

Genetic organization of the *KpnI* restriction – modification system

Deb K.Chatterjee, Alan W.Hammond, Robert W.Blakesley, Shelia M.Adams and Gary F.Gerard
Life Technologies, Inc., 8717 Grovemont Circle, Gaithersburg, MD 20877, USA

Received August 14, 1991; Revised and Accepted October 31, 1991

GenBank accession no. M76435

ABSTRACT

The *KpnI* restriction-modification (*KpnI* RM) system was previously cloned and expressed in *E. coli*. The nucleotide sequences of the *KpnI* endonuclease (*R·KpnI*) and methylase (*M·KpnI*) genes have now been determined. The sequence of the amino acid residues predicted from the endonuclease gene DNA sequence and the sequence of the first 12 NH₂-terminal amino acids determined from the purified endonuclease protein were identical. The *kpnIR* gene specifies a protein of 218 amino acids (MW: 25,115), while the *kpnIM* gene codes for a protein of 417 amino acids (MW: 47,582). The two genes transcribe divergently with a intergenenic region of 167 nucleotides containing the putative promoter regions for both genes. No protein sequence similarity was detected between *R·KpnI* and *M·KpnI*. Comparison of the amino acid sequence of *M·KpnI* with sequences of various methylases revealed a significant homology to N6-adenine methylases, a partial homology to N4-cytosine methylases, and no homology to C5-methylases.

INTRODUCTION

Type II restriction endonucleases and methylases are currently subjects of great interest. A number of methylase and endonuclease genes have been cloned and their nucleotide sequences determined [1–6]. Thus far three types of DNA modifications by methylases have been observed: methylation of cytosine at the 5 position (^{m5}C) or N4-position (^{N4m}C) and methylation of adenine at the N6 position (^{N6m}A) [7–11]. Conserved amino acid sequences have been found within these individual groups of methylases and also among all three groups [1,2,4,5,12]. However, ^{N4m}C methylases are now thought to be a part of the ^{N6m}A methylase family [4,13].

More sequence information is available for ^{m5}C methylases than for the other two types: a pro-cys pair has been proposed as the catalytic site [14] and the domain responsible for sequence specific DNA recognition has been identified [15]. More sequence information for ^{N6m}A and ^{N4m}C methylases is necessary to understanding the structure-function relationships of these enzymes. One of the recently cloned ^{N6m}A methylases is *M·KpnI* [16,17]. The recognition sequence of the *KpnI* RM system is 5'GGTACC3' [18]. *R·KpnI* cleaves the sequence to produce 4-base 3' sticky ends. The *KpnI* RM system has been

cloned and the approximate locations of both genes have been reported [16]. In this paper we described the genetic organization and the nucleotide sequence of the *KpnI* RM system. The amino acid sequence of *M·KpnI* deduced from the nucleotide sequence was compared with those of other methylases.

MATERIALS AND METHODS

Bacterial strains, plasmids and phages

Klebsiella pneumoniae strain ATCC55014 was used for cloning the *KpnI* RM system. Competent *E. coli* strains DH10B, DH5 α mc_r, and DH5 α F'IQ were obtained from Life Technologies, Inc. Plasmids pUC19 [19], pJJS1010 [20], and pBR322 [21] were used as cloning and expression vectors. M13mp18 and M13mp19 [19] were used for DNA sequencing.

Enzymes and chemicals

Restriction enzymes, Klenow fragment of *E. coli* DNA polymerase I, Taq DNA polymerase, and T4 DNA ligase were obtained from Life Technologies, Inc. T7 DNA polymerase (Sequenase) used for DNA sequencing was obtained from United States Biochemicals. [α ³⁵S]dCTP (>3000Ci/mmol) was purchased from Amersham. All other chemicals were of reagent grade.

DNA preparation

Total cellular DNA from 5 g of frozen *K. pneumoniae* was isolated by the method of Marmur [22]. Small- and large-scale plasmid DNA preparations were done by the alkaline-lysis procedure [23]. All other procedures were adapted from standard protocols [24].

