Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 5653–5656, September 1984
Biochemistry

Cloning and sequencing of liver cDNA coding for bovine protein C

(anticoagulant/vitamin K-dependent carboxylation/protein processing)

GEORGE L. LONG*, RAMA M. BELAGAJE*, AND ROSS T. A. MACGILLIVRAY†

*Division of Molecular and Cell Biology, Lilly Research Laboratories, Eli Lilly and Company, 307 East McCarty Street, Indianapolis, IN 46285; and †Department of Biochemistry, University of British Columbia, Vancouver, BC, Canada V6T 1W5

Communicated by K. M. Brinkhous, May 21, 1984

ABSTRACT **cDNA coding for protein C has been cloned from a bovine liver library in plasmid vector pBR322 and its sequence has been determined. Two overlapping clones code for the entire light and heavy chains of the mature protein, as well as a previously unreported connecting dipeptide (Lys-Arg) and a 39-amino acid leader peptide region. Identification and characterization of the clones establishes the liver as a site of protein C biosynthesis. A contiguous coding region reveals that a one-chain precursor protein is made that upon limited proteolysis yields both the mature light and heavy chains. The codon for aspartic acid is found at light chain amino acid position 71, showing that the β-hydroxyaspartic acid that exists in this position of the mature protein is the result of post-translational modification of an aspartic acid residue. Amino acid sequence homology in the amino-terminal region of the light chain with other vitamin K-dependent coagulation factors is continued into the leader peptide region.**

Protein C is a vitamin K-dependent blood plasma protein that plays a key role in hemostasis. A brief review of protein C has recently been published (1). Activated protein C in the presence of phospholipid and Ca^{2+} inactivates both factor Va (2–4) and factor VIIIc (5) by limited proteolysis. Bovine protein C has been purified and characterized (2, 6–8). In plasma it circulates as a two-chain (M_r = 21,000 and 41,000) serine protease zymogen precursor linked by a disulfide bridge. Stenflo and Fernlund (9, 10) have reported the bovine heavy and light chain amino acid sequences. Esmon and co-workers (11, 12) have shown that inactive protein C precursor is converted at the endothelial cell surface to an active protease by limited thrombin cleavage and that the activation process is dramatically potentiated by an additional protein cofactor, thrombomodulin. Activated protein C is identical to autoprothrombin IIA, described some time ago by Seegers *et al.* (13, 14). Human protein C has been purified and characterized by Kisiel (15).

Bovine protein C undergoes extensive post-translational modification. It contains 18% (wt/wt) carbohydrate (8). The 11 amino-terminal glutamate residues (16) are converted to γ-carboxyglutamate residues by a vitamin K-dependent process (17). Recently, it has been shown that residue 71 in mature bovine protein C exists as β-hydroxyaspartate (18). In addition to the above-mentioned limited proteolysis, it has been assumed that processing of nascent protein C requires cleavage of a signal peptide-containing precursor, to permit transport of the protein out of the cell. Also, it has been presumed that protein C is synthesized in the liver.

In this paper we report the isolation and characterization of bovine liver cDNA clones that contain the entire coding region of the mature protein, as well as a 39-amino acid amino-terminal segment, which presumably is cleaved during processing of the nascent protein. Our findings reveal that two-chain mature protein C is produced from a one-chain precursor and that during the course of the cleavage a previously unreported dipeptide (Lys-Arg) is removed.

MATERIALS AND METHODS

Construction of the Bovine Liver cDNA Library. Bovine liver mRNA was prepared and analyzed as described (19). cDNA synthesis, oligo(dC) tailing, and ligation into plasmid pBR322 that had been oligo(dG) tailed at the *Pst* I site was done according to the method of Land *et al.* (20) and is described elsewhere (21). Tailed cDNA was size fractionated by agarose gel electrophoresis and material estimated to be 1 kilobase or greater in size was used in the ligation. Chimeric plasmids were used to transform $CaCl_2$-treated (22) *Escherichia coli* strain RR1, and transformants were selected by tetracycline resistance.

High-Density Colony Screening. Transformed cells were plated in duplicate at an approximate density of 5000 colonies per nitrocellulose filter (Millipore HATF082, 82 mm diameter) on L broth/1.5% agar plates containing 12.5 μg of tetracycline per ml, amplified on plates containing 250 μg of chloramphenicol per ml, and hybridized with the mixed nucleotide probe as outlined elsewhere (23). Positive duplicate colonies were replated in duplicate and screened at a density of approximately 100–200 colonies per filter. At this point the hybridization signal is much more intense and individual colonies can be picked and grown in liquid media for plasmid preparation and analysis.

