Primary Structure of Human Pancreatic Elastase 2 Determined by Sequence Analysis of the Cloned mRNA†

Thomas S. Fletcher, Wei-Fang Shen, and Corey Largman*

Biochemistry Research Laboratory, Veterans Administration Medical Center, Martinez, California 94553, and Departments of Internal Medicine and Biological Chemistry, University of California, Davis, California 95616

Received May 15, 1987; Revised Manuscript Received June 30, 1987

ABSTRACT: A cDNA encoding elastase 2 has been cloned from a human pancreatic cDNA library. The cDNA contains a translation initiation site and a poly(A) recognition site and encodes a protein of 269 amino acids, including a proposed 16-residue signal peptide. The amino acid sequence of the deduced mature protein contains a 12-residue activation peptide containing a cysteine at residue 1 similar to that of chymotrypsin. The proposed active enzyme contains all of the characteristic active-site amino acids, including His-57, Asp-102, and Ser-195. The S1 binding pocket is bounded by Gly-216 and Ser-226, making this pocket intermediate in size between chymotrypsins and elastase 1 or protease E, consistent with the substrate specificity of elastase 2 for long-chain aliphatic or aromatic amino acids. Computer modeling studies using the amino acid sequence of elastase 2 superimposed on the X-ray structure of porcine elastase 1 suggest that a change of Gln-192 in elastase 1 to Asn-192 in elastase 2 may account for the lower catalytic efficiency of the latter enzyme. In addition, modeling studies have been conducted to attempt to identify basic amino acids in elastases which are absent in chymotrypsins, and which could account for the specific property of elastolysis. Several basic residues appear to be near the ends of the extended binding pocket of elastases which might serve to anchor the enzyme to the elastin substrate. These studies indicate that elastases 2 and elastase 1 both contain an Arg-65A as well as a basic dipeptide at 223/224 which is not present in chymotrypsins. In addition, Arg-217A is present in human elastase 2 but absent in a rat pancreatic protein which has been proposed to be an elastase 2 on the basis of sequence homology [MacDonald, R. J., Swift, G. H., Quinto, C., Swain, W., Pictet, R. L., Nikovits, W., & Rutter, W. J. (1982) *Biochemistry 21*, 1453–1463], but which was not isolated during screening of rat pancreatic tissue extracts for elastolytic activity [Largman, C. (1983) *Biochemistry 22*, 3763–3770].

Elastolytic enzymes have been implicated in diseases such as emphysema (Janoff, 1985) and atherosclerosis (Yamada et al., 1983) in which degradation of elastin fibers is observed. However, there is no clear understanding of the specific structural features which enable elastases to degrade elastin. The vast majority of structure/function studies on elastases have been carried out with pancreatic elastases. Two unique serine proteases with elastolytic activity have been isolated from pig (Hartley & Shotton, 1971; Gertler et al., 1977) and from human pancreas (Largman et al., 1976; Mallory & Travis, 1975). Porcine elastase 1 represents the classically described enzyme which possesses a substrate specificity for small hydrophobic amino acids (Kasafirek et al., 1976), highly basic charge (Hartley & Shotton, 1971), and tertiary structure which appears to contain a restricted S1[1] binding pocket to accommodate the small P1 amino acid side chain (Shotton & Watson, 1970). Although a similar elastase has been isolated from rat pancreas (Largman, 1983), the corresponding protein from human pancreas tissue (protease E) (Mallory & Travis, 1975) possesses little or no elastolytic activity.

We have previously suggested that the major elastolytic enzyme in human pancreatic tissue, elastase 2, is closely related to the chymotrypsin family (Largman et al., 1980). Thus, this highly basic protein possesses a broad specificity for substrates containing medium to large hydrophobic amino acids in the P1 position (Del Mar et al., 1980) and an activation peptide similar to those of the chymotrypsins (Largman et al., 1980). A similar elastase has been isolated from porcine pancreatic tissue (Gertler et al., 1977). Although cDNA cloning studies have indicated that a similar protein exists in rat pancreas (MacDonald et al., 1982), no enzymatic evidence of this protein was observed during fractionation of rat pancreatic tissue extracts (Largman, 1983).

In order to clarify the structural requirements for elastolysis, we have initiated studies on cloning the human pancreatic elastase-like enzymes, with the long-term goal of structure/function studies using site-directed mutagenesis and expression of recombinant enzymes. We have recently reported the cDNA sequence for human pancreatic protease E (Shen et al., 1987). Here we report the isolation of a full-length cDNA clone for human pancreatic elastase 2 and its deduced amino acid sequence. Comparison of human elastase 2 with rat, porcine, and mouse elastases 2 and with chymotrypsins in terms of amino acid homology and proposed three-dimensional structure suggests that the elastases 2 are separate members of the pancreatic serine endopeptidase family. Several residues which may be important for binding of elastases to elastin have been identified by comparison of the amino acid sequences of elastases and chymotrypsins.

