Philipp, M., & Maripuri, S. (1981) *FEBS Lett. 133*, 36–38.
Poland, D. C., & Scheraga, H. A. (1965) *Biopolymers 3*, 379–399.
Pollitt, S., & Zalkin, H. (1983) *J. Bacteriol. 153*, 27–32.
Privalov, P. L., & Khechinashvili, N. N. (1974) *J. Mol. Biol. 86*, 665–684.
Richardson, J. S. (1981) *Adv. Protein Chem. 34*, 167–339.
Sauer, R. T., Hehir, K., Stearman, R. S., Weiss, M. A., Jeitler-Nilsson, A., Suchanek, E. G., & Pabo, C. O. (1986) *Biochemistry 25*, 5992–5998.
Schellman, J. A. (1955) *C. R. Trav. Lab. Carlsberg, Ser. Chim. 29*, 223–240.
Schultz, S. C., & Richards, J. H. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 1588–1592.
Takahashi, K., & Sturtevant, J. M. (1981) *Biochemistry 20*, 6185–6190.
Tanford, C. (1970) *Adv. Protein Chem. 24*, 1–95.
Thornton, J. M. (1981) *J. Mol. Biol. 151*, 261–287.
Ueda, T., Yamada, H., Hirata, M., & Imoto, T. (1985) *Biochemistry 24*, 6316–6322.
Vasantha, N., Thompson, L. D., Rhodes, C., Banner, C., Nagle, J., & Filpula, D. (1984) *J. Bacteriol. 159*, 811–819.
Villafranca, J. E., Howell, E. E., Voet, D. H., Strobel, M. S., Ogden, R. C., Abelson, J. N., & Kraut, J. (1983) *Science (Washington, D.C.) 222*, 782–788.
Voordouw, G., Milo, C., & Roche, R. S. (1976) *Biochemistry 15*, 3716–3724.
Weber, P. C., Sheriff, S., Ohlendorf, D. H., Finzel, B. C., & Salemme, F. R. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 8473–8477.
Wells, J. A., & Powers, D. B. (1986) *J. Biol. Chem. 261*, 6564–6570.
Wells, J. A., Ferrari, E., Henner, D. J., Estell, D. A., & Chen, E. Y. (1983) *Nucleic Acids Res. 11*, 7911–7925.
Zoller, M. J., & Smith, M. (1983) *Methods Enzymol. 100*, 468–500.

Epidermal Growth Factor Binding Protein: Identification of a Different Protein†

Paul J. Isackson,‡,§ Robert E. Silverman,‖,⊥ Michael Blaber,‡ Alfred C. Server,#,△ Robert A. Nichols,#,○ Eric M. Shooter,# and Ralph A. Bradshaw*,‡,‖

Departments of Biological Chemistry and of Anatomy and Neurobiology, California College of Medicine, University of California, Irvine, California 92717, Department of Biological Chemistry, Washington University School of Medicine, St. Louis, Missouri 63110, and Department of Neurobiology, Stanford University School of Medicine, Stanford, California 94305

Received December 16, 1986; Revised Manuscript Received February 5, 1987

ABSTRACT: Partial amino acid sequence analysis of epidermal growth factor binding protein (EGF-BP), an arginine esteropeptidase that specifically associates with EGF to form a high molecular weight complex in male mouse submandibular glands, has revealed a single, distinct protein that is different from three previously reported forms of EGF-BP. This protein shows substantial sequence homology with these other putative forms of EGF-BP as well as with a large family of kallikreins expressed in the mouse submandibular gland. Purified EGF-BP contains three polypeptide chains as a result of two internal cleavages at residues 87–88 and 140–141. These modifications may represent processing events that are critical for determining the binding specificity of EGF-BP, since they occur within regions surrounding the substrate binding site.

