Green, C. L., Loechler, E. L., Fowler, K. W., & Essigmann, J. M. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 13–17.
Gross, M. L., Chess, E. K., Lyon, P. A., Crow, F. W., Evans, S., & Tuge, H. (1982) *Int. J. Mass Spectrom. Ion Phys. 42*, 243–254.
Hill-Perkins, M., Jones, M. D., & Karran, P. (1986) *Mutat. Res. 162*, 153–163.
Jay, E., Bambera, R., Padmanabhan, R., & Wu., R. (1974) *Nucleic Acids Res. 1*, 331–353.
Johnson, D. L., Reid, T. M., Lee, M.-S., King, C. M., & Romano, L. J. (1986) *Biochemistry 25*, 449–456.
Kadlubar, F. F., Beland, F. A. Beranek, D. T., Dooley, K. L., Heflich, R. H., & Evans, F. E. (1982) in *Environmental Mutagens and Carcinogens* (Sugimura, T., Kondo, S., & Takebe, H., Eds.) pp 385–396, Liss, New York.
Kuzmich, S., Marky, L. A., & Jones, R. A. (1983) *Nucleic Acids Res. 11*, 3393–3404.
Lee, M.-S., & King, C. M. (1981) *Chem.-Biol. Interact. 34*, 239–248.
Loechler, E., Green, C. L., & Essigmann, J. M. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 6271–6275.
Lundquist, R. C., & Olivera, B. M. (1982) *Cell (Cambridge, Mass.) 31*, 53–60.
Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) in *Molecular Cloning, a Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
McCann, J., Choi, E., Yamasaki, E., & Ames, B. N. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 5135–5139.
Melick, W. F., Naryka, J. J., & Kelly, R. E. (1971) *J. Urol. (Baltimore) 106*, 220–226.
Messing, J. (1983) *Methods Enzymol. 101*, 20–77.
Miller, E. C. (1978a) *Cancer Res. 38*, 1479–1496.
Miller, J. H. (1978b) in *The Operon* (Miller, J. H., & Reznikoff, W. S., Eds.) pp 31–88, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Miller, J. H. (1983) *Annu. Rev. Genet. 17*, 215–238.
Pai, V., Bloomfield, S. F., & Gorrod, J. W. (1985) *Mutat. Res. 151*, 201–207.
Patrianakos, C., & Hoffmann, D. (1979) *J. Anal. Toxicol. 3*, 150–154.
Rosenkranz, H. S., McCoy, E. C., Frierson, M., & Klopman, G. (1985) *Environ. Mutagen. 7*, 645–653.
Scribner, J. D., Fisk, S. R., & Scribner, N. K. (1979) *Chem.-Biol. Interact. 26*, 11–25.
Shapiro, R., Underwood, G. R., Zawadzka, H., Broyde, S., & Hingerty, B. E. (1986) *Biochemistry 25*, 2198–2205.
Sharma, M., & Box, H. C. (1985) *Chem.-Biol. Interact. 56*, 73–88.
Sproat, B. S., & Gait, M. J. (1984) in *Oligonucleotide Synthesis, a Practical Approach* (Gait, M. J., Ed.) pp 83–115, IRL Press, Washington, DC.
Sukumar, S., Notario, V., Martin-Zanca, D., & Barbacid, M. (1983) *Nature (London) 300*, 658–661.
Tabin, C. J., Bradley, S. M., Bargmann, C. I., Weinberg, R. A., Papageorge, A. G., Scolnick, E. M., Dhar, R., Lowy, D. R., & Chang, E. H. (1982) *Nature (London) 300*, 143–149.
Zarbl, H., Sukumar, S., Arthur, A. V., Martin-Zanca, D., & Barbacid, M. (1985) *Nature (London) 315*, 382–385.

Isolation and Characterization of a cDNA Encoding Rat Cationic Trypsinogen†

Thomas S. Fletcher,*,‡ Myriam Alhadeff,‡ Charles S. Craik,§ and Corey Largman‡
Biochemistry Research Laboratory, Veterans Administration Medical Center, Martinez, California 94553, Department of Internal Medicine, University of California, Davis, California 95616, and Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143

Received October 23, 1986; Revised Manuscript Received February 5, 1987

ABSTRACT: A cDNA encoding rat cationic trypsinogen has been isolated by immunoscreening from a rat pancreas cDNA library. The protein encoded by this cDNA is highly basic and contains all of the structural features observed in trypsinogens. The amino acid sequence of rat cationic trypsinogen is 75% and 77% homologous to the two anionic rat trypsinogens. The homology of rat cationic trypsinogen to these anionic trypsinogens is lower than its homology to other mammalian cationic trypsinogens, suggesting that anionic and cationic trypsins probably diverged prior to the divergence of rodents and ungulates. The most unusual feature of this trypsinogen is the presence of an activation peptide containing five aspartic acid residues, in contrast to all other reported trypsinogen activation peptides which contain four acidic amino acid residues. Comparisons of cationic and anionic trypsins reveal that the majority of the charge changes occur in the C-terminal portion of the protein, which forms the substrate binding site. Several regions of conserved charge differences between cationic and anionic trypsins have been identified in this region, which may influence the rate of hydrolysis of protein substrates.

The trypsins (EC 3.4.21.4) are important members of a large family of pancreatic serine proteases that share a common catalytic mechanism and posess similar tertiary structures. These enzymes are endopeptidases which are synthesized as proenzymes by pancreatic acinar cells and secreted into the gut. The trypsins are distinguished from the other pancreatic serine proteases both by their specificity for arginine or lysine residues and by their capability to activate the other pancreatic zymogens.

