Proc. Natl. Acad. Sci. USA
Vol. 80, pp. 7486–7490, December 1983
Biochemistry

Identification of cDNA clones encoding secretory isoenzyme forms: Sequence determination of canine pancreatic prechymotrypsinogen 2 mRNA

(immunoselection and hybrid selection of mRNA/*in vitro* protein synthesis and processing/two-dimensional isoelectric focusing/NaDodSO$_4$ gel electrophoresis/cDNA sequence determination/transport peptide sequences)

SETH D. PINSKY, K. STEVEN LAFORGE, VAN LUC, AND GEORGE SCHEELE

Laboratory of Cell and Molecular Biology, The Rockefeller University, New York, NY 10021

Communicated by Philip Siekevitz, September 7, 1983

ABSTRACT **A cDNA library has been constructed from canine poly(A)$^+$ mRNA. Clones containing cDNA inserts coding for prechymotrypsinogen 2 (isoelectric point = 7.1; M_r = 27,500), one of three canine pancreatic isoenzyme forms, were selected by colony hybridization using a cDNA probe synthesized from immunoselected prechymotrypsinogen 2 mRNA. To verify that cDNA clones code for prechymotrypsinogen 2 forms that translocate across rough endoplasmic reticulum membranes and fold into stable and identifiable secretory proteins, we conducted *in vitro* translation of hybrid-selected mRNA in the presence of microsomal membranes and optimal concentrations of glutathione and analyzed nascent translation products in their nonreduced state by two-dimensional isoelectric focusing/NaDodSO$_4$ gel electrophoresis and fluorography. A near full-length chymotrypsinogen 2 cDNA and its primed extension were used to determine the nucleotide sequence for the entire coding region of prechymotrypsinogen 2 mRNA and 87 residues, including a poly(A) addition signal, in the 3′ nontranslated region. The deduced amino acid sequence shows a 263-residue presecretory protein containing an 18-residue amino-terminal transport peptide (Met-Ala-Phe-Leu-Trp-Leu-Leu-Ser-Cys-Phe-Ala-Leu-Leu-Gly-Thr-Ala-Phe-Gly), which we have previously shown to mediate the translocation of chymotrypsinogen 2 across the rough endoplasmic reticulum membrane. Following the transport peptide is a 245-residue proenzyme, which shows 82% and 80% sequence identity with bovine chymotrypsinogens A and B, respectively. Conserved among the three zymogens are 10 Cys residues that form five disulfide bonds in bovine chymotrypsinogens A and B and the residues that are required for zymogen activation, substrate binding, and catalytic activity.**

Chymotrypsin(ogen) (EC 3.4.21.1) is synthesized and secreted by the exocrine pancreas and, upon delivery to the intestinal tract, is activated by trypsin to form chymotrypsin, a proteolytic enzyme important in the digestive process. In the past 20 years chymotrypsinogen and chymotrypsin have represented prototype molecules for the study of (pro)enzyme structure and function. The complete amino acid sequence of bovine chymotrypsinogen A was determined by Hartley (1) and that for bovine chymotrypsinogen B by Smillie *et al.* (2). X-ray crystallography structures of the zymogen (3) and the enzyme (4) were determined by Blow and collaborators. Scheele *et al.* (5) showed that chymotrypsinogen is synthesized as a presecretory protein containing an amino-terminal transport peptide responsible for translocation of the nascent protein across the membrane of the rough endoplasmic reticulum. The partial amino acid sequence of the transport peptide has been recently determined by Carne and Scheele (6).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