Epidermal growth factor (EGF)[1] has a number of different biological effects, including potent mitogenic activity for a variety of tissue culture cells, and appears to play important roles in development and growth control throughout life (Cohen, 1962; Carpenter & Cohen, 1979). Sequence analysis of the cDNA clone for mouse EGF (Gray et al., 1983; Scott et al., 1983b) has revealed a much larger precursor form containing 1217 amino acids compared to the 53 amino acid mature form. ProEGF is predicted to be a membrane-bound protein due to the presence of a transmembrane segment near the C-terminus (Doolittle et al., 1984). In addition to EGF, the precursor contains seven or eight regions that show sequence homology to EGF (Gray et al., 1983; Scott et al., 1983b; Doolittle et al., 1984). It is unclear if EGF generally acts as a membrane-bound protein in cell–cell interactions (or even as a receptor) or if it is normally processed and acts as

† This work was supported by USPHS Research Grants NS19964, DK32465 (formerly AM32465), and NS04270 and American Cancer Society Grant BC-273G.

* Author to whom correspondence should be addressed.
‡ Department of Biological Chemistry, University of California.
§ Department of Anatomy and Neurobiology, University of California.
‖ Department of Biological Chemistry, Washington University School of Medicine.
⊥ Present address: Diabetes Programs Branch, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892.
Department of Neurobiology, Stanford University School of Medicine.
△ Present address: Department of Neurology, University of Texas Medical School, Houston, TX 77030.
○ Present address: Laboratory of Molecular and Cellular Neuroscience, Rockefeller University, New York, NY 10021.

[1] Abbreviations: EGF, epidermal growth factor; EGF-BP, epidermal growth factor binding protein; β-NGF, β-subunit of mouse nerve growth factor; γ-NGF, γ-subunit of mouse nerve growth factor; α-NGF, α-subunit of mouse nerve growth factor; HMW-EGF, high molecular weight EGF; TAME, N^{α}-*p*-tosylarginine methyl ester; BAPNA, N^{α}-benzoylarginine-*p*-nitroanilide; DEAE, diethylaminoethyl; Tris, tris(hydroxymethyl)aminomethane; Bis-Tris, [bis(2-hydroxyethyl)amino]tris(hydroxymethyl)methane; TES, 2-[[2-hydroxy-1,1-bis(hydroxymethyl)ethyl]amino]ethanesulfonic acid; SDS, sodium dodecyl sulfate; HPLC, high-performance liquid chromatography.

0006-2960/87/0426-2082$01.50/0 © 1987 American Chemical Society

a soluble trophic factor (Rall et al., 1985). In view of these unusual features of the EGF precursor, elucidation of the proteolytic processing events that produce mature EGF will be important to understanding its mode of action.

EGF was originally isolated from male mouse submandibular glands (Cohen, 1962), and this remains the richest source of the protein (Rall et al., 1985). EGF is present in these extracts as a complex with a specifically bound arginine esteropeptidase, termed epidermal growth factor binding protein (EGF-BP) (Taylor et al., 1974). EGF-BP interacts with the C-terminal arginine residue of mature EGF (Server et al., 1976) and has been suggested to be responsible for proEGF processing in the mouse submandibular gland. Purified EGF-BP binds EGF in vitro (Server et al., 1976; Server & Shooter, 1976) but does not bind to the β-subunit of nerve growth factor (β-NGF). Conversely, γ-NGF, a similar esteropeptidase present in submandibular glands, binds β-NGF but not EGF (Server & Shooter, 1976). The abundance and activity of EGF-BP in other tissues containing EGF are not known.

There are several inconsistent reports in the literature regarding the sequence and hence the identity of EGF-BP. Two different forms (A and B) have been reported to exist by Anundi et al. (1982). Partial amino acid sequences have been reported for both forms (Anundi et al., 1982), and a complete cDNA sequence had been published for form B (Ronne et al., 1983; Lundgren et al., 1984). Although the sequence of form B has been the one generally reported in the literature as that of EGF-BP (Watt et al., 1986; Ashley & MacDonald, 1985), neither of the forms described by these workers (Anundi et al., 1982) has been shown to bind EGF in vitro. Protease D, isolated by Boesman et al. (1976), was reported to be the same as EGF-BP by immunological and kinetic criteria and has been shown to recombine with EGF in vitro (Hosoi et al., 1983).

