

Cloning and nucleotide sequence of the gene encoding the *Eca*I DNA methyltransferase

Vladimir Brenner+, Pál Venetianer and Antal Kiss*
Institute of Biochemistry, Biological Research Center, Hungarian Academy of Sciences, PO Box 521, 6701 Szeged, Hungary

Received November 10, 1989; Accepted December 4, 1989 EMBL accession no. X17111

ABSTRACT

The gene coding for the GGTNACC specific *Eca*I DNA methyltransferase (M.*Eca*I) has been cloned in *E. coli* from *Enterobacter cloacae* and its nucleotide sequence has been determined. The *ecaIM* gene codes for a protein of 452 amino acids (M_r: 51,111). It was determined that M.*Eca*I is an adenine methyltransferase. M.*Eca*I shows limited amino acid sequence similarity to other adenine methyltransferases. A clone that expresses *Eca*I methyltransferase at high level was constructed.

INTRODUCTION

DNA methyltransferases (MTases) fall in three groups according to the methylated base they generate: C5mC-MTases, N4mC-MTases and N6mA-MTases. The genes coding for several MTases have been cloned and sequenced (reviewed in 1). Comparison of the deduced amino acid sequences revealed that C5mC-MTases share extensive regions of homology (2). Similarly, N6mA-MTases were found to contain conserved sequences, most of which are unrelated to the conserved sequences of C5mC-MTases (3–5). Sequence similarity can be detected between M.*Pvu*II, an N4mC-MTase, and N6mA-MTases (6) and there is one motif which seems to be present in all DNA MTases sequenced to date (7,8).

Comparison of the derived amino acid sequences helped to identify the target recognizing domain of multispecific C5mC-MTases (9) and yielded data strongly suggesting the involvement of certain regions in the transmethylation reaction (10), in target recognition (11) and in S-adenosyl-methionine binding (8).

Here we report the cloning from *Enterobacter cloacae* and characterization of a gene encoding a sequence specific DNA methyltransferase. The cloned methyltransferase modifies *Eca*I sites in the DNA of the cloning host *E. coli*, therefore we think it is identical with the methyltransferase of the *Eca*I restriction-modification system (12). We present evidence that M.*Eca*I is an adenine methyltransferase. The clone carrying the *ecaIM* gene does not express *Eca*I endonuclease.

MATERIALS AND METHODS

Strains and Media

Enterobacter cloacae DSM 30056 (12) was obtained from J. Collins. For most plasmid cloning experiments *E. coli* ER1398 (13) was used as host. Cloning with the vector pER23S(−ATG) (T. Lukácsovich, unpublished) was done in *E. coli* JM107 (14) or ER1398 (pVH1). Plasmid pVH1 (15) carries the kanamycin resistance marker and the *lacI*^{Q} gene. M13 phage was grown in JM107. Bacteria were grown in LB medium (16) at 37°C.

Enzymes and chemicals

Restriction endonucleases were either prepared in this institute or were purchased from New England Biolabs. DNA polymerase I large fragment was from Vepex (Szeged) and modified T7 DNA polymerase (Sequenase) from United States Biochemical Corp. Deoxyadenosine 5′-α-(^{35}S)thiotriphosphate was purchased from Amersham.

Cloning methods

Preparation of bacterial genomic DNA, isolation of plasmids, transformation of *E. coli*, restriction mapping, agarose gel electrophoresis and subcloning of DNA fragments were done by standard procedures (16,17).

Determination of the nucleotide sequence

DNA fragments were cloned in mp18 and mp19 phage vectors (18) and were sequenced by the chain termination method (19,20) using deoxyadenosine 5′-α-(^{35}S)thiotriphosphate (21) and either DNA polymerase I large fragment or Sequenase.

SDS-polyacrylamide gel electrophoresis of proteins

Cells were sedimented by centrifugation, then dissolved in 1/5–1/3 volume of sample buffer (22). The extracts were heated to 100°C for 2 min, then run in a 10% SDS-polyacrylamide gel (22). Gels were stained with Coomassie Brilliant Blue R.