In this study we have developed a strategy to correlate the coding potentials of individual cDNAs with specific isoenzymic forms containing topogenic signals responsible for protein translocation across the rough endoplasmic reticulum membrane and export from the cell. This strategy includes *in vitro* translation of hybridization-selected mRNA in a reticulocyte lysate protein-synthesizing system in the presence of nuclease-treated microsomal membranes from canine (dog) pancreas and optimal concentrations of glutathione (5, 7), separation of translation products by two-dimensional urea isoelectric focusing (IEF)/NaDodSO$_4$ gel electrophoresis in the absence of reducing agents (7, 8), and analysis of radioactive products by fluorography. Comigration of nascent proteins with mature secretory proteins under these conditions depends on a number of sequential and overlapping biochemical events: (*i*) microsomal membrane recognition of the translation product as a presecretory protein containing an amino-terminal transport peptide (or its equivalent); (*ii*) functional translocation of the nascent product across the microsomal membrane; (*iii*) proteolytic cleavage of the transport peptide; (*iv*) sequestration within the microsomal lumen; and (*v*) development of conformation, including formation of the correct set of disulfide bonds resulting in the formation of a stable protein soluble in the aqueous cisternal medium (7, 9). Where studied, such pancreatic proteins (including chymotrypsinogen 2), synthesized and processed *in vitro*, have demonstrated biological and probiological activities (10). The resulting isoelectric point (pI) and apparent molecular weight of the final product, as judged by migration on two-dimensional gels, indicate the precise exportable protein that is encoded by the cDNA. After selection of a cDNA clone coding for one of the three isoenzymic forms of canine chymotrypsinogen, we have determined the nucleotide sequence for the entire coding region of prechymotrypsinogen 2 mRNA and noncoding bases in the 3′ flanking region.

METHODS

Construction of a Canine Pancreatic cDNA Library. Canine pancreatic RNA was extracted from a postnuclear supernate as described (5, 11) and poly(A)$^+$ mRNA was isolated by chromatography on oligo(dT)-cellulose (12). Double-stranded cDNA was synthesized (13) using the mixture of canine pancreas poly(A)$^+$ mRNA and inserted into pUC9 plasmids (14) at the *Pst* I site by G-C tailing (15). JM83 *Escherichia coli* were transformed with the pUC9 plasmid mixture by standard procedures (16) and clear colonies containing recombinant plasmid DNA

Abbreviation: IEF, isoelectric focusing.

