Volume 14 Number 21 1986 Nucleic Acids Research

Sequence organisation and transcriptional regulation of the mouse elastase II and trypsin genes

Brian J.Stevenson, Otto Hagenbüchle and Peter K.Wellauer

Swiss Institute for Experimental Cancer Research, 1066 Epalinges, Switzerland

Received 31 August 1986; Revised and Accepted 7 October 1986

ABSTRACT

Elastase II and trypsin mRNAs were cloned in form of their cDNAs from pancreas of strain A/J mice, and their complete nucleotide sequences were determined. The elastase II mRNA is 912 nucleotides long and encodes a protein of 271 amino acids. The cloned trypsin mRNA species is 814 nucleotides long and encodes a protein of 246 amino acids. The elastase II gene, which exists as a single copy in the haploid mouse genome, measures 11.2 kb from cap to poly(A) site and is interrupted by at least seven introns. Between 5 and 10 trypsin genes exist in the mouse genome. Five different trypsin genes, two of which are closely linked in a tail-to-tail manner, were studied in detail. They vary in size between 3.4 and 4.0kb, and all are interrupted by four introns. DNA sequence comparison of the elastase II, trypsin and Amy-2^a α-amylase genes reveals a conserved 13 nucleotide motif in their 5'-flanking regions. The differential accumulation of the elastase II and trypsin mRNAs in the cytoplasm of the acinar pancreatic cell is regulated predominantly at the transcriptional level.

INTRODUCTION

The cytoplasm of pancreatic cells of mouse contains a limited number of highly abundant mRNA species which encode pancreas-specific proteins [1]. We have shown that multiple genes encoding pancreatic α-amylase are regulated primarily at the transcriptional level by strong pancreas-specific promoters [2]. The high level accumulation of α-amylase mRNA in the cytoplasm of the exocrine pancreatic cell is thus the consequence of a combination of gene dosage and promoter efficiency. In order to investigate whether a common regulatory mechanism exists for the expression of pancreas-specific genes we decided to study the expression of additional genes which are active in this tissue, namely those encoding elastase II and trypsin.

Elastase and trypsin belong to a family of serine proteases which are selectively expressed to a high level in the exocrine pancreas of mammals [3]. They are digestive enzymes which are synthesized as inactive precursors [4, 5], containing a hydrophobic "signal" peptide sequence at the amino-

terminus which directs the nascent polypeptide across the membrane of the endoplasmic reticulum [6].

In this paper we present the sequences of the mRNAs and the structural organisation of the genes for mouse elastase II and trypsin. We show that these genes are highly expressed in the mouse pancreas and that they are regulated predominantly at the transcriptional level.

MATERIALS AND METHODS

Animals

Mice of the inbred strain A/J, 2 to 4 months old, were used in all experiments.

Extraction of RNA and DNA

Total polyadenylated RNA from pancreas was isolated according to Schibler et al. [7]. Extraction of high molecular weight DNA was as described by Wellauer & Dawid [8].

Cloning and identification of cDNAs for elastase and trypsin

The synthesis of full length, double-stranded cDNA and isolation of clones by colony hybridization was as described by Schibler et al. [7]. Double-stranded cDNA bands of 0.8 and 0.9kb were eluted from a 2% neutral agarose gel, and the cDNAs were cloned into the Pst I site of the pBR322 vector using the G-C tailing method. Two distinct classes of cDNA clones, representing abundant pancreatic mRNA species, were identified by restriction enzyme digestion. Clones pMPe7 and pMPt9, which contained the longest inserts in their respective class, were selected for further characterisation. Each clone was used to hybrid select mRNAs which were subsequently translated in vitro in the rabbit reticulocyte lysate system. In vitro translation products were partially digested with S. aureus protease V8 (Miles), and cleavage products were compared to those of commercially available pancreatic proteins on one-dimensional SDS-acrylamide gels [7]. A similar digestion pattern was obtained for the in vitro translated protein of pMPe7 selected mRNA and porcine elastase, while the translation product of pMPt9 selected mRNA gave bands which comigrated with those of porcine trypsin (data not shown). Subsequent DNA sequence analysis of the cloned inserts confirmed that pMPe7 and pMPt9 encode elastase and trypsin, respectively (see Results section).

Cloning of genomic DNA

Partial Mbo I liver DNA fragments, size selected (15 to 25kb) on a 10-

40% linear sucrose gradient, were ligated into the Bam H1 replacement lambda vector EMBL3 [9]. The recombinant lambda library contained about 3 mouse genome-equivalents of insert DNA. Screening of the recombinant phage library was performed according to Benton & Davis [10], using ^{32}P-labelled DNA probes nick-translated following the protocol of Maniatis et al. [11].

Restriction enzyme analysis and DNA sequencing

Restriction endonucleases were purchased from Boehringer-Mannheim or New England Biolabs and were used according to the suppliers' specifications. Restriction enzyme sites were mapped using the procedures of Smith & Birnstiel [12] and Schibler et al. [13]. DNA restriction fragments were sequenced by the enzymatic technique of Sanger et al. [14]. The 5'-terminal residues of the elastase and trypsin mRNAs were determined by using the Maxam & Gilbert [15] sequencing technique on cDNA generated by reverse transcriptase catalyzed primer extension [16]. The extended cDNA for elastase gave a unique DNA sequence. However, in several experiments the sequence ladder obtained for trypsin was ambiguous, with one main pattern superimposed on others. This suggests that the pMPt9-specific primer hybridizes to more than one trypsin mRNA (data not shown).

Electron microscopy

To form hybrids between cloned genomic DNA and mRNA, intact recombinant phage DNA or restriction fragments were denatured and hybridized in solution to total polyadenylated RNA from pancreas as described previously [13, 17]. Hybrids were spread for electron microscopy as described by Wellauer & Dawid [8]. Single- and double-stranded ϕX174 DNAs (5375 bases/base-pairs) were used as internal size standards.

Run-on transcription assays in isolated nuclei in vitro

Nuclei were purified from pancreas according to the method of Hewish & Burgoyne [18] as modified by Hagenbüchle et al. [19]. The in vitro elongation reactions were carried out according to Schibler et al. [20] in a cocktail containing 100mM Tris-HCl (pH 7.9), 50mM NaCl, 0.4mM EDTA, 0.1mM PMSF, 1.2mM DTT, 350mM $(NH_4)_2SO_4$, 1mg/ml heparin sulphate, 4mM $MnCl_2$, 2.5µM [α-^{32}P]UTP (400 Ci/mmol), 1mM each of ATP, CTP and GTP, 29% glycerol, 10mM creatine phosphate, 1600 U/ml RNasin, and about 10^8 nuclei/ml. Incubation was at 26°C for 15 mins. Labelled RNA was extracted from the reaction mixture and hybridized to filters as described by Hagenbüchle et al. [19]. To eliminate hybridization to repetitive sequences, large amounts of sheared, denatured mouse DNA (250µg/ml) were included during prehybridization and hybridization [20].

Other techniques

The procedure used for dot blot hybridization has been described [20]. Blotting of DNA was done as described previously [13, 21]. The subcloning of DNA restriction fragments into bacterial plasmid or filamentous phage vectors has been reported by Schibler et al. [7, 20].

RESULTS

Primary structures of the elastase and trypsin mRNAs.

We have isolated and characterised two recombinant cDNA clones, pMPe7 and pMPt9, which represent abundant pancreatic mRNA species. Preliminary analysis suggested that pMPe7 contained elastase cDNA and pMPt9 trypsin cDNA (see Materials and Methods). To confirm this tentative assignment, both recombinant clones were mapped for restriction enzyme sites. Appropriate restriction fragments spanning the entire cDNA insert of each clone were then subcloned in phage M13 vectors and their sequences determined for both strands. The two cDNA clones were found to be nearly complete copies of their respective mRNA. Primer extended cDNA synthesis was used to obtain those few 5'-terminal nucleotides of the mRNAs not present in the cDNAs (see Materials and Methods).