† This work was supported by the Research Service of the Veterans Administration and by a grant to C.L. from the National Institutes of Health (HL25408).

* Correspondence should be addressed to this author at the Biochemistry Research Laboratory, Veterans Administration Medical Center.

[1] The nomenclature introduced by Schechter and Berger (1967) is used to describe the positions of amino acids of a substrate. Amino acid residues are numbered P1, P2, P3, etc. in the N-terminal direction from the scissile bond. The corresponding subsites on the enzyme's active site are numbered S1, S2, S3, etc. in an analogous fashion.

0006-2960/87/0426-7256$01.50/0

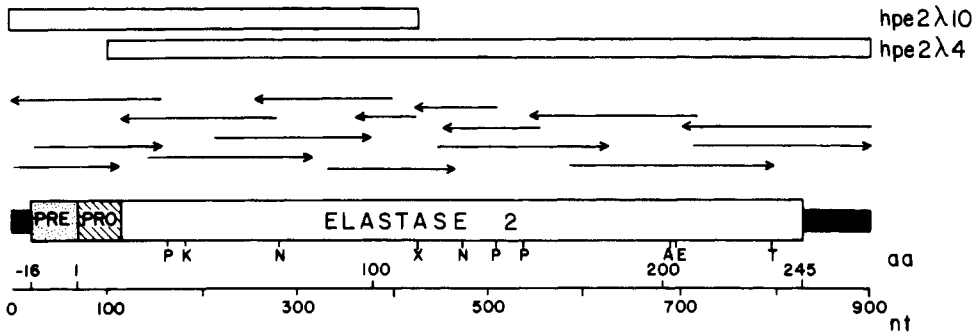

FIGURE 1: Sequencing strategy for human pancreatic elastase 2. The horizontal rectangles at the top of the figure represent the two clones which together yield the complete sequence. The rectangle labeled "Elastase 2" represents the coding region of the cDNA. The regions encoding the signal peptide (PRE) and the activation peptide (PRO) are shaded, and the 5′ and 3′ untranslated regions are represented by a black line. The partial clone hpe2-λ4 was originally isolated from the cDNA library using the rat elastase 2 probe. The direction and length of the sequencing runs are shown by the horizontal arrows. The numbers refer to amino acid residue positions (top) and nucleotide number (bottom). Cleavage sites for some restriction enzymes are denoted with the following abbreviations: A = *Aha*II; E = *Eag*I; K = *Kpn*I; N = *Nae*I; P = *Pst*I; T = *Taq*I; X = *Xho*I.

Experimental Procedures

Materials

A λ gt-11 human pancreatic cDNA library, which was kindly provided by Dr. Robert Weiss, was constructed from poly(A^+) mRNA isolated from human pancreas and ligated into bacteriophage λ gt-11 using *Eco*RI tails. The library contained a total of 10^5 independent clones containing inserts. A cDNA clone which contained the coding region for rat pancreatic elastase 2 was kindly provided by Dr. Ray MacDonald. All other reagents and enzymes were obtained from commercial sources.

Methods

Library Screening. An initial screening of 10^4 insert-containing plaques was performed with an 800 base pair (bp)[2] rat elastase 2 cDNA probe (MacDonald et al., 1982). Low-stringency hybridization conditions were initially used to screen the library (42 °C in 5% formamide, 5× SSPE, 5× Denhardt's solution, 0.1 mg/mL denatured salmon sperm DNA, and 0.1% SDS) to promote identification of elastase-related sequences. Filters were washed with a final wash of 0.5× SSC/0.1% SDS at 42 °C. The longest of seven positive fragments (~800 bp) was subcloned into M13 for sequence determination and was shown to code for an elastase 2 (see below). Since this clone (hpe2-λ4) was lacking a portion of the 5′ region, the library was rescreened with a 172 bp *Eco*RI/*Nae*I fragment of this clone as a probe under more stringent conditions (40% formamide instead of 5% formamide in the above hybridization solution). The final filter wash conditions were 0.1× SSC/0.1% SDS at 42 °C.

Elastase 2 Sequence Determination. The scheme for determining the full-length sequence of elastase 2 is summarized in Figure 1. Initial screening of the library with the rat elastase 2 probe yielded clone hpe2-λ4, which was then used to isolate clone hpe2-λ10. Both clones were sequenced in both directions using M13mp18 or M13mp19 and the dideoxy-^{35}S method described by Bankier and Barrell (1983), with wedge-shaped 5% polyacrylamide/urea gels.

Preparation of Pancreas mRNA and Northern Blotting. Total pancreas RNA was prepared by using the guanidinium thiocyanate procedure described by Chirgwin et al. (1979). mRNA was isolated by passage of total RNA through an oligo(dT) column. Electrophoresis was performed in 1.5% agarose gels containing 6% formaldehyde, 10 mM sodium phosphate, 1 mM EDTA, and 5 mM sodium acetate (pH 7.0). Samples were prepared in 1× gel buffer containing 50% formamide and 2.2 M formaldehyde and heated at 68 °C for 5 min prior to loading. RNA ladders (Bethesda Research Laboratories, Gaithersburg, MD) were run in an identical manner for molecular weight calibration.