Mixed Oligonucleotide Probe. The following mixed oligodeoxynucleotide was synthesized on a model 380A DNA synthesizer (Applied Biosystems, Foster City, CA) by the phosphite triester method as prescribed by the manufacturer and was purified by anion-exchange high-performance liquid chromatography:

$$5'\ \text{HO-T-C-}\tfrac{C}{T}\text{-T-C-C-A-T-}\tfrac{A}{G}\text{-C-A-}\tfrac{A}{G}\text{-T-A-}\tfrac{A}{G}\text{-T-G-OPO}_3^{2-}\ 3'.$$

The oligonucleotide was labeled by using T4 polynucleotide kinase and [γ-^{32}P]ATP from New England Nuclear and was purified by Sephadex G-25 column chromatography and stored frozen (−20°C) until the time of hybridization. The radioactive medium from the first hybridization was also frozen and reused for the second screening.

Large-Scale Plasmid Preparation. Large preparations of plasmid suitable for DNA sequencing were grown in M-9 medium in the presence of chloramphenicol and were banded by centrifugation in CsCl/ethidium bromide after cell lysis (24).

Restriction Mapping and Southern Hybridization. Plasmid DNA was prepared from 1-ml overnight liquid cultures by the method of Birnboim and Doly (25), cleaved with restriction endonucleases according to the supplier's recommendations, and Southern blotted from 1% agar gels by the modified method of Smith and Summers (26). Probes corresponding to the 5′ (bases 87–497, Fig. 2) and 3′ (bases 494–1149,

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: bp, base pair(s).

Fig. 2) ends of the cDNA translated region were separately purified from acrylamide gels and labeled by nick-translation using a standard protocol (27).

DNA Sequencing. Cesium chloride-purified plasmid or restriction fragments from mini-prepped DNA (25) suitable for DNA radiolabeling and sequencing were obtained by electrophoresis on 3.5% polyacrylamide gels and crush-elution (28). Blunt-end or 5'-overhanging DNA fragments were treated with bacterial alkaline phosphatase and T4 polynucleotide kinase according to the method of Maxam and Gilbert (28). 3'-Overhanging fragments were labeled with [^{32}P]-cordycepin according to the directions provided by the supplier (New England Nuclear). Labeled ends were separated on polyacrylamide gels after restriction endonuclease digestion and sequenced by the chemical modification technique of Maxam and Gilbert (28). The G reaction was performed for 7 min at 20°C, and the G+A reaction for 10 min at 37°C. Polyacrylamide sequencing gels (6% and 20%) were run according to Sanger and Coulson (29), fixed with 7% (wt/vol) acetic acid (6% gels dried), and exposed to x-ray film with the aid of intensifying screens. All DNA sequences were stored and analyzed with the aid of the Lilly Research Laboratories DNAce computing environment.

RESULTS

Screening of Bovine Liver cDNA Colonies for Protein C. Assuming that circulating levels of protein C [≈13 μg/ml of plasma (6)] and prothrombin give an approximation of the relative amounts of mRNA for the two proteins, it was estimated that protein C mRNA would represent approximately 0.1% of total mRNA, based on prothrombin representing 1% of total mRNA (19). Consequently, it was decided to screen approximately 50,000 colonies (10 duplicate filters). From this, 27 duplicate weak positives were observed and replated at a density of 100–200 colonies per filter (Table 1). In the second screening 6 clones gave a very intense, duplicate hybridization signal, 11 gave a weak signal, and 10 yielded no duplicate signal. Liberation of the cDNA inserts from positive clones by *Pst* I digestion yielded inserts ranging in size from 150 to 1250 bp (see Table 1), with the majority of them being around 800 bp. None of them appeared to contain inserts in which the *Pst* I cloning site was incorrectly reconstituted. Southern hybridization of *Pst* I-generated fragments from positive clones revealed hybridizing bands for only the 6 strongly hybridizing colony clones (data not shown). Consequently, the weakly hybridizing colonies have not been further characterized.

