Schuldiner, S., & Avron, M. (1971) *FEBS Lett. 14*, 233–236.
Schuldiner, S., Rottenberg, H., & Avron, M. (1972) *Eur. J. Biochem. 25*, 64–70.
Searle, G. F. W., & Barber, J. (1978) *Biochim. Biophys. Acta 302*, 309–320.
Senior, A. E., & Brooks, J. C. (1970) *Arch. Biochem. Biophys. 140*, 257–266.
Shafer, J., Baronowsky, P., Laursen, R., Finn, F., & Westheimer, F. H. (1966) *J. Biol. Chem. 241*, 421–427.
Singh, A., Thorton, E. R., & Westheimer, F. H. (1962) *J. Am. Chem. Soc. 237*, PC3006.
Treinin, A. (1971) in *The Chemistry of the Azido Group* (Patai, S., Ed.) pp 1–56, Interscience, New York.
Vanderkooi, J. M., & Chance, B. (1972) *FEBS Lett. 22*, 23–26.
Weber, W., & Osborn, M. (1969) *J. Biol. Chem. 244*, 4406–4412.
Wentrup, C., & Crow, W. D. (1970) *Tetrahedron 26*, 3965.

Primary Structure of Two Distinct Rat Pancreatic Preproelastases Determined by Sequence Analysis of the Complete Cloned Messenger Ribonucleic Acid Sequences†

Raymond J. MacDonald,* Galvin H. Swift,‡ Carmen Quinto,§ William Swain, Raymond L. Pictet,⊥ William Nikovits, and William J. Rutter

ABSTRACT: The mRNA sequences for two rat pancreatic elastolytic enzymes have been cloned by recombinant DNA technology and their nucleotide sequences determined. Rat elastase I mRNA is 1113 nucleotides in length, plus a poly(A) tail, and encodes a preproelastase of 266 amino acids. The amino acid sequence of the predicted active form of rat elastase I is 84% homologous to porcine elastase 1. Key amino acid residues involved in determining substrate specificity of porcine elastase 1 are retained in the rat enzyme. The activation peptide of the zymogen does not appear related to that of other mammalian pancreatic serine proteases. The mRNA for elastase I is localized in the rough endoplasmic reticulum of acinar cells, as expected for the site of synthesis of an exocrine secretory enzyme. Rat elastase II mRNA is 910 nucleotides in length, plus a poly(A) tail, and encodes a preproenzyme of 271 amino acids. The amino acid sequence is more closely related to porcine elastase 1 (58% sequence identity) than to the other pancreatic serine proteases (33–39% sequence identity). Predictions of substrate preference based upon key amino acid residues that define the substrate binding cleft are consistent with the broad specificity observed for mammalian pancreatic elastase 2. The activation peptide is similar to that of the chymotrypsinogens and retains an N-terminal cysteine available to form a disulfide link to an internal conserved cysteine residue.

The mammalian exocrine pancreas synthesizes, stores, and secretes approximately 15 enzymes and proenzymes for intestinal digestion. These secretory proteins account for greater than 80% of the protein synthesis of the gland (Jamieson & Palade, 1967; VanNest et al., 1980). A dominant fraction of the pancreatic secretions is a family of at least seven serine proteases: chymotrypsins A and B, two trypsins, elastases 1 and 2,[1] and kallikrein. The members of this gene family are probably related by evolution from a common ancestral protease (Neurath et al., 1967; deHaen et al., 1975) and have retained similar structure, size, and function. As secretory proteins, the serine proteases are expected to be synthesized initially as precursor proteins (Devillers-Thiery et al., 1975; Rutter et al., 1978) with an amino-terminal signal peptide (Blobel & Sabatini, 1971) that specifies the binding of polysomes to the rough endoplasmic reticulum and the subsequent vectorial transport into the cisternae of the reticulum and proteolytic processing of the precursors as the first steps in secretion. In addition, the pancreatic serine proteases are secreted as proenzymes (zymogens) that require selective enzymatic cleavage of an amino terminal "activation peptide" for conversion to the active enzymes. Analysis of the organization of the pancreatic serine protease genes and mRNAs is expected to contribute to an understanding of events in the evolution of this gene family and of the processes involved in the biosynthesis and maturation of the enzymes.

Two members of the pancreatic family of serine proteases are characterized by their ability to hydrolyze elastin. Elastase 1 action is largely limited to Ala–Ala and Ala–Gly bonds (Gertler et al., 1977). Determination of the three-dimensional structure of elastase 1 revealed that its cleavage preference can be explained in part by an occluded substrate binding pocket, relative to other serine proteases, that makes the binding of large amino acid side chains sterically impossible (Shotton & Watson, 1970). In contrast, elastase 2 has a preference for cleavage at amino acids with medium to large

† From the Division of Molecular Biology of the Department of Biochemistry (R.J.M. and G.H.S.), the University of Texas Health Science Center at Dallas, Dallas, Texas 75235, and the Department of Biochemistry and Biophysics (R.J.M., C.Q., W.S., R.L.P., W.N., and W.J.R.), The University of California, San Francisco, California 94143. *Received September 9, 1981.* This work was supported by the National Science Foundation (PCM 8006231 to R.J.M.) and the National Institutes of Health (AM21344 to W.J.R.).

‡ Virginia Lazenby O'Hara fellow.
§ Present address: Centro de Figacion de Nitrogeno, Cuernavaca, Morelos, Mexico.
⊥ Present address: Institute de Biologie Moleculaire, Universite Paris VII, Tour 43, 75221 Paris Cedex 05, France.

[1] We use Roman numerals to identify the rat pancreatic elastase mRNAs and the enzymes they encode to indicate that the enzyme activities have not been demonstrated directly. We retain the use of Arabic numerals in the discussion of the purified and well-characterized pancreatic elastases of pig and human.

0006-2960/82/0421-1453$01.25/0

hydrophobic side chains (Gertler et al., 1977; Del Mar et al., 1980). In this respect, elastase 2 more closely resembles the chymotrypsins. However, neither the structural basis of elastase 2 specificity nor its structural relationships with elastase 1 and chymotrypsin are known.

We report here the molecular cloning of the mRNA sequences for the two rat pancreatic elastases, the nucleotide sequences of the complete mRNAs, and the primary structure of the preproenzymes. Elastase I mRNA encodes an enzyme that represents the rat counterpart of the principal elastase isolated from porcine and human pancreas. Elastase II mRNA encodes an enzyme closely related to elastase 1 but, as judged from key amino acid substitutions, would have an altered and characteristic substrate specificity similar to that described for porcine and human elastase 2.

Experimental Procedures

Cloning Pancreatic mRNA Sequences as Double-Stranded Complementary DNA (ds-cDNA).[2] Total rat pancreatic polyadenylated RNA without enrichment for specific mRNA sequences was used for the construction of a ds-cDNA library. Details of the preparation of the library will be described elsewhere (W. Swain et al., unpublished results). The library comprised approximately 1000 recombinant clones obtained by inserting pancreatic ds-cDNA into linear pBR322 after the addition of either synthetic *Hin*dIII recognition site decamers (Goodman & MacDonald, 1980) or homopolymeric tails (Roychoudhury et al., 1976). All experiments with bacteria containing recombinant plasmids were conducted according to the NIH Guidelines for Recombinant DNA Research.

Preparation and Analysis of Pancreatic RNA. RNA was prepared from rat pancreas by the guanine thiocyanate procedure of Chirgwin et al. (1979). Polyadenylated RNA was isolated from total pancreatic RNA by binding twice to oligo(dT)–cellulose (Collaborative Research, Inc.; type 2) (Aviv & Leder, 1972). The RNA was dissolved in water, heated to 68 °C for 3 min, and diluted with an equal volume of twice-concentrated binding buffer immediately before each passage over oligo(dT)–cellulose to enhance the removal of rRNA.

Polyadenylated RNA was analyzed by electrophoresis in 1.5% agarose gels containing methyl mercuric hydroxide (Bailey & Davidson, 1976). Electrophoresis was for 17 h at 20 mA in 0.3 × 13 × 11 cm slab gels. RNA size standards were MS2 RNA (Miles Laboratories), rat 28S and 18S rRNAs, yeast 25S and 5S rRNAs (gifts from Dr. F. Masiarz), and *Escherichia coli* 23S and 16S rRNA. Linear DNA standards were obtained by digestion of bacteriophage λ DNA with *Hin*dIII and φX174 RF DNA with *Hae*III. After electrophoresis and staining with ethidium bromide, the RNA was transferred to DBM paper and hybridized to recombinant plasmid DNA labeled with ^{32}P by nick translation (Rigby et al., 1977) as described by Alwine et al. (1977).