The complete sequence of the mRNA deduced from the pMPe7 sequence, together with the derived hypothetical amino acid sequence, is shown in Figure 1. The mRNA contains 912 nucleotides, of which 21 nucleotides are 5'-noncoding region and 78 nucleotides 3'-noncoding region. The latter region contains a UAA stop codon, and an AAUAAA polyadenylation signal 17 nucleotides upstream of the polyadenylation site [22]. The coding region contains 813 nucleotides of open reading frame which encode 271 amino acid residues and specify a protein with a molecular weight of 28 913 daltons. The identification of pMPe7 mRNA as being the one encoding preproelastase is based on the following considerations. The presence and spacing of his-75, asp-123 and ser-218 in the mouse protein is characteristic of the charge relay system found in serine proteases of other animals [23]. The mouse protein is 58% homologous to porcine elastase I [24], and 48% homologous to the recently published partial sequence of porcine elastase II [25]. However, since the glycine residue at amino acid position 240 and the serine residue at amino acid position 252 are expected to be at the entrance to the substrate binding domain in the folded protein [26], the specificity of the mouse enzyme is expected to be more chymotrypsin-like than that of elastase I. This is consistent with the mouse enzyme being a type II elastase

Figure 1: Nucleotide sequence of elastase II mRNA from the pancreas of strain A/J mice. The sequence was deduced from sequencing of pMPe7 cDNA (positions 35–912) and from primer extension analysis (positions 1–34). X denotes bases not resolved on the sequencing gel of the primer extended cDNA. The translation start and stop codons, and the polyadenylation signal are boxed. The derived amino acid sequence for the preproenzyme is displayed above the nucleotide sequence. The inferred amino terminal residue (valine) of the mature protein chain, based on homology with the porcine enzyme [24], is underlined. Amino acid residues in the mouse protein which differ from those of rat preproelastase II [28] are designated by an asterisk. Residues mentioned in the text are designated by black bars above the protein sequence.

Figure 2: Nucleotide sequence of a trypsin mRNA from the pancreas of strain A/J mice. The sequence was deduced from sequencing of pMPt9 cDNA (positions 29-814) and from primer extension analysis (positions 1-28). X denotes bases not resolved on the sequencing gel of the primer extended cDNA. The translation start and stop codons, and the polyadenylation signal are boxed. The derived amino acid sequence for the preproenzyme is displayed above the nucleotide sequence. The inferred amino terminal residue (isoleucine) of the mature protein, based on homology with the porcine enzyme [29], is underlined. Amino acid residues in the mouse protein which differ from those of both rat pretrypsinogens I and II [31] are designated by an asterisk. Residues mentioned in the text are designated by a black bar.

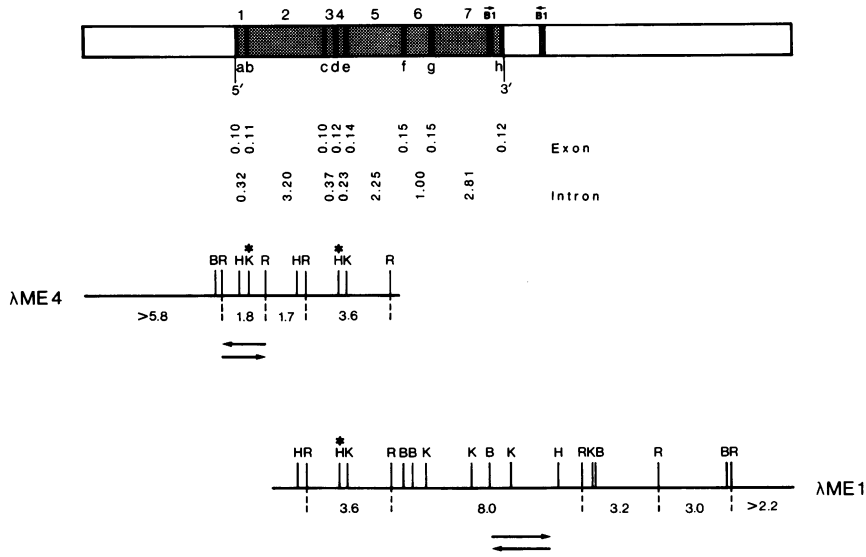

Figure 3: Sequence organisation of the elastase II gene of mouse. Restriction enzyme maps of the inserts of two overlapping clones, λME4 and λME1, are shown: Bam HI (B); Eco RI (R); Hind III (H); Kpn I (K). The sizes in kb of Eco RI fragments are given below the maps. Exonic restriction sites are designated by an asterisk. The extent of DNA sequence information is denoted by arrows below the restriction maps. The exon-intron map of the elastase II gene is presented above the cloned DNA inserts. It was established by measuring 16 to 25 RNA:DNA hybrids in the electron microscope. Exons are represented as black boxes and designated by lower case letters; introns are shown as stippled boxes and are numbered; flanking regions are denoted by open boxes. The sizes in kb of the exons and introns are shown below the map. The 5' and 3' ends of elastase II mRNA are indicated. Hatched regions represent B1 repetitive DNA elements found during sequence analysis of the 3'-terminal gene region; arrows indicate their relative orientations.

[25, 27]. A similar assignment has been made for the protein derived from the rat elastase II mRNA [28]. Indeed, mouse elastase II is 85% homologous to the rat elastase II enzyme (see Fig. 1), but only 58% homologous to rat elastase I [28].

The complete sequence of pMPt9 mRNA and its derived hypothetical amino acid sequence are shown in Figure 2. The mRNA is 814 nucleotides long and contains a short 5'-noncoding region of only 16 nucleotides. The 3'-noncoding region of 60 nucleotides contains a UAG translational stop codon, and the AAUAAA polyadenylation signal 19 nucleotides upstream of the polyadenylation site. The unique open reading frame of 730 nucleotides present in this mRNA encodes a protein of 246 amino acids with a molecular weight of 26 203 daltons. The protein is 80% homologous to, and shares identical

active site residues with porcine trypsin [29]. In addition, the sequence [val-$(asp)_4$-lys-ile], is present at residues 18 to 24 of the mouse protein. This sequence is the substrate for enteropeptidase, which cleaves after the lysine residue in trypsinogen and initiates the zymogen activation cascade in the lumen of the gut [30]. The protein is 88% and 89% homologous to the proteins derived from the rat trypsin I and II mRNAs, respectively [31; see Fig. 2]. Therefore, we conclude that pMPt9 mRNA encodes a trypsin-like enzyme in the mouse.

Structural analysis of the elastase II gene

The genomic DNA sequence specifying elastase II mRNA was isolated from a recombinant lambda phage library containing three genomic equivalents of mouse DNA. A screening of this library with radiolabelled pMPe7 plasmid yielded three positive plaques. The structural arrangement of these three clones harbouring elastase II sequences was then elucidated by restriction enzyme mapping, visualisation of RNA:DNA hybrids in the electron microscope and sequencing of selected genomic DNA restriction fragments.

