Isolation and characterization of the chicken trypsinogen gene family

Kai WANG,* Lu GAN, Inyoul LEE and Leroy HOOD

Department of Molecular Biotechnology GJ-10, University of Washington, Seattle, WA 98195, U.S.A.

Based on genomic Southern hybridizations and cDNA sequence analyses, the chicken trypsinogen gene family can be divided into two multi-member subfamilies, a six-member trypsinogen I subfamily which encodes the cationic trypsin isoenzymes and a three-member trypsinogen II subfamily which encodes the anionic trypsin isoenzymes. The chicken cDNA and genomic clones containing these two subfamilies were isolated and characterized by DNA sequence analysis. The results indicated that the chicken trypsinogen genes encoded a signal peptide of 15 to 16 amino acid residues, an activation peptide of 9 to 10 residues and a

INTRODUCTION

Trypsin (EC 3.4.2.1.4) is a secretory enzyme belonging to the large serine protease family. The enzyme is synthesized by the pancreatic acinar cells under the form of a zymogen, the trypsinogen. The zymogen is rapidly converted into the active form by releasing an N-terminal, six to eight amino acid residue activation peptide through the action of enterokinases. Trypsin catalyses the hydrolysis of peptide bonds on the carboxyl side of lysine (Lys) and arginine (Arg) residues of proteins in the intestinal lumen. The catalytic activity of trypsin, like all the other serine proteases, is mediated by the proper conformation of histidine (His), aspartate (Asp) and serine (Ser) residues at its substrate-binding site (reviewed in [1]).

The trypsinogen genes or proteins have been isolated and characterized from various species including human [2-4], bovine [5], dog [6], pig [7,8], rat [9–12], mouse [13], *Xenopus* [14], dogfish [15] and Atlantic cod [16]. Based on the isoelectric point of trypsins, two major forms of isoenzymes, cationic and anionic, have been identified from most of the species that have been studied. Genomic hybridization and cDNA sequence analyses, indicated that the trypsin isoenzymes are encoded by different genes in the genome rather than arising as a consequence of different post-translational modification processes.

Like the other pancreatic serine proteases, the trypsinogens are tightly regulated by a set of pancreatic-specific transcription factors. Using *in vitro* transfection with deletion clones, sitedirected mutagenesis and DNA footprinting on rat trypsinogen and chymotrypsinogen gene 5' regulatory regions, the sequence for a pancreatic tissue-specific enhancer has been identified and located approximately 100 to 200 bp upstream from the transcription start site [17–19]. The core regulatory sequence (CACC-TGT…TTTCCC) is pyrimidine rich and is required for the proper expression of pancreatic-specific serine proteases.

Recently the human and mouse trypsinogen genes have been located within the T-cell receptor (TCR) β locus by a large-scale sequencing project (L. Rowen, K. Wang, I. Lee and L. Hood, unpublished work). The location of the trypsinogen genes suggested that in addition to its well-studied digestive role in the

trypsin of 223 amino acid residues. The chicken trypsinogens contain all the common catalytic and structural features for trypsins, including the catalytic triad His, Asp and Ser and the six disulphide bonds. The trypsinogen I and II subfamilies share approximately 70% sequence identity at the nucleotide and amino acid level. The sequence comparison among chicken trypsinogen subfamily members and trypsin sequences from other species suggested that the chicken trypsinogen genes may have evolved in coincidental or concerted fashion.

intestine, the trypsins may be involved in immunologically related activities.

Through the natural selection process, peptide sequences that are conserved throughout evolution may be needed for the preservation of protein structure and functions. Therefore, comparing sequences from diverged lines may shed light on the possible roles of these proteins. Birds diverged from mammals approximately 300 to 350 million years ago. To study the evolution of trypsinogen gene sequences and their genomic organization and to shed light on their possible functional roles, the chicken trypsinogen genes have been isolated and studied in detail. In this report, we describe the isolation and characterization of two major types of trypsinogen genes from chicken.

MATERIALS AND METHODS

DNA and Southern blots

Chicken liver DNA was prepared using a standard protocol [20]. Digestions of genomic DNA with appropriate restriction endonucleases were carried out according to manufacturer's specifications. The digested DNAs were resolved on a 0.8% agarose gel (SeaKem, FMC, Rockland, ME, U.S.A.) in TAE buffer (40 mM Tris acetate and 2 mM EDTA) and transferred to nylon membranes (Zeta-probe, Bio-Rad Laboratories, Richmond, CA, U.S.A.) in 0.4 M NaOH [21].

cDNA and genomic library

A chicken cosmid library with 10 genome-equivalents was constructed from chicken liver genomic DNA in the pWE15A cosmid vector as described previously [22].

The total chicken pancreatic RNA was prepared from a freshly killed 3-month-old chicken using the guanidinium isothiocyanate extraction method [23]. The poly(A) mRNA was selected through an oligo(dT)-cellulose column [20] and cDNAs were synthesized and cloned into λ -ZAP directional cloning vector (Stratagene, La Jolla, CA, U.S.A.). A total of 1.5×10^6 independent recombinant phage plaques were obtained and screened with various chicken trypsinogen probes.

^{*}To whom correspondence should be addressed.

The nucleotide sequence data reported in this paper have been submitted to the GenBank nucleotide sequence database and have been assigned the accession numbers U15155-15157.

	Chicken Trypsinogen genes													
Exon 1 Exon 2 Exon 3 Exon 4 Exon 5														
	TRYEX15	TRYEX25	TRYEX35	TRYEX45	TRYEX55									
TRYL TRY	EX1	TRYEX2	TRYEX3	TRYEX4	TRYEX5									
For chic	eken trypsi:	nogen I fami	ly Fo	r chicken try	psinogen II family									
TRYL	асадатаатста	ACAGTTAAT	TRY	L ACTITCAC	CGGGCCTTGTACC									
TRYEX1	GAAATTCTTAGT	GTTAGTCGCC	TRY	EX1 TGGAAGCA	CCATGAAGTTTCT									
TRYEX15	CCAAGAAAGGCG	астаасаст	TRY	EX15 GAATCAGG	AAGAGAAACTTCAT									
TRYEX2	AGGATGATGACA	AGATTGTGG	TRY	EX2 CTGAACTC	TGGGTATCACTTC									
TRYEX25	GCAGCTGACAGC	ACCCATT	TRY	EX25 CAGTGAG	CAGCTGACAAGA									
TRYEX3	GCCCAAGACGGC	AGCGAG	TRY	EX3 GGTGGAAT	ACAGTGCCGACA									
TRYEX35	CAGCTGGTAGGC	AGAGGAA	TRY	EX35 TTGAATGI	CGGCACTGTATTC									
TRYEX4	CCAGATGTCCTG	CAGTGCC	TRY	EX4 CTGAACGO	TCCGATTCTGAG									
TRYEX45	CCTGGCAGGAGT	CTTTTCC	TRY	EX45 GCTGGCAG	TCTTGGTCACT									
TRYEX5	GGGATTCCGGTG	GTCCAG	TRY	EX5 GACTCGGG	TGGACCAGTT									
TRYEX55	AGTGTAAACACC	AGGATAACCT	TRY	EX55 AAGGGCAG	GAGGGAGC									

Figure 1 Diagrammatic summary of oligonucleotide primers used in chicken trypsinogen sequence analyses

The relative positions of the primer locations for each exon are showed in the top portion of the Figure. The arrows indicate the orientation of primers.

Northern blot analysis

Total chicken pancreatic RNA (15 μ g) was denatured in gel loading buffer containing 50% formamide for 3 min at 95 °C. The RNA sample was resolved in a 1.2% agarose/formaldehyde gel. The RNA was then transferred to a nitrocellulose filter.

Oligonucleotides

Four different oligonucleotide primers were synthesized from the conserved regions of all known trypsinogen sequences (Gen-Bank/EMBL release 74) (K. Wang, L. Gan and L. Hood, unpublished work). The trypsinogen-specific conserved region primers were as follows: TRYA 5'-TCCGGATCCTGATG-ACAAGATCGTTGGGGGG; TRYB 5'-TCCGGATCCTTCT-GTGGAGGCTCCCTCAT; TRYC 5'-TCCGGATCCATAG-CCCCAGGAGAC; and TRYD 5'-TCCGGATCCTTGGTG-TAGACACCAGG. All the primers contained a *Bam* HI linker (TCCGGATCC) at their 5' end.

