Primary Structure of Human Pancreatic Protease E Determined by Sequence Analysis of the Cloned mRNA†

Wei-fang Shen, Thomas S. Fletcher, 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 November 19, 1986; Revised Manuscript Received February 6, 1987

ABSTRACT: Although protease E was isolated from human pancreas over 10 years ago [Mallory, P. A., & Travis, J. (1975) *Biochemistry 14*, 722–729], its amino acid sequence and relationship to the elastases have not been established. We report the isolation of a cDNA clone for human pancreatic protease E and determination of the nucleic acid sequence coding for the protein. The deduced amino acid sequence contains all of the features common to serine proteases. The substrate binding region is highly homologous to those of porcine and rat elastases 1, explaining the similar specificity for alanine reported for protease E and these elastases. However, the amino acid sequence outside the substrate binding region is less than 50% conserved, and there is a striking difference in the overall net charge for protease E (6−) and elastases 1 (8+). These findings confirm that protease E is a new member of the serine protease family. We have attempted to identify amino acid residues important for the interaction between elastases and elastin by examining the amino acid sequence differences between elastases and protease E. In addition to the large number of surface charge changes which are outside the substrate binding region, there are several changes which might be crucial for elastolysis: Leu-73/Arg-73; Arg-217A/Ala-217A; Arg-65A/Gln-65A; and the presence of two new cysteine residues (Cys-98 and Cys-99B) which computer modeling studies predict could form a new disulfide bond, not previously observed for serine proteases. We also present evidence which suggests that human pancreas does not synthesize a basic, alanine-specific elastase similar to porcine elastase 1.

Despite substantial evidence linking elastolytic enzymes with emphysema and atherosclerosis (Rosenbloom, 1984), there is no clear understanding of the specific structural features which enable elastases to degrade elastin. Over the past 25 years, extensive research on porcine pancreatic elastase has resulted in a widely held view that this serine protease is representative of elastases. Thus, the characteristic properties of this enzyme, including a restricted binding pocket (Shotten & Watson, 1970) which results in alanine specificity (Kasafirek et al., 1976), as well as a highly basic charge (Hartley & Shotton, 1971), are taken to be defining features of elastases in a manner similar to those which define the specificity of trypsin (Stroud et al., 1974) and chymotrypsin (Mathews et al., 1967).

Although there are pancreatic elastases from rat (Largman, 1983) and dog (Geokas et al., 1980) which conform to this pattern, neither of the human pancreatic elastase-like enzymes shares these properties. Although human protease E (Mallory & Travis, 1975) has a specificity for alanine and is an efficient protease, this acidic enzyme shows little or no elastolytic activity. Furthermore, human pancreatic elastase 2 (Largman et al., 1976) is an efficient elastase but possesses a specificity for leucine and phenylalanine (Del Mar et al., 1980).

In order to elucidate 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. Here, we report the isolation of a full-length cDNA clone for human pancreatic protease E and its deduced amino acid sequence. Comparison of protease E, elastases, and chymotrypsins in terms of amino acid homology and proposed three-dimensional structure suggests that protease E is a separate member 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 protease E and elastases.

Experimental Procedures

Materials

A λgt11 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 λgt11 using *Eco*RI tails. The library contained a total of 10^5 independent clones containing inserts. A full-length cDNA clone for rat pancreatic elastase 1 was kindly provided by Dr. Ray MacDonald. All other reagents and enzymes were obtained from commercial sources.

Methods

Library Screening. An initial screening of 1.5×10^4 insert-containing plaques was performed with a 900 base pair (bp)[1] rat elastase I cDNA probe (MacDonald et al., 1982a) isolated by gel electrophoresis and labeled with ^{32}P by the method of Feinberg and Vogelstein (1983). Low stringency hybridization conditions were initially used to screen the library [42 °C in 5% formamide, 5× SSC, 5× Denhardt's solution,

† 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).

* Address correspondence to this author at the Biochemistry Research Laboratory, Veterans Administration Medical Center.

[1] Abbreviations: 20× SSC, 3.0 M sodium chloride/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); SDS, sodium dodecyl sulfate; EDTA, ethylenediaminetetraacetic acid; PTH, phenylthiohydantoin; kb, kilobase(s).

0006-2960/87/0426-3447$01.50/0

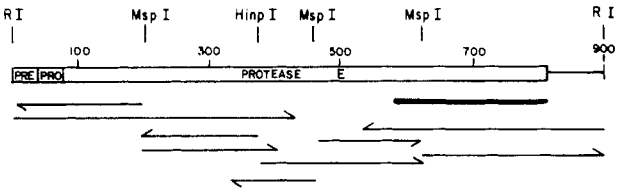

FIGURE 1: Sequencing strategy for human pancreatic protease E cDNA. The horizontal rectangle represents the amino acid coding region of protease E; the regions coding the signal peptide (PRE) and activation peptide (PRO) are shown. The horizontal line extending from the rectangle indicates 3′ untranslated sequence, which terminates in a poly(A) tail. The thick line represents the partial clone (6-3) which was initially isolated from the cDNA library by using the rat elastase 1 probe. The direction and length of each sequencing run are shown by the horizontal arrows, each beginning at the restriction site shown. The *Eco*RI sites were created during the library construction.

