Structure of the Two Related Elastase Genes Expressed in the Rat Pancreas

(Geared for publication, March 12, 1984)

Galvin H. Swift†‡, Charles S. Craik§, S. Jennifer Stary‡, Carmen Quintof‖, Raymond G. Lahaiet‡**, William J. Rutter†, and Raymond J. MacDonald‡

From the †,Division of Molecular Biology, Department of Biochemistry, The University of Texas Health Science Center at Dallas, Dallas, Texas 75235 and the ‡Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143

We have isolated and characterized rat genomic DNA fragments containing the two secretory elastase genes that are expressed in the exocrine pancreas. The complete exonic sequences for each of the genes as well as considerable intronic and flanking sequences are reported. Each elastase gene is interrupted by seven intervening sequences which are located at corresponding positions within the two genes, with one exception: the third intron of the elastase II gene has shifted one codon in the 5’ direction. The placement of introns within the amino acid coding domains in part may reflect the formation of the progenitor serine protease gene by the duplication of an exon encoding a characteristic polypeptide structure comprising three β sheets. The activation peptides of the zymogens and the signal peptides, which form discrete functional domains in the protein precursors, are encoded by separate exons. In addition to the TATA box, the two genes share considerable sequence similarity in the 5’O-proximal flanking regions (up to ~450 base pairs upstream); however, a number of gaps must be introduced to optimize the sequence alignment. The similarities are largely confined to six oligonucleotide regions with greater than 70% sequence conservation. The elastase I gene has a perfect repeating copolymer (GT)10 located 427–379 nucleotides upstream from the start of transcription. The elastase II gene has a similar GT-rich region (521/55 G or T) located 384–330 nucleotides upstream. Comparison of the 5’-flanking regions of the two elastase genes with those of pancreatic chymotrypsin and trypsin I and II reveals that one of the six conserved oligonucleotide regions is generally conserved for these genes as well. This conserved region contains putative enhancer core sequences.

Mammalian pancreatic elastases I and II (EC 3.4.21.11 and 3.4.21.1, respectively) are synthesized and secreted in large amounts by the exocrine pancreas. Both enzymes effectively digest elastin, although their preferred amino acid cleavage sites differ (Gertler et al., 1977; Largman et al., 1976). Determination of the mRNA sequence from cloned cDNAs for the two forms of rat pancreatic elastase revealed that the two enzymes are both initially synthesized as preproenzymes and that the proposed mature, active forms of the enzymes share 58% amino acid sequence identity (MacDonald et al., 1982a). These two secretory elastases are encoded by two prominent mRNAs of the exocrine pancreas (MacDonald et al., 1982a).

The two elastases are members of the pancreatic subfamily of simple serine proteases that comprise at least nine members in the rat: three isozymes of trypsin, three of chymotrypsin, one kallikrein, as well as the two elastases. These serine proteases are encoded by related genes that evolved from a common ancestral protease through gene duplication events (Neurath et al., 1967; deHaen et al., 1975). The structural genes encode related enzymes of similar size, structure, and function, but with variable proteolytic cleavage specificities. An important aspect of this subset of the extended serine protease gene family is their specific expression to high levels in pancreatic serinar cells.

In this report, we describe the relationship between the exon coding domains and the tertiary protein structure of the elastases. We also describe 5’-flanking sequences that may harbor gene regulatory elements involved in the pancreas-specific expression of the elastase genes. The accompanying papers describe the gene sequence organization for chymotrypsin B (Bell et al., 1984) and two trypsins (Craik et al., 1984) that are also expressed in the rat pancreas.

MATERIALS AND METHODS

Screening the Recombinant Phage Library, Phage Growth, and DNA Preparation—A recombinant DNA library of rat liver genomic fragments generated by partial EcoRI digestion and cloned into Charon 4A was the generous gift of T. D. Sargent, R. B. Wallace, and J. Bonner, California Institute of Technology (Sargent et al., 1979). The EcoRI library was screened (Benton and Davis, 1977) with nick-translated probes (Rigby et al., 1977) for elastase I or II (see below). Positive clones were isolated by three rounds of plaque purification. Large-scale preparations were done as described by Sain and Erdal (1981).

