Miksicek, R., Heber, A., Schmid, W., Danesch, U., Posseckert, G., Beato, M., & Schütz, G. (1986) *Cell (Cambridge, Mass.)* *46*, 283–290.
Mount, S. M. (1982) *Nucleic Acids Res. 10*, 459–472.
Podack, E. R., Kolb, W. P., & Müller-Eberhard, H. J. (1977) *J. Immunol. 119*, 2024–2029.
Podack, E. R., Kolb, W. P., & Müller-Eberhard, H. J. (1978) *J. Immunol. 120*, 1841–1848.
Poustka, A., Rackwitz, H.-R., Frischauf, A.-M., Hohn, B., & Lehrach, H. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 4129–4133.
Preissner, K. T., Wassmuth, R., & Müller-Berghaus, G. (1985) *Biochem. J. 231*, 349–355.
Pytela, R., Pierschbacher, M. D., Suzuki, S., & Ruoslahti, E. (1985) *Proc. Natl. Acad. Sci. U.S.A. 60*, 245–258.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.
Schmid, C. W., & Jelinek, W. R. (1982) *Science (Washington, D.C.) 216*, 1065–1070.
Stanley, K. K. (1986) *FEBS Lett. 199*, 249–253.
Stanley, K. K., & Luzio, J. P. (1984) *EMBO J. 3*, 1429–1434.
Suzuki, S., Pierschbacher, M. D., Hayman, E. G., Nguyen, K., Öhgren, Y., & Ruoslahti, E. (1984) *J. Biol. Chem. 259*, 15307–15314.
Suzuki, S., Oldberg, A., Hayman, E. G., Pierschbacher, M. D., & Ruoslahti, E. (1985) *EMBO J. 4*, 2519–2524.
Tschopp, J., & Mollnes, T.-E. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 4223–4227.
Tschopp, J., Masson, D., & Stanley, K. K. (1986) *Nature (London) 322*, 831–834.

A Complete cDNA Sequence for the Major Epidermal Growth Factor Binding Protein in the Male Mouse Submandibular Gland†

Michael Blaber,‡ Paul J. Isackson,‡,§ and Ralph A. Bradshaw*,‡
Departments of Biological Chemistry and of Anatomy and Neurobiology, California College of Medicine, University of California, Irvine, California 92717
Received March 30, 1987

ABSTRACT: The complete cDNA sequence of the major epidermal growth factor binding protein (EGF-BP), isolated from the mouse submandibular gland, has been determined. Oligonucleotide probes complementary to unique, nonconserved, regions of homogeneous preparations of EGF-BP were used to identify the correct cDNA clone from a male mouse submandibular gland cDNA library. The nucleotide sequence codes for a glandular kallikrein that is the main arginine esterase complexed with epidermal growth factor. The mRNA coding for this EGF-BP is estimated at 0.24% of the total mRNA of the adult male mouse submandibular gland, thus representing an abundant member of the kallikrein family in this tissue. In addition, the cDNA sequence defines a putative transcription start site. The reported cDNA sequence is clearly different from, and not an allelic form of, a previously reported cDNA sequence for EGF-BP. The present work reconciles conflicting information in the literature regarding the identity of EGF-BP.

The mouse submandibular gland is a rich source of both epidermal growth factor (EGF)[1] and nerve growth factor (NGF) (Cohen, 1960, 1962). Both are isolated from the mouse submandibular gland as high molecular weight complexes containing the arginine-specific esterases EGF binding protein (EGF-BP) and γ-NGF, respectively (Taylor et al., 1970; Varon et al., 1968). These esterases are serine proteases of the glandular kallikrein family, 25 closely related members of which are localized on mouse chromosome 7 (Mason et al., 1983; Evans et al., 1987). Glandular kallikreins have been defined as being able to cleave kininogen to release kinins, potent vasodilators (Orstavik, 1980). However, this specific kallikrein activity may vary considerably among the various members of this family. Since both EGF and NGF are translated as high molecular weight precursors that are processed to yield the mature active growth factor (Scott et al., 1983a,b; Ullrich et al., 1983; Gray et al., 1983), these specific esterases that are bound to the mature growth factors in the mouse submandibular gland have been postulated to be involved (Angeletti & Bradshaw, 1971; Frey et al., 1979; Berger & Shooter, 1977). Thus, EGF-BP and γ-NGF may be involved in the regulation of the levels of the mature growth factors.

The partial amino acid sequence of a preparation of EGF-BP has been reported by Anundi et al. (1982). Two forms of EGF-BP, type "A" and type "B", were reported by this group to be present in the male mouse submandibular gland. This preparation of EGF-BP was characterized as being similar to the EGF-BP described by Taylor et al. (1970) by the criteria of amino acid composition and cross-reactivity to a polyclonal anti-EGF-BP antiserum, although neither form was demonstrated to complex with EGF in vitro. Partial and full-length cDNA clones corresponding to type B EGF-BP have been reported (Ronne et al., 1983; Lundgren et al., 1984).

A comparison of the partial amino acid sequences reported by Anundi et al. (1982) with a partial sequence determined in our laboratory (Silverman, 1977) revealed significant discrepancies that were subsequently confirmed (Isackson et al.,

† This work was supported by USPHS Research Grant NS19964 and by American Cancer Society Grant BC273.
* Address correspondence to this author.
‡ Department of Biological Chemistry.
§ Department of Anatomy and Neurobiology.

[1] Abbreviations: EGF, epidermal growth factor; EGF-BP, epidermal growth factor binding protein; NGF, nerve growth factor; γ-NGF, γ-subunit of mouse 7S NGF; β-NGF, β-subunit of mouse NGF; bp, base pair(s); Tris, tris(hydroxymethyl)aminomethane; EDTA, ethylenediaminetetraacetic acid; SDS, sodium dodecyl sulfate; Da, dalton(s).