Table 1. Summary of clones showing hybridization with a mixed probe for bovine protein C

Clones	Intensity of colony hybridization	Southern hybridization	Approximate insert size, bp
2, 5	Strong	+	480 + 300
3, 4, 7	Strong	+	880 + 90
6	Strong	+	760
1, 8, 11, 13, 15	Weak	−	950 + 135
9	Weak	−	150
10	Weak	−	1250
12	Weak	−	385
14	Weak	−	1050
16	Weak	−	830
17	Weak	−	680

Clones are identified by arbitrary numbers. Those listed on the same line are believed to be identical as judged by restriction enzyme mapping (data not presented). Approximate insert size (bp, base pairs) is based upon electrophoretic migration of *Pst* I-liberated fragments on 1% agarose or 3.5% polyacrylamide gels.

Analysis of the cDNA. Two of the clones, pBC-2 and pBC-7, are overlapping and together represent the largest identified contiguous region of cDNA coding for protein C. Clones pBC-2 and pBC-7 are 787 and 1000 bp long, respectively (Fig. 1) and overlap by 333 coding bp, including an internal *Pst* I site. The homopolymeric tails are 22 and 30 G residues at the 5' ends and 16 and 15 C residues at the 3' ends of pBC-2 and pBC-7, respectively, which are the lengths anticipated from the chosen method of cDNA synthesis and tailing (20). Both inserts are in the opposite orientation of the vector ampicillin-resistance gene (data not presented). Fig. 1 also summarizes the strategy used in obtaining the cDNA sequence.

Analysis of the resulting nucleotide sequence and the corresponding amino acid sequence (Fig. 2) reveals that clones pBC-2 and pBC-7 together contain the entire coding region for both the heavy and light chains of mature protein C. Additionally, these clones contain a sequence corresponding to 39 "leader" amino acids at the "amino" terminus of the light chain, a dipeptide (Lys-Arg) that links the two chains in the nascent protein, and a stop codon and three untranslated nucleotides at the 3' end. The amino acid sequence shown in Fig. 2 agrees with that of Stenflo and co-workers (9, 10, 18) for the heavy and light chains except for the last two amino

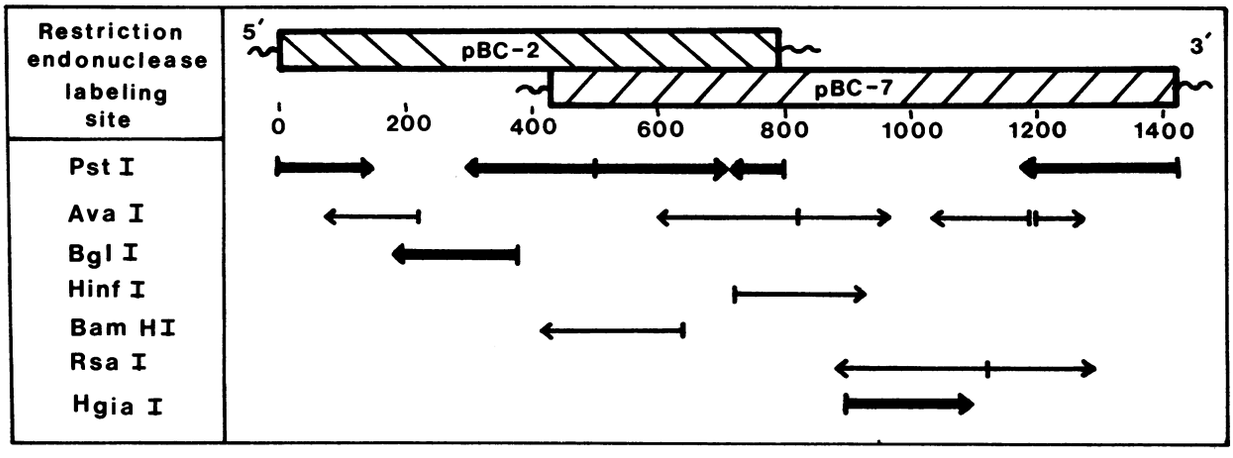

FIG. 1. Sequencing strategy for bovine liver protein C cDNA. Horizontal bars portray the sizes and relationship of the cDNA inserts in clones pBC-2 and pBC-7, which were used for sequencing. Horizontal arrows show the site of ^{32}P labeling (vertical line) and length of sequence obtained for individual fragments. In some cases fragments were sequenced more than once and in both inserts. Thick and thin arrows represent 3'-cordycepin- and 5'-ATP-labeled fragments, respectively. All sites within the inserts for the enzymes listed are shown except for *Bam*HI, *Hin*fI, and *Hgi*aI which also exist at positions 129, 86, and 190, respectively. Numbering refers to nucleotide position as shown in Fig. 2.