0.1 mg/mL sonicated salmon sperm DNA, 20 mM sodium phosphate (pH 6.5), 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. Following two rounds of plate purification, positive clones were isolated, cut with *Eco*RI, and subcloned into M13 for sequence determination. Following identification of a clone which appeared to code for an elastase-like protease (see below), the library was rescreened with this partial cDNA as a probe under stringent conditions (50% 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.

Protease E Sequence Determination. The scheme for determining the full-length sequence of protease E is summarized in Figure 1. Initial screening of the library with the rat elastase 1 probe yielded clone 6–3. This clone was then used to isolate clone HPE2 which was sequenced in both directions by subcloning in M13. Positive clones were cut with the appropriate restriction enzymes, and the resulting DNA fragments were isolated by agarose gel electrophoresis, electroelution, and purification on an Elutip-d column (Schleicher & Schuell, Keene, NH). DNA fragments were ligated into M13 mp18 or mp19 for sequence analysis using the deoxyadenosine 5′-[α-^{35}S]thiophosphate method described by Biggin et al. (1983). Following electrophoresis of sequencing reactions on 5% or 8% polyacrylamide/urea gels, the gels were dried and autoradiographed for 12–48 h at room temperature.

Preparation of Pancreas mRNA and Northern Blotting. Total pancreas RNA was prepared by using the guanidinium thiocyanate procedure described by Chrigwin 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. RNA was transferred to nitrocellulose by capillary action in 20x SSC, and the resulting filter was baked for 2 h in a vacuum oven, prehybridized and hybridized under stringent conditions as described above.

Protein Sequencing. A sample (5 nmol) of human pancreatic protease E (Largman et al., 1976) was subjected to 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 human protease E was predicted by using the coordinates of porcine elastase 1 (Brookhaven Protein Data Bank). The amino acid differences between the two sequences were substituted into the model of porcine elastase 1, and the structure was refined by 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 for 2 h with the appropriate restriction enzyme and subjected to electrophoresis in a 1% agarose gel. Genomic fragments containing protease E sequences were visualized by transfer to nitrocellulose (Maniatis et al., 1982) followed by probing under stringent conditions with ^{32}P-labeled protease E cDNA (Feinberg & Vogelstein, 1983) and autoradiography for 16 h without an enhancing screen at −70 °C. *Hin*dIII-cut λ molecular weight markers (New England Biolabs) were end labeled with [α-^{32}P]dCTP using Klenow fragment according to Maniatis et al. (1982).

RESULTS

Library Screening. Initial low-stringency screening of the human pancreas cDNA library with rat elastase 1 cDNA probe yielded two positive clones: an elastase-like clone and a partial cDNA with a nucleotide sequence identical with that recently reported for human trypsin (Emi et al., 1986). The elastase-like cDNA clone (6-3 in Figure 1) coded for the active-site Ser-195, a Ser-189, and two stretches of amino acid sequence (residues 209–217 and 223–233) which comprise the characteristic substrate binding pocket of porcine pancreatic elastase 1 (see Figure 4, below). This 243 bp clone was 63% homologous to rat elastase 1 and contained three stretches of high identity (22 of 23, 22 of 25, and 21 of 23 nucleotides, respectively). The 266 bp trypsin clone was 49% homologous with rat elastase 1 cDNA and included a sequence for the active-site Ser-195 region which was identical in 18 of 20 bases.

Clone 6-3 was used to rescreen the cDNA library at high stringency, resulting in identification of 46 positive clones. Clone HPE2, which was 900 bp long, was selected for sequence analysis. As shown in Figure 1, a series of restriction fragments covering the entire cDNA were subcloned into M13 and sequenced.

Nucleotide and Amino Acid Sequence. Figure 2 shows the sequence for clone HPE2, which we will refer to as protease E, and the deduced amino acid sequence. HPE2 contains an open reading frame that codes for a protein containing 267 amino acids. Amino acids 1–20 encoded by HPE2 are identical with the N-terminal 20 amino acid residues found for human protease E, and also the sequence reported for porcine protease E (Kobayashi et al., 1978) (Figures 2 and 3).

The 5′ portion of the cDNA codes for a hydrophobic region from amino acid −25 to −12, followed by amino acids −11 to −1 which are homologous to the porcine protease E activation peptide, as well as to that of lungfish elastase A (Figure 3). Although the assignment of the first residue of the activation peptide for human protease E is tentative, based on the predicted structure of the cleavage position of signal peptidase (Perlman & Halvorson, 1983), the cDNA open reading frame for protease E does not code for a cysteine in this region, suggesting that the activation peptide is not bound to the active enzyme following tryptic activation.