0006-2960/87/0426-6742$01.50/0

```
      141          145
     -Phe-Gln-Asn-Ala-Lys-

3'  -AAA-GTT-TTA-CGX-TT-   5'
      G   C   G

      17                        24
     -Ala-Val-Tyr-Arg-Tyr-Asn-Glu-Tyr-

3'  -CGA-CAC-ATG-GCA-ATG-TTG-CTT-ATA-   5'
```

FIGURE 1: (A) Synthetic mixed oligonucleotide utilized as a hybridization probe in screening for EGF-BP cDNA clones. The oligonucleotide is complementary to the coding strand for the indicated amino acid sequence. "X" indicates incorporation of all four nucleotide bases at that position. The numbering system of the amino acid sequence is used (Figure 2). (B) Synthetic oligonucleotide utilized as a sequencing primer for the 5'-terminal region of EGF-BP cDNA clones. The oligonucleotide is complementary to the coding strand for the indicated amino acid sequence (Figure 2).

1987). The EGF-BP described by Isackson et al. (1987) was isolated from the previously described high molecular weight complex (Taylor et al., 1970) and proved to be a unique, homogeneous, and distinctly different protein from the previously described EGF-BP type A (Anundi et al., 1982) or type B (Ronne et al., 1983; Lundgren et al., 1984). Most importantly, this glandular kallikrein was shown to complex with EGF in vitro.

Here we report the entire cDNA sequence corresponding to this protein and present evidence that it is the major arginine esterase complexed with EGF in the male mouse submandibular gland and not an allelic representative of the previously described EGF-BP's.

Experimental Procedures

Construction of a cDNA Library. RNA was isolated from submandibular glands of adult male Swiss Webster mice by homogenization in guanidine thiocyanate followed by centrifugation through a CsCl pad (Chirgwin et al., 1979). Poly(A+) RNA was purified by chromatography on oligo(dT)–cellulose (Aviv & Leder, 1972). cDNA was synthesized by using the method of Gubler and Hoffman (1983), followed by liquid chromatographic fractionation utilizing Bio-Gel A-50 (Bio-Rad Laboratories). Pooled fractions containing cDNAs from 500 to 1500 bp in length (representing approximately 80% of the total cDNA) were subsequently tailed with dC and annealed to dG-tailed pBR322 that had been *Pst*I-digested (Bethesda Research Laboratories). The annealed cDNA/vector was used to transform *Escherichia coli* strain DH5α (Bethesda Research Laboratories), and transformants were selected by using LB agar plates containing 12.5 μg/mL tetracycline (Maniatis et al., 1982). A "library" of approximately 5000 clones was subsequently replica-plated onto nitrocellulose filters (Schleicher & Schuell) and amplified on LB agar plates containing 200 μg/mL chloramphenicol prior to probe hybridization (Maniatis et al., 1982).

Oligonucleotide Probe Design and Colony Hybridization. A partial amino acid sequence of a specific, homogeneous EGF binding protein has been previously described (Silverman, 1977; Isackson et al., 1987). The region of amino acid sequence corresponding to residues 141–145[2] of the mature protein was chosen for (complementary) back-translation into a mixed 14-mer oligonucleotide (Figure 1A). This region of the mature protein corresponds to a highly variable "surface loop" of the three-dimensional structure of kallikreins (Bode et al., 1983) and may be expected to be unique for individual members of the family. The probe was chemically synthesized by using the solid phase phosphoramidite method (Caruthers et al., 1982) in an Applied Biosystems Inc. Model 380A DNA synthesizer. Replica-plated nitrocellulose filters containing the cDNA "library" were prehybridized overnight at room temperature in 100 mM Tris, 5 mM EDTA, 100 μg/mL poly(adenylic acid), 0.5% Nonidet P-40, 1 × Denhardt's solution, 1 M NaCl, 1 mM sodium pyrophosphate, and 0.1 mM ATP ("prehybridization buffer"). Colony hybridization with ^{32}P-labeled probe (Maniatis et al., 1982) was carried out overnight at room temperature in prehybridization buffer. Hybridized filters were washed under high stringency conditions using two 20-min washes at 45 °C in 3 M tetramethylammonium chloride, 50 mM Tris, 2 mM EDTA, and 0.1% SDS (Ullrich et al., 1984). Washed filters were then exposed overnight to X-ray film (Kodak XAR5).

Sequencing of Probe Positive Clones. Subcloning and sequencing utilized the M13 single-strand phage system (Messing et al., 1977) and the dideoxynucleotide method of Sanger et al. (1977). All sequencing reagents were purchased from New England Biolabs with the exception of [α-^{32}P]dATP, which was purchased from ICN Pharmaceuticals.

Other Materials. T4 DNA ligase and all restriction enzymes were purchased from New England Biolabs.

Results

Identification of a cDNA Clone Corresponding to the Known EGF-BP Amino Acid Sequence. A male mouse submandibular gland cDNA library of approximately 5000 clones was screened with a mixed-oligonucleotide probe (14-mer) complementary to the EGF-BP protein [as reported by Isackson et al. (1987)]. The sequence corresponds to a highly variable surface loop of the molecule and is unique to EGF-BP when compared to other kallikreins (Figure 2). Of the approximately 5000 colonies, 22 were positive when screened with this probe. Three of these were subjected to direct sequencing of an internal region of the cDNA inserts using the hybridization probe as a sequencing primer. Since it was complementary to the coding strand, this sequencing primer allowed the cDNA insert to be sequenced toward the 5' end of the molecule regardless of the orientation in the pBR322 plasmid. Sequence information was thus determined for a neighboring 5' region corresponding to another highly variable surface loop, the kallikrein loop (amino acids 88–91) (Bode et al., 1983), for which the amino acid sequence had also been determined (Isackson et al., 1987). In this fashion, one of the three positive clones (MB1-73) initially sequenced was identified as the cDNA corresponding to this amino acid sequence. The other two clones were not identifiable as members of the kallikrein family. The EGF-BP cDNA clone was sequenced (in both direction) in its entirety by subcloning into an M13 single-strand phage system (New England Biolabs). This cDNA insert, which was determined to be a partial clone that was approximately 27 nucleotides short of the initiation methionine (based on homology to other glandular kallikreins) (Drinkwater et al., 1987; Ullrich et al., 1984; Lundgren et al., 1984; Mason et al., 1983), was identified as the correct EGF-BP clone on the basis of identity with our partial amino acid sequence.

