site of pBR322 to construct a 13.3 kb recombinant plasmid pSF1 (Figure 1) which was found to be Su^r, Res^+ Mod^+ (r^+m^+). Plasmid pSF2 (Figure 1) was constructed by digesting pSF1 with BamHI and religating the largest fragment (approximately 8.75 kb). This clone was found to be

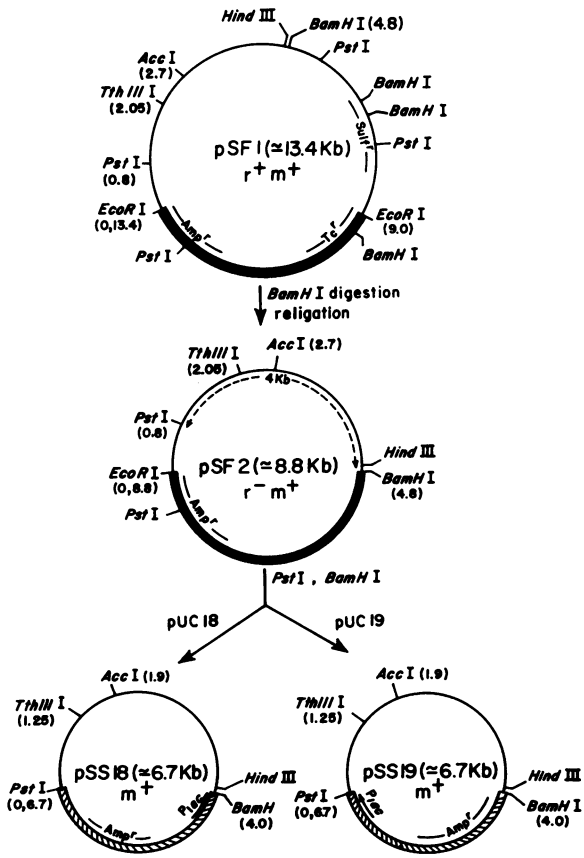

Figure 1. Structures of pSF1, pSF2, pSS18 and pSS19. The vector portions of the recombinants are represented by black (pBR322) or hatched boxes (pUC vectors). Selected restriction sites with relative positions in approximate kilobase pairs are noted.

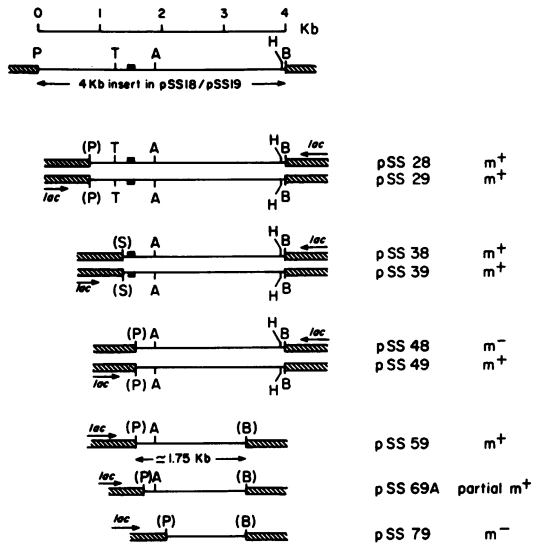

Figure 2. Localization of the EcoRII methylase gene and its promoter by serial Bal-31 deletion mapping. At the top is the 4 kb insert cloned in pSS18 or pSS19. The methylase activity is given on the right. Restriction sites: P, PstI; B, BamHI; S, SalI; H, HindIII; T, TthIIII and A, AccI. Sites created by linkers are shown in parentheses. The black box is the predicted promoter region. The hatched boxes represent the vector portions of the constructs. The vector used in construction of pSS28, pSS38 and pSS48 is pUC18, while pUC19 is the vector in pSS29, pSS39, pSS59 and pSS79. In pSS49 and pSS69A, pUC9 is the vector. Direction of transcription from the lac promoter is indicated.

devoid of endonuclease activity when tested for phage restriction, was Su sensitive, but expressed the methylase activity. Plasmid pSF2 was then digested with BamHI and PstI and the resulting fragment of approximately 4kb was ligated with BamHI and PstI cut pUC18 and pUC19 to give recombinants pSS18 and pSS19 respectively (Figure 1). When HB101 transformants were tested for methylase activity, strains carrying either of the plasmids were found to have the same activity. This suggests that the EcoRII methylase gene has its own promoter.