N-Terminal Protein Sequence Determination. A sample (10 nmol) of human pancreatic elastase 2 (Largman et al., 1976) was subjected to 20 rounds of automatic Edman degradation using a Beckman spinning-cup automatic protein sequenator in the UC Davis protein structure laboratory. The resulting PTH–amino acid derivatives were identified by high-pressure liquid chromatography.

Computer Modeling. The tertiary structure of porcine elastase 1 was modeled from the X-ray coordinates (Brookhaven Protein Data Bank) using the INSIGHT program (Dayringer et al., 1986).

Genomic Southern Gel. Genomic DNA was prepared from fresh human placental tissue as described by Maniatis et al. (1982). Aliquots of DNA (10 μg) were digested with the appropriate restriction enzyme and subjected to electrophoresis in a 1% agarose gel. Genomic fragments containing elastase 2 sequences were transferred to nitrocellulose (Maniatis et al., 1982) and probed under stringent conditions with a ^{32}P-labeled 800 bp fragment of elastase 2 cDNA (nucleotides 113–906) (Feinberg & Vogelstein, 1983). *Hin*dIII-cut λ molecular weight markers (New England Biolabs) were end-labeled with [α-^{32}P]pCTP according to Maniatis et al. (1982).

Results

Library Screening. Initial low-stringency screening of the human pancreas cDNA library with an 800 bp rat elastase 2 cDNA probe yielded seven positive clones. The longest clone encoded a protein, the N-terminus of which was identical with residues 2–20 of the N-terminal region of human pancreatic elastase 2. This cDNA clone (hpe2-λ4) was 82% homologous to rat elastase 2.

A 172 bp 5′ fragment of clone hpe2-λ4 was used to rescreen the cDNA library at moderate stringency, resulting in identification of six positive clones. Clone hpe2-λ10, which was 427 bp long, was shown by sequence analysis to contain the 5′ portion of the elastase 2 mRNA. As shown in Figure 1, a series of restriction fragments covering the entire cDNA were

[2] Abbreviations: 20× SSPE, 3.0 M NaCl, 0.23 M sodium phosphate, and 0.02 M EDTA (pH 7.4); 20× SSC, 3.0 M NaCl and 0.3 M sodium citrate (pH 7.5); 100× Denhardt's solution, 2% ficoll (M_r 400 000), 2% poly(vinylpyrrolidone) (M_r 360 000), and 2% bovine serum albumin; bp, base pair(s); kb, kilobase(s); nt, nucleotide(s); aa, amino acid(s); SDS, sodium dodecyl sulfate; EDTA, ethylenediaminetetraacetic acid; PTH, phenylthiohydantoin.

```
                          -16                 -10                                 1
                          met ile arg thr leu leu leu ser thr leu val ala gly ala leu ser cys gly asp pro thr tyr pro pro
    1 AAACAGTCCCAGGGACACACC   ATG ATA AGG ACG CTG CTG CTG TCC ACT TTG GTG GCT GGA GCC CTC AGT TGT GGG GAC CCC ACT TAC CCA CCT

          10                                  20                                  30
    tyr val thr arg val val gly gly glu glu ala arg pro asn ser trp pro trp gln val ser leu gln tyr ser ser asn gly lys trp
   94 TAT GTG ACT AGG GTG GTT GGC GGT GAA GAA GCG AGG CCC AAC AGC TGG CCC TGG CAG GTC TCC CTG CAG TAC AGC TCC AAT GGC AAG TGG

          40                                  50                                  60
    tyr his thr cys gly gly ser leu ile ala asn ser trp val leu thr ala ala HIS cys ile ser ser ser arg thr tyr arg val gly
  184 TAC CAC ACC TGC GGA GGG TCC CTG ATA GCC AAC AGC TGG GTC CTG ACG GCT GCC CAC TGC ATC AGC TCC TCC AGG ACC TAC CGC GTG GGG

          70                                  80                                  90
    leu gly arg his asn leu tyr val ala glu ser gly ser leu ala val ser val ser lys ile val val his lys asp trp asn ser asn
  274 CTG GGC CGG CAC AAC CTC TAC GTT GCG GAG TCC GGC TCG CTG GCA GTC AGT GTC TCT AAG ATT GTG GTG CAC AAG GAC TGG AAC TCC AAC

          100                 110                 120
    gln ile ser lys gly asn ASP ile ala leu leu lys leu ala asn pro val ser leu thr asp lys ile gln leu ala cys leu pro pro
  364 CAA ATC TCC AAA GGG AAC GAC ATT GCC CTG CTC AAA CTG GCT AAC CCC GTC TCC CTC ACC GAC AAG ATC CAG CTG GCC TGC CTC CCT CCT

          130                 140                 150
    ala gly thr ile leu pro asn asn tyr pro cys tyr val thr gly trp gly arg leu gln thr asn gly ala val pro asp val leu gln
  454 GCC GGC ACC ATT CTA CCC AAC AAC TAC CCC TGC TAC GTC ACG GGC TGG GGA AGG CTG CAG ACC AAC GGG GCT GTT CCT GAT GTC CTG CAG

          160                 170                 180
    gln gly arg leu leu val val asp tyr ala thr cys ser ser ser ala trp trp gly ser ser val lys thr ser met ile cys ala gly
  544 CAG GGC CGG TTG CTG GTT GTG GAC TAT GCC ACC TGC TCC AGC TCT GCC TGG TGG GGC AGC AGC GTG AAA ACC AGT ATG ATC TGT GCT GGG

          190                 200                 210
    gly asp gly val ile ser SER cys asn gly asp ser gly gly pro leu asn cys gln ala ser asp gly arg trp gln val his gly ile
  634 GGT GAT GGC GTG ATC TCC AGC TGC AAC GGA GAC TCT GGC GGG CCA CTG AAC TGT CAG GCG TCT GAC GGC CGG TGG CAG GTG CAC GGC ATC

          220                 230                 240
    val ser phe gly ser arg leu gly cys asn tyr tyr his lys pro ser val phe thr arg val ser asn tyr ile asp trp ile asn ser
  724 GTC AGC TTC GGG TCT CGC CTC GGC TGC AAC TAC TAC CAC AAG CCC TCC GTC TTC ACG CGG GTC TCC AAT TAC ATC GAC TGG ATC AAT TCG

          250   253
    val ile ala asn asn OC
  814 GTG ATT GCA AAT AAC TAA CCAAAAGAAGTCCCTGGGACTGTTTCAGACTTGGAAAGGTCACAGAAGGAAAATAATATAATAAAGTGACAACTC
```