Hybridization of [^{32}P]cDNA to Nitrocellulose-Bound Recombinant Plasmid DNAs. Individual recombinant plasmid DNAs were digested with *Bam*H1 endonuclease to generate linear molecules and subjected to electrophoresis in 1% agarose slab gels as described by Helling et al. (1974). DNAs were transferred to nitrocellulose (Schleicher & Schüll) according to the method of Southern (1975). Prehybridization and hybridization were at 68 °C in 5 × SSC, 0.1 M sodium phosphate, pH 6.8, 100 μg/mL *E. coli* DNA, and 10 × Denhardt's solution (Denhardt, 1966). The hybridization solution contained, in addition, 5 × 10^6 cpm of [^{32}P]cDNA in 10 mL. After hybridization for 1 day, the nitrocellulose sheet was washed twice in 0.1 × SSC with 0.1% $NaDodSO_4$ at 53 °C for 0.5 h and then twice in 0.1 × SSC at 53 °C for 0.5 h prior to autoradiography with No-Screen X-ray film.

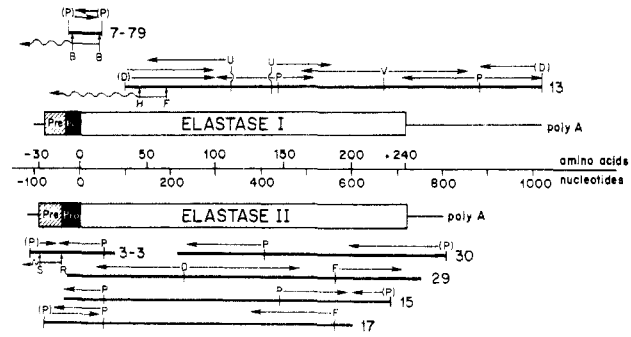

FIGURE 1: Sequencing strategies for rat pancreatic elastase I and elastase II mRNAs. The horizontal rectangles delineate the amino acid coding regions; the regions coding the "signal peptide" (Pre) and the activation peptides (Pro) are indicated. The horizontal lines extending from the rectangles represent the length of the noncoding regions of the mRNAs; a length of poly(A) is present at the 3′ end in addition to these noncoding regions. The thick horizontal lines represent the extent of each ds-cDNA insert; the numbers are the identification numbers for each recombinant plasmid. The direction and length of each sequencing run are indicated by the horizontal arrows, each starting at the position of the restriction endonuclease site noted (B, *Bst*N1; P, *Pst*I; U, *Pvu*II; D, *Hin*dIII; V, *Ava*I; H, *Hph*I; F, *Hin*fI; S, *Sau*3A; R, *Rsa*I). The *Hin*dIII and *Pst*I sites in parentheses were generated upon cloning the ds-cDNA by the addition of *Hin*dIII linkers and G–C homopolymeric tailing, respectively. The wavy horizontal lines delineate the position and length of sequences determined by "primer extension" analyses.

Preparation of Plasmid DNAs. Small amounts of recombinant plasmid DNAs bearing pancreatic ds-cDNA inserts were prepared according to the procedure described by Meagher et al. (1977) but with the addition of a final precipitation of the DNA from 0.05 M magnesium acetate and 0.25 M lithium acetate, pH 5, with 3 volumes of ethanol. Large plasmid DNA preparations were performed as described by Holland & McCarthy (1980).

DNA Sequence Analysis. Nucleotide sequence determinations were performed according to Maxam & Gilbert (1977), as described in detail (Maxam & Gilbert, 1980). Five sequencing reactions were always performed (G, dimethyl sulfate; G + A, formic acid; C + T, hydrazine; C, hydrazine plus NaCl; A > C, sodium hydroxide) to enhance sequence accuracy. DNA fragments were labeled at their 5′ ends with [γ-^{32}P]ATP and polynucleotide kinase (NEN or Boehringer) after treatment with calf intestine alkaline phosphatase (Boehringer). Labeled fragments were prepared by electrophoresis and eluted either by shaking for several hours in 0.3 M NaOAc, pH 7.0, for fragments shorter than 250 base pair (bp) or by electroelution in 0.05 M Tris, 0.05 M sodium borate, and 0.001 M EDTA, pH 8.3, for longer fragments.

Elastase I Sequencing. The scheme for determining the full length sequence of elastase I mRNA is summarized in Figure 1. The insert from recombinant plasmid pcXP13 was completely sequenced. The single available open reading frame indicated that the cloned sequence began with the codon specifying amino acid 33 of the active enzyme (nucleotide 202 of the mRNA) and extended to the 3′ end of the mRNA, including 15 residues of the poly(A) tail. To obtain a hybridization probe to screen for cloned sequences extending

[2] Abbreviations: ds-cDNA, double-stranded complementary DNA; cDNA, complementary DNA; 1 × SSC, 0.15 M sodium chloride with 0.015 M sodium citrate; $NaDodSO_4$, sodium dodecyl sulfate; Tris, tris(hydroxymethyl)aminomethane; EDTA, ethylenediaminetetraacetic acid; bp, base pairs.

further toward the 5′ end of the mRNA, a 60-bp *Hph*I–*Hin*fI fragment (nucleotides 234–294 of the mRNA sequence) of pcXP13 was prepared, and, after hybridization with total pancreatic polyadenylated RNA, was extended (see below) with reverse transcriptase in the presence of [α-^{32}P]dCTP (NEN) and unlabeled dTTP, dGTP, and dATP. After removal of the RNA by base hydrolysis, the ^{32}P-labeled extended primer was used to screen the pancreatic ds-cDNA library of approximately 1000 colonies grown on Whatman 540 paper according to the method of Craig et al. (1979). A single clone (pcXP7-79) that extended beyond the 5′ limit of pcXP13 was obtained, was sequenced completely, was shown to encode only the region between nucleotide 73 and 152 (indicated in Figure 1), and therefore did not overlap with the sequences in pcXP13. The sequence of the mRNA between the two cloned regions was obtained by 5′ end labeling the *Hph*I–*Hin*fI primer, extending the primer with reverse transcriptase in the presence of unlabeled deoxynucleotide triphosphates, and sequencing the extended primer after its purification by polyacrylamide gel electrophoresis. The sequence of the extended primer also confirmed the entire sequence of pcXP7-79. Similarly, the sequence of the 5′ end of elastase I mRNA was determined by extending a BstNI (nucleotides 82–141) primer fragment (prepared from pcXP7-79; see Figure 1) with reverse transcriptase and total pancreatic polyadenylated RNA as template.

Elastase II Sequencing. The scheme for determining the nucleotide sequence of elastase II mRNA from the ds-cDNA clones also is shown in Figure 1. As described under Results, pcXP29 was initially identified as elastase-like by preliminary sequence analysis. The large *Pst*I fragment of pcXP29 was isolated, labeled with ^{32}P by nick translation, and used to screen the ds-cDNA library for overlapping cloned sequences. The set of overlapping ds-cDNA inserts that were obtained and sequenced is shown in Figure 1. pcXP30 contained a poly(A) tail of 11 nucleotides and therefore represented a 3′ end of the mRNA. pcXP3-3, representing the 5′ end of the mRNA, was identified by screening the library with the 5′-most *Pst*I fragment (relative to the mRNA) of pcXP17. Although five overlapping ds-cDNA inserts were analyzed to obtain the complete mRNA sequence, no conflicting nucleotide assignments were obtained, indicating that each represented the same mRNA. The 5′ end of elastase II mRNA was determined by sequence analysis of a cDNA obtained by extending a 5′-end-labeled *Sau*3A–*Rsa*I fragment (nucleotides 28–77 of the mRNA). In contrast to other reports (e.g., Richards et al., 1979; Sippel et al., 1978), no anomalous sequences produced by cloning artifacts were observed at the 5′ ends of the ds-cDNA clones, including that of pcXP3-3, which has all but an estimated two nucleotides of the mRNA 5′ end.