Expression of *R·KpnI*

E. coli AH29 was grown in 2% tryptone, 1% yeast extract and 0.5% NaCl in the presence of 100 μ g/ml ampicillin and 50 μ g/ml spectinomycin at 30°C until the absorbance at 590 nm was between 1.0 and 2.0. The temperature of the culture was then raised to 37°C to maximize *R·KpnI* production due to runaway replication of p*KpnI*R. The cells were harvested after 2 h incubation at 37°C. The cells were stored at –70°C until used for *KpnI* purification.

Purification of *R·KpnI*

Eight grams of *E. coli* AH29 was suspended in 32 ml of suspension buffer containing 20 mM Tris-HCl, pH 7.5, 0.1 mM

EDTA, 0.1 mM PMSF and 10 mM β -mercaptoethanol (β ME) and sonicated. The cell debris was removed by centrifugation for 30 min at 14,000g. To the supernatant, freshly made 10% (w/v) streptomycin sulfate was added to a final concentration of 1%. After removal of a precipitate by centrifugation at 14,000 g for 30 min, solid ammonium sulfate was added to the supernatant to 53.5% (w/v) at 4°C. A pellet was recovered by centrifugation and dissolved in 10 ml of suspension buffer, and the suspension was dialyzed against 4 liters of the same buffer. The sample was loaded onto a heparin-agarose column (70 ml column volume) and the column was developed with a 700 ml linear NaCl gradient from 0.2 M to 0.8 M in suspension buffer. The fractions containing *KpnI* activity without any significant exonuclease activity were pooled and dialyzed against 4 l of Mono-Q buffer (same as suspension buffer except 10 mM β ME was replaced by 2 mM DTT). The sample was loaded onto a Mono-Q (Pharmacia, HR10/10) column and eluted with a linear gradient from 0.08 M to 0.48 M NaCl in Mono-Q buffer. The fractions containing *KpnI* activity without any detectable exonuclease activity were pooled and dialyzed against 10 mM Tris-HCl, pH 7.4, 0.1 mM EDTA, 0.2 M KCl, 1 mM DTT, and 50% glycerol. Most of the protein in the final pool was contained in a single polypeptide as determined by 10% SDS-PAGE.

N-terminal analysis of *KpnI* endonuclease

An aliquot of the final pool was purified through a second heparin-agarose column as above giving a homogenous protein which was used for N-terminal amino acid sequence analysis. Forty micrograms protein from an SDS gel of the enzyme was blotted onto a polyvinylidene difluoride (PVDF) membrane (Applied Biosystems, Inc.). The section of the membrane containing the band was cut out and loaded into an Applied Biosystems 470A sequencer for analysis. The phenylthiohydantoin (PTH) derivatives of the amino acids were identified with an Applied Biosystems 120A PTH analyzer connected to the sequencer.

DNA sequence determination and analysis

DNA sequencing was performed with Klenow fragment and Sequenase using the dideoxy chain termination method [25], and with Taq DNA polymerase using cycle sequencing [29]. DNA fragments to be sequenced were subcloned into M13mp18 and M13mp19. To ensure an unambiguous sequence, DNA templates

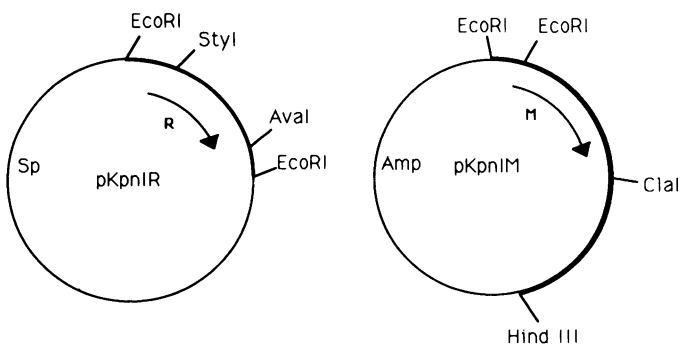


Fig. 1. Plasmid maps of *pKpnIM* and *pKpnIR*. Construction of these two plasmids is described elsewhere [16]. *pKpnIM* was constructed by subcloning the methylase gene in pBR322 while *pKpnIR* contains the endonuclease gene in runaway plasmid pJJS1010 [20]. The direction of transcription of both genes is shown by an arrow. Both the methylase gene and the endonuclease gene are transcribed from their own endogenous promoter.