A summary of the results for two overlapping recombinant clones, λME1 and λME4, which encompass the entire elastase II gene, is presented in Figure 3. The third clone, λME2, which was isolated from the recombinant phage library is identical to λME1 except that it lacks 5.4kb of 3' sequence (data not shown). Detailed exon-intron maps were obtained by measuring hybrid molecules between the elastase II mRNA and cloned genomic DNA in the electron microscope. The elastase II gene is 11.2 kb in size and is interrupted by at least seven intervening sequences which vary in size between 0.23 and 3.2kb. The eight exons, whose lengths vary between 0.1 and 0.15kb, constitute only 8.8 % of the total gene length. The combined size of the exons, as determined by electron microscopy, is 0.99kb, in good agreement with the size of the elastase II mRNA as determined by sequencing (912 nucleotides). In addition to gene sequences, flanking sequences extending 6.5kb upstream and 12kb downstream are present on λME4 and λME1, respectively.

The rationale for cloning pancreas-specific genes is to establish the structural basis for studying their regulation. Therefore, the sequence of the 5'-flanking region, which contains the promoter of the elastase II gene (see below), was determined. The DNA sequence extending 420 nucleotides upstream and 901 nucleotides downstream of the elastase II cap site is shown in Figure 4. This confirms the sequence of the 5'-noncoding region of the elastase II mRNA which was determined by primer extension, and places the

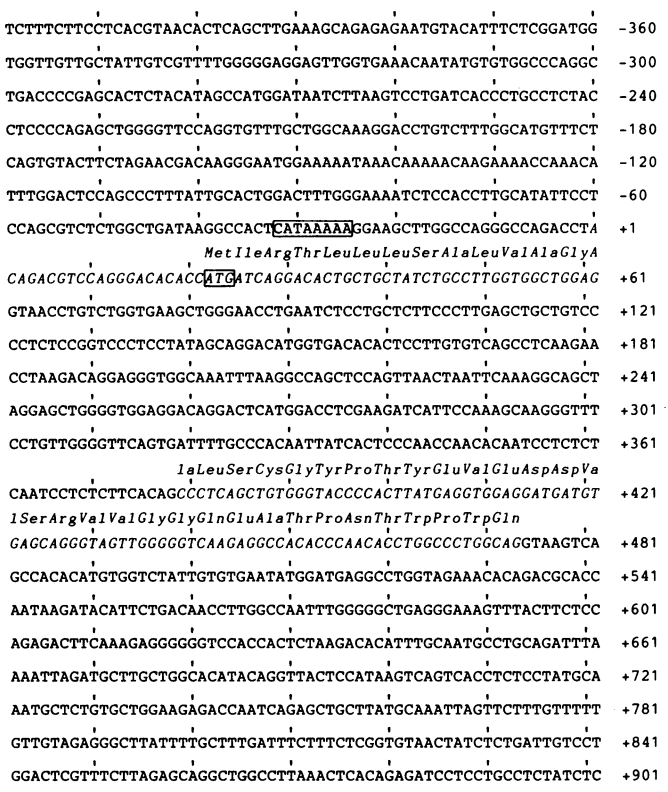

Figure 4: DNA sequence of the 5'-terminal region of the elastase II gene. This sequence was determined from the 1.8kb Eco RI fragment which contains the elastase II cap site (see Fig. 3). Numbering starts at the capped nucleotide (adenine) of the elastase II mRNA. Positive and negative numbers designate gene and 5'-flanking sequences, respectively. Exon sequences are represented by italic letters. The amino acid sequence is shown above the elastase II open reading frame in the first and second exons. The TATA box homology and the translation start codon are boxed.

capped nucleotide (adenine) 21 bases upstream of the AUG translation start codon. A sequence motif (CATAAAAA) reminiscent of the TATA box in front of other genes is present at nucleotide positions -25 to -32. There is also a sequence of high purine content (83%) in the -120 to -150 region. The sequences of the first and second exon sequences are identical to the elastase II mRNA sequence. The first exon, which is only 61 nucleotides long, encodes most but not all of the signal peptide of preproelastase II. The second exon is 95 nucleotides long and encodes the first 29 residues of the mature elastase II protein, including the 14 residue activation peptide

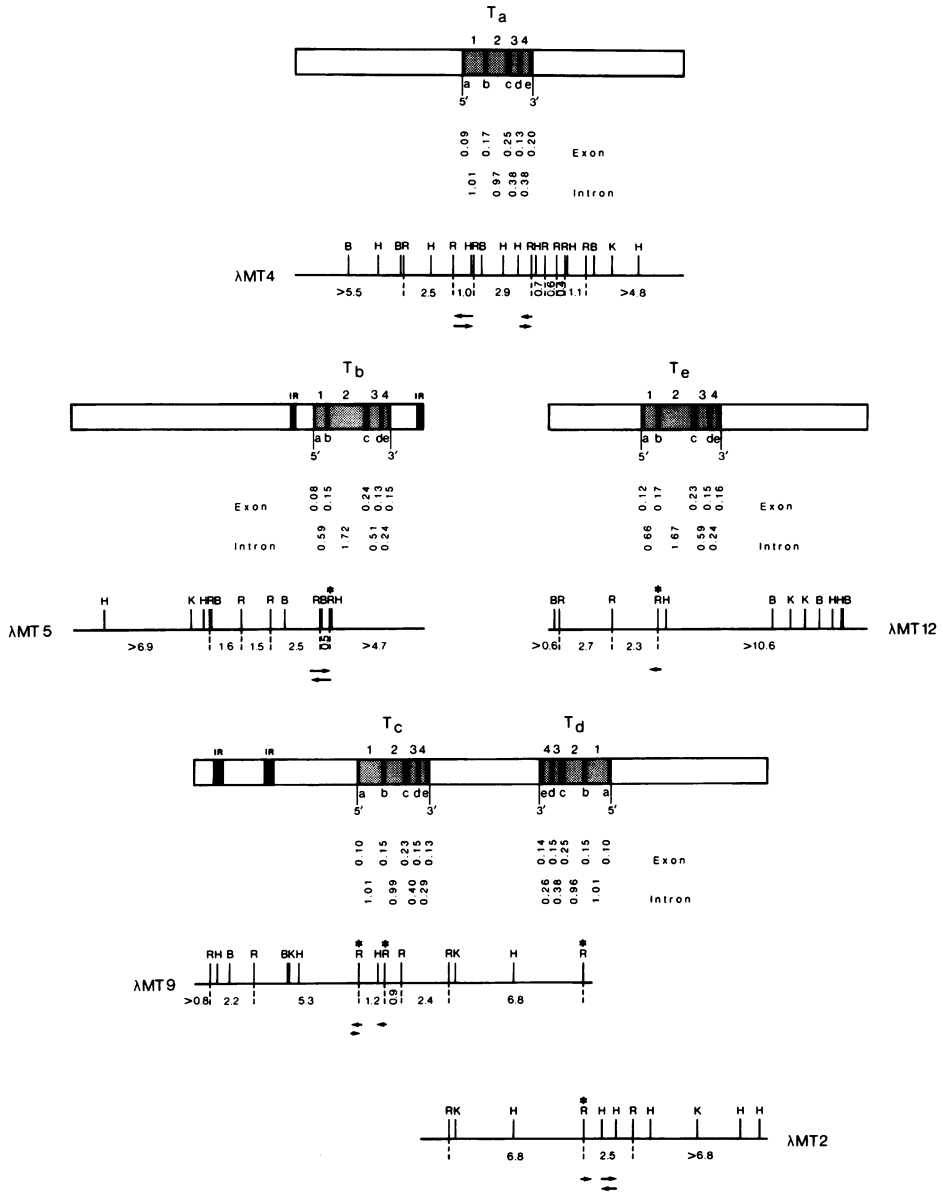

Figure 5: Sequence organisation of five different trypsin genes of mouse. Restriction enzyme maps of the inserts of five recombinant lambda clones are shown: Bam HI (B); Eco RI (R); Hind III (H); Kpn I (K). The sizes in kb of Eco RI fragments are given below the maps. Exonic restriction sites are designated by an asterisk. The extent of DNA sequence obtained for each

strand is denoted by arrows below the restriction maps. The exon-intron maps of the five trypsin genes (T_a to T_e) are presented above the cloned DNA inserts (λMT4 to λMT12, respectively). They were established by measuring 10 to 20 RNA:DNA hybrids in the electron microscope. Note that the overlapping clones, λMT9 and λMT2, contain a partial trypsin gene in addition to a complete one. Their relative orientation has been determined as described in the text. Exons are represented as black boxes and designated by lower case letters; introns are shown as stippled boxes and are numbered; flanking regions are denoted by open boxes. The sizes of the exons and introns are given in kb. The 5' and 3' ends of trypsin mRNA are indicated. Hatched regions represent inverted repeats (IR) seen in the electron microscope.

which may remain associated with the active enzyme through an interchain disulphide bond [25]. Both the first and second intron donor, and first intron acceptor sites follow the GT-AG rule for intron splicing [32].