A series of restriction endonuclease (*Alu*I, *Dde*I, *Hpa*II, *Nla*IV, and *Rsa*I) digestion maps of the partial EGF-BP clone were used to identify other EGF-BP sequences from among the remaining probe-positive clones, 11 of which mapped as EGF-BP. Thus, of the 5000-clone library, 12 clones or ap-

[2] Amino acid and nucleotide residue numbers are defined by beginning with the initiation methionine and the putative transcription initiation adenosine (see Figure 3).

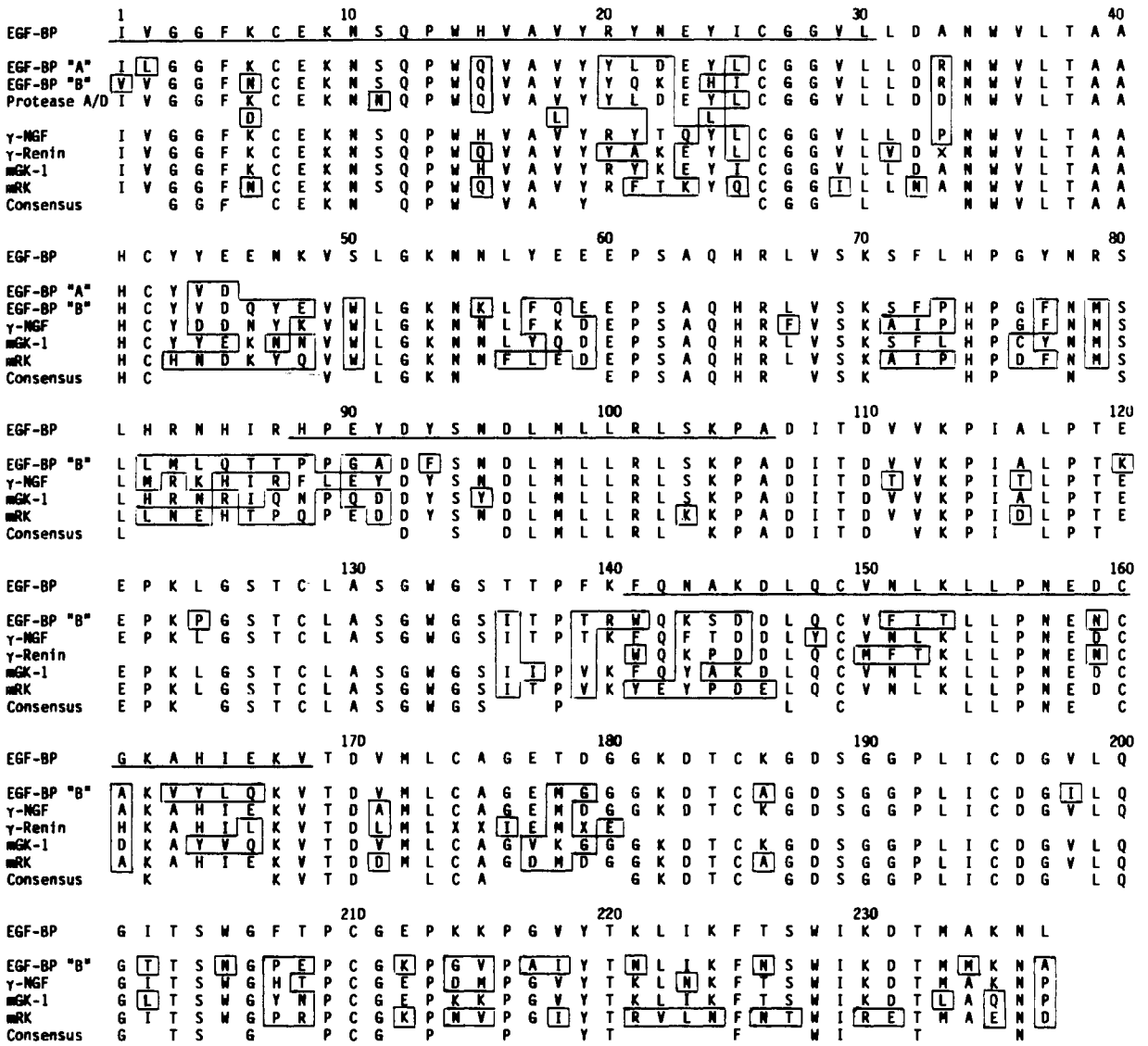

FIGURE 2: Amino acid alignments: EGF-BP (from cDNA; this report); EGF-BP type A (Anundi et al., 1982); EGF-BP type B (from cDNA; Lundgren et al., 1984); protease A/D (Schenkein et al., 1981); γ-NGF (from cDNA; Ullrich et al., 1984); γ-renin (Poe et al., 1983); mGK-1 (mouse glandular kallikrein) (from cDNA; Mason et al., 1983); mRK (mouse renal kallikrein) (from cDNA; van Leeuwen et al., 1986). The underlined regions of EGF-BP correspond to residues for which partial protein sequences were obtained (Isackson et al., 1987). Boxed regions correspond to amino acid positions which differ from those of EGF-BP. Consensus residues are presented in the bottom row.

proximately 0.24% of the mRNA molecules represented EGF-BP. The two largest of these inserts were restricted with *Pst*I and subcloned in their entirety into an M13 vector. The 5′ regions of the cDNA inserts were sequenced in one direction by utilizing the universal M13 primer (with clones of the correct orientation). Sequencing in the opposite direction was accomplished by using a synthetic oligonucleotide (24-mer) as a sequencing primer complementary to a unique 5′ region of clone MB1-73 (Figure 1B). One of these clones (MB2-20A) was an apparently complete clone (Figure 3), beginning with a putative transcription initiation adenylate residue (Drinkwater et al., 1987) and coding for the signal peptide (18 amino acids) and zymogen activation hexapeptide (Ullrich et al., 1984). The cDNA sequence of clone MB2-20A differed from the genomic sequence, mGK-9 (Drinkwater et al., 1987), at three positions: one occurs in the 5′ noncoding region and two in the 3′ noncoding region of the gene. We conclude that these changes are due to allelic polymorphisms or strain variations.