with significant overlap were sequenced and compared. The DNA sequence and the predicted amino acids were analyzed by Clone Manager 3 (Scientific and Educational Software) and PCGene. The amino acid sequence of *M·KpnI* and amino acid sequences of methylases were compared with the help of PCGene. The dot matrix analysis was done with the program DNA STAR.

RESULTS AND DISCUSSION

Cloning of the *kpnIRM* genes

The details of cloning the *kpnIRM* genes have been published elsewhere [16]. Final constructs containing the *kpnIM* and *kpnIR* genes are shown in Figure 1. *E. coli* AH29, deficient in *mcrA*, *mcrB* and *mrr* and harboring *pKpnIM* and *pKpnIR*, was used to produce the endonuclease. However, *pKpnIM* or *pKpnIR* and *pKpnIR* together can be stably maintained in *E. coli* proficient in *mcrA*, *mcrB* and *mrr*, because *M·KpnI* mediated modification of GGTACC sequences is not restricted by these systems [16].

Expression and purification of *R·KpnI*

KpnI was overexpressed and purified as described in Materials and Methods. Figure 2 shows an SDS-PAGE gel of the protein

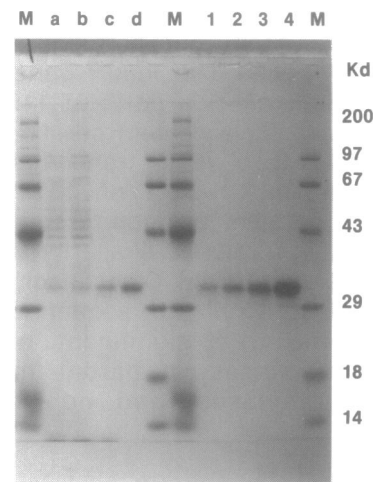


Fig. 2. Sodium dodecyl sulfate polyacrylamide gel electrophoresis of *R·KpnI* at different stages of purification. Lane a, crude extract; b, ammonium sulfate fraction; c, heparin-agarose pooled fraction; and d, Mono-Q pooled fraction. Lanes 1–4: 2, 5, 10 and 20 μ g of purified (Mono-Q fraction) *R·KpnI*. Lane M, molecular weight markers: myosin (Mr = 200 Kd), phosphorylase B (97 Kd), bovine serum albumin (67 Kd), ovalbumin (43 Kd), carbonic anhydrase (29 Kd), β -lactoglobulin (18 Kd) and lysozyme (14 Kd).

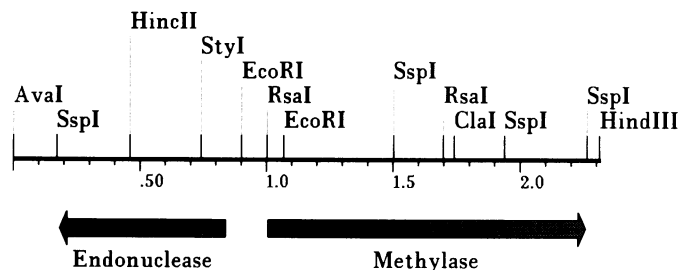


Fig. 3. Organization of the *KpnI* RM system. Restriction map of the cloned region of *K. pneumoniae* genomic DNA indicating the position and orientation of the endonuclease and methylase genes.