In view of these discrepancies, we have reexamined the homogeneity of EGF-BP purified from the high molecular weight EGF complex. We have found a single, unique protein sequence that is different from any of the proteins previously described. This protein is similar to but not identical with protease A (Schenkein et al., 1981) and to form A of Anundi et al. (1982); it is much less similar to form B for which the cDNA sequence has been reported (Ronne et al., 1983; Lundgren et al., 1984).

Materials and Methods

EGF Binding Protein. EGF-BP was isolated from a purified preparation of high molecular weight EGF (HMW-EGF). The latter was prepared according to the procedure of Taylor et al. (1970), starting with submandibular glands from 200 Swiss Webster adult male mice, and further purified on a Sephadex G-75 column (2.5 × 86 cm) in 24 mM sodium/potassium phosphate buffer, pH 6.7. EGF-BP and EGF were isolated from the complex by chromatography on DEAE-cellulose (1.5 × 25 cm) in 0.02 M Tris-HCl buffer, pH 7.5, using an NaCl gradient (Taylor et al., 1970).

The EGF-BP was characterized by its specific enzymatic activity, its amino acid composition, and especially its ability to recombine with EGF to form HMW-EGF. The arginine esteropeptidase activity was measured by using either N^{α}-*p*-tosylarginine methyl ester (TAME) or N^{α}-benzoylarginine-*p*-nitroanilide (BAPNA) as substrates for comparison with the known kinetic constants of EGF-BP (Nichols & Shooter, 1983). The preparations used in this work gave similar values for these constants. The amino acid composition was determined from acid hydrolysates with a Durrum D500 analyzer and agreed with the published values for EGF-BP originally reported by Taylor et al. (1974). The recombination of EGF-BP and EGF was monitored by either gel filtration or electrophoresis as described by Server et al. (1976). In the former, equal amounts (4 mg) of the two proteins were incubated in 24 mM sodium/potassium phosphate buffer, pH 6.7 (3.6 mL), for 16 h at 4 °C prior to analysis on Sephadex G-100 (2.5 × 90 cm) equilibrated in the same buffer. The complex formed in this experiment had the same elution volume as HMW-EGF, and electrophoretic analysis of this complex in Bis-Tris–TES at pH 7.05 (Server & Shooter, 1976) showed that its major protein component migrated at the same rate as that of HMW-EGF. In similar experiments where the reaction products were analyzed by electrophoresis in the above system (Server & Shooter, 1976), excess EGF (50 μg) when incubated with EGF-BP (50 μg) caused quantitative conversion of the latter into HMW-EGF. These experiments demonstrate that the EGF-BP preparations used in these studies meet the criteria originally defined by Taylor et al. (1970), i.e., the enzyme specifically associates with EGF to form the HMW-EGF complex.

Peptide Separations and Amino Acid Sequence Analysis. Purified EGF-BP (10 mg) was reduced and carboxymethylated as described previously (Angeletti et al., 1971). The modified protein was dialyzed against 50 mM ammonium bicarbonate with Spectrapor dialysis tubing (1000 molecular weight cutoff) and lyophilized. Polypeptide chains were separated by high-performance liquid chromatography (Beckman/Altex) on a Vydac C4 reverse-phase column. The purified polypeptides were sequenced with an Applied Biosystems Model 470A gas phase sequenator. Phenylthiohydantoins were analyzed quantitatively with a Beckman microsphere C18 reverse-phase column on a Hewlett-Packard high-performance liquid chromatograph, Model 1084B (Thomas et al., 1981b).

