

Cloning, expression in *Escherichia coli*, and reconstitution of human myoglobin

(gene expression/heme protein)

RAGHAVAN VARADARAJAN, ALEX SZABO, AND STEVEN G. BOXER

Department of Chemistry, Stanford University, Stanford, CA 94305

Communicated by John I. Brauman, May 6, 1985

ABSTRACT A full-length cDNA clone for human myoglobin has been isolated from a human skeletal muscle cDNA library. The clone as isolated has a cDNA insert approximately one kilobase long and has 5' and 3' untranslated regions of approximately 80 and 530 base pairs, respectively. The sequence of the translated region corresponds exactly to that predicted for human myoglobin. The cDNA was expressed in high yield in *Escherichia coli* as a fusion protein consisting of the first 31 amino acids of the phage λ cII gene, the tetrapeptide Ile-Glu-Gly-Arg, and the myoglobin sequence by following the approach of Nagai and Thogersen [Nagai, K. & Thogersen, M. C. (1984) *Nature (London)* 309, 810-812]. The fusion product was isolated, reconstituted with heme, cleaved with trypsin, and purified to generate a protein whose properties are indistinguishable from those for authentic human myoglobin. Myoglobin can be readily prepared on a gram scale by using these methods.

In the present communication we report the cloning of the cDNA for human myoglobin (Mb), the expression of this gene in high yield in *Escherichia coli*, simple procedures for purification of the gene product and reconstitution with heme, and limited comparisons of the isolated human Mb with Mbs from natural sources. This sets the stage for detailed physical and biochemical studies of Mbs modified by site-directed mutagenesis.

Mb is among a handful of the most studied proteins (1). Sperm whale Mb was the first protein to be characterized by x-ray crystallography (2-4), and the structures of several other Mbs and hemoglobins have been reported as well (5, 6). All Mbs studied to date have very similar secondary and tertiary structures. Because of their physiological importance, abundance, stability, and high-resolution structural characterization, these proteins have been subjected to characterization by essentially all tools available to protein chemists, biochemists, and biophysicists. Their features serve as paradigms for many aspects of globular protein structure, folding and dynamics, and ligand binding. It is straightforward to remove the heme group from Mb and reconstitute the apoMb with native heme (7), modified hemes (8), dye molecules (9), and even chlorophyll derivatives (10).

Recently Jeffreys and co-workers have isolated the genes for seal and human Mb (11-13). In each case the gene is in excess of 9 kilobase pairs (kb) long, whereas the coding regions total about 500 base pairs (bp). Because of the high level of homology between exon II of the seal and the human gene (87%), our strategy for isolating the cDNA for human Mb was to use a convenient restriction fragment containing exon II from the seal gene to probe a cDNA library from human skeletal muscle. By using the technique recently introduced by Nagai and Thogersen (14), this cDNA was

fused to a piece of the phage λ cII gene via a dodecamer linkage, leading to high-level expression in *E. coli* of a fusion protein. This gene product is readily isolated, combined with heme, and cleaved to produce high levels of human Mb.

MATERIALS AND METHODS

Unless mentioned otherwise, all DNA manipulations were as described elsewhere (15). Plasmid DNA preparation was by the method as described in ref. 16. Minilysates were prepared by the alkaline lysis method (17). DNA restriction fragments were separated by electrophoresis on agarose gels in Tris/acetate buffer. DNA fragments were recovered from low-melting agarose gels by using either Elutip-d columns (Schleicher & Schuell) or extraction with phenol and precipitation with ethanol. Authentic human Mb was obtained from Rockland Co. All enzymes were from New England Biolabs except for mung bean nuclease (P-L Biochemicals), lysozyme (Boehringer Mannheim), and trypsin (Worthington). DNA sequencing was performed according to the method of Maxam and Gilbert (18). *E. coli* strain MZ1 was a gift from M. Zuber. N99cI⁺ wild-type lysogen and the expression vector pAS1 were from Allan Shatzman. The β -globin expression system consisting of the vectors M13mp11FX and pLcIIFX β was generously provided by K. Nagai.

cDNA Clone Isolation. Plasmid pSM19.5 (12) containing the seal Mb gene was a generous gift from A. J. Jeffreys. The plasmid was digested with *EcoRI* (see Fig. 1), and the 2-kb *EcoRI/EcoRI* fragment containing exon II was isolated. One microgram of this fragment was nick-translated with [α -³²P]dATP (3000 Ci/mmol; 1 Ci = 37 GBq) to a specific activity of 2×10^7 cpm/ μ g and was used as a probe.