FIGURE 2: Nucleotide sequence of human pancreatic elastase 2 cDNA and the deduced amino acid sequence. The amino acid sequence is numbered sequentially from the first amino acid of the zymogen. The activation peptide comprises amino acids 1–12, and the proposed signal peptide is represented by amino acids −16 to −1. The underlined amino acids were shown to be identical with those determined for the activation peptide for human elastase 2 (Largman et al., 1980) and for residues 1–20 of active elastase 2 as described under Methods. The amino acids that comprise the charge relay system (His-57, Asp-102, and Ser-195 by the chymotrypsin numbering system; Hartley, 1970) are indicated by capital letters and superscript asterisks (residues 57, 105, and 200, respectively). The amino acids thought to limit the size of the S1 binding pocket (Gly-216 and Ser-226 in the chymotrypsin numbering system) are also indicated by asterisks (Gly-222 and Ser-234, respectively). The proposed polyadenylation site is underlined.

subcloned into M13 and sequenced. Although the entire elastase 2 cDNA sequence is derived from two separate clones, there was complete identity between the overlapping region.

Nucleotide and Amino Acid Sequence. Figure 2 shows the sequence for the two clones which represent the human pancreatic elastase 2 message and the deduced amino acid sequence. The 906 bp elastase 2 cDNA contains an open reading frame that encodes a protein containing 269 amino acids. Amino acids 1–12 are identical with those previously determined for the activation peptide of human elastase 2 (Largman et al., 1980), while residues 13–32 are identical with the N-terminal 20 amino acids determined for human pancreatic elastase 2 (Figure 2).

The 5′ portion of the cDNA encodes a Met residue at −16 immediately preceded by a five-base sequence (ACACC) highly homologous to the consensus sequence for eukaryotic translation initiation sites (Kozak, 1984), and followed by a hydrophobic amino acid region which conforms to the general pattern previously observed for leader sequences (Perlman & Halvorson, 1983). The active protein consists of amino acids 13–253 and contains all of the characteristic amino acids of the serine protease active site, which are denoted in the chymotrypsin numbering system as follows, with the amino acid residue of elastase 2 in parentheses: His-57, Asp-102 (Asp-105), Ser-195 (Ser-200), and Val-16 (Val-13), as well as a Ser-189 (Ser-195). The protein sequence also contains all of the cysteines previously shown to form the five disulfide bridges characteristic of chymotrypsins (Hartley, 1970), including a bridge between residue 1 of the activation peptide and residue 125 of the active enzyme.

The 3′ region consists of a TAA stop codon at position 829, followed by 76 nucleotides of untranslated message containing a polyadenylation recognition site (AATAAA). The nucleotide sequence which encodes the active protein (106–828) is 61% and 82% homologous with those of rat pancreatic elastases 1 and 2, respectively (MacDonald et al., 1982), and 77% identical with that of mouse elastase 2 cDNA (Stevenson et al., 1986).