Primer Extension. Primer extension by cDNA synthesis using restriction enzyme generated fragments of cloned ds-cDNA as primers and total pancreatic polyadenylated RNA as template was performed essentially as described by Hagenbuchle et al. (1980). Following treatment with calf intestine alkaline phosphatase, primer fragments were labeled at their 5′ ends with polynucleotide kinase and [γ-^{32}P]ATP (Maxam & Gilbert, 1980). The labeled fragments were purified by electrophoresis in polyacrylamide gels, eluted from the gel by shaking, concentrated by repeated extraction with 1-butanol, and precipitated with ethanol. Five to six picomoles of ^{32}P-labeled primer were hybridized with 30 μg of pancreatic polyadenylated RNA in 50 μL of 81% formamide, 0.4 M NaCl, 0.001 M EDTA, and 0.01 M Pipes, pH 6.4, for 60 min at 50 °C. With the assumption that either elastase mRNA accounted for approximately 5% of the total mRNA population, the R_0t of this hybridization (0.7 m s L^{-1}) exceeded the calculated $R_0t_{1/2}$ of the elastase mRNAs by 60-fold. The hybridization reaction was terminated by adding 125 μL of ethanol. The precipitate was washed once with cold 70% ethanol containing 20 mM Tris-HCl, pH 8.4, 10 mM NaCl, and 6 mM $MgCl_2$. The entire ethanol precipitate was included in a 40-μL reaction mix containing 20 mM Tris-HCl, pH 8.4, 10 mM NaCl, 6 mM $MgCl_2$, 20 mM dithiothrietol, and 2 mM each of the four deoxynucleotide triphosphates. After the addition of 60 units of reverse transcriptase (Life Sciences, Inc.) and incubation for 1 h at 42 °C, the reaction products were precipitated with ethanol. The precipitate was dissolved in 20 μL of 0.1 M NaOH containing 1 mM EDTA, incubated for 10 min at 68 °C, and layered directly on a polyacrylamide gel in TBE buffer. The extended primer was located by autoradiography, eluted from the gel, and sequenced as described above. In these experiments 20–25% of the primer strand molecules were extended. The predominant product was the maximally extended primer; only very minor amounts of shorter cDNAs, presumably premature termination products, were observed.

In Situ Hybridization. In situ hybridization was performed with cryostat sections of pancreas and parotid as described previously (Harding et al., 1977) with the following modifications: 200 000 cpm of the ^{3}H-labeled ds-cDNA cloned insert labeled by nick translation (at 5×10^6 cpm/μg) in 6 μL of 0.1 M Tris-HCl, pH 8, 0.1 M NaCl, 0.01 M EDTA, and 50% formamide was pipetted onto the section and covered with a glass cover slip. The probe was denatured in situ by heating to 80 °C for 10 min prior to hybridization overnight at 40 °C. Tissue sections were stained with toluidine blue and methylene blue.

Results

Strategy for the Isolation of Cloned ds-cDNA Sequences for Elastase mRNAs. A library of ds-cDNA sequences cloned in pBR322 was constructed by using total rat pancreatic polyadenylated RNA as the source of mRNA sequences. Because the mRNAs for the exocrine secretory proteins have been shown to account for approximately 90% of the total pancreatic mRNA population (Harding et al., 1977; Przybyla et al., 1979; VanNest et al., 1980), it was assumed that the majority of the recombinant plasmids contained ds-cDNA inserts for secretory protein mRNAs. For tentative identification of those recombinant plasmids likely to bear sequences for elastase mRNAs, we took advantages of two characteristics of pancreatic serine proteases.

(1) On the basis of the data of VanNest et al. (1980), synthesis of the serine proteases in the rat pancreas accounts for about 40% of the total pancreatic protein synthesis. Consequently, the mRNAs for serine proteases should be very prominent within the total population of pancreatic mRNAs.

(2) The serine proteases comprise a unique molecular weight class (M_r 25 000–30 000) of the pancreatic secretory proteins. Furthermore, in vitro translation of pancreatic polyadenylated RNAs resolved according to size by electrophoresis under denaturing conditions (Rutter et al., 1978; Chirgwin et al., 1979) demonstrated that only mRNAs between 1.0 and 1.25 kb in length directed the synthesis of polypeptides of the size expected for serine protease translation products. Thus, recombinant plasmids bearing elastase mRNA sequences should be contained within a class that hybridizes to prominent mRNAs between 1.0 and 1.25 kb in length.

A subset of 31 recombinant plasmids from the library was analyzed to determine whether they contained ds-cDNA in-

Table I: Amino Acid Homologies between Rat Elastases I and II and Other Pancreatic Serine Proteases

	identity[a] (%) to						
	rat elastase I[b]	rat elastase II[c]	porcine elastase 1	porcine kallikrein[d]	bovine chymotrypsin A[e]	bovine chymotrypsin B[e]	bovine trypsin
rat elastase I[b]		58% (1)[f]	84% (0)	30% (12)	37% (8)	37% (8)	37% (13)
rat elastase II[c]	58% (1)		58% (1)	33% (13)	38% (9)	40% (9)	36% (13)

[a] Percent of the minimum length required for alignment. The placement of gaps for optimizing the alignment was taken from Woodbury et al. (1978) and Figure 9. [b] For residues 1–240 of the predicted active form of elastase I; numbered according to Figures 4 and 9. [c] For residues 1–240 only (numbered according to Figures 5 and 9), although the predicted active enzyme, by analogy to chymotrypsin, also probably contains residues −14 through −1. [d] Completed sequence for porcine pancreatic kallikrein is from Tschesche et al. (1978). [e] For the chymotrypsin residues 16–245 only. [f] The numbers in parentheses are the minimum number of gaps required for sequence alignment (Woodbury et al., 1978).

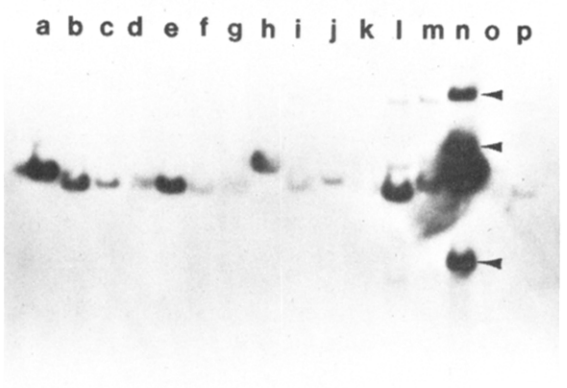

FIGURE 2: Identification of recombinant plasmids bearing ds-cDNA sequences for abundant pancreatic mRNAs. Linear recombinant plasmid DNAs were electrophoresed in a 1% agarose gel, transferred to nitrocellulose, hybridized with ^{32}P-labeled pancreatic cDNA (specific activity 2×10^8 cpm/μg), washed, and autoradiographed as described under Experimental Procedures. The arrrow heads indicate hybridization to nonlinearized forms of the plasmid DNAs that remain due to incomplete restriction endonuclease digestion. Lane h contains pcXP13 DNA discussed in the text.

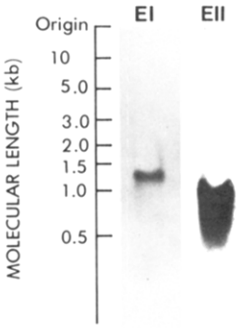

FIGURE 3: Estimates of the sizes of mRNAs encoded by recombinant plasmids pcXP13 (EI) and pcXP29 (EII). Pancreatic polyadenylated RNA was electrophoresed in a 1.5% agarose gel containing methyl mercuric hydroxide (Bailey & Davidson, 1976) and the position of mRNAs encoded by pcXP13 and pcXP29 determined by hybridization with the corresponding ^{32}P-labeled plasmid DNAs as described under Experimental Procedures. The two lanes were from a single gel. The scale for molecular length shown on the left was derived from the migration of several size standards described under Experimental Procedures.

serts for prominent mRNAs and whether those mRNAs fell into the correct size class. Equal amounts of recombinant plasmid DNAs were digested with *Bam*H1, electrophoresed in a 1% agarose gel, transferred to nitrocellulose, and hybridized to ^{32}P-labeled cDNA prepared from total pancreatic polyadenylated RNA by using reverse transcriptase with oligo(dT)$_{12-18}$ as primer. Under the conditions of hybridization chosen, the nitrocellulose-bound plasmid DNA was in local sequence excess over the small amount of the [^{32}P]cDNA probe. Therefore, to a first approximation, the intensity of the hybridization signal was directly related to the proportion of the complementary sequence in the mRNA population (Grunstein & Hogness, 1975; Biessmann et al., 1979). On the basis of the intensity of hybridization (Figure 2), recombinant plasmids a, b, e, h, l, and n were expected to contain sequences corresponding to the most prevalent pancreatic mRNAs including the serine protease mRNAs. [Subsequent analyses (W. Swain, C. Quinto, R. J. MacDonald, and W. Rutter, unpublished results; MacDonald et al., 1980) demonstrated that plasmids a, e, and n contained cloned sequences for the mRNA of amylase, the major rat pancreatic secretory protein.]