Results and Discussion

Purified EGF-BP showed only a single band on SDS–polyacrylamide gel electrophoresis in the absence of reducing agents. However, SDS gel electrophoresis following reduction or direct sequence analysis indicated the presence of three polypeptide chains in equal amounts. Accordingly, EGF-BP was reduced and carboxymethylated, and the three fragments were separated by reverse-phase high-performance liquid chromatography using a C4 column (Figure 1). Complete resolution was achieved as judged by electrophoresis and amino acid sequence determinations. In the latter analyses, identifications of the indicated residues (Figure 2) for each peak were made on samples of 600, 1800, and 1400 pmol for peaks 1, 2, and 3, respectively. No residue was assigned for cycle 14 of peak 1 and cycles 21 and 24 of peak 3. The sum of the three individual sequence determinations exactly matched the aggregate analysis, indicating that the three components separated by HPLC account for all of the protein present in the preparation.

The three N-terminal sequences can be readily aligned with the sequence of γ-NGF (Thomas et al., 1981a) (or any of the other full kallikrein sequences reported) (Figure 2). Peak 2 contains the amino terminus, peak 1, the middle section, and peak 3, the carboxyl terminus. This alignment is depicted schematically in Figure 3. Only chain B (peak 1) showed any heterogeneity; an HPLC peak that eluted close to the phenylthiohydantoin of phenylalanine was observed in the first cycle. Phenylalanine has been found in this position in the γ-NGF subunit but not in other proteins identified as EGF-BP. It may represent an allotypic variation or a sequencing artifact.

The N-terminal sequences of the three chains, compared in Figure 2, are clearly different from the sequences of

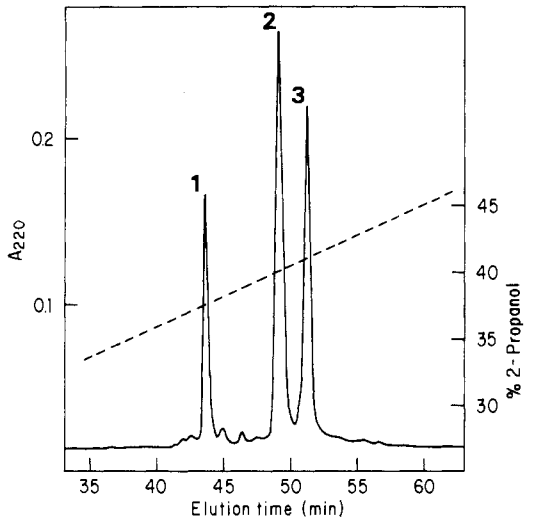

FIGURE 1: Elution profile of the separation of the constituent polypeptide chains of EGF-BP on reverse-phase HPLC. Reduced and carboxymethylated EGF-BP was applied to a Vydac C4 column equilibrated in 0.1% trifluoroacetic acid. The polypeptide chains were eluted with a linear gradient of 2-propanol.