Restriction-endonuclease Digestions and Mapping—Restriction endonucleases were purchased from New England Biolabs (Beverly, MA) or Bethesda Research Laboratories. Three different buffers were used, all containing 6 mM Tris-Cl, pH 7.5, 6 mM MgCl2, 3 mM 2-mercaptoethanol, but supplemented with different levels of NaCl: low salt = plus 6 mM NaCl for KpnI, BglII, BalI, and HpaII; medium salt = plus 60 mM NaCl for BamHI, HindIII, PstI, AsaI, PsaI, NciI, and XmaI; high salt = plus 150 mM NaCl for NdeI, StuI, and NcoI. The buffer for EcoRI was medium salt plus 100 mM Tris-Cl, pH 7.5.

Double digests of recombinant λ phage and pBR322 subclones requiring different buffers were performed first with the enzyme requiring lower salt, then increasing the salt concentration appropriately for the second enzyme. Mapping studies as well as positioning of the exons were aided by transferring restriction-endonuclease

* This work was supported by Grante AM27480 (R. J. M.) and AM21344 (W. J. R.) from the National Institutes of Health and Grant PCM8008221 (R. J. M.) from the National Science Foundation. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Virginia Lazenby O’Hara Fellow.
§ Present address, Centro de Fusigation de Nitrigeno, Cuernavaca, Morelos, Mexico.
‖ Present address, Centre de Recherches Cliniques, Hopital Saint Luc, Montreal, Canada.
digests that had been resolved by electrophoresis to paper and hybridization with 32P-labeled purified restriction-endonuclease fragments of the cloned ds-cDNAs representing regions of the appropriate mRNA.

Electrophoresis was performed either in agarose gels containing 40 mM Tris acetate, pH 8.2, 1 mM EDTA, 20 mM NaCl (Maniatis et al., 1982) or in polyacrylamide gels containing 0.1 M Tris borate, pH 8.3, 2 mM EDTA (Peacock and Dingman, 1969). For preparative electrophoresis, DNA fragments were electrophoresed from agarose or polyacrylamide gel slices in an ISCO (Lincoln, NE) concentrator according to Zassenhaus et al. (1982).

For mapping studies, DNA fragments in agarose gels were transferred to nitrocellulose (Sambrook, Schleicher & Schuell) or Gene Screen (New England Nuclear) by blotting (Southern, 1975) and DNA fragments in polyacrylamide gels were transferred to diazobenzyloxymethyl paper (Bio-Rad) by electrophoretic blotting according to Stollwag and Dahlberg (1980) with the modifications of Reidelhuber et al. (1980). Hybridization to filter-bound DNA was performed as described previously (MacDonald et al., 1982b). For digests of genomic DNA, transfer from agarose gels was performed by blotting (Southern, 1975) to nitrocellulose. The filter-bound genomic DNA was prehybridized for 24 h at 42 C in 50% formamide, 5 × SSC, 10 × Denhardt's solution, 50 mM NaPO4, pH 6.5, 250 μg/ml denatured sheared salmon sperm DNA, 0.1% SDS, and 0.1% sodium pyrophosphate, then hybridized for 8-10 h in the same solution containing approximately 2-5 × 106 cpm/ml 32P-labeled hybridization probe. Filters were washed twice for 10 min at room temperature in 2 × SSC with 0.5% SDS and then twice for 20 min at 55 C in each of the following solutions: 0.1 × SSC containing 0.5% SDS and 0.1% sodium pyrophosphate, 0.1 × SSC containing 0.1% SDS, and 0.1 × SSC.

32P-labeled hybridization probes were prepared from recombinant plasmid DNA or isolated ds-cDNA inserts by nick translation (Rigby et al., 1977). Nick-translated DNA was purified by phenol/chloroform extraction and microcentrifugal chromatography (Helmerhorst and Stokes, 1980; MacDonald et al., 1982b).

