Cimbala, M. A., van Lelyveld, P., & Hanson, R. W. (1981) in *Advances in Enzyme Regulation* (Weber, G., Ed.) pp 205–214, Pergamon Press, New York.

Evans, R. M., Birnberg, N. C., & Rosenfeld, M. G. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 7659–7663.

Exton, J. R. (1977) *J. Cyclic Nucleotide Res. 5*, 277–287.

Goodrich, A. G., Jenik, R. A., McDevitt, M. A., Morris, S. M., & Winberry, L. K. (1984) *Arch. Biochem. Biophys. 230*, 82–92.

Granner, D., Andreone, T., Sasaki, K., & Beale, E. (1983) *Nature (London) 305*, 549–551.

Iynedjian, P. B., & Hanson, R. W. (1977) *J. Biol. Chem. 252*, 655–662.

Kioussis, D., Hamilton, R., Hanson, R. W., Tilghman, S. M., & Taylor, J. M. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 4370–4375.

Lamers, W. H., Hanson, R. W., & Meisner, H. M. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 5137–5141.

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

Muller, M. J., Thomsen, A., Sibrowski, W., & Seitz, H. J. (1982) *Endocrinology (Baltimore) 111*, 1469–1475.

Seelig, S., Liaw, C., Towle, H. C., & Oppenheimer, J. H. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 4733–4737.

Sibrowski, W., Muller, M. J., & Seitz, H. J. (1982) *Arch. Biochem. Biophys. 213*, 327–333.

Spindler, S. R., Mellon, S. H., & Baxter, J. D. (1983) *J. Biol. Chem. 257*, 11627–11632.

Williams, R. S., & Lefkowitz, R. J. (1979) *J. Cardiovasc. Pharmacol. 1*, 181–189.

Wynshaw-Boris, A., Lugo, T. G., Short, J. M., Fournier, R. E. K., & Hanson, R. W. (1984) *J. Biol. Chem. 259*, 12161–12170.

Yafee, B. M., & Samuels, H. H. (1984) *J. Biol. Chem. 259*, 6284.

Yoo-Warren, H., Monahan, J. E., Short, J., Short, H., Bruzel, A., Wynshaw-Boris, A., Meisner, H. M., Samols, D., & Hanson, R. W. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 3656–3660.

Kallikrein-Related mRNAs of the Rat Submaxillary Gland: Nucleotide Sequences of Four Distinct Types Including Tonin[†]

Patricia L. Ashley and Raymond J. MacDonald*

Division of Molecular Biology, Department of Biochemistry, The University of Texas Health Science Center at Dallas, Dallas, Texas 75235

Received December 6, 1984

ABSTRACT: We have determined the nucleotide sequence of four submaxillary gland mRNAs, designated PS, S1, S2, and S3, that encode kallikrein and kallikrein-like serine proteases. The four enzymes share between 74% and 86% amino acid sequence identity and are identical in length with the exception of single two amino acid deletions in the S2 and S3 enzymes. The PS enzyme appears to be a true tissue kallikrein. The S1 enzyme shares 86% amino acid sequence homology with the PS enzyme and retains key amino acid residues thought to be primary determinants of kallikrein cleavage specificity. The S2 enzyme is rat submaxillary tonin. The amino acid sequence of the S3 enzyme is identical with tonin at 84% of its amino acid positions and retains the same amino acid substitutions at positions likely to determine substrate cleavage preferences.

Tissue kallikrein is a generic term for a family of serine proteases of closely related structure which are found in many mammalian tissues. The proteases of this family are more specific in cleaving substrates than other simple serine proteases of the pancreatic type. Each has very low activity on substrates such as collagen and casein (Fiedler, 1979; Bothwell et al., 1979). True tissue kallikreins (EC 3.4.21.8) are family members that cleave kininogen to selectively release the vasoactive peptides bradykinin or lysylbradykinin (Seki et al., 1972; Alhenc-Gelas et al., 1981; Yamada & Erdos, 1982). The enzymes are acidic glycoproteins comprising several forms with molecular weights between 25 000 and 40 000 [reviewed by Fiedler (1979)]. Isozymes may vary among tissues but

appear to differ only in their carbohydrate content (Fiedler, 1979). The vasodilatory effect produced by the kinin product of kininogen cleavage by kallikrein may play a role in the regulation of local blood flow in exocrine glands (Hilton, 1970; Hilton & Jones, 1968; Carretero & Scicli, 1981). Tissue kallikreins are distinct from plasma kallikrein, a high molecular weight, complex protease involved in the intrinsic blood coagulation system (Movat, 1979).

The extended family of kallikrein-like enzymes includes proteases of closely related structure with altered specificities and physiological roles not limited to the genesis of kinins. The murine submaxillary gland contains the highest level of kallikrein-like enzymes among the many tissues investigated (Frey et al., 1968; Brandtzaeg et al., 1976). The polypeptide substrates for these enzymes also are found in the submaxillary gland (Barka, 1980). Members of the kallikrein family include tonin, which cleaves angiotensin II from angiotensinogen (Boucher et al., 1974), the γ subunit of nerve growth factor

(NGF),[1] which processes the carboxyl terminus of the NGF precursor (Thomas et al., 1981), epidermal growth factor binding protein, which processes the carboxyl terminus of the EGF precursor (Frey et al., 1979), and rat submaxillary gland kallikrein (Lazure et al., 1981; Gutkowska et al., 1983). Comparison of the amino acid sequences of these and other glandular kallikreins reveals homologies of 60–75% (Fiedler & Fritz, 1981; Swift et al., 1982; Mason et al., 1983). This high level of sequence identity sets the kallikrein-related enzymes apart as a subfamily distinct from other simple serine proteases such as trypsin, chymotrypsin, and elastase with which they share 35–50% amino acid sequence identity.

To understand the role of multiple kallikrein-like enzymes, we have begun to define the number of expressed genes of the kallikrein family and the properties, the degree of relatedness, and the enzymatic function of the encoded proteases. In this report, we describe the molecular cloning of four distinct kallikrein-like messenger RNAs (designated PS, S1, S2/tonin, and S3) expressed in rat submaxillary gland, their nucleotide sequences, and the amino acid sequences of the enzymes they encode. On the basis of comparisons of the amino acid sequences of the encoded proteins, these kallikrein-like enzymes fall into two subclasses. One class contains the PS enzyme, which is identical with the rat pancreatic (Swift et al., 1982) and submaxillary gland (Lazure et al., 1981) enzyme that has true kallikrein enzyme activity (Gutkowska et al., 1983; C. Thibault, personal communication). The second enzyme (S1) of this subclass has a very high amino acid sequence homology (86%) to PS and retains the key amino acid residues thought to be primary determinants of kallikrein cleavage specificity. The second subclass contains the enzyme tonin, a serine protease that processes angiotensinogen directly to angiotensin II in vitro (Boucher et al., 1974; Lazure et al., 1984), and a closely related enzyme, S3.

EXPERIMENTAL PROCEDURES

Preparation of Rat Submaxillary Gland RNA. RNA was isolated from the submaxillary glands of adult male Sprague-Dawley rats by using a slightly modified version of the guanidine thiocyanate extraction procedure (Chirgwin et al., 1979). Samples containing 1–3 g of tissue were homogenized in 15 mL of 4 M guanidine thiocyanate, 0.1 M Tris-HCl, pH 7.5, and 0.1 M 2-mercaptoethanol for 20 s at high speed with a Tissumizer (Tekmar Industries). The homogenates were layered over 20 mL of 5.7 M CsCl dissolved in 0.1 M Na$_2$EDTA, pH 7.0 [treated with 0.1% diethyl pyrocarbonate (DEP) and autoclaved], in 40-mL Beckman SW 28 ultracentrifuge tubes. The samples were centrifuged for 22 h at 22000 rpm. The RNA pellets were resuspended in 1 mL of sterile, DEP-treated water. One-tenth volume of 2 M NaOAc, pH 7.0 (DEP-treated), and 2 volumes of EtOH were added, and the RNAs were precipitated at –20 °C overnight. After centrifugation at 8000 rpm for 20 min in a Sorvall HB4 rotor, the pellets were vacuum dried, resuspended in DEP-treated water, and stored at –80 °C. The total RNA from 10 glands was pooled to isolate polyadenylated RNA by oligo(dT)–cellulose chromatography (Aviv & Leder, 1972; MacDonald et al., 1982a).

Preparation and Screening of Rat Submaxillary Gland Double-Stranded Complementary DNA (ds-cDNA) Library.

A ds-cDNA library comprising approximately 20000 recombinant clones was constructed according to the method of Okayama & Berg (1982) using 1 μg of polyadenylated RNA from rat submaxillary gland.

The rat submaxillary gland ds-cDNA library was plated on nitrocellulose filters as described by Hanahan & Meselson (1980) at a density of approximately 3000 bacterial colonies per 137-mm filter. Replica filters made from the originals were incubated on chloramphenicol plates to amplify the plasmid DNA and processed by successive application to 2-mL pools of 0.5 N NaOH (twice), 1 M Tris-HCl, pH 8.0, and 1 M Tris-HCl (pH 8.0)/1.5 M NaCl. The filters were air-dried, and the DNA was fixed by baking at 80 °C for 2 h in a vacuum oven. The filters were prehybridized for 16 h in a solution containing 5× SSC, 50% formamide, 10× Denhardt's solution (Denhardt, 1966), 50 mM sodium phosphate, pH 7.0, 0.1% SDS, and 250 μg/mL sonicated and denatured salmon testes DNA at 42 °C. The same conditions were used for hybridization with the addition of 5 × 10^6 cpm per filter of a ^{32}P-labeled kallikrein cDNA probe. The probe sequences were derived from the cDNA plasmid pcXP39 containing the 3′ 550 bp of rat pancreatic kallikrein mRNA (Swift et al., 1982) which encode the carboxy-terminal 167 amino acids of kallikrein plus the 3′-untranslated region.

