J. Biochem. 120, 647-656 (1996)

The Primary Structure of the Major Pepsinogen from the Gastric Mucosa of Tuna Stomach[1]

Masao Tanji,[*,2] Etsuko Yakabe,[†,2] Takashi Kageyama,[‡] and Kenji Takahashi[*,§,3]

[*]Department of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, Bunkyo-ku, Tokyo 113; [†]Department of Chemistry, Faculty of Science and Technology, Aoyama Gakuin University, Setagaya-ku, Tokyo 157; [‡]Department of Cellular and Molecular Biology, Primate Research Institute, Kyoto University, Inuyama, Aichi 484; and [§]School of Life Science, Tokyo University of Pharmacy and Life Science, 1432-1 Horinouchi, Hachioji, Tokyo 192-03

Received for publication, May 14, 1996

The complete primary structure of the major component of tuna pepsinogens was determined by conventional protein chemistry methods. It was composed of a prosegment of 37 residues and a pepsin moiety of 323 residues, having a relative molecular mass of 39,364. The essential aspartyl residues in the active site and the three disulfide bonds common to other pepsinogens were conserved; however, several unique substitutions and/or deletions characteristic of tuna pepsinogen were found at various positions, especially in the prosegment and subsite regions, as compared with the sequences of other pepsinogens, which may affect the rate of activation of the zymogen, and/or the catalytic function and substrate specificity of the enzyme. Tuna pepsinogen is the least acidic among pepsinogens. The sequence identity between tuna pepsinogen and other pepsinogens ranged from 45 to 52%. A phylogenetic tree based on the primary structures suggested that tuna pepsinogen diverged from the pepsinogen A and prochymosin groups in an early period of pepsinogen evolution.

Key words: amino acid sequence, pepsin, pepsinogen, phylogenetic tree, tuna.

Pepsinogens are found in the gastric mucosae of all classes of vertebrates. They are typical aspartic proteinases with two essential aspartic acid residues in the active site. They are classified into three major groups, namely, pepsinogen A and pepsinogen C (progastricsin) in adult animals, and prochymosin in neonates, based on the differences in their enzymatic and immunochemical properties (*1, 2*). The primary structures are considerably different among these three groups. So far, the complete primary structures of porcine (*3-6*), chicken (*7*), and monkey (*8*) pepsinogens A, monkey pepsinogen C (*9*), and bovine prochymosin (*10*) have been determined by conventional protein sequencing. The structures of human (*11*) and rabbit (*12*) pepsinogens A, human (*13, 14*), rat (*15, 16*), guinea pig (*17*), and frog (*18*) pepsinogens C, and chicken embryonic pepsinogen (*19*) have been deduced from the results of base sequence analysis of their genomic DNAs and/or cDNAs. All these sequences except for those of chicken and frog represent those of mammalian pepsinogens. Therefore, our knowledge about the sequences of pepsinogens from non-mammalian species is rather scarce, and it is desirable to elucidate further those of non-mammalian pepsinogens to shed light on the molecular evolution and structure/function relationships of pepsinogens. In this respect, fish pepsinogens are drawing special attention since they are the lowest animals that have gastric pepsinogens.

Previously, we reported the purification and characterization of three pepsinogens from tuna (North Pacific bluefin tuna, *Thunnus thynuus orientalis*) (*20*). They are the smallest pepsinogens so far known, and along with some unique enzymatic properties, have NH_2-terminal structures that are fairly different from those of other animal pepsinogens. In the present paper, we report the complete primary structure of the major tuna pepsinogen, and discuss the relationship beween its structure and some enzymatic properties and molecular evolution.

EXPERIMENTAL PROCEDURES

Materials—Pepsinogens were purified from North Pacific bluefin tuna (*T. thynuus orientalis*) as described previously (*20*), the major zymogen, pepsinogen 2, being examined in the present study. Trypsin (tosyl-L-phenylalanylchloromethyl ketone-treated) was purchased from Sigma Chem., α-chymotrypsin from Worthington Biochem., thermolysin from Daiwa Kasei, lysylendopeptidase from Wako Pure Chem., endoproteinase Asp-N from Boehringer Mannheim, GmbH, and *Staphylococcus aureus* V8 protease from Miles Laboratories. Cyanogen bromide was purchased from Nacalai Tesque, and nitrophenylsulfenyl bromoskatole from Pierce. Gel filtration columns,

[1] This study was supported in part by Grants-in-Aid for Scientific Research from the Ministry of Education, Science, Sports and Culture of Japan.

[2] Present address: Department of Neurobiology and Anatomy, University of Texas Medical School at Houston, Houston, Texas 77030, USA.

[3] To whom correspondence should be addressed at: School of Life Science, Tokyo University of Pharmacy and Life Science, 1432-1 Horinouchi, Hachioji, Tokyo 192-03. Tel: +81-426-7146, Fax: +81-426-7149, E-mail:kenjitak@ls.toyaku.ac.jp.