Polymerase chain reactions

To isolate and identify the chicken trypsinogen genes, approximately 0.5 ng of chicken pancreatic cDNA were used in 50 μ l PCR reactions [24] that consisted of 50 mM NaCl, 20 mM Tris/HCl (pH 9.0), 2.5 mM MgCl₂, 0.5 mM dithiothreitol (DTT), 150 μ M of each nucleotide, 1 μ M of each trypsinogen-specific conserved region primer and 2 units of *Taq* DNA polymerase (Cetus, Norwalk, CT, U.S.A.). The reaction mixtures were subjected to 25 cycles of amplification. Each cycle consists of 20 s at 94 °C, 45 s at 50 °C and 90 s at 72 °C. The results of PCR reactions were analysed by gel electrophoresis and the

identity of amplified products was determined by direct sequencing with $[\gamma^{-33}P]$ ATP-labelled amplification primers.

Probes and hybridization

Trypsinogen cDNA and genomic clones were obtained by screening the pancreatic cDNA library and genomic cosmid library with trypsinogen gene PCR fragments.

The chicken trypsinogen PCR fragments were purified and labelled to a specific radioactivity greater than 1×10^8 c.p.m./mg of DNA by random priming [25]. Membranes were hybridized overnight in 50% formamide (v/v), $5 \times SSC$ ($1 \times SSC = 0.15$ M NaCl/0.15 M sodium citrate), 0.02 M sodium phosphate (pH 6.7), 100 μ g/ml denatured salmon sperm DNA, 1% SDS, 0.5% non-fat dry milk and 10% dextran sulphate at 37 °C. Following hybridization, these membranes were washed twice in $2 \times SSC$ with 0.1% SDS at 65 °C. The filters were then blotted dry with Whatman 3 MM filter paper and exposed to X-ray film overnight at -70 °C with intensifying screens.

DNA sequencing

The chicken trypsinogen cDNA and genomic flanking sequences were obtained directly from the λ cDNA and cosmid clones using cycle sequencing with *Taq* DNA polymerase. Primer sequences and their approximate locations on chicken trypsinogen gene are listed in Figure 1. Cycle sequencing reactions were performed with the Perkin Elmer thermal cycler (Norwalk, CT, U.S.A.).

	Signal peptideActivation peptide
Trypsinogen I	N K F L V L V A F V G V T V A F P I S D E D D D K I V G G Y S C A R S GCGAGGCT GTGTCAAGATC ATGAAATT CT TAG TG TTAGTCGCCTTTGT TGGTGTG A CTGTTGCCTTCCCCATC AGTGATGAGAAGAGTGATGACAAGATTGTGGGAGGCTACAGCTGTGGCAAGGAGTGC
Trypsinogen II	-TT GCA-CG-TC-TCCA-CTCCTGGCC
Trypsinogen I	A A P Y Q V S L N S G Y H P C G G S L I S S Q N V L S A A H C Y K S S I Q V K L G E Y CGCA CCTACCAGGTGTCTCTGAATTCTGGATATCACTTCTGTGGAGCTCTCCATCAGCAGCGCTGCGCCATGGGTGC TGTCAGCTGCTCACTGCTACAAGTCTTCCAT T CAAGTGAAGCTTGGGAG A TACAA
Trypsinogen II	ТТGTTТ
Trypsinogen I	N L A A Q D G S E Q T I S S S K V I R H S G Y N A N T L N N D I H L I K L S C C TG GCAGCCCAAGACGGCAGGAGGAG AC CATHAGTICATCAT AAGTCATCCGTCATCAGGCTACAATGC CAACAAC TGAACAATGATATCATGCTCATCAAGCTT TC C
irypsinogen ii	-AT-GA-TGTTGTGG
Trypsinogen I	KAATLNSYVNTVPLPTSCVT AGTTCLISGWGNTLSSGS AN AGCAG CC ACACTCAACTCCTACGCAACACAGTTCCTCTGCCTGACCAGCTGTGTGACC G CTGGAACA ACGTGCCTGATCTCTGGATGGGGGCAACACACTCAGCAGGGGCTTC TGTATCCAGA
irypsinogen ii	G-TTG-GT
Trypsinogen I	DVLQCLNAPVLS SSQCSSAYPGRITSNMICIGYLNGGKDSCQG TGTCCTGCAGTGCCTGAATGCCCCTGTACTCTCCT CAAGCCAGTGC AGCAGTGCTTATCCTGGAAGAAT TACCAGTAACATGATATGTATTGGATACCTGAATGGAGAAAAGACTCCTGCCAGGGGG
Trypsinogen II	АСССССССС
Trypsinogen I	D S G G P V V C N G Q L Q G I V S W G I G C A Q K G Y P G V Y T K V C N Y V S W I K T T ATTCCGGTGGTCCAGTTGTCTGCAATGGTCAG CTTCAGGGTATTGTTTCCTGGGGTATTGGATGCACAAAAAGGTTATCCTGGTGTTTACACTAAGGTTTGCAATTATGTTTCCTGGATT A AAACGACTA
irypsinogen ii	-CGÀCCCCC
Trypsinogen I	M S S N TGTCTTCCAACTGAGA C A C CC AT T TTACGTGACCTGCTTTTCCTCCT CCTCATT TCCTTCTGCT TTGGAATTGGATGTACACACCTTCACAAAAAA ATTAAA CATTTTCAGGCAAAAAAA
irypsinogen ll	КОТС-G-C-GA-GGTCGCC-CCССGG

Figure 2 Sequence alignment between chicken trypsinogen I and trypsinogen II subfamilies

The sequences from clones P1 and P29 are used in sequence alignment for trypsinogen I and II subfamilies respectively. The deduced amino acid sequences are listed with the corresponding cDNA clones. Gaps are introduced to maximize sequence identity and dashes are used to indicate sequence identity with the top line (trypsinogen I). The signal and activation peptides are indicated on top of the sequences. The polyadenylation signals are underlined and listed in bold typeface.

Expression of trypsinogen genes

To analyse the expression of trypsinogen genes from various tissues at mRNA level, sensitive PCR-based tests were used. To increase specificity and eliminate false PCR results, nested chicken trypsinogen gene-specific primer sets were used in succession. The primers used in PCR-based expression studies were TRYEX2, TRYEX3, TRYEX45, and TRYEX55 (Figure 1) for both trypsinogen I and II subfamilies. Control PCRs without template were included in all experiments. Approximately 0.5 ng of cDNA from various tissues were used in a 50 μ l PCR reaction as described above. The reaction mixtures were subjected to 20 cycles of amplification and each cycle consisted of 20 s of denaturation at 94 °C, 45 s of primer annealing at 60 °C and 90 s of extension at 72 °C. The results of PCR reactions were analysed by gel electrophoresis. A sample $(1 \mu l)$ of the PCR reaction mixtures was then removed and subjected to another 20 cycles of amplification with internal nested primer set (TRYEX3 and TRYEX45).

Sequence analysis

Sequence database search and multi-sequence alignments were performed with LaserGene sequence analysis program (DNAstar, Madison, WI, U.S.A.). The DNA and deduced amino acid sequences of chicken trypsinogen genes were compared with those found in GenBank/EMBL (release 80).

RESULTS

cDNA sequence analyses suggest that the chicken trypsinogen family encodes two major forms of trypsins

Approximately 10% of the plaques from the pancreatic cDNA library were hybridized with the chicken trypsinogen gene probes. Thirty different trypsin-hybridizing plaques were randomly picked, purified and analysed. Three of these clones carried very small inserts and the remaining 27 cDNA clones were characterized further by DNA sequencing. The sequencing results reveal that there are three different cDNA sequences and they can be grouped into two major forms of trypsinogen genes in chicken. These two forms of chicken trypsinogen gene cDNAs share approximately 69% sequence identity (including the 5' and 3' non-coding regions) at the nucleotide level (Figure 2) and do not cross-hybridize on genomic Southern blot under normal hybridization conditions.