The hybridization probe was prepared from an M13mp8 subclone of the ds-cDNA insert of pcXP39 (Ashley & MacDonald, 1984) in a 20-μL reaction containing 5 μg of single-stranded viral recombinant DNA, 100 μg/mL oligo(dT), 50 mM Tris-HCl, pH 8.3, 12 mM MgCl$_2$, 1 mM dATP, dGTP, and dTTP, 9 μM dCTP, 50 μCi of [α-^{32}P]dCTP (New England Nuclear), 20 mM DTT, and 20 units of reverse transcriptase. The mixture was incubated at 42 °C for 1 h; the unincorporated nucleotides were removed by centrifugation through a Bio-Gel P-30 column (MacDonald et al., 1982a) and denatured in 50% formamide by boiling for 5 min. Alternatively, a nick-translated (Rigby et al., 1977) ds-cDNA insert from a recombinant plasmid, pcXP130, bearing the 5′ 500 bp of rat pancreatic kallikrein mRNA (Swift et al., 1982) was used as the ^{33}P-labeled probe. After hybridization, the filters were washed at 55 °C in two 500-mL changes of 2× SSC and 0.1% SDS for 20 min each and two 500-mL changes of 2× SSC for 20 min each, air-dried, and exposed to X-ray film at –80 °C for 16 h with an intensifying screen. After autoradiography, colonies were selected that hybridized to the ^{32}P-labeled kallikrein cDNA probe.

Isolation of Plasmid DNAs. Recombinant plasmids bearing kallikrein cDNA inserts were prepared by the alkaline SDS lysis procedure (Birnboim & Doly, 1979). Large-scale preparations from 1-L cultures were further purified by ultracentrifugation through cesium chloride gradients containing ethidium bromide (Maniatis et al., 1982).

Nucleotide Sequence Analysis. The nucleotide sequences of the cloned kallikrein-related cDNAs were determined by using the base-specific chemical cleavage procedures of Maxam & Gilbert (1980). Five sequencing reactions (G, dimethyl sulfate; G + A, formic acid; C + T, hydrazine; C, hydrazine + NaCl; A > C, NaOH) were employed to enhance sequence accuracy. DNA fragments were labeled at their 5′ ends with [γ-^{32}P]ATP and polynucleotide kinase (New England Nuclear) after treatment with calf intestine alkaline phosphatase (Boehringer Mannheim) or labeled at their 3′ termini with Klenow polymerase (P-L Biochemicals) and the appropriate deoxyribonucleoside [α-^{32}P]triphosphate (Drouin, 1980). Labeled fragments were isolated by polyacrylamide gel electrophoresis and eluted from gel slices by using the

---

[1] Abbreviations: NGF, nerve growth factor; ds-cDNA, double-stranded complementary DNA; SSC, 0.15 M sodium chloride and 0.15 M sodium citrate; DTT, dithiothreitol; EGF-BP, epidermal growth factor binding protein; Tris-HCl, tris(hydroxymethyl)aminomethane hydrochloride; EDTA, ethylenediaminetetraacetic acid; DEP, diethyl pyrocarbonate; SDS, sodium dodecyl sulfate; bp, base pair(s).
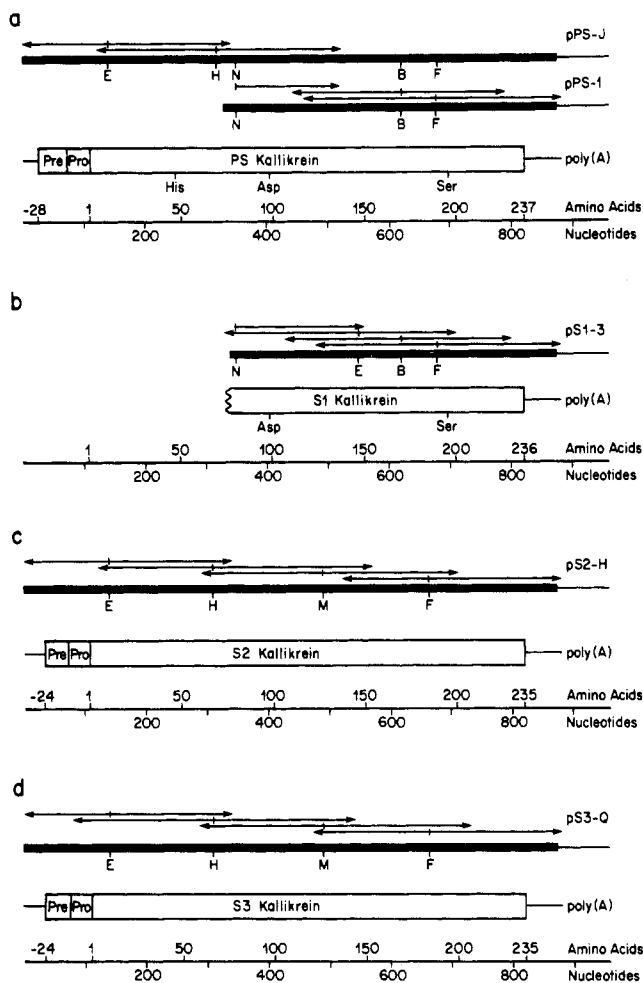
FIGURE 1: Sequencing strategies for the four rat submaxillary gland kallikrein mRNAs. The open horizontal rectangles delineate the amino acid coding regions; the proposed signal peptide (Pre) and activation peptide (Pro) regions are indicated. The lines extending from the open rectangles represent the lengths of the noncoding regions of each mRNA; poly(A) is present at the 3' end of each mRNA as indicated. The bold horizontal lines represent the extent of each cloned ds-cDNA; the designation for each ds-cDNA clone is indicated. The direction and length of each sequencing run are indicated by the horizontal arrows, each beginning at the restriction endonuclease site indicated (E, EcoRI; H, HindIII; N, NdeI; B, BglII; F, HinfI; M, MstII). Panel a illustrates the sequencing strategy for PS kallikrein mRNA, panel b for S1 mRNA, panel c for S2 mRNA, and panel d for S3 mRNA.

electroelution method of Zassenhaus et al. (1982).

The strategies for determining the nucleotide sequences of the four submaxillary gland kallikrein mRNAs are summarized in Figure 1. The full-length sequence of the pancreas-like kallikrein mRNA, denoted PS, was obtained from recombinant plasmids pPS-1 and pPS-J. Both strands were sequenced at least once over 78% of the length of the ds-cDNA inserts. The composite sequence begins at the 5'-noncoding region of the mRNA and includes a long poly(A) tail at the 3' end. The submaxillary gland kallikrein mRNA called S1 is represented by a single cDNA clone, pS1-3. Of the cloned length of S1 mRNA present in pS1-3, 88% of the sequence was determined by at least one run on both strands and 56% by two runs on both strands. The sequence of pS1-3 begins at codon 80 in the coding region and extends through the poly(A) tail. Rescreening of the submaxillary gland cDNA library with the 5' mRNA probe from pcXP130 did not yield any additional recombinant plasmids of the S1 type. The third class of submaxillary gland killikrein mRNA, S2, is represented by recombinant plasmid pS2-H. The cDNA insert of

this clone was sequenced over 75% of its length by at least one run on both strands. This sequence begins with the 5'-non-coding region of the S2 mRNA and extends through the poly(A) tail. The full-length sequence of a fourth submaxillary gland kallikrein mRNA, designated S3, was obtained from recombinant plasmid pS3-Q, which was sequenced at least once on both strands over 80% of the length of the ds-cDNA insert.

RESULTS

Isolation of Cloned ds-cDNA Sequences for Submaxillary Gland Kallikrein-Related mRNAs. A library comprising approximately 20 000 recombinants was constructed with the cDNA cloning vectors designed by Okayama & Berg (1982) and screened with a cDNA probe covering the 3' 550 bp of the rat pancreatic kallikrein mRNA. Of the recombinant plasmids that hybridized, 32 were analyzed by restriction enzyme digestion to determine the length of the mRNA sequences present in the cDNA inserts (data not shown). Of these, six appeared to be full-length or nearly full-length cDNA clones with insert sizes of approximately 900 bp. Three of these and two shorter cDNA clones were sequenced (Figure 1) as described below. The three remaining full-length cDNA clones were found to represent the pancreas-like (PS) class of kallikrein mRNA after one sequencing run and were not analyzed further. On the basis of the nucleotide sequence analyses (see below), the cloned cDNAs fell into four distinct classes representing unique kallikrein or kallikrein-related mRNAs: PS, S1, S2, and S3 mRNAs. The remaining 24 cDNA clones were mapped with restriction endonucleases and appeared to represent additional copies of the same PS, S2, and S3 mRNAs.

Complete Nucleotide Sequence of PS Kallikrein mRNA. The strategy to obtain the nucleotide sequence of the submaxillary gland kallikrein mRNA designated PS is shown in Figure 1a. Two recombinant plasmids containing overlapping insert sequences were used to derive the PS mRNA sequence through multiple sequencing runs which maximized sequence accuracy. One of these cDNA clones, pPS-J, contained an insert which represented the full length of the PS kallikrein mRNA. The other plasmid, pPS-1, contained insert sequences extending from codon 70 through the poly(A) tail of the mRNA.