Although multiple ionic forms of trypsin have been isolated from humans (Travis & Roberts, 1969; Mallory & Travis, 1973), cows (Louvard & Puigserver, 1974), rats (Brodrick et

† This work has been supported by grants to C.L. from the Research Service of the Veterans Administration and from the National Institutes of Health (HL25408).
* Address correspondence to this author at the Biochemistry Research Laboratory, Veterans Administration Medical Center.
‡ Veterans Administration Medical Center and University of California, Davis.
§ University of California, San Francisco.

0006-2960/87/0426-3081$01.50/0

al., 1980), dogs (Ohlsson & Tegner, 1973), and pigs (Louvard & Puigserver, 1974), little is known concerning the possible functional significance of the different trypsins. The classically described forms of bovine and porcine trypsins are cationic proteins (Walsh & Neurath, 1964; Travis & Liener, 1965), and in these animals, the anionic forms of trypsin represent less than 10% of the total trypsin content (Louvard & Puigserver, 1974). In contrast, two anionic trypsins represent approximately 60% of the trypsin synthesized in the rat pancreas (C. Largman, unpublished results). Two rat trypsin genes have been sequenced and shown to code for the anionic trypsinogens (Craik et al., 1984). The gene for rat pancreatic trypsin I was shown to code for the major anionic form produced by the pancreas by homology to the amino acid sequence of the N-terminus of purified rat anionic trypsin (unpublished results). Cloned rat pancreatic trypsinogen II has been expressed in mammalian cells and shown to also represent an anionic form of rat trypsin (Craik et al, 1984). To date, there has not been a sequence comparison between the anionic and cationic forms of trypsin from a single species.

In the current study, we have cloned the cDNA for rat cationic trypsinogen using an expression vector and antibodies directed against rat cationic trypsin. The deduced amino acid sequence for rat cationic trypsin and the amino acid sequences of bovine and porcine cationic trypsins have been compared to those of rat anionic trypsins with regard to possible differences in structure which might result in changes in catalytic activity. In addition, comparison of these five trypsinogen sequences has facilitated an assessment of the sequence divergence of the cationic and anionic trypsinogens.

Materials and Methods

Antibody Screening of a cDNA Expression Library. Rat cationic trypsin was purified as described by Brodrick et al. (1980). Rabbit antiserum to rat cationic trypsin was produced by a series of subcutaneous injections of rat cationic trypsin (100 μg) in Freund's complete adjuvant as previously described (Largman et al., 1981). The IgG fractions were prepared by chromatography on protein A–Sepaharose (Pharmacia, Piscataway, NJ).

A rat pancreatic cDNA library cloned into λgt11 was generously supplied by J. C. Edman (UCSF). This library contains about 20 000 independent clones inserted into the unique *Eco*RI site of λgt11 by the ligation of cDNA tailed with a linker sequence containing the restriction sites for *Eco*RI, *Bam*HI, and *Hin*dIII. The cDNA library was screened with polyclonal rabbit IgG raised against rat cationic trypsin, using ^{125}I-labeled goat anti-rabbit IgG as the second antibody (Young & Davis, 1983). The goat anti-rabbit IgG was radioiodinated by using Iodogen (Pierce, Rockford, IL) as described by Salacinski et al. (1981). Preliminary tests showed that this screening system was capable of detecting 1 ng of rat cationic trypsin.

Primer Extension. The 5′ portion of the mRNA encoding rat cationic trypsin was sequenced by hybridizing a ^{32}P-labeled oligonucleotide to rat pancreatic mRNA and extending in the presence of nucleotides and chain-terminating inhibitors (Craik et al, 1984). The reaction products were separated on a 10% polyacrylamide–urea gel. Rat pancreas RNA was extracted by using the technique of Chirgwin et al. (1979). Two rounds of chromatography of the pancreatic RNA on oligo(dT)–cellulose yielded mRNA which was subsequently used for primer extension. The priming reaction was performed as described in Craik et al. (1984), with the exception that the incubation temperature was 50 °C. The primed mRNA was redissolved in 55 μL of 0.5 mM ethylenediaminetetraacetic acid (EDTA) (pH 8.0) by heating to 90 °C for 2 min. Each of the extension reactions contained the following: 10 μL of primed mRNA, four deoxynucleoside triphosphates, one dideoxynucleoside triphosphate, and 15–20 units of avian myeloblastosis virus reverse transcriptase (Life Sciences, St. Petersburg, FL). The concentrations of the deoxynucleoside triphosphates were 800 μM except for the one to terminated which was 100 μM. The 2′,3′-dideoxynucleoside triphosphate concentration was 200 μM. A control reaction contained the same amount of primed mRNA and reverse transcriptase but only 500 μM deoxynucleoside triphosphates. The extension buffer contained 50 mM tris(hydromethyl)aminomethane hydrochloride (Tris-HCl) (pH 8.0), 50 mM KCl, 8 mM $MgCl_2$, and 4 mM dithiothreitol (DTT). The extension reaction was incubated for 1 h at 42 °C. Each reaction mix was phenol extracted, ethanol precipitated, and taken up in 2 μL of 50% formamide containing 0.05% bromophenol blue and 0.05% xylene cyanol. The tubes were then heated to 90 °C for 3 min and cooled on ice, and 2 μL of each mixture was loaded on a 10% polyacrylamide–urea gel. The gel was run for 12 h at 1500 V, dried, and autoradiographed.

DNA Hybridization and Dideoxy DNA Sequencing. DNA hybridizations were performed as described in Maniatis et al. (1982). Sequencing was performed as described in Messing (1983) using the bacteriophase M13mp11 and the 17-mer sequencing primer of New England Biolabs (no. 1212, New England Biolabs, Beverly, MA). Deoxynucleotides and dideoxynucleotides were from Pharmacia (Piscataway, NJ). The restriction enzymes were purchased from New England Biolabs (Beverly, MA) and BRL (Gaithersburg, MD). Polymerase I (Klenow) was from Boehringer Mannheim (Indianapolis, IN). DNA sequences were analyzed with computer programs supplied by Dr. H. Martinez (Biomathematics Computation Laboratory, Department of Biochemistry and Biophysics, UCSF, San Francisco, CA).