Comparison of Elastase 2 Protein Sequences. Figure 3 shows the alignment of the amino acid sequence of human elastase 2 with those derived from the cDNAs for rat and mouse elastases 2, as well as protein sequences of bovine α-chymotrypsin A and porcine elastases 1 and 2. The degree of identity over the active protein region between human elastase 2 and these proteins is 84%, 73%, 39%, 59%, and 55%, respectively. The maximum homology is maintained in the immediate N-terminus and in the C-terminal half of the molecule, which contains the P1 specificity pocket as well as most catalytic residues. Almost all of the variation between these proteins is observed between residues 48 and 180. It should be noted that human elastase 2 has the same amino acids as the other elastases 2 at Gly-216 and Ser-226 (chymotrypsin numbering system).

mRNA Size Determination. As shown in Figure 4A, when a sample of human pancreatic messenger RNA was electrophoresed on an agarose gel, transferred to nitrocellulose, and probed with ^{32}P-labeled rat elastase 2 cDNA, a single band was observed with a mobility consistent with a molecular size of 1100 nucleotides, indicating a single size of elastase 2 message is synthesized in the pancreas.

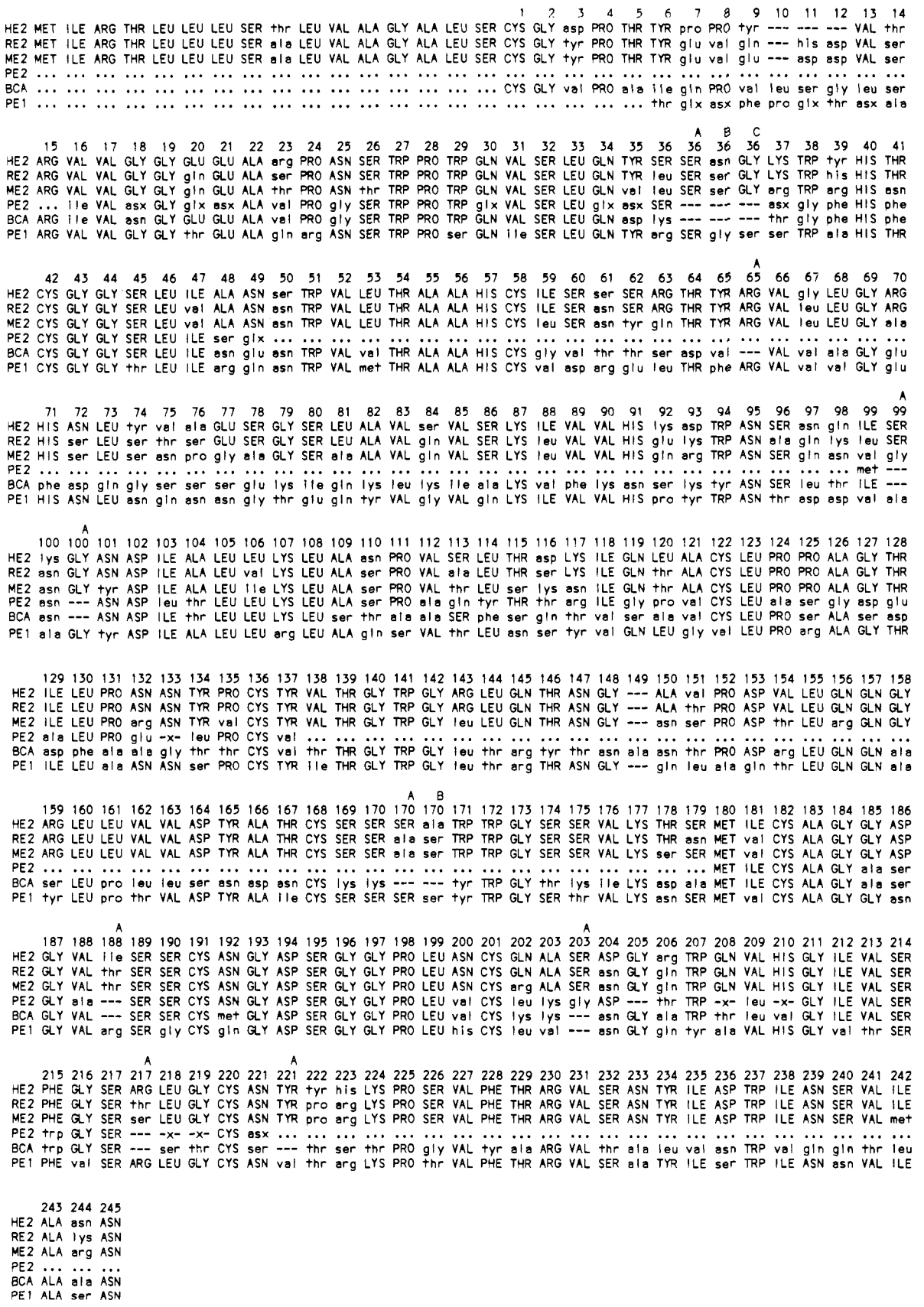

FIGURE 3: Amino acid sequence comparison of elastases and chymotrypsin. Amino acid sequence data are shown for human pancreatic elastase 2 (HE2), rat pancreatic elastase 2 (RE2; MacDonald et al., 1982), mouse elastase 2 (ME2; Stevenson et al., 1986), porcine elastase 2 (PE2; Vered et al., 1986), bovine chymotrypsin A (BCA; Hartley, 1970), and porcine elastase 1 (PE1; Hartley, 1970; Lamy et al., 1977). The chymotrypsin numbering system (Hartley, 1970) was used. For each protein sequence, known activation and signal peptides have been included. Amino acid residues which are shared between human pancreatic elastase 2 and at least one other enzyme are capitalized.