The 3′ region consists of a TAG stop codon at position 801, followed by 50 nucleotides of untranslated message, a repeated polyadenylation recognition site (AATAAA), and a poly(A) tail. The nucleotide sequence which encodes the active protein (43–801) is 55% and 58% homologous with that of rat pan-

```
            -20                                -10
GlyLeuLeuSerSerLeuLeuLeuValAlaValAlaSerGlyTyrGlyProProSerSer
GGGCTGCTCAGTTCCCTCCTCCTTGTGGCCGTTGCCTCAGGCTATGGCCCACCTTCCTCT

          -1 +1                                10
ArgProSerSerArgValValAsnGlyGluAspAlaValProTyrSerTrpProTrpGln
CGCCCTTCCAGCCGCGTTGTCAATGGTGAGGATGCGGTCCCCTACAGCTGGCCCTGGCAG

           20                                 30
ValSerLeuGlnTyrGluLysSerGlySerPheTyrHisThrCysGlyGlySerLeuIle
GTTTCCCTGCAGTATGAGAAAAGTGGAAGCTTCTACCACACGTGTGGCGGTAGCCTCATC

           40            ***                  50
GlyProAspTrpValValThrAlaGlyHISCysIleSerSerSerTrpThrTyrGlnVal
GGCCCCGACTGGGTTGTGACTGCCGGCCACTGCATCTCGAGCTCCTGGACCTACCAGGTG

           60                                 70
ValLeuGlyGluTyrAspArgAlaValLysGluGlyProGluGlnValIleProIleAsn
GTGTTGGGCGAGTACGACCGTGCTGTGAAGGAGGGCCCCGAGCAGGTGATCCCCATCAAC

           80                                 90          ***
SerGlyAspLeuPheValHisProLeuTrpAsnArgSerCysValGlyCysGlyAsnASP
TCTGGGGACCTCTTTGTGCATCCACTCTGGAACCGCTCGTGTGTCGGCTGTGGCAATGAC

          100                                110
IleAlaLeuIleLysLeuSerArgSerAlaGlnLeuGlyAspAlaValGlnLeuAlaSer
ATCGCCCTCATCAAGCTCTCACGCAGCGCCCAGCTGGGAGACGCCGTCCAGCTCGCCTCA

          120                                130
LeuProProAlaGlyAspIleLeuProAsnGluThrProCysTyrIleThrGlyTrpGly
CTCCCTCCCGCTGGTGACATCCTTCCCAACGAGACACCCTGCTACATCACCGGCTGGGGG

          140                                150
ProLeuTyrThrAsnGlyProLeuProAspLysLeuGlnGluAlaLeuLeuProValVal
CCTCTCTATACCAACGGGCCACTCCCAGACAAGCTGCAGGAGGCCCTGCTGCCCGTGGTG

          160                                170
AspTyrGluHisCysSerArgTrpAsnTrpTrpGlySerSerValLysLysThrMetVal
GACTATGAACACTGCTCCAGGTGGAACTGGTGGGGTTCCTCCGTGAAGAAGACCATGGTG

          180                          ***   190
CysAlaGlyGlyAspIleArgSerGlyCysAsnGlyAspSERGlyGlyProLeuAsnCys
TGTGCTGGAGGGGACATCCGCTCCGGCTGCAACGGTGACTCTGGAGGACCCCTCAACTGC

          200          *  *  *  *  *  *  * 210 *
ProThrGluAspGlyGlyTrpGlnValHisGlyValThrSerPheValSerAlaPheGly
CCCACAGAGGATGGTGGCTGGCAGGTCCATGGCGTGACCAGCTTTGTTTCTGCCTTTGGC

          220 *  *  *  *  *  *  *  *  *  * 230
CysAsnThrArgArgLysProThrValPheThrArgValSerAlapheIleAspTrpIle
TGCAACACCCGCAGGAAGCCCACGGTGTTCACTCGAGTCTCCGCCTTCATCGACTGGATT

          240      STOP
GluGluThrIleAlaSerHis
GAGGAGACCATAGCAAGCCACTAGAACCAAGGCCCAGCTGGCAGTTGCTTGATCCGATCC

CACATCCTTGAATAAAGAATAAAGATCTCTCAGAAAATTCCAAAAAAAAAAAAAAA
```

FIGURE 2: Nucleotide sequence of human pancreatic protease E cDNA and amino acid sequence of the encoded protein. The deduced amino acid sequence is numbered sequentially from the amino acid terminus of the predicted active enzyme. The predicted activation peptide comprises amino acids −11 to −1, and the partial sequence of a proposed signal peptide is represented by amino acids −25 to −12. The amino acid residues which 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 (His-45, Asp-94, and Ser-188, respectively). The amino acids which comprise the substrate binding pocket (residues 209–217 and 223–233 by the chymotrypsin numbering system) are denoted by superscript asterisks (residues 204–211 and 220–229). The nucleic acid sequence for the partial cDNA clone (6-3) is shown as underlined.

creatic elastases 1 and 2, respectively (MacDonald et al., 1982a).

As shown in Figure 4, the deduced amino acid sequence of protease E contains all of the residues which contribute to catalysis by the serine proteases: Val-16, His-57, Asp-102, Gly-193, Asp-194, and Ser-195 as well as all of the disulfide bonds present in porcine and rat pancreatic elastases 1 (Mathews et al., 1967; Hartley, 1970). Human protease E contains two additional cysteine residues at positions 98 and 99B, which are not present in other serine proteases. Computer modeling experiments indicate that these two cysteines are capable of forming a disulfide bond which would lock the short loop from residues 97 to 100 into a flat, rigid conformation, but with the same basic three-dimensional structure as porcine pancreatic elastase. The S1[2] specificity pocket of protease E contains a valine at position 216 and a threonine at position 226, suggesting that this pocket is sterically obstructed as observed for porcine pancreatic elastase 1 (Shotton & Watson, 1970).

mRNA Analysis. Electrophoresis of human poly(A+) RNA under denaturing conditions followed by hybridization with the full-length protease E cDNA probe yielded a single band with a size of approximately 1 kb (results not shown).