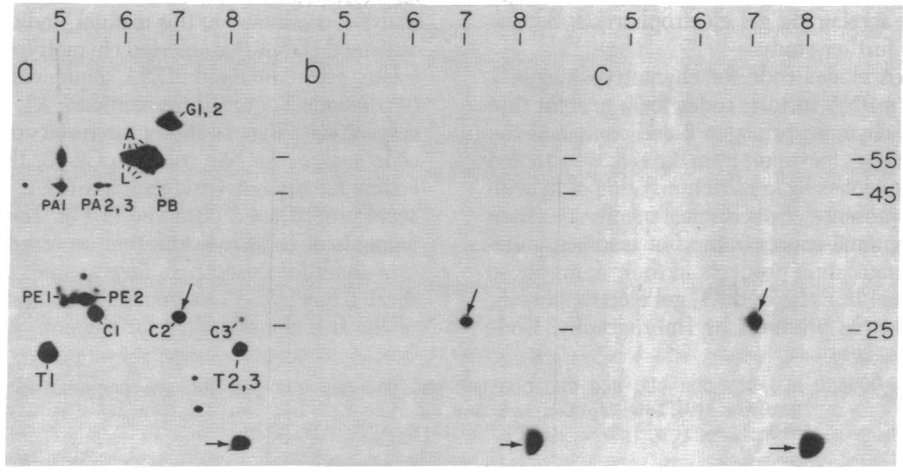

FIG. 1. Immunoselection of canine pancreatic C2 mRNA, colony selection by hybridization with a C2 cDNA probe, and code verification of cDNA clones by *in vitro* translation of hybrid-selected mRNA. *a–c* show fluorographic patterns of radioactive products synthesized in an *in vitro* reticulocyte lysate translation system in the presence of nuclease-treated microsomal membranes from canine pancreas (20 A_{260} units/ml) and optimal concentrations of glutathione and separated by two-dimensional urea IEF/NaDodSO$_4$ gel electrophoresis under nonreducing conditions (5, 7, 8). Under these *in vitro* conditions, transport peptides are removed from presecretory proteins by proteolytic cleavage and exportable proteins sequestered in microsomal vesicles develop conformational states indistinguishable from their counterparts synthesized *in vivo*. (*a*) Products derived from the translation of the entire mixture of canine pancreas poly(A)$^+$ mRNA. (*b*) Products synthesized *in vitro* in the presence of mRNA isolated from immunoselected C2 polysomes. Single-stranded ^{32}P-labeled cDNA (10^8 cpm/μg of DNA), synthesized from the purified C2 mRNA with avian myeloblastosis virus reverse transcriptase, was hybridized to 1,000 colonies previously grown in microtiter dishes, transferred to nitrocellulose filters overlaid on agar plates, and grown to a colony size of ≤1.5 mm. Prehybridization, hybridization, and washing conditions were as described by Su *et al.* (22). Identification of cDNA-containing clones was verified by code analysis as follows: plasmid DNA (1.8 μg) isolated from bacterial colonies by alkaline NaDodSO$_4$ extraction, purified on CsCl gradients, and bound to 3 × 3 mm nitrocellulose squares was used for hybrid selection of canine pancreas poly(A)$^+$ mRNA as described by Maniatis *et al.* (23). Hybrid-selected mRNA was translated *in vitro* and radioactive products were separated in two-dimensional gels and analyzed as described in the text. (*c*) Translation products derived from mRNA hybridized to pC2-13, which was used to determine the nucleotide sequence of canine prechymotrypsinogen 2. Numbers on the upper abscissae indicate pIs. Numbers on the right ordinate indicate molecular weight values × 10^{-3} for nonreduced proteins. In *a*, the proteins are labeled according to their actual or potential enzyme activities by the following abbreviations: T, trypsinogen; C, chymotrypsinogen; PE, proelastase; PA, procarboxypeptidase A; A, amylase; G, glycoprotein. Oblique arrows indicate chymotrypsinogen 2. Horizontal arrows indicate globin synthesized from residual globin mRNA in the reticulocyte lysate.

(ampicillin resistance, lac$^-$) were selected in the presence of 5-bromo-4-chloro-3-indolyl-β-D-galactoside. Colonies were transferred to microtiter dishes and grown in LB medium with ampicillin.

Characterization of cDNA Clones. Clones were screened by colony hybridization and colony identification was verified by hybrid selection/code analysis as described in the legend to Fig. 1. Recombinant plasmid DNAs, extracted with alkaline NaDodSO$_4$, were sized on agarose and acrylamide gels by using molecular weight markers. The nucleotide sequences of cDNA transcripts were determined by the chemical modification technique of Maxam and Gilbert (17) and the primer extension procedure of Hagenbuchle *et al.* (18) as modified by MacDonald *et al.* (19). Translation of nucleotide sequences into amino acid sequences and identification of restriction enzyme cleavage sites were conducted using computer programs (20).

Immunoselection of Canine Pancreatic C2 mRNA, Colony Selection by Hybridization, and Code Verification of cDNA Clones. Canine exocrine pancreatic proteins were extracted from a crude zymogen granule fraction and separated by two-dimensional IEF/NaDodSO$_4$ gel electrophoresis as described by Carne and Scheele (6). Monospecific polyclonal antibodies were prepared by excision of chymotrypsinogen 2 spots and repeated injection of an emulsified protein/acrylamide mixture into rabbits (6). Antiserum was applied to a protein A-Sepharose affinity column, washed extensively to remove serum RNase, and a purified immunoglobulin fraction was eluted in 0.1 M glycine (pH 3.0) (21). Antiserum was used to isolate chymotrypsinogen 2 polysomes from canine pancreas using the procedure of Kraus and Rosenberg (21).

RESULTS

Fig. 1*a* shows the majority of exportable proteins in the canine pancreas as separated by two-dimensional urea IEF/NaDodSO$_4$ gel electrophoresis. Among the serine protease zymogens, three forms of chymotrypsinogen were identified: C1, C2, and C3.*

To study the cDNA and genomic clones that code for chymotrypsinogen 2 mRNA, we excised the chymotrypsinogen 2 spot (pI = 7.1; M_r = 27,500) from two-dimensional gels, raised monospecific antibodies to this protein in rabbits, and used the IgG fraction from the antiserum to immunoselect chymotrypsinogen 2 polysomes from a canine pancreas postmitochondrial supernatant fraction. The success in the production of antibody and in the immunoselection procedure is demonstrated in Fig. 1*b*, which shows a fluorogram of the translation products directed by the purified mRNA during *in vitro* protein synthesis. In addition to globin that is synthesized by the residual undegraded globin mRNA in the reticulocyte lysate, only a single product appeared, which, judged by its pI (7.1) and M_r (27,500), is chymotrypsinogen 2.