Isolation of Rat Nuclear DNA—Rat liver and spleen were finely chopped and then homogenized in 250 mM sucrose (3 ml/g of tissue) containing 10 mM Tris-CI, pH 7.4, 2 mM CaCl2, and 30 mM NaCl using a Sorvall Omni-Mixer at half-maximal speed for 30-40 s. After filtering the homogenate through three layers of 0.8-mm mesh Nitex screen, Triton X-100 was added to 1% (v/v) and nuclei were collected by centrifugation for 5 min at 1000 x g. The crude nuclei were resuspended in the homogenization buffer using 0.5 ml/g of starting tissue. With gentle swirling, the nuclei were lysed by the addition of 3 volumes of a solution containing 10 mM Tris-CI, pH 8.0, 15 mM NaCl, 10 mM EDTA, 1.3% SDS, and 70 μg/ml proteinase K. After incubation at 40 C for 6-12 h, the DNA was gently extracted with phenol and then chloroform/1-butanol (4:1, v/v), dialyzed against 10 mM Tris-CI, pH 8, with 2 mM EDTA, and stored at 4 C over chloroform.

Selection into pBR322 and isolation of Plasmid DNA—EcoRI fragments from recombinant λ DNA were subcloned into the EcoRI site of pBR322. pBR322 was digested with EcoRI, treated with calf intestine alkaline phosphatase (Worthington), mixed with appropriately equal molar amounts of EcoRI-digested λE1a, λE1b, or λE2 DNA, ligated with T4 DNA ligase, and used to transform Escherichia coli R1. The procedures used are described by Goodman and MacDonald (1979). In some instances, it was necessary first to isolate individual EcoRI fragments of recombinant λ DNA by preparative electrophoresis and electrophoresis as deoxyribosomes. Bacterial colonies containing recombinant plasmids bearing the appropriate EcoRI fragments were identified from restriction-endonuclease digests of plasmid DNA isolated according to Birnboim and Doly (1979).

For large-scale preparation of plasmid DNA, 1-liter cultures were amplified with chloramphenicol and cleared lysates were prepared as described by Holland and McCarthy (1980), treated with 100 μg/ml boiled RNase A at 37°C for 4 h, extracted with phenol/chloroform, and precipitated with ethanol. Plasmid DNA was then obtained by chloroform extraction and ethanol precipitation.

Nucleic Acid Sequencing—Nucleotide-sequence determinations were performed according to the protocols of Maxam and Gilbert (1980). DNA fragments were end-labeled with [γ-32P]ATP and T4 polynucleotide kinase and isolated by preparative gel electrophoresis and electroelution as described above. Five sequencing reactions (G, dimethyl sulfoxide, G + A, formic acid; C + T; hydrazine; C, hydrazine + salt; and A + C, sodium hydroxide) were employed routinely to enhance sequence accuracy. Misreadings of sequence data were minimized by two independent readings of all sequencing-gel autoradiograms and rechecking these readings directly against a computer record derived from the original readings. Except for a limited number of discrepancies discussed under “Results,” the exon coding sequences were confirmed by the previously determined mRNA sequences of both genes (MacDonald et al., 1982a). Compilation and analysis of nucleotide-sequence data were performed with the programs and facilities of Dr. Hugo Martinez, Biomatics Computing Laboratory, Department of Biochemistry and Biophysics, University of California, San Francisco, CA.

RESULTS

Identifying and Mapping Cloned Genomic Sequences for Rat Elastase I and II genes—Hybridization probes for screening a λ rat genomic DNA library were derived from cloned double-stranded cDNA sequences of elastase I and II mRNAs (MacDonald et al., 1982a). The recombinant plasmid pcXP13 contained a ds-cDNA insert of 920 base pairs comprising the 3’ 80% of the elastase I mRNA. pcXP30 contained a ds-cDNA insert of 590 base pairs comprising the 3’ 65% of elastase II mRNA. The library of rat genomic DNA fragments derived from partial digestion with EcoRI and cloned in Charon 4A by Sargent et al. (1979) was screened separately by hybridization with pcXP13 and pcXP30 DNA labeled with 32P by nick translation. Under the hybridization conditions employed, the cloned elastase cDNAs (62% nucleotide-sequence similarity) do not cross-hybridize.

The elastase I hybridization probe detected 26 recombinant λ clones that could be grouped into two classes (represented by λE1a and λE1b) based upon the number and sizes of EcoRI fragments. Restriction-endonuclease mapping, Southern-blot analysis using ds-cDNA probes encoding various regions of elastase I mRNA, and nucleotide-sequence analysis (see below) demonstrated that λE1b contained exons 1-6 and λE1a contained exons 7 and 8 of the elastase I gene.