* To whom correspondence should be addressed

+Present address: Institute of Microbiology, Department 112, Videnská 1083, Prague 4, 14200, Czechoslovakia

RESULTS AND DISCUSSION

Cloning of the *ecaIM* gene

The *ecaIM* gene was cloned by a method which selects for a 'self-modifying' clone (23,1). This method works best if the vector contains many recognition sites in essential parts of the plasmid. Since the available plasmid vectors did not contain *Eca*I sites, we first constructed a suitable vector. A 4289 bp *Aat*II fragment of lambda phage DNA (nucleotides 5101−9394, ref. 24) that contains four *Eca*I sites was cloned in the unique *Aat*II site of pUC18 plasmid (18) to yield plasmid pVB40 (Fig. 1). The 0.9 kb *Bam*HI fragment between the *Bam*HI site at 5505 in the lambda sequence and the *Bam*HI site in the multiple cloning site of pUC18 was deleted. The resulting plasmid (pVB44) containing only one *Bam*HI site was used to clone the *Eca*I methyltransferase gene.

Enterobacter cloacae DNA was partially digested with *Sau*3AI and ligated to pVB44 which had been cleaved with *Bam*HI and dephosphorylated with bacterial alkaline phosphatase. The ligated DNA was transformed into *E. coli* ER1398. Plasmid DNA isolated from a mixed liquid culture of approximately 30,000 Amp^R transformants was digested to completion with *Eca*I endonuclease, then the digested DNA was used to transform ER1398.

Most Amp^R clones contained plasmids which were smaller than the vector. The appearance of these clones is probably due to the fact that the *Eca*I sites of pVB44 are in a non-essential region of the plasmid. Presumably, the DNA fragments resulting from *Eca*I digestion are circularized *in vivo* at low frequency (25) and the largest fragment carrying the origin of replication and the Amp^R marker can be rescued using this type of selection. We found two clones which contained plasmids larger than the vector and the lambda-derived DNA region seemed to be intact. DNA isolated from these clones was resistant to *Eca*I digestion suggesting that it carries the *ecaIM* gene. The two plasmids were identical by restriction analysis. One of the plasmids named pEca1 was chosen for further investigation. In the course of this work pEca1 DNA was digested with several restriction enzymes known to be sensitive to either adenine or cytosine methylation and there was no indication that pEca1 carried any modification other than *Eca*I, thus we think that the cloned enzyme is the methyltransferase of the *Eca*I restriction-modification system. The clone carrying pEca1 did not show phage restriction *in vivo* and no *Eca*I endonuclease could be detected in partially purified cell extracts.

In the approximately 2.2 kb insert of pEca1 there is a 1.6 kb *Hin*dIII fragment. This *Hin*dIII fragment was recloned in the plasmid vector pVB40. The resulting plasmids (pEca4 and pEca5) contain the *Hin*dIII fragment in opposite orientation (Fig. 1). pEca4 DNA is completely resistant to *Eca*I digestion (even if no IPTG was added to the culture) whereas pEca5 is only partially resistant (not shown). The methylation properties of pEca4 and pEca5 suggest that the coding region is located in the *Hin*dIII fragment but the *ecaIM* promoter lies outside of the fragment. This would explain why the expression is dependent on the orientation.

