Mouse Glandular Kallikrein Genes: Identification and Characterization of the Genes Encoding the Epidermal Growth Factor Binding Proteins†

Catherine C. Drinkwater,* Bronwyn A. Evans, and Robert I. Richards
Howard Florey Institute of Experimental Physiology and Medicine, University of Melbourne, Parkville, Victoria, 3052 Australia
Received March 30, 1987

ABSTRACT: Previously, three proteins have been separately identified as the mouse epidermal growth factor binding protein (EGF-BP). We have identified and sequenced the coding regions of three distinct genes encoding these EGF-BPs from the BALB/c strain. The genes are all members of the glandular kallikrein gene family, which encodes a highly homologous group of serine proteases. Expression of the EGF-BP genes was detected in mouse salivary gland only and was at a relatively similar level for each gene. The isolation of three distinct genes from the one mouse strain indicates that the conflicting data previously reported in the literature are not a result of allelic polymorphisms or strain differences.

Mouse epidermal growth factor (EGF)[1] is a 53 amino acid protein that stimulates the proliferation and differentiation of various mammalian cells (Carpenter & Cohen, 1979; Schlessinger et al., 1983). The major source of mouse EGF is the adult male salivary gland (Cohen, 1962), where it is almost entirely found in a high molecular weight complex of approximately 74 000 (Taylor et al., 1970). This complex consists of two molecules of mature EGF bound to two molecules of a 29 000-dalton arginyl esterase known as the EGF binding protein (EGF-BP). Prepro-EGF, a 130 000-dalton protein, is processed by a series of steps to a 9000-dalton EGF precursor (Burmeister et al., 1984), which is then cleaved at the carboxy terminus by EGF-BP to produce the mature growth factor (Frey et al., 1979). The interaction of EGF and its binding protein is highly specific. Although the γ-subunit of the 7S nerve growth factor (NGF) complex is structurally related to EGF-BP, both proteins being members of the mouse glandular kallikrein family (Evans et al., 1987), neither it nor trypsin has any effect on the 9000-dalton EGF precursor (Frey et al., 1979). Likewise, EGF-BP is unable to substitute for γ-NGF in the analogous β-NGF processing system (Server & Shooter, 1976).

It has been reported that purification of EGF-BP from salivary glands of adult male NMRI mice (an outbred strain) yields two distinct polypeptides, designated EGF-BP type A and EGF-BP type B (Anundi et al., 1982; Ronne et al., 1983). It was not determined, however, whether these two proteins represent alleles or are encoded by separate genes. More recently, another protease, whose sequence differs from both EGF-BP type A and EGF-BP type B, was reported as the only EGF-BP purified from the high molecular weight complex in male Swiss ICR mouse salivary glands (Isackson et al., 1987). Furthermore, this protein was shown to be able to associate with EGF and reconstitute the high molecular weight complex, a property not demonstrated for either EGF-BP type A or EGF-BP type B. For the purposes of this paper, the EGF-BP isolated by Isackson et al. (1987) shall be referred to as EGF-BP type C.

In view of these anomalies in the identity of the EGF processing enzyme(s), we have characterized the genes encoding the respective proteins. We have previously reported a systematic analysis of the mouse glandular kallikrein gene family, in which 24 nonallelic genes were isolated, mapped, and partially sequenced, thus defining the limits of functional diversity (Evans et al., 1987). In this paper, we present the further characterization of those members of the kallikrein gene family that encode EGF-BP type A, type B, and type C. The complete predicted amino acid sequences from all three genes were compared in an effort to identify residues which specify binding to the 9000-dalton EGF precursor. The identification of three distinct genes in the one mouse strain (BALB/c) indicates that the different proteins are not the result of allelic polymorphism.

Experimental Procedures

Preparation and Isolation of Genomic Clones. Kallikrein clones were isolated from a mouse genomic library constructed by using the EMBL3A vector as described previously (Evans et al., 1987).

Nucleotide Sequence Analysis. Restriction fragments of the λ clones were subcloned into M13 mp 18 or 19 (Messing, 1983) and sequenced by the chain termination method (Sanger et al., 1977) using [α-^{32}P]dATP and either the standard M13 17-mer or specific oligodeoxyribonucleotides (Evans et al., 1987) as primers.

Oligodeoxyribonucleotides. Oligodeoxyribonucleotides corresponding to variable regions of the DNA sequence from coding regions of the genes mGK-9, -13, and -22, to residues 3315–3344 of the mouse prepro-EGF gene (Scott et al., 1983; Gray et al., 1983), and to regions conserved between all the kallikrein genes (Evans et al., 1987) were synthesized by the solid-phase phosphoramidite procedure as previously described (van Leeuwen et al., 1986).

Specificity of the oligodeoxyribonucleotides for mGK-9, -13, and -22 was tested by hybridization of the probes to nitrocellulose filters prepared in the following manner. Two microliters each of λ clones (approximately 10^8 PFU/mL) containing either one or two kallikrein genes and representing

† This work was supported by grants to the Howard Florey Institute from the National Health and Medical Research Council of Australia, the Ian Potter Foundation, and the Myer Family Trusts. C.C.D. is the recipient of a Commonwealth Postgraduate Research Award.
* Address correspondence to this author.

[1] Abbreviations: EGF, epidermal growth factor; EGF-BP, epidermal growth factor binding protein; NGF, nerve growth factor; γ-NGF, γ-subunit of the 7S NGF complex; PFU, plaque-forming unit(s); kDa, kilodalton(s).