Purified chymotrypsinogen 2 mRNA was used to prepare a cDNA probe for the selection of colonies containing plasmids that code for chymotrypsinogen 2 mRNA. Colony hybridization conducted on 1,000 colonies from a cDNA library revealed 22 colonies with strong hybridization signals. On the basis of insert

* Isoenzyme forms are numbered consecutively from anode to cathode following the recommendations of the IUPAC-IUB commission on the biochemical nomenclature of multiple forms of enzymes separated by polyacrylamide gel electrophoresis (24).

size as determined by acrylamide gel electrophoresis, 8 colonies were selected for further study.

To verify that cDNA clones code for chymotrypsinogen 2 mRNA and that such mRNA in turn codes for a protein that could be identified as chymotrypsinogen 2 and contained the appropriate structural signals for export from the cell, we checked each of the eight cDNA clones by *in vitro* translation of hybrid-selected mRNA in the presence of microsomal membranes from canine pancreas and optimal concentrations of oxidized glutathione, separation of translation products in their nonreduced state by two-dimensional IEF/NaDodSO$_4$ gel electrophoresis, and analysis of radioactive products by fluorography. Code analysis, conducted in this manner, indicated that five colonies contained cDNA that selected chymotrypsinogen 2 mRNA. One colony each contained cDNA sequences that selected chymotrypsinogen 1, procarboxypeptidase A1, and amylase mRNAs, respectively. Fig. 1*c* shows the results of hybrid selection and code analysis for one colony, pC2-13, that contained a cDNA coding for chymotrypsinogen 2 mRNA. A small amount of nascent proelastase 2 observed in Fig. 1*c* indicates that minimal amounts of cross-hybridization occurred under the hybridization conditions used (23). Size analysis on acrylamide gels indicated that pC2-13 contained the longest insert (937 base pairs) of the five clones coding for chymotrypsinogen 2 mRNA. Ac-

```
CGUACC AUG GCU UUC CUC UGG CUC CUC UCC UGC UUC GCC CUC CUG GGC ACA GCC UUC GGC
       Met Ala Phe Leu Trp Leu Leu Ser Cys Phe Ala Leu Leu Gly Thr Ala Phe Gly
       -18                             -10                                -1

UGC GGG GUC CCU GCC AUC CAG CCG GUG UUA AGU GGC CUG UCC AGG AUC GUC AAU GGG GAA
Cys Gly Val Pro Ala Ile Gln Pro Val Leu Ser Gly Leu Ser Arg Ile Val Asn Gly Glu
+1                                  10                                      20

GAC GCU GUC CCG GGC UCC UGG CCC UGG CAG GUG UCC CUG CAG GAC AGC ACC GGC UUC CAC
Asp Ala Val Pro Gly Ser Trp Pro Trp Gln Val Ser Leu Gln Asp Ser Thr Gly Phe His
21                                  30                                      40

UUC UGC GGG GGG UCC CUC AUC AGC GAG GAC UGG GUG GUC ACU GCU GCC CAC UGC GGG GUC
Phe Cys Gly Gly Ser Leu Ile Ser Glu Asp Trp Val Val Thr Ala Ala His Cys Gly Val
41                                  50                                      60

AGA ACC ACC CAC CAG GUU GUA GCC GGG GAG UUC GAC CAG GGC UCA GAU GCU GAG AGC AUC
Arg Thr Thr His Gln Val Val Ala Gly Glu Phe Asp Gln Gly Ser Asp Ala Glu Ser Ile
61                                  70                                      80

CAG GUG CUG AAG AUU GCC AAG GUU UUC AAG AAC CCC AAG UUC AAC AUG UUC ACC AUC AAC
Gln Val Leu Lys Ile Ala Lys Val Phe Lys Asn Pro Lys Phe Asn Met Phe Thr Ile Asn
81                                  90                                      100

AAC GAC AUC ACC CUG CUG AAG CUG GCC ACA CCC GCC CGC UUC UCC AAG ACC GUG UCC GCC
Asn Asp Ile Thr Leu Leu Lys Leu Ala Thr Pro Ala Arg Phe Ser Lys Thr Val Ser Ala
101                                 110                                     120

GUG UGC CUG CCC CAG GCG ACC GAC GAC UUC CCU GCU GGG ACC CUG UGU GUC ACC ACG GGC
Val Cys Leu Pro Gln Ala Thr Asp Asp Phe Pro Ala Gly Thr Leu Cys Val Thr Thr Gly
121                                 130                                     140

UGG GGC CUG ACC AAA CAC ACC AAU GCC AAC ACC CCC GAC AAG CUA CAG CAG GCG GCC CUG
Trp Gly Leu Thr Lys His Thr Asn Ala Asn Thr Pro Asp Lys Leu Gln Gln Ala Ala Leu
141                                 150                                     160

CCC CUC CUG UCC AAC GCC GAG UGC AAG AAG UUC UGG GGC AGC AAG AUC ACC GAU CUU AUG
Pro Leu Leu Ser Asn Ala Glu Cys Lys Lys Phe Trp Gly Ser Lys Ile Thr Asp Leu Met
161                                 170                                     180

GUC UGU GCG GGU GCC AGU GGC GUC UCC UCC UGC AUG GGU GAC UCU GGU GGC CCC CUG GUC
Val Cys Ala Gly Ala Ser Gly Val Ser Ser Cys Met Gly Asp Ser Gly Gly Pro Leu Val
181                                 190                                     200

UGC CAG AAG GAU GGA GCC UGG ACU CUG GUG GGC AUC GUG UCC UGG GGC AGC GGU ACC UGC
Cys Gln Lys Asp Gly Ala Trp Thr Leu Val Gly Ile Val Ser Trp Gly Ser Gly Thr Cys
201                                 210                                     220

UCC ACC UCC ACU CCU GGC GUG UAC GCC CGU GUC ACC AAG CUC AUA CCC UGG GUA CAG CAG
Ser Thr Ser Thr Pro Gly Val Tyr Ala Arg Val Thr Lys Leu Ile Pro Trp Val Gln Gln
221                                 230                                     240

AUC CUA CAA GCC AAC UGA GUCCCCCACCCCCACCCUGACCUCUCUGCAGACCCUGCUUCCACAGAGUCUCCAU
Ile Leu Gln Ala Asn ---
241             245

AAAGCAAUAAAAGACACAUUGAUGAUGAA
```