Discussion

The mouse glandular kallikreins comprise a highly homologous 25-member gene family located on chromosome 7 (Mason et al., 1983; Evans et al., 1987). As a group, they have been estimated to represent approximately 5–10% of the male mouse submandibular gland mRNA (Ullrich et al., 1984; Richards et al., 1982). Thus, EGF-BP (as reported here), representing approximately 0.24% of the mRNA in this tissue, is present as one of the major kallikreins. Two possible transcription start sites are identified in the genomic sequence (Drinkwater et al., 1987). The EGF-BP clone MB2-20A starts precisely at the adenosine of the second transcription start site, suggesting that this may be the actual position utilized.

The identification and characterization of the individual kallikreins are complicated by their high degree of identity. The different proteins may be expected to yield similar amino acid compositions (Table I) as well as share regions of extensive sequence similarity (Figure 2). This relatedness extends to the immunological cross-reactivity between the kallikreins; polyclonal antisera raised against one member of the family may be expected to cross-react with a significant number of epitopes shared by other members (Bothwell et al., 1979; Anundi et al., 1982). Further, classification of glandular kallikreins on the basis of enzymatic activity toward amino

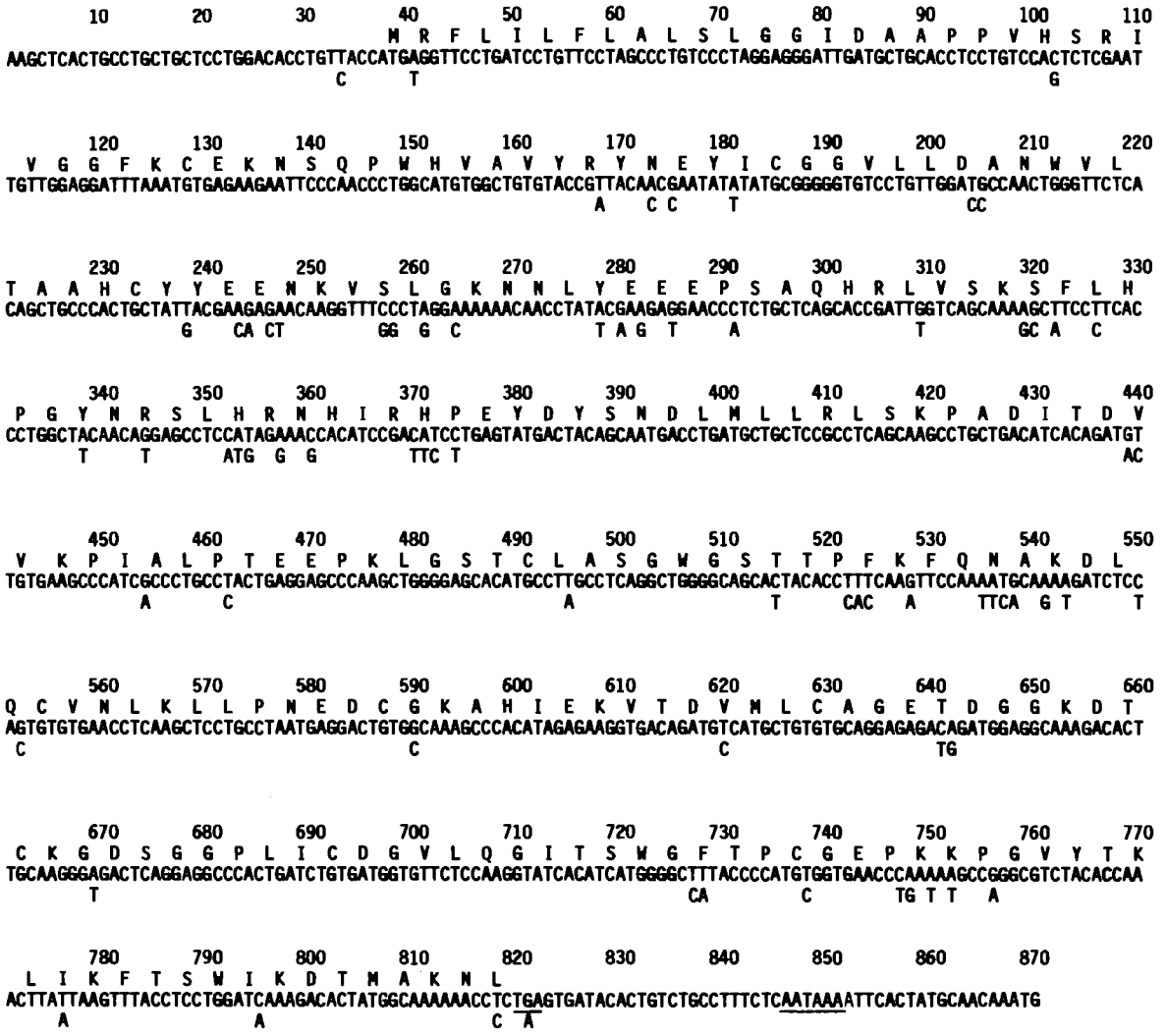

FIGURE 3: Nucleotide sequence of EGF-BP (middle lines), its translation (upper lines), and alignment with γ-NGF cDNA (lower lines) (Ullrich et al., 1984). The comparison with γ-NGF does not cover the 3′ nontranslated region; only those nucleotides of γ-NGF which differ from those of EGF-BP are reported. The start and stop codons and the poly(A) signal sequence are underlined. The one-letter amino acid code is the following: 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; Y, Tyr.

Table I: Amino Acid Compositions of Several Mouse Glandular Kallikreins[a]

	EGF-BP[b]	EGF-BP[c]	EGF-BP "B"[d]	γ-NGF[e]	β-NGF endopeptidase[f]	protease A[g]	mGK-1[h]	mRK[i]	γ-renin[j]
lysine	20	21	17	20	18	16	22	15	18
histidine	6	8	4	7	5	5	5	6	7
arginine	4	6	4	5	3	3	5	6	4
aspartate	28	26	25	29	30	27	28	35	28
threonine	14	14	15	17	13	14	10	13	17
serine	14	14	14	12	16	16	13	10	14
glutamate	21	19	20	14	17	19	18	21	19
proline	15	14	18	16	19	19	15	18	16
glycine	22	21	22	20	20	23	20	19	20
alanine	13	12	13	12	12	14	12	13	13
cysteine	11	10	10	10	12		11	10	7
valine	12	14	16	12	13	11	17	14	13
methionine	4	3	7	7	4	6	3	5	6
isoleucine	9	10	10	10	12	9	10	11	9
leucine	26	24	23	23	28	27	25	23	24
tyrosine	8	10	6	9	7	7	13	7	9
phenylalanine	6	6	7	8	5	8	4	5	9
tryptophan	5	5	6	6	4		6	6	4
total	238	237	237	237	238	224	237	237	237