0006-2960/87/0426-6750$01.50/0

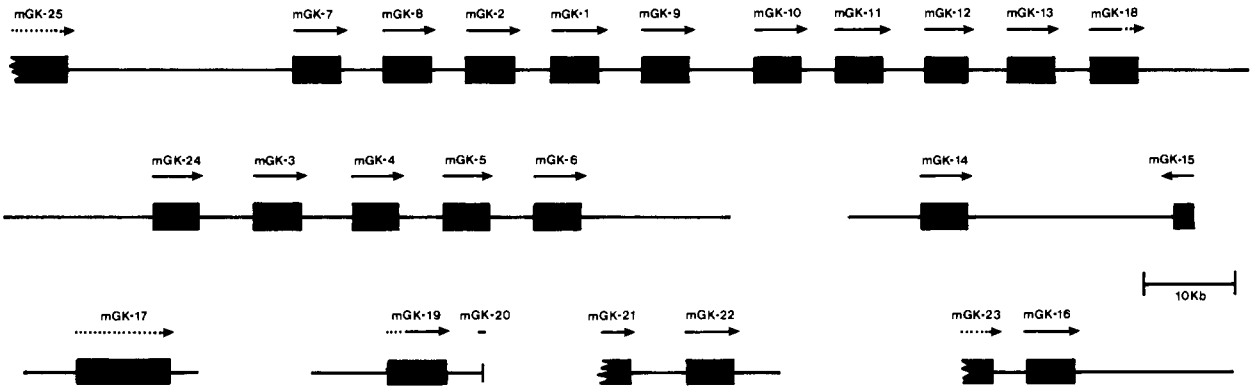

FIGURE 1: Linkage map of the mouse glandular kallikrein gene family. Genes (five exons and four introns) are denoted by solid bars, and the arrows above each gene show the direction of transcription. The relative positions of each cluster of genes within the kallikrein locus have not been determined [adapted from Evans et al. (1987)].

the complete gene family was spotted onto a lawn of LE392 cells and incubated at 37 °C overnight. The phage were then transferred to 0.45-μm nitrocellulose filters and lysed, and the DNA was denatured by the method of Maniatis et al. (1982). The filters were washed in 2 × SSC, air-dried, and baked in vacuo for 2 h at 80 °C. After hybridization to the oligodeoxyribonucleotide probe, the filters were washed in 2 × SSC at successively increasing temperatures to determine the uniqueness of and appropriate stringency for each probe.

mRNA Preparation and Northern Blots. Polyadenylated RNA was isolated from mouse salivary gland, kidney, pancreas, brain, and testis as detailed elsewhere (van Leeuwen et al., 1986). Ten-microgram aliquots were then denatured in formaldehyde, electrophoresed on 1% agarose–formaldehyde gels, and transferred to nitrocellulose as described by Maniatis et al. (1982).

Hybridization Analysis. Oligodeoxyribonucleotide probes were labeled at the 5′ terminus using T4 polynucleotide kinase and [γ-^{32}P]ATP by a modification (van Leeuwen et al., 1986) of the method of Maniatis et al. (1982). Nitrocellulose filters were prehybridized for 2–4 h at 42 °C in a solution of 5 × SSC, 50 mM sodium phosphate, pH 6.8, 1 mM sodium pyrophosphate, 50 μg/mL sonicated herring sperm DNA, 0.02% bovine serum albumin, 0.02% Ficoll, 0.02% poly(vinylpyrrolidone), 100 μM ATP, and 20% formamide. Hybridization to the oligodeoxyribonucleotide probes was performed at room temperature overnight, and the filters were washed in 2 × SSC at the appropriate temperature as determined by the specificity test described earlier. Filters were then air-dried and exposed to Kodak X-AR film at −80 °C with intensifying screens.

RESULTS

Identification of the Genes Encoding EGF-BP Types A, B, and C. A simplified map of the mouse glandular kallikrein gene family is shown in Figure 1. Partial sequence from the coding region of all members of the gene family has been obtained (Evans et al., 1987). By comparing these coding regions with partial amino acid sequences (Anundi et al., 1982; Isackson et al., 1987), we were able to identify the genes encoding the EGF-BPs. The complete coding sequence of each of these genes was subsequently determined (Figure 2).

In the case of EGF-BP type A, there is only one amino acid residue difference between the N-terminal protein sequence (27-kDa peptide; Anundi et al., 1982) and the predicted amino acid sequence of the gene mGK-22 (Figure 3A). No other kallikrein gene has less than five differences from the 27-kDa peptide sequence (Evans et al., 1987), and we suggest that the observed difference is possibly due to polymorphism between the different strains of mice used (outbred NMRI and inbred BALB/c). An oligodeoxyribonucleotide complementary to residues 1148–1174 (Figure 2a), corresponding to amino acids 19–27 of the protein sequence (Anundi et al., 1982), was subsequently synthesized and used to screen the complete kallikrein gene family. Only mGK-22 hybridized to this probe at the relatively low washing stringency of 37 °C (Figure 4), further supporting its designation as the gene encoding EGF-BP type A. The nucleotide sequence of the coding regions of the gene was determined, as well as 541 base pairs of 5′ flanking DNA (Figure 2a).

An oligodeoxyribonucleotide complementary to nucleotide residues 334–363 of an EGF-BP type B cDNA clone (Lundgren et al., 1984) was synthesized and used to screen an array of λ clones representing the complete mouse glandular kallikrein gene family. When washed in 2 × SSC at 60 °C, only one gene, mGK-13, remained bound to the probe (Figure 4). The five exons of this gene were subsequently sequenced, as well as 362 base pairs of 5′ flanking DNA (Figure 2b). Comparison of the predicted amino acid sequence of mGK-13 with the N-terminal sequence of EGF-BP type B (17-kDa peptide; Anundi et al., 1982) showed only one amino acid difference (Figure 3). All other kallikrein genes have three or more differences from the 17-kDa peptide sequence (Evans et al., 1987). Although mGK-13 differs at 11 nucleotide residues from the EGF-BP type B cDNA clone determined by Lundgren et al. (1984), homology between the gene and the cDNA is 99%, which is significantly greater than that between any two members of the kallikrein gene family (80–90%). We again assume that these changes are a result of polymorphic strain differences.