A region containing part of intron 7, all of exon h (which specifies the carboxy terminal 4 amino acids of the protein and the 3'-noncoding region of the mRNA), and about 2kb downstream of the polyadenylation site was also subjected to sequence analysis (data not shown). The 93 nucleotide exon sequence was found to be identical to that of the 3' end of the elastase II mRNA. Two regions with high homology to the published sequence of the mouse B1 element [33] are present, in opposite orientations, in the last intron and about 1.5kb downstream of the polyadenylation site. The location of these B1 elements is shown in Figure 3.

The haploid copy number of the elastase II gene was determined by Southern and quantitative dot-blot analysis which indicate that there is a single elastase II gene per haploid mouse genome (data not shown).

Multiple trypsin genes exist in the mouse genome

About 30 to 40 positive plaques were obtained when radiolabelled trypsin cDNA plasmid pMPt9 was used to screen the recombinant lambda library. Of the twenty phage plaques selected for further screening, nineteen contained sequences that hybridized to the cloned trypsin cDNA. These positive clones were divided into four groups according to their restriction enzyme patterns. Each group contained at least two clones with overlapping DNA inserts.

The structural arrangement of five different recombinant lambda clones, which are representative of the four classes that contain trypsin sequences, was deduced from restriction enzyme mapping and from the analysis of RNA:DNA hybrids in the electron microscope. A summary of the restriction endonuclease and electron microscopy data is presented in Figure 5.

The five cloned trypsin genes vary in length between 3.4kb (T_d) and

4.0kb (T_e). Each gene is interrupted by at least four intervening sequences, some of which vary in size between different genes. This variability in intron length accounts for most of the observed gene length heterogeneity and provides a means by which two trypsin gene classes can be distinguished. T_a, T_c and T_d belong to one class of trypsin genes whose largest intron is intron one, while T_b and T_e belong to a second class whose largest intron is intron two.

No hybrid formation was observed in the electron microscope when single strands of two overlapping clones, λMT9 and λMT2, were hybridized to pancreatic mRNA, due to extensive self-annealing of the DNA. This resulted in the formation of characteristic intramolecular stem and loop stuctures (data not shown). Since both clones hybridize to trypsin cDNA, the stem and loop structures suggested the existence of two closely linked trypsin genes which are located on opposite strands in the genome. The orientation of the two closely linked genes was elucidated by hybridizing a 5' cDNA probe specific for the first and second exons of trypsin genes to filter bound restriction fragments of λMT2 (data not shown). Only one 5' end, that of the complete copy, was detected. These results are consistent with the existence of two closely linked trypsin genes, termed T_c and T_d, arranged in a tail-to-tail configuration in the mouse genome (Fig. 5).

The five cloned trypsin gene copies were further characterised by DNA sequence analysis. The regions sequenced, and the extent of DNA sequence obtained, are denoted by arrows below each lambda clone in Figure 5. A comparison between the DNA sequences of T_a, T_b, T_c and T_d, extending from about 300 nucleotides upstream of each cap site, through the first exon and into the first intron, is presented in Figure 6. The various trypsin genes have a colinear DNA sequence extending from the 5' boundary of the first intron (+70) through the conserved TATA motif 28 nucleotides upstream of the cap site. Beyond this point gaps must be introduced to maximize homology, which becomes random upstream of nucleotide position -200. A region of high homology, flanked by purine rich sequences, is revealed in the -90 to -110 region. Only the first exon (Fig. 6) and the fifth exon (data not shown) of T_a are identical in sequence to the cDNA sequence of pMPt9. Furthermore, duplexes between T_a and pMPt9 were stable under conditions (0.165M salt; 73°C) which were sufficient to melt duplexes between pMPt9 and the four other trypsin genes (data not shown). These data strongly suggest that the cDNA clone pMPt9 is derived from a trypsin mRNA transcribed from the T_a gene. The sequence of the T_a gene confirms the sequence of the 5'-noncoding

Figure 6: Comparison of 5'-terminal sequences of four trypsin genes of mouse. Sequences of the 5'-terminal regions of the T_b (b), T_c (c) and T_d (d) genes are compared to that of the T_a (a) gene. Numbering for each gene sequence starts at the position of the capped nucleotide (adenine) of the T_a trypsin mRNA. Positive and negative numbers designate gene and 5'-flanking sequences, respectively. Hyphens indicate identity to the T_a sequence. Note that spaces have been introduced to maximize homology between the sequences. Exon sequences are represented by italic letters. The amino acid sequence is derived from the open reading frame in the first exon of the T_a gene. Nucleotide changes in T_b, T_c or T_d which would give rise to amino acid residues different from the T_a protein are boxed. The TATA motif and the translation start codon are also boxed.

region of the trypsin mRNA which was determined by primer extension, and places the capped nucleotide (adenine) 16 bases upstream of the AUG translation start codon.

T_b, T_c and T_d, if transcribed, would specify mRNAs and proteins which differ in sequence from those of T_a. These differences are indicated in Figure 6. However, the length of the first exon appears to be identical