The elastase II hybridization probe detected seven identical recombinant λ clones. One, designated λE2, was shown by restriction-endonuclease mapping, Southern-blot analysis using ds-cDNA probes encoding the 5′, middle, and 3′ regions of elastase II mRNA, and nucleotide-sequence analyses (see below) to encode the mRNA sequence for pancreatic elastase II mRNA.

Restriction-endonuclease maps of the cloned elastase I and II genes are presented in Fig. 1. Each gene is interrupted by seven intervening sequences that extend the coding sequences from the 1.1 kb of the mRNA for elastase I to 12 kb in the genome and from 0.92 kb of the mRNA for elastase II to 10 kb. Analysis of the λ clones for each elastase gene established that all eight exons for elastase I were contained within the two λ clones λE1a and λE1b and that all eight exons for elastase II were contained within λE2. Comparison of the Southern-blot hybridization of the two elastase I λ clones and genomic DNA utilizing the entire mRNA sequence as a probe (Fig. 2, left panel) revealed no additional bands in the genomic DNA beyond those predicted by the λ clone map. Therefore, we conclude that there are no cross-hybridizing genes closely related to elastase I. Similar evidence (Fig. 2, right panel) indicates that elastase II also represents a unique gene in the rat genome.

Because the genomic regions of two recombinant clones for
Fig. 1. Restriction-endonuclease map and exon positions of the two pancreatic elastase genes. Restriction-endonuclease sites along the cloned genomic sequences are: E, EcoRI; H, HindIII; B, BamHI; G, BglII; P, PstI; K, KpnI; and N, NcoI. Within each map, the positions of exons (numbered 1 to 8) are indicated by narrow boxes. The length scale marks the distance from the predicted first nucleotide of exon 1 of both genes. The elastase I gene and flanking sequences are spanned by the two recombinant phages shown, \( \lambda E1a \) and \( \lambda E1b \). The dashed line in \( \lambda E1a \) indicates the presence of four EcoRI fragments of 1.6, 1.5, 0.6, and 0.5 kb that have not been ordered. The wavy lines indicate the phage arms. The recombinant phage designated \( \lambda E2 \) contains the entire genomic region shown for elastase II. The region deleted in approximately 20% of the preparation of \( \lambda E2 \) is delimited by the short horizontal bar between exons 7 and 8 (see also Fig. 2). The precise limits of the deletion and the nature of the sequences are unknown.

![Restriction-endonuclease map and exon positions of the two pancreatic elastase genes.](image)

Fig. 2. Southern-blot analysis of the mRNA coding domains of the elastase I gene (left) and the elastase II gene (right) in \( \lambda \) clones and in total genomic DNA. Recombinant clones \( \lambda E1a \), \( \lambda E1b \), and \( \lambda E2 \), digested with either EcoRI or BglII, were cut to approximately the level of a single-copy gene in rat nuclear DNA using yeast DNA as carrier and electrophoresed alongside EcoRI- and BglII-digested rat nuclear DNA. The left panel was hybridized with \( ^{32}P \)-labeled pcXP13A DNA (a recombinant plasmid containing a full-length elastase I cDNA) to detect elastase I exon sequences. The right panel was hybridized with a mixture of \( ^{32}P \)-labeled pcXP17 and pcXP30 DNA, which together include virtually all elastase II mRNA sequences (MacDonald et al., 1982a). The exons present in each hybridizing band are indicated by the numbers. The lane marked \( \lambda \) contained the HindIII-digested \( \lambda 1857 \) S7 DNA size standards indicated by the length scales marked in kilobase pairs. The asterisks denote hybridizing bands from \( \lambda E2 \) that result from a subset (approximately 20%) of the recombinant \( \lambda \) clone that has a 1.6-kb deletion within intron 6 of elastase II. The position and length of this deletion are shown in Fig. 1.

![Southern-blot analysis of the mRNA coding domains of the elastase I gene (left) and the elastase II gene (right).](image)

the elastase I gene did not overlap, it was necessary to determine whether additional DNA sequences exist in genomic DNA between the two regions represented in \( \lambda E1a \) and \( \lambda E1b \). If the two EcoRI sites abut in genomic DNA, then the BglII fragment that contains this site is predicted to be 3.3 kb (see Fig. 1); additional sequences would make this BglII fragment correspondingly longer. The data of Fig. 2 (left panel) show that this fragment, which contains exon 6, is approximately 3.3 kb. Therefore, it appears that all DNA sequences within the elastase I gene are contained within \( \lambda E1a \) and \( \lambda E1b \). However, the presence of an additional EcoRI fragment 0.3 kb or shorter would not be discerned by this analysis.