The sequence of the submaxillary gland PS kallikrein mRNA and the derived amino acid sequence of the enzyme are shown in Figure 2. The mRNA is 873 nucleotides long plus a poly(A) tract and includes a 5'-untranslated region of about 30 nucleotides and a 3'-untranslated region of 48 nucleotides. It is identical with the kallikrein mRNA expressed in pancreas (Swift et al., 1982), with the exception of six additional nucleotides at the 5' end. A second cDNA clone of the PS class isolated from the submaxillary gland library, pPS-G, lacked the six 5'-nucleotides of pPS-J and, therefore, was identical with the pancreatic mRNA sequence. These two classes of submaxillary gland and pancreatic PS cDNA clones that differ by six nucleotides may represent mRNAs transcribed from a single gene which has two transcription start sites six nucleotides apart (see Discussion).

The submaxillary gland PS mRNA prescribes a preproenzyme of 265 amino acids with a molecular weight of 29 285. This glandular kallikrein has the amino acid residues of the serine protease catalytic triad at the appropriate positions—histidine-41, aspartate-96, and serine-189—and other amino acids characteristic of serine proteases (Young et al., 1978).

Partial Nucleotide Sequence of S1 mRNA. The recombinant plasmid pS1-3 represents a second class of mRNA

```
                                          -28                                    -20                                        -10
                    UAGGAAGGCGU          met pro val thr met trp phe leu  Ile leu phe leu ala leu ser leu gly arg asn asp ala ala
AGCUCCAAGCUCAGCACCUGCUGCUCCUGC           AUG CCU GUU ACC AUG UGG UUC CUG  AUC CUG UUC CUC GCC CUG UCC CUG GGA CGG AAU GAU GCU GCA   96

         -1  +1                                              10                                     20
pro pro val gln ser arg val val gly gly tyr asn cys glu met asn ser gln pro trp gln val ala val tyr tyr phe gly glu tyr
CCU CCC GUC CAG UCU CGG GUU GUU GGA GGA UAU AAC UGU GAG AUG AAU UCC CAA CCC UGG CAA GUG GCU GUG UAC UAC UUC GGC GAA UAC   186

                30                                              40
leu cys gly gly val leu Ile asp pro ser trp val Ile thr ala ala his cys ala thr asp asn tyr gln val trp leu gly arg asn
CUA UGU GGG GGU GUC CUG AUA GAC CCC AGC UGG GUG AUC ACA GCU GCU CAC UGC GCA ACC GAC AAU UAC CAG GUU UGG CUG GGC CGA AAC   276

asn leu tyr glu asp glu pro phe ala gln his arg leu val ser gln ser phe pro his pro gly phe asn gln asp leu Ile trp asn
AAC CUA UAU GAA GAU GAA CCC UUU GCU CAG CAC CGG CUU GUC AGC CAA AGC UUC CCU CAC CCC GGU UUC AAC CAG GAC CUC AUA UGG AAC   366

                90                                              100                                    110
his thr arg gln pro gly asp asp tyr ser asn asp leu met leu leu his leu ser gln pro ala asp Ile thr asp gly val lys val
CAC ACC CGA CAA CCU GGG GAC GAC UAC AGC AAU GAU UUG AUG CUG CUC CAC CUC AGC CAG CCU GCG GAC AUC ACA GAU GGU GUG AAG GUC   456

                120                                            130                                    140
Ile asp leu pro Ile glu glu pro lys val gly ser thr cys leu ala ser gly trp gly ser Ile thr pro asp gly leu glu leu ser
AUC GAU CUG CCC AUU GAG GAG CCC AAG GUG GGG AGC ACC UGC CUU GCC UCG GGC UGG GGC AGC AUC ACA CCU GAC GGA UUG GAA UUA AGU   546

                150                                            160                                    170
asp asp leu gln cys val asn Ile asp leu leu ser asn glu lys cys val glu ala his lys glu glu val thr asp leu met leu cys
GAU GAU CUC CAG UGU GUG AAC AUC GAU CUU CUG UCU AAU GAG AAG UGC GUC GAG GCA CAC AAA GAA GAG GUG ACA GAU CUC AUG CUG UGU   636

                180                                            190                                    200
ala gly glu met asp gly gly lys asp thr cys lys gly asp ser gly gly pro leu Ile cys asn gly val leu gln gly Ile thr ser
GCA GGA GAG AUG GAU GGU GGC AAA GAC ACU UGC AAG GGU GAC UCA GGA GGC CCC CUC AUC UGU AAU GGU GUG CUC CAA GGC AUC ACG UCC   726

                210                                            220                                    230
trp gly phe asn pro cys gly glu pro lys lys pro gly Ile tyr thr lys leu Ile lys phe thr pro trp Ile lys glu val met lys
UGG GGC UUU AAC CCA UGC GGU GAA CCC AAG AAG CCA GGC AUC UAC ACC AAA CUU AUU AAG UUC ACC CCC UGG AUA AAA GAA GUU AUG AAG   816

        237
glu asn pro stop
GAA AAC CCC UGA GUGUCACACAGUCCCCUGGUCUCAAUAAAACCCACCAUGCAGCAC-poly(A)                                                    873
```
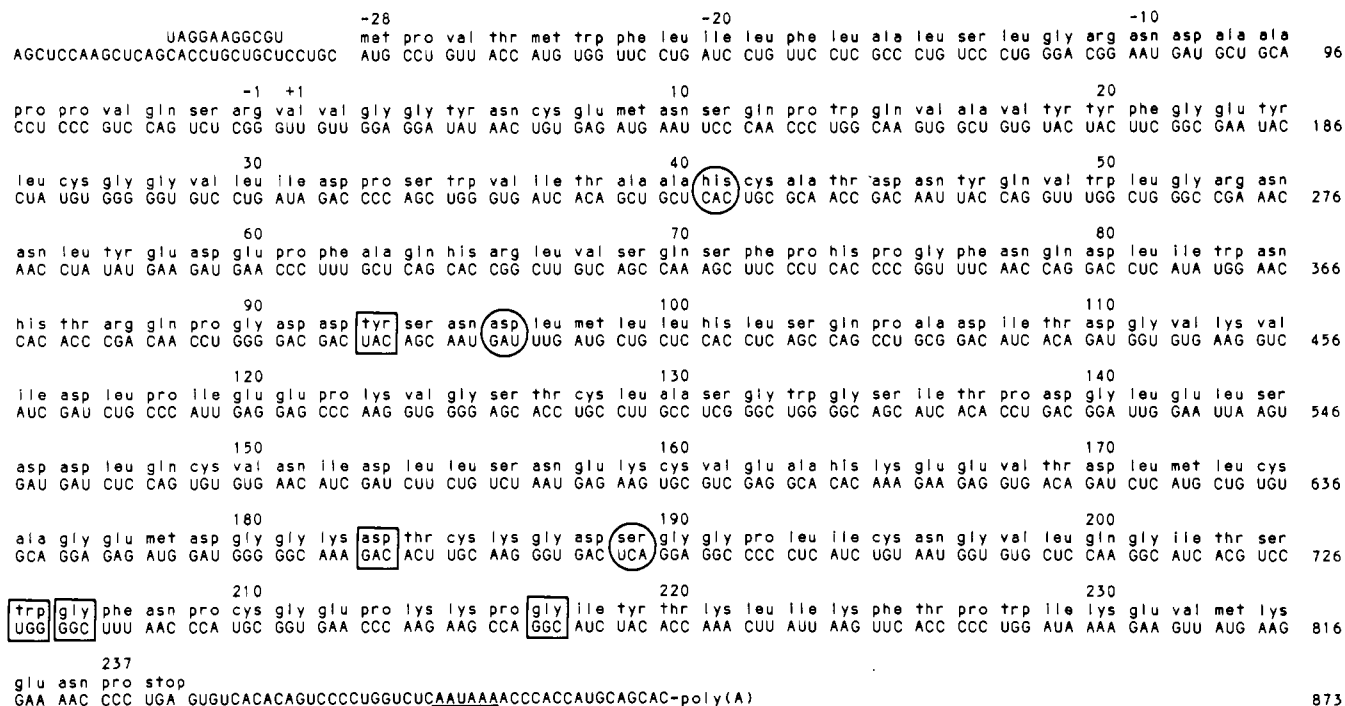
FIGURE 2: Complete nucleotide sequence of PS kallikrein mRNA and the amino acid sequence of the encoded preproenzyme. Amino acid numbering starts at the first residue of the mature enzyme. The amino acid residues of the serine protease catalytic triad are *circled*. Amino acid residues important in determining substrate specificity that are discussed in the text are *boxed*. The conserved AAUAAA in the 3'-noncoding region of the mRNA is *underlined*. The conserved nucleotide sequence near the 3' end of eukaryotic 18S rRNA (Hagenbuchle et al., 1978) is shown at the position with the greatest number of base pairings within the 5'-noncoding region of the mRNA.

```
      81                                    90                                    100
      leu Ile trp asn his thr arg lys pro gly asp tyr ser asn asp leu met leu leu his leu ser glu pro ala asp Ile thr asp
AC    CUC AUA UGG AAC CAC ACC CGA AAA CCU GGG GAC UAC AGC AAU GAC CUG AUG CUG CUC CAC CUC AGC GAG CCU GCG GAC AUC ACA GAU   89

      110                                    120                                    130
gly val lys val Ile asp leu pro thr lys glu pro lys val gly ser thr cys leu val ser gly trp gly ser thr asn pro ser glu
GGG GUG AAG GUC AUC GAU CUG CCC ACG AAG GAG CCC AAG GUG GGG AGC ACC UGC CUU GUC UCA GGC UGG GGC AGC ACC AAC CCC UCU GAG   179

      140                                    150                                    160
trp glu phe pro asp asp leu gln cys val asn Ile his leu leu ser asn glu lys cys Ile lys ala tyr lys glu lys val thr asp
UGG GAA UUC CCU GAU GAU CUC CAG UGU GUG AAC AUC CAC CUA CUG UCU AAU GAG AAG UGC AUC AAA GCC UAC AAA GAA AAG GUG ACA GAU   269

      170                                    180                                    190
leu met leu cys ala gly glu leu glu gly gly lys asp thr cys arg gly asp ser gly gly pro leu Ile cys asp gly val leu gln
CUG AUG CUG UGU GCA GGA GAG UUG GAA GGG GGC AAA GAC ACU UGC AGG GGU GAC UCA GGA GGC CCC CUC AUC UGU GAU GGU GUG CUC CAA   359

      200                                    210                                    220
gly Ile thr ser trp gly ser val pro cys gly glu pro asn lys pro gly Ile tyr thr lys leu Ile lys phe thr ser trp Ile lys
GGC AUC ACA UCA UGG GGC UCU GUC CCA UGC GGU GAA CCC AAU AAG CCA GGC AUC UAC ACC AAA CUU AUU AAG UUC ACC UCC UGG AUA AAA   449

      230                        236
glu val met lys glu asn pro stop
GAA GUU AUG AAG GAA AAC CCC UGA GUGUCACACUGUACCCUGAUCUCAAUAAAACCCACCAUGCAGCAC-poly(A)                                    518
```
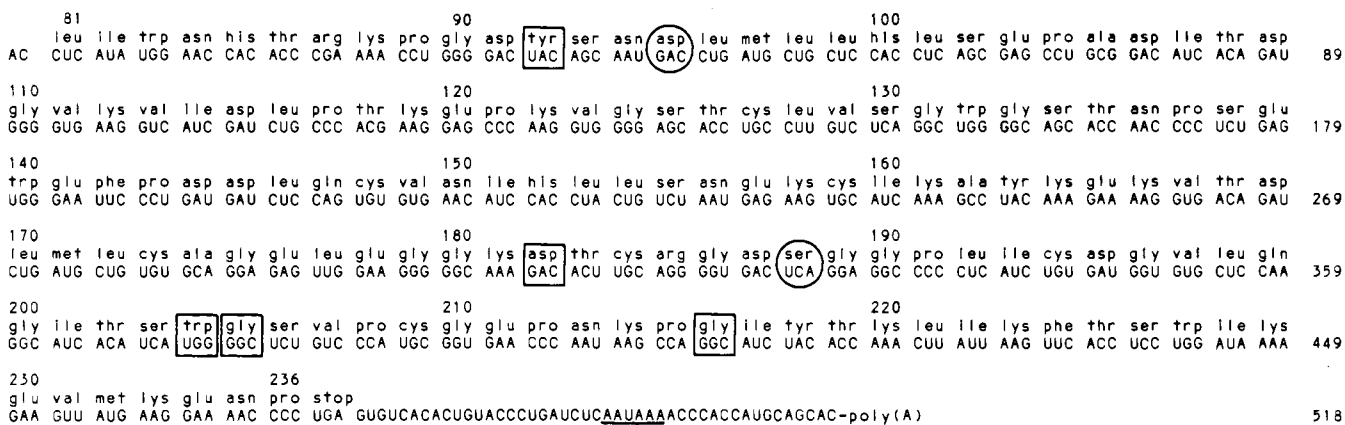
FIGURE 3: Nucleotide sequence for S1 mRNA and the amino acid sequence of the encoded protein fragment. The partial amino acid sequence begins at the amino acid residue corresponding to number 80 of the PS enzyme. Catalytically important amino acids and conserved nucleotide sequences are indicated as described in the legend to Figure 2.