The gene mGK-9 was identified as encoding EGF-BP type C by comparison of the protein sequence (Isackson et al., 1987) with the predicted amino acid sequences of the kallikrein genes. The nucleotide sequence of the coding regions and 505 base pairs of 5′ flanking DNA was determined (Figure 2c). The sequence of mGK-9 is virtually identical with a cDNA clone for EGF-BP type C (Blaber et al., 1987) with four nucleotide changes, two in the 5′ untranslated region and two in the 3′ untranslated region. An oligodeoxyribonucleotide corresponding to nucleotides 2277–2306 was synthesized and proved specific to the gene when washed in 2 × SSC at 25 °C (Figure 4).

Structure of the Genes. The exon–intron boundaries of mGK-9, -13, and -22 are identical with those of all other members of the kallikrein gene family (Evans et al., 1987). All three genes have the variant 5′-TTTAAA-3′ box typical to kallikrein promoter regions (Mason et al., 1983; Evans & Richards, 1985; van Leeuwen et al., 1986) and polyadenylation

```
a
GAATTCAGCTCCTAATAGGAAAAGTAAAATCAGTGATCCAACTTCCAAGTCCAAAATGAGGCATTTTGGCATTCTCCAAGGAAGGATTTGTATCTACAGGGAGCCTTGCAGCCCTATAGG  120

CTACCACTTGTGCAAAGACTCTGTTCTAACTGCAACTTTCTGGGATAATCTTGGACTTCAGGTAGAGGACCTTCCTTAGAGGTGTGGGTTAGGCAGCACTCGCTCCACTACCCTCGGTTA  240

ACACCATTTCACCATTCCCAGGACCTTGACATCTGGCTAAGGTCCCTGGAACACTAACAGCTCCAGAAACAAAAGTCCATGCCCCACCCGATTTTATCCTTGGGAAGGTATCCAGGGTGA  360

CCTAGGGCCTACTGGACATCAGGTGCAACAGGACCGCCCCTTACCACAGCAGGGAGGGTAAGTGCAGGGAATCCACCTCTGTCAGCAAGGGAAGAGCTTTGGTCAGCATTTGGCTGCCAG  480

                                                                     ↓         ↓                                 MetArgPheLeuIleLe
GAGAGGGGTGGGGCTGTGGGGGAGAATGAGGGTTTTTAAAGTCTCCCTCAGGAGCCTCCACAGCTCCAAGCTGACCTCCTGCAGCTCCTGGACAGCTGTTACCATGAGGTTCCTGATCCT  600

uPheLeuThrLeuSerLeuGlyGlyIleA
GTTCCTAACCCTGTCCCTAGGAGGGATTGGTGAGATAGGGGAAGGGGAGGGGTCACAGCCCTGACTCTCCTGATGCTATTGGCCCTGGCCATACCTTTTCCCTCTCGCCTGAAGCCCAGT  720

AGATCATGAGCACCACCTGCTCCTCCTCCTCCTGTGCCTTATGCAGCGATGGGACCTTACAGTGCTTGTCCAGGG---------1.85kb--------AAGCTTCTTTGCCATCCCTTCA  840

CCTAAAATTCTAAAGTGCTCTGACACTATACCCGATTCGTCATTCAAGAGTTCTTCCTGTACCCACAGCATACATCTCCCTATGCAGGACCAGCCATATATATTGGTCCCTACTCCAGAT  960

                                                                                                                 *
                                                                                                           spAlaAlaProProVa
CCCCTGGCCATGTCCCCAGATCCCAGCCCTGCTTAGCTGTGTCCCCCAATGGTACACCTCCTTACATGTGCCAAATCTCCCCTTCCCCATTGCGCCTGTTGCAGATGCTGCACCTCCTGT 1080

          1                                                   20
lGlnSerArgIleLeuGlyGlyPheLysCysGluLysAsnSerGlnProTrpGlnValAlaValTyrTyrLeuAspGluTyrLeuCysGlyGlyValLeuLeuAspArgAsnTrpValLe
CCAGTGTCGAATACTTGGAGGATTTAAATGTGAGAAGAATTCCCAACCCTGGCAGGTGGCTGTGTACTACTTAGATGAGTACCTATGCGGGGGAGTCCTGTTGGACCGCAACTGGGTTCT 1200

    40
uThrAlaAlaHisCysTyrGluAs
CACAGCTGCGCACTGCTATGAAGAGTAAGTAAAGGTGGAGACAGGAAAGGAAGATGGCAGCCAGAGATCATGACTCCAGGACAGGCTGGGGTACTGAAGGGCAGGGGAGTGGGACTGGCT 1320

GAAACTATCCATGACCTCCTGGGTCTCTCTTTACTCCATCCTCCAGGTTTGTGTTCAGAGTCCTGTTTCTTTTTTTTTTTCTCTGTTATTATCTGTCTCTCTCCTGTCTGTATTCACATA 1440

TGAGTG--------0.4kb--------GATGTGTGTGTGTGTGTGATTGTGTGTGTGTGTGATTGTGTGTGTGTGTGTATCTTTCTGTCTGTGTGTGTTTCTGTCTCTGACTCTCCCTTC 1560

                                                                                         pLysTyrAsnIleTrpLeuGlyLysAsnLy
CTCTTTGCCTCCTCCAGCTCTTCTGCTGATGCTGACCAATCTCCCTGTTTCTGTCCACTCACCCCCTCCATCTCCTTCCCTGCCCCACAGCAAGTATAATATTTGGCTGGCAAAAAGAA 1680

           60                                                80
sLeuPheGlnAspGluProSerAlaGlnHisArgLeuValSerLysSerPheProHisProAspPheAsnMetSerLeuLeuGlnSerValProThrGlyAlaAspLeuSerAsnAspLe
GCTATTCCAAGATGAACCCTCTGCTCAGCACCGATTGGTGAGCAAAAGCTTCCGTCATCCTGAGTTCAACATGAGCCTCCTCCAAAGTGTACCTAGTGGGGCGGACTTAAGCAATGACGT 1800

          100                                               120
uMetLeuLeuArgLeuSerLysProAlaAspIleThrAspValValLysProIleAspLeuProThrThrGluProLysLeuGlySerThrCysLeuAlaSerGlyTrpGlySerIleAs
GATGCTGCTGCGCCTCAGGAAGCCTGCTGACATCACAGATGTTGTGAAGCCCATCGACCTGCCCACTACGGAGCCCAAGCTGGGGAGCACATGCCTAGCCTCAGGCTGGGGCAGCATTAA 1920

                                                                                                                140
nGlnLeuIleT                                                                                                    yrGlnAsnProAs