Nucleotide Sequences of the Elastase I and II Genes—Fig. 3 illustrates the sequencing schemes for the two cloned elastase genes. Rather than determining the entire nucleotide sequence of each gene, three regions were emphasized: 5'-flanking sequences, exons, and exon/intron boundaries. The sequencing strategies generally involved initiating sequencing runs from within exons utilizing infrequent restriction-endonuclease sites previously determined from the mRNA sequences. Comparison of the nucleotide sequences of the exons with the sequences of the mRNAs showed that these cloned genes clearly represent the two elastase mRNAs expressed in the pancreas. The elastase I mRNA sequence derived from the cloned cDNA (MacDonald et al., 1982a) is identical to the exon sequences in \( \lambda E1a \) and \( \lambda E1b \) with one exception (see Fig. 4). The difference alters codon 78 from methionine in the mRNA to valine. This near-identity also indicates that \( \lambda E1a \) and \( \lambda E1b \) represent halves of the same elastase gene. The single-nucleotide discrepancy may be due to a cloning artifact or to an allelic difference that reflects the lack of genetic homogeneity in the Sprague-Dawley strain, since the cloned mRNA and the gene sequences were derived from different stocks of Sprague-Dawley rats. The sequence of elastase II mRNA is identical to that of the exons in \( \lambda E2 \) (see Fig. 5).

Table I summarizes the intron positions within the coding sequences of elastase I and II mRNAs. The amino acid lengths encoded by the first and last exons are short, 13% and 5 amino acids, respectively, but also include mRNA untranslated regions. The amino acid coding lengths of the five internal exons are fairly uniform, varying between 27% and 50 amino acids. The intervening sequences in the two elastase genes occur at corresponding positions, with one exception. In this instance, the position of intron 3 in elastase II is displaced one codon upstream.

In contrast to the conserved amino acid coding exons that share 62% nucleotide-sequence identity (MacDonald et al., 1982a), the third intron is not conserved between elastase I and II. The intron positions and their relative lengths are summarized in Table I.

![Restriction-endonuclease map and exon positions of the two pancreatic elastase genes.](image)

S. Clift and R. J. MacDonald, unpublished data.
to elastase I, but which has substitutions at key amino acid positions that should significantly alter amino acid cleavage preferences (MacDonald et al., 1982a).

The two elastase genes contain intervening sequences at identical positions with the exception of the position of intron 3. The precise positions of the introns are determined by comparing the known mRNA sequences with the interrupted gene sequences reported here and by applying the Breathnach-Chambon rules which require GT at the 5' end and AG at the 3' end of each intervening sequence (Breathnach and Chambon, 1981). The splice points that are predicted for the elastase genes are also consistent with optimum-potential base pairing across intronic sequences with U1 small nuclear RNA as proposed by Lerner et al. (1980) and Rogers and Wall (1980).

The position of the junction between exons 3 and 4 differs by three nucleotides between elastase I and II. Even when splice-site redundancy is taken into account, it is clear that the intron position has shifted. The position of intron 3 in the genes for the two rat trypsins (Craik et al., 1984), rat chymotrypsin (Bell et al., 1984), and mouse kallikreins (Mason et al., 1983) is identical to that of the elastase I gene. Therefore, we propose that the position of this intron in the ancestral serine protease gene corresponds to that of elastase I. The movement of the position of this intron in elastase II requires the sliding of both the upstream (donor) and downstream (acceptor) splice junctions one codon toward the 5' end of the gene.

_Intron Boundary Sliding—_Intron movement by junction sliding might involve a two-step mechanism. First, the sliding (as a result of mutations) of one exon/intron boundary three nucleotides upstream would maintain the reading frame and generate an encoded protein with one more or one less amino acid. The second step must include a compensatory sliding of the other exon/intron boundary of the pair by three nucleotides in the same direction. The nucleotide sequences retained at the 5' and 3' boundaries of the third intron of the elastase I and II genes are consistent with this sliding mechanism of intron movement.