A human skeletal muscle cDNA library (20) was linearized with both *EcoRI* and *HindIII*. One microgram of each digest was run on a 0.7% agarose gel. Blotting onto nitrocellulose and subsequent hybridization of the probe were carried out as described in ref. 21. A single band appeared on the autoradiogram, indicating the presence of a cDNA insert approximately 1 kb long.

One microgram of the library was used to transform *E. coli* HB101; 5000 ampicillin-resistant colonies were picked and screened according to the method of Grunstein and Hogness (22). One positive clone was obtained. Plasmid DNA from this clone (pMb0, see Fig. 2) was isolated, a restriction map was prepared as shown in Fig. 1a, and the region from the distal *Ava I* site to the message start codon was sequenced.

Construction of pMb1. Twenty micrograms of the expression vector pAS1 (23) was linearized with *BamHI* and then digested with 75 units of mung bean nuclease under conditions described in ref. 23. Twenty micrograms of pMb0 partially digested with *Nco I* was treated with 60 units of mung bean nuclease. These digested plasmids were both

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.

Abbreviations: kb, kilobase(s) or kilobase pair(s); bp, base pair(s).

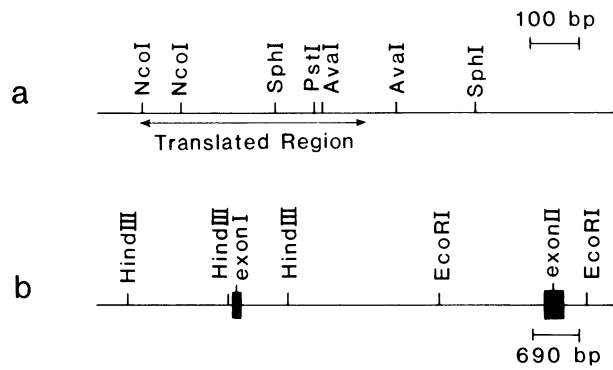


FIG. 1. (a) Restriction map of the insert in pMb0, which is a full-length human Mb cDNA. The insert is cloned into the Berg-Okayama vector (19). (b) Restriction map of the insert in pSM19.5 (12), which contains the first two exons of the seal Mb gene. The *EcoRI/EcoRI* restriction fragment containing exon II was used as a probe.

further digested with *Sph* I. The 637-bp *Nco* I/*Sph* I fragment from pMb0 was ligated to the 5.69-kb *Bam*HI/*Sph* I fragment from pAS1 and used to transform *N99c*⁺. Transformants were selected on L broth ampicillin plates, and plasmid from one such transformant was isolated and sequenced to ensure that the first glycine codon of the Mb cDNA was preceded by the ATG codon from pAS1.

Construction of pMb3. As outlined in Fig. 2, the 993-bp *Hind*III/*Sph* I fragment from pMb0 was ligated to the 3.83-kb *Hind*III/*Sph* I fragment from pAS1 to give pMb2. pMb2 was partially digested with *Nco* I and then with mung bean nuclease as described above. The resulting fragments were further digested with *Sal* I. The 726-bp *Nco* I/*Sal* I fragment was then ligated to *Stu* I/*Sal* I-cleaved M13mp11FX (14). The ligation mixture was used to transform *E. coli* strain JM101 (24). Cells were plated in the presence of isopropyl β -D-thiogalactopyranoside and 5-bromo-4-chloro-3-indolyl β -D-galactopyranoside (BMB). Six colorless plaques were picked and the double-stranded replicative form of the recombinant phage was isolated. Double digestion with *Hind*III and *Bam*HI confirmed the presence of the desired insert. The replicative form from one such plaque was selected for further use and was named M13mp11FXMb. The 759-bp *Bam*HI/*Hind*III fragment from M13mp11FXMb was ligated to *Bam*HI/*Hind*III-cleaved pLcIIIFX β (14) and used to transform *N99c*⁺. Plasmid from one recombinant clone was sequenced to ensure that the first glycine codon of Mb was preceded by the DNA sequence coding for Ile-Glu-Gly-Arg. This plasmid was named pMb3.