Southern Blot Analysis. As shown in Figure 4B, Southern blot analysis reveals that for each of the restriction enzymes used, human genomic DNA yields a limited number of bands when probed with an 800 bp elastase 2 cDNA fragment. Summation of the molecular weights of the fragments generated by each of the enzymes indicates a total of less than

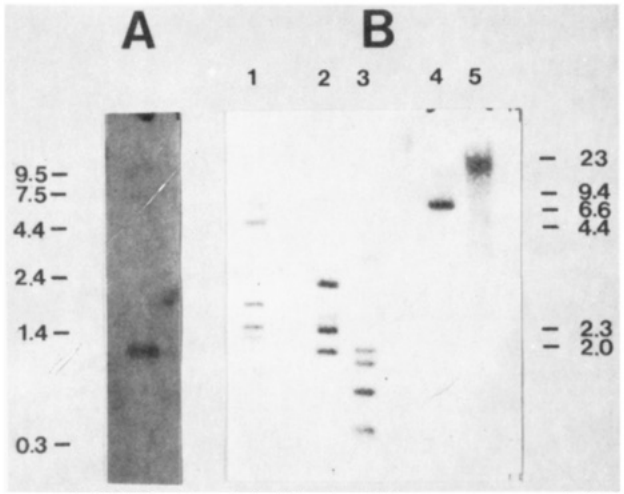

FIGURE 4: (A) Northern hybridization. Human pancreas poly(A+) mRNA was electrophoresed in a formaldehyde gel, transferred to nitrocellulose, and probed with a labeled human pancreatic elastase 2 probe as described under Methods. The size of the mRNA was estimated by using RNA standards. (B) Genomic Southern hybridization. Placental genomic DNA was digested with the restriction enzymes indicated below, electrophoresed, and transferred to nitrocellulose as described under Methods. The size of the fragments was estimated by using *Hin*dIII-cut λ DNA as standards. Lane 1, *Sst*I; lane 2, *Pvu*II; lane 3, *Pst*I; lane 4, *Bgl*II; lane 5, *Bal*I. The probe (hpe2-λ4) used corresponds to the portion of the cDNA encoding the active enzyme and 3′ untranslated region.

10–12 kb, respectively. Since previous studies have shown that both the rat (Swift et al., 1984) and mouse (Stevenson et al., 1986) elastase 2 genes are single-copy genes of approximately 12 kb, our data suggest that the human elastase 2 gene is also present as a single- or low-copy number gene.

Discussion

Determination of the amino acid sequence of human pancreatic elastase 2 permits correlation of substrate specificity with structure for this class of elastases. There appear to be two separate aspects of elastase 2 activity for which the amino acid sequence should be interpreted: (a) What structural features result in the broad substrate specificity and relatively inefficient catalytic activity observed (Del Mar et al., 1980), and (b) what particular structural characteristics confer the unusual property of elastolysis on these enzymes?