Recently, Craik et al. (1983) correlated internal-polypeptide length differences among serine proteases with the position of intervening sequences, which suggested that sliding of one of a pair of junctions was responsible for the loss or acquisition of polypeptide domains during the evolution of this gene family. The apparent movement of the position of intron 3 of elastase II, which requires the movement of both members of the pair of splice junctions, is additional evidence that boundary sliding did occur in duplicated members of the serine protease gene family. Similar evidence can be found for boundary sliding in other gene families as well (Craik et al., 1983; Tucker et al., 1981) and may be a general mechanism for protein evolution in eukaryotes.

_Exon-encoded StructuralDomains—_Several of the eight exons of each elastase gene encode structural or functional domains of the enzyme (Figs. 4 and 5). The first exon contains the signal peptide, a common feature of gene organization for secretory proteins (Breathnach and Chambon, 1981). The second elastase exon encodes the activation peptide and the amino-terminal 16 amino acids of the mature (active) enzyme. The activation peptide and the amino-terminal 16 amino acids of the mature (active) enzyme. The activation peptide and the amino-terminal 16 amino acids of the mature (active) enzyme. The activation peptide and the amino-terminal 16 amino acids of the mature (active) enzyme. The activation peptide and the amino-terminal 16 amino acids of the mature (active) enzyme.

**DISCUSSION**

We have described the sequence organization of two rat genes encoding elastase-like serine proteases. Both are single-copy genes selectively expressed to high levels in the exocrine pancreas as part of the complement of digestive hydrolytic enzymes. The product of the elastase I gene is the rat counterpart of the well-characterized porcine pancreatic elastase (MacDonald et al., 1982a; Largman, 1983). The second pancreatic elastase gene encodes a serine protease closely related...
Pancreatic Elastase Gene Structure

4. Partial sequence of the elastase I gene. Nucleotide sequences are shown for the eight exons and adjoining intron regions. The proposed TATAA box is overlined. The proposed mRNA cap site (MacDonald et al., 1982a) is denoted by +1. Exon/intron boundaries are determined from the known mRNA sequence and the GTP/AG rule. Strings of colons indicate gaps in the sequence. The numbering of the amino acid sequence starts at the proposed amino terminus of the mature, active enzyme. The under lined nucleotides are corrections of the mRNA sequence previously reported (MacDonald et al., 1982a); the single nucleotide difference between the cloned cDNA and the gene is circled.
Fig. 5. Partial sequence of the elastase II gene. Legend is as for Fig. 4.
McLachlan (1979) has proposed that the progenitor serine protease gene was derived from the duplication of a genetic region encoding a three-β sheet motif, followed by a second duplication that yielded a complete primitive enzyme comprising two similar hydrophobic domains. This evolutionary scheme involves recruitment of genetic information through exon duplication. The internal five exons of the elastase gene, however, only partly correlate with the four subdomains containing three β sheets each. While exon 3 encodes the three-β sheet subdomain nearest the amino terminus and exon 7 encodes the carboxyl-terminal subdomain, the correspondence between the three-β sheet subdomains and exons is disrupted in the three internal exons. This could mean that the genesis of the progenitor serine protease gene did not involve duplications of a single exon encoding a three-β sheet structural motif. However, the loss or acquisition of introns during evolution could obscure the original exon/intron organization of the progenitor gene that indeed may have arisen through exon duplications. Comparison of the organization of chymotrypsin (Bell et al., 1984), trypsin (Craik et al., 1984), kalikrein (Mason et al., 1983), and elastase genes reveals that introns have been either gained or lost (or both) during the divergent evolution of these members of the pancreatic serine protease gene family. Chymotrypsin has one fewer and trypsin and kalikrein have three fewer introns than the elastases. Of the four remaining introns, only one is located at precisely the equivalent position for all five serine protease genes; the other intron positions vary by one to five codons among the serine proteases, similar to the altered position of intron 3 in the two elastases. Thus, introns appear to have been mobile in this gene family. It is likely that changes in the serine protease gene organization have obscured the original exon/intron arrangement. Therefore, the genesis of the primordial serine protease gene through exon duplication remains a possibility. A phylogenetic analysis of the organization of pancreatic serine protease genes might reveal other exon/intron arrangements compatible with the postulated duplication events.