Expression. Plasmids pMb1 and pMb3 were used to transform the *E. coli* strain MZ1 as follows. One hundred nanograms of plasmid in buffer (10 mM Tris-HCl, pH 8.0/1 mM EDTA) was added to 200 μ l of competent MZ1 cells (prepared as described in ref. 15). The mixture was incubated on ice for 30 min and then heat shocked at 34°C for 5 min. Then 2.5 ml of warmed L broth was added, and the mixture was incubated at 28°C for 60 min, then plated on L broth/ampicillin plates and incubated at 28°C. Ampicillin-resistant colonies were picked and grown in L broth to an OD₆₅₀ of 0.75. The temperature was then rapidly elevated to 42°C by adding an equal amount of 65°C L broth with rapid swirling, and the culture was grown at 42°C. After a 2-hr induction the cells were spun down and frozen at -70°C. Cells were lysed by keeping them at room temperature for 1-2 min and then adding 100 μ l of 2 \times Laemmli buffer (25). Samples were then heated to 90°C for 5 min and analyzed by NaDodSO₄/PAGE. Protein was detected by either Coomassie blue staining (Fig. 3) or immunoblotting using horseradish

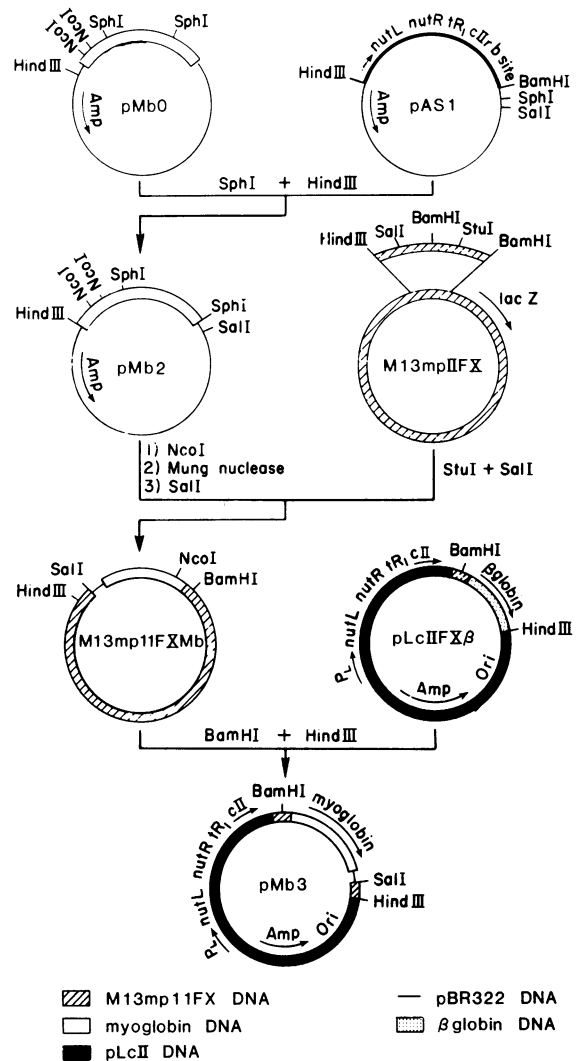


FIG. 2. Construction of pMb3. Indicated above is the strategy for replacement of the cDNA for β -globin by that for human Mb in the expression vector pLcIIIFX β (14). Mb cDNA in pMb0 has an *Nco* I site at the initiation codon. Cleavage at this site and subsequent digestion by mung nuclease trims back the DNA to the Gly-1 codon and makes a blunt end. cDNA thus treated was attached to the FX linker sequence by subcloning in M13mp11FX (14). The linker-Mb sequence was then inserted into *Bam*HI/*Hind*III-cleaved pLcIIIFX β as shown to generate pMb3. [Note that the region between the *Hind*III site and the start of the Mb cDNA is simian virus 40 DNA (19), not pBR322. This was not indicated, for clarity.]

peroxidase coupled with antiserum to human Mb (Cappel Laboratories, Cochranville, PA) (26).

Isolation of Fusion Product, Cleavage, and Purification. Purification of the fusion product was essentially as described by Marston *et al.* (27). Cultures were grown as described and cells were isolated by centrifugation. Cells were lysed and fusion protein-containing inclusion bodies were isolated and solubilized as described in ref. 27. The solubilized denatured protein was dialyzed against 50 mM Tris-HCl, pH 8.0/100 mM NaCl/1 mM CaCl₂. Approximately 10 equivalents of heme dissolved in a minimal amount of 0.01 M NaOH was added to reconstitute the fusion protein. This product was then subjected to 12-hr digestion by trypsin (18 mg of trypsin per g of protein) at 25°C, followed by dialysis against 10 mM Tris-HCl, pH 8.0. The resulting moderately pure human Mb was concentrated, loaded onto a DEAE-Sephacel column (Pharmacia), and eluted with 50 mM Tris-HCl, pH 8.0. Mb-containing fractions were pooled and applied to a Seph-

acryl S-200 column in 10 mM Tris·HCl, pH 8.0, for final purification.