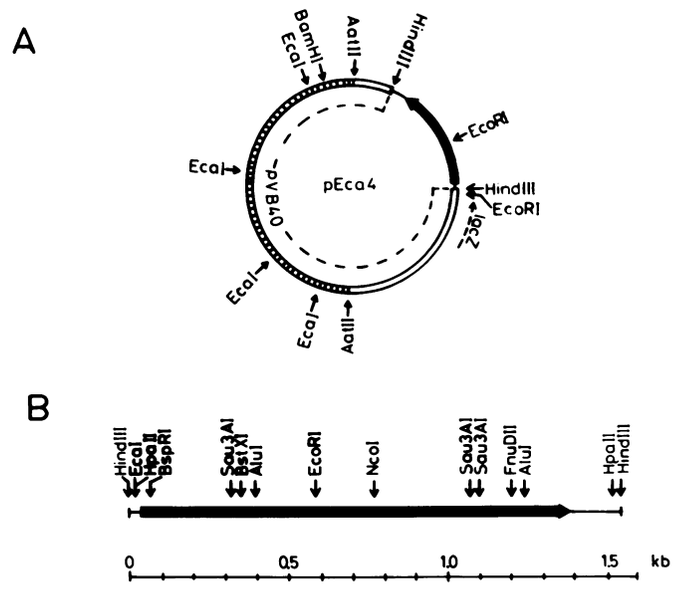

Figure 1. Map of the DNA region coding for the *Eca*I methyltransferase. The coding sequence is indicated by the thick arrow.
A: pEca4 (striped segment: lambda DNA *Aat*II fragment, empty segment: pUC18).
B: Restriction map of the 1.6 kb *Hin*dIII fragment cloned in pEca4.
Restriction sites which were used to subclone fragments for sequencing are shown.

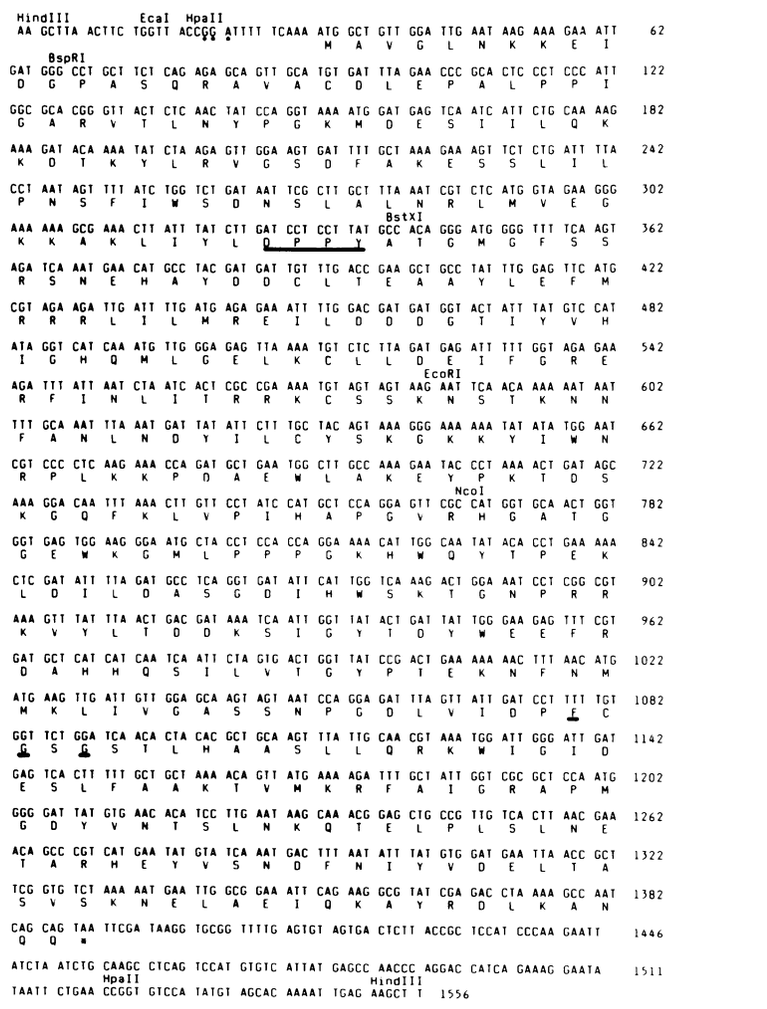

Figure 2. Nucleotide sequence of the *ecaIM* gene and the amino acid sequence of the methyltransferase protein. The DPPY and F.G.G blocks are underlined. Dots indicate a Shine−Dalgarno site.

Nucleotide sequence

The nucleotide sequence of the 1.6 kb *Hin*dIII fragment was determined (Fig. 2). Only one long open reading frame was found, it begins with ATG at 33 and ends at 1389 with TAA. The ATG at 33 is followed by other potential (in-frame) start codons downstream in the sequence (the closest is at 155).