[a] Composition values derived from actual proteins were taken from the original data and adjusted for a total of 237 residues, the consensus for mature glandular kallikreins, and rounded to the nearest integer. [b] Taylor et al. (1974). [c] This report. [d] Lundgren et al. (1983). [e] Ullrich et al. (1984). [f] Wilson et al. (1979). [g] Boesman et al. (1976). [h] Mason et al. (1983). [i] van Leeuwen et al. (1986). [j] Poe et al. (1983).

acid ester substrates is complicated by the extent of proteolytic processing of the enzyme which may alter its kinetic parameters toward such substrates (Nichols & Shooter, 1983). Additionally, various members of the kallikrein family share similar internal proteolytic cleavage sites, although the susceptibility of these sites to proteolysis (via autolysis or the action of other kallikreins) may vary for individual kallikreins (Isackson et al., 1987; Poe et al., 1983; Thomas et al., 1981; Anundi et al., 1982; Server & Shooter, 1976). Primarily for these reasons, the literature contains reports of a variety of differently named kallikreins with similar, if not identical, functional characteristics (Boesman et al., 1976; Schenkein et al., 1981; Hosoi et al., 1983; Wilson & Shooter, 1979; Bothwell et al., 1979).

Epidermal growth factor (EGF) was first reported to be present in crude homogenates of the submandibular gland of the male mouse as part of a high molecular weight complex of approximately 74000 by Taylor et al. (1970). This complex was shown to consist of a dimeric association of the 6400-Da EGF protein and an arginine-specific esterase of approximately 29000 Da. The complex could be reversibly dissociated in vitro (Taylor et al., 1970); the esterase was termed epidermal growth factor binding protein (EGF-BP). Homogeneous preparations of EGF-BP were isolated by both Taylor et al. (1974) and Server et al. (1976), and nearly identical amino acid compositions were reported by both groups. The substrate specificity of EGF-BP was investigated by Taylor et al. (1974) and found to be specific for esters formed by arginine, with the esters of lysine and citrulline having approximately 10% and 1% the catalytic rate of those of arginine, respectively. The peptide chain composition of EGF-BP was investigated by Server and Shooter (1976) and determined to consist of three distinct peptides with molecular weights of approximately 6800, 9600, and 10700.

Utilizing the purification procedure of Taylor et al. (1974), Anundi et al. (1982) reported an amino acid composition (Table I) and a partial amino acid sequence of EGF-BP. Two esterase components (EGF-BP types A and B) were identified by this group. Amino acid sequence analysis gave mixed sequences for each kallikrein at a number of cycles, indicating further heterogeneity. Neither of the isolated kallikreins was reported to reassociate with EGF in vitro. Type A EGF-BP was characterized as a 27000-Da single-chain molecule, whereas type B was present in various stages of proteolytic modification. The individual peptides of EGF-BP type B were aligned, on the basis of homology with the amino acid sequence of the γ-subunit of NGF (Thomas et al., 1981), to "reconstruct" an intact molecule (Anundi et al., 1982) (Figure 4).

EGF-BP type B, as proposed by Anundi et al. (1982), was described as being similar to the partial amino acid sequences of two peptides isolated by Silverman (1977) for a putative preparation of EGF-BP (Figure 4). This material contained a homogeneous kallikrein that yielded three peptides upon reduction and carboxymethylation [in agreement with Server and Shooter (1976)]. However, the amino-terminal fragment was not recovered after ion-exchange chromatography, and amino acid sequence information was obtained only for the central (residues 88–140) and carboxy-terminal (residues 141–237) peptides. A comparison of the sequences obtained in these two studies reveals that the central peptide of EGF-BP type B gave a single sequence (Anundi et al., 1982) that was identical with the partial amino acid sequence for the central peptide of EGF-BP as described by Silverman (1977) but that the carboxy-terminal peptide of EGF-BP type B gave mixed sequences at 6 out of 37 positions determined. It did contain the appropriate residues at these ambiguous positions to produce the corresponding carboxy-terminal sequence described by Silverman (1977) (Figure 4). This suggests that the preparation of EGF-BP type B of Anundi et al. (1982) might represent a mixture of two kallikreins, one of which is the same as that initially reported by Silverman (1977) and ultimately defined in this report.

Further support for this conclusion is provided by the cloning experiments of this group (Ronne et al., 1983; Lundgren et al., 1984). Partial and full-length cDNA clones have been reported for the type B EGF-BP, and the deduced amino acid sequence of these clones is identical with the amino-terminal peptide sequence of type B EGF-BP as described by Anundi et al. (1982). However, major discrepancies exist between these clones, and the partial amino acid sequence of the central peptide region reported by Anundi et al. (1982) and the carboxy-terminal peptide of EGF-BP type B (Anundi et al., 1982) can be matched up with the amino acid sequence predicted from the cDNA of Lundgren et al. (1984) only if the *alternative* amino acids at the mixed positions are used, i.e., the ones that did not align with the partial amino acid sequence of Silverman (1977) (Figure 4).

The identification of these unique sequences in the EGF-BP preparation of Anundi et al. (1982) is entirely consistent with the independent genomic analyses of Drinkwater et al. (1987). There studies identify EGF-BP type A as corresponding to genomic clone mGK-22, EGF-BP type B (as defined by the full-length cDNA) (Lundgren et al., 1984) as mGK-13, and the EGF-BP defined in this report as mGK-9 (Figure 4). Thus, the EGF-BP type B defined by protein sequencing (Anundi et al., 1982) is a mixture of two gene products: mGK-9 and mGK-13.