Localization of Methylase Coding Region by Bal-31 Deletion

The size of EcoRII methylase protein is reported to be 55 kd by Buryanov et al (28) and 60 kd by Friedman (5), suggesting an effective coding region of approximately 1.45 to 1.6 kb in length. To localize the proper coding region in the 4 kb insert of pSS18/pSS19 we gradually shortened the insert length by serial Bal-31 deletions and subcloned the deleted derivatives into pUC18 and pUC19. Figure 2 shows these constructions. A deletion of

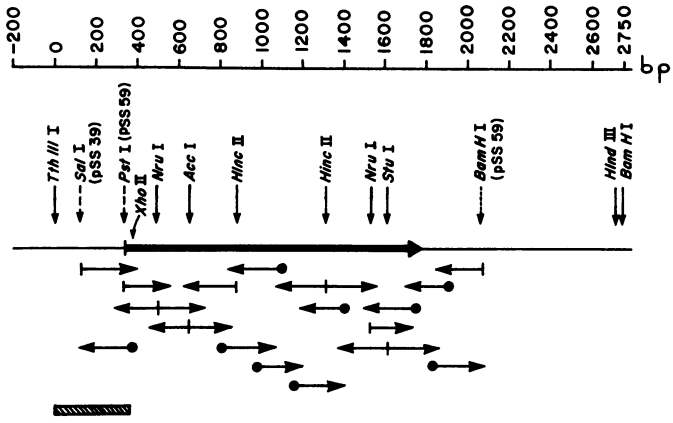

Figure 3. Restriction map of the region coding for the EcoRII methylase gene and the sequencing strategy. Selected restriction sites shown were used to generate fragments for sequencing. Sites indicated by dashed arrows are not present in the original N3 DNA. The hatched segment shown at the bottom is the Tth111I-XhoII fragment used as probe in S_1 mapping. ⟶, fragments prepared by Bal-31 deletion; ⊢⟶, prepared by restriction digestion.

approximately 1.35 kb from the PstI site of pSS18/pSS19 (Figure 1) did not affect the expression of the methylase gene when subcloned in either orientation (plasmids pSS38 and pSS39, Figure 2). However, further removal of approximately 200 bp from the same end resulted in expression of activity only when the fragment was placed under the control of the lac promoter (pSS49, Figure 2). Transcription of the methylase gene must proceed from methylase to restriction enzyme and the endogenous promoter of the methylase gene is deleted in plasmids pSS48 and pSS49. A removal of approximately 650 bp from the other end (i.e., BamHI site) of the insert in pSS49 allowed us to construct pSS59 (Figure 1) which was found to be the most active among all pUC recombinants and had an insert size of about 1.75 kb (Table 2). Deletion from the PstI site of pSS59, of approximately 100 bp (pSS69A) and 450 bp (pSS79) resulted in partial and complete loss of methylase activity, respectively.

Nucleotide Sequence and Codon Usage

The sequence of 1933 base pairs was determined for both strands starting 120 bp downstream from the Tth111I site (position 1.25 of pSS18/pSS19, Figure 1). The sequencing strategy is described in Figure 3 and the complete sequence is shown in Figure 4. Only one open reading frame large enough to code for the EcoRII methylase was found. The 1431 bp region from the first ATG codon