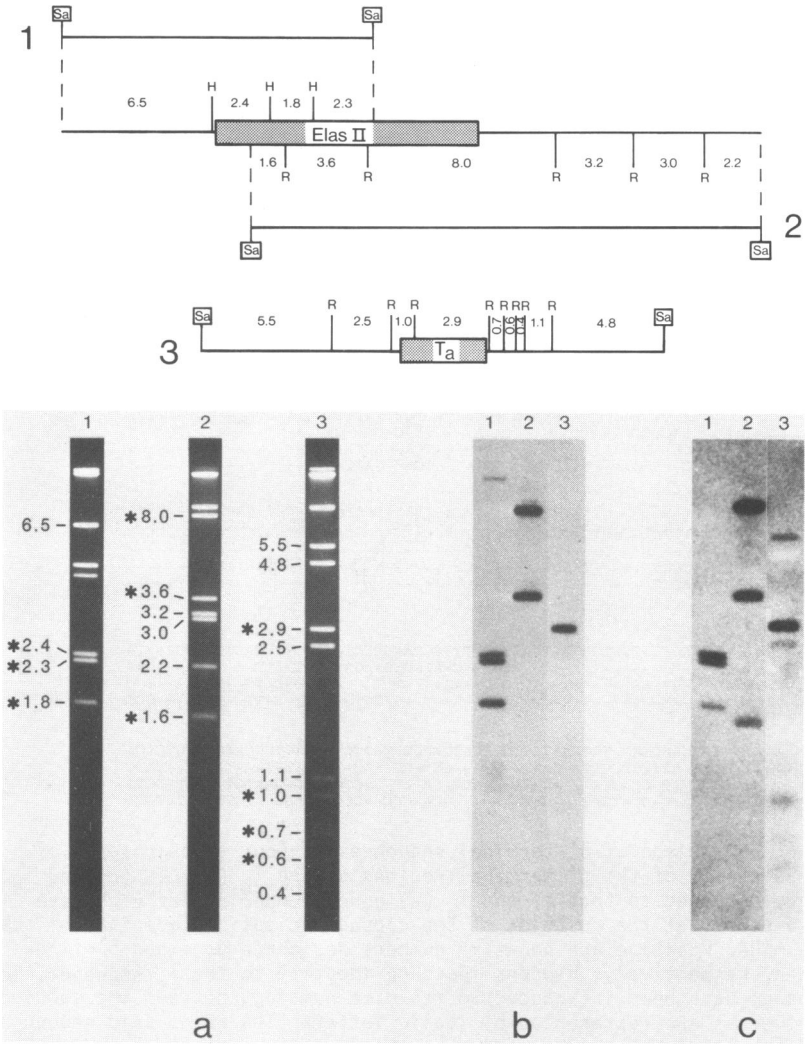

Figure 7: Transcription analysis of the elastase II and trypsin genes. DNA restriction fragments of the elastase II and T_a trypsin genes were separated by electrophoresis on a 1% agarose gel. DNA fragments were transferred to nitrocellulose filters (Schleicher and Schuell) and hybridized to (b) 5 x 10^5 cts/min of ^{32}P-labelled, nick-translated elastase II (lanes 1,2) or trypsin (lane 3) cDNA or (c) 10^7 cts/min of nascent RNA chains elongated in vitro in the presence of [α-^{32}P]UTP in isolated pancreatic nuclei. The EtBr stained gel is shown in (a). Lane 1, Hind III / Sal I digest of 2µg of λME4 DNA containing the 5'-terminal part of the elastase II gene; lane 2, Eco RI / Sal I digest of 2µg of λME1 DNA containing the 3'-terminal part of the elastase II gene; lane 3, Eco RI / Sal I digest of 2µg of λMT4 DNA contaning the entire T_a trypsin gene. The size of the fragments is given in kb; those

which give a clear signal in (c) are designated by an asterisk. The origin of the fragments is shown in the maps at the top: Hind III (H); Eco RI (R). Boxed Sal I (Sa) sites cut within the polylinker of the EMBL3 vector [9]. Gene sequences are represented by stippled boxes.

between the genes, namely 56 nucleotides. This short exon encodes most but not all of the signal peptide of pretrypsinogen.

The number of trypsin genes per haploid mouse genome was determined by quantitative dot-blot hybridization, using a restriction fragment of pMT9 cDNA as a probe (data not shown). The results from this experiment indicated that 4 to 5 trypsin gene copies exist per haploid mouse genome. However, qualitative Southern blot analysis of mouse genomic DNA reveals 10 Eco RI restriction fragments, of which five can be accounted for by the cloned trypsin DNAs (data not shown). From this data we conclude that the value obtained by dot-blot analysis is an underestimate of the trypsin gene copy number, probably due to non-uniform hybridization efficiencies between the cDNA probe and various trypsin genes, whose exons exhibit sequence heterogeneity. Therefore, we estimate that there are between 5 and 10 trypsin genes per haploid mouse genome.

Transcriptional analysis of the elastase II and trypsin genes

The cloned genomic sequences and the cDNAs were used to study the expression of the elastase II and trypsin genes. Since T_a is the only trypsin gene among the cloned copies which has been shown to be transcribed in the mouse pancreas, its DNA was used in all experiments.

The regions of the genomic clones which harbour elastase II and trypsin mRNA sequences were determined by hybridizing radiolabelled elastase II or trypsin cDNAs to filter bound restriction fragments of the appropriate genomic clones. This experiment localises the exon containing regions within the cloned DNAs. The results obtained by hybridizing ^{32}P-labelled elastase II or trypsin cDNA to gel-fractionated restriction fragments of λME4 and λME1 (elastase II) or λMT4 (trypsin), respectively, are shown in Figure 7. Only the 2.4, 1.8 and 2.3kb Hind III fragments of λME4 and the 3.6 and 8kb Eco RI fragments of λME1 hybridize and therefore contain elastase II exon sequences. The gene internal 1.6kb Eco RI fragment of λME1 does not hybridize and, therefore, contains only intron sequences.

A single Eco RI fragment (2.9kb) of λMT4 hybridizes with the trypsin cDNA probe. The 1kb fragment which spans the T_a cap site gives a signal only on very long autoradiographic exposures (data not shown) because the cDNA contains only 28 nucleotides of hybridizable first exon sequence.

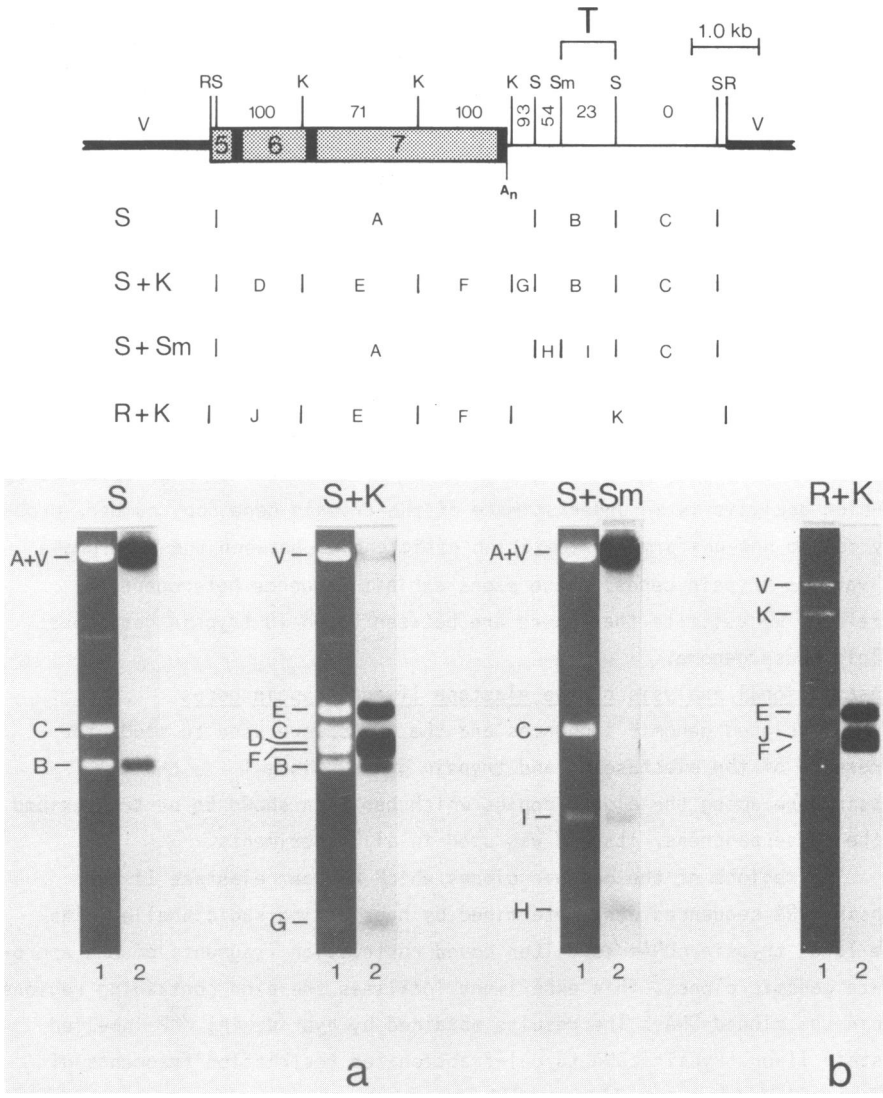

Figure 8: Analysis of transcription termination in the elastase II gene. Nascent RNA chains elongated in vitro in the presence of [α-^{32}P]UTP in isolated pancreatic nuclei (a) or ^{32}P-labelled, nick-translated elastase II cDNA (b) were hybridized to filter bound DNA fragments generated by digesting a recombinant plasmid, containing the 3'-terminal 8kb Eco RI fragment of the elastase II gene, with a variety of restriction endonucleases. DNA fragments (A to K) were separated on a 1% agarose gel. The EtBr stained restriction patterns and the corresponding autoradiograms are shown in lanes 1 and 2, respectively. The origin of the fragments is depicted in the map at the top: Eco RI (R); Kpn I (K); Sac I (S); Sma I (Sm); vector sequences (V). Exons are represented as black boxes; introns are shown as stippled boxes

and numbered. Thin and thick lines represent 3'-flanking and vector sequences, respectively. Numbers above the map are in percent and indicate the relative polymerase loading on each fragment (see the text). The region of transcription termination is indicated (T).