In order to characterize the multi-member chicken trypsinogen gene family in a systematic fashion, like other multi-member gene families, the trypsinogen genes are grouped into two subfamilies based on sequence similarity; trypsinogen I and trypsinogen II. Results from Northern hybridization (data not shown) and the number of sequences represented in cDNA clones suggest that the chicken trypsinogen I subfamily genes (24 from a total of 27 sequenced chicken trypsinogen cDNA clones) have a higher expression level than trypsinogen II (three from a total of 27 sequenced chicken trypsinogen cDNA clones) sub-

		Signal peptide	Activation peptide	Enzyme			
<u></u>							
EUMAN-TRY1	Het Asn Pro Leu	Leu Ile Leu Thr Phe Val Ala Ala Ala Leu Ala Ala	Pro Phe Asp Asp Asp Lys	The val Gly Gly			
EUMAN-TRY3	Met Asn Leu Leu	Leu Ile Leu Thr Phe Val Ala Ala Ala Val Ala Ala	Pro Phe Asp Asp Asp Lys	ile wai diy diy			
EUNAN-TRYJ	Het Asn Pro Phe	Leu Ile Leu Ala Phe Val Gly Ala Ala Val Ala	Val Pro Phe Asp Asp Asp Lys	Ile val GIY GIY			
BOVINE-TRY2 (J	A) Het His Pro Leu	Leu Ile Leu Ala Phe Val Gly Ala Ala Val Ala	Phe Pro Ser Asp Asp Asp Lys	Ile Val Gly Gly			
D00-TRY1 (C	C) Net Lys Thr Phe	Ile Phe Leu Ala Leu Leu Gly Ala Thr Val Ala	Phe Pro Ile Asp Asp Asp Asp Lys	Ile Val Gly Gly			
DOG-TRY2 (J	A) Net Asn Pro Leu	Lou Ile Leu Ala Phe Lou Gly Ala Ala Val Ala	Thr Pro Thr Asp Asp Asp Lys	Ile Val Gly Gly			
NOUSE - TRY	Net Ser Ala Leu	Lou Ile Leu Ala Leu Val Gly Ala Ala Val Ala	Phe Pro Val Asp Asp Asp Lys	Ile Val Gly Gly			
RAT-TRYL (J	(A) Het Ser Ala Leu	Leu Ile Leu Ala Leu Val Gly Ala Ala Val Ala	Phe Pro Leu Glu Asp Asp Lys	Ile Val Gly Gly			
RAT-TRY2 (J	A) Net Arg Ala Leu	Leu Ile Leu Ala Leu Val Gly Ala Ala Val Ala	Phe Pro Val Asp Asp Asp Lys	Ile Val Gly Gly			
BAT-TRY3 (C	(C) Het 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			
	(C) Het Lys Ile Ser	Ile Phe Phe Ala Phe Leu Gly Ala Ala Val Ala	Leu Pro Val Asn Asp Asp Lys	Ile Val Gly Gly			
CRICKIN-PRV T //	(C) Net Ive Phe Lev	Val Imu Val Ala Phe Val Gly Val Thr Val Ala	Phe Pro Ile Ser Asp Glu Asp Asp Asp Lys	Ile Val Gly Gly			
	(b) Not Ive Do Iou	The Ley Ser Or Ley Cly Ma has be vel als	Phe Pro Gly Gly Ala Amp Amp Amp Ive	Ile Val Gly Gly			
			Dhe hen hen hen lys	The The Gly Gly			
ABBUFUE-TRI	Het Lys Phe Let			The Val Civ Civ			

Figure 3 Trypsinogen signal and activation peptides from various species

The cationic and anionic trypsinogen isoenzymes are indicated as 'A' and 'C' respectively, following the name of each trypsinogen. Gaps are introduced to optimize the sequence alignment. Sequences are from the following: Human-try1 and try2, [2]; Human-try3, [3]; Bovine-try2, [5]; Dog-try1 and Dog-try2, [6]; Mouse-try, [13]; Rat-try1, [9]; Rat-try2, [10]; Rat-try3, [11]; Rat-try4, [12]. Chicken-try I and II are from this study; *Xenopus*-try, [14]; and Atlantic Cod-try, [16].