Protein Sequencing. A sample (10 nmol) of rat cationic trypsinogen was subjected to automatic Edman sequence analysis using a Beckman 890C automatic sequencer at the U.C. Davis Protein Structure Laboratory. Amino acid–phenylthiohydantoin(PTH) derivatives were identified by high-pressure liquid chromatography.

Results and Discussion

Cloning and Sequencing of the Cationic Trypsinogen cDNA. The primary screening of the rat pancreas cDNA library with specific antibodies against rat cationic trypsin yielded 19 positive clones. Five of these were successfully purified, and two of these five had easily detectable inserts of 200–300 base pairs (bp). The largest of the two was subcloned into the *Hin*dIII site of M13mp11. Use of the linker with multiple restriction sites permitted subcloning with *Hin*dIII rather than *Eco*RI. Dideoxy sequence analysis showed that this clone extended from position 316 to position 600 (Figure 1). The DNA sequence encoded for a portion of a protein which was highly homologous to pancreatic trypsinogens and which included aspartic acid residues at amino acids 93 and 180. These two amino acid residues are homologous to the Asp-102 of the serine protease charge relay system and the Asp-189 that confers the Arg/Lys substrate specificity on trypsins.

In order to isolate a full-length cDNA, the cDNA fragment was isolated, labeled with ^{32}P by nick translation, and employed to probe the full library. Approximately 5% of the phage in the cDNA library hybridized to this sequence under stringent conditions. An 800 bp clone was isolated, subcloned into M13mp11, and sequenced. This clone contained an open reading frame extending from position 49 (amino acid 2) to

```
   -15                -10                              -1  1                              10
   met lys ala leu ile phe leu ala phe leu gly ala ala val ala leu pro leu asp asp asp asp lys ile val gly gly tyr thr
 1 ATG AAG GCC TTA ATT TTC CTT GCT TTC CTT GGA GCT GCT GTT GCT CTC CCT CTG GAT GAT GAT GAC AAG ATT GTT GGA GGC TAC ACC

                    20                               30                              40
   cys gln lys asn ser leu pro tyr gln val ser leu asn ala gly tyr his phe cys gly gly ser leu ile asn ser gln trp val val
91 TGC CAG AAG AAT TCT CTC CCA TAC CAG GTG TCT CTG AAT GCT GGC TAC CAT TTT TGT GGA GGC TCC CTC ATC AAT TCC CAG TGG GTT GTT

                    50                               60                              70
   ser ala ala his cys tyr lys ser arg ile gln val arg leu gly glu his asn ile asp val val glu gly gly glu gln phe ile asp
181 TCA GCC GCT CAC TGC TAC AAA TCC CGA ATT CAG GTG CGC CTG GGA GAA CAC AAC ATT GAT GTC GTT GAG GGT GGT GAG CAA TTC ATT GAT

                    80                               90                              100
   ala ala lys ile ile arg his pro ser tyr asn ala asn thr phe asp asn asp ile met leu ile lys leu asn ser pro ala thr leu
271 GCA GCT AAA ATC ATC CGC CAC CCC AGT TAT AAT GCA AAC ACC TTT GAC AAT GAT ATT ATG TTG ATT AAG CTG AAT TCA CCT GCC ACC CTC

                    110                              120                             130
   asn ser arg val ser thr val ser leu pro arg ser cys gly ser ser gly thr lys cys leu val ser gly trp gly asn thr leu ser
361 AAT TCT CGA GTG TCC ACT GTC TCT CTG CCA AGA TCT TGT GGA TCA TCT GGT ACT AAG TGC CTT GTG TCT GGC TGG GGC AAC ACC CTG AGC

                    140                              150                             160
   ser gly thr asn tyr pro ser leu leu gln cys leu asp ala pro val leu ser asp ser ser cys lys ser ser tyr pro gly lys ile
451 TCT GGC ACG AAC TAC CCT TCA CTG CTT CAG TGT CTT GAT GCC CCT GTC CTC TCT GAC AGT TCT TGC AAA AGT TCT TAC CCA GGC AAG ATC

                    170                              180                             190
   thr ser asn met phe cys leu gly phe leu glu gly gly lys asp ser cys gln gly asp ser gly gly pro val val cys asn gly gln
541 ACT AGC AAC ATG TTC TGT CTG GGC TTT CTG GAG GGC GGA AAG GAC TCC TGC CAG GGT GAC TCT GGT GGC CCT GTA GTC TGC AAT GGC CAG

                    200                              210                             220
   leu gln gly val val ser trp gly tyr gly cys ala gln lys gly lys pro gly val tyr thr lys val cys asn tyr val asn trp ile
631 CTC CAG GGT GTT GTT TCC TGG GGT TAT GGC TGT GCT CAG AAA GGA AAA CCT GGT GTC TAT ACC AAG GTA TGC AAC TAC GTG AAC TGG ATT

                    230  232
   gln gln thr val ala ala asn STOP
721 CAG CAG ACC GTC GCT GCC AAC TAA AACGCCTTTATGTCTTTATACTTTGTAGTCAAATCCACCTTCTTTGTATGCCCCAAGTCATTCTCACAAGAAAAACA
```

FIGURE 1: Nucleotide sequence and the deduced amino acid sequence of rat cationic trypsinogen. The underlined amino acids correspond to His-57, Asp-102, Asp-189, Ser-195, Gly-216, and Gly-226 (chymotrypsin numbering system). The predicted prepeptide consists of amino acid residues −15 to −1. Residue 1 is the first residue of the zymogen.

position 741 (Figure 1). The deduced sequence of the N-terminal 19 amino acids was identical with that found for purified rat cationic trypsinogen (amino acids 2–20). It continued for 80 bp beyond the 3′ end of the coding region but did not include a polyadenylation site. Although this clone contained the coding information for the active enzyme and activation peptide, the 5′ terminus was incomplete and lacked the signal sequence found in other pancreatic secretory proteins (Watson, 1984).