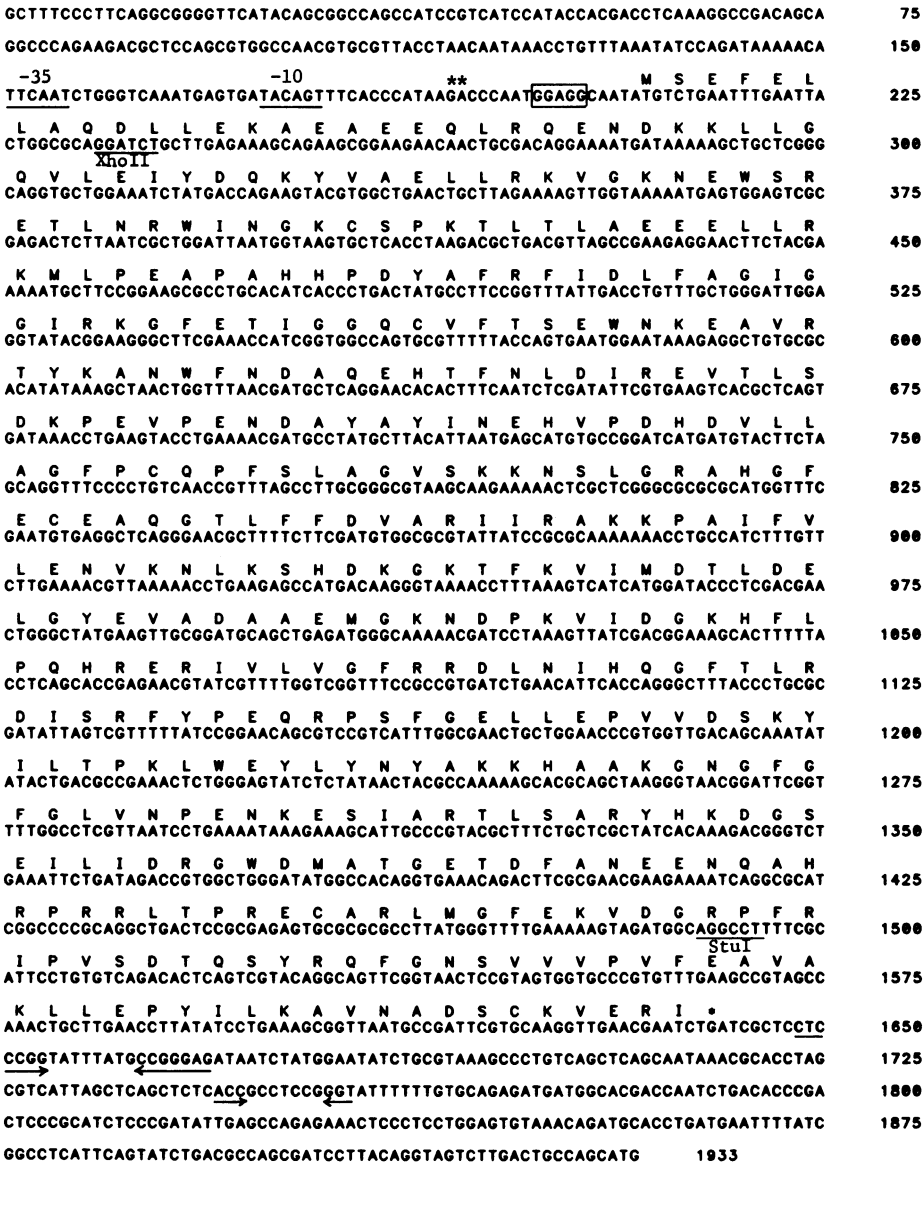

Figure 4. Nucleotide sequence of the EcoRII methylase gene and derived amino acid sequence. Shine-Dalgarno sequence is indicated by a box 5' to the coding region. Transcriptional start sites are indicated by asterisks. The -10 and -35 region promoter sequences are underlined. The XhoII site used to prepare the probe for S_1 mapping is marked. Regions of inverted repeats at the 3' end are indicated by horizontal arrows.

Table 1. Comparison of the codon usage of the EcoRII methylase gene with highly expressed E. coli genes

			frequency per 1000					frequency per 1000					frequency per 1000	
		[a]M	M	[b]HE			M	M	HE			M	M	HE
Arg	CGU	8	17	44	Pro	CCU	12	25	2	Asn	AAU	11	23	2
	CGC	11	23	18		CCC	4	8	1		AAC	12	25	35
	CGA	5	10	0		CCA	0	0	6					
	CGG	3	6	0		CCG	7	15	25	Gln	CAA	2	4	5
	AGA	1	2	1							CAG	13	27	31
	AGG	3	6	0	Ala	GCU	10	21	64					
						GCC	10	21	8	His	CAU	6	13	6
Leu	CUU	12	25	3		GCA	6	13	33		CAC	7	15	13
	CUC	8	17	2		GCG	11	23	23					
	CUA	2	4	0						Glu	GAA	35	73	51
	CUG	19	40	60	Gly	GGU	13	27	45		GAG	10	21	17
	UUA	3	6	2		GGC	10	21	31					
	UUG	1	2	3		GGA	4	8	1	Asp	GAU	17	36	16
						GGG	4	8	1		GAC	11	21	35
Ser	UCU	3	6	21										
	UCC	1	2	15	Val	GUU	11	23	38	Tyr	UAU	11	21	3
	UCA	3	6	1		GUC	3	6	6		UAC	4	8	15
	UCG	3	6	0		GUA	6	13	27					
	AGU	4	8	3		GUG	10	21	15	Cys	UGU	2	4	1
	AGC	5	10	8							UGC	4	8	4
					Ile	AUU	11	23	12					
Thr	ACU	4	8	28		AUC	9	19	46	Phe	UUU	16	34	6
	ACC	5	10	24		AUA	3	6	0		UUC	11	23	20
	ACA	3	6	4										
	ACG	6	13	2	Lys	AAA	24	50	62	Met	AUG	6	13	21
						AAG	12	25	25					
										Trp	UGG	6	13	6

[a]M, EcoRIIM; [b]HE, highly expressed genes in E.coli; values are taken from Gouy and Gautier (34).

at position 208 to the termination codon TGA at position 1638 codes for 477 amino acids (Figure 4). The size of this protein would be 54,564 daltons. The translational start codon is preceeded by a strong Shine-Dalgarno sequence GGAGG (positions 199-203, Figure 4).

An analysis of the nucleotide sequence revealed almost equal distribution of A+T (51.8%) and G+C (48.2%) content. Unlike the A+T rich (65%) EcoRI (29,30) and EcoRV (31) restriction-modification genes, base composition of the EcoRII methylase is similar to that of E. coli (32).

Analysis of codon usage (Table 1) has revealed a pattern typical for weakly expressed genes in E.coli. Genes coding for major proteins show a

narrow and selective preference in their codon usage (33,34). In strongly expressed genes, 23 codons with an individual occurence of less than 5 per 1000 appear with a total frequency of 36 per 1000 (3.6%), whereas in the EcoRIIM gene the frequency of occurrence of these codons is 25.5%. A broad preference for codons in the EcoRIIM gene can be seen for almost all amino acids. In particular the rare codons for Leu, Ser, Pro and Tyr are used more than 50% of the time.