CCAGTTAATATGTGAGTCAGGTCCAAGCACAACAGCTTGGTGTGGAGGAGGGGCAGACGGGACTAGATTGGGCCCTGCTTACCACCCACTTTCTTCCTGATCCAAAGACCAAAACCCAAA 2040

                                   160                                                         180
nAspLeuGlnCysValSerIleLysLeuHisProAsnGluValCysValLysAlaHisIleLeuLysValThrAspValMetLeuCysAlaGlyGluMetAsnGlyGlyLysAspThrCy
TGATCTCCAGTGTGTGTCCATCAAGCTCCATCCTAATGAGGTGTGTGTGAAAGCCCATATAGTGAAGGTGACAGATGTCATGCTGTGTGCAGGAGAGATGAATGGAGGCAAAGACACTTG 2160

sLys
TAAGGTGAGACAGCCCCTCCCTGCAGTGAGGTGAAGGGCTGAAAGAGGGAACTGGGTTTCTTGGTTCCCACTTCAACACCCTGACTAGGCAGGCTCTTTTTATGCTTCCCTGTGGAAGGC 2280

ATGTTCCCTGGGAGGGGAGGACCGTGTAGCTGGTACTATTTCTCCTCTGTGACATGTCTTAGGGGACTGTTTTGTGAGGTCCTCACCTCAACCCAATGTCCTCACAAAATCTGGATCCCA 2400

AGAGCTGGATTCCCTCACCTATTGAACAGAGCTGCCTCTTAAGAGTCCGTGTATTCCTCCCTGGGAGCCAGTTCCAGATTGTATACTAGAATCCAACTGGGAGGCCCTGGCTTACTGCTG 2520

                                                         200
                 GlyGlySerGlyGlyProLeuIleCysAspGlyValLeuGlnGlyIleThrSerTrpGlySerThrProCysGlyGluProAsnAlaProAlaIleTyrThr
TACCTGTCCTTCTATCAGGGAGACTCAGGAGGCCCACTGATCTGTGATGGTGTTCTACAAGGTATCACATCATGGGGCTCTACCCCATGTGGTGAACCCAATGCACCGGCCATCTACACC 2640

   220
LysLeuIleLysPheThrSerTrpIleLysAspThrMetAlaLysAsnPro***                                                      ▼
AAACTTATTAAGTTTACCTCCTGGATAAAAGACACTATGGCCAAAAACCCCTGAGTGTCACATTATCTGCTGTTCTCAATAAAATATACCATGCAACATATGAGTTCAAGGTCTGACATT 2760

CTCTGTGCTACTGGAGTGGGGTAGCAAATCAAGGATCC                                                                                    2798


b
ATAGGTCTAGGTAACTTCCTTAGTACTCGGGGTTGGAGTCACTTGTGCCTCCATCCTCCTTTAACACCAACTCACCATTGTCCCTGACTTTGCATGTGGCTAAGGTCCCCGGAACTCGAA  120

ACAGCATTGCCCAACAGCTCCAGGAACACCCAATATCATCCCTGGGAAGTATCCAGGGTGACTTAGGGCCTACTGTACAGAGGGTGCAATGAGGCGGTCCCATTCCACAGCCAGGAGGGT  240

GCAAGCAGGGAGTCCACCTCTGTCACCAATGGCAGGGCTTAGTCAGCCATCTGGGTGTCAACAGGACAGGGGAGGGACTATGGAGAGAATGAGGGTATTTAAAATCTCCTGAGAGCCTCA  360

↓    ↓                                          MetTrpPheLeuIleLeuPheLeuAlaLeuSerLeuGlyGlyIleA
ACAGCTCCAAGCTCACTGCCTGCAGTTCCAGGACACCCTGTTACCATGTGGTTCCTGATCCTGTTGCTAGCGCTGTCCCTAGGAGGGATTGGTGAGATGAGGAAGGGGAGGGGTCACGC  480

CTGACTCTCCTGCTGCTTTGGCCCTGGACCTCCCTTTCACCTCTCCCCTTACCCCAGCTACATCATGAGCAAGCCTGCTCCTCCTTCTCTGCCCTGCACACTGCCAGGTCCCTCACAGTC  600

CTTGCTCACAGGCCAGGCCCCTGTCCACTCCCACTTTCCAGAATTTTCTCCTCTTCCTTAGACACCTGATCATCTCTGTGCTTCTCCAGTCTTTCTAAGGTTCCC---------1.9kb-  720

--------AGCACTAAGGTTACTACCATACCTTAACCCAAAACTATAAAATGACCTTATACTATACCCAATTCCTGATTCAAGAGTTCTTTCTGTCAGCTACAGCATACATATCCATATC  840

CAGGACCAGCCATATATTGGTCTCTACCCCAGATCCCCTGGACATGTCCCCAGATCCCAGTCCTGCTTAGCTGTGTCCCGCCAATGGTAGACGTCCCTACATGTGCCAACTCTCCCCTCC  960

                          *                1                                                         20
                     spAlaAlaProProLeuGlnSerArgValValGlyGlyPheAsnCysLysLysAsnSerGlnProTrpGlnValAlaValTyrTyrGlnLysGluHisIleC
CCATTGCTCCTGTTGCAGATGCTGCACCTCCTCTCCAGTCTCGGGTGGTTGGACGATTTAACTGTAAGAAGAATTCCCAACCCTGGCAGGTCGCTGTGTACTACCAAAAGGAACACATTT 1080

                                     40
ysGlyGlyValLeuLeuAspArgAsnTrpValLeuThrAlaAlaHisCysTyrValAs
GTGGGGGTGTCCTGTTGGACCGCAACTGGGTTCTCACAGCTGCCCACTGCTATGTCGAGTGAGTAAGGGTGGAGAAAGAAAAGCAGATTGGCAGCCAAGGACATGACTCCAGGCCAGGCT 1200

GGGTACTGAAGGACAGTGGGGTGGGACTGGCTGAAATCTACCCATGCCTCCTGACTCTCTTTCCTCCTCCATCTTGGTTTGTTTCAGATCCTGTTTCTTTCTTTTCTGCCTTATTGTCTG 1320

TCTGTCTCTTGTCTGTATTCACACTTGCTGTCTCATGCCACTGTCATGTCTTGTTTACCAAGAAAACAAATAAGGCTAGAAAGTCTTCTATTATTAGAAAATATCAAATCTAATTATAGA 1440

AA-----------0.35kb----------TATCTGTGTATGTATGTGTGTGTGTCTGTGTATGTATGTGATTTTCTGTTTCTGCCTGTTTTATTATATCTGTTTCTTTGAGTCTGTTTGT 1560

GTGTGTATCTTTCTGTCTGTGTCTGTTTATGTATTTGATCTATCTGTTACTTCCTCCCTTCCTCTTGCCCACGGGTTCCTATTTATTCTAACCAATCTCCCCATTGCTCCTCTTTTTGGC 1680

                                                                      60
                           pGlnTyrGluValTrpLeuGlyLysAsnLysLeuPheGlnGluGluProSerAlaGlnHisArgLeuValSerLysSerPheProHisProGlyP
TCCATGTCCCTCCCTGTTGTCCCAGCCAGTATGAGGTTTGGCTGGGCAAAAACAAGTTATTCCAAGAGGAACCCTCTGCTCAGCACCGATTGGTCAGCAAAAGCTTCCCTCACCCTGGCT 1800

        80                                                 100
heAsnMetSerLeuLeuMetLeuGlnThrIleProProGlyAlaAspPheSerAsnAspLeuMetLeuLeuArgLeuSerLysProAlaAspIleThrAspValValLysProIleAlaL
TCAACATGAGCCTCCTGATGCTTCAAACAATACCTCCTGGGCCTGACTTCAGCAATGACCTGATGCTCCTCCGCCTCAGCAAGCCTGCTGACATCACAGATGTTGTGAAGCCCATCGCCC 1920
```