In an attempt to isolate the 5′ terminus, six independent clones of rat cationic trypsinogen were isolated by using a 73 bp 5′ fragment from the "full-length" clone. These clones were subcloned into M13mp11 using *Eco*RI and sequenced. Two of these clones extended to position 5, lacking only the methionine codon and the first base of the lysine codon.

The final portion of the sequence was determined by primer extension. A 30-mer primer extending from position 46 to 75 was synthesized. This particular sequence was selected because it was the most diffferent from the anionic mRNAs in this highly conserved region, having four mismatches between the anionic trypsinogens and this primer, one of which was at the 3′ end of the primer, as well as a deletion of 3 bp. This should have ensured that the mRNA for cationic trypsinogen was the only sequence primed (Reyes & Wallace, 1984). However, although there was a major pattern corresponding to cationic trypsin, minor patterns coding for the two anionic trypsins could be detected. Fortunately, the portion of the sequence that was unknown for cationic trypsin is shared by all three sequences and is thus unambiguous (results not shown).

Sequence Homology to Other Trypsin-like Proteases. The sequence of rat cationic trypsinogen contains all of the major structural features common to trypsinogens (Figure 2). In addition to the obligatory Asp-189 and the catalytic triad of His-57, Asp-102, and Ser-195 (chymotrypsin numbering system; Hartley, 1970), rat cationic trypsinogen contains the following conserved features: an activation peptide containing a poly(aspartic acid) sequence as described below; and the six conserved disulfide bonds found in other rat trypsinogens. In addition, residues 199 and 200 of rat cationic trypsinogen are valines, as are the homologous residues in cow, pig, sheep, turkey, spiny dogfish, and rat anionic trypsins. This feature is unlike trypsin-like enzymes including rat pancreatic kallikrein (Leu-Val) (Swift et al., 1982), human kidney kallikrein (Leu-Met) (Baker & Shine, 1985), adipsin (Leu-Val) (Cook et al., 1985), and hanakah factor (Leu-Leu) (Gershenfield & Weissman, 1986).

The N-terminal sequence is similar to those of other trypsinogens (Figure 3). All trypsinogens have a polyanionic cluster immediately preceding the activation cleavage position (Lys-15/Ile-16), with the possible exception of human cationic trypsinogen which is reported by Brodrick et al. (1978) to contain an Asp-Lys activation peptide. Rat cationic trypsinogen is unique in having five negatively charged residues in a row compared to the four acidic residues reported for most other trypsins (Stroud et al., 197[illegible]). The initial few residues which immediately precede the polyanionic sequence in the activation peptides are notable in that they seem to generally consist of either a single valine residue (cow and turkey) or three residues of which one is proline [pig, sheep, all rat trypsinogens, and the human trypsinogens reported by Guy et al. (1978)].

The signal peptide sequence follows the usual pattern of a charged residue in the first few residues followed by regions of hydrophobic residues and small, uncharged amino acids. It is of interest that the signal peptides of both anionic trypsinogen II and cationic trypsinogen have a basic amino acid immediately after the methionine whereas the signal peptide of anionic trypsinogen I has a serine at that position (Figure 2). Although the presence of a charged amino acid is more typical (Watson, 1984), anionic trypsinogen I is the most abundant of the three proteins.

Rat cationic trypsinogen shows a high degree of overall homology with the two rat anionic trypsinogens. At the DNA level, rat cationic trypsinogen is 76% and 75% homologous to

																8	9	10	11	11A	12	13	14	15	16	17	18	19	20	21
RCT	met	lys	ala	leu	ile	phe	leu	ala	phe	leu	gly	ala	ala	val	ala	leu	pro	leu	asp	asp	asp	asp	asp	lys	ile	val	gly	gly	tyr	thr
BCT	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	val	asp	---	asp	asp	asp	lys	ile	val	gly	gly	tyr	thr
PCT	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	phe	pro	tyr	asp	---	asp	asp	asp	lys	ile	val	gly	gly	tyr	thr
AN1	met	ser	ala	leu	leu	ile	leu	ala	leu	val	gly	ala	ala	val	ala	phe	pro	leu	glu	---	asp	asp	asp	lys	ile	val	gly	gly	tyr	thr
AN2	met	arg	ala	leu	leu	phe	leu	ala	leu	val	gly	ala	ala	val	ala	phe	pro	val	asp	---	asp	asp	asp	lys	ile	val	gly	gly	tyr	thr
		•			*	*			*	*						*	*	*	*	*										

	22	23	24	25	26	27	28	29	30	31	32	33	34	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53
RCT	cys	gln	lys	asn	ser	leu	pro	tyr	gln	val	ser	leu	asn	ala	gly	tyr	his	phe	cys	gly	gly	ser	leu	ile	asn	ser	gln	trp	val	val
BCT	cys	gly	ala	asn	thr	val	pro	tyr	gln	val	ser	leu	asn	ser	gly	tyr	his	phe	cys	gly	gly	ser	leu	ile	asn	ser	gln	trp	val	val
PCT	cys	ala	ala	asn	ser	I/V	pro	tyr	gln	val	ser	leu	asn	ser	gly	ser	his	phe	cys	gly	gly	ser	leu	ile	asn	ser	gln	trp	val	val
AN1	cys	pro	glu	his	ser	val	pro	tyr	gln	val	ser	leu	asn	ser	gly	tyr	his	phe	cys	gly	gly	ser	leu	ile	asn	asp	gln	trp	val	val
AN2	cys	gln	glu	asn	ser	val	pro	tyr	gln	val	ser	leu	asn	ser	gly	tyr	his	phe	cys	gly	gly	ser	leu	ile	asn	asp	gln	trp	val	val
		*	*	*	*	*								*		•										***				