Abbreviation: HPLC, high performance liquid chromatography.

Superose 12 (1.6×50 cm, preparative grade) and Superose 12 HR10/30, were obtained from Pharmacia LKB Biotech., and reverse-phase high performance liquid chromatography (HPLC) columns, TSKgel ODS-120T (0.4×25 cm) and C8 AG120 (4.6×250 mm), were from Tosoh, and Shiseido, respectively. Reagents for amino acid analysis and manual amino acid sequencing were obtained from Wako Pure Chem. Other chemicals were of reagent grade and used without further purification.

Peptide Nomenclature—Peptides are designated by three letters and subsequent numbers. The peptides derived from heat-denatured and reduced-carboxymethylated samples of pepsinogen are designated by the first letters of "H" and "R," respectively. The second letters: B, Sk, T, C, Th, Ly, V, and E, indicate peptides derived on treatment with cyanogen bromide, BNPS-skatole, trypsin, α-chymotrypsin, thermolysin, lysyl-endopeptidase, *S. aureus* V8 protease, and endoproteinase Asp-N, respectively. The third letters: S, Pa, and Pb, indicate peptides soluble in 10, 10-50, and 50-70% aqueous formic acid, respectively. Insoluble peptides are indicated by the letter, P. Arabic numerals following the three letters show the order of elution on gel filtration and/or HPLC. Since additional fragmentation was carried out for some peptides, the same nomenclature was adopted for each subfragmented peptide. Arabic numerals in parentheses indicate peptides obtained in the same fraction, which were sequenced directly as a mixture.

Fragmentation of Proteins and Purification of Peptides—In the course of purification, the peptides were monitored by measuring the absorbance at 280, 259, 230, or 215 nm, or determined by the fluorometric method using fluorescamine according to de Bernardo *et al.* (*21*) with leucine as a standard. Unless otherwise specified, gel filtration was performed on a Superose 12 HR10/30 column, and HPLC on a TSKgel ODS-120T column. On HPLC, peptides were eluted with a linear gradient of acetonitrile, from 0 to 50%, in 0.1% trifluoroacetic acid at the flow rate of 0.4 ml/min, unless otherwise specified.

Cyanogen bromide cleavage: Reduced and carboxymethylated pepsinogen (10 mg, 260 nmol), prepared essentially according to Crestfield *et al.* (*22*), was treated at 37°C for 24 h in 0.8 ml of 70% formic acid containing 8.5 mg cyanogen bromide (an about 300-fold molar excess over protein). The reaction mixture was diluted with distilled water and then lyophilized. The lyophilized material was suspended in 6% formic acid. The soluble peptides (RBS) were purified by gel filtration on a Superose 12 column (1.6×50 cm) equilibrated with 6% formic acid (Fig. 1a) and subsequent HPLC (Fig. 1, b and c). The insoluble peptides (RBP) were purified by gel filtration on a Superose 12 column (1.6×50 cm) equilibrated with 0.2 M ammonium acetate buffer, pH 10 (Fig. 1d) and subsequent HPLC on a Shiseido C8 AG120 column eluted with a linear gradient of acetonitrile from 0 to 50% in the same buffer (Fig. 1, e and f). In addition, HTPb1 and HTPa-8 were subjected to cyanogen bromide cleavage in a similar manner and the cleavage mixture of each peptide was directly sequenced.

Tryptic digestion: Heat-denatured pepsinogen (7.9 mg, 200 nmol) was digested with 25 μg (1 nmol) of trypsin in 2.0 ml of 0.05 M ammonium bicarbonate, pH 7.8, containing 0.01 M $CaCl_2$, at 37°C for 14 h. The peptides were separated into three fractions according to the solubility in a formic acid solution, namely HTS, HTPa, and HTPb. The peptides in a quarter of the HTS fraction were purified by HPLC (Fig. 1g). Purification of some peptides was achieved by HPLC (Fig. 1h) after gel filtration using a quarter of the HTS fraction. Peptides in half of the HTPa fraction were purified by HPLC (Fig. 1i) and subsequent gel filtration in 50% formic acid. HTPb1 and HTPb2 were obtained on gel filtration in 75% formic acid (Fig. 1j). They were subjected to further cleavage as follows.

HTPb1 (4 nmol) was cleaved with cyanogen bromide and the resulting peptides were purified by gel filtration in 75% formic acid. One major peptide, HTPb1-BP6, was obtained.

HTPb1 (5 nmol) was dissolved in 22 μl of 70% acetic acid containing 187 μg (about 500 nmol) of nitrophenylsulfenyl bromoskatole and then left for 42 h in the dark at room temperature. The reaction was stopped by adding 200 μl of distilled water. One major peptide, HTPb1-SkS5, was obtained on gel filtration in 70% formic acid. HTPb1 (10 nmol) was digested with 2 μg endoproteinase Asp-N in 50 μl of 0.1 M sodium phosphate buffer, pH 8.0, at 37°C for 20 h. One major peptide, HTPb1-EP2, was obtained on gel filtration in 75% formic acid. This peptide was digested with *S. aureus* V8 protease and then the mixture was directly sequenced with an automatic sequencer.