	IΥ	"	ψı	16	£	E B	5	¥.I	<u> </u>	q	¥.S		L	5.6	<u> </u>	Ľ	3.4	ς	6	5	1	I.N	\$	ىھ	L	¥	s	•	Ľ	L	Y	ĶS	5 R	L	<u>0 y</u>	R	1	i f	н.	11	0	V. L	E	<u>6 N</u>	ţ.	E	IN		A K	I	I.Ŗ	H P		LH S	. N.	Major	ity	
			20	•			-			30						_	<u>.</u>	÷	_	_			1	50			_		1									70			r-r-				80								121.					
24	iv	2 2	ΥĽ	Ϊċ	Ē	ĒN	ŝ	¥	Ÿ	ą	v 3	ł.	N	56	1-1	Y	A F	č	66	ŝ	Ľ	ז ו ז ו ז	Ľ	Q I		v	ŝ	î١			Ŷ	K S	ĒŔ	i	ąν	Ř	L	E		ii	Ē	νι	Ē	6 N 6 N	E	i.	1 8	2	A K A K	i	I R I R	H F	R			HUMAN	-TRY1	
24 37	IV	66	Ţ.	ŢĊ	F.	5.	S	Ļ	ΡΥ	8		L.		56	E	3	4 E	ç	6.6	ŝ	Ł	1 5	Ę	Q			ş			l C	î	Ľ٦		ļ	Q V N V	1	L	F		I	K.	V L	1	6 N 6 N	E	Ì.	IN	A .	A K	Į	IR	HF	5			HUMAN		
9	i v	è è	Ŷ	İÈ	Ř,	A B	ŝ	i.	ΡŸ	ğ	v s	i	Ň	ŝ	Ι.	الك	H F	è	ċ	ŝ	ì.	iï	ŝ	Q.	i v	v	ŝ	Ā.	A H	Ċ	Ý,	ĸ	<u>s</u> R	ĵ.	ò v	ŝ	ĩ	È	Ë,	Î	P 1	νĩ	Ē	¢ ji	È	i e	ij	Ā,	Ã, È	i	įđ) H I	Ľ.		Ľ,	PIG-T	RY	
24 7	1 V 1 V	66	¥.	ŤĊ	۱ <u>۹</u>	Ŧ,	ĥ	v	ΡΥ ΡΥ	8	v	L	NL N		1.	¥!	X F H F	ĉ	66	ŝ	L	IN IN	L S	Q I		ł	s	A :	A H A H		YL	<u>e</u> 1	516	I	Q V 0 V	R		E	Ď.		ឝ	٩Ļ	າຄິ	6 <u>6</u>	E	35	IIS	11	SK	L.	IR	1 8 8	S.	נעטי וווי		BOVIN	E-TRY1 E-TRY2	
24	ΪV	ŝ	3	ţç	<u>۽</u> ا	Ŗ	Ş	Š.	P Y	ģ	ý s	Ĩ.	Ň,	<u>s</u> s	-	Ŷ	Ë	ç	Ģ	ş	ĩ.	Į,	ŝ	ò		Ś.	ŝ	Ă .	Ă Ì	ļ	Ŷ	Ķ	S R	ļ	ğ y	Ë	ĩ.	E	Y.	ļ	إلها	<u>ڊ</u> ان	١Ē.	يقاغ	Ē	į,	ĪN	A	ĀK	ī	IR	Ĥ	إليا	ΠŪ	μü.	DOG-T	RY1	
24	i v	6 6	Ý	i č	Ò	έſŝ	ηs	v :	ΡŸ	ě	v	ιì	ΠĽ		FI	Ϋ́	ĤF	č	Ğ	ŝ	ĩ	iï	ηő	ě		v	ŝ	Â.	Ā	ìč	Ý	κŪ	Ð,	i	ð i	R	ti	Ē	₩ï	ñÎ	Ň٩.	νì	È	č ji	È	i i	Ŷ Ď	s	ÂŔ	Ч	iĵ	, Ĥ i	i.	L.		HOUSE	-TRY	
24 24	IV	66	Ŷ	ŤĊ	6	20	US IS	š	ΡΥ ΡΥ	8	v s	1 L 1 L	N '	s e s e		¥!	HF	ĉ	60	5	L L			8		/ ¥	ş	\$	A }	10	¥	K 1	SR SR	I	ŝ١	R R	L			N I N I		ΥL	E	60	IE (8 F	IN	A	A X		IK		, N	감망		RAT-T RAT-T	IRY1 IRY2	
25	ÎÝ	6 6	Y.	ţ	lè	ξĻ	ļŝ	Û	PY	ò	ý s	i i	ij.	Ā	j.	4	ËĖ	ŝ	ŝ	ŝ	ŗ.	i	15	ĝ.		Ż,	ŝ	Ă.	A 1	įç	Ŷ	Ŝ,	<u>s</u> 8	Ľ,	<u>à</u> i	R	ĩ	Ē	Ň	N I	æ	٧D	١ę.	ŝ	Ē	ÌF	i	14	Å.5	ć į	1 8		19		Ę.	RAT-1	1113	
1	iv	6 6	۲C	a da	A.	ì		Å	ΡŸ	ě	v	51		ŝ	11	Ϋ́,	H F	č	ŝ	: ;	ĩ	ili	s	ġ.		lί	ŝ	Â	21	ič	Ý	κ.	ŝ	٢ï	٩ì	Ċ	12	iŧ	ក៉ារ៉	۳¢		Å	, b	3	E.	រំកា	is	3	s i	٢	İÂ	HC	28		ίď.	Chic	ien Try	I
1	1 9	66	Ϋ́	Εð	P		LS.	¥	P	- 8	v	5 L 5 L	N	S a		I¥ '	HF	ĉ	6 1	6 S 6 S	÷L.	10		움	W \ W \	/₩	ļŞ	\$	* ;	4 C	Y	K :	S R Ri R	I	8	/ R / R	L	6 E 6 E	뙈	틞I	ß	Ň	E	e s	E	V V T YI	ĥ	S	SN	4	IR		S KI	¥ S S	ĮΪ	Chick	ken Try ISH-TRy	II
21	iΩ		Ū,	75	4	K S	S	X	ŗ	μ	v :	i i	5	ş		13	HF	ç	<u>ġ</u>	ŝŝ	ĥ	ĮΪ	Ņ	ធ្តេ		<u>.</u>	ş	٨	ĸ	1 9	ĩ	κl,	1 5	ų.	ġ	/ R	Ē	ĢĒ	22	N,İ	A.	ĥ	E	G T	E	3.6	IS	ş	ŝ	Q V	I.	HE	Ē,	Y H S	S Y	XENO	PUS-TRY	f
19								ы. ,							. e	-				• •				- 100				2					эц 						-					•11		,									аш •	ACT.		y
		100	1					1	<u>10</u>	-			•			120	<u>, 1</u>	-	<u>r</u>		•	1						1	2		14)		-	2		150	LP		~	<u> </u>	÷	<u>ب</u> ت د	<u>, , , , , , , , , , , , , , , , , , , </u>		-			÷	17	<u>, 1</u>				7 I	мазон	rity	
103	Τι	ŵ,	P	D	<u>c</u> r	11	κι	ş	ŝū	A	۷	1	A	R	rs	力	1	ĩ	P	T A	P	P]/	ι Π	Ģ	Ť	٢Ç	ι	I	5 (G	6	N	TL	JS	5 (Ø	Y P	D	EL	. 9	C (D	A P	۷	LS	91	K	C I	EĂ	5 1	P	G K	I T	S N	HUNA	H-TRY1	
103	÷ι			13	ιi	ł	κι κι	ŝ	31	â	V		Å	R	/ 5	h	IS	1	P:	HA	P	P	1 1	ĉ	눼		ľ	I	5	6 H 6 H	6		Ť	S	南		8	Y P Y P	D	E	ផ្ដា	ci	. D	A P	÷.	۲Å	å /	E	S	E A	30	5			L.S.	HUNA	H-TIRYZ H-TIRY3	
115	ŦĽ	01			1 L	I	K L K L	s	51	1	싺		â		냚	ιŦ	Ϋ́	ł	P	ĮĄ	ę	21	1	ŝ	邗		ł	ł	S	61	6	2	ΤL	a s	Ę	Â	jő	Y P Y P	뭐	E	١¥	61	ĥ	AF	: .	١Ų	84	ĮĮ	2	<u>لا م</u>	Si	P	للالع	단	<u>ון</u> א	HUNAL PTG-1	N-TRY4 TRY	
103	ŢĨ	Å.	Ď	įÓ	Ď,	ļ	ĸ	ŝ	Į,	Â	Д	Dï	Ā	R	រនី	Ţ	ţŢ	βį	pl	Ş,	ķ	٩ľ	Ŝ.	Ģ	ţŲ	ļ	ĩ	i	ş	Ģ	Ģ		įĮ	5	ş	s v	Ū,	ŶP	Ĩ,	ιī	à	çi	. Y	Â	Ľ.	ī š	μĽ	Ņ	ŝ	ĒĀ	5	È.	ġ	i il	ij.	BOVI	E-TRY	Ļ
103	цţ	١٣ i	N D	ii	ii	i	κ i	ŝ	ŝ	Â	H	1	ŝ		15	JÅ,	ij	51	۴ſ	k s	, či	Ê	Ň	ċ	ť	à à	t	i	ŝ	ċi	Ğ	i.	ŧΖ	ŝ	ф	ġ	Ľ,	ŶP	ě	٧Ľ	ĽΫ	ĉ		Â	i	ιs	le i	Ļ۷	č	<u>R H</u>		P	çõ	id	5 N	DOC-1	TRY1	
103 103	Ψί	DI	ND		1 L	I	K L K L	à	5 I 5 I	₽	Ť	L N L N	\$	R	/ A	멳	뷞	51	밁	R A Sís	i c	Â		ŝ	Ŧ		ł	I	s	6 1 6 1		N	Ŧ	S	ĥ	ŝ	#r	¥Р,	w	ιι ιι	lg	a	28		ςΨ		18 /	<u>ا</u> عا	i c i	EA	51	r P r P	ŝ	H	E N M N	DOG-	TRYZ E-TRY	
103	Ŧ			H	1 L	I	K L	ş	ŝ	١X	K		•		<u>؛</u> ار	Ľ₽	31	41	2	ş, آ	ļċ	1		ç	Ŧ	Q Q	÷	ŗ	ş	61	16	8	Ţ	ş	بلا	c V	1	N P	2		ģ	<u>e</u> h	d b	. :	ሕ	1.5	18		21	ĒÅ	٩	P	E	ŢŢ	រុទ្ធ	RAT-	TRY1	
104	ŢŪ	Jo I	ŇĎ	ij	Ĩ	Ţ	Ë I	Ð	ŝ	Ā	ιĝ	L H	ទ្រី	<u>Ř</u>	Ś	Ţ	ÿ,	Ĩ	P	RS	ç	Ľ٢,	5 5	l¢.	į,	È à	ĩ	Ň	ş	ç i	ġ	ÿ.	İ.	ŝ	Å.	Ġ	Ц	ŶP	١Ā	ī ī	Ìĝ	ç	Ĭ,	Ā	ž	ĩ ş	D	ŝŝ	ķ	K S	<u>ן ז</u> ו	Ċ P.	لياغ	iΪ	S N	RAT-	TRYS	
10	ίì	M	ŇĎ	ii	ΪÌ	i	R i	5	È,	٩.	π		ß	¥.	ý,	ų,	ЪЙ,	ին	P	Ĩŝ	à	£۵	τh	6	ŧ	ħλ	Ξt	Ŧ	ŝ	ċ	16	8	ł٩	ŝ	ጟ	c s	للم	Ϋ́́́́Р	Ð	b١	ιğ	č	L N	Â	Ϋ́,	L S	ŝ	5 Q	Z	ŝŝ	4	ŕÞ	٥Ň	ił	ŝï	Chic	ken Try	y I
80 87	ΨГ	u u	N D N D	I	N L N L	ī	K 1 K 1	ŝ	£1	X	1Å[<u>Y S</u>	븉	D.	ĥ	11	Ы	11 L 5 L	PL	븱	ĥĉ		K A Y A	6	Ŧ	E C Mic	Ľ	I	ŝ	6 1 6 1	16	N	Н	L S	J NL	Ç ۲	ų	Į P	16	հե	LQ	ĉ	ᄈ		зw	LS	6	D E	2	QE KG		Y P.		Ιł	S H	Chic	ken Try ISH-TR	y II Y
100	H	-	N D	H		I	K 1	r \$	<u>s</u>		ş	Ĺ	A	\$	V N	Ľ,	V.	<u>]</u> [P	ş	ç	ব্		Ś	Ţ	S	Ļ	Ĵ	ş	Ġ i c i	İĠ	Ľ.	Ŧ		ų	Ş s	R	۲P	0	Π.	ĪÒ	ç	Ī	Ą	Ċ,	Ī		<u>م</u> ا ا	ļ	N S	لم	Y P	G E	ΞÌ	Ä	XENO	PUS-TR	Ý
-			~ c								<u>د</u>	 n .		~					Ś			,	~			~ .									-						- 4 - 1		н. т.									•••	- 12	• ••		Maia		
	180	•							190	•		1			-	200	<u>بد</u>		×	21		•					22		*					-	2	30	•					2	40	-			-											
183		Ş	V ¢	Ę	L	ç	6 1	D	ş	5	ç	0 5	ç	ç	2	Ū.	<u>F</u>		ይ	Ļ	ŞĢ	Ŵ	V S		Ģ	Ð.	5	A	Q	K	K	P	5	v v	Ţ	K \	۲Ľ	NY	(yl	Ķ,	ļ	K.	N T	ļ		MIS									-	HUNA	H-TRY1	
183	- Fili	ċ	v č	Ē	ì	ç	ŝ	Ď	ŝ	ù	h	Ď	č	è	ΡŸ	i v	Ę,	Ň	ě	tè	ŝ	Ŵ	vš		š	南		Â	m	R I	NR	F	è i	v v	t	κï	Ý	N 1	r v	Ďi	W İ	R.	ĎŤ	1	ÂÂ	ñ s										HUNA	W-TRY3	
195	10		V 6 V 6	F	ι Ε ι Ε	Ĝ	6		s	čč	Ĩ	0 1	6	ŝ	ΡV PY		č	N G N G	8	l	36	Ψ	V 5 V 5	5 W 5 W	e	₩,		•	<u>اللا</u>	KI	N LB N K	P	5	ŶŸ	Ŧ	K \ K \	ųΨ	N 1 N 1	ŕv	å'	W I W I	삓	<u> </u>	I	A A A A	NIS										PIG-	VI-TRY4 ·TRY	
183	. N. 1	١٤	Â	Ψ	LE	ŝ	6		ş	ç	į	0 9	ę	ŝ	PV	ιQ	iç,	i c	Å	Ĺ	į	ļ	¥ s	S W	ŝ	÷.	5		ğ	ĶΕ	ŝ	2	ŝ	Ϋ́	Ţ	£ (ļ		r vi	Ā	ŧ Į	لفي	ÈŢ	Į.	A A	NS	•									BOVI	NE-TRY	1
183	ΞŪ.	ç	î e	Ľ۲.	Ì	¢	ŝ	ÈĎ	ŝ	èè	Ģ	Ď	Ģ	ē	ē j	i v	č		Ê	ĩ	ġ	Ŵ	v i	ŝ	ç	Å.	Ġ	Ā	ğ	ŝ	Ĉ,	P	ē	vIS	ė	R 1	i èl	Ē,	L V	إليا	īi	٩,	ğİ	i	ĀĀ	ï.										DOG-	TRY1	٤.
183	i.		V G	1		č	è		ŝ		16	0 1	6	ĉ	PI	÷ *	č		Ē	l	ŝ	i	v	5 8	ŝ	Ŷ	50	Â	å	P	<u> </u>	ŀ	5	v v	Ŧ	K I		"u	ŸŸ	5	H I	ଥ	빏Ŧ	I	ÂÂ	N 3	5									HOUS	-TRYZ Se-try	
183	N	Į,	ve		11	2	ŝ		ş	ĉ	į	0 9	5	ŝ	2	::	ŝ		R	1	Q G	1	¥ s	5 1	ŝ	Ĵ.	6	1	Ŀ	2		12	ę	Ϋ́,	1	K 1	::	ЪЦ	ÊÂ	ផ្តោ	! !	8	DT	1.		2										RAT	TRY1	
184	H	٢Ś	Ē	F	í i	Ģ	Ğ.	Ċ	ş	Ś	į	Ď	Ģ	ĕ	P 1	Į į	č	i i	Ģ	Ŧ,	ò	Ŵ	ý i	s i	Ğ	ż.	ŝ	Â	<u>e</u>	Ķ	ŝ	2	ĕ	ý ý	İ	Ĕ.	ç		ý,Ý	١	ŧ i	è	9 I	Ø	Ă	<u>.</u>										RAT	TRYS	
16	i ii	۲ċ	وليا	Ġ	td	D ¢	è	È D	ŝ	č	ŝ	Ď	ŝ	č	5	νv	č		5	ť	ě é	i	v	5 1	Ğ	۱ĩ٢	5	i â	2	Ŕ	54	hé	č	v i	÷	ř.	vč		YW	ß	# i	Ľ	ŧł		33		•									Chi	cken Tr	γI
160		E C	VG	÷	늷	6	5	K 0 K 0	s	ĉ	56	D	56	6	PI	÷۷	ĉ	N (Ľ		Q Q	I	¥.	5 W 5 W	6	ų,	60		Ľ	¥	6) D	16	ŝ	v v v v	Ŧ	ĥ	V C V C	₽	¥ ¥ ¥∙¥	เริ่า	W I W I	<u>ل</u> م الا	E	I		JY S 1	,									Chi DOG	cken Tr FISH-TF	<u>ү</u> п
18		1 c	т e	ĬŶſ	M	6	ŝ	K 0 K 0	S	ĉ	ŚŚ	D	ŝ	ê	PI	Ϋ́	ŝ		.0	Ē	ġ ġ	Ø	Y.	s i	Ś	Ť.	ŝ		2	5		15	ŝ	v i	L.	K	v ć	N	λĺ	A	١, ٣	9	ШŢ	Ā	1.											XEN	OPUS-TI	Υ.