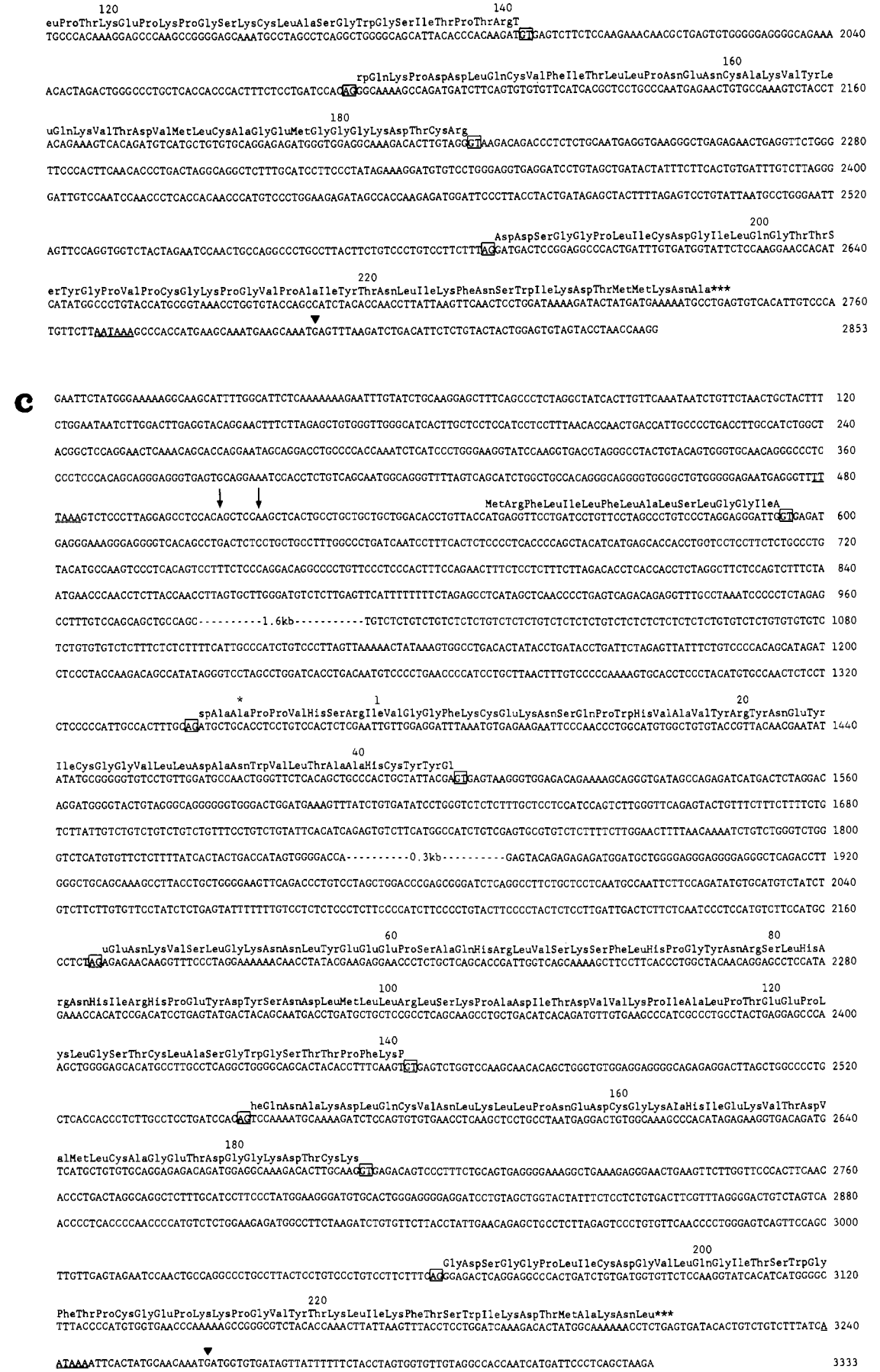

FIGURE 2: Sequence of the EGF-BP genes. The nucleotide and encoded amino acid sequences of the EGF-BP type A [(a) mGK-22], EGF-BP type B [(b) mGK-13], and EGF-BP type C [(c) mGK-9] genes are shown. Numbers to the right refer to nucleotide residues. The splice sites around each exon and the initiation and termination codons are boxed. The coding sequences refer to the position relative to the N-terminal isoleucine (a, c) or valine (b) of the mature proteins. Asterisks denote the first residue of the zymogen peptides, and the amino acids preceding this constitute the signal peptides. The variant TATA boxes (Corden et al., 1980) and the polyadenylation signals (Proudfoot & Brownlee, 1976) are underlined. Putative transcription initiation sites are shown by arrows, and the triangles indicate the transcription termination site (Mason et al., 1983).

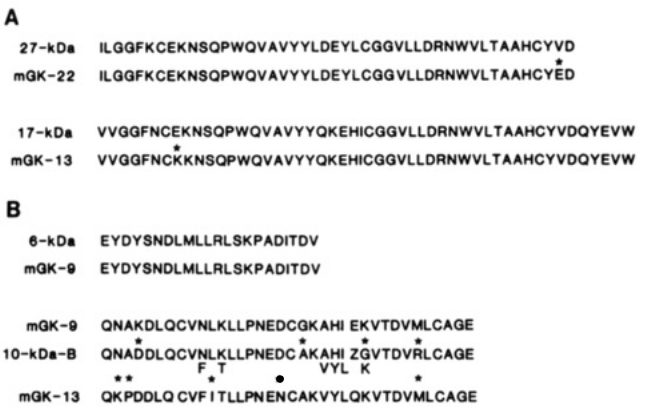

FIGURE 3: Comparison of EGF-BP amino acid sequences. (A) The deduced amino acid sequences of mGK-22 and mGK-13 are compared to the 27-kDa peptide fragment of EGF-BP type A and the 17-kDa peptide fragment of EGF-BP type B as determined by Anundi et al. (1982). (B) The deduced amino acid sequences of mGK-9 and -13 are compared to the 6- and 10-kDa B fragments of Anundi et al. (1982). Asterisks denote differences between residues.

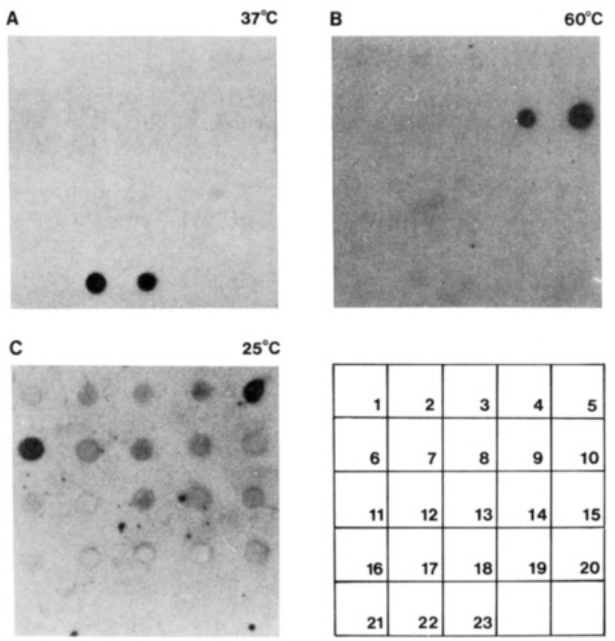