To test if the first ATG is the functioning start codon of the *Eca*I methyltransferase, a pair of plasmids (pEca6 and pEca7) was constructed using the expression vector pER23S(−ATG) (Fig. 3). DNA fragments cloned in the *Sal*I site of pER23S(−ATG) are transcribed from the ribosomal RNA B gene P_2 promoter, but translation can only occur if the cloned gene contains an initiator codon. Expression of the cloned gene can be regulated via the *lac* operator which is located downstream of the promoter (T. Lukácsovich, pers. comm., Fig. 3). The *ecaIM* gene contains a unique *Bsp*RI site (nucleotide 67) which separates the first and second ATG codons (Fig. 2). The large *Bsp*RI-*Hin*dIII fragment containing the second but not the first ATG was inserted between the filled-in *Sal*I site and the *Hin*dIII site of pER23S(−ATG). The resulting plasmid was named pEca6 (Fig. 3). To construct the other plasmid (pEca7), the ends of the *Hpa*II-*Hpa*II fragment containing the whole open reading frame were filled in by DNA polymerase I large fragment then it was ligated into the filled-in *Sal*I site of pER23S(−ATG). Plasmids pEca6 and pEca7 contain no *Eca*I site, therefore total DNA was isolated from the clones and tested for *Eca*I methylation. Figure 4 demonstrates that DNA isolated from JM107 (pEca6) was digestible with *Eca*I, whereas DNA from JM107 (pEca7) was resistant, suggesting that translation starts at the first ATG. Translation cannot start further upstream, because there is an in-frame stop codon (TAA) at position 6. The reading frame beginning at 33 predicts a protein of 452 amino acids (M_r: 51,111). A protein of the expected size can be detected, upon IPTG induction, in cells containing pEca7 (Fig. 5). The level of M.*Eca*I production directed by pEca7 (Fig. 5) seems to be sufficiently high to facilitate purification of the enzyme.

The sequenced region upstream of the open reading frame does not contain structures resembling a typical *E. coli* promoter. The lack of such structure may explain why the level of expression of the *ecaIM* gene is different in pEca4 and pEca5.

Comparison with other DNA methyltransferases

M.*Eca*I is the first DNA methyltransferase with GGTNACC specificity for which the amino acid sequence has become known. The analysis of the deduced amino acid sequence revealed that it contains the motif F.G.G which, in a more or less conserved form, seems to be present in all DNA-MTases (7, 8) and has been suggested to be part of the S-adenosyl-methionine binding site (8). The identification of this motif in M*Eca*I gives further support to the notion that this conserved element must be responsible for a general step in the methylation process, possibly for binding S-adenosyl-methionine. The *Eca*I sequence does not show the motifs specific for C5mC-MTases (2), it does contain, however, the sequence DPPY common to adenine methyltransferases (26−42). This motif, in a modified form, was also found in N4mC-methyltransferases, therefore it was

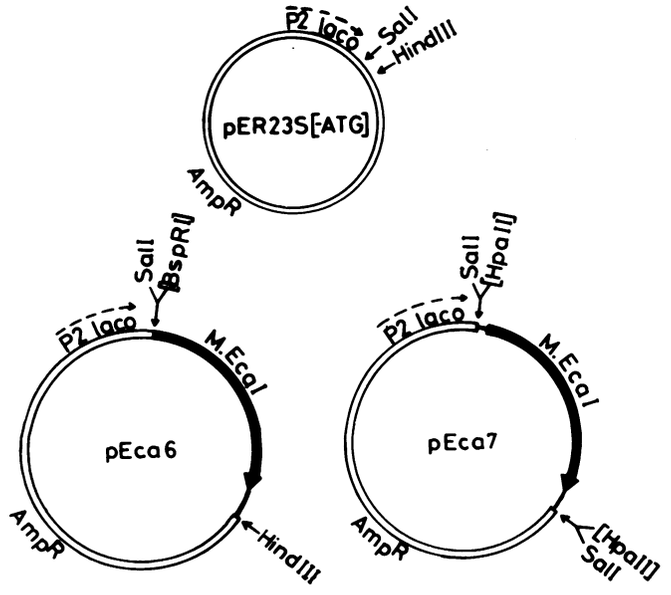

Figure 3. Schematic map of pER23S (−ATG), pEca6 and pEca7. Brackets mark restriction sites which were modified by filling-in and/or ligation and cannot be cleaved in pEca6 and pEca7 (see text).