encoding a kallikrein-like enzyme (Figure 1b). pS1-3 contained an insert of a truncated mRNA sequence extending from codon 80 through the poly(A) tail. Attempts to find additional clones encoding the 5' end of this mRNA were unsuccessful. The partial mRNA sequence (Figure 3) is 518 nucleotides long, including a 3'-nontranslated region of 48 nucleotides, and is 94% homologous to the corresponding region of the PS kallikrein mRNA. The encoded protein fragment of 158 amino acids contains the aspartate and serine residues of the catalytic triad as well as other amino acids which contribute to the substrate binding pocket. The partial amino acid sequence is identical at 86% of the residues with that of the comparable region of PS kallikrein (Table I) with a single amino acid deletion at residue 91. No changes occur between PS and S1 at key amino acid residues that have been proposed to be principal determinants of kallikrein-like cleavage specificity. Aspartate-183 at the bottom of the binding pocket, glycines-206 and -217 at the mouth of the binding pocket, and tyrosine-93 plus tryptophan-205, which

Table I: Relatedness of the Four Submaxillary Gland Kallikrein-like Enzymes and Their mRNAs[a]

	PS	S1	S2	S3
S1[b]	94/86			
S2	85/74	92/78		
S3	85/76	91/76	91/84	
trypsin[c]	/43	/49	/40	/41

[a] Numbers preceding the slant, percent nucleotide sequence identity of the entire mRNA sequence; numbers following the slant, percent amino acid sequence identity of the preproenzymes. [b] Comparison only for the partial S1 mRNA sequence comprising nucleotides 351–873 and S1 protein amino acid residues 80–237. [c] For rat trypsin [from MacDonald et al. (1982a)], only the amino acid sequence relatedness is shown.

have been proposed to define the P2[2] specificity of kallikrein (Bode et al., 1983), are conserved.