Southern Blot Hybridization. As shown in Figure 5, Southern blot analysis of human genomic DNA reveals a small number of bands for each restriction enzyme used, when probed with a full-length cDNA for protease E. When a partial cDNA probe is used (Figure 5), selected bands disappear from each restriction pattern, suggesting that there are probably one or a limited number of protease E genes.

Attempts To Identify cDNA Related to Elastase 1. Since the partial protease E cDNA clone isolated with rat elastase 1 cDNA contained the alanine specificity binding region, but was only 63% identical at the DNA level, we hypothesized that if a cDNA representing an alanine-specific enzyme with basic charge (similar to elastases 1) were present in the library, it should hybridize to rat elastase 1 as well as, or better than, protease E. The pancreas cDNA library was rescreened with the rat elastase 1 probe in an attempt to isolate such a clone. Nine positive clones were isolated, and duplicate filters were hybridized under stringent conditions with both the partial protease E cDNA clone (6-3) and the 226 bp trypsin cDNA clone described above. Six of these clones hybridized strongly to protease E cDNA and not to trypsin, while the other three clones hybridized strongly to the trypsin cDNA and not to protease E. There were no clones which were not highly homologous to either protease E or trypsin, under the stringent conditions used, indicating that rat elastase 1 cDNA detected only these two types of cDNA in the human pancreas library. Since rat elastase 1 is only 49% homologous with the 266 bp trypsin partial cDNA, there does not appear to be any cDNA in the library which is highly related to the elastase 1 family.

DISCUSSION

We have employed a cDNA encoding rat pancreatic elastase 1 (MacDonald et al., 1982a) to screen a human pancreas cDNA library in order to clone human elastase-like enzymes. Under low-stringency conditions, protease E was the major cDNA isolated. Several clones for trypsin were also obtained, but no cDNA encoding a human homologue of porcine pancreatic elastase 1 was detected. The partial cDNA initially isolated by this method encoded a protein with a very highly conserved active-site serine region (residues 189–201), as well as two stretches of amino acids (residues 209–217 and residues 223–233) thought to form the S1 specificity pocket which confers P1 alanine specificity, and portions of the extended substrate binding region for elastases (Figure 4). However, comparison of the amino acid sequences of protease E and porcine elastase 1, rat elastase 1, porcine elastase 2, and bovine α-chymotrypsin reveals that protease E is significantly different from these other enzymes. The most obvious difference is that protease E is highly acidic, with a net charge of 6− (assigning histidine as $^1/_2$+ at neutral pH), while porcine and rat elastases

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

```
                               1   2   3   4   5   6   7   8   9  10  11  12 13  14  15   16  17  18  19  20  21  22  23  24  25  26  27  28  29  30  31  32  33 34  35
      Human Protease E                         tyr-GLY-PRO-PRO-SER-ser-arg-PRO-ser-SER-ARG VAL-VAL-ASN-GLY-GLU-ASP-ALA-VAL-PRO-TYR-SER-TRP-PRO-TRP-GLN-VAL-ser-LEU-GLN-TYR
    Porcine Protease E                          x -GLY-arg-PRO-SER-tyr-asn-PRO-ala-ala-ARG VAL-VAL-ASN-GLY-GLU-ASP-ALA-VAL-PRO-TYR-SER-TRP-PRO-TRP-GLX-VAL- x -LEU-GLX-TYR
     Bovine Protease E                                                                     VAL-VAL-ASX-GLY-GLX-ASX-ALA-VAL-PRO-TYR
   Lungfish Elastase A                         cys-GLY-val-PRO-SER-tyr-pro-PRO-thr-ala-ARG VAL-VAL-ASN-GLY- x - x -ALA-VAL-PRO-TYR- x -TRP-PRO-TRP-GLN-ile
    Porcine Elastase 1                             thr-glx-asx-phe-pro-glx-thr-asx-ala-ARG VAL-VAL-gly-GLY-thr-glu-ALA-glu-arg-asn-SER-TRP-PRO-ser-GLN-ile-arg-ser-gly-ser
Bovine Chymotrypsin A cys-gly-val-pro-ala-ile-PRO-PRO-val-leu-ser-gly-leu-SER-ARG ile-VAL-ASN-GLY-GLU-glu-ALA-VAL-PRO-gly-SER-TRP-PRO-TRP-GLN-VAL-lys
```