Figure 3 shows the results when ^{32}P-labeled pcXP13, chosen as containing sequences for a prominent mRNA (Figure 2, lane h), and another recombinant plasmid, pcXP29, were hybridized to pancreatic polyadenylated RNA that had been resolved by electrophoresis and transferred to DBM paper. pcXP13 hybridized to a 1.25-kb mRNA and pcXP29 to a 1.0-kb mRNA, both within the size range expected for serine protease mRNAs. Preliminary nucleotide sequence analysis of the ds-cDNA inserts of these two recombinant plasmids revealed that each contained an mRNA sequence for a different elastase-like serine protease (see below).

Complete Nucleotide Sequence of the Two Cloned Elastase mRNAs. The sequencing strategy for the two cloned elastase mRNAs is shown in Figure 1. To obtain the complete elastase I mRNA sequence, it was necessary to sequence a cDNA obtained by the use of reverse transcriptase to extend a short primer DNA derived from pcXP13 (see Experimental Procedures). To obtain the full length sequence of elastase II mRNA, it was necessary to sequence multiple, overlapping ds-cDNA fragments from independent recombinant plasmids that were identified by cross-hybridization to pcXP29 (see Experimental Procedures). In addition, the 5′-end sequences of both mRNAs were obtained by extension of short primers derived by restriction endonuclease digestions of cloned sequences near the 5′ end of each mRNA (see Experimental Procedures).

The sequence of elastase I mRNA and the derived amino acid sequence are shown in Figure 4. Elastase I mRNA contains a single coding frame uninterrupted by termination codons that prescribes a 266 amino acid protein with a molecular weight of 28 981. The relative spacing of amino acid residues histidine-45, aspartate-93, and serine-188 of the charge relay system and the conserved amino acids surrounding them characterize the encoded protein as a serine protease. A comparison of the amino acid sequence with that for other pancreatic serine proteases (Table I) shows the highest hom-

RAT PANCREATIC ELASTASE I mRNA

```
                                    -26                 -20                                 -10
            -UAGGAAGGCGU-   met leu arg phe leu val phe ala ser leu val leu tyr gly his ser thr gln asp
      GUGGUCUACUCUCUCCACAAC  AUG CUG CGC UUC CUG GUG UUC GCU UCC CUG GUC CUG UAU GGA CAC AGU ACC CAG GAC

                                    1                                   10
phe pro glu thr asn ala arg val val gly gly ala glu ala arg arg asn ser trp pro ser gln ile ser leu gln
UUU CCG GAA ACU AAU GCC CGC GUG GUU GGA GGG GCU GAA GCC CGG AGG AAC UCC UGG CCA UCU CAG AUU UCC CUC CAG

 20                                  30                                  40
tyr leu ser gly gly ser trp tyr his thr cys gly gly thr leu ile arg arg asn trp val met thr ala ala his
UAC CUG UCC GGA GGA UCA UGG UAC CAC ACC UGU GGA GGG ACC CUC AUC CGA CGC AAC UGG GUG AUG ACC GCU GCC CAC

                  50                                  60                                  70
cys val ser ser gln met thr phe arg val val val gly asp his asn leu ser gln asn asp gly thr glu gln tyr
UGU GUG AGC AGC CAG AUG ACU UUC CGA GUG GUC GUC GGA GAC CAC AAC CUG AGU CAG AAU GAC GGC ACG GAG CAG UAC

                                 80                                  90
val ser val gln lys ile met val his pro thr trp asn ser asn asn val ala ala gly tyr asp ile ala leu leu
GUG AGC GUG CAG AAG AUC AUG GUG CAC CCC ACC UGG AAC AGC AAC AAU GUG GCU GCA GGC UAU GAC AUC GCC CUA UUG

        100                                 110                                 120
arg leu ala gln ser val thr leu asn asn tyr val gln leu ala val leu pro gln glu gly thr ile leu ala asn
CGC UUG GCC CAG AGU GUC ACA CUC AAU AAC UAC GUC CAG CUG GCU GUU UUG CCC CAG GAG GGA ACC AUC CUG GCU AAC

                        130                                 140
asn asn pro cys tyr ile thr gly trp gly arg thr arg thr asn gly gln leu ser gln thr leu gln gln ala tyr
AAC AAU CCC UGC UAU AUC ACA GGC UGG GGG AGA ACC AGA ACC AAU GGG CAG CUG UCU CAG ACC CUG CAG CAG GCG UAC

150                                 160                                 170
leu pro ser val asp tyr ser ile cys ser ser ser ser tyr trp gly ser thr val lys thr thr met val cys ala
CUG CCC AGC GUG GAC UAC AGC AUC UGC UCC AGC UCC UCU UAC UGG GGC UCC ACG GUG AAG ACG ACC AUG GUG UGU GCU

                180                                 190                                 200
gly gly asp gly val arg ser gly cys gln gly asp ser gly gly pro leu his cys leu val asn gly gln tyr ser
GGU GGA GAC GGA GUU CGC UCU GGG UGC CAG GGU GAU UCU GGG GGA CCC CUC CAU UGC UUG GUG AAC GGC CAG UAU UCU

                                210                                 220
val his gly val thr ser phe val ser ser met gly cys asn val ser lys lys pro thr val phe thr arg val ser
GUC CAC GGA GUG ACC AGC UUU GUG UCC AGU AUG GGC UGU AAU GUC UCC AAG AAG CCC ACA GUC UUC ACC CGA GUC UCU

        230                                 240
ala tyr ile ser trp met asn asn val ile ala tyr thr stop
GCU UAC AUU UCC UGG AUG AAC AAU GUC AUU GCC UAC ACC UGA ACGUCUUCCUGAGUCCAGUGGCCUCCCCAAGAUGGUGCUUAGCUCUG

CAAUAAGAUCUGAAGUAAGCGAGAAGGAGUUAGGGCUGGGGUUAUCGCUCAGCGGCAGAACACUUUCUAAUAUGUGCGAAGCCCUGGUUCCAUCCCCCAAAUG

CUGCAGAGCGGGGGCAGGGGAGUGCUGCUGGUGUCUCUCUGCUCUUUCUCCAGUAUCUGGGGCAAAGGUGGUUAAUGAAAAACAGCUCUGAGACUGAGCCAGA

UACAGAAAGGCAAAUAAAACUCAAUGUGUUAUC(A)n
```

FIGURE 4: Complete nucleotide sequence of rat pancreatic elastase I mRNA and the amino acid sequence of the encoded preproenzyme. The scheme for sequencing is diagrammed in Figure 1 and detailed under Experimental Procedures. The deduced amino acid sequence is numbered sequentially from the amino terminus of the predicted active enzyme. The predicted activation peptide comprises amino acid residues −10 through −1 and the prepeptide residues −26 through −11. The conserved sequence AAUAAA, in the 3′-noncoding region, is underlined. The conserved sequence near the 3′ end of eukaryotic 18S rRNA (italic) that can base pair with the 5′-noncoding region of elastase I mRNA is shown at the position with the greatest number of base pairings.

ology (84%) with porcine elastase 1.

The sequence of elastase II mRNA and the derived amino acid sequence are given in Figure 5. Elastase II mRNA contains a single coding frame uninterrupted by termination codons that prescribes a protein of 271 amino acids with a molecular weight of 28 890. The conservation of the amino acids of the charge relay system and contiguous sequences identifies the encoded protein as a serine protease. The amino acid sequence has greater homology to porcine elastase 1 (58% sequence identity and a single amino acid insertion) than to other pancreatic serine proteases (Table I).

5′ Termini of the mRNAs. The length and sequence of the 5′ ends of the two elastase mRNAs were determined by primer extension (see Experimental Procedures). Figure 6 shows the results of the sequence analysis of the extended primers. For elastase II this analysis confirmed the nucleotide sequence obtained from pcXP3-3 and revealed that the mRNA extended only two nucleotides further than the cloned sequence. Comparison of the sequences obtained by primer extension analyses vs. direct sequence analyses of the 5′ ends of other mRNAs [e.g., human globin (Barelle, 1977), chicken ovalbumin (McReynolds et al., 1978; Malek et al., 1981), and mouse amylase (Hagenbuchle et al., 1980)] reveals that the primer extension is capable of delineating all but the first (methylated) nucleotide and the ^{7}meG cap structure of the mRNA. Therefore, it is likely that the nucleotide sequences presented for elastases I and II (Figures 4 and 5, respectively) include all but the first 5′ nucleotide of the entire mRNA coded length. Thus, we estimate the length of elastases I and II mRNAs as 1113 and 910 nucleotides, respectively, plus the 5′-cap structure and poly(A). The difference in length between the two is due to the long 3′-noncoding region of elastase I mRNA. With the assumption of a poly(A) tract of 60 nucleotides, similar to the number average poly(A) length for total mRNA of canine pancreas (Harding et al., 1977), the actual mRNA length is within 6% of that estimated by electrophoretic analysis in agarose gels containing methyl mercuric hydroxide (Figure 2).