The classic X-ray crystallographic studies of the serine proteases demonstrated a highly conserved backbone structure with small changes in specific amino acid residues thought to result in peptide bond specificity for trypsin, chymotrypsin, and elastase 1 (Stroud et al., 1971). In particular, X-ray studies showed that Val-216 and Thr-226 of elastase 1 partially occlude the large S1 binding pocket observed for chymotrypsin, which is bounded by Gly-216 and Gly-226 (Shotton & Watson, 1970). The sequences of human, rat, and mouse elastases 2 all contain a Gly-216/Ser-226 pattern, suggesting that these enzymes would possess an S1 binding pocket of intermediate size between elastases 1 and chymotrypsins. Carlson et al. (1986) have modeled the structure of rat elastase 2 and have predicted that the S1 binding pocket would be of intermediate dimensions. Since essentially all of the residues identified in this model as interacting with substrates are conserved between rat and human elastases 2, the proposed enlarged substrate binding pocket would appear to explain the observed specificity of human elastase 2 for substrates containing either extended aliphatic or aromatic side chains in the P1 amino acid, while small side chains such as alanine would not bind tightly, in agreement with the observed reduced activity toward alanine substrates (Del Mar et al., 1980). The proposed model of rat elastase 2 has also been interpreted to explain why human elastase 2 will not rapidly cleave substrates with a branched carbon in the P1 position (Carlson et al., 1986). These authors suggest that there are substantial steric contacts between the binding pocket and the branched amino acid side chain. It is less clear why the elastases 2 appear to possess lower inherent catalytic efficiency toward peptide substrates. Studies of human elastase 2 reveal that the binding constant, K_m, of the enzyme toward its best substrate is elevated approximately 10-fold compared to the best substrates for either bovine α-chymotrypsin or porcine elastase 1 (Del Mar et al., 1980). In addition, the catalytic rate constant (k_{cat}) is decreased approximately 10–30-fold in elastase 2 compared with the other two enzymes, resulting in an overall reduction in catalytic efficiency of approximately 100-fold compared to the other classes of serine proteases possessing extremely related active sites. Carlson et al. (1986) have suggested that reduced hydrolysis rates may be due to the fact that Gln-192, which is thought to interact with the substrate P1 side chain (Shotton et al., 1971), is Asn in the rat elastase 2 cDNA. Since residue 192 is also an asparagine in human elastase 2, it is possible that this difference is important for proper substrate binding or orientation for efficient hydrolysis, resulting in the lower efficiency of human elastase 2. Since the rat and mouse elastases 2 have not been isolated, the catalytic efficiency of these proteases is unknown. However, porcine elastase 2, which also contains Asn-192, exhibits similar catalytic rate constants for peptide substrates as those reported for chymotrypsin (Gertler et al., 1977). It should be noted that chymotrypsin contains Met-192, making interpretation of the contribution of residue 192 to efficiency difficult.

Atlas (1975) attempted to define the amino acids which contribute to the extended binding region of elastase 1 by X-ray crystallographic analysis of an elastase 1–substrate complex. Although human elastase 2 exhibits a modest increase in substrate binding for longer substrates (Del Mar et al., 1980), many of the residues denoted by Atlas (1975) are changed in elastase 2, including Val-99/Ile-99, Gln-150/Ala-150, Leu-151/Val-151, and the change of Gln-192/Asn-192 discussed above. The importance of each of these residues for increased catalytic efficiency remains, unclear but can now be addressed by using site-directed mutagenesis (Craik et al., 1985).

In contrast to the numerous studies of elastase and chymotrypsin specificity toward peptide substrates, relatively few studies have attempted to identify the amino acid residues important for elastolysis. Gertler (1971a) provided evidence for nonspecific electrostatic adsorption of basic proteins to the acidic elastin molecule and demonstrated that maleylation of the three lysines of porcine elastase 1 prevented both adsorption to elastin and elastolysis but did not affect hydrolysis of peptide or general protein substrates (Gertler, 1971b). Specific modification of two arginine residues in porcine elastase 1 reduced both elastolytic and esterolytic activity by 85% (Davril et al., 1984). These authors proposed that Arg-65A and Arg-217A were the modified residues. However, these experiments did not differentiate between effects due to specific modification of lysine or arginine residues, which might participate in some type of binding to elastin, and general lowering of the isoelectric point of the enzyme. Comparison of the amino acid sequences of porcine elastase 1 and rat and human elastases 2 with bovine chymotrypsin A (Figure 3) shows that although all of these proteins are basic, the three elastases are substantially more cationic with net charges of

8+, 8+, and 11+, respectively, compared to a net charge of 3+ for chymotrypsin (assigning His a value of $^1/_2$+ at neutral pH).

Examination of the model of the porcine elastase 1 crystal structure reveals that only a few of the basic residues are close to the substrate binding cleft, in a position which might facilitate binding to the elastin molecule, while being outside the previously described peptide binding sites. In particular, Arg-217A, Arg-223/Lys-224, and Arg-65A appear to be candidates for anchoring the active-site catalytic cleft of porcine elastase 1 to elastin. Comparison of the sequences shown in Figure 3 as well as that of chymotrypsinogen B (Hartley, 1970; data not shown) reveals that all elastases contain Arg-65A and a basic dipeptide structure at positions 223/224, while chymotrypsins do not have a 65A residue and have neutral amino acids at 223/224. It is of interest that human protease E, an alanine-specific, anionic protease which does not rapidly hydrolyze elastin, has a Gln-65A but possesses a basic tripeptide at residues 222–224 (Shen et al., 1987). Human elastase 2 contains an Arg at position 217A, while rat and mouse elastases 2 do not have a basic amino acid at this position. It should be noted that neither of the latter two proteins has been purified, despite attempts to isolate elastolytic enzymes from rat pancreatic tissue extracts (Largman, 1983). It is possible that the rat enzyme has reduced capacity for elastolysis due to the Arg-217A to Thr-217A change. These differences suggest that Arg-65A, Arg-217A, and perhaps basic amino acids at 223/224 may be important for binding elastases to elastin and suggest these residues as candidates for modification by site-directed mutagenesis.