―――――――――――――

[2] The nomenclature of Schecter & Berger (1967) is used to identify amino acid positions of the polypeptide substrate.

```
                              -24           -20                                  -10
        UAGGAAGGCGU           met trp leu gln ile leu ser leu val leu ser val gly arg ile asp ala ala pro pro gly
AGAUCACCACCUGCUUCUCCUGGACACCUGAUACC AUG UGG CUC CAG AUC CUG UCC CUA GUC CUG UCC GUG GGA CGA AUU GAU GCU GCA CCU CCU GGC   98

     -1  +1                              10                                      20
gln ser arg ile val gly gly tyr lys cys glu lys asn ser gln pro trp gln val ala val ile asn glu tyr leu cys gly gly val
CAG UCU CGG AUU GUU GGA GGA UAU AAG UGU GAG AAG AAU UCC CAA CCC UGG CAA GUG GCU GUC AUC AAU GAA UAC CUC UGC GGG GGU GUC   188

          30                              40                                    50
leu ile asp pro ser trp val ile thr ala ala(his)cys tyr ser asn asn tyr gln val leu leu gly arg asn asn leu phe lys asp
CUG AUA GAC CCC AGC UGG GUG AUC ACG GCU GCC CAC UGC UAU AGC AAC AAU UAC CAG GUU UUG CUG GGC CGA AAC AAC CUA UUU AAA GAU   278

          60                              70                                    80
glu pro phe ala gln arg arg leu val arg gln ser phe arg his pro asp tyr ile pro leu ile val thr asn asp thr glu gln pro
GAA CCC UUU GCU CAG CGC CGG CUU GUC AGG CAA AGC UUC CGA CAC CCU GAC UAU AUC CCA CUC AUC GUG ACG AAU GAC ACU GAA CAA CCU   368

          90                              100                                   110
val his asp[his]ser asn(asp)leu met leu leu his leu ser glu pro ala asp ile thr gly gly val lys val ile asp leu pro thr
GUG CAU GAC CAC AGC AAU GAC CUG AUG CUG CUC CAC CUC AGC GAG CCU GCU GAC AUC ACA GGU GGU GUG AAG GUU AUC GAU CUG CCC ACU   458

          120                             130                                   140
lys glu pro lys val gly ser thr cys leu ala ser gly trp gly ser thr asn pro ser glu met val ser his asp leu gln cys
AAG GAG CCC AAG GUG GGG AGC ACC UGC CUU GCC UCA GGC UGG GGC AGC ACC AAC CCC UCU GAG AUG GUA GUA AGU CAU GAU CUC CAG UGU   548

          150                             160                                   170
val asn ile his leu leu ser asn glu lys cys ile glu thr tyr lys asp asn val thr asp val met leu cys ala gly glu met glu
GUG AAC AUC CAC CUA CUG UCU AAU GAG AAG UGC AUC GAA ACC UAC AAA GAC AAU GUG ACA GAU GUC AUG CUG UGU GCA GGA GAG AUG GAA   638

          180                             190                                   200
gly gly lys[asp]thr cys ala gly asp(ser)gly gly pro leu ile cys asp gly val leu gln gly ile thr ser[gly][gly]ala thr pro
GGA GGA AAA GAC ACU UGU GCG GGU GAC UCA GGA GGC CCC CUC AUC UGU GAU GGU GUG CUC CAA GGC AUC ACA UCG GGG GGC GCU ACC CCA   728

          210                             220                             230       235
cys ala lys pro lys thr pro[ala]ile tyr ala lys leu ile lys phe thr ser trp ile lys lys val met lys glu asn pro stop
UGU GCU AAA CCC AAA ACA CCG GCA AUC UAU GCC AAA CUU AUU AAG UUC ACC UCC UGG AUA AAA AAA GUU AUG AAG GAA AAC CCC UGA GUG   818

UCACACUGUGCCCGGUCUCAAUAAAACCCACCAUGCAGC-poly(A)                                                                             858
```
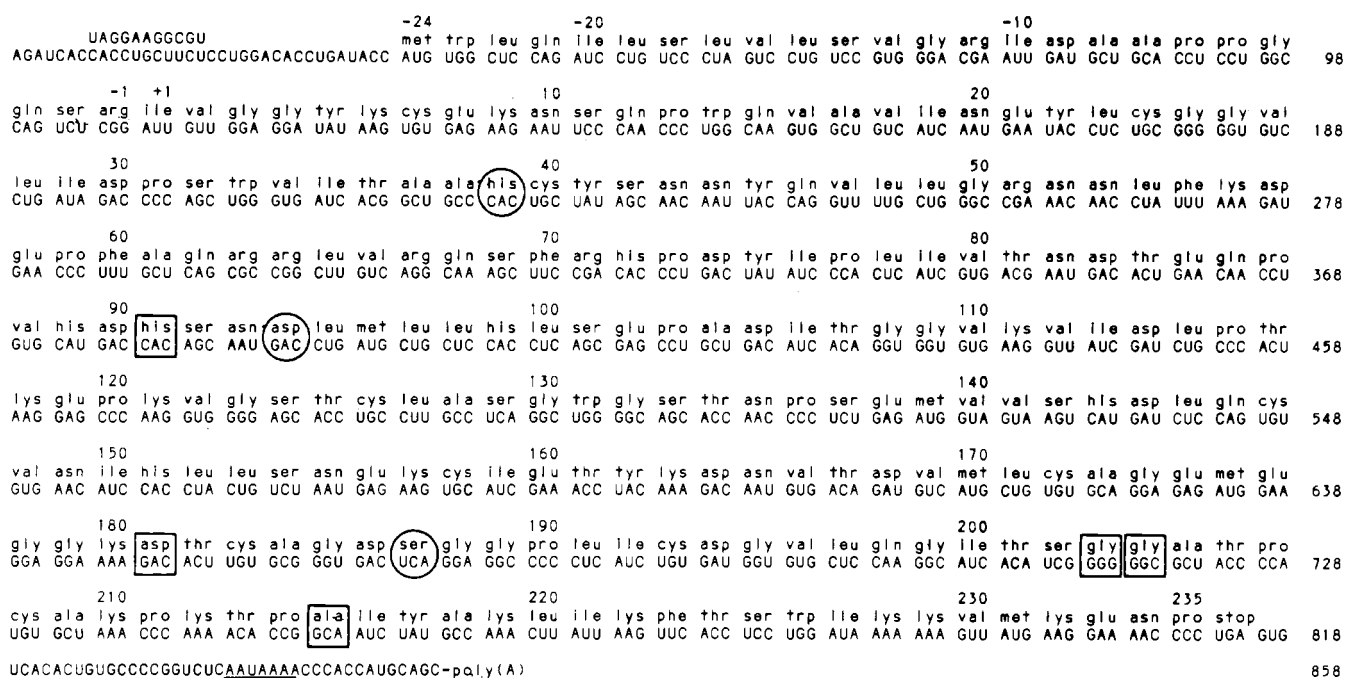
FIGURE 4: Complete nucleotide sequence for S2 mRNA and the amino acid sequence of the encoded preproenzyme. Important amino acids and conserved nucleotide sequences are indicated as described in the legend to Figure 2.
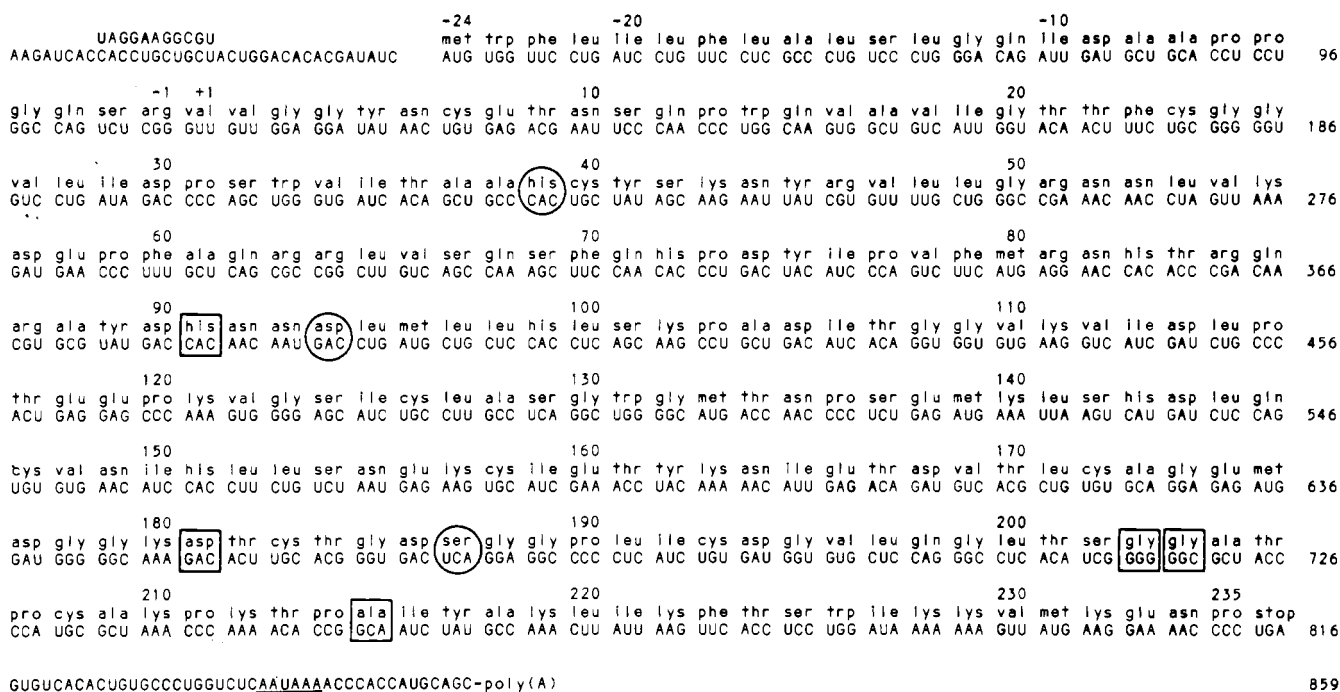
```
                              -24           -20                                  -10
        UAGGAAGGCGU           met trp phe leu ile leu phe leu ala leu ser leu gly gln ile asp ala ala pro pro
AAGAUCACCACCUGCUGCUACUGGACACACGAUAUC AUG UGG UUC CUG AUC CUG UUC CUC GCC CUG UCC CUG GGA CAG AUU GAU GCU GCA CCU CCU     96

     -1  +1                              10                                      20
gly gln ser arg val val gly gly tyr asn cys glu thr asn ser gln pro trp gln val ala val ile gly thr thr phe cys gly gly
GGC CAG UCU CGG GUU GUU GGA GGA UAU AAC UGU GAG ACG AAU UCC CAA CCC UGG CAA GUG GCU GUC AUU GGU ACA ACU UUC UGC GGG GGU   186

          30                              40                                    50
val leu ile asp pro ser trp val ile thr ala ala(his)cys tyr ser lys asn tyr arg val leu leu gly arg asn asn leu val lys
GUC CUG AUA GAC CCC AGC UGG GUG AUC ACA GCU GCC CAC UGC UAU AGC AAG AAU UAU CGU GUU UUG CUG GGC CGA AAC AAC CUA GUU AAA   276

          60                              70                                    80
asp glu pro phe ala gln arg arg leu val ser gln ser phe gln his pro asp tyr ile pro val phe met arg asn his thr arg gln
GAU GAA CCC UUU GCU CAG CGC CGG CUU GUC AGC CAA AGC UUC CAA CAC CCU GAC UAC AUC CCA GUC UUC AUG AGG AAC CAC ACC CGA CAA   366

          90                              100                                   110
arg ala tyr asp[his]asn asn(asp)leu met leu leu his leu ser lys pro ala asp ile thr gly gly val lys val ile asp leu pro
CGU GCG UAU GAC CAC AAC AAU GAC CUG AUG CUG CUC CAC CUC AGC AAG CCU GCU GAC AUC ACA GGU GGU GUG AAG GUC AUC GAU CUG CCC   456

          120                             130                                   140
thr glu glu pro lys val gly ser ile cys leu ala ser gly trp gly met thr asn pro ser glu met lys leu ser his asp leu gln
ACU GAG GAG CCC AAA GUG GGG AGC AUC UGC CUU GCC UCA GGC UGG GGC AUG ACC AAC CCC UCU GAG AUG AAA UUA AGU CAU GAU CUC CAG   546

          150                             160                                   170
cys val asn ile his leu leu ser asn glu lys cys ile glu thr tyr lys asn ile glu thr asp val thr leu cys ala gly glu met
UGU GUG AAC AUC CAC CUU CUG UCU AAU GAG AAG UGC AUC GAA ACC UAC AAA AAC AUU GAG ACA GAU GUC ACG CUG UGU GCA GGA GAG AUG   636

          180                             190                                   200
asp gly gly lys[asp]thr cys thr gly asp(ser)gly gly pro leu ile cys asp gly val leu gln gly leu thr ser[gly][gly]ala thr
GAU GGG GGC AAA GAC ACU UGC ACG GGU GAC UCA GGA GGC CCC CUC AUC UGU GAU GGU GUG CUC CAG GGC CUC ACA UCG GGG GGC GCU ACC   726

          210                             220                             230       235
pro cys ala lys pro lys thr pro[ala]ile tyr ala lys leu ile lys phe thr ser trp ile lys lys val met lys glu asn pro stop
CCA UGC GCU AAA CCC AAA ACA CCG GCA AUC UAU GCC AAA CUU AUU AAG UUC ACC UCC UGG AUA AAA AAA GUU AUG AAG GAA AAC CCC UGA   816

GUGUCACACUGUGCCCUGGUCUCAAUAAAACCCACCAUGCAGC-poly(A)                                                                         859
```

FIGURE 5: Complete nucleotide sequence of S3 mRNA and the amino acid sequence of the encoded preproenzyme. Important amino acids and conserved nucleotide sequences are indicated as described in the legend to Figure 2.