Domain I											Domain II																				
KpnI	L	I	L	D	F	F	A	G	S	G	T	T	A	H	A	KpnI	I	D	F	C	Y	I	D	P	P	Y	N	T	G	N	K
EcoP15	I	<u>V</u>	L	D	F	F	A	G	S	G	T	T	A	H	T	EcoP15	<u>V</u>	K	<u>M</u>	I	Y	I	D	P	P	Y	N	T	G	K	D
EcoP1	L	<u>V</u>	L	D	F	F	A	G	S	G	T	T	A	E	A	EcoP1	<u>V</u>	N	<u>M</u>	I	Y	I	D	P	P	Y	N	T	G	K	D
HinfI	<u>V</u>	<u>V</u>	L	D	P	F	F	G	T	G	T	T	G	A	<u>V</u>	HinfI	I	D	L	I	F	A	D	P	P	Y	F	M	Q	T	E
EcaI	L	<u>V</u>	I	D	P	F	C	G	S	G	<u>S</u>	T	L	H	A	EcaI	<u>A</u>	K	L	I	Y	L	D	P	P	Y	A	T	G	M	G
T4 dam	R	E	<u>V</u>	D	L	F	C	G	<u>G</u>	L	<u>S</u>	V	S	L	N	T4 dam	G	D	F	V	Y	<u>V</u>	D	P	P	Y	L	I	T	V	A
EcoRV	<u>V</u>	<u>W</u>	<u>V</u>	E	P	F	M	G	T	G	V	V	A	F	N	EcoRV	D	D	<u>V</u>	V	Y	C	D	P	P	Y	I	<u>G</u>	R	H	V
Ecodam	C	<u>L</u>	<u>V</u>	E	P	F	<u>V</u>	G	A	G	<u>S</u>	V	E	L	N	Ecodam	<u>A</u>	S	<u>V</u>	V	Y	C	D	P	P	Y	A	P	L	<u>S</u>	A
DpnII	R	Y	E	E	P	F	<u>V</u>	G	<u>G</u>	G	A	L	E	F	D	DpnII	G	D	F	V	Y	E	D	P	P	Y	I	P	L	<u>S</u>	E
HhaII	L	I	L	D	P	<u>A</u>	S	G	<u>G</u>	<u>Y</u>	<u>S</u>	V	E	E	C	HhaII	<u>V</u>	K	I	A	F	E	D	P	Q	Y	R	<u>G</u>	V	L	D
BamHI	<u>V</u>	<u>V</u>	L	D	P	F	G	G	S	G	T	T	E	A	<u>V</u>	BamHI	<u>V</u>	D	T	I	F	<u>A</u>	D	P	P	F	N	L	D	K	E
BamHI viral	I	<u>V</u>	E	D	P	F	M	G	S	G	T	T	A	K	M	BamHI viral	I	D	L	T	V	T	S	P	P	Y	D	D	L	R	N
Cfr9I	Y	<u>V</u>	L	D	P	F	F	G	S	G	T	V	G	V	<u>V</u>	Cfr9I	<u>V</u>	R	C	I	V	T	S	P	P	Y	W	<u>G</u>	L	R	D
SmaI	K	<u>V</u>	L	D	P	F	F	G	S	G	T	T	G	V	<u>V</u>	SmaI	<u>E</u>	D	C	V	V	T	S	P	P	Y	W	<u>G</u>	L	R	D
MvaI	K	<u>L</u>	Y	D	P	F	M	G	S	G	T	<u>S</u>	L	V	E	MvaI	<u>E</u>	D	I	V	V	T	S	P	P	Y	<u>G</u>	D	<u>S</u>	K	T
PvuII	L	<u>V</u>	<u>V</u>	D	I	F	G	S	N	T	T	G	L	<u>V</u>	PvuII	I	S	<u>L</u>	V	M	T	S	P	P	F	A	L	Q	R	K	

Fig. 4. Homologous domains of ^{N6m}A methylases and ^{N4m}C methylases. Consensus sequences of ten ^{N6m}A methylases (upper panel) and six ^{N4m}C methylases (lower panel) are shown. Amino acids that are identical to those in M·KpnI are in bold letters and amino acids that are functionally similar are underlined. Published sequences of the following methylases were used: EcoP15 and EcoPI [28], HinfI [30], EcaI [31], T4dam [32], EcoRV [33], Ecodam [34], DpnII [35], HhaII [36], BamHI [37], BamHI viral [38], Cfr9I and MvaI [4], SmaI [39], and PvuII [13].