	54	55	56	57	58	59	60	61	62	63	64	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84
RCT	ser	ala	ala	his	cys	tyr	lys	ser	arg	ile	gln	val	arg	leu	gly	glu	his	asn	ile	asp	val	val	glu	gly	gly	glu	gln	phe	ile	asp
BCT	ser	ala	ala	his	cys	tyr	lys	ser	gly	ile	gln	val	arg	leu	gly	glu	asp	asn	ile	asn	val	val	glu	gly	asp	glu	gln	phe	ile	ser
PCT	ser	ala	ala	his	cys	tyr	lys	ser	arg	ile	gln	val	arg	leu	gly	glu	his	asn	ile	asp	val	leu	glu	gly	asn	glu	gln	phe	ile	asn
AN1	ser	ala	ala	his	cys	tyr	lys	ser	arg	ile	gln	val	arg	leu	gly	glu	his	asn	ile	asn	val	leu	glu	gly	asp	glu	gln	phe	ile	asn
AN2	ser	ala	ala	his	cys	tyr	lys	ser	arg	ile	gln	val	arg	leu	gly	glu	his	asn	ile	asn	val	leu	glu	gly	asn	glu	gln	phe	val	asn
									*								*			•		*			*				*	*

	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114
RCT	ala	ala	lys	ile	ile	arg	his	pro	ser	tyr	asn	ala	asn	thr	phe	asp	asn	asp	ile	met	leu	ile	lys	leu	asn	ser	pro	ala	thr	leu
BCT	ala	ser	lys	ser	ile	val	his	pro	ser	tyr	asn	ser	asn	thr	leu	asn	asn	asp	ile	met	leu	ile	lys	leu	lys	ser	ala	ala	ser	leu
PCT	ala	ala	lys	ile	ile	thr	his	pro	asn	phe	asn	gly	asn	thr	leu	asp	asn	asp	ile	met	leu	ile	lys	leu	ser	ser	pro	ala	thr	leu
AN1	ala	ala	lys	ile	ile	lys	his	pro	asn	tyr	ser	ser	trp	thr	leu	asn	asn	asp	ile	met	leu	ile	lys	leu	ser	ser	pro	val	lys	leu
AN2	ala	ala	lys	ile	ile	lys	his	pro	asn	phe	asp	arg	arg	thr	leu	asn	asn	asp	ile	met	leu	ile	lys	leu	ser	ser	pro	val	lys	leu
		•		•		•			*	*	*	*	*		*	*									*		*	***	*	

	115	116	117	118	119	120	121	122	123	124	125	127	128	129	130	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146
RCT	asn	ser	arg	val	ser	thr	val	ser	leu	pro	arg	ser	cys	gly	ser	ser	gly	thr	lys	cys	leu	val	ser	gly	trp	gly	asn	thr	leu	ser
BCT	asn	ser	arg	val	ala	ser	ile	ser	leu	pro	thr	ser	cys	ala	ser	ala	gly	thr	gln	cys	leu	ile	ser	gly	trp	gly	asn	thr	lys	ser
PCT	asn	ser	arg	val	ala	thr	val	ser	leu	pro	arg	ser	cys	ala	ala	ala	gly	thr	glu	cys	leu	ile	ser	gly	trp	gly	asn	thr	lys	ser
AN1	asn	ala	arg	val	ala	pro	val	ala	leu	pro	ser	ala	cys	ala	pro	ala	gly	thr	gln	cys	leu	ile	ser	gly	trp	gly	asn	thr	leu	ser
AN2	asn	ala	arg	val	ala	thr	val	ala	leu	pro	ser	ser	cys	ala	pro	ala	gly	thr	gln	cys	leu	ile	ser	gly	trp	gly	asn	thr	leu	ser
		***			*	*	*	***			*	*		*	*	*			•			•							*	

	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176
RCT	ser	gly	thr	asn	tyr	pro	ser	leu	leu	gln	cys	leu	asp	ala	pro	val	leu	ser	asp	ser	ser	cys	lys	ser	ser	tyr	pro	gly	lys	ile
BCT	ser	gly	thr	ser	tyr	pro	asp	val	leu	lys	cys	leu	lys	ala	pro	ile	leu	ser	asp	ser	ser	cys	lys	ser	ala	tyr	pro	gly	gln	ile
PCT	ser	gly	ser	ser	tyr	pro	ser	leu	leu	gln	cys	leu	lys	ala	pro	val	leu	ser	asp	ser	ser	cys	lys	ser	ser	tyr	pro	gly	gln	ile
AN1	asn	gly	val	asn	asn	pro	asp	leu	leu	gln	cys	val	asp	ala	pro	val	leu	ser	gln	ala	asp	cys	glu	ala	ala	tyr	pro	gly	glu	ile
AN2	ser	gly	val	asn	glu	pro	asp	leu	leu	gln	cys	leu	asp	ala	pro	leu	leu	pro	gln	ala	asp	cys	glu	ala	ser	tyr	pro	gly	lys	ile
	*		*	*	*		*	*		•		*	*			•		•	***	***	***		***	***	•				•	