FIG. 2. The nucleotide sequence of chymotrypsinogen 2 mRNA and the deduced amino acid sequence of the preproenzyme. Shown are 882 nucleotides of canine pancreas C2 mRNA, which include 6 bases in the 5′ noncoding region and 87 bases including a polyadenylylation signal (underlined) in the 3′ noncoding region. The numbered amino acid sequence for prechymotrypsinogen 2 derived from the large open reading frame is shown below the nucleotide sequence. The zymogen, chymotrypsinogen 2, is comprised of amino acid residues 1 through 245. The transport peptide, comprised of residues −1 through −18, begins with the AUG initiation codon. The two vertical arrows indicate sites where proteolytic processing occurs. Proteolytic cleavage of the transport peptide occurs at the Gly-Cys linkage. Activation of the zymogen occurs by tryptic cleavage at the Arg-Ile bond.

cordingly, it was chosen for nucleotide sequence analysis.

Although the sequence data revealed a large open reading frame that encoded an amino acid sequence similar to that of bovine chymotrypsinogen A, the AUG initiation codon was not identified. A *Sau*96I–*Ava* I restriction fragment was labeled at the 5′ end and used as a primer for the synthesis of a single-stranded cDNA extension toward the 5′ end of the mRNA. The extended fragment was isolated on an acrylamide gel, electrophoresed onto DE81 paper, and eluted into buffer containing 1 M NaCl. Sequence analysis showed, in addition to the 66-base primer, an extension that overlapped 54 bases with the 5′ end of the pC2-13 cDNA and a further 14-base sequence that revealed the AUG initiation codon and 7 residues into the 5′ noncoding region of the mRNA.