HTPb2 (10 nmol) was digested with 3 μg (0.5 nmol) *S. aureus* V8 protease in 30 μl 0.1 M ammonium bicarbonate, pH 7.8, containing 2 mM EDTA at 37°C for 20 h. The reaction mixture was lyophilized and then directly sequenced with an automatic sequencer.

Chymotryptic digestion: Heat-denatured pepsinogen (4 mg, 100 nmol) was digested with 50 μg (about 2 nmol) of α-chymotrypsin in 1 ml of 50 mM sodium phosphate buffer, pH 6.5, at 37°C for 2 h. The digest was gel-filtered on a Superose 12 column (1.6×50 cm) in 5% acetic acid (Fig. 1k). Four fractions, HCS1, HCS2, HCS3, and HCS4, were pooled, and the peptides contained in each fraction were purified by HPLC (Fig. 1l).

Thermolysin digestion: Heat-denatured pepsinogen (4 mg, 100 nmol) was digested with 200 μg (about 5.8 nmol) of thermolysin in 2.0 ml of 0.1 M Tris-HCl buffer, pH 6.7, containing 10 mM $CaCl_2$ at 37°C for 9 h. The peptides soluble in 5% acetic acid (HThS) were purified by gel filtration (Fig. 1m) and subsequent HPLC (Fig. 1n).

Lysylendopeptidase digestion: Heat-denatured pepsinogen (1.6 mg, 40 nmol) was digested with 60 μg (about 2 nmol) of lysylendopeptidase in 400 μl of 0.1 M ammonium bicarbonate, pH 8.2, at 37°C for 4 h. The resulting peptides were fractionated according to their solubility in formic acid. The peptides soluble in 10-50% formic acid (HLyP) were gel-filtered in 50% formic acid. One major peptide, HLyP14, was obtained.

Amino Acid Analysis—Samples for analysis were hydrolyzed with 100 μl of 6 N HCl containing 0.05% phenol at 110°C for 24 h in evacuated sealed tubes. The amino acids were determined with a Hitachi model 835 amino acid analyzer essentially according to Spackman *et al.* (*23*).

Amino Acid Sequence Determination—The amino acid sequences were determined by a modification (*24*) of the manual Edman degradation method (*25*) or by using an automatic sequencer (Applied Biosystems Model 477A/120A). On manual sequencing, the conversion of thiazolinone derivatives to thiohydantoin ones was carried out in methanol/acetyl chloride (7 : 1, v/v) at 50°C for 15 min,

TABLE I. **Amino acid compositions of acid hydrolysates of peptides (residues/molecule).** The values in parentheses are those obtained on sequencing.

Amino acid	RBS1-24	RBS2-2	RBS2-9	RBP2-5	RBP3-10	HTS-20	HTS-21	HTS-22	HTS-29	HTS-31
Lys	0.9 (1)		2.9 (4)			1.0 (1)	1.7 (2)	1.0 (1)		
His			1.2 (1)		0.5 (1)	0.8 (1)				
Arg				0.5 (1)						0.7 (1)
Asx	4.9 (5)		11.1 (8)	3.8 (5)	1.8 (2)		0.8 (1)	0.8 (1)		1.0 (1)
Thr			4.9 (5)	2.5 (4)	2.4 (4)		1.0 (1)	0.8 (1)		
Ser	1.6 (2)		11.2 (12)	5.5 (4)	8.4 (10)		1.7 (2)	1.6 (2)		
Glx	1.4 (1)		10.0 (7)	4.8 (2)	7.4 (8)	1.3 (1)	2.2 (2)	2.0 (2)	1.1 (1)	1.1 (1)
Pro	nd (1)		5.0 (5)	nd (1)	nd (2)	nd (2)	nd (1)	nd (1)		
Gly	2.3 (2)		8.6 (6)	5.0 (5)	6.1 (5)				1.0 (1)	1.0 (1)
Ala	3.1 (3)	1.1 (1)	3.5 (2)	2.2 (2)	1.7 (1)	0.8 (1)			1.2 (2)	
Cys			nd (2)	nd (2)						
Val	3.0 (3)		4.0 (5)	3.1 (3)	6.1 (5)				0.9 (1)	
Met	nd (1)	nd (1)	nd (1)	nd (1)	nd (1)	nd (1)				
Ile	2.0 (2)		2.7 (3)	3.2 (5)	2.8 (4)				0.9 (1)	
Leu	2.0 (2)		3.8 (3)	2.8 (1)	6.4 (4)				1.5 (1)	1.1 (1)
Tyr		0.9 (1)	2.2 (4)	0.2 (1)	nd (3)	0.8 (2)				0.7 (1)
Phe	1.7 (2)		3.9 (4)		2.8 (2)		2.0 (2)	2.0 (2)		
Trp			(2)		(2)					(1)
Position	117-141	114-116	9-82	196-232	142-195	28P-36P	57-67	58-67	317-323	20P-26P
Yield (%)	1	2	3	5	7	39	18	23	60	24