Electronic Absorption Spectra. Cyanmetmyoglobin was prepared by adding a small crystal of potassium cyanide to a cuvette containing the metquo form. Reduction to the ferrous deoxygenated form was accomplished by adding a few crystals of sodium dithionite to the cyanmetmyoglobin. Carbonmonoxymyoglobin was prepared by slowly bubbling carbon monoxide into the cuvette containing the deoxygenated form prepared above. All absorption spectra (Fig. 4) were obtained on a Varian 2300 spectrophotometer.

RESULTS AND DISCUSSION

The human Mb cDNA clone isolated had an insert of approximately 1 kb and has 5' and 3' untranslated regions of approximately 80 and 530 bp, respectively. The entire coding region was sequenced and found to be identical with the sequence derived from the exons of the human Mb gene (13). This in turn is consistent with the known protein sequence (28).

Several different approaches were taken to express human apoMb in *E. coli*. To this end the human Mb cDNA was cloned in the expression vector pAS1 (23), which has been shown to give relatively high yields of a number of eukaryotic proteins. In the desired recombinant, pMb1, the Gly-1 codon of human Mb was preceded by the initiation codon of the *cII* gene, as demonstrated by sequence analysis. Expression was under the control of the λ P_L promoter. pMb1 was used to transform the defective λ lysogen MZ1, which has a temperature-sensitive mutation (*cI857*) in the λ *cI* gene. When these cells are grown at 30°C the repressor synthesized by this gene is active, and transcription from the P_L promoter is repressed. At 42°C, however, the repressor is no longer active and transcription from the P_L promoter is possible.

MZ1 cells containing pMb1 were induced as described in *Materials and Methods*, and protein expression was analyzed by NaDodSO₄/PAGE. No apoMb was detectable by Coomassie blue staining above the background of native *E. coli* proteins of comparable molecular weight. Immunoblotting experiments with antibodies to human Mb revealed a band at apparent molecular mass 17 kDa, demonstrating that apoMb was expressed in induced cells at a low level (no more than 0.2% of total protein estimated crudely from the sensitivity of immunoblots using known quantities of sperm whale Mb). No fragments of lower molecular weight were detected. Although this experiment demonstrates that human apoMb itself can be expressed in *E. coli*, the levels of expression in our hands were not sufficient for biophysical studies, and this approach was not optimized or pursued further.

Very recently Nagai and Thogersen (14) reported an ingenious approach to the expression of β -globin in *E. coli*. It has often been observed that eukaryotic gene expression in *E. coli* can be enhanced considerably by fusing the gene of interest to a portion of an *E. coli* or phage λ gene (29). Nagai and Thogersen have generalized this approach by inserting a linker between a fragment of the *cII* gene (DNA coding for the first 31 amino acids of *cII*) and the gene of interest. The linker codes for the tetrapeptide sequence Ile-Glu-Gly-Arg, which serves as the recognition site for the protease factor Xa. Many previous workers have been frustrated in their attempts to express β -globin at high levels in *E. coli*. However, Nagai and Thogersen demonstrated both expression of the fusion product and cleavage by factor Xa to β -globin by NaDodSO₄/PAGE (yields and reconstitution were not reported).

The construction of pMb3 outlined in Fig. 2 replaces the cDNA for β -globin with that for human Mb in the expression vector pLcIIFX β . The region of the sequence in the vicinity of the Gly-1 codon was sequenced, demonstrating that the *cII*

gene fragment and the ATC-GAG-GGT-AGG sequence coding for the cleavage site (abbreviated FX) were as desired. pMb3 was introduced into MZ1 by transformation as described above and whole cells were analyzed by NaDodSO₄/PAGE (Fig. 3) revealing the presence of a new, intense band at apparent molecular mass about 21 kDa. As shown in the second and third lanes in Fig. 3, this band was present in cells that had been induced and not in those kept at 30°C. In addition to being at the correct molecular mass for the fusion product, this band was shown by immunoblotting to react with antibodies to human Mb. From these gels we estimate that on the order of 10% of total cell protein is the fusion product.