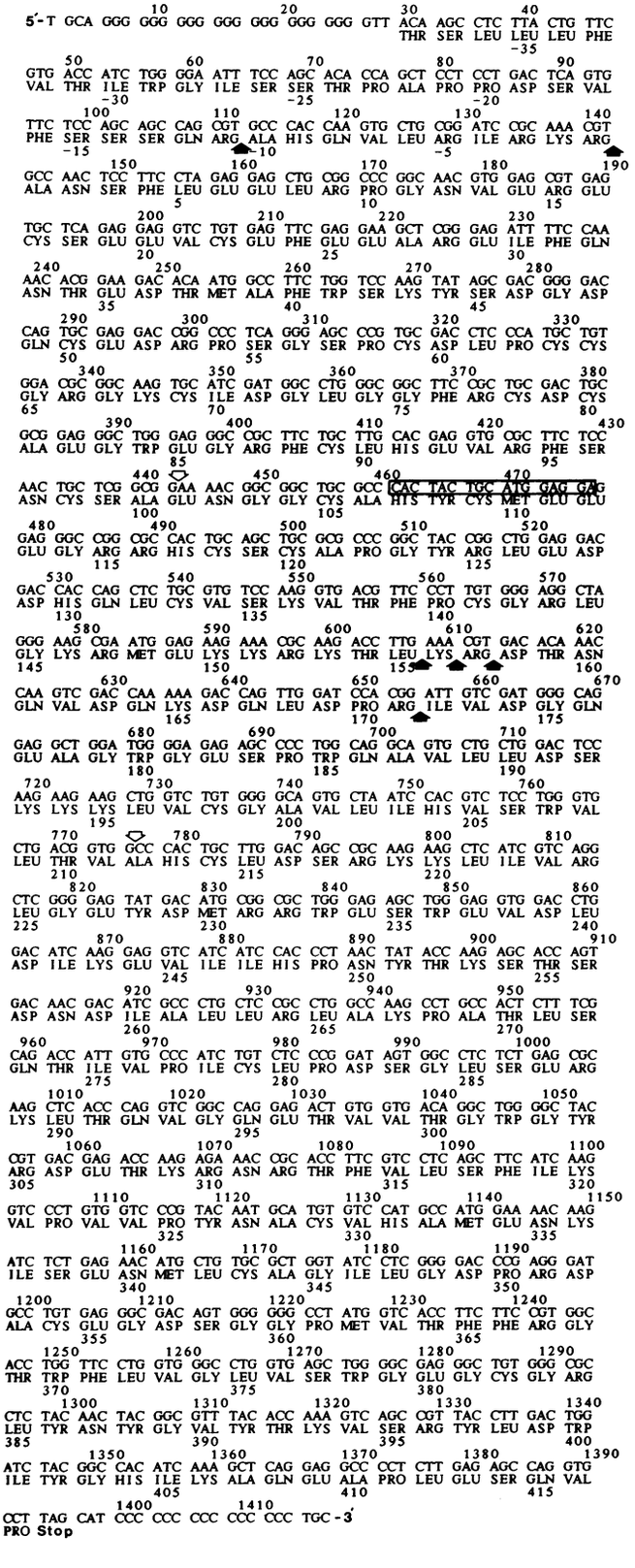

FIG. 2. Nucleotide and amino acid sequences derived from cloned bovine liver cDNA coding for protein C. The amino acid sequence for the precursor of protein C is shown below the determined nucleotide sequence. Open arrows above the DNA sequence at positions 441 and 774 mark the ends of clones pBC-2 and pBC-7, respectively, excluding the homopolymeric tails. Proposed sites for limited proteolytic processing of the precursor are shown by bold vertical arrows below the amino acid sequence and are discussed in the text. Cleavage at amino acid position 171 results in activation of protein C and the formation of a 14-amino acid activation peptide and has been previously reported (8). Glutamate residues 6, 7, 14, 16, 19, 20, 23, 25, 26, 29, and 35 exist as γ-carboxyglutamate residues in the mature protein (14). Amino acid residue 71 exists in the mature protein as β-hydroxyaspartate (18). The boxed region in the nucleic acid sequence is complementary to the mixed probe utilized in the initial screening.

acids of the heavy chain, which are reversed from the previously published assignment (10).