Figure 4 Trypsin sequence alignment from various species

The amino acid sequences that differ from the majority are boxed. † represents residues involved in catalytic activities and * indicates the residues involved in disulphide linkages. The amino acid sequences are numbered according to the standard bovine chymotrypsin numbering system. The GenBank accession numbers for the sequences are: Human-try1 (A25852), Human-try2 (B25852), Human-try3 (X15505), Human-try4 (X72781), Pig-try (A00947), Bovine-try1 (A00946), Bovine-try2 (X54703), Dog-try1 (B26273), Dog-try2 (A26273), Mouse-try (B25528), Rat-try1 (J00778), Rat-try2 (A22657), Rat-try3 (A27547), Rat-try4 (S05494), Dogish-try (A00950/B27719), *Xenopus*-try (X53485) and Atlantic Cod-try (X76887/X75998).

family genes. There are two different but very similar cDNA sequences (more than 99% sequence similarity) from the 24 trypsinogen I subfamily cDNA sequences and one from the three trypsinogen II subfamily cDNA clones. The possibility of the

two highly conserved trypsinogen I sequences actually being contributed by different alleles can not be excluded, since the genomic DNA and mRNA are not obtained from inbred chickens. The complete cDNA sequences and deduced amino Trypsinogen gene 5' non- coding sequences Initiation codon

Human TRY1	ACCACC	ATG
Human TRY3	ACAC TC T ACCACC	ATG
Bovine TRY2	TC T CCACC	ATG
Dog TRY1	ACACCCCAGACCAGATCCAGGG AG CAACC	ATG
Dog TRY2	ATACTTC TG CCATC	ATG
Mouse TRY	GACTTTC TG TCACC	ATG
Rat TRY1	CCTTC TG CCACC	ATG
Rat TRY2	TACCTTC TG CCACA	ATG
Rat TRY 4	GCAC T AG G CACC	ATG
Chicken TRY I	GCGAGGCTGTGTCA AG ATC	ATG
Chicken TRY II	CTTGGTGGTGGA AG CACC	ATG
Xenopus TRY	TACTGA TC T CCACC	ATG
Atlantic cod TRY	CGGCACG AG G CGACC	ATG

Figure 5 Comparison of 5' non-coding regions of 13 different trypsinogen genes

Gaps have been introduced to maximize homology between the sequences. The initiation codon (ATG) for trypsinogen genes is indicated on top of the sequences. The origins of the sequences are listed in the legend of Figure 3.



Figure 6 Genomic Southern blot analysis of the chicken trypsinogen gene family

Chicken liver DNA is digested with restriction enzyme *Hin*dIII, electrophoresed through a 0.8% agarose gel, transferred to nylon membrane, and probed with chicken trypsinogen I and trypsinogen II cDNA probes as indicated on top of the Figure. DNA size marker (1 kb marker) from BRL is used as size standard.

acid sequences for trypsinogen genes from trypsinogen I and II subfamilies are listed in Figure 2.

(1) Coding region

The trypsinogen I and II cDNA clones contain 741 bp of coding region, a short 5' non-coding region and approximately 100 bp of 3' non-coding region (Figure 2). The 741 bp coding regions translate into 247 amino acids in length. Like all the known trypsins, the chicken trypsins are also synthesized as preproenzymes which contain an N-terminal signal peptide (pretrypsinogen) followed by a short (six to eight amino acid residues) activation peptide. The coding region cDNA sequence similarity, excluding the signal and activation peptides, between trypsinogen I and II subfamilies is 71 % which is, as expected, slightly higher than the overall cDNA sequence similarity (69 %).