FIGURE 3: N-Terminal sequences of protease E, elastases, and chymotrypsin. The N-terminal sequences are aligned to maximize homology with the constraint that all of the proteins are cleaved by trypsin at Arg-15 (chymotrypsin numbering system; Hartley, 1970) during activation. The porcine elastase 1 proenzyme sequence is from Lamy et al. (1977); the lungfish elastase 1 sequence is from deHaen and Gertler (1974); the bovine and porcine protease E sequences are from Kobayashi et al. (1980) and Kobayashi et al. (1978), respectively; and the bovine chymotrypsinogen A sequence is from Hartley (1970). Amino acid residues which are identical between protease E and at least one of the other enzymes are capitalized. Residues marked "x" were reported as unidentified.

```
                                                                                         A   B   C
     16  17  18  19  20  21  22  23  24  25  26  27  28  29  30  31  32  33  34  35  36  36  36  36  37  38  39  40  41  42
HPE VAL VAL ASN GLY GLU asp ALA VAL PRO tyr SER TRP PRO TRP GLN VAL SER LEU GLN TYR glu lys ser GLY SER phe TYR HIS THR CYS
PE1 VAL VAL gly GLY thr glu ALA gln arg asn SER TRP PRO ser GLN ile SER LEU GLN TYR arg ser gly ser SER trp ala HIS THR CYS
RE1 VAL VAL gly GLY ala glu ALA arg arg asn SER TRP PRO ser GLN ile SER LEU GLN TYR leu ser gly GLY SER trp TYR HIS THR CYS
PE2 ile VAL ASX GLY GLX asx ALA VAL PRO gly SER TRP PRO TRP GLX VAL SER LEU GLX asx ser --- --- --- asx gly phe HIS phe CYS
BCA ile VAL ASN GLY GLU glu ALA VAL PRO gly SER TRP PRO TRP GLN VAL SER LEU GLN asp lys --- --- --- thr gly phe HIS phe CYS

                                 *                                                              *A                       *
     43  44  45  46  47  48  49  50  51  52  53  54  55  56  57  58  59  60  61  62  63  64  65  65  66  67  68  69  70  71
HPE GLY GLY SER LEU ILE gly pro asp TRP VAL VAL THR ALA gly HIS CYS ile ser SER ser trp THR tyr gln VAL VAL leu GLY GLU tyr
PE1 GLY GLY thr LEU ILE arg gln asn TRP VAL met THR ALA ala HIS CYS val asp arg glu leu THR phe arg VAL VAL val GLY GLU his
RE1 GLY GLY thr LEU ILE arg arg asn TRP VAL met THR ALA ala HIS CYS val ser SER gln met THR phe arg VAL VAL val GLY asp his
PE2 GLY GLY SER LEU ILE ser glx
BCA GLY GLY SER LEU ILE asn glu asn TRP VAL VAL THR ALA ala HIS CYS gly val thr thr ser asp val --- VAL VAL ala GLY GLU phe

     *                                                           •   A  *B
     72  73  74  75  76  77  78  79  80  81  82  83  84  85  86  87  87  87  88  89  90  91  92  93  94  95  96  97  98  99
HPE ASP arg ala val lys glu GLY pro GLU GLN val ile pro ILE asn ser gly asp leu PHE VAL HIS PRO leu TRP ASN arg ser cys VAL
PE1 asn leu asn gln asn asn GLY thr GLU GLN tyr val gly val gln lys --- --- ile val VAL HIS PRO tyr TRP ASN thr asp asp VAL
RE1 asn leu ser gln asn asp GLY thr GLU GLN tyr val ser val gln lys --- --- ile met VAL HIS PRO thr TRP ASN ser asn asn VAL
PE2                                                                                                                     met
BCA ASP gln gly ser ser ser glu lys ile GLN lys leu lys ILE ala lys --- --- val PHE lys asn ser lys tyr ASN ser leu thr ile

     A   B
     99  99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127
HPE ALA cys GLY ASN ASP ILE ALA LEU ile LYS LEU SER arg SER ALA GLN LEU gly asp ala VAL GLN LEU ALA ser LEU PRO pro ALA GLY
PE1 ALA ala GLY tyr ASP ILE ALA LEU leu arg LEU ala gln SER val thr LEU asn ser tyr VAL GLN LEU gly val LEU PRO arg ALA GLY
RE1 ALA ala GLY tyr ASP ILE ALA LEU leu arg LEU ala gln SER val thr LEU asn asn tyr VAL GLN LEU ALA val LEU PRO gln glu GLY
PE2 --- --- asn ASN ASP leu thr LEU leu LYS LEU ala ser pro ALA GLN tyr thr thr arg ile gly pro val cys LEU ala ser gly asp
BCA --- --- asn ASN ASP ILE thr LEU leu LYS LEU SER thr ala ALA ser phe ser gln thr VAL ser ala val cys LEU PRO ser ALA ser

                         *                                                                                               *
    128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157
HPE ASP ILE LEU PRO ASN glu THR PRO CYS TYR ILE THR GLY TRP GLY pro leu --- TYR THR ASN GLY pro LEU PRO ASP lys LEU GLN glu
PE1 thr ILE LEU ala ASN asn ser PRO CYS TYR ILE THR GLY TRP GLY leu thr arg --- THR ASN GLY gln LEU ala gln thr LEU GLN gln
RE1 thr ILE LEU ala ASN asn asn PRO CYS TYR ILE THR GLY TRP GLY arg thr arg --- THR ASN GLY gln LEU ser gln thr LEU GLN gln
PE2 glu ala LEU PRO glu -x- leu PRO CYS val
BCA ASP asp phe ala ala gly THR thr CYS val thr THR GLY TRP GLY leu thr arg TYR THR ASN ala asn thr PRO ASP arg LEU GLN gln

                                     *                   A   B
    158 159 160 161 162 163 164 165 166 167 168 169 170 170 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185
HPE ALA leu LEU PRO val VAL ASP TYR glu his CYS SER arg trp asn trp TRP GLY SER ser VAL LYS lys THR MET VAL CYS ALA GLY GLY
PE1 ALA tyr LEU PRO thr VAL ASP TYR ala ile CYS SER ser ser ser tyr TRP GLY SER thr VAL LYS asn ser MET VAL CYS ALA GLY GLY
RE1 ALA tyr LEU PRO ser VAL ASP TYR ser ile CYS SER ser ser ser tyr TRP GLY SER thr VAL LYS thr THR MET VAL CYS ALA GLY GLY
PE2                                                                                                 MET ile CYS ALA GLY ala
BCA ALA ser LEU PRO leu leu ser asn asp asn CYS lys lys --- --- tyr TRP GLY thr lys ile LYS asp ala MET ile CYS ALA GLY ala

                 A                                                              *A
    186 187 188 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 203 204 205 206 207 208 209 210 211 212 213
HPE ASP --- ile ARG SER GLY CYS ASN GLY ASP SER GLY GLY PRO LEU asn CYS pro thr glu asp GLY gly TRP gln VAL HIS GLY VAL THR
PE1 asn gly val ARG SER GLY CYS gln GLY ASP SER GLY GLY PRO LEU his CYS leu val --- asn GLY gln tyr ala VAL HIS GLY VAL THR
RE1 ASP gly val ARG SER GLY CYS gln GLY ASP SER GLY GLY PRO LEU his CYS leu val --- asn GLY gln tyr ser VAL HIS GLY VAL THR
PE2 ser gly ala --- SER ser CYS ASN GLY ASP SER GLY GLY PRO LEU val CYS leu lys gly asp --- thr TRP -x- leu -x- GLY ile val
BCA ser gly val --- SER ser CYS met GLY ASP SER GLY GLY PRO LEU val CYS lys lys --- asn GLY ala TRP thr leu val GLY ile val

                     A                   A                                                                       *   *
    214 215 216 217 217 218 219 220 221 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241
HPE SER PHE VAL SER ala phe GLY CYS ASN thr arg ARG LYS PRO THR VAL PHE THR ARG VAL SER ALA phe ILE asp TRP ILE glu glu THR
PE1 SER PHE VAL SER arg leu GLY CYS ASN val thr ARG LYS PRO THR VAL PHE THR ARG VAL SER ALA tyr ILE ser TRP ILE asn asn val
RE1 SER PHE VAL SER ser met GLY CYS ASN val ser lys LYS PRO THR VAL PHE THR ARG VAL SER ALA tyr ILE ser TRP met asn asn val
PE2 SER trp gly SER --- -x- -x- CYS ASX
BCA SER trp gly SER --- ser thr CYS ser --- thr ser thr PRO gly VAL tyr ala ARG VAL thr ALA leu val asn TRP val gln gln THR

    242 243 244 245
HPE ILE ALA SER his
PE1 ILE ALA SER asn
RE1 ILE ALA tyr thr
PE2
BCA leu ALA ala asn
```

