Biochem. J. (1995) 307, 471–479 (Printed in Great Britain)

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 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.

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, site-directed 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 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.

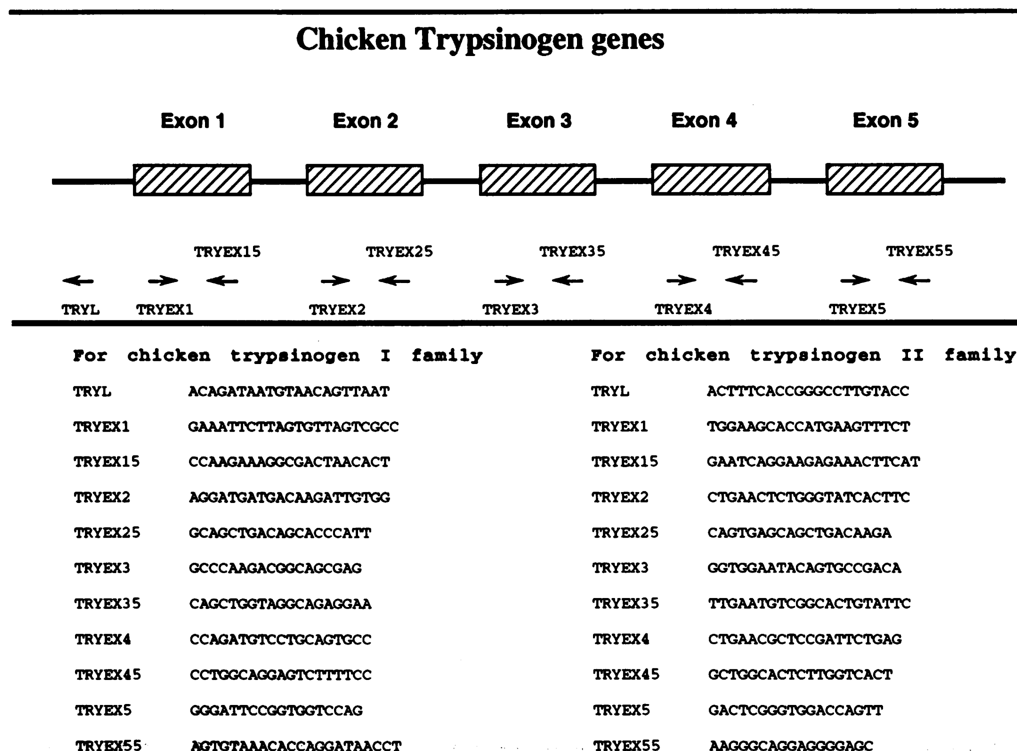


Figure 1 Diagrammatic summary of oligonucleotide primers used in chicken trypsinogen sequence analyses

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 μg) 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 (GenBank/EMBL release 74) (K. Wang, L. Gan and L. Hood, unpublished work). The trypsinogen-specific conserved region primers were as follows: TRYA 5′-TCCGGATCCTGATGACAAGATCGTTGGGGG; TRYB 5′-TCCGGATCCTTCTGTGGAGGCTCCCTCAT; TRYC 5′-TCCGGATCCATAGCCCCAGGAGAC; and TRYD 5′-TCCGGATCCTTGGTGTAGACACCAGG. 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 μl PCR reactions [24] that consisted of 50 mM NaCl, 20 mM Tris/HCl (pH 9.0), 2.5 mM $MgCl_2$, 0.5 mM dithiothreitol (DTT), 150 μM of each nucleotide, 1 μM 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 [γ-^{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 × SSC (1 × SSC = 0.15 M NaCl/0.15 M sodium citrate), 0.02 M sodium phosphate (pH 6.7), 100 μg/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 × SSC with 0.1 % SDS at room temperature and twice in 0.5 × 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.).