The limited proteolysis that results in the introduction of selective breaks in the polypeptide chain provides an additional means of identification. The gene product of mGK-13, corresponding to the cDNA described by Ronne et al. (1983) and Lundgren et al. (1984) as EGF-BP type B, appears to be present in the male mouse submandibular gland as a two-chain form, containing a disulfide-linked 17000-Da amino-terminal peptide and 10000-Da carboxy-terminal fragment (Figure 5). This is consistent with the deduced amino acid sequence for the cDNA clone of EGF-BP type B as reported by Lundgren et al. (1984); this kallikrein lacks basic residues, typical sites for proteolytic processing (Kreil, 1981), at positions 83 or 87 (Figure 2). Thus, the 6000-Da central peptide would be unlikely to be generated by proteolytic processing, and, consequently, no peptide sequence information for this region would be expected for this particular kallikrein, and none was found.

On the other hand, the mGK-9 gene product (as described herein) is present in the male mouse submandibular gland as a three-chain form containing a 10000-Da amino-terminal peptide, a 6000-Da central peptide, and a 10000-Da carboxy-terminal peptide (Silverman, 1977; Isackson et al., 1987). The preparation of Anundi et al. (1982) is consistent with this structure, assuming the amino-terminal peptide was lost in chromatographic isolation, as occurred in the study of Silverman (1977). The amino acid sequence of EGF-BP type A, the third kallikrein present, as deduced from the genomic sequence of mGK-22 (Drinkwater et al., 1987) contains no basic residues at the typical sites for proteolytic processing (residues 83, 87, and 140) (Figure 2). Therefore, this kallikrein would be expected to be present in the male mouse submandibular gland as a single-chain molecule, in agreement with the analysis of Anundi et al. (1982). It would contribute no

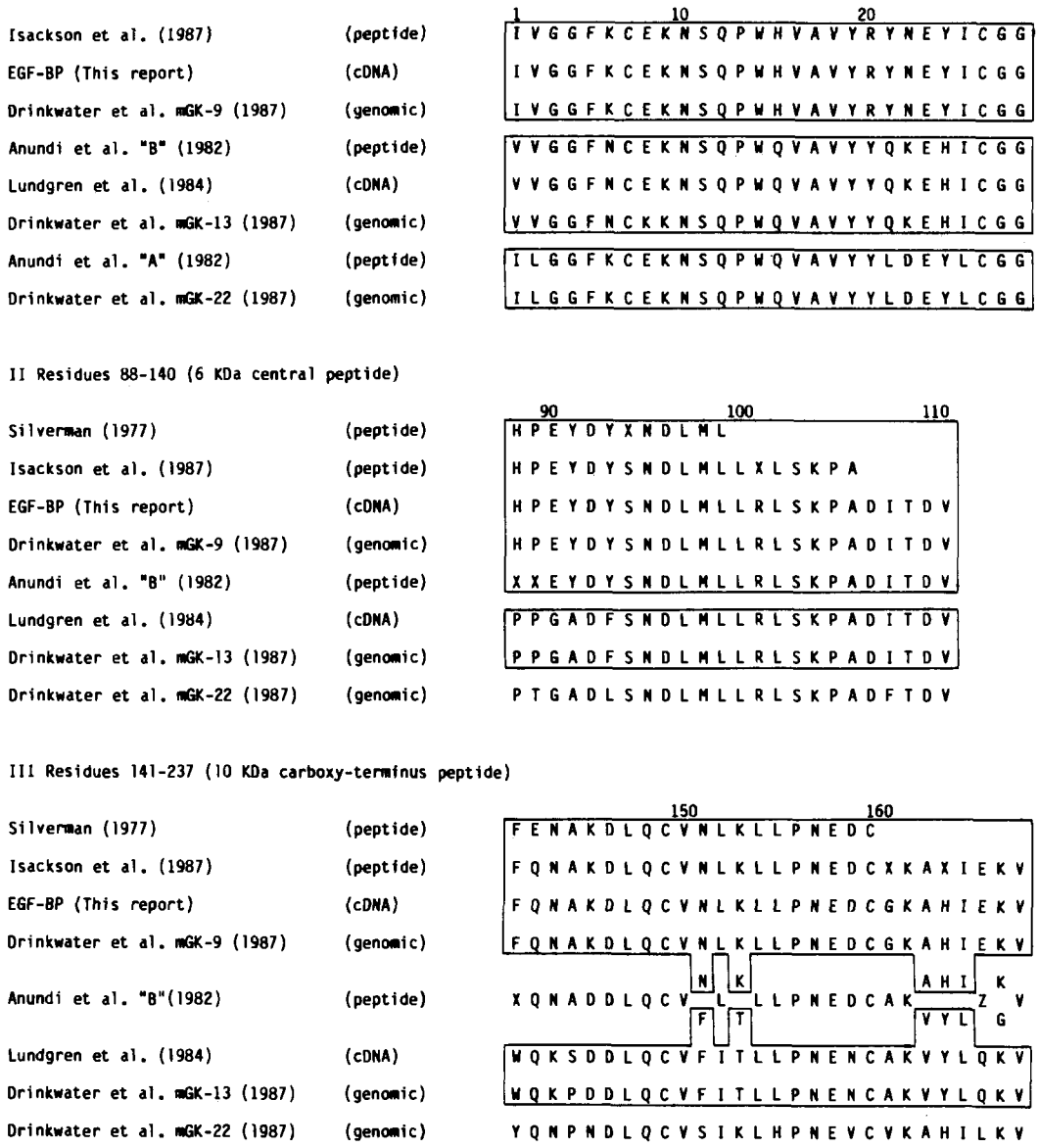

FIGURE 4: Relationship of peptide fragments, comprising the various reported EGF binding proteins, with the corresponding amino acid sequences derived from cDNA and genomic sequences. The isolated peptides of EGF-BP as described by Silverman (1977) and Isackson et al. (1987) correspond in all cases with the amino acid sequence derived from the mGK-9 gene. The alignment by Anundi et al. (1982) of the various peptides forming EGF-BP type B appears to be a chimeric construction of various peptides derived from the products of the mGK-13 and mGK-9 genes: The amino-terminal 10000-Da peptide (residues 1–87) is derived from the mGK-13 gene, the 6000-Da central peptide (residues 88–140) is derived from the mGK-9 gene, and the carboxy-terminal 10000-Da peptide appears to be a mixed sequence containing peptides derived from both the mGK-9 and mGK-13 genes (see text for explanation). The cDNA clone described by Lundgren et al. (1984) for EGF-BP type B corresponds to mGK-13.