Amino acid	HTS-32	HTS-40	HTS2-34	HTPa-5	HTPa-8	HTPa-9[a]	HTPa-10	HTPb1	HTPb2
Lys		0.9 (1)		1.0 (1)		2.0 (1,0,0)			0.6 (1)
His				0.7 (1)	0.8 (1)	0.8 (0,1,0)		1.0 (1)	
Arg	1.0 (1)		0.3 (1)		0.9 (1)	2.9 (0,1,1)	1.8 (2)	0.7 (1)	
Asx	1.3 (1)	4.0 (4)	1.0 (1)	3.4 (3)	7.3 (8)	13.0 (3,8,2)	4.0 (3)	10.0 (10)	3.4 (3)
Thr	1.8 (2)	2.5 (3)	0.7 (1)	1.3 (1)	5.0 (6)	6.9 (0,6,1)	2.3 (2)	8.4 (8)	
Ser	2.1 (2)	3.2 (3)	1.4 (1)	4.1 (5)	4.6 (5)	8.2 (4,5,0)	1.9 (1)	13.8 (16)	2.8 (4)
Glx	3.0 (2)	3.5 (3)	2.1 (4)	2.6 (2)	6.0 (5)	13.2 (4,5,3)	7.2 (6)	15.0 (11)	3.5 (4)
Pro	nd (1)	nd (3)		nd (1)	nd (2)	nd (1,2,0)		nd (4)	nd (1)
Gly	3.3 (4)	3.4 (3)	1.5 (2)	2.0 (1)	4.2 (4)	11.5 (3,4,5)	6.5 (6)	13.1 (12)	3.7 (3)
Ala		1.5 (1)	0.8 (2)	1.5 (1)	5.3 (6)	11.7 (3,6,0)		6.2 (6)	2.7 (3)
Cys				nd (2)	nd (2)	nd (0,2,0)		nd (2)	
Val		2.0 (2)	0.4 (1)	3.0 (3)	3.9 (4)	8.8 (4,4,1)	1.9 (1)	11.6 (11)	3.1 (4)
Met	0.1 (1)	nd (1)			1.2 (2)	nd (2,2,0)		0.2 (2)	1.2 (2)
Ile	1.1 (1)	1.8 (1)	1.0 (2)	1.1 (1)	2.8 (3)	7.5 (3,3,2)	3.1 (3)	10.9 (10)	2.4 (3)
Leu	0.9 (1)	2.1 (1)	0.6 (1)	1.2 (1)	1.2 (1)	5.7 (1,1,2)	2.0 (2)	7.6 (7)	1.2 (1)
Tyr	1.1 (1)	2.5 (3)	0.7 (2)	0.6 (1)	3.3 (4)	3.7 (1,4,0)	1.6 (2)	1.0 (3)	0.8 (1)
Phe		2.2 (2)	0.7 (1)	0.7 (1)	2.2 (2)	5.6 (2,2,2)	3.0 (3)	5.5 (4)	2.0 (2)
Trp	(1)			(1)	(1)	(0,1,1)	(1)	(2)	
Position	68-85	37P-30	305-323	31-56	228-284	86-117 228-284 285-304	285-316	118-227	86-117
Yield (%)	72	3	2	4	28	3 4 5	17	7	3

Amino acid	HTPb1-BP6	HCS1-9[b]	HCS1-12	HCS1-13	HCS1-15	HCS1-18	HCS1-20[b]	HCS1-22[b]	HThS3-13
Lys		2.6 (1,2)						0.9 (1,0)	1.0 (1)
His	nd (1)	0.9 (1,1)							
Arg			nd (1)			0.7 (1)	1.0 (0,1)		
Asx	3.7 (2)	2.0 (0,2)	1.3 (1)		3.3 (4)	1.4 (1)	6.6 (4,1)	1.8 (1,0)	1.1 (1)
Thr	2.2 (4)	1.0 (0,1)			1.5 (2)	1.4 (1)	5.2 (4,1)	3.0 (0,2)	
Ser	8.1 (10)	3.6 (0,4)	1.0 (1)	1.6 (2)	1.2 (1)	1.0 (1)	2.6 (2,0)		
Glx	5.2 (8)	5.4 (0,4)		2.1 (2)	2.1 (2)	2.5 (2)	7.8 (4,3)	1.6 (0,1)	1.2 (1)
Pro	nd (2)	nd (1,1)	nd (1)	nd (1)	nd (1)		nd (1,0)	nd (0,1)	nd (1)
Gly	9.9 (5)		nd (1)	1.0 (1)	1.3 (1)	2.0 (2)	7.0 (2,5)	3.0 (2,1)	0.9 (1)
Ala	1.6 (1)	2.4 (1,1)		1.2 (1)	1.0 (1)	1.4 (1)	2.2 (2,0)	2.3 (2,0)	
Cys		nd (0,2)					nd (1,0)		
Val	5.0 (5)		nd (1)				8.0 (3,0)		
Met	nd (1)	nd (1,0)			nd (1)		nd (1,0)		
Ile	1.7 (4)		2.0 (2)	1.0 (1)		1.4 (2)	2.4 (1,0)	2.5 (1,2)	
Leu	3.4 (4)				1.3 (1)	0.7 (1)	2.2 (0,1)	2.4 (2,1)	
Tyr	nd (3)		nd (1)		1.3 (2)				
Phe	nd (2)	1.7 (1,2)		1.0 (1)		0.7 (1)	2.0 (1,1)	nd (1,0)	0.7 (1)
Trp	(2)							(0,1)	(1)
Position	142-195	P32-P37 47-66	222-230	104-112	2-17	308-320	236-261 282-296	117-126 177-185	66-72
Yield (%)	1	5 30	8	23	17	6	1 1	1 2	17