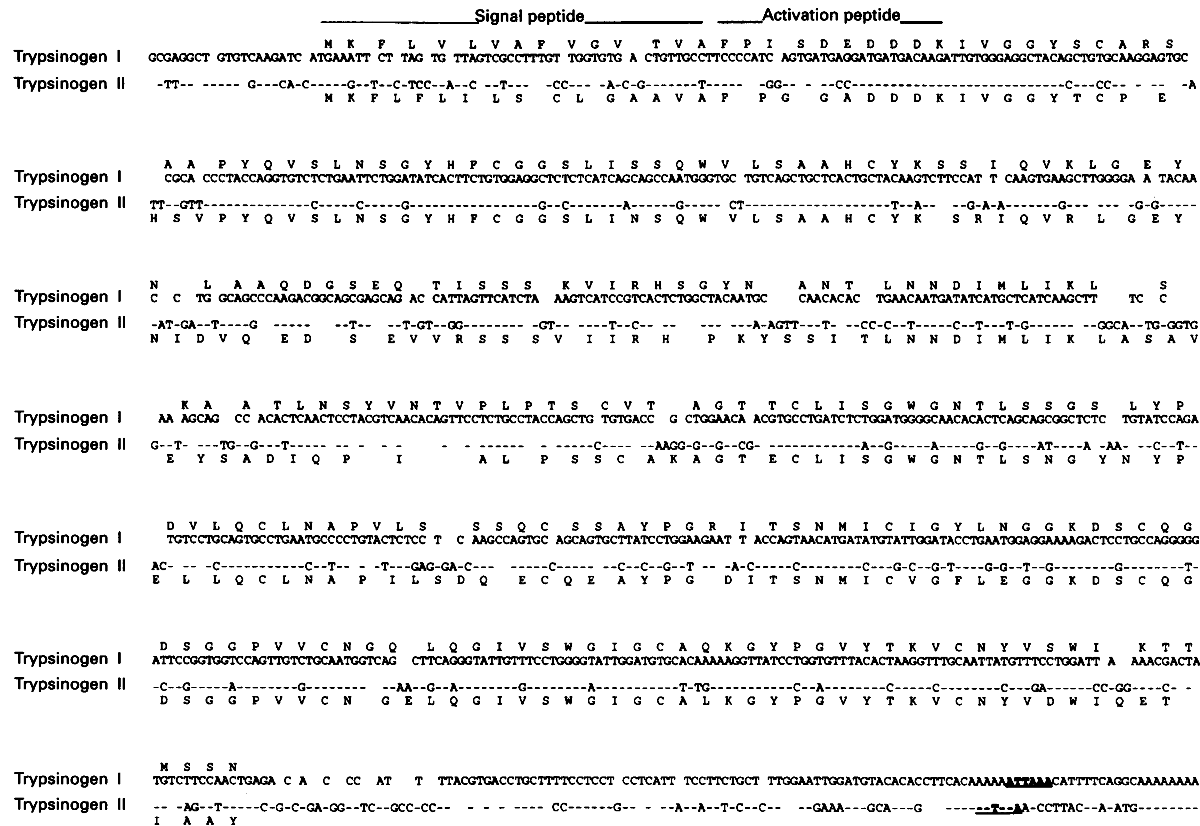


Figure 2 Sequence alignment between chicken trypsinogen I and trypsinogen II subfamilies