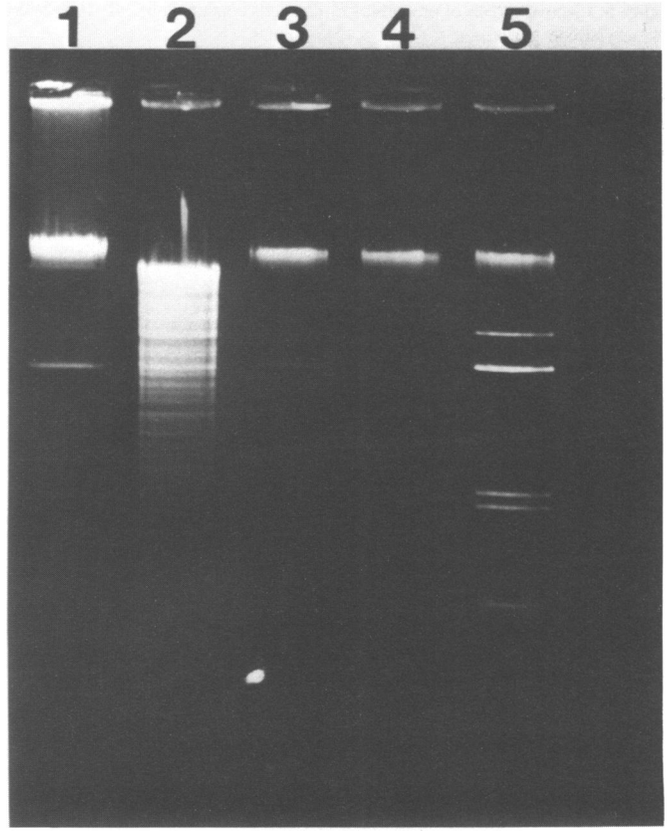

Figure 4. Digestion with *Eca*I of bacterial DNA isolated from *E. coli* clones containing either pEca6 or pEca7. 0.8% agarose gel.
1. JM107 (pEca6)
2. JM107 (pEca6) digested with *Eca*I
3. JM107 (pEca7)
4. JM107 (pEca7) digested with *Eca*I
5. JM107 (pEca7) and pVB40 digested with *Eca*I. pVB40 DNA was added to the digestion to exclude the possibility of enzyme inhibition.

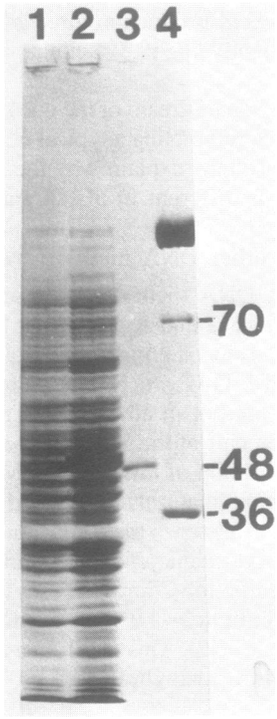

Figure 5. Electrophoresis of proteins of *E. coli* cells containing pEca7. 10% SDS-polyacrylamide gel.
1. uninduced
2. 3 hrs after adding 1 mM IPTG
3. *Bsp*RI methyltransferase (a 48 kDa protein)
4. *E. coli* RNA polymerase
Molecular weights ($M_r \times 10^{-3}$) are shown at right.