Transcriptional Start and Stop Signals

In order to determine the transcriptional start point(s) of the gene, RNA from HB101(pSS18) was hybridized to the terminally labeled minus strand of an XhoII-Tth111I fragment (see Figure 3) from pSS18 covering the 5' end and the promoter region of the methylase gene. The length of the S_1 resistant DNA fragment was determined by gel electrophoresis. Two strong signals of protected DNA, 48 and 49 nucleotides in length along with some weaker bands, were detected in the autoradiogram (Figure 5). The purines G and A at positions 191 and 190 (Figure 4) corresponded to the 48th and 49th nucleotides respectively. These sites are upstream from the proposed Shine-Dalgarno sequence GGAGG (positions 199-203, Figure 4). Positions 174 to 179 contain the sequence TACAGT which resembles the -10 region of a typical E. coli promoter. The underlined bases match the consensus -10 region promoter of Rosenberg and Court (35). This hexanucleotide has been reported to be the -10 region promoter sequence for ϕ X174B (36). This sequence is 11 and 12 bp respectively, upstream of the A and G start sites, 2 bp more than usually found for E. coli promoters. The hexanucleotide TTCAAT, 17 bp proximal of the proposed -10 box contains 3 out of the six bases reported for the E.coli -35 consensus sequence, TTGACA (35).

There are two sets of sequences downstream from the methylase gene that are likely candidates for a transcription termination signal. One is a GC rich region (starting at position 1648, Figure 4) of inverted repeats in the vicinity of the translational stop codon TGA (position 1639) which can form a very stable secondary stem and loop structure, but it is not followed by a series of T residues. The other is a region 102 bp downstream to the stop codon, where an inverted repeat of three base pairs is followed by a series of T residues (Figure 4).

Homology between EcoRII Methylase and other Methylases

The DNA methylases whose sequences have been reported are EcoRI (29,30), EcoRV (31), PstI (37), HhaII (38), BspRI (39), BsuRI (40), SPR (41,42), PaeR7 (43) DpnII (44), ϕ3T (45) and dam (46). While dam, EcoRI, EcoRV, HhaII,

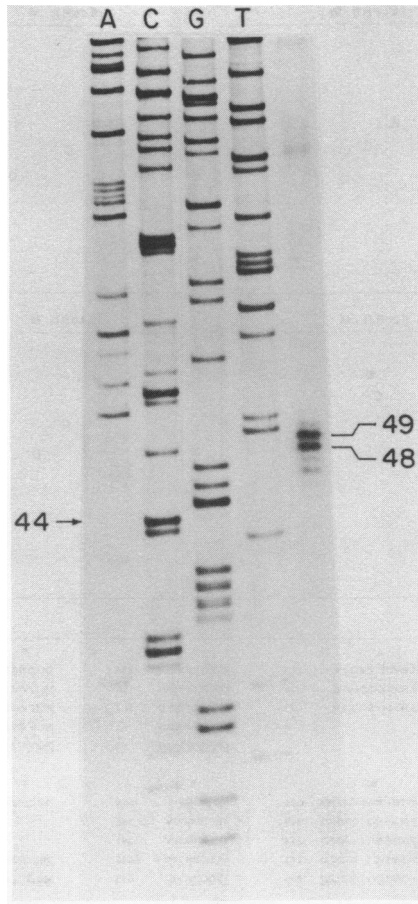

Figure 5. Nuclease S_1 mapping of the transcriptional start point. The Autoradiogram shows the S_1 protected fragments in the extreme right lane. Sequencing products from an M13mp19 recombinant having an insert at the BamHI site are used as molecular weight markers. The sixth C in the BamHI sequence GGATCC is the 44th nucleotide in the ladder.

PstI, PaeR7 and DpnII enzymes methylate adenine residues, the others are cytosine methylases. Partial homologies were reported among the cytosine methylases BsuRI, BspRI, SPR and ϕ3T (40,45). A computer search of the published methylase sequences did not identify significant homology between EcoRII methylase and any of the above mentioned adenine methylating enzymes. But comparison with known cytosine methylase sequences indicated some degree of similarity with the EcoRII methylase (Figure 6). Although the homologies between the EcoRII enzyme and any of these three enzymes are weaker than