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 μl 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 μl) 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 (DNA-star, 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
HUMAN-TRY1		Met Asn Pro Leu Leu Ile Leu Thr Phe Val Ala Ala Ala Leu Ala Ala	Pro Phe Asp Asp Asp Asp Lys	Ile Val Gly Gly
HUMAN-TRY2		Met Asn Leu Leu Leu Ile Leu Thr Phe Val Ala Ala Ala Val Ala Ala	Pro Phe Asp Asp Asp Asp Lys	Ile Val Gly Gly
HUMAN-TRY3		Met Asn Pro Phe Leu Ile Leu Ala Phe Val Gly Ala Ala Val Ala	Val Pro Phe Asp Asp Asp Asp Lys	Ile Val Gly Gly
BOVINE-TRY2	(A)	Met His Pro Leu Leu Ile Leu Ala Phe Val Gly Ala Ala Val Ala	Phe Pro Ser Asp Asp Asp Asp Lys	Ile Val Gly Gly
DOG-TRY1	(C)	Met 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	(A)	Met Asn Pro Leu Leu Ile Leu Ala Phe Leu Gly Ala Ala Val Ala	Thr Pro Thr Asp Asp Asp Asp Lys	Ile Val Gly Gly
MOUSE-TRY		Met Ser Ala Leu Leu Ile Leu Ala Leu Val Gly Ala Ala Val Ala	Phe Pro Val Asp Asp Asp Asp Lys	Ile Val Gly Gly
RAT-TRY1	(A)	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
RAT-TRY2	(A)	Met Arg Ala Leu Leu Ile Leu Ala Leu Val Gly Ala Ala Val Ala	Phe Pro Val Asp Asp Asp Asp Lys	Ile Val Gly Gly
RAT-TRY3	(C)	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
RAT-TRY4	(C)	Met Lys Ile Ser Ile Phe Phe Ala Phe Leu Gly Ala Ala Val Ala	Leu Pro Val Asn Asp Asp Asp Lys	Ile Val Gly Gly
CHICKEN-TRY I	(C)	Met Lys Phe Leu Val Leu Val Ala Phe Val Gly Val Thr Val Ala	Phe Pro Ile Ser Asp Glu Asp Asp Asp Lys	Ile Val Gly Gly
CHICKEN-TRY II	(A)	Met Lys Phe Leu Phe Leu Ile Leu Ser Cys Leu Gly Ala Ala Val Ala	Phe Pro Gly Gly Ala Asp Asp Asp Lys	Ile Val Gly Gly
XENOPUS-TRY		Met Lys Phe Leu Leu Leu Cys Val Leu Leu Gly Ala Ala Ala Ala	Phe Asp Asp Asp Lys	Ile Ile Gly Gly
ATLANTIC COD-TRY		Met Lys Ser Leu Ile Phe Val Leu Leu Gly Ala Val	Phe Ala Glu Glu Asp Lys	Ile Val Gly Gly

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].