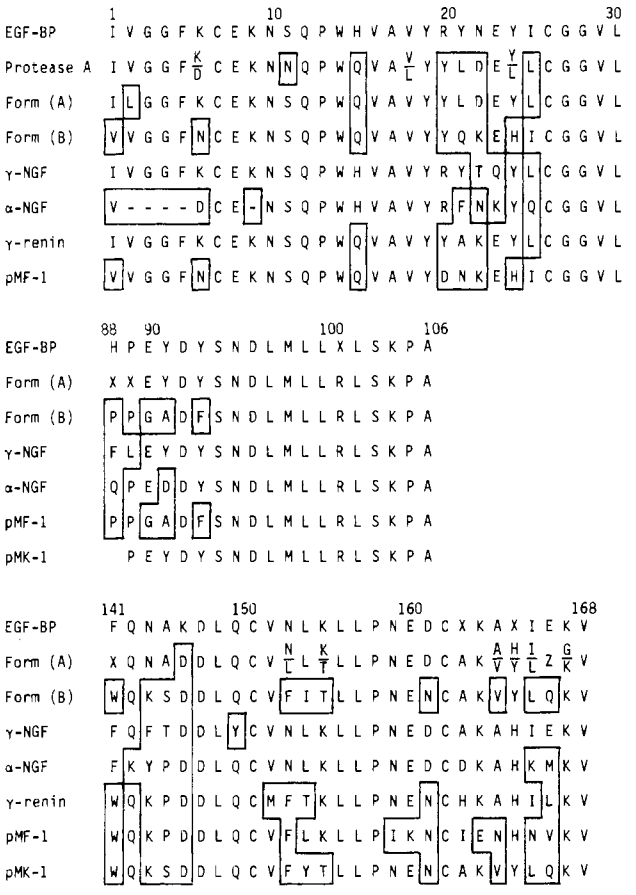

FIGURE 2: Comparison of the partial amino acid sequences of several mouse submandibular gland kallikreins. Boxes indicate residues that are different from EGF-BP isolated in this study. Sequences shown are as follows: EGF-BP (this study); protease A (Schenkein et al., 1981); form A and form B (Anundi et al., 1982); γ-NGF (Thomas et al., 1981a,b); α-NGF (Isackson et al., 1984); γ-renin (Poe et al., 1983); pMF-1 (Fahnestock et al., 1986); pMK-1 (Richards et al., 1982). One-letter code used for amino acids: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; X, unknown; Y, Tyr; Z, Glx. Residue numbers correspond to γ-NGF (Ullrich et al., 1984).[2]

EGF-BP reported by others (Anundi et al., 1982; Lundgren et al., 1984; Schenkein et al., 1981). Our sequence is similar

FIGURE 3: Schematic representation of the polypeptide chains of EGF-BP. Peptides A, B, and C correspond to peaks 2, 1, and 3, respectively, which were separated by HPLC (see Figure 1). Solid regions represent portions that were sequenced.

to that of form A of Anundi et al. (1982), with only seven differences in the 71 residues compared, but much different from that of form B (Lundgren et al., 1984), which has 21 differences. There are 6 differences between our EGF-BP and protease A (Schenkein et al., 1981) of the 30 N-terminal amino acids that were compared. Interestingly, protease A and form A of Anundi et al. (1982) appear to be very similar proteins since they differ in only 3 of their 40 N-terminal amino acids. Since the differences are at amino acid residues where the assignments could easily be incorrect, they may, in fact, be identical.

The mouse genome has been found to contain a cluster of 25–30 genes encoding arginine esteropeptidases that are homologous to each other and to glandular kallikrein (Mason et al., 1983). Most of these genes appear to be expressed in the submandibular gland (Ullrich et al., 1984; Isackson et al., 1986). Members of this family include γ-NGF (Thomas et al., 1981a), α-NGF (Isackson & Bradshaw, 1984), γ-renin (Poe et al., 1983), protease A (Schenkein et al., 1981), and the two putative forms of EGF-BP reported by Anundi et al. (1982). On the basis of sequence comparisons, the EGF-BP characterized in this study is clearly a distinct member of this family. However, it does appear to be functionally identical with protease D (Boesman et al., 1976) in that both recombine with EGF to form a high molecular weight complex (Server & Shooter, 1976; Hosoi et al., 1983) and both are completely processed to three polypeptide chains as isolated from HMW-EGF. No sequence data have been reported for protease D.

The three-chain composition of EGF-BP may be a critical structural feature required for its binding specificity. Amino acid sequence analysis of EGF-BP indicates that the internal cleavages occur before residues 88 and 141 (Figure 2).[2] These cleavage sites are within regions of the sequence predicted to form surface loops characteristic of glandular kallikreins (Bode et al., 1983). X-ray crystal structure analysis of porcine pancreatic kallikrein (Chen & Bode, 1983) has shown five surface loops, which include amino acid residues 18–26, 52–63, 89–93, 133–145, and 157–167 [using the same numbering scheme as for γ-NGF (Ullrich et al., 1984)]. The surface loop regions exhibit the greatest sequence variability within the kallikrein family (Mason et al., 1983; Isackson et al., 1986). These surface loops partially obstruct the substrate binding site and may be involved in determining the specificity of binding to large protein substrates (Bode et al., 1983; Chen & Bode, 1983). Amino acid residues 93, 143, and 145 [corresponding to amino acid residues 93, 141, and 143 of γ-NGF (Ullrich et al., 1984)] of porcine kallikrein, which are immediately adjacent to the cleavage sites, have been predicted by X-ray crystal structure analysis to make contacts with