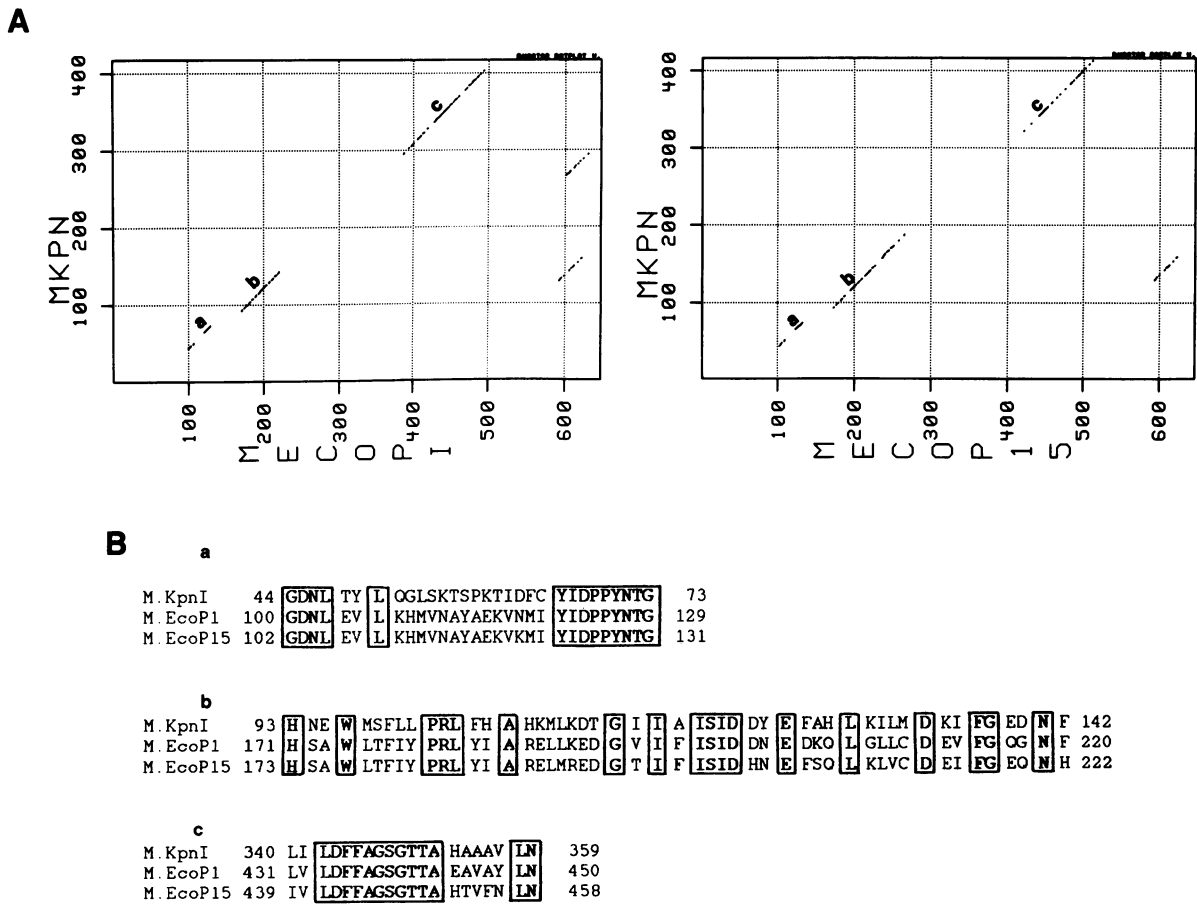


Fig. 5. Comparison of amino acid sequence of M·KpnI, M·EcoP1 and M·EcoP15. (A) Dot matrix analysis of these three methylases generated by the program DNA STAR. Window size and stringency values were 30 and 33, respectively. (B) Amino acid sequences in three strongly homologous regions (a-c). Amino acids identical to those in M·KpnI are indicated in boxes.

at different stages of purification. The final step yielded nearly homogeneous R·*KpnI*. Total activity obtained was 10^7 units per g of cells, which was 1000-fold more than obtained from *K. pneumoniae*. Fifty milligrams of purified *KpnI* were obtained from 12 g of *E. coli* AH29. The molecular weight (Mr) of *KpnI* purified from the clone was estimated from SDS-PAGE to be approximately 31,000 daltons.