In Situ Localization of Elastase I mRNA. The cloned sequence of elastase I (pcXP13) hybridizes intensely in situ

RAT PANCREATIC ELASTASE II mRNA

```
                                -30                                     -20
 -UAGGAAGGCGU-                  met ile arg thr leu leu leu ser ala leu val ala gly ala leu ser cys gly tyr pro
ACAGACAUCCACGGACACACC           AUG AUC AGG ACA CUG CUG CUA UCU GCC UUG GUG GCU GGA GCU CUC AGC UGC GGG UAC CCC

-10                                        -1  +1                                  10
thr tyr glu val gln his asp val ser arg val val gly gly gln glu ala ser pro asn ser trp pro trp gln val
ACU UAU GAG GUC CAG CAU GAU GUG AGC AGG GUA GUU GGG GGU CAA GAG GCC AGU CCC AAC AGC UGG CCC UGG CAG GUC

             20                                        30                                        40
ser leu gln tyr leu ser ser gly lys trp his his thr cys gly gly ser leu val ala asn asn trp val leu thr
UCC CUG CAG UAC CUU UCC UCU GGA AAG UGG CAC CAC ACC UGC GGA GGC UCC CUG GUG GCC AAC AAC UGG GUU CUG ACC

                               50                                        60
ala ala his cys ile ser asn ser arg thr tyr arg val leu leu gly arg his ser leu ser thr ser glu ser gly
GCU GCC CAU UGU AUC AGC AAC UCU AGG ACC UAC CGC GUG CUG CUG GGC CGA CAC AGC CUC UCC ACC UCG GAA UCC GGC

     70                                        80                                        90
ser leu ala val gln val ser lys leu val val his glu lys trp asn ala gln lys leu ser asn gly asn asp ile
UCA CUG GCU GUU CAG GUC UCC AAG CUU GUG GUC CAU GAG AAG UGG AAC GCU CAA AAG CUC UCC AAC GGG AAU GAC AUU

                   100                                       110                                       120
ala leu val lys leu ala ser pro val ala leu thr ser lys ile gln thr ala cys leu pro pro ala gly thr ile
GCC CUG GUC AAA CUG GCU AGC CCA GUG GCC CUG ACC AGC AAG AUC CAG ACG GCU UGC CUC CCA CCC GCU GGU ACC AUU

                               130                                       140
leu pro asn asn tyr pro cys tyr val thr gly trp gly arg leu gln thr asn gly ala thr pro asp val leu gln
CUG CCA AAC AAC UAC CCC UGC UAC GUC ACA GGC UGG GGC CGG CUG CAG ACC AAC GGA GCC ACU CCU GAC GUC CUG CAG

             150                                       160                                       170
gln gly arg leu leu val val asp tyr ala thr cys ser ser ala ser trp trp gly ser ser val lys thr asn met
CAG GGC CGU CUG CUG GUU GUG GAC UAU GCC ACC UGC UCC AGC GCU AGC UGG UGG GGA AGC UCC GUG AAG ACC AAC AUG

                           180                                       190
val cys ala gly gly asp gly val thr ser ser cys asn gly asp ser gly gly pro leu asn cys gln ala ser asn
GUG UGC GCU GGU GGC GAC GGU GUG ACC UCC AGC UGC AAU GGG GAC UCC GGC GGA CCA CUG AAU UGC CAG GCA UCU AAU

     200                                       210                                       220
gly gln trp gln val his gly ile val ser phe gly ser thr leu gly cys asn tyr pro arg lys pro ser val phe
GGC CAG UGG CAG GUG CAC GGU AUU GUG AGC UUU GGC UCC ACU CUG GGC UGC AAC UAC CCC CGU AAG CCA UCU GUC UUC


thr arg val ser asn tyr ile asp trp ile asn ser val ile ala lys asn stop
ACC AGG GUC UCC AAC UAC AUC GAC UGG AUC AAC UCG GUG AUU GCA AAG AAC UAA GCAAAGGCAUCUCUGUGACUGUCACCUUGGG

AACGCACGGAAGGGACACAGUAAUAAAGUGACUUACGAAUU(A)n
```

FIGURE 5: Complete nuleotide sequence of rat pancreatic elastase II mRNA and the amino acid sequence of the encoded preproenzyme. The scheme for sequencing is diagrammed in Figure 1 and detailed under Experimental Procedures. The deduced amino acid sequence is numbered sequentially from the amino terminus of the predicted major chain of the active enzyme to facilitate comparison with elastase I. The predicted activation peptide comprises amino acid residues −14 through −1 and the prepeptide residues −30 through −15. The conserved sequence AAUAAA in the 3′-noncoding region is underlined. The conserved 18S rRNA sequence (italic) is shown at the position with the greatest number of base pairings with the 5′-noncoding region.

to the acinar cells of pancreatic tissue slices (Figure 7). The autoradiographic grains define individual pancreatic acini, each containing a number of cells organized around a small central lumen. All acinar cells are labeled. The autoradiographic grains are concentrated within the basal regions of the acinar cells and reflect the distribution of rough endoplasmic reticulum within these cells. No hybridization is seen to connective tissue, blood vessels, or islets of Langerhans. The hybridization of the cloned elastase I cDNA to pancreatic tissue slices is specific for elastase I mRNA, because the cloned elastase I sequence does not cross-hybridize under normal conditions with other pancreatic mRNAs, including those for the other serine proteases (elastase II, chymotrypsin, and trypsin) (Figure 3 and unpublished observations with ds-cDNA clones for each that show the absence of cross-hybridization). These results clearly demonstrate that, in the pancreas, elastase I mRNA is present at high levels exclusively in the acinar cells. In addition, the subcellular localization within the rough endoplasmic reticulum is consistent with the known site of synthesis of pancreatic secretory proteins (Jamieson & Palade, 1967).

The result of hybridizing the cloned elastase I sequence to parotid tissue slices is shown in Figure 8. Because the parotid gland does not secrete elastase, it was anticipated that the specificity of the elastase I probe would be demonstrated by the lack of hybridization to the parotid exocrine tissue. The absence of autographic grains over the exocrine cells confirmed this prediction. Surprisingly, however, a few isolated cells within connective tissue were heavily labeled, indicating the presence of high levels of an mRNA closely related to elastase mRNA. The labeled cells were identified as mast cells by their metachromatic staining with toluidine blue, their intense staining by Giemsa, and their proximity to blood vessels.

Discussion

Two cloned sequences were initially selected from the rat pancreatic ds-cDNA library because each corresponded to sequences for prominent (i.e., secretory protein) mRNAs and hybridized to mRNAs of the size predicted for the serine protease mRNAs. Nucleotide sequence analyses of the cloned inserts demonstrated that they were derived from elastase mRNAs. Elastase I mRNA encodes a serine protease zymogen with a secretory protein leader peptide, an activation peptide at the amino terminus of the mature secreted zymogen, and an amino acid sequence 84% homologous to the sequence of porcine elastase 1. The homology increases to 88% if chemically similar amino acids [as defined by Shotton &

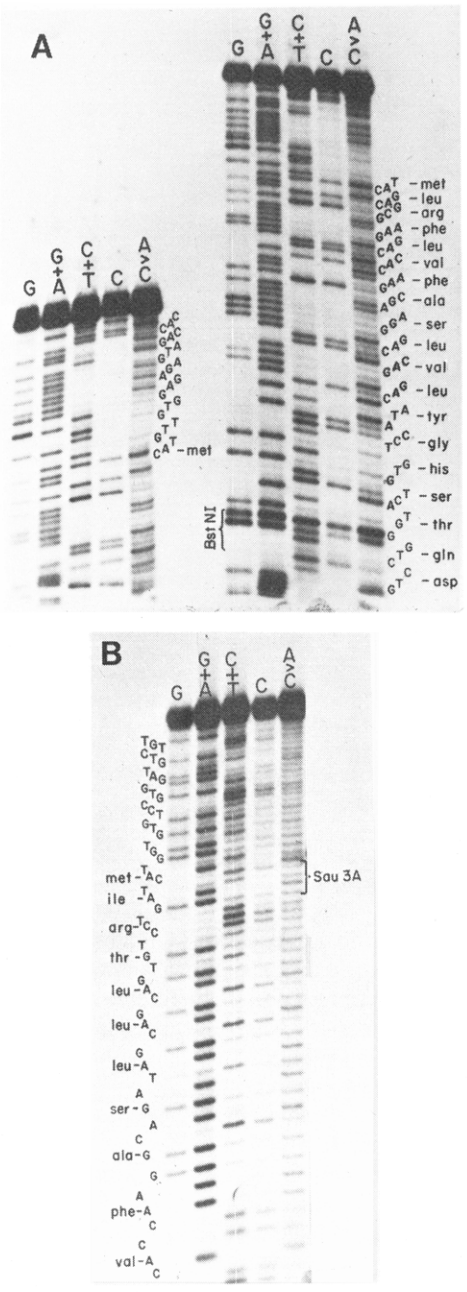

FIGURE 6: Autoradiograms showing the nucleotide sequence of the 5′ end of elastase I (panel A) and II (panel B) mRNAs determined by "primer extension" analyses. Because the cDNAs that were sequenced were generated by copying the mRNAs, the nucleotide sequences shown are the mRNA complement. The translation for this region of the mRNA is given. (Panel A) The 5′ end of elastase I mRNA. The end of the DNA primer is indicated by the *Bst*N1 bracket. The ambiguity of the sequence at the *Bst*N1 site is probably due to the presence of a small amount of contaminating unextended primer; the sequence of these two nucleotides was independently confirmed from the sequence of pcXP3-3 that overlaps this region. (Panel B) The 5′ end of elastase II mRNA. The end of the DNA primer is indicated by the *Sau*3A bracket.