To determine where the polymerase II molecules initiate and terminate transcription *in vivo*, the same filter bound restriction fragments of cloned DNAs were also hybridized to ^{32}P-labelled nascent RNA chains obtained by run-on transcription in isolated pancreatic nuclei *in vitro*. The hybridization signal obtained for each restriction fragment is an indirect measure of the number of polymerase II molecules actively transcribing that fragment [2, 20]. The results of run-on transcription on the elastase II gene are shown in Figure 7. All elastase II restriction fragments which hybridized with the cDNA also give strong signals with the nascent chain probe. As expected, the 1.6kb intron fragment of λME1 now also gives a signal. The 2.4kb fragment, which contains the cap site 21 nucleotides downstream of the 5'-terminal *Hind* III site, hybridizes stoichiometrically with the nascent chain probe. No hybridization is detected with the 6.5kb *Hind* III upstream fragment indicating that transcription initiates at or very close to the elastase II cap site. Transcription termination must occur somewhere within the 8kb *Eco* RI fragment, since fragments further downstream do not hybridize.

To localise more precisely the region(s) of transcription termination in the elastase II gene, the polymerase loading was determined for restriction fragments generated by further digestion of the 8kb *Eco* RI fragment, which had been subcloned in pBR322. Figure 8 shows the result of hybridizing nascent chains to filter bound restriction fragments of this subclone, obtained by digestion with a variety of restriction enzymes. Signal strengths were determined by quantitative densitometry. Values were normalised to that of fragment D, which lies within the body of the gene. The gene internal fragment E gives a lower signal than fragments D or F because of hybrid instability between the nascent RNA chains and the DNA sequence, which is highly AT rich (data not shown). Since fragment C does not give a signal, termination of transcription is expected to occur somewhere in fragment I, whose 5' boundary lies 0.85 kb downstream of the polyadenylation site.

Run-on analysis on the T_a trypsin gene yields a strong signal for the 2.9kb *Eco* RI fragment of λMT4 upon hybridization to nascent chains (Fig. 7). In addition the 1kb *Eco* RI fragment, which contains the trypsin cap site,

TABLE 1. Quantitation of elastase II and trypsin mRNAs and nascent transcripts in the pancreas

cDNA[1]	Hybridization probe (^{32}P)	Total cts/min in hybridization	Cts/min hybridized[2]	Cts/min per kb[3]	No. of molecules per cell[4]	Relative promoter strength[5]
Amylase	End-		16063	10363	(1.0×10^5)	
Elastase II	labelled	5×10^5	3143	3572	3.5×10^4	
Trypsin	poly(A)$^+$		2229	2822	2.7×10^4	
Protein L7	RNA		25	48	4.6×10^2	
Amylase	Nascent		2116	1365		(100)
Elastase II	RNA chains	10^7	175	199		15
Trypsin	elongated		95	120		9
Protein L7	*in vitro*		18	35		3

[1] 4µg aliquots of recombinant cDNA plasmid were heated to 80°C for 3 mins. in 50mM NaOH, neutralised with 1/10 volume 0.5M HCl / 0.5M Tris HCl (pH 8.0) and spotted on Gene-screen (NEN) membranes. Amylase cDNA plasmid is pMPa21 [1]; elastase II and trypsin cDNA plasmids are pMPe7 and pMPt9, respectively (this paper); protein L7 cDNA plasmid is pL7 [49].
[2] Filters were hybridized in formamide solutions at 42°C as described previously [19].
[3] Sizes of cDNA inserts are: pMPa21, 1.55kb; pMPe7, 0.88kb; pMPt9, 0.79kb; pL7, 0.52kb.
[4] The values for elastase II, trypsin and protein L7 mRNAs were calculated using the previously determined number of amylase mRNA molecules per cell (1.0×10^5) as a standard [1]. The number of trypsin mRNA molecules per cell is a minimal estimate since several different trypsin mRNA species may be synthesised in the pancreas (see text).
[5] Promoter strengths were calculated from "cts/min per kb" and are expressed in percent, relative to that of the *Amy*-2^a α-amylase gene. The value for the trypsin promoter may reflect the concerted activity of several genes.

and the 0.7 and 0.6kb *Eco* RI fragments, which are located downstream of the polyadenylation site, yield a distinct signal. No transcription is observed of sequences in the 1.1 or in the 4.8kb downstream fragments. These results suggest that transcription initiates at or near the T_a cap site and continues downstream of the polyadenylation site. The 5.5 and 2.5kb *Eco* RI fragments containing 5'-flanking sequences are weakly transcribed, but to a much lower extent than the 2.9kb *Eco* RI fragment. This low level transcription suggests that some initiation occurs upstream of the cap site. We have not determined whether this is a general phenomenon for all trypsin genes or for only a subset of them. Similar low level transcription has been reported of sequences upstream of the mouse *Amy*-1^a α-amylase cap site [20].

To estimate the relative transcription rate of the elastase II and trypsin genes in the pancreas, we hybridized radiolabelled nascent RNA chains elongated *in vitro* to dot blots containing elastase II and trypsin cDNAs. α-amylase and ribosomal protein L7 cDNAs were included as controls for pancreas-specific and housekeeping functions, respectively. The *Amy*-2^a α-amylase gene has been shown previously to be regulated at the transcrip-

tional level [2]. Therefore, the signals obtained from hybridization to elastase II and trypsin cDNAs were compared to the signal produced from α-amylase cDNA. Nascent transcripts of the elastase II and trypsin genes yield a 7- and 11-fold weaker hybridization signal, respectively, than Amy-2[a] nascent chains (Table 1). These values reflect the different number of polymerase II molecules engaged in transcription and are thus an indirect measure of the relative efficiency of the elastase II and trypsin promoters. The relative abundance of elastase II and trypsin mRNAs in the cytoplasm of the exocrine pancreatic cell was estimated by hybridizing ^{32}P-endlabelled poly(A)$^+$ RNA from pancreas to the respective cDNAs. The hybridization signals obtained in this experiment were again compared to the signal produced from hybridization to α-amylase cDNA, since the steady state level of α-amylase mRNA in the pancreas has been determined previously [1]. The elastase II and trypsin mRNAs both yield approximately 3-fold lower hybridization signals than α-amylase mRNA (Table 1). Since the relative hybridization signals obtained with poly(A)$^+$ RNA and nascent chains are comparable, promoter efficiency is the major factor in determining the steady state levels of elastase II and trypsin mRNAs. The number of ribosomal protein L7 mRNA molecules is almost 20-fold lower than would be expected from the relative promoter strength of the L7 gene. This suggests that the L7 gene is regulated mainly at the post-transcriptional level in the mouse pancreas.