Nucleotide sequence of the *KpnI* RM System

Previously we determined the physical map and the approximate boundaries of the *kpnIM* and *kpnIR* genes within cloned fragments [16]. The genetic organization of the genes and nucleotide sequence of the 2316 bp *AvaI-HindIII* fragment containing the *KpnI* RM system have been determined (Figure 3).

Analysis of the predicted amino acid sequence demonstrated two large open reading frames (ORF), presumably corresponding to the methylase and endonuclease genes. The two genes transcribe divergently. The shorter ORF of 654 bp (ATG⁸³⁸ to TAA¹⁸⁴) codes for a protein of 218 amino acids. Twelve N-terminal amino acids of purified R·*KpnI* determined by sequence analysis, met-asp-val-phe-asp-lys-val-tyr-ser-asp-asp-asn, matched the amino acid sequence deduced from the DNA sequence starting at ATG⁸³⁸. Therefore, the shorter ORF was assigned to the endonuclease. The predicted molecular weight (Mr) of R·*KpnI* from the nucleotide sequence is 25,112 daltons. This value is less than the 31,000 daltons derived by SDS-PAGE analysis. The discrepancy could be due to an abnormal mobility for the endonuclease in SDS-PAGE.

The longer ORF of 1251 bp (ATG¹⁰⁰⁶ to TAG²²⁵⁶) was assigned to M·*KpnI* because several clones containing this region expressed methylase but not endonuclease activity. In addition, deletion derivatives within this ORF displayed no methylase activity. This ORF codes for a protein 417 amino acids in length (Mr: 47,582).

Both open reading frames are preceded by consensus ribosome binding sites (RBS) and promoter-like sequences. Five nucleotides upstream of the initiation codon of the *kpnIR* gene is the sequence 5'-AAGGAGGT^{3'}, which is predicted to be the RBS [26]. This sequence is complementary to the 3'-end of 16S RNA of *E. coli*. A search for a RBS for the *kpnIM* gene showed a sequence 5'-AGGAGG^{3'} four nucleotides upstream of the ATG initiation codon. It is interesting that both genes have long ribosome binding sites. However, higher complementarity of the RBS of the *kpnIR* gene (8 bases) than the *kpnIM* gene (6 bases) with 16S RNA may result in more efficient translation of the *kpnIR* gene. That may explain why the *kpnIRM* system cannot be cloned in one step [16]. The *kpnIM* gene must be cloned first to ensure proper methylation (protection) of the host DNA before expression of the *kpnIR* gene. However, other factors such as promoter strength, mRNA stability and enzyme turnover may also influence the supremacy of the *kpnIR* gene expression.

Nucleotide composition and codon usage

It is interesting to note that despite the low genomic A and T composition (43%) of *K. pneumoniae*, which produces M·*KpnI* and R·*KpnI* [11], the A+T composition of the *kpnIR* and *kpnIM* gene is about 62% and 66%, respectively. Moreover, the utilization of A and T at the third position of codons is about 73% and 74% for *kpnIR* and *kpnIM* gene, respectively. The reason for the relatively high A+T content of these genes is not clear.

Comparison of amino acid sequences

The amino acid sequences of the *KpnI* proteins were compared with other known endonucleases and methylases. As expected, no significant homology was detected between R·*KpnI* and M·*KpnI* sequences. Similarly, no sequence homology could be detected between R·*KpnI* and other endonucleases. This is not unusual, because except for *EcoRI* and *RsrI* [27], which are isoschizomers, no significant homology has been detected among the endonucleases.