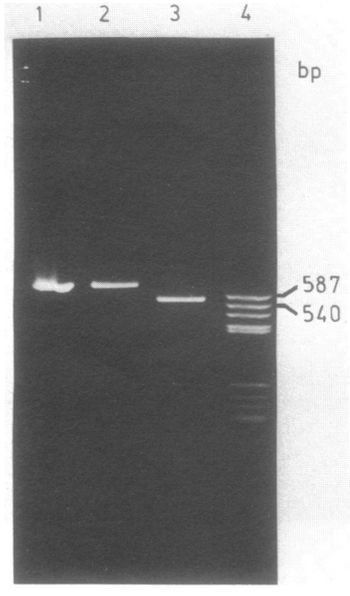

Figure 6. Digestion of the 640 bp *Eco*RI fragment of pEca4 with *Hpa*II.
1. undigested
2. digested with *Eca*I
3. digested with *Hpa*II
4. pBR322 digested with *Bsp*RI

suggested to be characteristic for DNA-MTases modifying extracyclic nitrogen (6, 42). The lack of resemblance to C5mC-MTases and the presence of the typical DPPY motif suggested that M.*Eca*I would be an adenine methyltransferase.

M.*Eca*I is an adenine methyltransferase

We have indirect experimental evidence suggesting that M.*Eca*I is indeed an adenine methyltransferase. The sequenced region contains a unique *Eca*I site which partially overlaps with a *Hpa*II site: GGTTACCGG (at position 14). Methylation, whether C5 or N4, of either cytosine in the sequence CCGG, is known to block cleavage by *Hpa*II (43,44). To test if methylation by M.*Eca*I protects the overlapping *Hpa*II site from *Hpa*II cleavage, a 640 bp *Eco*RI-*Eco*RI fragment was isolated from pEca4. This fragment contains 584 bp of the 1.6 kb *Hin*dIII fragment (with the 5'-end of the *ecaIM* gene) and the multiple cloning site of pUC18 (approximately 50 bp).

The appearance, upon *Hpa*II digestion, of an approximately 560 bp fragment indicates (Fig. 6) that *Eca*I methylation did not block *Hpa*II cleavage at the overlapping site, thus M.*Eca*I must be an adenine methyltransferase.

ACKNOWLEDGEMENTS

We thank Drs. J.Collins, E.Raleigh, U.Günthert, T.Lukácsovich for strains and E.Csorba for her excellent technical assistance.

REFERENCES

1. Lunnen,K.D., Barsomian,J.M., Camp,R.R., Card,C.O., Chen,S.-Z., Croft,R., Looney,M.C., Meda,M.M., Moran,L.S., Nwankwo,D.O., Slatko,B.E., Van Cott,E.M. and Wilson,G.G. (1988) *Gene* **74**, 25–32.
2. Pósfai,J., Bhagwat,A.S., Pósfai,G. and Roberts,R.J. (1989) *Nucl. Acids Res.* **17**, 2421–2435.
3. Hattman,S., Wilkinson,J., Swinton,D., Schlagman,S., Macdonald,P.M. and Mosig,G. (1985) *J. Bacteriol.* **164**, 932–937.
4. Lauster,R., Kriebardis,A. and Guschlbauer,W. (1987) *FEBS Lett.* **220**, 167–176.
5. Chandrasegaran,S. and Smith,H.O. (1988) Structure and expression. Vol. 1: From proteins to ribosomes. Eds. Sarma, R.H. and Sarma,M.H., Adenine Press, NY, pp. 149–156.
6. Tao,T., Walter,J., Brennan,K.J., Cotterman,M.M. and Blumenthal,R.M. (1989) *Nucl. Acids Res.* **17**, 4161–4175.
7. Lauster,R. (1989) *J. Mol. Biol.* **206**, 313–321.
8. Klimasauskas,S., Timinskas,A., Menkevicius,S., Biruliene,D., Butkus,V. and Janulaitis,A. *Nucl. Acids Res.* in press.
9. Balganesh,T.S., Reiners,L., Lauster,R., Noyer-Weidner,M., Wilke,K. and Trautner,T.A. (1987) *EMBO J.* **6**, 3543–3549.
10. Wu,J.C. and Santi,D.V. (1987) *J. Biol. Chem.* **262**, 4778–4786.
11. Lauster,R., Trautner,T.A. and Noyer-Weidner,M. (1989) *J. Mol. Biol.* **206**, 305–312.
12. Hobom,G., Schwarz,E., Melzer,M. and Mayer,H. (1981) *Nucl. Acids Res.* **9**, 4823–4832.
13. Raleigh,E.A. and Wilson,G. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 9070–9074.
14. Yanisch-Perron,C., Vieira,J. and Messing,J. (1985) *Gene* **33**, 103–119.
15. Haring,V., Scholz,P., Scherzinger,E., Frey,J., Derbyshire,K., Hatfull,G., Willetts,N.S. and Bagdasarian,M. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 6090–6094.
16. Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
17. Silhavy,T.J., Berman,M.L. and Enquist,L.W. (1984) Experiments with gene fusions. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
18. Norrander,J., Kempe,T. and Messing,J. (1983) *Gene* **26**, 101–106.
19. Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
20. Messing,J. (1983) *Methods Enzymol.* **101**, 20–79.
21. Biggin,M.D., Gibson,T.J. and Hong,G.F. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3963–3965.