DISCUSSION

The elastase II and trypsin mRNAs of mouse have only about 10% of their lengths occupied by noncoding regions. Short 5'-noncoding regions present on these mRNAs appear to be a common feature of pancreatic mRNAs, since other pancreas-specific mRNAs such as those specifying mouse α-amylase [16], rat chymotrypsin B [34] and carboxypeptidase A [35] all contain short leader sequences. The functional significance of this finding remains unclear since no obvious sequence conservation for hybridization to the 3' end of the 18S rRNA [36] is observed between these mRNA species. It is conceivable that the length of the 5'-noncoding region may influence mRNA stability or translation efficiency. An exception to this rule is the rat RNase mRNA, which contains both long 5'- and 3'-noncoding sequences [37]. However, secretory RNase is also expressed in the rat salivary glands [38], and therefore its mRNA may not be specific to the acinar pancreas.

While the mRNAs for elastase II and trypsin are of similar length, the

transcription units which encode these mRNAs differ greatly in their size and structural organisation. The single copy elastase II gene is 11.2kb long and is interrupted by seven intervening sequences. The exon-intron arrangement closely resembles that of the slightly smaller rat elastase II gene described by Swift and co-workers [39]. Indeed, the mouse elastase II mRNA shares extensive sequence homology with its rat counterpart, and the proteins derived from these mRNAs are 85% homologous (Fig. 1).

As many as ten trypsin genes may exist in the haploid mouse genome. The five cloned copies vary between 3.4 and 4.0kb in length, and all are interrupted by four introns. DNA sequencing shows that the trypsin genes share extensive sequence homology in the first exon and in the immediate 5'-flanking region, but that the extent of homology is close to random upstream of nucleotide position -200 (Fig. 6). Comparing individual gene sequences, however, reveals that T_c and T_d, the two closely linked copies, are more related to each other than to either T_a or T_b. Moreover, the sequenced region of T_b (see Fig. 6) is almost identical to that of T_e, even within intron regions (data not shown). These data suggest that the trypsin gene family has evolved by multiple duplication events from a common ancestor. T_b and T_e may be the products of a recent duplication event, while the two classes of trypsin genes would be the result of a more ancient separation.

Craik and co-workers [40] have isolated two trypsin genes of rat. The exon-intron arrangement of the rat trypsin I gene is very similar to that of T_a, whereas the rat trypsin II gene, which is more than 5kb long, has a structure that is very different from any of the cloned mouse genes. In addition, sequence analysis of the mouse trypsin T_a gene reveals significantly higher homology to the rat trypsin I gene in the noncoding and intronic regions, suggesting that T_a might be the mouse equivalent of the rat trypsin I gene. While at least two trypsin genes are transcriptionally active in the rat pancreas [31, 40], only the mRNA transcript of the T_a gene has been isolated in mouse (Fig. 2). However, the ambiguous pattern obtained from sequence analysis of primer extended cDNA may be the result of cDNA synthesis on more than one trypsin mRNA (data not shown). The sequence heterogeneity detected in the 5'-noncoding regions of various cloned trypsin genes would give rise to trypsin mRNAs with different leader sequences. In addition, RNase treatment reduced the signal given by the hybridization of radiolabelled nascent RNA chains to T_a, but not to elastase II restriction fragments (data not shown). These observations all suggest that more than one trypsin gene may be active in the mouse pancreas.

Mapping of nascent RNA chains indicates that transcription of the elastase II and trypsin genes starts at their cap sites, while termination of transcription occurs about 1 and 1.5kb downstream of their respective polyadenylation sites. Two B1 elements are located on opposite strands in the 3' region of the elastase II gene (Fig. 3). We do not know at the present time whether these B1 elements are transcriptionally active. It is conceivable that transcription termination of the elastase II gene may be influenced by convergent transcription from the B1 element which is located downstream of the termination region on the elastase II non-coding strand. However, the B1 element which is located on the coding strand within the last intron has apparently little or no effect on the polymerase loading of the immediate downstream region. Previous work has shown that termination of transcription in the mouse Amy-2^a α-amylase gene occurs at multiple sites in a region between 2.5 and 4kb downstream of the polyadenylation site [19]. The dilution of polymerase II molecules downstream of the elastase II polyadenylation site suggests that transcription termination in this gene also occurs at multiple sites.

Elastase II and trypsin mRNAs represent abundant transcripts in the pancreas, each accumulating to about 30 000 copies per cell (Table 1). It is not possible to compare directly the strengths of the elastase II promoter and an individual trypsin promoter since we do not know how many trypsin genes are transcriptionally active. However, the single copy elastase II promoter is less than two-fold stronger than the concerted activity of the trypsin promoters as measured by hybridization, and the levels of steady state mRNA can be accounted for by these relative promoter strengths. This suggests that the accumulation of elastase II and of trypsin transcripts in the pancreas is regulated predominantly at the level of transcription. A similar conclusion has been reached for the single Amy-1^a and the multiple Amy-2^a α-amylase genes [2, 20; Table 1]. Thus, transcriptional regulation appears to be a common mechanism for genes which express abundant products of the pancreas. In contrast, the accumulation of rare and moderately abundant mRNAs, which are expressed in several tissues, appears to be regulated mainly at the post-transcriptional level [41]. Our finding, that the low level accumulation of ribosomal protein L7 mRNA in the cytoplasm of acinar pancreatic cells is the result of post-transcriptional regulation, is compatible with this idea.

Several studies have characterised conserved sequences in promoter regions which are involved in the transcriptional regulation of eukaryotic

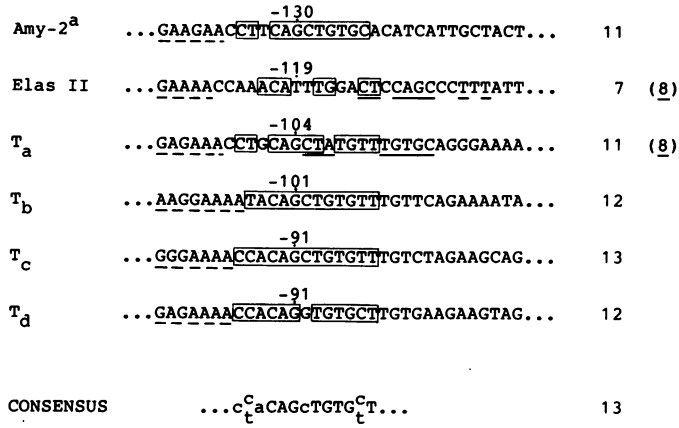

Figure 9: Identification of a conserved sequence motif in the 5'-flanking regions of the Amy-2^a α-amylase, elastase II and trypsin genes of mouse. The distance of these sequences from the cap site is indicated. The consensus sequence of 13 nucleotides was derived by counting the number of identical bases present at each position. Nucleotides which are homologous to the consensus are boxed in the individual sequences. The extent of homology of each sequence to the consensus is indicated on the right. Alternative alignments in the elastase II and T_a genes are designated by solid lines. Purine-rich regions 5' to the sequence motif are denoted by broken lines.

mRNA genes (for a review, see ref. 41). Sequence analysis of the 5'-flanking regions of the elastase II and trypsin genes reveals conserved TATA motifs [43], but no CAAT boxes [44] or Sp1 binding sites [45]. A comparison of the sequences upstream of the Amy-2^a α-amylase, elastase II and trypsin genes reveals a conserved sequence element, at similar positions relative to their cap sites (Fig. 9). A closely related sequence element also occurs in front of several pancreas-specific genes of rat [39]. In the rat chymotrypsin, elastase I and amylase genes it is part of the pancreas-specific transcriptional enhancer [46, 47, 48]. Deletion analysis of the mouse Amy-2^a α-amylase gene has revealed that the region -90 to -160, which contains this conserved element, has the properties of a transcriptional enhancer whose activity is indirectly dependent on glucocorticoid hormones (unpublished observations). Therefore, the activity of the Amy-2^a enhancer appears to play a major role in determining steady state levels of α-amylase mRNA in the exocrine pancreas of mouse. It is conceivable that this conserved element is also involved in the observed transcriptional regulation of elastase II and trypsin mRNA levels. The differential accumulation of α-amylase, elastase II and trypsin mRNAs in the cytoplasm of pancreatic cells may thus be the consequence of differential enhancer activities.