Hartley (1970)] are accepted. In addition to the overall amino acid sequence homology, the presence of valine at residue 210 and threonine at residue 222 would be expected to limit the activity of the enzyme to substrates that contain amino acid residues with small side chains, such as alanine or glycine. This is a characteristic of porcine elastase 1 and distinguishes it from enzymes that have chymotrypsin-like or trypsin-like activities. Within acinar cells, in situ hybridization of the cloned elastase I was limited to the region of endoplasmic reticulum, consistent with its role as a secretory digestive enzyme. Clearly elastase

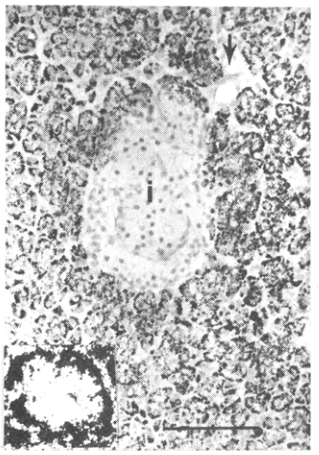

FIGURE 7: Hybridization in situ of the cloned elastase I mRNA sequence to cells of pancreatic tissue. In situ hybridization was performed with ^{3}H-labeled ds-cDNA insert of pcXP13 as described under Experimental Procedures. Acinar cells are highly labeled. The islet of Langerhans (i) and connective tissue, including blood vessels (arrow), are unlabeled. Exposure was for 10 days. The bar represents 100 μm. (Insert) 4× higher magnification of the cross section of a single acinus showing the concentration of autoradiographic grains in the basal regions of the acinar cells. The clear central area represents the lumen and the apical region of the cells that is mostly void of rough endoplasmic reticulum.

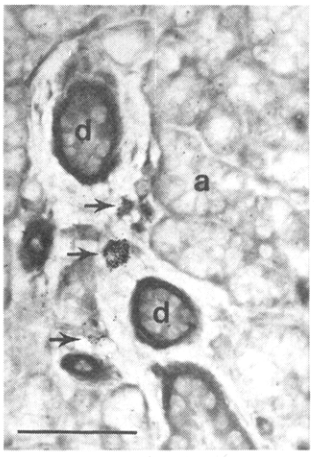

FIGURE 8: Hybridization in situ of the cloned elastase I mRNA sequence to mast cells within the parotid gland. In situ hybridization was as for Figure 7. Acini (a) and ducts (d) show no hybridization. Individual cells (arrows) within the connective tissue are labeled. Exposure was for 10 days. The bar represents 50 μm.

I represents the rat counterpart of the major elastolytic secretory (pro)enzyme described for pig (Lewis et al., 1956), human (Largman et al., 1976), and other vertebrates [reviewed by Bieth (1978)].

The hybridization in situ of the cloned elastase I sequence to all acinar cells indicates that all pancreatic acinar cells synthesize elastase 1. In similar experiments, cloned amylase mRNA sequences also hybridized to all the pancreatic acinar cells (MacDonald et al., 1981). These observations are indicative of the uniform synthesis and secretion of the entire complement of the secretory digestive enzymes by all pancreatic acinar cells and are consistent with the immunocytochemical localization of multiple secretory (pro)enzymes in all bovine pancreatic acinar cells (Kraehenbuhl et al., 1977).

The hybridization in situ of the cloned pancreatic elastase I sequence to parotid mast cell sequence demonstrates the presence of an elastase-like mRNA in tissue mast cells. These results indicate the utility of in situ hybridization to identify a few dispersed cells that contain a specific mRNA within

ALIGNMENT OF PANCREATIC ELASTASES

```
          1                                  20                                  40       (45)
EI Rat    V V G G A E A R R N S W P S Q I S L Q Y L S G G S W Y H T C G G T L I R R N W V M T A A H C V S
EII Rat   V V G G Q E A S P N S W P W Q V S L Q Y L S S G K W H H T C G G S L V A N N W V L T A A H C I S
E Pig     V V G G T E A Q R N S W P S Q I S L Q Y R S G S S W A H T C G G T L I R Q N W V M T A A H C V D

                                 60                                  80                  (93)
EI Rat    S Q M T F R V V V G D H N L S Q N D G T E Q Y V S V Q K I M V H P T W N S N N V A A G Y D I A L
EII Rat   N S R T Y R V L L G R H S L S T S E S G S L A V Q V S K L V V H E K W N A Q K L S N G N D I A L
E Pig     R E L T F R V V V G E H N L N Q N N G T E Q Y V G V Q K I V V H P Y W N T D D V A A G Y D I A L

               100                                 120                                 (141)
EI Rat    L R L A Q S V T L N N Y V Q L A V L P Q E G T I L A N N N P C Y I T G W G R T R T N G Q L S Q T
EII Rat   V K L A S P V A L T S K I Q T A C L P P A G T I L P N N Y P C Y V T G W G R L Q T N G A T P D V
E Pig     L R L A Q S V T L N S Y V Q L G V L P R A G T I L A N N S P C Y I T G W G L T R T N G Q L A Q T

                                 160                                 180(182) (185) (188)
EI Rat    L Q Q A Y L P S V D Y S I C S S S S Y W G S T V K T T M V C A G G D G V R S G C Q G D S G G P L
EII Rat   L Q Q G R L L V V D Y A T C S S A S W W G S S V K T N M V C A G G D G V T S S C N G D S G G P L
E Pig     L Q Q A Y L P T V D Y A I C S S S S Y W G S T V K N S M V C A G G N G V R S G C Q G D S G G P L

                        200                 (210)                 220(222)                        241
EI Rat    H C L - V N G Q Y S V H G V T S F V S S M G C N V S K K P T V F T R V S A Y I S W M N N V I A Y T
EII Rat   N C Q A S N G Q W Q V H G I V S F G S T L G C N Y P R K P S V F T R V S N Y I D W I N S V I A K N
E Pig     H C L - V N G Q Y A V H G V T S F V S R L G C N V T R K P T V F T R V S A Y I S W I N N V I A S N
```

FIGURE 9: Alignment of the amino acid sequences of rat elastase I, rat elastase II, and porcine elastase 1. The alignment beings at the homologous amino termini of the active form of porcine elastase 1, the predicted active form of rat elastase I, and the predicted major chain of the activated rat elastase II to facilitate comparisons. The numbering is for elastase II, one amino acid (residue 196) longer than the others. The numbers in parentheses indicate the position of residues discussed in the text.

a large population that does not. The absence of cross-hybridization between the elastase I ds-cDNA and any other pancreatic mRNA indicates that the pancreatic elastase I mRNA sequence is more closely related to the mast cell sequence than to elastase II, chymotrypsinogen, or trypsinogen mRNAs. Rat tissue mast cells contain large amounts of a serine protease, termed mast cell protease. The amino acid sequence of mast cell protease is known (Woodbury et al., 1978), and it differs significantly from elastase, chymotrypsin, and trypsin. The degree of sequence identity of the rat mast cell protease with rat elastase I is only 31%. Unless localized regions of the mRNA sequence have been highly conserved, the mast cell protease mRNA would not hybridize with the elastase probe under the conditions of in situ hybridization. An alternative explanation is the presence in rat tissue mast cells of an as yet unreported (pro)elastase that is homologous to the pancreatic enzyme.