Comparison of the M·*KpnI* sequence to the sequences of several ¹⁴C methylases showed no significant homology. M·*KpnI* is an N6-adenine specific methylase [16,17]. In all N6-adenine specific methylases analyzed to date, there are at least two common homologous domains. The tetrapeptide DPPY is suggested to be involved in the methylation of the exocyclic amino group, and a FXGXG motif is thought to be the binding site of S-adenosyl methionine [4]. As expected, these two conserved domains were detected within the M·*KpnI* amino acid sequence (Figure 4). Relatedness between M·*KpnI* and other N6-adenine and N4-cytosine methylases around the two common regions is shown in Figure 4. Dot matrix comparisons were done between M·*KpnI* and several N6-adenine and N4-cytosine specific methylases. Strong homology was detected in the DPPY and FXGXG regions of most N6-adenine methylases and only the FXGXG region of N4-cytosine methylases. However, similar comparison of M·*KpnI* with M·*EcoP15* and M·*EcoP1* showed additional homologous regions (Figure 5A). The amino acid sequences of three strongly homologous regions are shown in Figure 5B. It is interesting that the M·*KpnI* sequence is more closely related to *EcoP15* and *EcoP1* methylases, which belong to type III RM systems. A stretch of 4(GDNL) and 9(YIDPPYNTG) amino acids of *EcoP15* and *EcoP1* are perfectly matched to M·*KpnI* in Region a. In addition, a stretch of 12 and 11 amino acids of *EcoP15* and *EcoP1* methylase, respectively, are identical to M·*KpnI* Region c (Figure 5B). A significant homology in Region b was also found. The function of this region is unknown. Sequence recognition for *EcoP15* and *EcoP1* methylase is 5'-CAGCAG-3' and 5'-AGACC-3' [28], respectively, and the endonucleases cut some 25 residues to the right of the recognition sequence. It is possible that M·*KpnI*, M·*EcoP1*, and M·*EcoP15* are of common ancestral origin and have common general mechanism by which they catalyze the transfer of a methyl group.

ACKNOWLEDGEMENTS

We like to thank Ira Palmer for purification of *KpnI*, Dr. William Lane (Harvard Microchem) for the protein sequencing, Molly Craxton for providing the Cycle Sequencing protocol prior to publication and Jill Bradshaw for preparation of the manuscript.

REFERENCES

- Chandrasegaran, S., and Smith, H.O. (1988) in *Structure and Expression*. Vol. 1, From Proteins to Ribosomes. eds. R.H. Sarma and M.H. Sarma. (Adenine Press) pp. 149-156.
- Posfai, J., Bhagwat, A.S., Posfai, G., and Roberts, R.J. (1989) *Nucl. Acids Res.* **17**, 2421-2435.
- Som, S., Bhagwat, A.S., and Friedman, S. (1987) *Nucl. Acids Res.* **15**, 313-332.
- Klimasauskas, S., Timinskas, A., Menkevicius, S., Butkiene, D., Butkus, V., and Janulaitis, A. (1989) *Nucl. Acids Res.* **17**, 9823-9832.
- Card, C.O., Wilson, G.G., Weule, K., Hasapes, J., Kiss, A., and Roberts, R.J. (1990) *Nucl. Acids Res.* **18**, 1377-1383.