When induced cells were lysed by using a French press and spun down the desired protein was found to be in the pellet. This suggests that the protein aggregates and inclusion bodies are formed, so we followed the isolation procedure recently described in ref. 27. As shown in the fourth lane of the gel in Fig. 3, this procedure leads to isolation of the fusion protein at quite a high level of purity, and NaDodSO₄/PAGE analysis of the supernatant (not shown) demonstrates that very little of the fusion product is not sedimented. Reconstitution of the partially purified fusion product with heme was successful, as demonstrated by characteristic features in the electronic absorption spectrum (see below). It is well known that reconstituted Mbs are insensitive to proteolytic enzymes, whereas the apoproteins are rapidly degraded (1). Thus, we chose to use trypsin in place of factor Xa, as it will also cleave at the arginine residue preceding Gly-1 of human Mb and is much less expensive. As shown in lane 5 of Fig. 3, treatment of the reconstituted fusion protein with trypsin produces a protein of apparent molecular mass 17 kDa and digests essentially all other protein impurities as a bonus, minimizing the need for further purification. This protein (predicted molecular mass without heme = 17,053 Da) was mixed with sperm whale Mb (17,199 Da), horse heart Mb (16,951 Da), or authentic human Mb and subjected to NaDodSO₄/PAGE. The human protein migrated as if it had a slightly higher apparent molecular mass than the horse heart protein and about the same apparent molecular mass as the sperm whale protein, and it was indistinguishable from authentic human Mb (data not shown). From the amino acid sequence the closest undesirable tryptic fragments are predicted to have molecular masses of 17,936 or 14,982 Da, far outside the range observed. Thus, we are confident that the correct protein has been produced. All of the procedures described

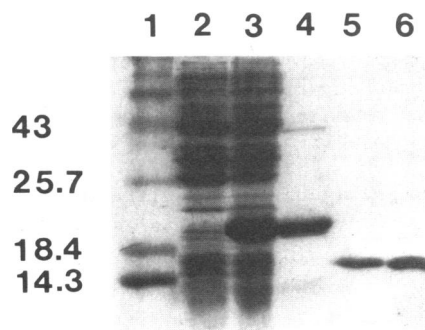


FIG. 3. Expression of human apoMb in *E. coli*. NaDodSO₄/PAGE (17.5%) of the following: Lane 1, protein molecular mass standards (kDa); lane 2, whole MZ1 cells containing pMb3 before induction; lane 3, whole cells after induction at 42°C for 2 hr, showing fusion protein at apparent molecular mass 21 kDa that cross-reacts with antibodies to human Mb; lane 4, precipitated fusion protein resolubilized in 8 M urea; lane 5, final product after reconstitution with heme, 12-hr digestion with trypsin, and purification; lane 6, sperm whale apoMb (17,199 Da).

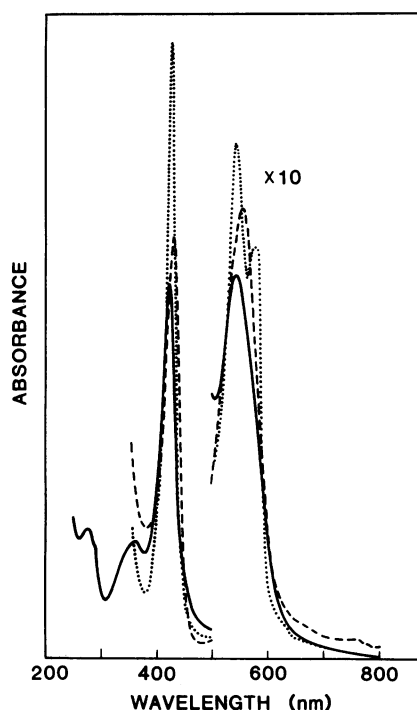


FIG. 4. Electronic absorption spectra of human Mb produced in *E. coli*. Solid line, cyanmet form; dashed line, deoxy form produced by reduction with excess sodium dithionite (excess dithionite obscures the near UV region, which is not shown for clarity); dotted line, carbonmonoxy form. Spectra were obtained on the same sample in the order cyanmet, deoxy, carbonmonoxy, so relative intensities are meaningful. These spectra are indistinguishable from those for authentic human Mb.

above have been successfully scaled up to produce human Mb on a gram scale.