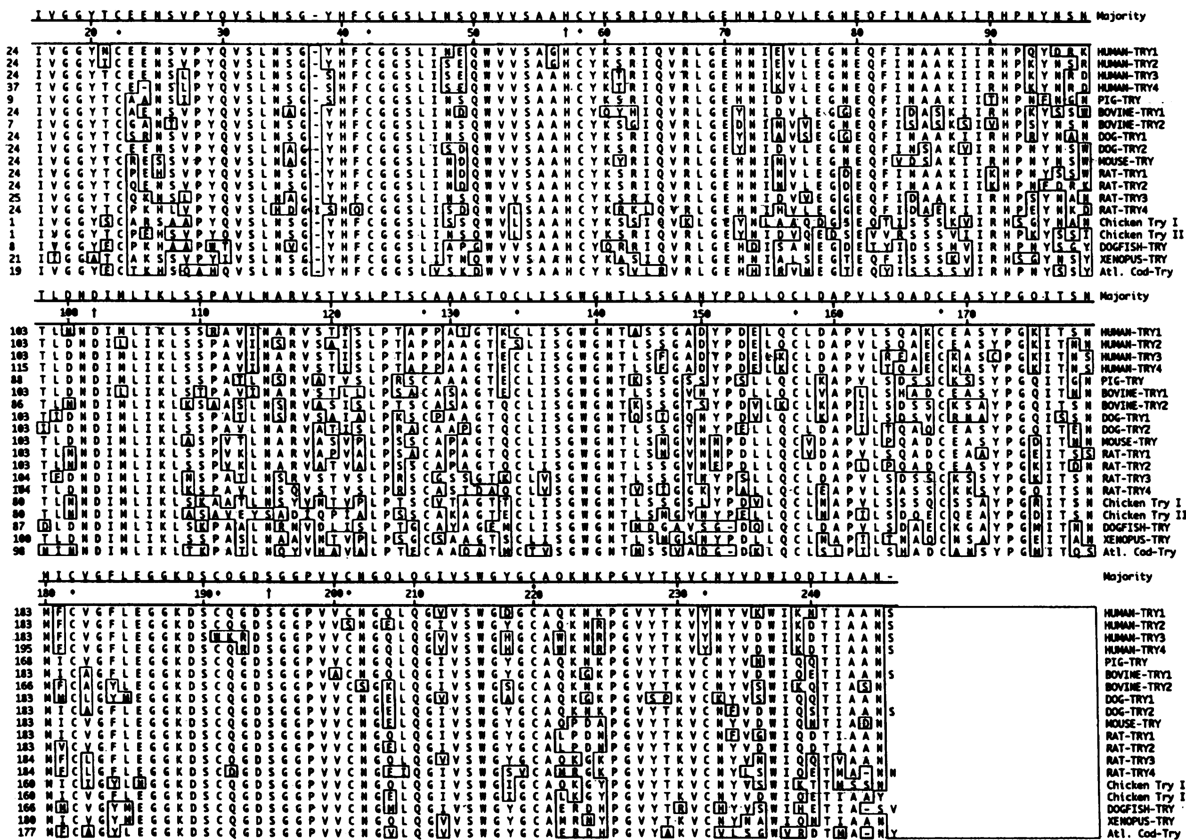


Figure 4 Trypsin sequence alignment from various species

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), Dogfish-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.

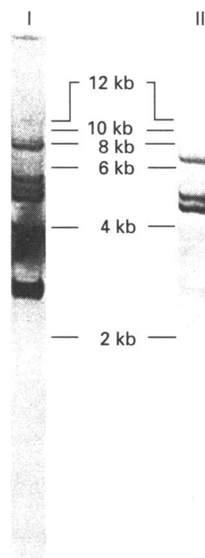


Figure 6 Genomic Southern blot analysis of the chicken trypsinogen gene family

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 pre-proenzymes which contain an N-terminal signal peptide (pre-trypsinogen) 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 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.

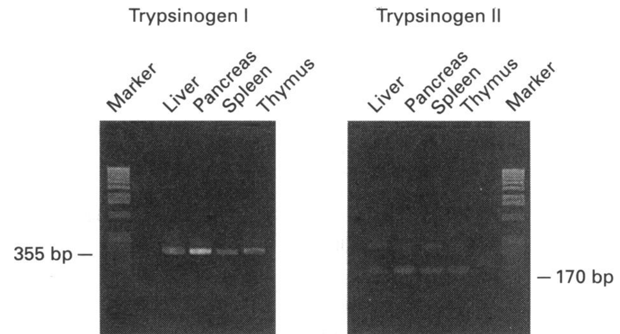


Figure 7 Amplification of trypsinogen genes with cDNA extracted from various chicken tissues

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.).