DISCUSSION

The cDNA library used in this study has previously been shown to contain ≈0.4% prothrombin cDNA (21). Estimates of the amount of protein C circulating in the blood (≈13 μg/ml) are about 1/10th that of prothrombin (6). Six positive clones for protein C were obtained from approximately 50,000 transformed bacterial colonies. Assuming that the frequency of positive clones reflects the level of mRNA in the cell, protein C mRNA represents ≈0.01% of total mRNA. The reason for this discrepancy in the apparent level of protein C mRNA is not known. Two possible factors are differences in levels of messenger translation and protein degradation and clearance.

The derived amino acid sequence reported in this paper is in complete agreement with that previously published for the mature heavy and light chains (9, 10, 18) with the exception of the two carboxyl-terminal residues of the heavy chain, which we observe to be Val-Pro rather than Pro-Val. It is unlikely that the difference is due to a true genetic variant or a result of the cDNA synthesis and cloning since the change would require a minimum change of four nucleic acid bases. We believe it is more likely that the original assignment (10), based solely on carboxypeptidase Y digestion, was incorrectly made. Drakenberg *et al.* (18) have reported the existence of β-hydroxyaspartate at position 71 of the mature light chain. They point out that this residue could be derived from aspartate, serine, or glycine. Our findings are consistent with their report and clearly establish that the modification involves post-translational modification of an aspartate residue.

This study establishes that protein C is synthesized in the liver and that the mature two-chain protein arises from a single-chain precursor by limited proteolysis. It additionally establishes the existence in the nascent protein of a Lys-Arg dipeptide at the junction between the light and heavy chains and a leader peptide of at least 39 amino acids at the amino terminus. Fig. 3 shows that the previously noted (17) sequence homology between protein C and other vitamin K-dependent coagulation factors is continued into the leader peptide region and may reflect an important role this region plays in the biosynthesis of this unique class of proteins. On the basis of the known structural homology of protein C with other vitamin K-dependent coagulation factors (17) and continued homology with factors X (22) and IX (30) and prothrombin (21) in the leader peptide region (Fig. 3), we predict that nascent protein C is cleaved by signal peptidase after amino acid −11 to yield an extracellular prozymogen and a typical signal peptide (31). Amino acids −39 through −11 contain a distal highly hydrophobic region and a polar region proximal to the proposed signal peptidase site. The putative propeptide (residues −10 thru −1) is highly basic in composition. Extracellular cleavage by a trypsin-like enzyme(s) after amino acids −1, 156, and 157, followed by a sequence-specific carboxypeptidase cleavage of the basic lysine residue from the resulting light chain, would give the peptides observed in the circulating zymogen. The existence of a similar linking highly basic (Arg-Arg) dipeptide has also been observed by Fung *et al.* (23) for bovine factor X, and they have proposed a similar scheme of limited proteolysis.

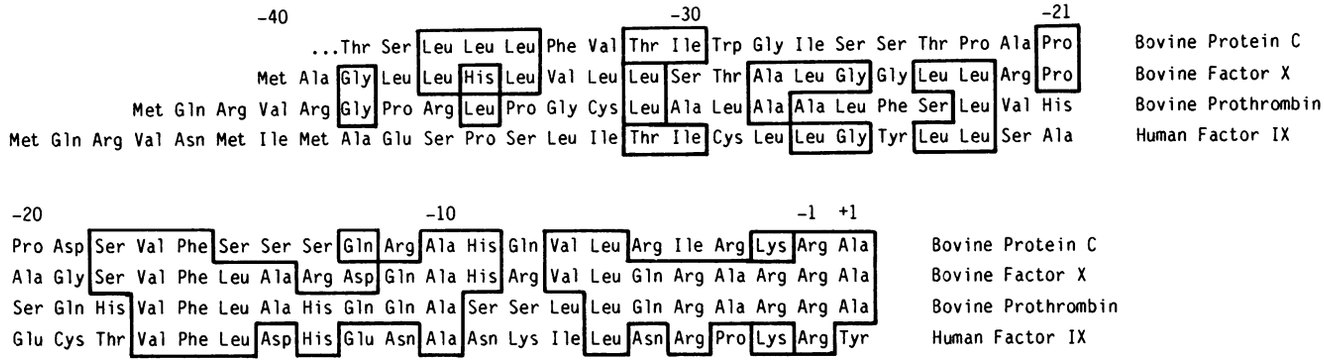

FIG. 3. Comparison of the leader region amino acid sequence for bovine protein C, factor X, prothrombin, and human factor IX. Sequences are aligned according to the amino-terminal amino acid of the mature protein and the structural homology of the mature proteins reported elsewhere (9). Identical residues are boxed.