Like most of the species studied, chicken pancreatic trypsins can also be divided into cationic and anionic trypsin isoenzymes



Figure 7 Amplification of trypsinogen genes with cDNA extracted from various chicken tissues

Trypsinogen I and II subfamily-specific TRYEX2 and TRYEX55 primers were used for the initial amplification, followed by 20 cycles with the subfamily-specific TRYEX3 and TRYEX45. The expected size for trypsinogen I and II subfamily amplification products are 355 bp and 170 bp respectively. The intensity of the bands does not reflect the transcript concentration, since these were from PCR amplifications with nested primer sets used in succession. Size standards were from BRL (1 kb marker; BRL, Gaithersburg, MD, U.S.A.).

based on their isoelectric points. The deduced amino acid sequences suggest that the chicken cationic trypsin isoenzymes are encoded by the trypsinogen I subfamily and the anionic trypsins are encoded by the trypsinogen II subfamily.

(i) Signal peptides

The chicken signal peptides for trypsinogens I and II are 15 to 16 amino acid residues in length, which is similar to that of rat, bovine, dog and human pre-trypsinogen signal peptides (Figure 3). The sequences of signal peptides follow the usual pattern of a charged residue (Lys at second amino acid position for chicken trypsinogen genes) followed by hydrophobic residues and small uncharged amino acids [26]. It is interesting to note that all the anionic trypsinogen signal peptides including chicken, contain a conserved Leu-Ile-Leu hydrophobic region (Figure 3). This sequence conservation may direct the anionic trypsinogen to a particular secretion site or influence the rate of secretion or translation.

(ii) Activation peptide

The typical activation peptides for trypsinogens are six- to eightamino-acid residues in length and composed of a cluster of anionic amino acid residues. The chicken trypsinogen I and II activation peptides contain ten and nine amino acid residues respectively. This makes the chicken trypsinogens possess the longest activation peptides among all the known trypsinogens. Unlike most of the trypsinogen activation peptides, the activation peptide for chicken trypsinogen II contains three rather than four anionic residues; one glutamyl and two aspartyl residues. The chicken trypsinogen I and II activation peptides also preserve proline (Pro) as the second amino acid (Figure 3).

(iii) Trypsin

The chicken trypsins are composed of 223 amino acid residues. No significant number of gaps are required to align the amino acid sequences of the chicken trypsins to other known trypsins (Figure 4). The overall sequence similarity to other known trypsins is about 70 %. The deduced molecular mass for chicken trypsin is 23334 Da for trypsin I and 24042 Da for trypsin II. Both chicken trypsins contain all the major catalytic and

(a) Exon 3 Exon 4 Exon 5 Exon 1 Exon 2 11111 11111 -// // Intron A Intron B Intron C Intron D 254 137 248 59 166 Trypsinogen I 890 700 560 600 137 61 163 254 250 Trypsinogen II 800 720 280 720 (b) Trypsinogen I genomic Trypsinogen I genomic Trypsinogen I genomic Trypsinogen I genomic acagcacagagagateteaaggggttactatgaacaagtgetetga aaaocttggggttaacgataagocatttcagaagatgtgtttctagaagctcaagttgccagtccaccagtctgggatcgaggctgtgtcaagatcATGAA Trypsinogen I gene Trypsinogen I P 1 Trypsinogen I P36 ------Trypsinogen I geno Trypsinogen I P 1 Trypsinogen I P38 Trypsinogen I geno V A P P I B D B D D D E I V G G Y S C A R S A A P Y G V tattet tattetgragtfioccttccccatcastgatgacgatgatgacgacgatgatggacgaggctacaggtgccccaaggtgcccccaccctaccaggtg Trypsinogen I gene Trypsinogen I P 1 Trypsinogen I P38 Trypsinogen I geno Trypsinogen I P 1 Trypsinogen I P30 Trypsinogen I gene Trypsinogen I P 1 Trypsinogen I P38 Trypsinogen I geno Trypsinogen I P 1 Trypsinogen I P36 N L I E L & E A A Y L & & Y V Y Y V L J Y & C V Y G Y G Y C L I S A(1914) GCTEATEAMACHTECHAGCAGCAEACHCTECTACGAEACHGTECTECTACCAECTIGTAGCAECTIGAACAAGTGGCCTAATCTCTG GCTEATEAMACHTECHAGCAGCAECAECTIGTAGCAECTIGTAECTIGTAECTIGTAECTIGTAECTIGTAECTIGTAECTIGTAECTIGTAEC Trypsinogen I geno Trypsinogen I P 1 Trypsinogen I P 1 Trypsinogen I geno Trypsinogen I P 1 Trypsinogen I P38 Typsinogen I gen Typsinogen I P 1 Typsinogen I P 1 _____ V L S S C C S A Y P G R I P S M M I C I G Y L M G G R D S C Q ACTCTCCTCAAGCCAGTGCTAACCCTGGAAGAATTACCAGTAACATGATGATGATGCAGAAGAAGACTCCTCCAGgtaa Trypsinogen I gene Trypsinogen I P 1 Trypsinogen I P38 ------Trypsinogen I genomic Trypsinogen I P 1 Trypsinogen I P38 W & I & C A D K & Y P & V Y T K V C H Y V S W I K T T M S S H * GOOGTATTOGATGTGCACAAAAAGGTTATCCTGGGTTTACACTAAGGTTGCAATTATGTTACCTGGATTAAAACGACTATGTCTTCCAACtgagacact Trypsinogen I genomic Trypsinogen I P 1 Trypsinogen I P30 -----Trypsinogen I geno Trypsinogen I P 1 Trypsinogen I P38 _____ den T debi t<u>tgtgcagt</u>tectctggttacttccatggg (c) Trypsinogen II genomic Trypsinogen II genomic Trypsinogen II genomic gccactgaatggagggcacctgtgttagatgttaccaaggtcaacaaggcaacgtcagcttgcatttctaaaaatacgtggtcctgctttagTrypsinogen II genomic Trypsinogen II P39 Trypsinogen II genouic Trypsinogen II 729 ctgcattt//ggagtcaggcctggtgctgtgaaagcccattgtctatcttgcagTTocTTTCCCTGGAGGTGCCGATGATGACAAGATTGTGGGAGGCTAC Trypsinogen II genomic Trypsinogen II P29 ACTGOCCAGAGCATTCOGTACCAGGTGTCCCTGAACTCTOGGTATCACTTCTGTGGAGGGTCCCTCATCAACAGCCAGTGGTGTCCTGTCAGCTGC Trypsinogen II genomic Trypsinogen II P29 A H C Y K S A I Q TCACTOCTATAAATCgtaagtatggagegggtttegt//etgeceaacceaeggetaactgetgttttteeetgetggeaattttgeagCGGTATACAAGT Trypsinogen II genomic Trypsinogen II P29 TREAST TRANSPORTAGE ACATEGATOTIC CAGAAAAACAAGTGAAAGTAGTCAAGAGTTCAATAACAGTTCAATAAGTTCAATAGTGAAGTTCAATAGTTCAATAGTTCAATAGTTCAATAGTTCAATAGTGAAGTTCAATAGTTCAATAGTGAAGTTCAATAGTTGAATAGTTCAATAGTTGTTGAATAGTTGTTGTTGAATAGTTGTTGAATAGT Trypsinogen II genomic Trypsinogen II P29 N D I N L I E L A S A V E T S A D I Q P I A L P S C A E A G T S ATAATGACATTATGTTGATCAAGTTGGCATCTGCGGGGGATACAGTGCGAGATTCAAGCCGAGCTCCTGTGCCAAGGGGGGGACGGGG Trypsinogen II genomic Trypsinogen II P29 C L I S G W G H 7 L S H G TOCTTGATCTCAOGGTGOGGAAACACGCCTGAOCAATGGCTgtaagtatcaacacgcgctgccagtaccggtttgcacttct//ttttgttaagtgacaga Trypsinogen II genomic Trypsinogen II P29 T H Y P H L L Q C L H A P I L S D Q B actgggcccaggagaaagacaaatgttcgctactctcccggtagACAATTACCCTGAACTTCTCAGTGCCTGAACGCTCCGATTCTGAGTGACCAAGAGT Trypsinogen II genomic Trypsinogen II P29 C Q B A Y P G D I T R B M I C Y G P L B G G E D S C Q GCCNGCAGGCCTACCCGGGTGATATCACCAGGAACATGATCTGTGCGGGTTCCTGGAGGGTGGGAAAGACTCGTGCCAGgtaagagttgaatgttggtgc Trypsinogen II genomic Trypsinogen II P29 G D S G G F V V C B G E L G agagecetgggccacacacecec//ttagaagttcatggccctaatttttttcctccagGGTGACTCGGGTGGACCAGTTGTGCAATGGAGAACTGCAA Trypsingen II genomic Trypsingen II P29 GIVENCIA CALEGYPOVYTEVCENTACCOAGTTACCCAGAGTCTGCAATACGTTGACTGACCAGAAACCATGCAG Trypsinogen II genomic Trypsinogen II P29 A I CTACtgacagecagaeggeeteatgeteettggetgeteeeeteetgeeettateatttteetettaagaaaatggeaacageaca<u>aataaa</u>aacettaet Trypsinogen II genomic Trypsinogen II 929