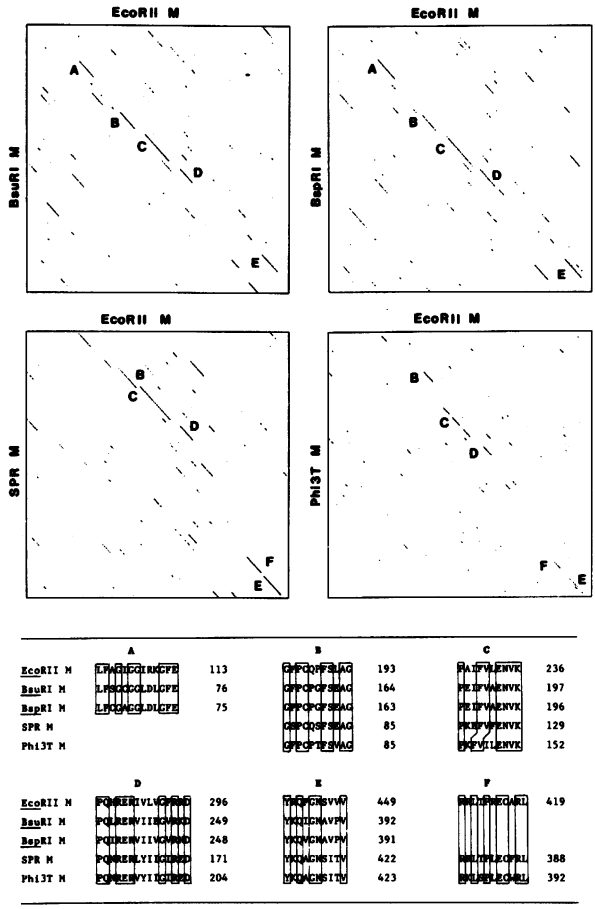

Figure 6. Comparison of amino acid sequence of EcoRII methylase with BsuRI, BspRI, SPR and ϕ3T methylases. At the top is a comparison of the sequences by a Dot Matrix evaluation using the program DIAGON (19) based on Dayhoff's similarity score between amino acids. Regions of partial homology are represented by diagonal stretches. Regions where partial homologies are most apparent are marked and the sequences in these regions are shown in the bottom half. The conserved amino acids are boxed. The comparison was done using the program SEQHP (20).

the similarities between the latter, the regions where partial homology is most apparent are aligned in identical fashion in all of them. However, in the EcoRII methylase protein, these domains are shifted slightly toward the C-terminus of the molecule. Comparison with the SPR methylase reveals

a region where the DNA homology is also quite apparent:

EcoRII M	TTTGTTCTTGAAAACGTTAAA	916
	:::::: ::::::: ::::::	
SPR M	TTTGTTTTTGAAAATGTTAAA	674

A striking similarity was found between the protein sequences of EcoRII methylase and MspI methylase (Lin, P.M. and Roberts, R.J., manuscript in preparation). While some of the regions common in EcoRII.M, BsuRI.M, BspRI.M, SPR.M and ϕ3T.M are also present in the MspI.M sequence, EcoRII and MspI methylases have greater overall homology with each other (data not shown).

A global search for homology in protein sequences carried out between EcoRII methylase and sequences in the PIR, Doolittle and translated Genbank data bases did not reveal significant homology with any other sequences.

Effect of N-Terminal Deletion on Enzyme Activity

Partial methylase activity was found in pSS69A (Figure 2), in which the insert was shortened by Bal-31 deletion of 101 bp from the proposed translational start codon (position 208, Figure 4) and cloned in pUC9, downstream to the lac promoter. Sequence analysis revealed that the fragment was fused in-frame with the N-terminal lacZ sequence of pUC9 adding 11 amino acids at position 35 of the methylase protein :

```
                                  Hind III   Pst I
pUC9....ATG ACC ATG ATT ACG CCA AGC TTG GCT GCA GCG GAA ATC TAT GAC....RII.M
        M   T   M   I   T   P   S   L   A   A   A   E   I   Y   D
                                                    35  36  37  38
```

The activity was one fourth that of pSS59 (see Table 2). Thus the N-terminal 34 amino acids are not necessary for enzyme activity.

Cloning of the EcoRII Methylase under control of P_L-λ Promoter

Although we found an increased expression of methylase by cloning the gene under the influence of the lac promoter in pUC vectors, in order to further increase the yield of the enzyme we cloned it downstream to the strong inducible P_L promoter of bacteriophage λ. Figure 7 shows the construction of pSS113 and pSS114 in which the position of the gene downstream to the P_L promoter are opposite to one another in orientation. When transformed into E. coli N4830 containing the temperature sensitive cI857 repressor, only the strain carrying pSS114 could be induced to express methylase at 42°C.

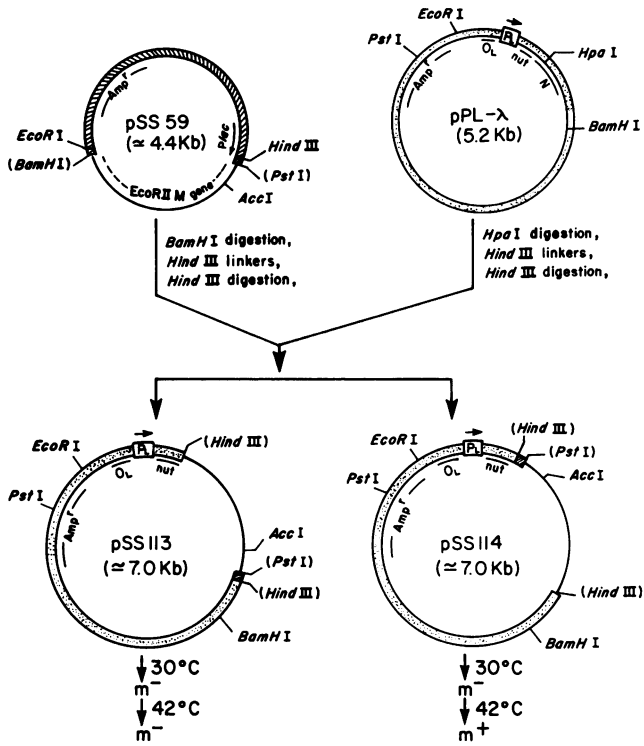

Figure 7. Cloning strategy of the EcoRII methylase gene in expression vector pPL-λ . The thin line represents the insert region. Linker created sites are those shown in parentheses.