	177	178	179	180	181	182	183	184	184A	185	186	187	188	188A	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204
RCT	thr	ser	asn	met	phe	cys	leu	gly	phe	leu	glu	gly	gly	lys	asp	ser	cys	gln	gly	asp	ser	gly	gly	pro	val	val	cys	asn	gly	gln
BCT	thr	ser	asn	met	phe	cys	ala	gly	tyr	leu	glu	gly	gly	lys	asp	ser	cys	gln	gly	asp	ser	gly	gly	pro	val	val	cys	ser	gly	lys
PCT	thr	gly	asn	met	ile	cys	val	gly	phe	leu	glu	gly	gly	lys	asp	ser	cys	gln	gly	asp	ser	gly	gly	pro	val	val	cys	asn	gly	gln
AN1	thr	ser	ser	met	ile	cys	val	gly	phe	leu	glu	gly	gly	lys	asp	ser	cys	gln	gly	asp	ser	gly	gly	pro	val	val	cys	asn	gly	gln
AN2	thr	asp	asn	met	val	cys	val	gly	phe	leu	glu	gly	gly	lys	asp	ser	cys	gln	gly	asp	ser	gly	gly	pro	val	val	cys	asn	gly	glu
		*	*		•		*		*																			•		•

	209	210	211	212	213	214	215	216	217	219	220	221	221A	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238
RCT	leu	gln	gly	val	val	ser	trp	gly	tyr	gly	cys	ala	gln	lys	gly	lys	pro	gly	val	tyr	thr	lys	val	cys	asn	tyr	val	asn	trp	ile
BCT	leu	gln	gly	ile	val	ser	trp	gly	ser	gly	cys	ala	gln	lys	asn	lys	pro	gly	val	tyr	thr	lys	val	cys	asn	tyr	val	ser	trp	ile
PCT	leu	gln	gly	ile	val	ser	trp	gly	tyr	gly	cys	ala	gln	lys	asn	lys	pro	gly	val	tyr	thr	lys	val	cys	asn	tyr	val	asn	trp	ile
AN1	leu	gln	gly	ile	val	ser	trp	gly	tyr	gly	cys	ala	leu	pro	asp	asn	pro	gly	val	tyr	thr	lys	val	cys	asn	phe	val	gly	trp	ile
AN2	leu	gln	gly	ile	val	ser	trp	gly	tyr	gly	cys	ala	leu	pro	asp	asn	pro	gly	val	tyr	thr	lys	val	cys	asn	tyr	val	asp	trp	ile
				*					*				***	***	*	***										*		*		

	239	240	241	242	243	244	245
RCT	gln	gln	thr	val	ala	ala	asn
BCT	lys	gln	thr	ile	ala	ser	asn
PCT	gln	gln	thr	ile	ala	ala	asn
AN1	gln	asp	thr	ile	ala	ala	asn
AN2	gln	asp	thr	ile	ala	ala	asn
	*	***		*		*	

FIGURE 2: Amino acid sequences of five trypsinogens. The amino acid residues are numbered by using the standard chymotrypsin numbering system. Positions underscored with a single asterisk are not identical in all sequences. Those underscored with three asterisks are different in cationic and anionic trypsinogens but identical within each group. Positions underlined are those thought to interact with one of the following trypsin inhibitors. The specific interactions of each of the inhibitors are as follows: pancreatic trypsin inhibitor with bovine cationic trypsin residues 39, 40, 41, 97, 189, 190, 193, 195, 214, 215, 216, and 219 (Janin & Chothia, 1976); soybean trypsin inhibitor with porcine cationic trypsin residues 40, 57, 60, 189, 190, 192, 193, 195, 214, 215, 216, and 217 (Janin & Chothia, 1976); Japanese Quail ovomucoid third domain protease inhibitor and bovine cationic trypsin residues 41, 96, 143, 149, 151, 189, 190, 193, 195, 214, and 217 (Papamokos et al., 1982); and pancreatic secretory trypsin inhibitor and bovine cationic trypsin residues 39, 41, 60, 97, 175, 189, 190, 193, 195, 214, 215, 216, 219, and 224 (Bolognesi et al., 1982). Abbreviations: RCT, rat cationic trypsinogen; BCT, bovine cationic trypsinogen; PCT, porcine cationic trypsinogen; AN1, rat anionic trypsinogen I; AN2, rat anionic trypsinogen II. The sequence references are RCT (this work), BCT (Walsh, 1970; Hartley, 1970), PCT (Hermodson et al., 1973), and AN1 and AN2 (Craik et al., 1984).

rat anionic trypsinogen I and rat anionic trypsinogen II, respectively, while the corresponding amino acid sequence is 75% homologous to anionic trypsinogen I and 77% homologous to anionic trypsinogen II. One interesting observation is that the amino acid sequence of rat cationic trypsinogen is more homologous to bovine cationic trypsinogen (79%) and porcine cationic trypsinogen (86%) than it is to the rat anionic trypsinogens. This would imply that the divergence of a primordial trypsin gene into anionic and cationic forms probably occurred before the divergence of rodents and ungulates.

Charge Distribution. The most obvious difference between the rat trypsins is their difference in overall net charge. The

	8	9	10	11	11A	12	13	14	15
RCT	Leu	Pro	Leu	Asp	Asp	Asp	Asp	Asp	Lys
AN1	Phe	Pro	Leu	Glu	—	Asp	Asp	Asp	Lys
AN2	Phe	Pro	Val	Asp	—	Asp	Asp	Asp	Lys
PCT	Phe	Pro	Tyr	Asp	—	Asp	Asp	Asp	Lys
ST	Phe	Pro	Tyr	Asp	—	Asp	Asp	Asp	Lys
HT1	Ala	Pro	Phe	Asp	—	Asp	Asp	Asp	Lys
HT2	Ala	Pro	Phe	Asp	—	Asp	Asp	Asp	Lys
TT			Val	Asp	—	Asp	Asp	Asp	Lys
BCT			Val	Asp	—	Asp	Asp	Asp	Lys
HCT								Asp	Lys

FIGURE 3: Activation peptides for a number of trypsinogens. Abbreviations: RCT, rat cationic trypsinogen; AN1, rat anionic trypsinogen I; AN2, rat anionic trypsinogen II; PCT, porcine cationic trypsin; ST, sheep trypsinogen; HT1, human trypsinogen I; HT2, human trypsinogen II; TT, turkey trypsinogen; BCT, bovine cationic trypsinogen; HCT, human cationic trypsinogen. Sequences are from the following: AN1 and AN2, Craik et al. (1984); PCT, Hermodson et al. (1973); ST, Dayhoff (1969); HT1 and HT2, Guy et al. (1978); TT, Kishida & Liener (1968); BCT, Hartley (1970); and HCT, Brodrick et al. (1978).