[a]A mixture containing three major peptides. [b]A mixture containing two major peptides. nd, not determined.

essentially according to Tarr (*26*), and phenylthiohydantoin derivatives of amino acids were identified by HPLC on a Microsorb C18 (4×250 mm) column according to Tsunasawa *et al.* (*27*). Throughout this paper, residues are numbered according to the numbering of porcine pepsinogen (*1*) and pepsin (*3*), except for the residues in Table

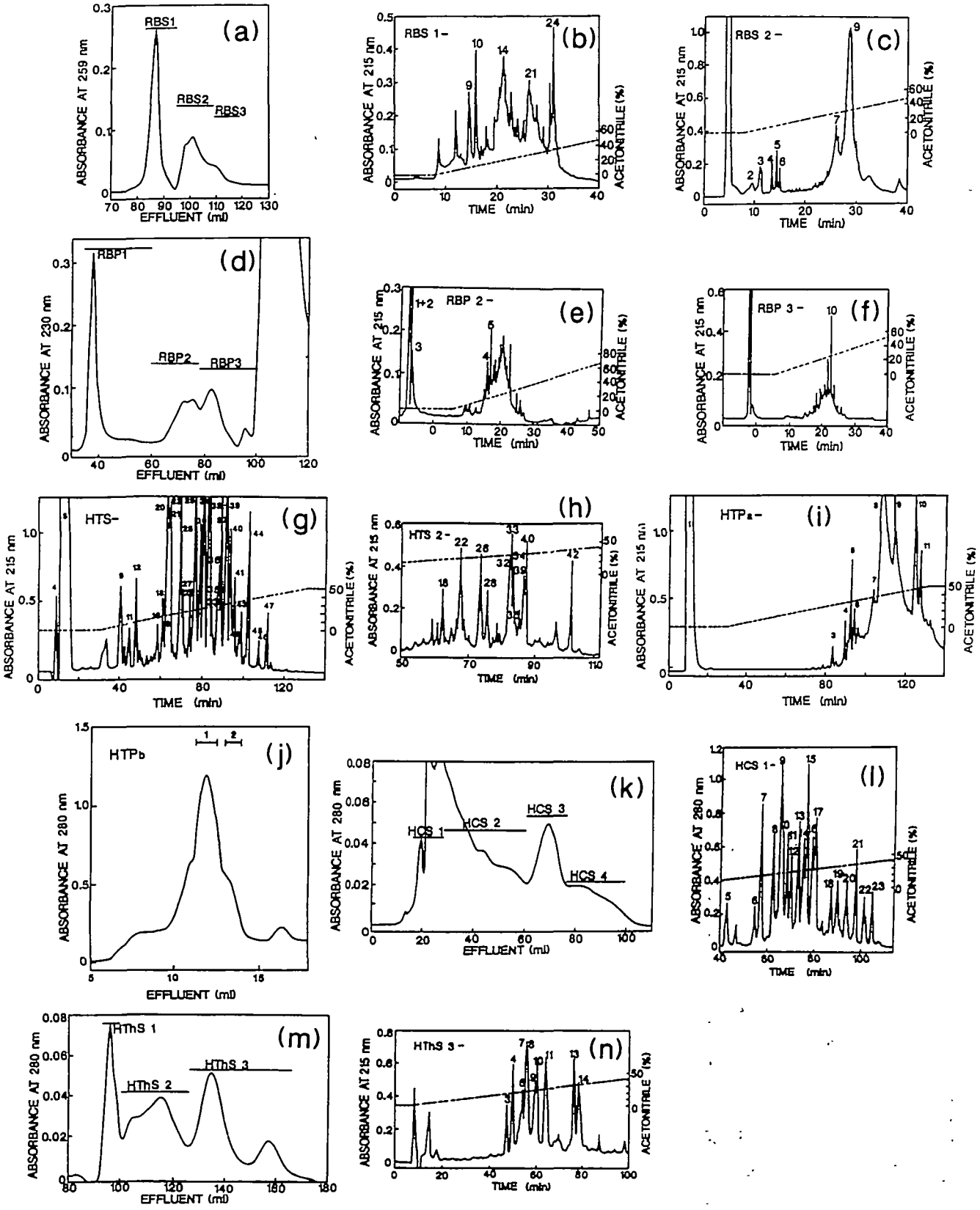

Fig. 1. **Fractionation of peptides derived from heat-denatured and reduced-carboxymethylated tuna pepsinogen 2 on enzymatic and chemical cleavage.** A minimum number of chromatograms are shown. In chromatograms a, d, j, k, and m, the fractions under the bars were pooled. For details, see under "EXPERIMENTAL PROCEDURES."

I and Fig. 2, in which the numbering of tuna pepsinogen and pepsin is used.

RESULTS

Figure 1 and Table I present a minimum amount of data on the fractionation and amino acid compositions, respectively, of peptides obtained from heat-denatured or reduced-carboxymethylated tuna pepsinogen 2 on enzymatic and/or chemical cleavage. From a tryptic digest of heat-denatured pepsinogen, 18 major peptides were isolated and completely or partially sequenced. These peptides accounted almost fully for the amino acid composition of the original protein (*20*). Tryptic peptides HTPb1 (residues 118-227, Fig. 2) and HTPa8 (residues 228-284, Fig. 2) were further fragmented by cyanogen bromide cleavage and/or by digestion with endopeptidase Asp-N and V8 protease, and the fragments were separated and sequenced. In addition, 29 chymotryptic peptides, 1 lysylendopeptidase peptide, and 27 thermolysin peptides derived from heat-denatured pepsinogen, and 13 cyanogen bromide peptides from RCM-pepsinogen were isolated and sequenced completely or partially. The NH_2-terminal 72-residue sequence had been determined previously (*20*), which was confirmed in the present study. In these analyses, the residues in the pepsin sequence were identified twice or more times by Edman degradation except for residues 186-195 and 214-221 (Fig. 2), which were sequenced only once.

The peptides other than the trypsin series contained a number of peptides overlapping the tryptic peptides, and thus the tryptic peptides could be aligned unequivocally. The locations of the disulfide bonds, Cys47-Cys52, Cys207-Cys211, and Cys250-Cys283, were deduced from the result that only three different cystine-containing peptides, HTPa6(or 5), HTPb1, and HTPa8, were isolated from the tryptic digest of heat-denatured pepsinogen. The positions of cystine residues in the second and third peptides were deduced only from the amino acid compositions and sequence homology with other pepsinogens, since no phenylthiohydantoin derivative of an amino acid was obtained for these positions. These results led to deduction of the complete primary structure of tuna pepsinogen (Fig. 2). Tuna pepsinogen is composed of 360 amino acid residues, comprising a 37-residue propeptide and a 323-residue pepsin moiety with the three disulfide bonds. It has a calculated molecular weight of 39,364 and the following