Elastase II mRNA encodes a serine protease zymogen with a secretory protein leader peptide, an "activation peptide" related to the A chain of chymotrypsin, and an amino acid sequence with 58% sequence identity with porcine elastase 1. The primary structure of rat pancreatic proelastase II that was deduced from the nucleotide sequence of its mRNA clearly establishes that this enzyme is a serine protease. Alignment of the amino acid sequence with other serine proteases reveals sequence identity of 40% with bovine chymotrypsin B, 36% with bovine trypsin, 33% with porcine kallikrein (Table I), and 31% with the rat mast cell protease.

A preference for proteolytic cleavage at certain amino acid residues may be inferred from the derived amino acid sequence of elastase II. Relative to elastase 1, the replacement of Val-209 and Thr-221 with Gly-210 and Ser-22 (see Figure 9) should enlarge the binding pocket and accommodate binding of substrates with large side chains to the S_1 position.[3] The retention of significant sequence homology with elastase 1 might be expected to retain significant activity toward elastase-specific substrates containing alanine and glycine. The conservation of Ser-182 (Ser-189 of chymotrypsin) would maintain a preference for amino acids with hydrophobic side chains.

The presence of glutamine-185 (position 192 in chymotrypsin) in elastase 1 is thought to facilitate binding of substrates with small side chains in the P_1 position by movement over the binding cleft after substrate binding (Shotton et al., 1971). In the case of alanine in the P_1 position, this induced-fit movement traps the methyl side chain between Gln-185 and Val-210. The substitution in rat elastase II of asparagine at this position (Figure 9) may cause the enzyme to be less effective in binding substrates with small side chains.

Shotton et al. (1971) proposed that the side chain of Leu-141 in elastase 1 maintains a hydrophobic interaction with the substrate residues in the P_4 position and is an important consideration for efficient catalysis. In rat elastase II, the replacement of Leu-141 by threonine would eliminate this potential interaction. As a consequence of these and other amino acid substitutions, the elastase II specificity is predicted to be more like chymotrypsin.

The predicted substrate preferences of the rat enzyme correlate well with the observed substrate specificity of a second pancreatic elastolytic enzyme, elastase 2, isolated from porcine (Ardelt, 1975; Gertler et al., 1977) and human (Largman et al., 1976; Del Mar et al., 1980) pancreas. Gertler et al. (1977) demonstrated thar porcine elastase 2 preferentially cleaves elastin at bonds involving leucine, phenylalanine, and

[3] The nomenclature of Schechter & Berger (1967) is used to identify subsites on the enzyme (Sn) and amino acid positions within the polypeptide substrate (Pn).

	-10												+1			
Rat Proelastase I							T	Q	D	F	P	E	T	N	A	R	V	V	G	G
Lungfish Proelastase B							S	E	M	Y	L	E	E	M	E	R	V	V	G	G
Rat Proelastase II	C	G	Y	P	T	Y	E	-	V	Q	H	D	V	S	R	V	V	G	G	
Lungfish Proelastase A	C	G	V	P	S	Y	P	P	T	-	-	-	-	A	R	V	V	G	G	
	+1								+10											
Bovine Chymotrypsinogen A	C	G	V	P	A	I	Q	P	V	L	S	G	L	S	R	I	V	N	G	
Porcine Chymotrypsinogen A	C	G	V	P	A	I	Q	P	V	L	S	G	L	A	R	I	V	N	G	
Bovine Chymotrypsinogen B	C	G	V	P	A	I	Q	P	V	L	S	G	L	A	R	I	V	N	G	
Bovine Chymotrypsinogen C	C	G	A	P	I	F	Q	P	N	L	S	-	-	A	R	V	V	G	G	
Porcine Chymotrypsinogen C	C	G	V	P	S	F	P	P	N	L	S	-	-	A	R	V				

FIGURE 10: Alignment of the amino-terminal sequences of rat proelastases I and II with those of other related pancreatic serine protease zymogens. The data for the other related serine protease zymogens were compiled by deHaen & Gertler (1974). The numbering for proelastases I and B is that for the predicted active enzyme with an amino-terminal sequence homologous to porcine elastase 1. An alternative numbering system based upon bovine chymotrypsinogen is also given. The known and predicted cleavage sites for activation are aligned with the +1 position for rat proelastase I.

tyrosine in the P_1 position in contrast to elastase 1 which prefers Ala–Ala and Ala–Gly bonds of elastin. Elastase 2 more readily hydrolyzes specific synthetic substrates of chymotrypsin than those of elastase 1 (Ardelt, 1975; Gertler et al., 1977; Del Mar et al., 1980). Human elastase 2 preferentially hydrolyzes synthetic substrates with medium to large hydrophobic side chains, except those with side chains that are disubstituted on the first carbon (Del Mar et al., 1980). Del Mar et al. (1980) also demonstrated that elastase 2 has a relatively high K_m for its best synthetic substrate compared to the best substrates reported for other serine proteases, indicating less efficient substrate binding and suggesting a low inherent catalytic efficiency. Although it appears that the rat elastase II reported in this paper is the rat counterpart of elastase 2 isolated from other mammalian pancreases, definitive proof could be obtained by the expression of the cloned elastase II mRNA sequences in other host cells, the purification of the enzyme, and an analysis of its preferred cleavage sites in elastin and synthetic substrates.

mRNA and Protein Domains. 5′-Noncoding Regions. Both elastase I and II mRNAs have relatively short untranslated sequences at the 5′ end. Although no significant homology exists between the two noncoding regions, both contain sequences complementary to a highly conserved sequence at the 3′ end of eukaryotic 18S rRNA (Hagenbuchle et al., 1978). It has been proposed that the efficiency of initiation of translation on eukaryotic mRNAs is enhanced by the presence of a short 5′-noncoding leader sequence (Young et al., 1981) and the presence of a region in the 5′ leader capable of base pairing with the conserved 3′ region of 18S rRNA (Hagenbuchle et al., 1978).

3′-Noncoding Regions. The 3′-noncoding region of elastase I consists of 289 nucleotides starting with the termination codon UGA; that of elastase II is 75 nucleotides following the termination codon UAA. The hexanucleotide AAUAAA, which is conserved near the 3′ end of polyadenylated eukaryotic mRNAs (Proudfoot & Brownlee, 1976), is present 14 and 15 nucleotides upstream from the poly(A) in elastase I and elastase II mRNAs, respectively. Aside from this conserved hexanucleotide, there are no regions of significant homology between the two 3′-noncoding regions. This is in contrast to the extensive amino acid (58%) and nucleotide (62%) sequence homologies in the coding regions.

Prepeptides of the Secretory Proteins. Elastases I and II each contain a prepeptide (signal peptide) 16 amino acids long, similar in length and in hydrophobic character to the prepeptides of other secretory proteins (for example, Habener et al., 1978) that are processed to generate the secreted protein. On the basis of the characteristics of the amino acid residues at the cleavage site of other prepeptides and the anticipated amino-terminal sequences of the activation peptides of the elastase zymogens (see below), cleavage for the prepeptides is predicted to occur between Ser-(−11) and Thr-(−10) in preproelastase I (Figure 4) and between Ser-(−15) and Cys-(−14) in preproelastase II (Figure 5).

Activation Peptides of the Zymogens. The activation peptide of rat proelastase I includes amino acids −10 through −1 (Figure 10). The carboxy terminus of the activation peptide is determined from the highly conserved amino-terminal sequence of the active serine proteases. The amino terminus of the activation peptide cannot be identified unequivocally without an independent determination of the amino terminus of the proenzyme by protein sequencing techniques. Nonetheless, the predicted activation peptide is the same length as that for lungfish proelastase B (deHaen & Gertler, 1974). The sequence of the rat proelastase I activation peptide does not resemble significantly the sequence of any other serine protease activation peptide (Figure 10). The differences in the amino acid composition between porcine elastase 1 and its zymogen (Gertler & Birk, 1970) indicate an activation peptide of the same length as rat proelastase I and a similar (Ser, Lys, Asp, Phe, Pro_3, Glu, Ala, and Arg), though clearly distinct, composition.

The activation peptide of rat elastase II is 14 amino acids long and is related to the A chain of the chymotrypsins and the amino-terminal sequence of lungfish proelastase A (Figure 10). In addition, the internal cysteine residue of chymotrypsin that forms a disulfide bridge with the amino-terminal cysteine of the A chain is conserved in rat elastase II (Cys-113). Therefore the amino-terminal peptide of proelastase II may be associated with the enzyme after activation. This is consistent with the preliminary evdience of Gertler et al. (1977) which suggested that porcine elastase 2 is composed of a short (A) chain with an amino-terminal sequence Cys–Gly linked to a major (B) chain by a disulfide bond. Activation of proelastase II by cleavage after Arg-(−1) would allow the amino-terminal Val-(+1) to form an electrostatic bond with Asp-187 (Asp-194 in chymotrypsin), as occurs upon chymotrypsinogen activation (Freer et al., 1970).