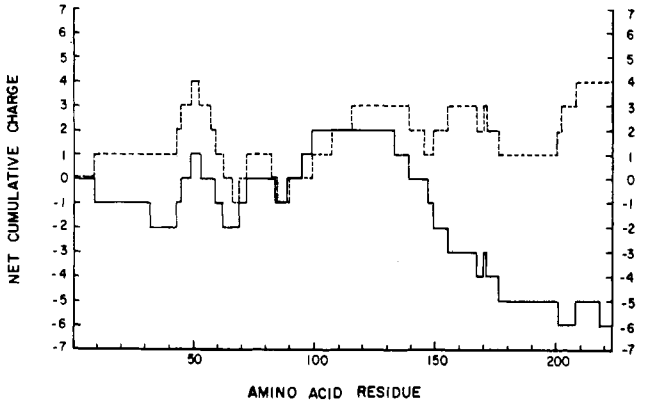

FIGURE 4: Net cumulative charge of rat cationic trypsin (dotted line) and rat anionic trypsin I (solid line). The net cumulative charge is summed, starting at Ile-16, over the whole active trypsin sequence. For the sake of clarity, the other trypsins were not included in this figure. These two profiles are representative of the cationic and anionic trypsins, respectively.

cumulative net charge is shown for rat anionic trypsin I and rat cationic trypsin in Figure 4. These two patterns are representative of those seen for other anionic and cationic molecules. Rat cationic trypsin has a net charge of +4 whereas both anionic trypsins have a net charge of −6. This is as would be expected on the basis of their isoelectric points (anionic = 5.3, cationic = 8.5; C. Largman et al., unpublished results). One intriguing point is that the cumulative net charges of the three rat trypsins, bovine cationic trypsin, and porcine cationic trypsin are all within one charge of each other in the region from amino acids 16 to 165 and most of the charge difference is located in the C-terminal 78 residues of the trypsin molecule (Figure 4). It should be noted that the substrate binding site is formed by the residues in this C-terminal part of trypsin.

If one assumes that differences in charge are functionally important, then they should be conserved among cationic trypsins and occur in regions of the molecule that are involed in either binding specificity or catalysis. When one compares the charge differences between the rat anionic trypsins and rat, porcine, and bovine cationic trypsin, there are several substitutions that are shared in all cationic trypsins. Two amino acid clusters, residues 167–169 and 222–224 (Figure 2), which are adjacent to each other in the tertiary structure of active trypsins show consistent charge changes between anionic and cationic trypsins. The net charge difference between cationic and anionic trypsins is 3+ at residues 167–169 and 3+ at residues 222–224, resulting in an overall net change of 6 charges. In addition, at residues 166, 170, and 221A, which flank these two regions, there is a pattern of substitution of nonpolar residues in anionic trypsins for polar residues in cationic trypsins. While this region is not known to be critical to the activity of the enzyme, it should be noted that amino acids 222–224 are adjacent to the binding pocket. In addition to the changes noted above, there are consistent charge changes at residues 49 (Asp to Ser) and at position 113 (Lys to Thr) that result in net charge changes of 1+ and 1−, respectively. However, both of these changes are distant from the active site and do not appear to be of obvious functional significance.

Two cDNAs which apparently code for human trypsinogens have recently been sequenced (Emi et al., 1986). These clones, which are 89% homologous, have not been conclusively identified as coding for either anionic or cationic trypsinogens, and we have chosen to not include them in the discussion of the sequence differences between anionic and cationic trypsins. However, it is of interest that these two sequences do not follow the pattern of the other trypsins, in that both proteins resemble the anionic trypsins at residues 165–170 but the cationic trypsins at residues 221A–224.

Residues Important for Substrate Binding. Although other members of the serine protease family have an extended substrate binding region which spans up to seven amino acids (Nakajima et al., 1979; Thompson & Blout, 1973; Harper et al., 1984; Baumann et al., 1973), the specificity of trypsins is usually thought to be dominated by the selectivity of the P_1 position with little influence of secondary binding sites. However, small variations in hydrolysis rates of synthetic peptides varying in the P_2 and P_3 positions have been observed (Tanaka et al., 1983), implying the presence of an extended binding site for trypsin. Furthermore, X-ray crystallographic studies of trypsin–protein inhibitor complexes reveal numerous interactions which have been interpreted as an extended binding region (Bolognesi et al., 1982). These studies have led to identification of several residues which appear to be involved in the interaction of trypsin with Japanese Quail ovomucoid protease inhibitor (Papamokos et al., 1982), pancreatic secretory trypsin inhibitor (Bolognesi et al., 1982), pancreatic trypsin inhibitor (Janin & Chothia, 1976), or soybean trypsin inhibitor (Janin & Chothia, 1976). The amino acid residues of trypsin that appear to be in close contact with one or more of these inhibitors are underlined in Figure 2. Of these 23 residues, 9 are polymorphic in the trypsins listed in Figure 2. The sequence of one region (residues 96–99) is quite variable among the trypsins and appears to be partially conserved within the cationic trypsins. This suggests that the interaction of cationic and anionic trypsins with these inhibitors may be different due to charge and steric bulk changes in this region. In addition to this particular difference, there are changes at residues 39, 149, 151, 175, and 217 that could also affect substrate binding. If these interactions are analogous to the interactions of trypsin with its protein substrates, then it would be reasonable to hypothesize that changes at these residues may affect substrate specificity.