FIGURE 4: Specificity of the synthetic oligodeoxyribonucleotide probes. DNA from λ clones containing all the kallikrein genes was transferred to nitrocellulose filters. These were hybridized to (A) the mGK-22 27-mer, (B) the mGK-13 30-mer, and (C) the mGK-9 30-mer and washed in 2 × SSC at the temperatures shown. The λ clones contain the kallikrein genes: (1) mGK-7 and -8; (2) mGK-8 and −2; (3) mGK-8, −2, and −1; (4) mGK-2 and -1; (5) mGK-1 and -9; (6) mGK-9; (7) mGK-10; (8) mGK-11 and -12; (9) mGK-12 and -13; (10) mGK-13; (11) mGK-3 and -4; (12) mGK-4; (13) mGK-4 and -5; (14) mGK-5 and -6; (15) mGK-6; (16) mGK-14; (17) mGK-15; (18) mGK-16 and -23; (19) mGK-17; (20) mGK-18; (21) mGK-19; (22) mGK-21 and -22; (23) mGK-22. Clones containing mGK-24 and -25 were tested separately by Southern blotting and showed no hybridization to any of the probes.

signals (Proudfoot & Brownlee, 1976) in the 3′ untranslated region. Exon three of mGK-22 is 6 base pairs shorter than the other kallikrein genes, with the resultant deletion of two amino acid residues from EGF-BP type A occurring within the region referred to as the autolysis loop (Tschesche et al., 1979). Four amino acids are removed from this region of γ-NGF (Thomas et al., 1981). This cleavage step is not thought necessary for protease activity, since other members of this homologous family, such as mouse renal kallikrein, are active in a 27 000-dalton single-chain form (Bothwell et al., 1979). Thus, the structure of this region is not crucial to the function of the resultant serine protease, and the absence of two amino acids should not affect the activity of EGF-BP type A.