We thank Charles Brush (Lilly Research Laboratories) for preparing the mixed oligonucleotide probe and Dr. Colin Hay for growing the colonies and preparing the nitrocellulose filters used in the initial high-density screening. This work was supported in part by a grant from the Medical Research Council of Canada to R.T.A.M.

1. Esmon, C. T. (1983) *Blood* **62,** 1155–1158.
2. Kisiel, W., Canfield, W. M., Ericsson, L. H. & Davie, E. W. (1977) *Biochemistry* **16,** 5824–5831.
3. Seegers, W. H., Marlar, R. A. & Walz, D. A. (1978) *Thromb. Res.* **13,** 233–243.
4. Walker, F. J., Sexton, P. W. & Esmon, C. T. (1979) *Biochim. Biophys. Acta* **571,** 333–342.
5. Vehar, G. A. & Davie, E. W. (1980) *Biochemistry* **19,** 401–410.
6. Stenflo, J. (1976) *J. Biol. Chem.* **251,** 355–363.
7. Esmon, C. T., Stenflo, J., Suttie, J. W. & Jackson, C. M. (1976) *J. Biol. Chem.* **251,** 3052–3056.
8. Kisiel, W., Ericsson, L. H. & Davie, E. W. (1976) *Biochemistry* **15,** 4893–4900.
9. Fernlund, P. & Stenflo, J. (1982) *J. Biol. Chem.* **257,** 12170–12179.
10. Stenflo, J. & Fernlund, P. (1982) *J. Biol. Chem.* **257,** 12180–12190.
11. Esmon, C. T. & Owen, W. G. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 2249–2252.
12. Esmon, N. L., Owen, W. G. & Esmon, C. T. (1982) *J. Biol. Chem.* **257,** 859–864.
13. Seegers, W. H. & Ulutin, O. N. (1961) *Thromb. Diath. Haemorrh.* **6,** 270–281.
14. Seegers, W. H., Novoa, E., Henry, R. L. & Hassouna, H. I. (1976) *Thromb. Res.* **8,** 543–552.
15. Kisiel, W. (1979) *J. Clin. Invest.* **64,** 761–769.
16. Fernlund, P., Stenflo, J. & Tufvesson, A. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 5889–5892.
17. Olson, R. E. & Suttie, J. W. (1978) *Vitam. Horm. (N.Y.)* **35,** 59–108.
18. Drakenberg, T., Fernlund, P., Roepstorff, P. & Stenflo, J. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 1802–1806.
19. MacGillivray, R. T. A., Degen, S. J. F., Chandra, T., Woo, S. L. C. & Davie, E. W. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 5153–5157.
20. Land, H., Grez, M., Hansen, H., Lindenmaier, W. & Schultz, G. (1981) *Nucleic Acids Res.* **10,** 2253–2266.
21. MacGillivray, R. T. A. & Davie, E. W. (1984) *Biochemistry* **23,** 1626–1634.
22. Dagert, M. & Ehrlich, S. D. (1979) *Gene* **6,** 23–28.
23. Fung, M. R., Campbell, R. A. & MacGillivray, R. T. A. (1984) *Nucleic Acids Res.* **12,** 4481–4492.
24. Katz, L., Williams, P. H., Sato, S., Leavill, R. W. & Helinski, D. R. (1973) *J. Bacteriol.* **114,** 577–591.
25. Birnboim, H. C. & Doly, J. (1979) *Nucleic Acids Res.* **7,** 1513–1523.
26. Smith, G. E. & Summers, M. D. (1980) *Anal. Biochem.* **109,** 123–129.
27. Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* **113,** 237–251.
28. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560.
29. Sanger, F. & Coulson, A. R. (1978) *FEBS Lett.* **87,** 107–110.
30. Kurachi, K. & Davie, E. W. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 6461–6464.
31. Jackson, R. C. & Blobel, G. (1980) *Ann. N.Y. Acad. Sci.* **343,** 391–403.