[2] Residue numbers have been assigned by alignment with the amino acid sequence of γ-NGF, as deduced from the cDNA sequence (Ullrich et al., 1984). This numbering differs from that of Thomas et al. (1981a) in the C-terminal portion of the molecule (by four residues) due to the excision of a tetrapeptide that occurs during the formation of the mature protein (as found in the 7S complex).

bovine pancreatic trypsin inhibitor (Chen & Bode, 1983). We suggest that proteolytic cleavage in these regions may cause movement of neighboring amino acid side chains, which results in altered specificity toward large protein substrates. EGF-BP may thus require processing at internal sites before achieving a conformation that recognizes EGF. A similar event, but with opposite consequences, occurs with thrombin, in that two cleavages occur at positions equivalent to those in EGF-BP and produce γ-thrombin, which has a greatly reduced ability to cleave fibrinogen (Chang et al., 1979).

The role of EGF-BP in the processing of proEGF is not well-defined. It apparently binds the C-terminal arginine residue of mature EGF (Server et al., 1976) and has been shown to cleave the C-terminal extension of an EGF precursor of M_r 9000 (Frey et al., 1979). The mature sequence of EGF is flanked by single basic (arginine) residues within proEGF (Gray et al., 1983; Scott et al., 1983b), and kallikreins are good candidates to be the physiological processors of such sites (Mason et al., 1983). This differs from NGF (Scott et al., 1983a; Ullrich et al., 1983), which is flanked by dibasic residues in its precursor structure and is likely to be processed by a different type of enzyme (Loh & Gainer, 1982).

The high degree of sequence identity among these proteases (80–95%) makes this an attractive system for elucidating the structural features of glandular kallikreins that confer binding specificity. We are currently isolating the cDNA clone of EGF-BP for a more thorough comparison.

ACKNOWLEDGMENTS

We thank Eva Lujan and Chris E. Bradshaw for technical assistance with portions of this work. The sequence analyses were performed under the direction of Stephen S. Disper in the Protein/Nucleic Acid Analysis Laboratory, Department of Biological Chemistry, University of California, Irvine, CA 92717. We also thank Marja Uskali for expert preparation of the manuscript.

REFERENCES

Angeletti, R. H., Bradshaw, R. A., & Wade, R. D. (1971) *Biochemistry 10*, 463–469.

Anundi, H., Ronne, H., Peterson, P. A., & Rask, L. (1982) *Eur. J. Biochem. 129*, 365–371.

Ashley, P. L., & MacDonald, R. J. (1985) *Biochemistry 24*, 4512–4520.

Bode, W., Chen, Z., & Bartels, K. (1983) *J. Mol. Biol. 164*, 237–282.

Boesman, M., Levy, M., & Schenkein, I. (1976) *Arch. Biochem. Biophys. 175*, 463–475.

Carpenter, G., & Cohen, S. (1979) *Annu. Rev. Biochem. 48*, 193–216.

Chang, T.-L., Feinman, R. D., Landis, B. H., & Fenton, J. W. (1979) *Biochemistry 18*, 113–119.

Chen, Z., & Bode, W. (1983) *J. Mol. Biol. 164*, 237–282.

Cohen, S. (1962) *J. Biol. Chem. 237*, 1555–1562.

Doolittle, R. F., Feng, D. F., & Johnson, M. D. (1984) *Nature (London) 307*, 558–560.