Expression of Gene Product

In vivo protein synthesis was monitored using ^{35}S-methionine pulse labeling in *E. coli* N4830 harboring recombinant plasmids pSS113 and pSS114. Extracts were loaded onto an SDS-polyacrylamide gel before and after thermal induction. A protein of apparent molecular weight 60 kd was found in the 42°C extract from the cells bearing pSS114 (the recombinant properly oriented for P_L-directed transcription) (Figure 8). This confirms both the direction of transcription of the methylase gene and the size of the protein deduced from the DNA sequence data.

Assessment of Methylase Activity in Different Constructions

DNA methylase activity was determined in whole cell extracts of HB101 and N4830 containing different plasmids described in Figures 1, 2 and 7. Results are given in Table 2. Cloning of the restriction-modification genes from the low copy number plasmid N3 into the multicopy plasmid pBR322 (pSF1,

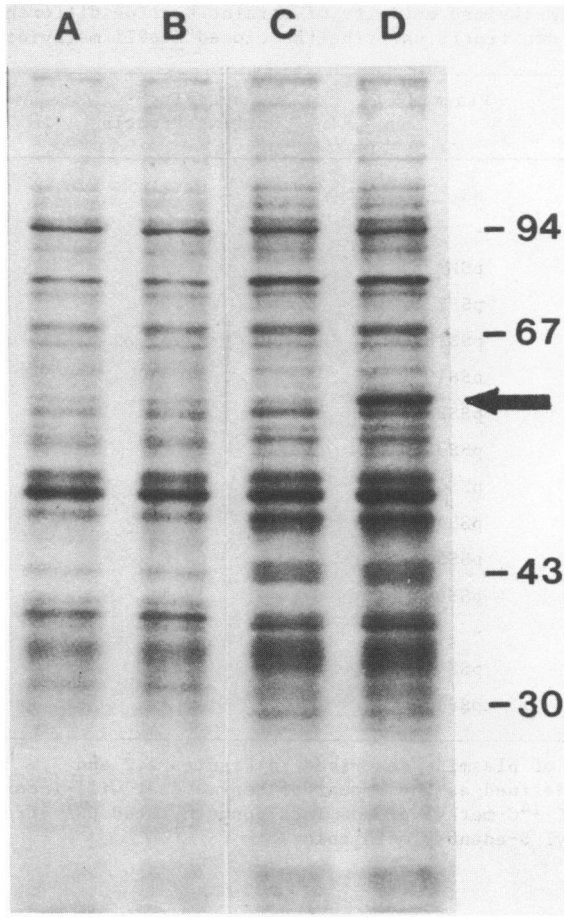

Figure 8. Expression of the EcoRII methylase gene. Cultures of TS mutant strain E. coli N4830 harboring plasmid pSS113 (lanes A & C) or pSS114 (lanes B & D) were labeled with ^{35}S-methionine at 30°C (A & B) or after an induction of 90 min at 42°C (C & D). Extracts were loaded onto a 5-15% gradient SDS-polyacrylamide gel. The autoradiogram shows bands of ^{35}S-labeled proteins. Molecular weight markers are indicated. The band indicated by an arrow corresponds to a molecular weight of 60,000.

Figure 1) resulted in only a small increase (1.8 fold) in methylase activity. Removal of the endonuclease gene (pSF2, Figure 1) did not increase methylase expression. Enzyme activity was increased by removing approximately 800 bp and cloning into pUC vectors (pSS18 and pSS19). A much greater increase occured with pSS59 (Figure 2), in which the methylase gene was placed under the control of the lac promoter. Sequence analysis of the 5' end of the

Table 2. DNA methylase activity of strains bearing different plasmid constructs carrying the cloned EcoRII methylase gene

Strain	Plasmid[a]	Activity[b] Unit/mg Protein	Amplification
F1100	N3	4.5	1
HB101	-	0.7	-
HB101	pSF1	7.9	1.8
HB101	pSF2	8.0	1.8
HB101	pSS18	27.5	6.1
HB101	pSS19	31.5	7.0
HB101	pSS29	47.2	10.5
HB101	pSS38	33.3	7.4
HB101	pSS48	1.0	-
HB101	pSS49	53.3	11.8
HB101	pSS59	80.0	17.8
HB101	pSS69A	19.5	4.3
N4830	-	0.4	-
N4830	pSS113	1.1	-
N4830	pSS114	176.7	39.3

[a]Construction of plasmids described in Figures 1,2 and 7. [b]One unit of enzyme is defined as the amount of enzyme that will incorporate 1 pmole/min of ^{14}C-methyl groups into nonmethylated DNA (from E.coli B) from ^{14}C-methyl S-adenosylmethionine.

insert in pSS59 revealed that the fragment was ligated to the PstI site of pUC19 DNA at the beginning of the Shine-Dalgarno sequence.

Cloning the same fragment under control of the P_L promoter of bacteriophage λ(recombinant pSS114, Figure 7) gave a further 2.2 fold increase in enzyme activity. Overall, the methylase activity of N4830 (pSS114) is 40 fold higher than that of the parental plasmid N3 in strain E. coli F1100.