Complete Nucleotide Sequence of S2 mRNA. The recombinant plasmid pS2-H contained the mRNA sequence (Figure 1c) encoding the kallikrein-like enzyme tonin (see Discussion). The cloned S2 mRNA sequence is 858 nucleotides, with a 5'-noncoding region of 35 nucleotides and a 3'-untranslated region of 46 nucleotides (Figure 4), and is 85% homologous to the PS mRNA (Table I).

The S2 mRNA encodes a preproenzyme of 259 amino acids (M_r 28 375) that is 74% homologous to the PS kallikrein. Compared to the PS kallikrein, the S2 enzyme has a deletion of two amino acid residues between valine-18 and glutamate-23 (see Figure 8). In addition, changes occur at key amino acid residues. While aspartate-183 at the bottom of the substrate binding pocket is conserved, glycine-217 is replaced by alanine, and tyrosine-93 and tryptophan-205 are replaced by histidine-91 and glycine-203 in the S2 protein.

Complete Nucleotide Sequence of S3 mRNA. A full-length cDNA clone, pS3-Q, represented the fourth kallikrein-related mRNA (Figure 1d). The S3 mRNA is 859 nucleotides long, including a 5'-untranslated region of 36 nucleotides and a 3'-untranslated region of 46 nucleotides (Figure 5). The S3 mRNA prescribes a preproenzyme of 259 amino acids and molecular weight 28 372. The mRNA is 85% homologous to the PS mRNA and the encoded protein 76% homologous to PS kallikrein (Table I). The S3 enzyme has a different two amino acid deletion between valine-18 and cysteine-26 (relative to the PS kallikrein) than does the S2 enzyme (see Figure 8). Key amino acid residues that are the principal determinants

```
              3'...UAGGAAGGCGU...5'                    18S rRNA

       *    *                    (MetProValThr)Met
AGCUCCAAGCUCAGCACCuGCuGCUCCUGCAUGCCUGUUACC AUG...PS mRNA (24 to 42)
        AGAUCACCACCuGCuUCUCCUGGACACCUGAUACC AUG...S2 mRNA (35)
        AAGAUCACCACCuGCuGCUACUGGACACACGAUAUC AUG...S3 mRNA (36)
AGCUCCAAGCTCACUGCCuGCuGCUCCUGAACACCUGUUACC AUG...mGK-1 mRNA (36 or 42)
       *    *
```

FIGURE 6: Comparison of the 5'-untranslated regions of murine kallikrein-related mRNAs. Numbers in *parentheses* indicate the nucleotide length of each untranslated region. Potential base pairings with a conserved 3'-oligonucleotide sequence of eukaryotic 18S mRNA (Hagenbuchle et al., 1978) are *underlined*. Lower case *u* indicates potential G·U pairings. *Asterisks* indicate two potential transcription start sites. The mGK-1 sequence is taken from a mouse submaxillary gland kallikrein gene (Mason et al., 1983).

of the size, shape, and ionic character of the substrate binding pocket (e.g., aspartate-183, alanine-215, histidine-91, and glycine-203) are identical with those of the S2 enzyme. The S2 and S3 submaxillary gland kallikrein-like enzymes are 91% homologous to one another at the nucleotide level and share 84% amino acid sequence identity.

DISCUSSION

Screening a rat submaxillary gland cDNA library with a rat pancreatic kallikrein cDNA probe detected 38 cross-hybridizing recombinant plasmids. Nucleotide sequence analysis of a subset of these revealed the presence of four different, closely related mRNAs that encode kallikrein-like enzymes. The most abundant of these mRNAs, designated PS, is identical with the kallikrein mRNA expressed in the pancreas (Swift et al., 1982). The cloned nucleotide sequence of a second mRNA (S1) extends from the codon specifying amino acid 80 of a kallikrein-like protein to the 3' end of the mRNA; the PS and S1 proteins share 86% amino acid sequence similarity in their overlap. S2 and S3 kallikrein-related mRNAs encode preproenzymes that are identical at 74% and 76%, respectively, of the amino acid positions when compared to rat PS kallikrein. The generally high level of amino acid sequence homology shared by these four submaxillary gland serine proteases extends to the other members of the multigene family of kallikrein-like enzymes having a cleavage preference for arginyl residues (Figure 8).

The PS mRNA encodes the rat submaxillary gland kallikrein described by Lazure et al. (1981), which has the substrate cleavage specificity of a true kallikrein (Gutkowska et al., 1983; G. Thibault, personal communication). Because the S1 protein is identical with PS kallikrein at 86% of its amino acid sequence and retains key amino acid residues thought to be principal determinants of peptide cleavage specificity, it may represent a second true kallikrein.

The S2 mRNA encodes the rat submaxillary enzyme tonin. The reported amino acid sequence of tonin (Lazure et al., 1984) and the predicted amino acid sequence of the S2 protein (aside from the pre- and propeptides) are identical, with the exception of a 16-residue deletion in the tonin sequence that includes the aspartate of the catalytic triad (Lazure et al.,

1984). Recently, the presence of the 16 amino acids predicted from the mRNA sequence has been confirmed by protein sequencing techniques (C. Lazure, personal communication). Therefore, tonin appears not to be an unusual, short kallikrein but rather a representative member of the family with a normal polypeptide length and with the aspartate of the catalytic triad in the appropriate position. The S3 enzyme is identical with tonin at 84% of its amino acid sequence, retains the same amino acid substitutions relative to PS kallikrein at key residues thought to determine kallikrein cleavage specificity, and therefore may represent a second tonin.

mRNA and Protein Domains. (A) 5'- and 3'-Untranslated Regions of mRNA. The length of the 5'-untranslated region of the PS mRNA is estimated to be between 24 and 42 nucleotides in length. The length of this untranslated region depends upon the choice of 2 potential transcription start sites 6 nucleotides apart and upon the choice of 2 methionine initiator codons 12 nucleotides apart (see Figure 6). The 5'-untranslated regions for the cloned S2 and S3 mRNAs are 35 and 36 nucleotides, respectively. The nucleotide sequences of these three 5'-untranslated regions are strikingly related; this region of the PS mRNA shares 75% and 72% homology with the same regions of S2 and S3, respectively, and the 5'-untranslated region of S2 is 81% homologous to that of S3. Each 5'-untranslated region contains a short sequence complementary to a conserved sequence near the 3' end of 18S and rRNA (Hagenbuchle et al., 1978). Base pairing of oligonucleotide sequences present in the 5'-untranslated region of some eukaryotic mRNAs has been suggested to promote mRNA selection by ribosomes (Hagenbuchle et al., 1978; Sargan et al., 1982; MacDonald et al., 1982a).

The 3'-untranslated regions of the four submaxillary gland mRNAs are approximately 90% homologous to one another. This is somewhat surprising since noncoding regions of related mRNAs are often quite divergent (MacDonald et al., 1982a,b). This level of conservation is slightly greater than the average of the protein coding domains (80%) and suggests either an important function for the 3'-noncoding sequences or recent gene conversion events that did not discriminate between amino acid coding and noncoding sequences.