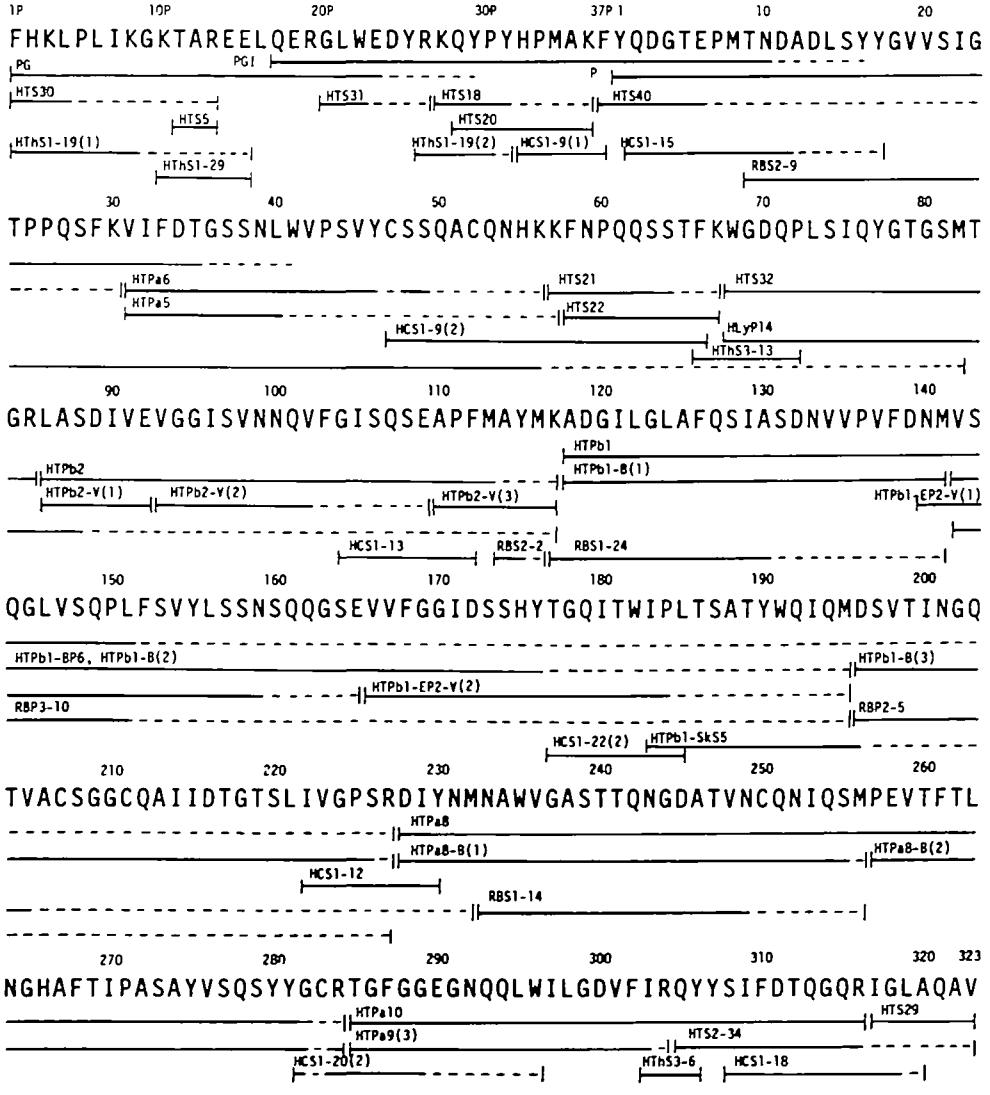

Fig. 2. **Alignment of peptides.** All the major tryptic peptides obtained from heat-denatured pepsinogen are shown. As for other peptides, only a minimum number of important peptides are shown. The amino acids are expressed in the single letter code. Solid lines indicate the sequences established by Edman degradation. Dashed lines show the sequences of which only the amino acid compositions are known. PG, PGI, and P indicate pepsinogen, pepsinogen activation intermediate, and pepsin, respectively.