DISCUSSION

The EcoRII modification enzyme has been sequenced. One open reading frame of 1515 nucleotide pairs has been found which can code for a protein of 477 amino acids and a calculated molecular weight of 54,564 daltons. This is in good agreement with the reported molecular weight of the EcoRII methylase; 55,000 when determined by Sephadex G-100 chromatography (28) and

60,000 by SDS-PAGE (5). Using the hybrid plasmid pSS114 which contains a λ-P_L promoter, we found expression of a gene product of apparent molecular weight of 60,000 (determined by SDS-PAGE) after heat induction. The identity of this product as the EcoRII methylase is supported by: 1) the only difference between the proteins synthesized by strains carrying plasmid pSS114 or pSS113 is the 60,000 MW band (Figure 8); 2) emergence of this band after temperature induction is concomitant with the expression of enzyme activity measured by incorporation of methyl group into DNA.

The assignment of the reading frame is consistent with the activity of a deletion mutant at the N-terminus of the sequence. Deletion of 101 bp from ATG start codon resulted in a clone pSS69A which was found to retain one fourth the enzyme activity of the parent plasmid pSS59. The assigned reading frame was found to be in-frame with the lac sequence in that construct. The partial activity of this deletion mutant is likely to be due to a protein fusion between the N-terminal 11 amino acids of β-galactosidase and 443 C-terminal amino acids of the methylase. The first 34 amino acid residues are therefore not essential for enzymatic function.

The region for which sequence has been determined extends 292 bp downstream from the methylase termination codon. Since the EcoRII methylase and endonuclease genes lie adjacent to each other (ref. 27 and Figure 1), a part of the latter might be included in the sequence presented in this paper. Inactivation of the endonuclease gene by insertion of transposon Tn3 at a position approximately 300 bp downstream from the StuI site at 1489 (Figure 4) (A.B., unpublished data) suggests this possibility. The only reading frame that extends to the end of the sequence we have determined is on the strand complementary to the methylase gene strand. If this turns out to be the correct reading frame for the endonuclease, the two EcoRII genes are coded on opposite strands with only 33 bp between the 3'-ends of the two genes.

A number of DNA methylases have been sequenced. These include both adenine and cytosine methylases. Similar to other cytosine methylating modification enzymes (39-42,45), the EcoRII enzyme does not bear any noticeable homology with any of the adenine methylating enzymes. The cytosine methylases that have been sequenced are all from the species Bacillus or from phages of this species. Moreover, all of them recognize and methylate the sequence GGCC. Not surprisingly, strong similarities were found in their primary protein structures. In this paper, we have reported that EcoRII shares several partially homologous domains with methylases that do not recognize the same

sequence as EcoRII. These enzymes include BsuRI, BspRI and ϕ3T methylases. EcoRII methylase is also partially homologous to the SPR methylase. The latter enzyme recognizes the sequence CC(A/T)GG, as is recognized by EcoRII in addition to the sequences GGCC and CCGG (45). Thus all cytosine methylases for which primary sequences are known share several common domains.

As some of the homologous domains are shared by enzymes recognizing very dissimilar DNA sequences (EcoRII & BsuRI, for example), these regions are unlikely to contain the sequences required for their site-specific interaction with DNA. As such, they may represent regions involved in methylation of the position 5 of cytosine or in binding of S-adenosyl methionine. Site-directed mutagenesis of these regions may help in the determination of the functional role of these regions.

Purification of large quantities of EcoRII methylase should aid structural and enzymatic studies of the protein . For this purpose, we have made plasmid constructs that substantially overproduce the enzyme. As no methylases have been crystallized, the crystallization of EcoRII methylase protein and its structural analysis should be of great value in understanding how methylases function.

ACKNOWLEDGEMENTS

The authors want to express their gratitude to Dr. R.J. Roberts who made his laboratory available for part of this work and to Christopher Keller who performed the computer searches. Part of this work was supported by Grant GM27787 from the NIGMS and NSF Grant DMB 8217553.

REFERENCES

1. May, M.S. and Hattman, S. (1975) J. Bacteriol. 122, 129-138.
2. Roberts, R.J. (1985) Nucl. Acids Res. 13, r165-r200.
3. Friedman, S. (1981) Mol. Pharmacol. 19, 314-320.
4. Santi, D.V., Norment, A. and Garrett, C.E. (1984) Proc. Nat. Acad. Sci. U.S.A. 81, 6993-6997.
5. Friedman, S. (1985) J. Biol. Chem. 260, 5698-5705
6. Friedman, S. (1986) Nucl. Acids Res. 14, 4543-4556.
7. Boyer, H.W. and Rouland-Dussoix, D. (1969) J. Mol. Biol. 41, 459-472.
8. Bolivar, F., Rodriguez, 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.
9. Vieira, J. and Messing, J. (1982) Gene 19, 259-268.
10. Celeste, Y.-P., Vieira, J. and Messing, J. (1985) Gene 33, 103-119.
11. Hattman, S., Schlagman, S. and Cousens, L. (1973) J. Bacteriol. 115, 1103-1107.
12. Gottesman, M.E., Adhya, S. and Das, A. (1980) J. Mol. Biol. 140, 57-75.
13. Watanabe, T., Nishida, H., Ogata, C., Thoshihiko, A. and Sato, S. (1964) J. Bacteriol 86, 1235-1241.
14. Maniatis, T., Fritsch, E.F. and Sambrook. J. (1982) in Molecular Cloning:

A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
15. Hanahan, D. (1983) J. Mol. Biol. 166, 557-580.
16. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 5463-5467.
17. Biggin, M.D., Gibson, T.J. and Hong, G.F. (1983) Proc. Natl. Acad. Sci.U.S.A. 80, 3963-3965.
18. Delaney, A.D. (1982) Nucl. Acids Res. 10, 61-67.
19. Staden, R. (1982) Nucl. Acids Res. 10, 2951-2961.
20. Goad, W.B. and Kanehisa, M.I. (1982) Nucl. Acids Res. 10, 247-263.
21. Brosius, J., Cate, R.L. and Perlmutter, A.P. (1982) J. Biol. Chem. 257, 9205-9210.
22. Maxam, A.M. and Gilbert, W. (1980) in Methods in Enzymology, Grossman, L. and Moldave, K., ed., Vol 65, Part I, pp 499-560, Academic Press, New York.
23. Friedman, S. (1982) J. Bacteriol. 151, 262-268.
24. Lowry,O.H., Rosenbrough, N.J., Farr, A.L. and Randall, R.J. (1951) J. Biol. Chem. 193, 265-275.
25. Rosenberg, M., Ho, Y.-S. and Shatzman, A. (1983) in Methods in Enzymology, Wu, R., Grossman, L. and Moldave, K., ed., Vol 101, Part C, pp 123-138, Academic Press, New York.
26. Takano, T., Watanabe, T. and Fukasawa, T. (1968) Virology 34, 290-302.
27. Kosykh, V.G., Buryanov, Y.I. and Bayev, A.A. (1980) Mol. Gen. Genet. 178, 717-718.
28. Buryanov, Y.I., Nesterenko, V.F., Kosykh, V.G. and Bayev, A.A. (1981) Dokl. Akad. Nauk. SSSR. 257, 495-497.
29. Newman, A.K., Rubin, R.A., Kim, S. and Modrich, P. (1981) J. Biol. Chem. 256, 2131-2139.
30. Greene, P.J., Gupta, M., Boyer, H.W., Brown, W.E. and Rosenberg, J.M. (1981) J. Biol. Chem. 256, 2143-2153.
31. Bougueleret, L., Schwarzstein, M., Tsugita, A. and Zabeau, M. (1984) Nucl. Acids Res. 12, 3659-3676.
32. Shapiro, H.S. (1970) in CRC Handbook of Biochemistry, Selected Data for Molecular Biology, Sober, H.A., ed., 2nd Ed., pp H80-H88, The Chemical Rubber Co., Cleveland.
33. Grantham, R., Gautier, C., Gouy, M., Jacobzone, M. and Mercier, R. (1981) Nucl. Acids Res. 9, r43-r74.
34. Gouy, M. and Gautier, C. (1982) Nucl. Acids Res. 10, 7055-7074.
35. Rosenberg, M. and Court, D. (1979) Ann. Rev. Genet. 13, 319-353.
36. Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, J.C., Hutchinson, C.A., Slocombe, P.M. and Smith, M. (1977) Nature 265, 687-695.
37. Walder, R.Y., Walder, J.A. and Donelson, J.E. (1984) J. Biol. Chem. 259, 8015-8026.
38. Schoner, B., Kelly, S. and Smith, H.O. (1983) Gene 24, 227-236.
39. Posfai, G., Kiss, A., Erdei, S., Posfai, J. and Venetianer, P. (1983) J. Mol. Biol. 170, 597-610.
40. Kiss, A., Posfai, G., Keller, C.C., Venetianer, P. and Roberts, R.J. (1985) Nucl. Acids Res. 13, 6403-6421.
41. Buhk, H.J., Behrens, B., Tailor, R., Wilke, K., Prada, J.J., Gunthert, U., Noyer-Weidner, M., Jentsch, S. and Trautner, T.A. (1984) Gene 29, 51-61.
42. Posfai, G., Baldauf, E., Erdei, S., Posfai, J., Venetianer, P. and Kiss, A. (1984) Nucl. Acids Res. 12, 9039-9049.
43. Theriault, G., Roy, P.H., Howard, K.A., Benner, J.S., Brooks, J.E., Waters, A.F. and Gingeras, T.R. (1985) Nucl. Acids Res. 13, 8441-8461.

44. Mannarelli, B., Balganesh, T.S., Greenberg, B., Springhorn, S.S. and Lacks, S.A. (1985) Proc. Natl. Acad. Sci. U.S.A. 82, 4468-4472.
45. Tran-Betcke, A., Behrens, B., Noyer-Weidner, M. and Trautner, T.A. (1986) Gene 42, 89-96.
46. Brooks, J.E., Blumenthal, R.M. and Gingeras, T.R. (1983) Nucl. Acids Res. 11, 837-851.