(B) Prepeptides and Activation Peptides. Alignment of the amino acid sequences derived from the PS, S2, and S3 mRNA sequences with the amino acid sequences of the mature porcine pancreatic kallikrein (Tschesche et al., 1979) and rat submaxillary gland kallikrein (Lazure et al., 1981) indicates that pre- and propeptide regions of 28 or 24 residues for PS and 24 residues for S2 and S3 are encoded by these kallikrein-related mRNAs (Figure 7). The boundary between the signal peptide and the activation peptide of the kallikrein zymogens cannot be determined unambiguously from the sequence data. Signal peptidase cleavage sites have been predicted for several kallikrein family members, using only amino acid sequence information derived from cloned mRNA sequences (see Figure 7).

```
-28          -24            -20            -15              -10            -5            -1  +1
(Met Pro Val Thr)Met Trp Phe Leu Ile Leu Phe Leu Ala Leu Ser Leu Gly Arg Asn Asp Ala Ala Pro Pro Val Gln Ser Arg Val ... Rat PS type

                 Met Trp Leu Gln Ile Leu Ser Leu Val Leu Ser Val Gly Arg Ile Asp Ala Ala Pro Pro Gly Gln Ser Arg Ile ... Rat S2 type

                 Met Trp Phe Leu Ile Leu Phe Leu Ala Leu Ser Leu Gly Gln Ile Asp Ala Ala Pro Pro Gly Gln Ser Arg Val ... Rat S3 type

                 Met Trp Phe Leu Ile Leu Phe Leu Ala Leu Ser Leu Gly Gly Ile Asp Ala Ala Pro Pro Val Gln Ser Arg Ile ... Mouse mGK-1

                 Met Trp Phe Leu Ile Leu Phe Pro Ala Leu Ser Leu Gly Gly Ile Asp Ala Ala Pro Pro Leu Gln Ser Arg Val ... Mouse EGF-BP

             Met Ser Ala Leu Leu Ile Leu Ala Leu Val Gly Ala Ala Val Ala Phe Pro Leu Glu Asp Asp Asp Lys Ile ... Rat Trypsin I
```

FIGURE 7: Comparison of the pre- and propeptides of murine kallikrein-like enzymes and trypsin. Open *arrows* indicate the position of prepeptide cleavages proposed previously (see text); closed *arrows* indicate the cleavage positions proposed in this study. Sequence sources: mGK-1, Mason et al., 1983; mouse EGF-BP, Ronne et al., 1983; rat trypsin I, MacDonald et al., 1982a.
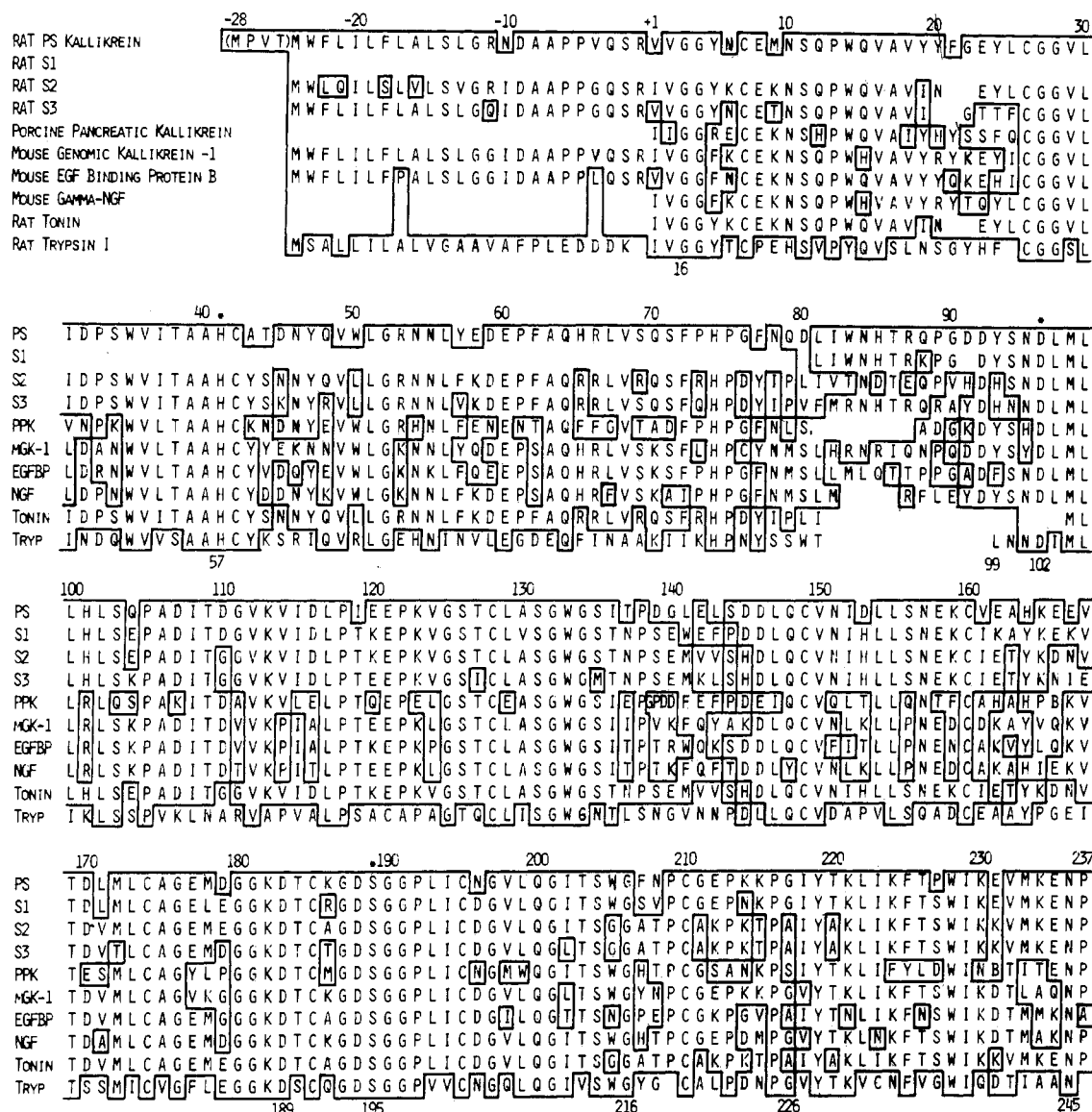
FIGURE 8: Amino acid sequence comparison of kallikrein-like enzymes and trypsin. The numbering scheme starts at the amino terminus of the predicted active enzyme form of rat kallikrein. Chymotrypsinogen numbering is included for key amino acid positions to facilitate comparison with other serine proteases. The boxes delineate the amino acids shared by the greatest number of proteins at each position. Small gaps are introduced to optimize sequence alignments. The amino acid sequences for porcine pancreatic kallikrein (PPK; Tschesche et al., 1979), the γ subunit of mouse nerve growth factor (γ-NGF: Thomas et al., 1981), mouse epidermal growth factor binding protein (EGF-BP; Lundgren et al., 1984), rat tonin (Lazure et al., 1984), and rat trypsin (MacDonald et al., 1982) are included for comparison. The asterisks denote the positions of the histidine, aspartate, and serine of the catalytic triad.

The prepropeptide sequence is highly conserved among these proteins and suggests that the signal peptide cleavage occurs at the same site in each. The cleavage site consensus sequence derived by Perlman & Halvorson (1983) predicts two possible prepeptide cleavage sites in the kallikrein sequences. The first occurs after the glycine at -12, which is present in all five kallikrein-related enzymes shown (Figure 7). However, for S2, S3, mGK-1, and the EGF binding protein, this cleavage would yield a prepeptide of only 13 amino acids, shorter than any yet reported. The second predicted cleavage site occurs after the alanine at -8, also conserved among all five kallikrein-like enzymes compared (Figure 7). As Ronne et al. (1983) noted, this cleavage site would generate an amino-terminal sequence (Ala-Pro-Pro-Xaa-Gln-Ser-Arg) of the zymogens nearly identical with that of mouse α-NGF (Ala-Pro-Pro-Val-Glu-Ser-Arg), a kallikrein-related protein without protease activity (Ronne et al., 1984). While this prepeptide cleavage point yields an acceptable precleavage sequence (Ile-Asp-Ala) for the S2, S3, EGF-BP, and mGK-1 prepro-

enzymes, it does not (Asn-Asp-Ala) for PS kallikrein. If the four additional N-terminal amino acids encoded by the PS mRNA are expressed, they may compensate for a foreshortened signal peptide generated by cleavage between -12 and -11. These proposed endopeptidase cleavage sites for the preprokallikreins (Figure 7) would define signal peptides which meet minimum length requirements and would leave a short activation peptide as predicted for prokallikreins (Schachter, 1980).

(C) Enzymes. The four submaxillary gland serine proteases encoded by the cloned mRNA sequences described here represent a family of enzymes sharing 74–86% amino acid homology. Amino acid residues histidine-41, aspartate-96, and serine-189 of the serine protease catalytic triad and their contiguous sequences are conserved. All four proteins retain an aspartate residue at sequence position 183 (the bottom of the substrate binding pocket), a characteristic of proteases with trypsin-like cleavage preference. The presence of glycine or alanine residues at positions 206 and 217 in all four enzymes

is consistent with an open binding pocket to accommodate bulky amino acid side chains. X-ray crystallographic analysis (Bode et al., 1983) has revealed that the kallikrein specificity pocket is larger than that of trypsin due to the insertion of an additional residue at 209; this proline residue is conserved in PS, S1, S2/tonin, and S3, as well as all other kallikrein family members (see Figure 8). In porcine kallikrein, the side chain of serine-217, which affects the geometry of the P1 binding pocket and partly shields the P1 side chain from the negative charge of aspartate-183, is absent in the four rat enzymes. The replacement of glycine for serine at position 217 in rat PS kallikrein demonstrates that the presence of the serine near the bottom of the binding pocket is not necessary for kallikrein specificity. The presence of glycine at this position may accommodate benzamidine binding (Bode et al., 1983) and explain the greater affinity of rat kallikrein (Mares-Guia & Diniz, 1967) than porcine kallikrein (Geratz et al., 1979) for this serine protease inhibitor.

Chen & Bode (1983) have proposed that a principal determinant of the specific cleavage of the kininogen precursor by kallikrein is the presence of a hydrophobic sandwich formed by tyrosine-93 and tryptophan-205 that facilitates the binding of bulky, hydrophobic residues in the substrate position adjacent to the cleavage site (the P2 position). This may explain in part the selective cleavage by tissue kallikrein at the arginine following a phenylalanine and at the methionine following a leucine to release lysylbradykinin from the kininogen precursor. Tyrosine-93 and tryptophan-205, present in PS kallikrein, are retained in the S1 enzyme; together with the overall sequence conservation between PS kallikrein and S1, this suggests that the S1 enzyme may retain kallikrein cleavage specificity as well. Tyrosine-93 and tryptophan-205 are replaced by histidine-93 and glycine-205 is the S2/tonin and S3 enzymes and may contribute toward the altered substrate specificity of tonin, and the S3 enzyme as well.