Active Enzymes. The molecular weight of the predicted

active form of elastase I is 25 985. The molecular weight of the active form of elastase II would be 27 278, assuming that the activation peptide remains associated with the enzyme. This is similar to the molecular weight of 26 800 determined for porcine elastase 2 (Gertler et al., 1977).

The presence of an activation peptide similar to the chymotrypsins and the presence of key amino acid residues that suggest a chymotrypsin-like substrate preference indicate that elastase 2 is an evolutionary link between the chymotrypsins and elastase 1. The amino acid sequence of rat elastase II is equally related (58% sequence identity with an amino acid inserted at residue 196; see Figure 9) to both porcine and rat elastase 1 (Table I). In contrast, sequence alignment of rat elastase II with bovine chymotrypsins A and B is only 38% and 40%, respectively, and requires making nine insertions or deletions. Analyses of the amino-terminal sequences of two proelastases (A and B) isolated from African lungfish pancreas revealed that proelastase B was related to porcine elastase 1, whereas proelastase A was a clearly different enzyme that appeared to be more closely related to the chymotrypsins (deHaen & Gertler, 1974). If rat elastase II is the equivalent of lungfish proelastase A, as it appears to be, then its closer overall sequence homology with elastase I than with chymotrypsin points out the danger of assigning evolutionary relationships on the basis of a short sequence fragment such as the activation peptide.

The preparation of ds-cDNA clones for serine proteases now permits the engineering and production of altered enzymes by site-specific mutagenesis and expression in alternative host cells. This should facilitate further probing of the catalytic mechanism and the proposed determinants of substrate preference of specific serine proteases, as well as creating potentially useful proteases with altered specificity.

Acknowledgments

We thank Drs. H. W. Boyer for providing pBR322, R. Curtiss for providing X1776, A. Ullrich, J. Shine, P. Seeburg, and E. Craig for helpful discussions on ds-cDNA cloning, H. Martinez for providing the computer programs for DNA sequence analyses, R. Sandhaus and Z. Werb for help in identifying tissue mast cells, G. Bell for helpful discussions and criticisms, and Sabina Dougherty and Leslie Spector for preparing the manuscript.

References

Alwine, J. C., Kemp, D. J., & Stark, G. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5350–5354.

Ardelt, W. (1975) *Biochim. Biophys. Acta 393*, 267–273.

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

Bailey, J. M., & Davidson, N. (1976) *Anal. Biochem. 70*, 75–85.

Baralle, F. E. (1977) *Cell (Cambridge, Mass.) 12*, 1085–1095.

Biessmann, H., Craig, E. A., & McCarthy, B. J. (1979) *Nucleic Acids Res. 7*, 981–996.

Bieth, J. (1978) *Front. Matrix Biol. 6*, 1–82.

Blobel, G., & Sabatini, D. (1971) *Biomembranes 2*, 193–195.

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

Craig, E. A., McCarthy, B. J., & Wadsworth, S. C. (1979) *Cell (Cambridge, Mass.) 16*, 574–588.

deHaen, C., & Gertler, A. (1974) *Biochemistry 13*, 2673–2677.

deHaen, C., Neurath, H., & Teller, D. C. (1975) *J. Mol. Biol. 92*, 225–259.

Del Mar, E. G., Largman, C., Brodrick, J. W., Fassett, M., & Geokas, M. C. (1980) *Biochemistry 19*, 468–472.

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

Devillers-Thiery, A., Kindt, T., Scheele, G., & Blobel, G. (1975) *Proc. Natl. Acad Sci. U.S.A. 72*, 5016–5020.

Freer, S. T., Kraut, J., Robertus, J. D., Wright, H. T., & Xuong, Ng. H. (1970) *Biochemistry 9*, 1997–2009.

Gertler, A., & Birk, Y. (1970) *Eur. J. Biochem. 12*, 170–176.

Gertler, A., Weiss, Y., & Burstein, Y. (1977) *Biochemistry 16*, 2709–2716.

Goodman, H. M., & MacDonald, R. J. (1980) *Methods Enzymol. 68*, 75–90.

Grunstein, M., & Hogness, D. S. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 3961–3965.

Habener, J. F., Rosenblatt, M., Kemper, B., Kronenberg, H. M., Rich, A., & Potts, J. T., Jr. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 2616–2620.

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

Hagenbuchle, O., Bovery, R., & Young, R. A. (1980) *Cell (Cambridge, Mass.) 21*, 179–187.

Harding, J. D., MacDonald, R. J., Przybyla, A. E., Chirgwin, J. M., Pictet, R. L., & Rutter, W. J. (1977) *J. Biol. Chem. 252*, 7391–7397.

Helling, R. B., Goodman, H. M., & Boyer, H. W. (1974) *J. Virol. 14*, 1235–1244.

Holland, L. J., & McCarthy, B. J. (1980) *Biochemistry 19*, 2965–2976.

Jamieson, J. D., & Palade, G. E. (1967) *J. Cell Biol. 34*, 597–615.

Kraehenbuhl, J. P., Racine, L., & Jamieson, J. D. (1977) *J. Cell Biol. 72*, 406–423.

Largman, C., Brodrick, J. W., & Geokas, M. C. (1976) *Biochemistry 15*, 2491–2500.

Lewis, U. J., Williams, D. E., & Brink, N. G. (1956) *J. Biol. Chem. 222*, 705–720.

MacDonald, R. J., Crerar, M. M., Swain, W. F., Pictet, R. L., Thomas, G., & Rutter, W. J. (1980) *Nature (London) 287*, 117–122.

MacDonald, R. J., Crerar, M. M., Swain, W. F., Pictet, R. L., & Rutter, W. J. (1981) *Cancer (Amsterdam) 47*, 1497–1504.

Malek, L. T., Eschenfeldt, W. H., Munns, T. W., & Rhoads, R. E. (1981) *Nucleic Acids Res. 9*, 1657–1673.

Maxam, A., & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 560–564.

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

McReynolds, L., O'Malley, B. W., Nisbet, A. D., Fothergill, J. E., Gival, D., Fields, S., Robertson, J., & Brownlee, G. G. (1978) *Nature (London) 273*, 723–728.

Meagher, R. B., Tait, R. C., Betlach, M., & Boyer, H. W. (1977) *Cell (Cambridge, Mass.) 10*, 521–536.

Neurath, H., Walsh, K. A., & Winter, W. P. (1967) *Science (Washington, D.C.) 158*, 1638–1644.

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

Przybyla, A. E., MacDonald, R. J., Harding, J. D., Pictet, R. L., & Rutter, W. J. (1979) *J. Biol. Chem. 254*, 2154–2159.

Richards, R. I., Shine, J., Ullrich, A., Wells, J. R. E., & Goodman, H. M. (1979) *Nucleic Acids Res. 7*, 1137–1146.

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

Roychoudhury, R., Jay, E., & Wu, R. (1976) *Nucleic Acids Res. 3*, 101–116.

Rutter, W. J., Przybyla, A. E., MacDonald, R. J., Harding, J. D., Chirgwin, J. M., & Pictet, R. L. (1978) in *Differentiation and Neoplasia* (Saunders, G. F., Ed.) pp 487–508, Raven Press, New York.

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

Shotton, D. M., & Hartley, B. S. (1970) *Nature (London) 225*, 802–806.

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

Shotton, D. M., White, N. J., & Watson, H. C. (1971) *Cold Spring Harbor Symp. Quant. Biol. 36*, 91–105.

Sippel, A. E., Land, H., Lindenmaier, W., Nguyen-Huu, M. C., Wurtz, T., Timmis, K. N., Giesecke, K., & Schutz, G. (1978) *Nucleic Acids Res. 5*, 3275–3294.

Southern, E. M. (1975) *J. Mol. Biol. 98*, 503–517.

Tschesche, H., Mair, G., Godec, G., Fiedler, F., Ehret, W., Hirschauer, C., Lemon, M., & Fritz, H. (1978) in *Kinins-II* (Fujii, S., Moriya, H., & Suzuki, T., Eds.) pp 245–260, Plenum Press, New York and London.

VanNest, G. A., MacDonald, R. J., Raman, R. K., & Rutter, W. J. (1980) *J. Cell Biol. 86*, 784–794.

Woodbury, R. G., Katunuma, N., Kobayashi, K., Titani, K., & Neurath, H. (1978) *Biochemistry 17*, 811–819.

Young, R. A., Hagenbuchle, O., & Schibler, U. (1981) *Cell (Cambridge, Mass.) 23*, 451–458.