**Sequence Similarities in 5’ Flanking Regions**—In the 5’ flanking regions that may be expected to harbor regulatory sequences, the two elastase genes share notable sequence similarities (Fig. 6). Although a number of gaps must be

### Table 1

Introns within the amino acid coding regions of elastase I and II genes

<table>
<thead>
<tr>
<th>Introns</th>
<th>Elastase I</th>
<th>Elastase II</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-13(-3)</td>
<td>-17(-3)</td>
</tr>
<tr>
<td>2</td>
<td>15/16(30/31)</td>
<td>16/16(30/31)</td>
</tr>
<tr>
<td>3</td>
<td>49/61</td>
<td>48/61</td>
</tr>
<tr>
<td>4</td>
<td>91/100b</td>
<td>91/100b</td>
</tr>
<tr>
<td>5</td>
<td>137(146)</td>
<td>137(146)</td>
</tr>
<tr>
<td>6</td>
<td>185/186(192/193)</td>
<td>185/186(192/193)</td>
</tr>
<tr>
<td>7</td>
<td>235/236(240/241)</td>
<td>236/237(240/241)</td>
</tr>
</tbody>
</table>

**Fig. 6. Comparison of the 5′ flanking sequences of the elastase I and II genes.** Gaps are introduced to optimize sequence homology. Asterisks indicate identical nucleotides. The imperfect repeats at the ends of the GT-rich segments are underlined.

**Fig. 7. Related upstream sequences of pancreatic serine protease genes.** The relative distances of these conserved sequences from the start of transcription for each of the genes are indicated by the negative numbers centered above one of the nucleotides common to each type of sequence. For the consensus sequence, x denotes no nucleotide preference, lower case denotes 3/5, upper case 4/5, and the underlined upper case 5/5. Underlining indicates the positions of potential enhancer core sequences; the gaps in the underlining indicate differences with the proposed core sequence GTGC<sub>AAA</sub> TTTG.
inserted to optimize the alignment, the two genes share 39% sequence relatedness up to about 450 nucleotides upstream from the start of transcription. This degree of relatedness may be compared to the high (62%) nucleotide-sequence conservation between the two elastase amino acid-coding domains and the absence of significant relatedness between the sequenced introns of the two genes, between the 3' untranslated regions of the two mRNAs, or between regions further upstream of the genes. Included within the partly conserved upstream sequences are GT-rich regions. The perfect repeating copolymer (GT)$_{24}$ upstream of the elastase I gene might be capable of Z-DNA conformation, whereas the 55-nucleotide GT-rich stretch upstream of the elastase II gene would not. The regions between the GT-rich domains and the start of transcription contain six oligonucleotide islands 9–21 nucleotides in length with greater than 70% sequence identity.

If these conserved sequences have function, they might be found in the 5'-flanking regions of other pancreatic serine proteases as well. The gene sequences of chymotrypsin B reported by Bell et al. (1984) and of trypsins I and II reported by Craik et al. (1984) were searched for the presence of sequences related to the six conserved islands of the elastases. An oligonucleotide region centered at −111 of elastase I is largely conserved at a similar position for these other pancreatic serine protease genes as well, and the island AGCTTGGG at −217 in elastase I is also present in chymotrypsin B at −354 (Fig. 7). These two oligonucleotide sequences, largely conserved among the five pancreatic serine protease genes, may represent conserved regulatory sequences. We have searched the two elastase genes for potential enhancer core sequences, GTGGTTTTAG, (Weiber et al., 1983). The elastase I gene contains the candidate core sequence in the first intron. The sequenced regions of the elastase II gene do not contain a core sequence. Notably, the conserved regions corresponding to the −111 island of elastase I contain potential enhancer core sequences with single-nucleotide discrepancies for all five pancreatic serine protease genes (Fig. 7). A functional assay for developmental regulatory sequences is now necessary to determine whether these conserved sequences determine the timing, extent, or site of expression of these developmentally regulated genes.

Acknowledgments—We thank Drs. T. Sargent, B. Wallace, and J. Bonner for the Charon 4A rat genomic DNA library, Dr. Hugo Martinez for the DNA analysis programs, Drs. G. Bell and R. Butow for many helpful discussions, and Ms. Marie Rotondi for assisting with the manuscript.

REFERENCES

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