ACKNOWLEDGEMENTS

We thank Dr. U. Schibler and D. E. Kelley for cDNA cloning and DNA sequencing during early stages of this work, and Dr. B. Hirt for his continuing support. We thank also R. Bovey for expert technical assistance and M. Cockell for critical reading of the manuscript. This work was supported by a grant from the Swiss National Science Foundation.

REFERENCES

1. Schibler, U., Tosi, M., Pittet, A.-C., Fabiani, L. and Wellauer, P. K. (1980) J. Mol. Biol. **142**, 93-116.
2. Hagenbüchle, O., Schibler, U., Petrucco, S., Van Tuyle, G. C. and Wellauer, P. K. (1985) J. Mol. Biol. **185**, 285-293.
3. Van Nest, G. A., MacDonald, R. J., Raman, R. K. and Rutter, W. J. (1980) J. Cell. Biol. **86**, 784-794.
4. Freer, S. T., Kraut, J., Robertus, J. D., Wright, H. T. and Xuong, N. H. (1970) Biochemistry **9**, 1997-2009.
5. Fehlhammer, H., Bode, W. and Huber, R. (1977) J. Mol. Biol. **111**, 415-438.
6. Blobel, G. and Dobberstein, B. (1975) J. Cell. Biol. **67**, 835-851.
7. Schibler, U., Tosi, M., Pittet, A.-C., Fabiani, L. and Wellauer, P. K. (1980) J. Mol. Biol. **142**, 93-116.
8. Wellauer, P. K. and Dawid, I. B. (1977) Cell **10**, 193-212.
9. Frischauf, A. M., Lehrach, H., Poustak, A. M. and Murray, N. J. (1983) J. Mol. Biol. **170**, 827-842.
10. Benton, W. D. and Davis, R. W. (1977) Science **156**, 180-182.
11. Maniatis, T., Jeffrey, A. and Kleid, D. K. (1975) Proc. Natl. Acad. Sci. U.S.A. **72**, 1184-1188.
12. Smith, H. O. and Birnstiel, M. L. (1976) Nucl. Acids Res. **3**, 2387-2398.
13. Schibler, U., Pittet, A.-C., Young, R. A., Hagenbüchle, O., Tosi, M., Gellman, S. and Wellauer, P. K. (1982) J. Mol. Biol. **155**, 247-266.
14. Sanger, F., Nicklen, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463-5467.
15. Maxam, A. M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 560-564.
16. Hagenbüchle, O., Bovey, R. and Young, R. A. (1980) Cell **21**, 179-187.
17. Bodary, S., Grossi, G., Hagenbüchle, O. and Wellauer, P. K. (1985) J. Mol. Biol. **182**, 1-10.
18. Hewish, D. R. and Burgoyne, L. A. (1973) Biochem. Biophys. Res. Commun. **52**, 504-510.
19. Hagenbüchle, O., Wellauer, P. K., Cribbs, D. L. and Schibler, U. (1984) Cell **38**, 737-744.
20. Schibler, U., Hagenbüchle, O., Wellauer, P. K. and Pittet, A.-C. (1983) Cell **33**, 501-508.
21. Southern, E. M. (1975) J. Mol. Biol. **98**, 503-517.
22. Proudfoot, N. J. and Brownlee, G. G. (1976) Nature **263**, 211-214.
23. Kraut, J. (1977) Ann. Rev. Biochem. **46**, 331-358.
24. Shotton, D. M. and Hartley, B. S. (1970) Nature **225**, 802-806.
25. Vered, M., Gertler, A. and Berstein, Y. (1986) Int. J. Peptide Protein Res. **27**, 183-190.
26. Shotton, D. M. and Watson, H. C. (1970) Nature **225**, 811-816.
27. Del Mar, E. G., Largman, C., Brodrick, J. W., Fassett, M. and Geokas, M. C. (1980) Biochemistry **19**, 468-472.
28. MacDonald, R. J., Swift, G. H., Quinto, C., Swain, W., Pictet, R. L., Nikovits, W. and Rutter, W. J. (1982) Biochemistry **21**, 1453-1463.

29. Hermodoson, M. A., Ericsson, L. H., Neurath, H. and Walsh, K. A. (1973) Biochemistry **12**, 3146-3153.
30. Maroux, S., Baratti, J. and Desnuelle, P. (1971) J. Biol. Chem. **246**, 5031-5039.
31. MacDonald, R. J., Stary, S. J. and Swift, G. H. (1982) J. Biol. Chem. **257**, 9724-9732.
32. Breathnach, R. and Chambon, P. (1981) Ann. Rev. Biochem. **50**, 349-383.
33. Krayev, A. S., Kramerov, D. A., Skryabin, K. G., Bayev, A. A. and Georgiev, G. P. (1980) Nucl. Acids Res. **8**, 1201-1215.
34. Bell, G. I., Quinto, C., Quiroga, M., Valenzuela, P., Craik, C. S. and Rutter, W. J. (1984) J. Biol. Chem. **259**, 14265-14270.
35. Quinto, C., Quiroga, M., Swain, W. F., Nikovits, W. C., Standring, D. N., Pictet, R. L., Valenzuela, P. and Rutter, W. J. (1982) Proc. Natl. Acad. Sci. U.S.A. **79**, 31-35.
36. Hagenbüchle, O., Santer, M., Steitz, J. A. and Mans, R. J. (1978) Cell **13**, 551-563.
37. MacDonald, R. J., Stary, S. J. and Swift, G. H. (1982) J. Biol. Chem. **257**, 14582-14585.
38. Rabinovitch, M. R., Sreebny, L. M. and Smuckler, E. A. (1968) J. Biol. Chem. **243**, 3441-3446.
39. Swift, G. H., Craik, C. S., Stary, S. J., Quinto, C., Lahaie, R. G., Rutter, W. J. and MacDonald, R. J. (1984) J. Biol. Chem. **259**, 14271-14278.
40. Craik, C. S., Choo, Q.-L., Swift, G. H., Quinto, C., MacDonald, R. J. and Rutter, W. J. (1984) J. Biol. Chem. **259**, 14255-14264.
41. Carneiro, M. and Schibler, U. (1984) J. Mol. Biol. **178**, 869-880.
42. Serfling, E., Jasin, M. and Schaffner, W. (1985) Trends Genet. **1**, 224-230.
43. Goldberg, M. L. (1979) Ph.D. Thesis, Stanford University, U.S.A..
44. Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980) Nucl. Acids Res. **8**, 127-142.
45. Gidoni, D., Dynan, W. S. and Tjian, R. (1984) Nature **312**, 409-413.
46. Walker, M. D., Edlund, T., Boulet, A. M. and Rutter, W. J. (1983) Nature **306**, 557-561.
47. Swift, G. H., Hammer, R. E., MacDonald, R. J. and Brinster, R. L. (1984) Cell **38**, 639-646.
48. Boulet, A. M., Erwin, C. R. and Rutter, W. J. (1986) Proc. Natl. Acad. Sci. U.S.A. **83**, 3599-3603.
49. Meyuhas, O. and Perry, R. P. (1980) Gene **10**, 113-129.