taaatgaaccttacgatagttac<u>tttggagg</u>ctgaagctcatcagagccaagttgggttccc

Trypsinogen II genomic Trypsinogen II P29

structural features of trypsin. In addition to the catalytic triad, His-57, Asp-102 and Ser-195, and the obligatory Asp-189 (in bovine chymotrypsin numbering system) [27], 12 cysteine residues are found in the same positions as those in bovine, dog, pig, rat and Atlantic cod trypsins. Although the pairing of half-cystine residues in chicken trypsins has not been determined, it seems likely that the six disulphide bridges linking cysteine residues 15-145, 33-49, 117-218, 124-191, 156-170 and 181-205 (in bovine chymotrypsin number) [28] are conserved in chicken. Since catalytic activity of trypsins is due to the ability of proton transfer among His-57, Asp-102 and Ser-195, as expected the sequences surrounding these amino acids are highly conserved in trypsin (Figure 4). The amino acid residues which involve the calcium-binding site, including Glu-70, Asn-72, Val-75, Glu-77 and Glu-80 are also conserved in the chicken trypsins except that the chicken trypsin II has conservative substitutions at Val-75

(2) Non-coding regions

Most of the cDNA 5' non-coding regions for known trypsin genes are about 12 to 15 nucleotides in length. Even though the exact transcription start sites for chicken trypsinogen I and II genes have not been determined, the cDNA sequences reported in this paper could be near full-length since the sequences reported already have longer 5' non-coding regions than most of the known trypsinogen genes (Figure 5).

(from Val to Ala) and Glu-77 (from Glu to Asp) (Figure 4).

In accordance with the previous findings on 5' non-coding sequences of eukaryotic mRNA [29], the nucleotide positions three and four bases upstream of the initiation codon (ATG) are an adenosine (A) and cytosine (C) for most of the trypsinogen genes (Figure 5). It has been suggested that the presence of a short 5' non-coding leader sequence may be essential for the stabilization and efficiency of initiation of translation of eukaryotic mRNAs [29]. There is little sequence conservation among the trypsinogen 5' non-coding regions (Figure 5). Therefore, it is unlikely that this region plays any regulatory roles for trypsinogen gene expression or translation besides the stabilization of the initiation machinery for translation.

The 3' non-coding regions for chicken trypsins I and II are 103 and 105 nucleotides respectively, and are longer than the 3' noncoding regions of most of the trypsinogen genes reported.

The chicken trypsinogen gene family may be encoded by as many as nine different members

Using trypsinogen-subfamily-specific probes, six chicken trypsinogen I-hybridizing bands and three trypsinogen II-hybridizing bands can be visualized with Southern hybridization on *Hin*dIII-(Figure 6) and *Bam*HI (data not shown)-digested genomic DNA. Therefore, the chicken trypsinogen gene family may contain as many as nine different genes. Six of these belong to the larger trypsinogen I subfamily and the other three belong to the trypsinogen II subfamily.

The expression of trypsinogen genes can be detected in various tissues from chicken

Using PCR with nested PCR primer sets, the chicken trypsinogen

477

genes from both chicken trypsinogen I and II subfamilies can be detected in various tissues that including liver, pancreas, spleen and thymus (Figure 7). In pancreas, the trypsinogen-specific amplification products can be observed by gel electrophoresis within 10 cycles of the initial amplification process (with TRY-EX2 and TRYEX55 primer set). For other tissues, the trypsinogen amplification products can only be observed after two rounds of amplifications with nested specific primer sets (a total of 40 cycles of amplification). This indicates that the expression levels of trypsinogen genes in liver, spleen and thymus are much lower than in pancreas.

By screening a chicken spleen cDNA library with a probe containing both the trypsinogen I and II genes, four trypsinogenhybridizing plaques were obtained from a total of approximately 1×10^6 independent plaques. This makes the trypsinogen gene expression level in spleen close to 0.0004% of the total mRNA. These trypsinogen-containing cDNA clones from spleen were not characterized in detail due to their small insert sizes. However, based on PCR analyses with subfamily-specific primers, these clones all belong to the trypsinogen I subfamily.

An analysis of genomic chicken trypsinogen genes reveals they are encoded by five exons

The partial sequences of germline trypsinogen genes I and II are given in Figure 8 and aligned against their cDNA counterparts. Like the human and rat trypsingen genes, the chicken pretrypsinogens are also encoded by five separate exons. The trypsinogen I and II cDNAs are separated by five exons of 59 bp, 166 bp, 254 bp, 137 bp and 248 bp, and 61 bp, 163 bp, 254 bp, 137 bp and 250 bp respectively (Figure 8a). The differences in the length of exons between trypsinogens I and II are due to the length variation in the signal peptides, activation peptides (Figure 3) and 5' and 3' non-coding regions. Otherwise, the introns interrupt the coding regions in exactly the same position between trypsinogens I and II. Germline sequences for the exons and part of the introns (Figures 8b and 8c) are determined by direct sequencing from cosmid clones with primers listed in Figure 1. The three trypsinogen I subfamily sequences (one from genomic and two from cDNA) share more than 98% of their sequences (Figure 8b). The exon 1 for both trypsinogens I and II codes for the large portion of the signal peptide. The activation peptide and a small part of the signal peptide are encoded by exon 2. The intron/exon splice sites (--]GT.....AG]--) and their flanking sequences in the chicken trypsinogen genes are closely related to the consensus splicing signals (- - - JGTAAGT......CAGJ- - -) [30]

The chicken trypsinogen polyadenylation signals, like all the other eukaryotic genes, are composed of two relatively close spaced sequences. One is the virtually invariant AATAAA sequence and the other is a GT- or T-rich segment at 25 to 30 bp downstream (Figures 8b and 8c). The chicken trypsinogen polyA sequence, as expected, occurs at 12 to 15 bp downstream from the first polyadenylation signal (AATTAA for trypsin I and AATAAA for trypsin II) (Figure 2 and Figures 8b and 8c).

No obvious sequence similarity in the introns and the 5' noncoding regions between trypsinogens I and II can be found. A

Figure 8 Genomic organization (a) and the partial genomic sequences of the chicken trypsinogen I and II genes (b and c)

The exons are represented with hatched boxes. The size of introns are estimated by PCRs on cosmid DNA clones with the exon primers listed in Figure 1. The intron and exon sizes are listed under the corresponding positions (a). The genomic sequences for chicken trypsinogen I and II are listed in (b) and (c) respectively. The coding regions are in capital letters and the deduced amino acids are listed on top of the sequences. Identity to the germline sequence is indicated by dashes. Polyadenylation signals are underlined and the putative pancreatic enhancer sequences are listed in italic and underlined.

cis-acting pancreatic-specific serine-protease-enhancer-like sequence is located in the promoter region of trypsinogen I (TATCTGT...TTTCCA) and trypsinogen II (CAGCTGT... TGTCCA) (the canonical sequence for pancreatic serine protease is CACCTGT...TTTCCC). This suggests that the chicken may be using a similar set of *trans*-acting transcription factors to regulate the expression of trypsinogen genes.

DISCUSSION

The chicken trypsinogen gene family may contain as many as nine different members and can be divided into two subfamilies, a six-member trypsinogen I subfamily which encodes the cationic isoenzymes and a three-member trypsinogen II subfamily which encodes the anionic trypsinogen isoenzymes. Like the rat cationic and anionic trypsinogen genes [10], the chicken trypsinogen subfamilies do not cross-hybridize on genomic Southern blots (Figure 6).