6. Wilson, G.G. (1988) *Gene* **74**, 281–289.
7. Dunn, D.B., and Smith, J.D. (1958) *Biochem J.* **68**, 627–636.
8. Adams, R.L.P., and Burdon, R.H. (1985) in *Molecular Biology of DNA Methylation*. Springer-Verlag, New York.
9. Janulaitis, A.A., Klimasauskas, S., Petrusyte, M., and Butkus, V. (1983) *FEBS Lett.* **161**, 131–134.
10. Janulaitis, A.A., Stakenas, P., Petrusyte, M., Bitinaite, J., Klimasauskas, S., and Butkus, V. (1984) *Mol. Biol. (Moscow)* **18**, 115–129.
11. Ehrlich, M., Wilson, G.G., Kuo, K.C., and Gehrke, C.W. (1987) *J. Bacteriol.* **169**, 939–943.
12. Lauster, R., Kreibardis, A., and Guschlbauer, W. (1987) *FEBS Lett.* **220**, 167–176.
13. Tao, T., Walter, J., Brennan, K.J., Cotterman, M.M., and Blumenthal, R.M. (1989) *Nucl. Acids Res.* **17**, 4161–4175.
14. Wu, J.C. and Santi, D.V. (1987) *J. Biol. Chem.* **262**, 4778–4786.
15. Balganes, T.S., Reiners, L., Lauster, R., Noyer-Weidner, M., Wilke, K. and Trautner, T.A. (1987) *EMBO J.* **6**, 3543–3549.
16. Hammond, A.W., Gerard, G.F., and Chatterjee, D.K. (1991) *Gene* **97**, 97–102.
17. Kiss, A., Finta, C., and Venetianer, P. (1991). *Nucl. Acids Res.* **19**, 3460.
18. Tomassini, J., Roychoudhury, R., Wu, R., and Roberts, R.J. (1978) *Nucl. Acids Res.* **5**, 4055–4064.
19. Yanisch-Perron, C., Vieira, J., and Messing, J. (1985) *Gene* **33**, 103–109.
20. Sninsky, J.J., Uhlin, B.E., Gustafsson, P., and Cohen, S.N. (1981) *Gene* **16**, 275–286.
21. Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L., Boyer, H.W., Crossa, J.H., and Falkow, S. (1977) *Gene* **2**, 95–113.
22. Murrar, J. (1961) *J. Mol. Biol.* **3**, 208–218.
23. Birnboim, H.C. (1980) *Methods in Enzymol.* **100**, 243–255.
24. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
25. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
26. Shine, J., and Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. USA*, **71**, 1342–1346.
27. Kaszubska, W., Aiken, C., O'Connor, C.D. and Gumpert, R.I. (1989) *Nucleic Acid Res.* **17**, 10403–10425.
28. Hümbelin, M., Suri, B., Rao, D.N., Hornby, D.P., Eberle, H., Pripfl, T., Kenel, S., and Bickle, T.A. (1988) *J. Mol. Biol.* **200**, 23–29.
29. Craxton, M. (1991) *Methods: A companion to Methods in Enzymology*, **1** (in Press).
30. Chandrasegaran, S., Lunnen, K.D., Smith, H.O. and Wilson, G.G. (1988) *Gene* **70**, 387–392.
31. Brenner, V., Venetianer, P. and Kiss, A. (1990) *Nucleic Acids Res.* **18**, 355–359.
32. MacDonald, P.M. and Mosig, G. (1984) *EMBO J.* **3**, 2863–2871.
33. Bougueleret, L. Schwarzstein, M., Tsugita, A. and Zabeau, M. (1984) *Nucleic Acids Res.* **12**, 3659–3677.
34. Brooks, J.E., Blumenthal, R.M. and Gingeras, T.R. (1983) *Nucleic Acids Res.* **11**, 837–851.
35. Mannareli, B.M., Balganes, T.S., Greenberg, B., Springhorn, S.S. and Lacks, S.A. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 4468–4472.
36. Chandrasegaran, S., Wu, L.P., Valdo, E. and Smith, H.O. (1988) *Gene* **74**, 15–21.
37. Vanek, P.G., Connaughton, J.F., Kaloss, W.D. and Chirikjian, J.G. (1990) *Nucleic Acids Res.* **18**, 6145.
38. Connaughton, J.F., Kaloss, W.D., Vanek, P.G., Nardone, G.A. and Chirikjian, J.G. (1990) *Nucleic Acids Res.* **18**, 4002.
39. Heidman, S., Seifert, W., Kessler, C. and Domdey, H. (1989) *Nucleic Acids Res.* **17**, 9783–9796.