Fig. 2 shows 882 nucleotides of canine pancreatic chymotrypsinogen 2 mRNA, including 735 nucleotides representing the amino acid coding region, 6 nucleotides in the 5′ noncoding region, and 87 nucleotides including a polyadenylylation signal (A-A-U-A-A-A) in the 3′ noncoding region.

DISCUSSION

Characterization of exocrine pancreatic proteins from five species [guinea pig (25), rat (26), rabbit (27), dog (5), and human (28)] by two-dimensional IEF/NaDodSO$_4$ gel electrophoresis indicates that multiple discrete forms are generally present in each species for each of the protease zymogens: trypsinogen, chymotrypsinogen, proelastase, and procarboxypeptidase A. Each of these multiple forms is apparently derived from separate mRNA species (29). Among the three forms of chymotrypsinogen, identified in the dog (5), form 1 shows a pI of 5.5 and an M_r of 27,500 and accounts for 2.4% of the exportable protein mass, as judged by the incorporation of a mixture of 15 ^{14}C-labeled amino acids into protein contained in pancreatic slices. Form 2 shows a pI of 7.1 and an M_r of 27,500 and accounts for 4.7% of the exportable protein mass; form 3 shows a pI of 9.5 and an M_r of 27,500 and accounts for 0.4% of the exportable mass.

The importance of using techniques capable of identifying cDNA clones that code for specific secretory enzymes and their isoenzymic forms is demonstrated in the data presented in this paper. Despite the use of a cDNA probe synthesized from purified chymotrypsinogen 2 mRNA and standard hybridization conditions, colony hybridization studies identified not only chymotrypsinogen 2 clones within the cDNA library but also, in several cases, clones that represented other secretory proteins. These findings indicate the limitation in colony hybridization studies and the importance of code verification as described here. Using procedures described in the *Methods* and in the legend to Fig. 1, we have verified the identification of cDNA clones that code for chymotrypsinogen 2 mRNA in a biological system that synthesizes three chymotrypsinogen forms. Separation of translation products by one-dimensional polyacrylamide gel electrophoresis in NaDodSO$_4$ fails to distinguish among the three chymotrypsinogen forms because they show identical molecular weight values.

Previous work describing the cDNA structures coding for rat pancreatic procarboxypeptidase A (30), proelastases I and II (31), and trypsinogen 1 (19) identified these structures in terms of the extent of their sequence identity with corresponding bovine zymogens but failed to relate individual cDNA clones to specific rat secretory proteins. Ultimately, correlations between 5′ flanking sequences of genes coding for pancreatic enzymes and developmental and regulatory changes in synthetic rates for specific secretory proteins will require that such identifications be made.

The nucleotide sequence presented in Fig. 2 reveals a single amino acid reading frame uninterrupted by termination codons, which codes for a presecretory protein with 263 amino acids and a M_r of 27,785. This molecular weight value is in agreement with the M_r 27,500 value estimated for canine chymotrypsinogen 2 by polyacrylamide gel electrophoresis in NaDodSO$_4$ (5). The presecretory protein includes an 18-residue amino-terminal transport peptide and a 245-residue zymogen or proenzyme. The canine zymogen sequence shows 82% amino acid sequence identity with bovine chymotrypsinogen A and 80% sequence identity with bovine chymotrypsinogen B (Fig. 3). Based on numbers of charged amino acid residues, bovine chymotrypsinogen A shows a net charge of +3, canine chymotrypsinogen 2 shows a net charge of −1, and bovine chymotrypsinogen B shows a net charge of −6. These findings agree with the known isoelectric points of chymotrypsinogen A (pI = 9.1, ref. 32), C2 (pI = 7.1, ref. 5), and chymotrypsinogen B (pI = 5.2, ref. 32).