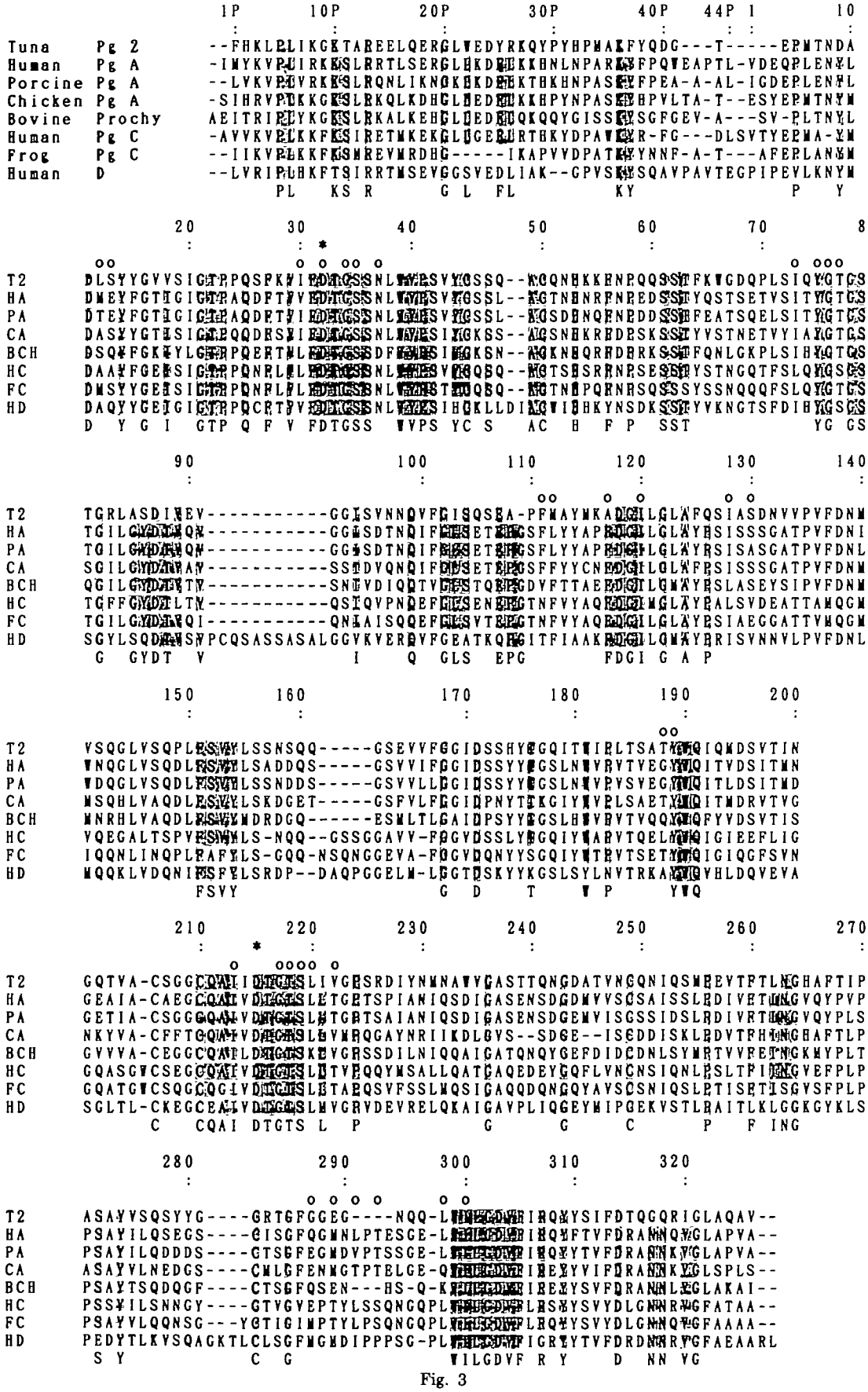

Fig. 3

Residue No.	11	29	72	110	116	127	187	212	286	297
Tuna pepsin 2	DLSY	VIFDTGSSNL	SIQYGTG	PFMA	KADGIL	SIASD	ATYW	AIIDTGTSLIVG	FGGEG----	-LWIL
Human pepsin A	·ME·	·V········	··T····	SFLY	PF····	··S·S	EG··	··V······LT·	·Q·MNLPTE	-····
Monkey pepsin A	·VE·	··········	··T····	SFLY	PF····	··S·S	EG··	··V······LT·	··Q·MDVPTE	-····
Porcine pepsin A	·TE·	··········	··T····	SFLY	PF····	··SAS	EG··	··V······LT·	·E·MDVPTS	-····
Chicken pepsin A	·A··	··········	Y·A····	SFFY	NF····	··S·S	E···	··V······L·M	·ENM·TPTE	-Q···
Bovine prochymosin	·SQ·	·L······DF	··H····	DVFT	EF····	·L··E	QQ··	··L·····KL··	·QS·N---H	-K···
Human pepsin C	·AA·	·L········	·L···S·	TNFV	QF···M	ALSV·	EL··	··V······LTV	VEPTYLSSQ	P····
Monkey pepsin C	·AA·	·L········	·L···S·	TNFV	QF···M	TLSV·	EL··	··V······LTV	VEPTYLSAQ	P····
Rat pepsin C	·A··	·L········	·L·····	TNFV	QF···M	GLS·G	EL··	G·V······L·M	LESISLTSE	P····
Human cathepsin D	·AQ·	·V········	D·H··S·	ITFI	·F····	R·SVN	KA··	··V······M··	·M·MDIPPP	P····
Human renin	·TQ·	·V·······V	TLR·S··	LPEM	EF··VV	EQ·IG	TGV·	·LV···A·Y·S·	IHAMDIPPP	PT·A·
Mouse renin	NSQ·	·······A··	T·H··S·	IPFM	QF··V·	RS·VG	TDS·	VVV···S·F·SA	LHAMDIPPP	PV·V·
Subsite S	43	1 1 3	121 2	11 3	1 1 3			1234 2 4	2 2	
Subsite S'		22 2	2 2			3 2	31 3	1	1 1	1 1