FIGURE 4: Amino acid sequence comparison of protease E, elastases, and chymotrypsin. Amino acid sequence data are shown for the active enzyme portions of human protease E (HPE), porcine elastase 1 (PE1; Hartley, 1970), rat pancreatic elastase 1 (RE1; MacDonald et al., 1982a), porcine pancreatic elastase 2 (PE2; Vered et al., 1986), and bovine chymotrypsin A (BCA; Hartley, 1970). The chymotrypsin numbering system (Hartley, 1970) is used for comparison. Amino acid residues which are shared between protease E and at least one other enzyme are capitalized. Amino acid residues at which there is a negative charge change between protease E and both porcine and rat elastases 1 are marked with an asterisk. The conserved stretches of amino acids which comprise the S1 specificity pocket and portions of the extended binding region are boxed.

Table I: Amino Acid Homologies between Human Protease E and Pancreatic Serine Proteases[a]

	porcine[b] elastase 1	porcine[c] elastase 2	rat[d] elastase 1	rat[d] elastase 2	bovine[a] chymotrypsin A	bovine[a] chymotrypsin B	bovine[b] cationic trypsin	rat[e] anionic trypsin 1
protease E	55	45	56	57	40	44	37	36

[a] Percent identity, following optimal alignment of gaps following Hartley (1970), using for alignment the chymotrypsin A sequence from residues 16–245. [b] Sequence data from Hartley (1970). [c] Partial sequence data for 109 amino acids reported by Vered et al. (1986). Asx and Glx assignments were counted as mismatches. [d] Sequence data from MacDonald et al. (1982a). [e] Sequence data from MacDonald et al. (1982b).