Added in Proof

A recent publication (Kawashima et al., 1987) reports that two elastase 2 mRNAs are expressed in the pancreas. One of their sequences is identical with the one reported here (elastase IIA) and codes for the protein identified as elastase 2. No protein has been identified as corresponding to the second cDNA they report (elastase IIB).

Acknowledgments

We thank Al Smith of the UC Davis Protein Structure Laboratory for determination of the N-terminal amino acid sequence for elastase 2. We also thank Ray MacDonald for the rat elastase 2 cDNA probe and Bob Weiss for the human pancreatic cDNA library.

Registry No. DNA (human clone hpe2-λ10/hpe2-λ4 elastase 2 messenger RNA complementary), 110743-50-5; preproelastase 2 (human clone hpe2-λ10/hpe2-λ4 reduced), 107528-53-0; proelastase 2 (human clone hpe2-λ10/2-λ4 reduced), 107528-55-2; elastase 2 (human clone hpe2-λ10/hpe2-λ4 reduced), 110071-54-0; elastase 2, 75603-19-9; proelastase 2, 74505-35-4; preproelastase 2, 80966-01-4; elastase 1, 9004-06-2; chymotrypsin, 9004-07-3; pancreatic proteinase E, 68073-27-8.

References

Atlas, D. (1975) *J. Mol. Biol. 93*, 39–53.

Bankier, A. T., & Barrell, B. G. (1983) *Tech. Life Sci.: Biochem. B5 B508*, 1–34.

Carlson, G. M., MacDonald, R. J., & Meyer, E. F. (1986) *J. Theor. Biol. 119*, 107–124.

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

Craik, C., Largman, C., Fletcher, T., Roczniak, S., Barr, P. J., Fletterick, R., & Rutter, W. J. (1985) *Science (Washington, D.C.) 228*, 291–297.

Davril, M., Jung, M.-L., Duportail, G., Lohez, M., Han, K. K., & Bieth, J. (1984) *J. Biol. Chem. 259*, 3851–3857.

Dayringer, H. E., Tramontano, A., Sprang, S. R., & Fletterick, R. J. (1986) *J. Mol. Graphics 4*, 82–87.

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

Feinberg, A., & Vogelstein, B. (1983) *Anal. Biochem. 132*, 6–13.

Gertler, A. (1971a) *Eur. J. Biochem. 23*, 36–40.

Gertler, A. (1971b) *Eur. J. Biochem. 20*, 541–546.

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

Hartley, B. S. (1970) *Philos. Trans. R. Soc. London, B257*, 77–87.

Hartley, B. S., & Shotton, D. M. (1971) *Enzymes (3rd Ed.) 3*, 323–373.

Janoff, A. (1985) *Am. Rev. Respir. Dis. 132*, 417–433.

Kasafirek, E., Fric, P., Slaby, J., & Malis, F. (1976) *Eur. J. Biochem. 69*, 1–13.

Kawashima et al. (1987) *DNA 6*, 163–172.

Kozak, M. (1984) *Nucleic Acids Res. 12*, 857–873.

Lamy, F., Gibson, D., Ledoux, M., & Moreux, J. C. (1977) in *Elastin and Elastic Tissue* (Sanburg, L., Gray, W. R., & Franzblau, C., Eds.) pp 165–175, Plenum Press, New York.

Largman, C. (1983) *Biochemistry 22*, 3763–3770.

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

Largman, C., Brodrick, J. W., & Geokas, M. C. (1980) *Biochim. Biophys. Acta 623*, 208–212.

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

Mallory, P. A., & Travis, J. (1975) *Biochemistry 14*, 722–729.

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

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

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

Shen, W.-F., Fletcher, T. S., & Largman, C. (1987) *Biochemistry 26*, 3447–3452.

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.

Stevenson, B. J., Hagenbuchle, O., & Wellauer, P. K. (1986) *Nucleic Acids Res. 14*, 8307–8330.

Stroud, R. M., Kay, L. M., & Dickerson, R. E. (1971) *Cold Spring Harbor Symp. Mol. Biol. 36*, 125–140.

Swift, G. H., Craik, C. S., Stary, S. J., Quinto, C., Lahaie, R. G., Rutter, W. J., & MacDonald, R. J. (1984) *J. Biol. Chem. 259*, 14271–14278.

Vered, M., Gertler, A., & Bernstein, Y. (1986) *Int. J. Pept. Protein Res. 27*, 183–190.

Yamada, E., Hazama, F., Kataoka, H., Amano, S., Sasahara, M., Kayembe, K., & Katayama, K. (1983) *Virchows Arch., B 44*, 241–245.