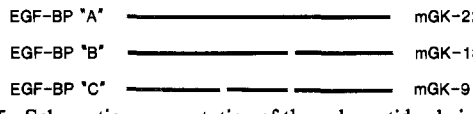

FIGURE 5: Schematic representation of the polypeptide chain structure of the products of mGK-9, mGK-13, and mGK-22 as they occur in adult male mouse submaxillary gland isolates. The designations EGF-BP "A", "B", and "C" are defined by Drinkwater et al. (1987).

sequence information for either the central 6000-Da peptide or the carboxy-terminal 10000-Da peptide. Since the mixed residues of the carboxy-terminal 10000-Da peptide were reported (Anundi et al., 1982) to be in approximately equimolar amounts and the amino-terminal peptide and the 6000-Da central peptide were present in quantitatively similar amounts, the EGF-BP preparation as described by Anundi et al. (1982), in addition to containing mGK-22 (type A), appears to have approximately equivalent amounts of mGK-13 and mGK-9, which together comprised the composite type B structure proposed.

The sum of these findings clearly suggests that mGK-9, the cDNA sequence of which is describe herein, which is present in the sample of Anundi et al. (1982), as well as those of Silverman (1977) and Isackson et al. (1987), is the principal EGF-BP. They do not rule out that the products of mGK-13 and mGK-22 also bind to EGF (and therefore are legitimate EGF-BP's) although no evidence for entities with their chain structure was found by Server and Shooter (1976). In view of the fact that mGK-9 has been demonstrated to bind to EGF in vitro and the others have not, its identification as the major EGF-BP in mouse submandibular gland, as originally described by Taylor et al. (1970), seems justified.

When comparing the amino acid sequences of several glandular kallikreins, the heterogeneity is localized to certain regions (Figure 2). These variable regions, unique for each

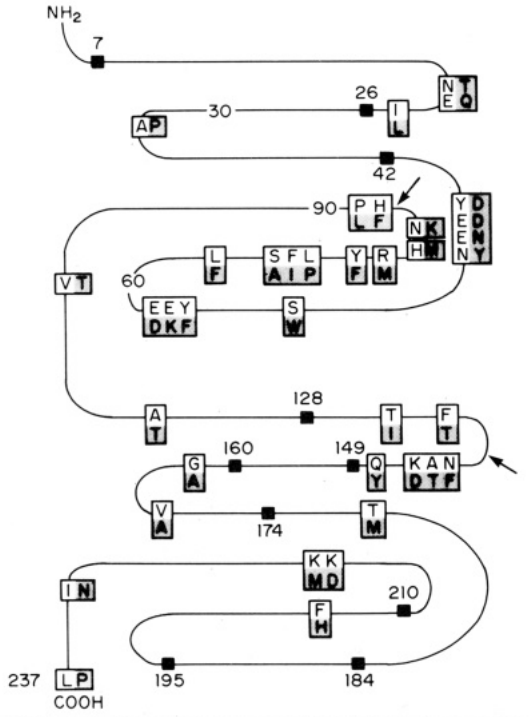

FIGURE 6: Diagrammatic representation of the primary structures of EGF-BP and γ-NGF showing the general organization of the surface loops of the molecules (Bode et al., 1983; Shotton & Watson, 1970). Only those amino acid positions which differ between EGF-BP and γ-NGF are indicated, and the residues of γ-NGF are shaded. Half-cystines are indicated by solid squares. Arrows point to known proteolytic processing sites of the molecules. The numbering system corresponds to that of Figure 2.

kallikrein, constitute surface loops of the three-dimensional structure (Bode et al., 1983; Shotton & Watson, 1970) (Figure 6). The invariant amino acid sequences comprise structurally conserved regions among the kallikreins which are responsible for the commonly shared structural framework (Greer, 1981). Included in the conserved regions of the kallikreins in Figure 2 are His-41, Asp-96, and Ser-189 which constitute the catalytic center of the serine-type proteases (Young et al., 1978). Also conserved is Asp-183 which confers a general substrate specificity for basic residues via electrostatic interactions with either arginine or lysine in the substrate P_1 position[3] (Hartley, 1970). The specificity for larger polypeptide substrates is determined by the structure of the substrate binding pocket which was composed of several of the highly variable surface loops of the molecule as well as structurally conserved regions (Tschesche et al., 1979; Bode et al., 1983).

The similarity between these kallikreins is an excellent opportunity to study the mechanism of the enzyme binding pocket specificity. The amino acid sequences of the EGF-BP presented here and the closely related γ-NGF (Thomas et al., 1981; Ullrich et al., 1984) are diagrammed with respect to the organization of their surface loops in Figure 6. The two proteins are 85% identical (92% at the nucleotide level), yet they will not substitute for one another in complexing with (and therefore perhaps processing of) their respective growth factors (Server & Shooter, 1977). Since both EGF-BP and γ-NGF are able to reassociate with the mature forms of EGF and β-NGF, respectively, the specific interactions of these complexes within the region of the enzyme binding pocket involves only the P_4, ..., P_1 "substrate" residues since residues corresponding to P_{-1}, P_{-2}, etc. are absent. Not surprisingly, EGF and β-NGF have different amino acid sequences at their carboxyl termini (EGF being -Trp-Glu-Leu-Arg and β-NGF, -Lys-Ala-Thr-Arg). In this regard, the interaction of EGF within the binding pocket of EGF-BP may involve electrostatic interactions between the P_3 glutamic acid of EGF and a basic residue of EGF-BP. However, in β-NGF, the P_3 position is an alanine, indicating that the putative electrostatic interaction of the EGF complex might be replaced by hydrophobic or steric interactions in the β-NGF complex. A basic residue in the region of the binding pocket of EGF-BP, which is replaced by an uncharged, hydrophobic, or sterically large residue in γ-NGF, may thus contribute to the distinct specificities of EGF-BP and γ-NGF. In this regard, residue positions 79, 82, 88, or 215 (which in EGF-BP are Arg-, His-, His-, and Lys- and in γ-NGF are Met-, Met-, Phe-, and Met-, respectively) may affect the enzyme specificity in this manner if this type of replacement occurs at a single position. Preliminary model-building studies utilizing trypsin as a comparative molecular model (M. Blaber, R. Blevins, K. A. Thomas, P. J. Isackson, and R. A. Bradshaw, unpublished observations) suggest that Lys-215 of EGF-BP may interact electrostatically with the P_3 glutamate of EGF. In EGF-BP type A (mGK-22) and type B (mGK-13), position 215 is occupied by an alanine and valine, respectively. Thus, if a lysine is important at this position for the specific binding of EGF, then the A and B forms would be expected to have considerably reduced binding affinity for EGF than the EGF-BP presented here. A lysine at position 215 is conserved at the homologous positions for trypsin and mGK-1, which interestingly is situated 5′ to the EGF-BP gene (mGK-9) on mouse chromosome 7 (Drinkwater et al., 1987). Trypsin appears unable to process the EGF precursor (Frey et al., 1979), and the mGK-1 gene product, although expressed in the adult male submandibular gland (Mason et al., 1983), has not been identified as an EGF binding protein. Thus, a lysine at position 215 may be necessary but not sufficient to confer EGF binding. In this regard, interactions affecting specificity that are dependent on more extensive protein–protein interactions may occur at some distance from the substrate binding pocket, involving other surface loops. Further molecular modeling and expression of EGF-BP and γ-NGF, along with hybrid variants, will allow a careful examination of the substrate binding sites and elucidate the molecular basis of the observed specificites.