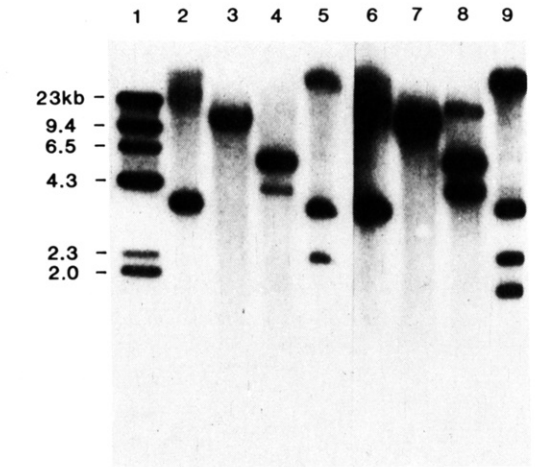

FIGURE 5: Genomic Southern hybridization. Placental genomic DNA was digested with the indicated restriction enzymes, subjected to gel electrophoresis, and transferred to nitrocellulose as described under Methods. Lanes 2–5 were probed with a partial cDNA (6-3) shown in Figures 1 and 2; lanes 6–9 were probed with a full-length cDNA for protease E. The following restriction enzymes were used: lanes 2 and 6, *Bam*HI; lanes 3 and 7, *Eco*RI; lanes 4 and 8, *Hin*dIII; and lanes 5 and 9, *Kpn*I. Lane 1 contains λ molecular weight markers end filled with [^{32}P]deoxycytidine as described under Methods.

1 have net charges of 8+. This observation is in agreement with the measured isoelectric points of human protease E of 4.9 (unpublished results) and rat and porcine elastases 1 of approximately 10 (Largman, 1983). In addition, there are substantial regions of homology between protease E and bovine chymotrypsin which are not present in porcine or rat elastases 1. This is reflected in the overall amino acid homology between protease E and other proteases. As shown in Table I, protease E does not possess a high degree of homology with any other pancreatic proteases (<57%). In fact, its homology is highest with the amino acid sequence encoded by rat elastase 2 cDNA (MacDonald et al., 1982a). Although this protein has not been isolated, it is predicted to be closely related to the chymotrypsin family. It should be noted that when compared with other serine proteases, protease E exhibits the pattern of amino acid insertions and deletions which is observed for elastases and not that found for chymotrypsins (Figure 4). In addition, protease E has a two amino acid insertion at 87A and 87B.

Mallory and Travis (1975) isolated human pancreatic protease E on the basis of substrate P1 specificity for alanine and reported that this acidic protein possessed high proteolytic activity but no elastolytic activity. Protease E was also isolated from both porcine and bovine pancreatic tissue (Kobayashi et al., 1978, 1980). In the same time period, investigators reported isolation of a protein with properties similar to those of protease E but which showed low elastolytic activity, which was referred to as "human pancreatic elastase 1" (Feinstein et al., 1974; Largman et al., 1976; Fujimoto et al., 1980). The results of the current study show that the N-terminal amino acid sequence of the protein referred to as human pancreatic elastase 1 (Largman et al., 1976) is identical with those of bovine and porcine proteases E and thus establishes the fact that "human elastase 1" and protease E are the same protein (Figure 3). It is of interest that an identical N-terminal amino acid sequence was reported for lungfish elastase A, an apparently cationic protein with limited elastolytic activity (de Haen & Gertler, 1974). There is limited homology between the human protease E activation peptide and those of porcine protease E and lungfish elastase A, but no homology between the proteases E and porcine elastase 1 in this region. These data, taken together with the net acidic charge and the low homology found between protease E and either elastases or chymotrypsins (Figure 4, Table I), indicate that the pancreatic proteases E represent a distinct subfamily of the serine proteases characterized by a specificity for small, neutral amino acids in the P1 position and an acidic isoelectric point.

Many investigators have searched unsuccessfully for a human pancreatic protease which would be similar to the classically described elastase 1, possessing a specificity for alanine and a basic net charge (Feinstein et al., 1974; Mallory & Travis, 1975; Largman et al., 1976). When we screened the human pancreas cDNA library, the rat elastase 1 probe detected both protease E and trypsin which had relatively low homology with the probe, but there was no evidence for a cDNA highly related to elastase 1. In addition, Southern gel analysis indicates that there is probably one copy of the protease E gene and no related genes. These results appear to confirm the earlier protein fractionation studies that there is no human homologue of pancreatic elastase 1.

A major goal of this work is the identification of the enzyme structure(s) which is(are) necessary for elastolysis. It was initially thought that one of the prerequisites was a specificity for alanine (Hartley & Shotton, 1971). However, the current study reveals that the amino acid sequence which comprises the S1 binding pocket for alanine substrates is essentially unchanged for protease E compared to porcine or rat elastases 1 (Figure 4, amino acid residues 209–217 and 223–233). It has long been known that porcine pancreatic elastase 1 is tightly bound to insoluble elastin (Hall, 1970). The residues of elastase which are important for interaction with elastin have not been identified. Gertler (1971a) demonstrated that maleylation of the three lysine residues (Lys-87, Lys-177, and Lys-224) prevents binding to elastin. While none of these lysines appears on the surface of porcine elastase near the proposed substrate binding site (Atlas, 1975), the fact that Lys-87 is changed to Ser-87 in protease E suggests that this residue might play a role in attachment to elastin.