Amino acid residues previously shown to be important for zymogen and enzyme function in bovine chymotrypsinogens A and B are conserved in our sequence of canine chymotrypsinogen 2. Zymogen activation occurs by tryptic cleavage of the ^{15}Arg-^{16}Ile bond (for a review, see ref. 33). Of interest is our finding that the segment ^{182}Cys to ^{201}Cys, which is largely hydrophobic (^{194}Asp is an exception) and contains the substrate binding site, is conserved among the three chymotrypsinogen sequences. This 20-residue segment contains 11 small amino acid residues (Ala, Gly, Ser), which may provide the flexibility necessary for formation of the hydrophobic pocket. Catalytic activity relies on the ability of ^{57}His to transfer protons from ^{102}Asp to ^{195}Ser. All 8 tryptophan residues and a number of hydrophobic segments, which confer thermodynamic stability to the molecule, are conserved among the three sequences (see Fig. 3) as well as 10 Cys residues, which form 5 disulfide bonds in chymotrypsinogen A and

```
          10        20        30        40
           *         *         *         *
                     E                   K           N N
  1 CGVPAIQPVLSGLSRIVNGEDAVPGSWPWQVSLQDSTGFHFCGGSLISED
                A

            T SDV          SS K  K         S Y SL
 51 WVVTAAHCGVRTTHQVVAGEFDQGSDAESIQVLKIAKVFKNPKFNMFTIN
            T SDV          LET DT     G         SIL VR

          S A S  Q        S S   A   T          RY
101 NDITLLKLATPARFSKTVSAVCLPQATDDFPAGTLCVTTGWGLTKHTNAN
              Q  E        S DE     M  A     K  YNALK

     R    S      TN   Y  T  K A I
151 TPDKLQQAALPLLSNAECKKFWGSKITDLMVCAGASGVSSCMGDSGGPLV
          T  IV  TD R Y   RV  V I

    K N             S                A VN    T A
201 CQKDGAWTLVGIVSWGSGTCSTSTPGVYARVTKLIPWVQQILQAN
      N     A       S       A       A M    ET A
```

FIG. 3. Comparison of the amino acid sequence of canine chymotrypsinogen 2 with bovine chymotrypsinogens A and B. Amino acid residues found in the bovine chymotrypsinogen A and chymotrypsinogen B sequences that differ from those shown for the canine C2 sequence are shown above and below the canine sequence, respectively. The majority of these residue substitutions can be conservatively explained by single base changes. Where two base changes are required, the corresponding amino acid residue is underlined. Amino acids are shown in the one-letter code as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr. The vertical arrow indicates the site where activation of the zymogen occurs by tryptic cleavage.

B (positions 1–122, 42–58, 136–201, 168–182, and 191–220, ref. 34).

The 18-residue amino-terminal transport peptide begins at the initiator methionine and contains 16 hydrophobic residues and two polar residues, Ser and Thr. Nine of the residues are hydrophobic amino acids containing bulky side chains (Phe, Leu, Trp). Previous amino acid sequence analysis studies on canine pancreatic prechymotrypsinogen 2 and chymotrypsinogen 2 synthesized *in vitro* in the absence and presence of microsomal membranes, respectively (6), determined the site of cleavage of the transport peptide. This site occurs at the $\overset{-1}{\text{Gly}}\text{-}\overset{+1}{\text{Cys}}$ bond shown in Fig. 2. Despite the incomplete nature of the previously described amino acid sequence, the positions of 9 residues agree with those assigned by the nucleotide sequence analysis studies reported here. As described by Steiner and coworkers (35) and Carne and Scheele (9), a region with a high probability of forming a β turn occurs at or near the transport peptide cleavage site of a number of presecretory proteins. The sequence Pro-Ala-Ile-Gln at positions 4–7 in the chymotrypsinogen 2 sequence shows such probability and its existence in chymotrypsinogen A has been confirmed by x-ray crystallography (4). The importance of this β turn is underscored by the fact that the first 13 residues, including the β turn region, are conserved in the three chymotrypsinogen sequences (see Fig. 3).

The authors thank Mr. Russell F. Jacoby for the preparation of immunoselected chymotrypsinogen 2 mRNA and Drs. Richard Broglie, Anthony Cashmore, and Ken Bernstein for helpful discussions. This research was supported by research Grant AMDD 18532 from the National Institutes of Health.