Extensive polypeptide regions are highly conserved among all four of the encoded preproenzymes while other regions diverge to a great extent. Upon consideration of the structure of porcine pancreatic kallikrein (Bode et al., 1983), it is evident that the clusters of sequence changes are localized to external loops of the protein and would not be expected to jeopardize enzymatic function. The first of these variable regions comprises amino acid residues 19–25. The small amino acid deletions in S2 and S3 occur in this region (Figure 8). The kallikrein autolysis loop, which includes residues 79–91, is the site of the second variable region. In the single-chain form of the kallikrein-like enzymes, this loop contains 11 more residues than does trypsin (an exception is γ-NGF with only 7 additional residues). The third variable region (amino acid residues 139–145) corresponds to an external loop that interacts with threonine-208 and proline-209 (Bode et al., 1983). These interactions draw the loop partly over the binding pocket and may affect the specificity of substrate and inhibitor binding (Bode et al., 1983; Chen & Bode, 1983). Therefore, the large variations in this region among the kallikrein-like enzymes (Figure 8) may modulate substrate selection.

The kallikrein subfamily of serine proteases has been implicated in the processing of a number of polypeptide hormones. Several of these proteases are present in the murine submaxillary gland. The kallikrein-like enzymes encoded by two of the mRNAs described herein encode submaxillary kallikrein and tonin, known members of the kallikrein family that cleave kininogen and angiotensinogen, respectively, to release bioactive peptides. The two enzymes S1 and S3, encoded by the other two mRNAs, are closely related to kalli-

krein and tonin, respectively. Although these enzymes may have redundant functions to kallikrein and tonin, it is possible that each has a unique function to process as yet unknown precursors. The characterization of the cleavage specificity of the two new enzymes awaits purification of the enzymes either from rat submaxillary glands or from alternative hosts in which the cloned cDNAs could be expressed.

REFERENCES

Alhenc-Gelas, F., Marchetti, J., Allegrini, J., Corvol, P., & Menard, J. (1981) *Biochim. Biophys. Acta 677*, 477–488.

Ashley, P. L., & MacDonald, R. J. (1984) *Anal. Biochem. 140*, 95–103.

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

Barka, T. (1980) *J. Histochem. Cytochem. 28*, 836–859.

Birnboim, H. C., & Doly, J. (1979) *Nucleic Acids Res. 7*, 1513–1523.

Bode, W., Chen, Z., Bartels, K., Kutzbach, C., Schmidt-Kastner, G., & Bartunik, H. (1983) *J. Mol. Biol. 164*, 237–282.

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

Boucher, R., Asselin, J., & Genest, J. (1974) *Circ. Res. 34 (Suppl. 1)*, 203–209.

Brandtzaeg, P., Gautvik, K. M., Nustad, K., & Pierce, J. V. (1976) *Br. J. Pharmacol. 56*, 155–167.

Carretero, O. A., & Scicli, A. G. (1981) *Hypertension (Dallas) 3 (Suppl. 1)*, 4–12.

Chen, Z., & Bode, W. (1983) *J. Mol. Biol. 164*, 283–311.

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

Denhardt, D. T. (1966) *Biochem. Biophys. Res. Commun. 23*, 641–646.

Drouin, J. (1980) *J. Mol. Biol. 140*, 15–34.

Fiedler, F. (1979) *Handb. Exp. Pharmacol. 25*, 103–161.

Fiedler, F., & Fritz, H. (1981) *Hoppe-Seyler's Z. Physiol. Chem. 362*, 1171–1175.

Frey, E. K., Kraut, H., Werle, E., Vogel, R., Zickgraf-Rudel, G., & Trautschold, I. (1968) *Das Kallikrein-Kinin-System und Seine Inhibitoren*, Enke Verlag, Stuttgart, FRG.

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

Geratz, J. D., Stevens, F. M., Polakosi, K. K., Parrish, R. F., & Tidwell, R. R. (1979) *Arch. Biochem. Biophys. 197*, 551–559.

Gutkowska, J., Thibault, G., Cantin, M., Garcia, R., & Genest, J. (1983) *Can. J. Physiol. Pharmacol. 61*, 449–465.

Hagenbuchle, O., Santer, M., Steitz, J. A., & Mans, R. J. (1978) *Cell (Cambridge, Mass.) 13*, 551–563.

Hanahan, D., & Meselson, M. (1980) *Gene 10*, 63–67.

Hilton, S. M. (1970) *Handb. Exp. Pharmacol. 15*, 389–399.

Hilton, S. M., & Jones, M. (1968) *J. Physiol. (London) 195*, 521–533.

Lazure, C., Seidah, N. G., Thibault, G., Genest, J., & Chretien, M. (1981) in *Proceedings of the VIIth American Peptide Symposium* (Rich, D. H., & Gross, M. R., Eds.) pp 517–520, Pierce Chemical Co., Rockford, IL.

Lazure, C., Leduc, R., Seidah, N. G., Thibault, G., Genest, J., & Chretien, M. (1984) *Nature (London) 307*, 555–558.

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

MacDonald, R. J., Stary, S. J., & Swift, G. H. (1982a) *J. Biol. Chem. 257*, 9724–9732.

MacDonald, R. J., Swift, G. H., Quinto, C., Swain, W., Pictet, R. L., Nikovits, W., & Rutter, W. J. (1982b) *Biochemistry 21*, 1453–1463.

Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning*, p 93, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Mares-Guia, M., & Diniz, C. R. (1967) *Arch. Biochem. Biophys. 121*, 750–756.

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

Maxam, A., & Gilbert, W. (1980) *Methods Enzymol. 65*, 499–560.

Movat, H. Z. (1979) *Handb. Exp. Pharmacol. 25*, 1–89.

Okayama, H., & Berg, P. (1982) *Mol. Cell. Biol. 2*, 161–170.

Perlman, D., & Halvorson, H. O. (1983) *J. Mol. Biol. 167*, 391–409.

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

Rigby, P. J. W., Dieckman, M., Rhodes, C., & Berg, P. (1977) *J. Mol. Biol. 113*, 237–251.

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

Ronne, H., Anundi, H., Rask, L., & Peterson, P. A. (1984) *Biochemistry 23*, 1229–1234.

Sargan, D. R., Gregory, S. P., & Butterworth, P. H. W. (1982) *FEBS Lett. 147*, 133–136.

Schachter, M. (1980) *Pharmacol. Rev. 31*, 1–17.

Schechter, I., & Berger, A. (1967) *Biochem. Biophys. Res. Commun. 27*, 157–162.

Seki, T., Nakajima, T., & Erdös, E. G. (1972) *Biochem. Pharmacol. 21*, 1227–1235.

Swift, G. H., Dagorn, J.-C., Ashley, P. L., Cummings, S. W., & MacDonald, R. J. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 7263–7267.

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

Tschesche, H., Mair, G., Godec, G., Fiedler, F., Ehret, W., Hirschauer, C., Lemon, M., & Fritz, H. (1979) *Adv. Exp. Med. Biol. 120*, 245–260.

Yamada, K., & Erdös, E. G. (1982) *Kidney Int. 22*, 331–337.

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.

Zassenhaus, H. P., Butow, R. A., & Hannon, Y. P. (1982) *Anal. Biochem. 125*, 125–130.

Tissue-Specific Expression of Kallikrein-Related Genes in the Rat[†]

Patricia L. Ashley and Raymond J. MacDonald*

Division of Molecular Biology, Department of Biochemistry, The University of Texas Health Science Center at Dallas, Dallas, Texas 75235

Received December 6, 1984

ABSTRACT: Four distinct kallikrein-related mRNAs (PS, S1, S2, and S3), encoded by members of a multigene family, are selectively expressed in various combinations in several rat tissues. Although closely related along most of the mRNA sequence, the four mRNAs can be selectively detected with synthetic oligonucleotide probes complementary to highly variable mRNA subregions. PS mRNA, which encodes an enzyme with true kallikrein activity, is present at high levels in the submaxillary gland, pancreas, and kidney. S1 mRNA, which encodes an enzyme similar to the PS kallikrein, is detected only in the submaxillary gland and is present at one-fifth the PS mRNA level. S2 mRNA, which encodes the enzyme tonin, is present in the submaxillary gland at half the PS mRNA level and at a slightly higher level in the prostate. S3 mRNA, which encodes an enzyme very similar to tonin, is present in the submaxillary gland at one-tenth the PS mRNA level and in the prostate at about the same level as tonin mRNA.

The kallikrein-like enzymes are homologous serine proteases encoded by a multigene family. True kallikreins (EC 3.4.21.8) selectively cleave the protein kininogen, found in the blood and in the interstitium of most tissues, to release the vasodilatory peptides bradykinin or lysylbradykinin [reviewed by Fiedler (1979)]. Other members of the kallikrein-like enzyme family cleave other prohormones: tonin cleaves angiotensinogen (Boucher et al., 1974), the γ subunit of nerve growth factor processes the precursor of nerve growth factor (Thomas et al., 1981), and the epidermal growth factor binding protein processes the precursor of epidermal growth factor (Frey et al., 1979). The kallikrein-like enzymes most likely function within the organ in which they are synthesized, and their presence