[3] The standard substrate nomenclature (P_4, ..., P_1; P_{-1}, ..., P_{-4}) is used, with peptide cleavage occurring between P_1 and P_{-1} (Schechter & Berger, 1967).

ACKNOWLEDGMENTS

We gratefully acknowledge the exchange of sequence information with Dr. Rob Richards and his colleagues of the Howard Florey Institute, Melbourne, Australia, prior to publication. The oligonucleotides used in this study were prepared by the Protein/Nucleic Acid Analysis Laboratory under the direction of Steve S. Disper. Marja Uskali provided expert assistance in the preparation of the manuscript.

REFERENCES

Angeletti, R. H., & Bradshaw, R. A. (1971) *Proc. Natl. Acad. Sci. U.S.A. 68*, 2417–2420.

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

Aviv, H., & Leder, P. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 1408–1412.

Berger, E. A., & Shooter, E. M. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 3647–3651.

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

Boesman, M., Levy, M., & Schenkein, I. (1976) *Arch. Biochem. Biophys. 175*, 463–476.

Bothwell, M. A., Wilson, W. H., & Shooter, E. M. (1979) *J. Biol. Chem. 254*, 7287–7294.

Caruthers, M. H., Beaucage, S. L., Becker, C., Efcavitch, H. W., Fisher, E. F., Galluppi, G., Goldman, R., de Haseth, P., Martin, F., Matteucci, M. D., & Stabinsky, Y. (1982) in *Genetic Engineering* (Setlow, J. K., & Hollaender, A., Eds.) pp 1–17, Plenum Press, New York.

Chirgwin, J. M., Przbygla, A. E., MacDonald, R. J., & Rutter, W. J. (1979) *Biochemistry 18*, 5294–5299.

Cohen, S. (1960) *Proc. Natl. Acad. Sci. U.S.A. 46*, 302–311.

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

Drinkwater, C. C., Evans, B. A., & Richards, R. I. (1987) *Biochemistry* (following paper in this issue).

Evans, B. A., Drinkwater, C. C., & Richards, R. I. (1987) *J. Biol. Chem. 262*, 8027–8034.

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.

Greer, J. (1981) *J. Mol. Biol. 153*, 1027–1042.

Gubler, U., & Hoffman, B. J. (1983) *Gene 25*, 263–269.

Hartley, B. S. (1970) *Philos. Trans. R. Soc. London, B 257*, 77–87.

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

Isackson, P. J., Silverman, R. E., Blaber, M., Server, A. C., Nichols, R. A., Shooter, E. M., & Bradshaw, R. A. (1987) *Biochemistry 26*, 2082–2085.

Kreil, G. (1981) *Annu. Rev. Biochem. 50*, 317–348.

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

Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) in *Molecular Cloning: A Laboratory Manual*, pp 72, 122, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

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

Messing, J., Groenenborn, B., Muller-Hill, B., & Hofschneider, P. H. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 3642–3646.

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

Orstavik, T. B. (1980) *J. Histochem. Cytochem. 28*, 881–889.

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

Richards, R. I., Catanzaro, D. F., Mason, A. J., Morris, B. J., Baxter, J. D., & Shine, J. (1982) *J. Biol. Chem. 257*, 2758–2761.

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

Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.

Schechter, I., & Berger, A. (1967) *Biochem. Biophys. Res. Commun. 27*, 156–162.

Schenkein, I., Franklein, 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-Pescadoe, 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., & Shooter, E. M. (1977) *Adv. Protein Chem. 31*, 339–409.

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

Shotton, D. M., & Watson, H. C. (1970) *Nature (London) 225*, 811–816.

Silverman, R. E. (1977) Ph.D. Thesis, Washington University, St. Louis, MO.

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. (1981) *J. Biol. Chem. 256*, 9156–9166.

Tschesche, H., Mair, G., Goedec, G., Fiedler, F., Ehret, W., Hirschauer, C., Lemon, M., Fritz, H., Schmidt-Kastner, G., & Kutzbach, K. (1979) *Adv. Exp. Med. Biol. 120A*, 245–260.

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

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

van Leeuwen, B. H., Evans, B. A., Tregear, G. W., & Richards, R. I. (1986) *J. Biol. Chem. 261*, 5529–5535.

Varon, S., Nomura, J., & Shooter, E. M. (1968) *Biochemistry 7*, 1296–1303.

Wilson, W. H., & Shooter, E. M. (1979) *J. Biol. Chem. 254*, 6002–6009.

Young, C. L., Barker, W. C., Tomaselli, C. M., & Dayhoff, M. D. (1978) in *Atlas of Protein Sequence and Structure* (Dayhoff, M. O., Ed.) Vol. 5, Suppl. 3, pp 73–93, National Biomedical Research Foundation, Silver Spring, MD.