The number of expressed chicken trypsinogen genes may be limited

Though multiple members for each trypsinogen subfamily have been identified from cosmid clones and genomic Southern hybridizations (Figure 2), the number of expressed trypsinogen genes in chicken is limited, since only two different trypsinogen I sequences and one trypsinogen II sequence were found among the 27 sequenced cDNA clones. This indicates that either there are a limited number of functional trypsinogen genes in the genome or some of the subfamily members are expressed at very low levels. No obvious pseudogenes can be found from the limited genomic and cDNA sequencing results presented here (Figures 8b and 8c).

Low-level expression of trypsinogen genes in tissues besides the pancreas can be observed by PCRs (Figure 7) and library screening with trypsinogen gene-specific primers or probes. This low-level and non-tissue-specific trypsinogen gene expression can also be observed in other species including human, mouse, *Xenopus* and fish (K. Wang, L. Gan and L. Hood, unpublished work). The human trypsinogen IV has been found specifically expressed in brain and pancreas recently [4]. Thus, the expression of trypsinogen genes in other tissues may play some unknown biological roles. However, the low expression level of trypsinogen genes in these tissues makes it difficult to determine its possible biological implications.

Trypsin is a highly conserved enzyme throughout evolution

The amino acid sequence alignment with 19 different trypsins from 10 different species (Figure 4) reveals 80 absolutely conserved positions which occupy approximately 35% of the entire enzyme sequence. These positions include the residues for the charge relay system, most of the disulphide bridges and residues needed to form the substrate-binding site. The alignment also reveals 24 highly variable positions (Figure 4). According to the results of chemical modifications and the trypsin three-dimensional structure, those highly variable positions are located on the exterior of the enzyme.

Chicken trypsin may be evolved in coincidental or concerted fashion

Sequence alignments of various proteins from different species provide valuable information regarding structures, functions and evolutionary relationships. It has been shown that within a limit error, the rate of change of amino acid sequence has remained essentially unchanged for each protein [31]. From the study of rat, pig and bovine trypsinogens, the evolution rate of trypsin has been calculated to be approximately 1% amino acid change for every 3 million years [7,9,11]. After the correction of multiple hits at the same location [31], the actual number of replacements per 100 residues for chicken trypsins from bovine trypsin 1 can be calculated to be 37.7 and 28.9 respectively. According to the trypsin evolution rate calculated by rat and pig trypsin, the divergence between mammal (bovine) and avian (chicken) molecules should have occurred approximately 90 to 100 (28.9 or 37.7 multiplied by 3) million years ago. However, fossil and other molecular evolutionary studies suggested that the avian and mammal diverged approximately 300 million years ago.

This contradiction may be explained by (1) the mutation rate for birds and mammals may be different according to their living environment (higher mutation rate for mammals); (2) the observed mutation rate may be greatly altered (limited) by the sites available. Since large portions (approximately 35%) of the sequences are required to be conserved in order to maintain proper function, birds may actually have a mutation rate close to mammals (due to the limited sites that can be changed, the observed mutation rate in avian seems much lower); (3) there are more trypsinogen genes in the genome and the sequences used in the comparison may not be the proper orthologous sequences, therefore, the observed mutation rates may not be used directly in evolutionary studies; and, most likely, (4) the observed mutations may be the result of multiple horizontal gene correction or fixation events (gene conversion). Therefore, the chicken trypsinogen gene family, like other multi-member gene families may have evolved in coincidental fashion [32,33].

It is interesting to note that the gene conversion mechanism has been used extensively in chicken immunoglobulin genes to generate its V gene diversity [34]. Thus, the chicken may possess a very active gene conversion machinery and the similarity among trypsinogen subfamily members (greater than 98 % in the available trypsinogen I subfamily sequences) is the result of extensive gene conversion events.

We thank the entire sequencing group for their help on the work of trypsinogen cDNA sequencing. This work was supported by DOE genome award to L.H.

REFERENCES

- 1 Kraut, J. (1977) Annu. Rev. Biochem. 46, 331-358
- 2 Emi, M., Nakamura, Y., Ogawa, M., Yamamoto, T., Nishide, T., Mori, T. and Matsubara, K. (1986) Gene 41, 305–310
- 3 Tani, T., Kawashima, I., Mita, K. and Takiguchi, Y. (1990) Nucleic Acids Res. 18, 1631
- 4 Wiegand, U., Corbach, S., Minn, A., Kang, J. and Muller-Hill, B. (1993) Gene 136, 167–175
- 5 Le Huerou, I., Wicker, C., Guilloteau, P., Toullec, R. and Puigserver, A. (1990) Eur. J. Biochem. 193, 767–773
- 6 Pinsky, S. D., LaForge, K. S. and Scheele, G. (1985) Mol. Cell. Biol. 5, 2669-2676
- 7 Hermodson, M. A., Ericsson, L. H., Neurath, H. and Walsh, K. A. (1973)
- Biochemistry 12, 3146-3153
- 8 Puigserver, A. and Louvard, M. N. (1974) Biochim. Biophys. Acta 371, 177-185
- 9 MacDonald, R. J., Stary, S. J. and Swift, G. H. (1982) J. Biol. Chem. **267**, 9724–9732
- Craik, C. S., Choo, Q. L., Swift, G. H., Quinto, C., MacDonald, R. J. and Rutter, W. J. (1984) J. Biol. Chem. **259**, 14255–14264
- 11 Fletcher, T. S., Alhadeff, M., Craik, C. S. and Largmen, C. (1987) Biochemistry 26, 3081–3086
- 12 Luetcke, H., Rausch, U., Vasiloudes, P., Scheele, G. A. and Kern, H. F. (1989) Nucleic Acids Res. 17, 6736
- 13 Stevenson, B. J., Hagenbuchle, O. and Wellauer, P. K. (1986) Nucleic Acids Res. 14, 8307–8330
- 14 Shi, Y. B. and Brown, D. D. (1990) Genes Dev. 4, 1107-1113
- 15 Titani, K., Ericsson, L. H., Neurath, H. and Walsh, K. A. (1975) Biochemistry 14, 1358–1366

- 16 Gudmundsdottir, A., Gudmundsdottir, E., Oskarsson, S., Bjarnason, J. B., Eakin, A. K. and Craik, C. S. (1993) Eur. J. Biochem. 217, 1091–1097
- 17 Meister, A., Weinrich, S. L., Nelson, C. and Rutter, W. J. (1989) J. Biol. Chem. 264, 20744–20751
- 18 Bell, G. I., Quinto, C., Quiroga, M., Valenzuela, P., Craik, C. S. and Rutter, W. J. (1984) J. Biol. Chem. **259**, 14265–14270
- 19 Davis, B. P., Hammer, R. E., Messing, A. and MacDonald, R. J. (1992) J. Biol. Chem. 267, 26070–26077
- 20 Maniatis, T., Fritsch, W. F. and Sambrook, J. (1989) Molecular Cloning: A Laboratory Manual. 2nd edn., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
- 21 Reed, K. C. and Mann, D. A. (1985) Nucleic Acids Res. 13, 7207-7221
- 22 Wang, K., Klotz, J. L., Kiser, G., Bristol, G., Hays, E., Lai, E., Gese, E., Kronenberg, M. and Hood, L. (1994) Genomics **20**, 419–428
- 23 Chomczynski, P. and Sacchi, N. (1987) Anal. Biochem. 162, 156-159

Received 5 October 1994; accepted 28 November 1994

- 24 Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, D. B. and Erlich, H. A. (1988) Science 239, 487–491
- 25 Feinberg, A. P. and Vogelstein, B. (1983) Anal. Biochem. 132, 6-13
- 26 Watson, M. E. E. (1984) Nucleic Acids Res. 12, 5145-5164
- 27 Walsh, K. A. and Neurath, H. (1964) Proc. Natl. Acad. Sci. U.S.A. 52, 884-889
- 28 Kauffman, D. L. (1965) J. Mol. Biol. 12, 929-932
- 29 Kozak, M. (1981) Nucleic Acids Res. 9, 5233-5253
- 30 lida, Y. (1990) J. Theor. Biol. 145, 523-533
- 31 Dickerson, R. E. (1971) J. Mol. Evol. 1, 26-45
- 32 Hood, L., Campbell, J. H. and Elgin, S. C. R. (1975) Annu. Rev. Genetics 9, 305-353
- 33 Zimmer, E. A., Martin, S. L., Beverley, S. M., Kan, Y. W. and Wilson, A. C. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 2158–2162
- 34 Carlson, L. M., McCormack, W. T., Postema, C. E., Humphries, E. H. and Thompson, C. B. (1990) Genes Dev. 4, 536–547