Another amino acid which may be important for substrate specificity is Phe-99. In kallikreins, a tyrosine is usually present at this position and is thought to form a hydrophobic sandwich with Trp-215, resulting in an increased affinity for peptides such as Phe-Arg-X (Swift et al., 1982). The Phe-99 in rat cationic trypsin is also capable of forming a sandwich structure with Trp-215 and the substrate aromatic ring. Anionic rat trypsins, as well as bovine and porcine cationic trypsin, contain a leucine at this position. This difference suggests that rat

cationic trypsin should exhibit a higher affinity for Phe-Arg substrates.

Comparisons of the selectivity of anionic or cationic trypsins toward peptide or protein substrates have not been reported, but our results suggest the hypothesis that some of the amino acid differences between anionic and cationic trypsins, particularly residues 165–170, 221A–224, and 96–99, might result in preferential hydrolysis of protein substrates as the result of secondary binding site interactions. Such a hypothesis could be tested by constructing hybrid trypsins containing portions of the anionic and cationic enzymes. This could be most easily done with the rat trypsins, since the respective cDNAs share a considerable number of homologous restriction sites.

ACKNOWLEDGMENTS

We thank Jeff Edman and William J. Rutter for kindly supplying the rat pancreas cDNA library, Steve Gardell and Yosuke Ebina for their helpful advice, and Al Smith for performing the protein sequencing.

REFERENCES

Baker, A. P., & Shine, J. (1985) *DNA 4*, 445–450.

Baumann, W. K., Bizzozero, S. A., & Dutler, H. (1973) *Eur. J. Biochem. 39*, 381–391.

Bolognesi, M., Gatti, G., Menegatti, E., Guarneri, M., Marquart, M., Papamokos, E., & Huber, R. (1982) *J. Mol. Biol. 162*, 839–868.

Brodrick, J. W., Largman, C., Hsiang, M. W., Johnson, J. H., & Geokas, M. C. (1978) *J. Biol. Chem. 253*, 2737–2742.

Brodrick, J. W., Largman, C., Geokas, M. C., O'Rourke, M., & Ray, S. B. (1980) *Am. J. Physiol. 239*, G511–G515.

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

Cook, K. S., Groves, D. L., Min, H. Y., & Spiegelman, B. M. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 6480–6484.

Craik, C. S., Choo, Q. L., Swift, G. H., Quinto, C., MacDonald, R. J., & Rutter, W. J. (1984) *J. Biol. Chem. 259*, 14255–14264.

Dayhoff, M. O. (1969) *Atlas of Protein Sequence and Structure 4*, D224.

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

Gershenfeld, H. K., & Weissman, I. L. (1986) *Science (Washington, D.C.) 232*, 854–861.

Guy, O., Lombardo, D., Bartlelt, D. C., Amie, J., & Figarella, C. (1978) *Biochemistry 17*, 1669–1675.

Harper, J. W., Cook, R. R., Roberts, C. J., McLaughlin, B. J., & Powers, J. C. (1984) *Biochemistry 23*, 2995–3002.

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

Hermodson, M. A., Ericsson, L. H., Neurath, H., & Walsh, K. A. (1973) *Biochemistry 12*, 3146–3153.

Janin, J., & Chothia, C. (1976) *J. Mol. Biol. 100*, 197–211.

Kishida, T., & Liener, I. E. (1968) *Arch. Biochem. Biophys. 126*, 111–120.

Largman, C., Brodrick, J. W., & Geokas, M. C. (1981) *Methods Enzymol. 74*, 272–290.

Louvard, M. N., & Puigserver, A. (1974) *Biochem. Biophys. Acta 371*, 177–185.

Mallory, P. A., & Travis, J. (1973) *Biochemistry 12*, 2847–2851.

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

Messing, J. (1983) *Methods Enzymol. 101*, 20–79.

Nakajima, K., Powers, J. C., Ashe, B. M., & Zimmerman, M. (1979) *J. Biol. Chem. 254*, 4027–4032.

Ohlsson, K., & Tegner, H. (1973) *Biochim. Biophys. Acta 317*, 328–337.

Papamokos, E., Weber, E., Bode, W., Huber, R., Empie, M. W., Kato, I., & Laskowski, M., Jr. (1982) *J. Mol. Biol. 158*, 515–537.

Reyes, A. A., & Wallace, R. B. (1984) *Genet. Eng. 6*, 157–173.

Salacinski, P. R. P., McLean, C., Sykes, J. E. C., Clement-Jones, V. V., & Lowry, P. J. (1981) *Anal. Biochem. 117*, 136–146.

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

Swift, G. H., Dagorn, J. C., Ashley, P. L., Cummings, S. W., & MacDonald, R. J. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 7263–7267.

Tanaka, T., McRae, B. J., Cho, K., Cook, R., Fraki, J. E., Johnson, D. A., & Powers, J. C. (1983) *J. Biol. Chem. 258*, 13522–13557.

Thompson, R. C., & Blout, E. R. (1973) *Biochemistry 12*, 57–65.

Travis, J. & Liener, I. E. (1965) *J. Biol. Chem. 240*, 1967–1973.

Travis, J., & Roberts, R. C. (1969) *Biochemistry 8*, 2884–2889.

Walsh, K. A. (1970) *Methods Enzymol. 19*, 41–63.

Walsh, K. A., & Neurath, H. (1964) *Proc. Natl. Acad. Sci. U.S.A. 52*, 884–889.

Watson, M. E. E. (1984) *Nucleic Acids Res. 13*, 5145–5164.

Young, R. A., & Davis, R. W. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 1194–1198.