(i) Signal peptides

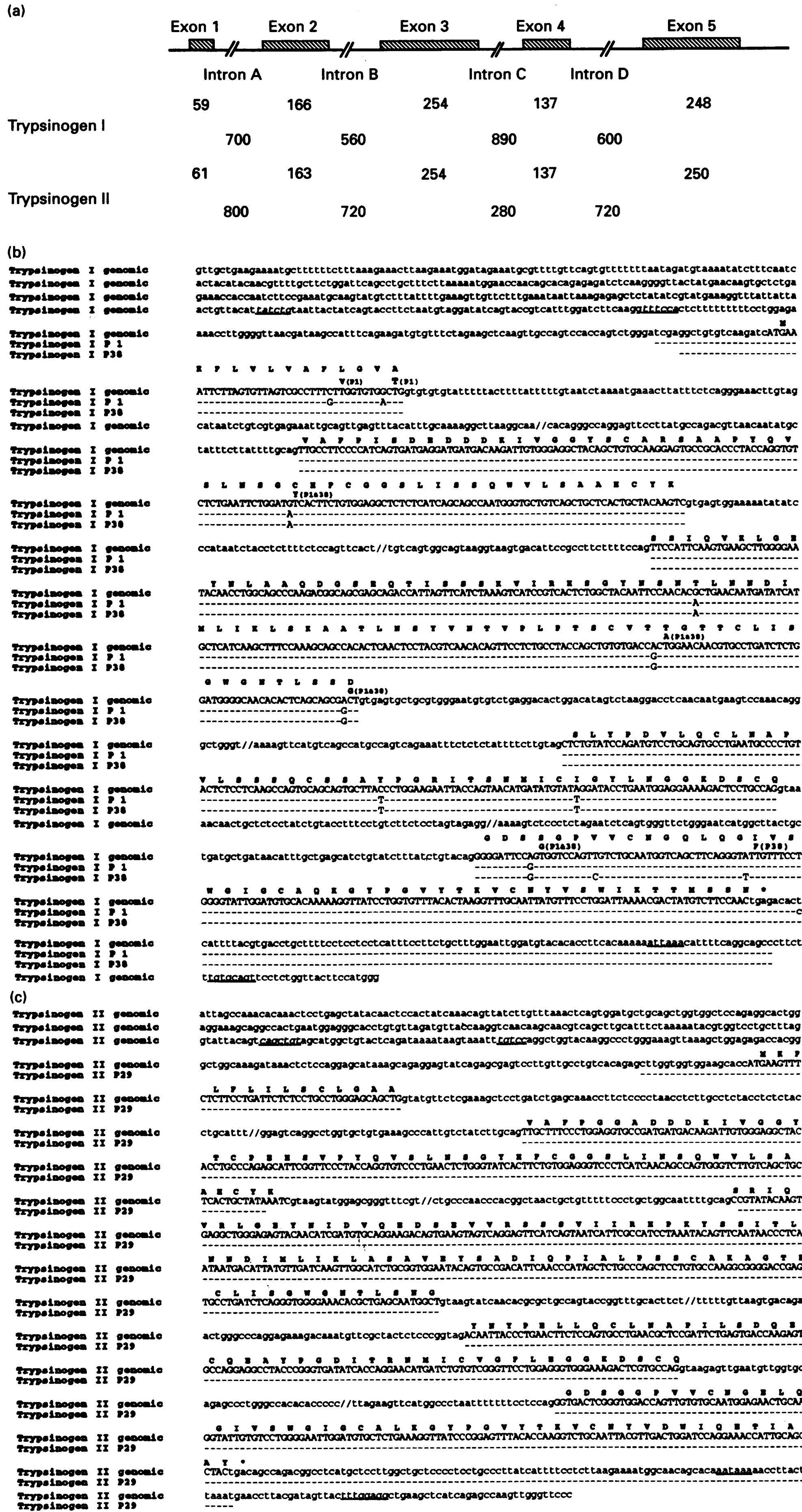
The chicken signal peptides for trypsinogens I and II are 15 to 16 amino acid residues in length, which is similar to that of rat, bovine, dog and human pre-trypsinogen signal peptides (Figure 3). The sequences of signal peptides follow the usual pattern of a charged residue (Lys at second amino acid 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 eight-amino-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 1
Exon 2
Exon 3
Exon 4
Exon 5
Intron A
Intron B
Intron C
Intron D
Trypsinogen I
59
166
254
137
248
700
560
890
600
Trypsinogen II
61
163
254
137
250
800
720
280
720
(b)
(c)

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

(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).

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′ non-coding 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 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 trypsinogen-hybridizing 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 trypsinogen genes, the chicken pre-trypsinogens 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 (- - -↓GTAAGT......CAG↓- - -) [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′ non-coding 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. **257**, 9724–9732
10 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
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 Iida, 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

Received 5 October 1994; accepted 28 November 1994