Since it is thought that the overall net cationic charge is important for binding to elastin (Gertler, 1971b), we have noted in Figure 4 each of the amino acids at which there is a consistent anionic charge change between human protease E and the elastases. X-ray crystallographic studies of complexes of porcine pancreatic elastase 1 with substrates (Atlas, 1975) and inhibitors (Shotton et al., 1971; Hughes et al., 1982) indicate that all of these changes are outside the substrate binding region. These studies do, however, suggest several amino acid changes in protease E which may be important for binding to elastin. First, Leu-73, which is present in all elastases, is replaced by Arg-73 in protease E (Figure 4). Leu-73 lies in the bottom of the S4 subsite of the substrate binding cleft of porcine pancreatic elastase 1, in a hydrophobic

environment. Computer modeling studies suggest that the guanidinium group of Arg-73 would rotate into the polar solvent and thus disrupt the binding of substrate in the S4 site. Another potentially important change is Arg-217A/Ala-217A, since in porcine elastase 1 Arg-217A was identified as important for binding in the S3′ subsite (Atlas, 1975), and Davril et al., (1984) suggested that modification of Arg-217A blocked elastolytic activity. Finally, a loop of amino acids including Val-99 appears to be important for substrate/enzyme interaction. This is the region of protease E in which an extra pair of cysteines is located (Cys-98 and Cys-99B). Computer modeling studies indicate that these residues are positioned correctly for formation of a disulfide bond which would "lock" the two residues between these cysteines into a flat rigid intrachain loop, but no changes in the protein chain conformation were observed.

Dimicoli and Bieth (1977) have shown that porcine elastase 1 has a high-affinity Ca^{2+} binding site, supporting earlier evidence (Hall, 1970) that the binding of elastase to elastin is mediated by Ca^{2+} bridges. Arginine modification (Arg-65A and Arg-217A) has been shown to reduce both the metal ion binding capacity and the ability of porcine elastase to degrade elastin (Davril et al., 1984). These authors suggest that Ca^{2+} binds to the region bounded by Glu-70 and Glu-80 and that modification of Arg-65A disrupts the Ca^{2+} binding site. Protease E exhibits major differences from other elastases in this region: Arg-65A is changed to Gln-65A, and eight of the nine amino acids between Glu-70 and Glu-80 are different.

The question of which structural features confer the property of elastolysis on elastases can now be addressed by site-directed mutagenesis (Craik et al., 1985). Among the likely candidates for substitution would be Arg/Leu-73, Ala/Arg-217A, and the proposed disulfide bond between 98 and 100. Changes of this type should experimentally demonstrate if single amino acid changes control elastolysis or if this process depends on the overall net charge of the enzyme.

ACKNOWLEDGMENTS

We thank Myriam Alhadeff for sharing reagents and techniques and Bob Fletterick and Virginia Rath in the Department of Biochemistry, University of California, San Francisco, for helping and allowing us to use the computer modeling equipment. We also thank Ray MacDonald for providing the cDNA for rat pancreatic elastase 1 and Bob Weiss for providing the human pancreatic cDNA library.

REFERENCES

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

Biggin, M. D., Gibson, T. J., & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 3963–3965.

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.

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

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

Dimicoli, J. L., & Bieth, J. (1977) *Biochemistry 16*, 5532–5537.

Emi, M., Nakamura, Y., Ogawa, M., Yamamoto, T., Nishide, T., Mori, T., & Matsubara, K. (1986) *Gene 41*, 305–310.

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

Feinstein, G., Hofstein, R., Koifmann, J., & Sokolovsky, M. (1974) *Eur. J. Biochem. 43*, 569–581.

Fujimoto, K., Ogawa, M., Saito, N., Kosaki, G., Minamiura, N., & Yamamoto, T. (1980) *Biochim. Biophys. Acta 612*, 262–267.

Geokas, M. C., Largman, C., Brodrick, J. W., & Fassett, M. (1980) *Am. J. Physiol. 238*, G238–G246.

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

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

Hall, D. A. (1970) *Gerontologia (Basel) 16*, 325–339.

Hartley, B. S. (1970) *Philos. Trans. R. Soc. London, B 257*, 77–87.

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

Hughes, D. L., Sieker, L. C., Bieth, J., & Dimicoli, J.-L. (1982) *J. Mol. Biol. 162*, 645–658.

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

Kobayashi, R., Kobayashi, Y., & Hirs, C. H. W. (1978) *J. Biol. Chem. 253*, 5526–5530.

Kobayashi, Y., Kobayashi, R., & Hirs, C. H. W. (1980) *J. Biol. Chem. 256*, 2466–2470.

Lamy, F., Gibson, D., Ledoux, M., & Moreux, J. C. (1977) in *Elastin and Elastic Tissue* (Sandburg, 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.

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

MacDonald, R. J., Stary, S. J., & Swift, G. H. (1982b) *J. Biol. Chem. 252*, 9724–9732.

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.

Mathews, B. W., Sigler, P. B., Henderson, R., & Blow, D. M. (1967) *Nature (London) 214*, 652–656.

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

Rosenbloom, J. (1984) *Lab. Invest. 51*, 605–623.

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

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.

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

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