Fahnestock, M., Brundage, S., & Shooter, E. M. (1986) *Nucleic Acids Res. 14*, 4823–4835.

Frey, P., Forand, R., Maciag, T., & Shooter, E. M. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 6294–6298.

Gray, A., Dull, T. J., & Ullrich, A. (1983) *Nature (London) 303*, 722–725.

Hosoi, K., Kamiyama, S., Atsumi, T., Nemoto, A., Tanaka, I., & Ueha, T. (1983) *Arch. Oral Biol. 28*, 5–11.

Isackson, P. J., & Bradshaw, R. A. (1984) *J. Biol. Chem. 259*, 5380–5383.

Isackson, P. J., Ullrich, A., & Bradshaw, R. A. (1984) *Biochemistry 23*, 5997–6002.

Isackson, P. J., Dunbar, J. C., & Bradshaw, R. A. (1986) *J. Cell. Biochem. 33*, 65–75.

Loh, Y. P., & Gainer, H. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 108–112.

Lundgren, S., Ronne, H., Rask, L., & Peterson, P. A. (1984) *J. Biol. Chem. 259*, 7780–7784.

Mason, A. J., Evans, B. A., Cox, D. R., Shine, J., & Richards, R. I. (1983) *Nature (London) 303*, 300–307.

Nichols, R. A., & Shooter, E. M. (1983) *J. Biol. Chem. 258*, 10296–10303.

Poe, M., Wu, J. K., Florance, J. R., Rodkey, J. A., Bennett, C. D., & Hoogsteen, K. (1983) *J. Biol. Chem. 258*, 2209–2216.

Rall, L. B., Scott, J., Bell, G. I., Crawford, R. J., Penschow, J. D., Niall, H. D., & Coghlan, J. P. (1985) *Nature (London) 313*, 228–231.

Richards, R. I., Catanzaro, D. F., Mason, A. J., Morris, B. J., Baxter, J. D., & Shine, J. (1982) *J. Biol. Chem. 257*, 1453–1463.

Ronne, H., Lundgren, S., Severinsson, L., Rask, L., & Peterson, P. A. (1983) *EMBO J. 2*, 1561–1564.

Schenkein, I., Franklin, E. C., & Frangione, B. (1981) *Arch. Biochem. Biophys. 209*, 57–62.

Scott, J., Selby, M., Urdea, M., Quiroga, M., Bell, G. I., & Rutter, W. J. (1983a) *Nature (London) 302*, 538–540.

Scott, J., Urdea, M., Quiroga, M., Sanchez-Pescador, R., Fong, N., Selby, M., Rutter, W. J., & Bell, G. I. (1983b) *Science (Washington, D.C.) 221*, 236–240.

Server, A. C., & Shooter, E. M. (1976) *J. Biol. Chem. 251*, 165–173.

Server, A. C., Sutter, A., & Shooter, E. M. (1976) *J. Biol. Chem. 251*, 1188–1196.

Taylor, J. M., Cohen, S., & Mitchell, W. M. (1970) *Proc. Natl. Acad. Sci. U.S.A. 67*, 164–171.

Taylor, J. M., Mitchell, W. M., & Cohen, S. (1974) *J. Biol. Chem. 249*, 2188–2194.

Thomas, K. A., Baglan, N. C., & Bradshaw, R. A. (1981a) *J. Biol. Chem. 256*, 9156–9166.

Thomas, K. A., Silverman, R. E., Jeng, I., Baglan, N. C., & Bradshaw, R. A. (1981b) *J. Biol. Chem. 256*, 9147–9155.

Ullrich, A., Gray, A., Berman, C., & Dull, T. J. (1983) *Nature (London) 303*, 821–825.

Ullrich, A., Gray, A., Wood, W. I., Hayflick, J., & Seeburg, P. H. (1984) *DNA 3*, 387–392.

Watt, K. W. K., Lee, P.-J., M'Timkulu, T., Chan, W.-P., & Loor, R. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3166–3170.