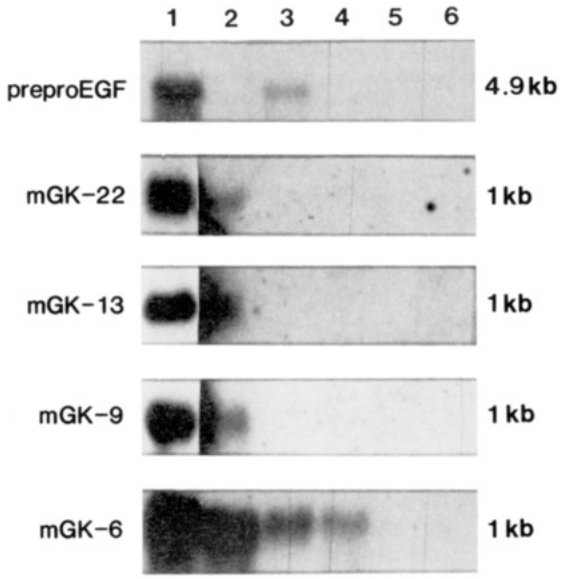

FIGURE 5: Expression of EGF-BP genes in mouse tissues. Northern blots were prepared from 10 μg of polyadenylated RNA from mouse male (lane 1) and female (lane 2) salivary glands, male kidney (lane 3), pancreas (lane 4), brain (lane 5), and testis (lane 6). Blots were hybridized to the pro-EGF 30-mer, mGK-22 27-mer, mGK-13 30-mer, mGK-9 30-mer, and mGK-6 30-mer (van Leeuwen et al., 1986). Sizes of the mRNA bands are shown on the right. Hybridization conditions for all blots were as detailed under Experimental Procedures, with washes in 2 × SSC at 60 °C. Tracks containing male salivary gland mRNA hybridized to mGK-9, -13, and -22 were exposed for 5 h, while the remaining five tracks in each case were exposed for 30 h in order to detect any low levels of hybridization.

Expression of the EGF-BP Genes in Mouse. For production of mature EGF, there must be coordinate expression of the prepro-EGF and EGF-BP genes. Though prepro-EGF mRNA synthesis has been described in a number of tissues (Rall et al., 1985), processed EGF has only been isolated from adult mouse salivary gland (Cohen, 1962), implying that EGF-BP is only expressed in this tissue. Consequently, we analyzed the expression profiles of the three EGF-BP genes in an attempt to resolve the discrepancies in the reported identity of the protein. If any of the genes were inappropriately expressed in tissues other than salivary gland, we could then infer that the resultant gene product is not the specific EGF-processing enzyme. Northern blots of mRNA from mouse salivary glands, kidney, pancreas, brain, and testis were screened with the gene-specific oligodeoxyribonucleotide probes. 30-mers specific for the prepro-EGF gene and the renal kallikrein gene, mGK-6 (van Leeuwen et al., 1986), were also used to probe the RNA blots (Figure 5). Expression of the EGF gene was detected in salivary gland and kidney, whereas all three EGF-BP genes were only expressed in salivary gland. This is in contrast to another member of the family, renal kallikrein, which was expressed in salivary gland, kidney, and pancreas. Levels of expression of each of the EGF-BP genes in the mouse salivary gland were investigated. From RNA blot analysis, each of the genes appears to be expressed at a similar level. However, the isolation of clones from a cDNA library, using the gene-specific oligodeoxyribonucleotide probes, showed that EGF-BP type B mRNA is about twice as abundant as that of type A or type C (type A, 8 clones; type B, 14 clones; type C, 8 clones[2]). Expression of both EGF and the binding

[2] Catherine C. Drinkwater, unpublished results.

proteins in the salivary glands is sexually dimorphic; however, no difference in the expression of EGF between male and female kidney is detected.[2]

Discussion

Confusion has surrounded the identity of mouse EGF-BP, with three separate proteins being purified from different strains, namely, EGF-BP type A and type B (Anundi et al., 1982) and the protein we have designated EGF-BP type C (Isackson et al., 1987). In this study, we have identified three distinct members of the mouse glandular kallikrein gene family that encode these proteins.

Comparison of the N-terminal amino acid sequence of EGF-BP type A and type B (Anundi et al., 1982) with the predicted amino acid sequence of the corresponding genes, mGK-22 and mGK-13, shows only one amino acid residue difference in each case. These differences may be due to strain polymorphism, as we have assumed, or to errors in amino acid assignment. In this regard, it is worth noting that each of the discrepancies corresponds to the analogous residue in the sequence of the copurified peptide (Figure 3A). Anundi et al. (1982) reported additional amino acid sequence data from both an internal peptide (6 kDa) and a heterogeneous C-terminal fragment (10 kDa B). The sequence of the 6-kDa peptide shows significant differences from a cDNA sequence for EGF-BP type B reported by Lundgren et al. (1984) in the same laboratory and from our EGF-BP type A sequence. However, the sequence is identical with that encoded by mGK-9 (Figure 3B). Furthermore, the heterogeneous C-terminal peptide appears to be a mixture of EGF-BP type B and type C, not, as claimed, type A. Again, most of the differences between the protein sequence and the predicted amino acid sequence correspond to the analogous residue in the copurified protein (Figure 3B).

The genes mGK-9 (EGF-BP type C) and mGK-13 (EGF-BP type B) are separated by 2 pseudogenes and 1 potentially functional gene in a group of 11 linked kallikrein genes [see Figure 1 and Evans et al. (1987)]. Though mGK-22 (EGF-BP type A) is not directly linked to mGK-9 or -13, its gene product is only 77% homologous to either of their gene products. This falls well within the normal range of homology between members of the mouse glandular kallikrein family (70–90%); accordingly, mGK-22 is not an allele of either mGK-9 or mGK-13. Furthermore, mGK-22 is linked to another kallikrein gene, mGK-21, whose sequence homology to either mGK-1 or mGK-12, the genes adjacent to mGK-9 and -13, respectively, is no greater than normal. Therefore, though two of the three proteins described as EGF-BPs were identified in a different mouse strain than the third, all three are actually encoded by nonallelic genes in the BALB/c strain.