The electronic absorption spectrum of Mbs provides a very sensitive measure of the integrity of the heme-apoprotein interaction and specific interactions with ligands. The absorption spectrum of the purified protein was examined under a variety of conditions, including the cyanmet, deoxy, and carbonmonoxy forms, as shown in Fig. 4. The absorption maxima and relative band intensities are identical to those of authentic human Mb and other related myoglobins (1). The ^1H NMR spectra at 500 MHz of the cyanmet forms of authentic Mb and that produced in *E. coli* were compared (data not shown). These spectra were identical for both hyperfine shifted heme and resolvable protein resonances and will be the subject of subsequent communications. These results demonstrate that the protein is fully functional with respect to ligand binding, the electronic structure of the heme, and the protein tertiary structure, setting the stage for detailed biophysical studies.

We are very grateful to David Lambright, who assisted with the isolation and reconstitution of the protein, Dr. Peter Gunning for the

gift of the cDNA library and assistance with screening, and J. Williamson, who assisted with the NMR spectroscopy. Drs. A. J. Jeffreys, L. Kedes, K. Nagai, A. Shatzman, and M. Zuber generously made available essential materials used in this work. We also thank Drs. D. Shortle, R. Fox, and P. Kim for very useful discussions. The 500-MHz NMR spectra were obtained at the Stanford Magnetic Resonance Laboratory, which is supported by National Science Foundation Grant GR23633 and National Institutes of Health Grant RR0711. This work was supported by National Science Foundation Grant PCM8303776, National Institutes of Health Grant GM27738, and Gas Research Institute Grant 82-260-0089. S.G.B. is a Dreyfus Teacher-Scholar Fellow.

1. Antonini, E. & Brunori, M. (1971) *Hemoglobin and Myoglobin in Their Reactions with Ligands* (North-Holland, Amsterdam).
2. Kendrew, J. C. (1962) *Brookhaven Symp. Biol.* **15**, 216-227.
3. Takano, T. (1977) *J. Mol. Biol.* **110**, 537-568.
4. Takano, T. (1977) *J. Mol. Biol.* **110**, 569-584.
5. Scouloidi, H. & Baker, E. N. (1978) *J. Mol. Biol.* **126**, 637-660.
6. Ladner, R. C., Heidner, E. J. & Perutz, M. F. (1977) *J. Mol. Biol.* **114**, 385-414.
7. Teale, F. W. J. (1959) *Biochim. Biophys. Acta* **35**, 543.
8. Rossi-Fanelli, A., Antonini, E. & Caputo, A. (1964) *Adv. Protein Chem.* **19**, 73-222.
9. Stryer, L. J. (1965) *J. Mol. Biol.* **13**, 482-495.
10. Wright, K. A. & Boxer, S. G. (1981) *Biochemistry* **20**, 7546-7556.
11. Wood, D., Blanchetot, A. & Jeffreys, A. J. (1982) *Nucleic Acids Res.* **10**, 7133-7144.
12. Blanchetot, A., Wilson, V., Wood, D. & Jeffreys, A. J. (1983) *Nature (London)* **301**, 732-734.
13. Weller, P., Jeffreys, A. J., Wilson, V. & Blanchetot, A. (1984) *EMBO J.* **3**, 439-446.
14. Nagai, K. & Thogersen, M. C. (1984) *Nature (London)* **309**, 810-812.
15. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
16. Gunsalus, R. P., Zurawski, G. & Yanofsky, C. (1979) *J. Bacteriol.* **140**, 106-113.
17. Birnboim, H. C. & Doly, J. (1979) *Nucleic Acids Res.* **7**, 1513-1523.
18. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499-560.
19. Okayama, H. & Berg, P. (1982) *Mol. Cell. Biol.* **2**, 161-170.
20. Gunning, P., Ponte, P., Blau, M. & Kedes, L. (1983) *Mol. Cell. Biol.* **3**, 1985-1995.
21. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503-517.
22. Grunstein, M. & Hogness, D. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 3961-3965.
23. Rosenberg, M., Ho, Y.-S. & Shatzman, A. (1983) *Methods Enzymol.* **101**, 123-164.
24. Messing, J. (1983) *Methods Enzymol.* **101**, 21-98.
25. Laemmli, U. K. (1970) *Nature (London)* **227**, 680-685.
26. Towbin, H., Staehelin, T. & Gordon, J. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 4350-4354.
27. Marston, F. A. O., Lowe, P. A., Doel, M. T., Schoemaker, J. M., White, S. & Angal, S. (1984) *Bio/Technology* **2**, 800-804.
28. Romeo-Herrera, A. E. & Lehmann, H. (1974) *Proc. R. Soc. London Ser. B* **186**, 249-279.
29. Harris, T. J. R. (1983) *Genetic Engineering* **4**, 127-185.