1. Hartley, B. S. (1964) *Nature (London)* **201,** 1284–1287.
2. Smillie, L. B., Furka, A., Nagabushan, N., Stevenson, K. J. & Parks, C. O. (1968) *Nature (London)* **218,** 343–346.
3. Kraut, J. (1971) in *The Enzymes*, ed. Boyer, P. D. (Academic, New York), pp. 165–184.
4. Blow, D. M. (1971) in *The Enzymes*, ed. Boyer, P. D. (Academic, New York), pp. 185–212.
5. Scheele, G., Jacoby, R. & Carne, T. (1980) *J. Cell Biol.* **87,** 611–628.
6. Carne, T. & Scheele, G. (1982) *J. Biol. Chem.* **257,** 4133–4140.
7. Scheele, G. & Jacoby, R. (1982) *J. Biol. Chem.* **257,** 12277–12282.
8. Bieger, W. & Scheele, G. (1980) *Anal. Biochem.* **257,** 222–230.
9. Carne, T. & Scheele, G. (1983) in *The Secretory Process*, ed. Cantin, M. (Karger, Basel, Switzerland), in press.
10. Scheele, G. & Jacoby, R. (1983) *J. Biol. Chem.* **258,** 2005–2009.
11. Lizardi, P. & Engelberg, A. (1979) *Anal. Biochem.* **98,** 116–122.
12. Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 1408–1412.
13. Wickens, M. P., Buell, G. N. & Schimke, R. T. (1978) *J. Biol. Chem.* **258,** 2483–2495.
14. Vieira, J. & Messing, J. (1982) *Gene* **19,** 259–268.
15. Otsuka, A. (1981) *Gene* **13,** 339–346.
16. Schleif, R. F. & Wensink, P. C. (1981) *Practical Methods in Molecular Biology* (Springer, New York), pp. 134–136.
17. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560.
18. Hagenbuchle, O., Bovery, R. & Young, R. A. (1980) *Cell* **21,** 179–187.
19. MacDonald, R. J., Stary, S. J. & Swift, G. H. (1982) *J. Biol. Chem.* **257,** 9724–9732.
20. Pustell, J. & Kafatos, F. C. (1982) in *The Application of Computers to Research on Nucleic Acids*, eds. Soll, D. & Roberts, R. J. (IRL, Washington, DC), pp. 51–60.
21. Kraus, J. P. & Rosenberg, L. E. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 4015–4019.
22. Su, T. S., Bock, H. O., O'Brien, W. E. & Beaudet, A. L. (1981) *J. Biol. Chem.* **256,** 11826–11831.
23. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning, A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 329–343.
24. IUPAC-IUB Commission on Biochemical Nomenclature (1978) *Eur. J. Biochem.* **82,** 1–3.
25. Scheele, G. (1975) *J. Biol. Chem.* **250,** 5375–5385.
26. Scheele, G. (1982) in *The Secretory Granule*, eds. Poisner, A. & Trifaro, J. (Elsevier Biomedical, Amsterdam), pp. 213–246.
27. Iacino, D., Scheele, G. & Liebow, C. (1980) *Am. J. Physiol.* **239,** G247–G254.
28. Scheele, G., Bartelt, D. & Bieger, W. (1981) *Gastroenterology* **80,** 461–473.
29. Scheele, G. (1981) in *Methods in Cell Biology*, eds. Hand, A. & Oliver, C. (Academic, New York), Vol. 23, pp. 345–358.
30. Quinto, C., Quiroga, M., Swain, W. F., Nikovits, W. C., Standring, D. N., Pictet, R. L., Valenzuela, P. & Rutter, W. J. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 31–35.
31. MacDonald, R. J., Swift, G. H., Quinto, C., Swain, W., Pictet, R. L., Nikovits, W. & Rutter, W. J. (1982) *Biochemistry* **21,** 1453–1463.
32. Kubacki, V., Brown, K. D. & Laskowski, M. (1949) *J. Biol. Chem.* **180,** 73–78.
33. Hess, G. P. (1971) in *The Enzymes*, ed. Boyer, P. D. (Academic, New York), pp. 213–249.
34. Brown, J. R. & Hartley, B. S. (1966) *Biochem. J.* **101,** 214–228.
35. Chan, S. J., Duguid, J. R., Quinn, P., Labrecque, A., Noyes, B., Keim, P., Heinrikson, R. L. & Steiner, D. F. (1979) in *From Gene to Protein: Information Transfer in Normal and Abnormal Cells*, eds. Russel, T. R., Brew, K., Faber, H. & Schultz, J. (Academic, New York), pp. 361–378.