We have previously isolated and partially sequenced 24 glandular kallikrein genes, forming a single genetic locus on chromosome 7 in the mouse (Evans et al., 1987). Sequence and expression data obtained indicate that in BALB/c mice there are at least 10 genes that are expressed in the salivary gland, in addition to 10 pseudogenes and another 4 genes that have the potential to encode functional proteins. All members of the gene family exhibit a high degree of homology, though they contain localized regions of variability. These divergent regions often contain residues thought to be involved in the formation of the substrate binding site, which may therefore be important in determining substrate specificity (Mason et al., 1983). A comparison of these regions in the EGF-BP genes could therefore establish which residues may be important in binding the 9000-dalton EGF precursor. Consequently, we sequenced the entire coding region of each gene and determined the structure of the proteins they encode.

All three genes encode the histidine, serine, and aspartate residues essential for the formation of a serine protease catalytic site (Young et al., 1978) as well as the aspartate residue required for cleavage at basic amino acids (Krieger et al., 1974). No residues were observed to be identical between the three proteins, apart from those in regions conserved between the majority of the kallikreins. Thus, there are no primary structural similarities between the three EGF-BPs which would indicate residues that are important in specifying the binding of the 9000-dalton EGF precursor to the enzyme(s). Blaber et al. (1987) have postulated that the amino acid residue at position 215 is important for specificity. This residue, however, is different in all three genes, being Lys in mGK-9, Ala in mGK-22, and Val in mGK-13.

The 5′ flanking regions of mGK-9, -13, and -22 were also compared to each other and to mGK-1 (Mason et al., 1983), mGK-3 and -4 (Evans & Richards, 1985), and mGK-5[2] and mGK-6 (van Leeuwen et al., 1986). Again there were no obvious similarities which could be implicated in regulation of the three genes, and the only changes from the other kallikrein genes consisted of point mutations and single base insertions and deletions (data not shown).

We had previously shown that, as well as in salivary glands, there are kallikrein genes expressed in mouse kidney, pancreas, brain, liver, testis, and spleen (van Leeuwen et al., 1986). To further characterize the EGF-BP genes, we examined their expression in various tissues. While the prepro-EGF gene is expressed in salivary gland and kidney, mGK-9, -13, and -22 are only expressed in the salivary glands. Immunologically reactive forms of EGF and prepro-EGF mRNA have been detected in several tissues (Byyny et al., 1972; Rall et al., 1985); however, mature EGF as determined by bioassay has only been isolated from salivary glands in the mouse (Cohen, 1962). Thus, even though another member of the mouse glandular kallikrein gene family (mGK-6) is expressed in kidney, this enzyme is not able to process pro-EGF. In the mouse, biologically active EGF is only found where the EGF-BP genes are expressed. All three genes appear to be expressed in relatively similar amounts.

The most obvious explanation of the differing reports on the identity of EGF-BP is that it comprises a mixture of three proteins—type A, type B, and type C, encoded by the genes mGK-22, -13, and -9, respectively. All three proteins appear to be present in the EGF-BP mixture isolated and partially sequenced by Anundi et al. (1982). As yet, we have no means of differentiating whether one, two, or all of the binding proteins are directly involved in the cleavage of mouse EGF. The stoichiometry of the high molecular weight complex (two molecules of EGF/two molecules of EGF-BP) precludes the combined interaction of all three binding proteins with the substrate. Thus, it may be that each protein is capable of acting as an EGF-processing enzyme. This apparent redundancy is somewhat surprising, given the high substrate specificity of other members of the mouse glandular kallikrein family. For example, γ-NGF cannot process the 9000-dalton EGF precursor (Frey et al., 1979), yet is no more divergent from the EGF-BPs than type A, type B, and type C are from each other. Were all three proteins active, this redundancy could account in part for the observed variation in the number of kallikrein genes between different species (Evans et al., 1987). Expression of the EGF-BP genes in cultured cell lines would permit the elucidation of the exact role of their products.

Acknowledgments

We thank Leah Brown, Lucy Duncan, and Drs. Geoff

Tregear and Jim Haralambidis for synthesis of oligodeoxyribonucleotides and Dr. Ralph Bradshaw and his co-workers at the University of California, Irvine, for making sequence information available prior to publication.

References

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

Blaber, M., Isackson, P. J., & Bradshaw, R. A. (1987) *Biochemistry* (preceding paper in this issue).

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

Burmeister, M., Avivi, A., Schlessinger, J., & Soreq, H. (1984) *EMBO J. 3*, 1499–1505.

Byyny, R. L., Orth, D. N., & Cohen, S. (1972) *Endocrinology (Philadelphia) 90*, 1261–1266.

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

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

Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C., & Chambon, P. (1980) *Science (Washington, D.C.) 209*, 1406–1414.

Evans, B. A., & Richards, R. I. (1985) *EMBO J. 4*, 133–138.

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.

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.

Krieger, M., Kay, L. M., & Stroud, R. M. (1974) *J. Mol. Biol. 83*, 209–230.

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*, 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. (1983) *Methods Enzymol. 101*, 20–78.

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

Proudfoot, N. J., & Brownlee, G. G. (1976) *Nature (London) 263*, 211–214.

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.

Ronne, H., Lundgren, S., Sverinsson, 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.

Schlessinger, J., Schreiber, A. B., Levi, A., Lax, I., Libermann, T., & Yarden, Y. (1983) *CRC Crit. Rev. Biochem. 14*, 93–111.

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

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

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

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

Tschesche, H., Mair, G., & Godec, G. (1979) *Adv. Exp. Med. Biol. 120A*, 245–260.

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

Young, C. L., Barker, W. C., Tomaselli, C. M., & Dayhoff, M. O. (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.