22. Laemmli,U.K. (1970) *Nature (London)* **227**, 680–685.
23. Szomolányi,É., Kiss,A. and Venetianer,P. (1980) *Gene* **10**, 219–225.
24. Sanger,F., Coulson,A.R., Hong,G.F., Hill,D.F. and Petersen,G.B. (1982) *J. Mol. Biol.* **162**, 729–773.
25. Conley,E.C., Saunders,V.A. and Saunders,J.R. (1986) *Nucl. Acids Res.* **14**, 8905–8917.
26. Greene,P.J., Gupta,M., Boyer,H.W., Brown,W.E. and Rosenberg,J.M. (1981) *J. Biol. Chem.* **256**, 2143–2153.
27. Newman,A.K., Rubin,R.A., Kim,S.-H. and Modrich,P. (1981) *J. Biol. Chem.* **256**, 2131–2139.
28. Schoner,B., Kelly,S. and Smith,H.O. (1983) *Gene* **24**, 227–236.
29. Brooks,J.E., Blumenthal,R.M. and Gingeras,T.R. (1983) *Nucl. Acids Res.* **11**, 837–851.
30. Walder,R.Y., Walder,J.A. and Donelson,J.E. (1984) *J. Biol. Chem.* **259**, 8015–8026.
31. Bougueleret,L., Schwarzstein,M., Tsugita,A. and Zabeau,M. (1984) *Nucl. Acids Res.* **12**, 3659–3676.
32. Macdonald,P.M. and Mosig,G. (1984) *EMBO J.* **3**, 2863–2871.
33. Mannarelli,B.M., Balganesh,T.S., Greenberg,B., Springhorn,S.S. and Lacks,S.A. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 4468–4472.
34. 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.
35. Slatko,B.E., Benner,J.S., Jager-Quinton,T., Moran,L.S., Simcox,T.G., Van Cott,E.M. and Wilson,G.G. (1987) *Nucl. Acids Res.* **15**, 9781–9796.
36. Narva,K.E., Wendell,D.L.,Skrdla,M.P. and Van Etten,J.L. (1987) *Nucl. Acids Res.* **15**, 9807–9823.
37. Loenen,W.A.M., Daniel,A.S., Braymer,H.D. and Murray,N.E. (1987) *J. Mol. Biol.* **198**, 159–170.
38. Chandrasegaran,S., Lunnen,K.D., Smith,H.O. and Wilson,G.G. (1988) *Gene* **70**, 387–392.
39. Lacks,S.A., Mannarelli,B.M., Springhorn,S.S. and Greenberg,B. (1986) *Cell*, **46**, 993–1000.
40. Miner,Z. and Hattman,S. (1988) *J. Bacteriol.* **170**, 5177–5184.
41. 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.
42. Looney,M.C., Moran,L.S., Jack,W.E., Feehery,G.R., Benner,J.S., Slatko,B.E. and Wilson,G.G. (1989) *Gene* **80**, 193–208.
43. Butkus,V., Petrauskiene,L., Maneliene,Z., Klimasauskas,S., Laucys,V. and Janulaitis,A. (1987) *Nucl. Acids Res.* **15**, 7091–7102.
44. McClelland,M. and Nelson,M. (1988) *Gene* **74**, 291–304.