Fig. 4. **Comparison of the amino acid sequences of the subsite regions of pepsins, cathepsins D and E, and renins.** The subsite regions in porcine pepsin (*32*) and the corresponding regions in others are compared. Dots indicate residues identical with those in tuna pepsin, and dashes indicate deletions.

amino acid composition: Lys 10, His 5, Arg 8, Asp 17, Asn 17, Thr 24, Ser 38, Glu 10, Gln 32, Pro 16, Gly 36, Ala 21, 1/2Cys 6, Val 26, Met 9, Ile 24, Leu 19, Tyr 18, Phe 17, and Trp 7. The sequence contains no potential *N*-glycosylation site. Very recently the above sequence was confirmed by the cloning and sequencing of a cDNA coding for tuna pepsinogen 2 (Tanji, M. *et al.*, to be published).

DISCUSSION

The amino acid sequence of tuna pepsinogen is compared with those of other typical pepsinogens and human procathepsin D (*28*) in Fig. 3. Functionally important residues, such as the two aspartic acid residues in the active site, Asp32 and Asp215 (porcine pepsin numbering used hereafter), and Tyr 75 in the tyrosine flap, and the sequences around them, and the three disulfide bonds are highly conserved, as in other pepsinogens. However, there are several interesting features, including unique substitutions and deletions, in the tuna pepsinogen sequence as compared with in other pepsinogens, especially in light of the known three-dimensional structures of porcine pepsinogen (*29-31*) and pepsin (*32, 33*). The total numbers of residues in tuna pepsinogen (360 residues) and pepsin (323 residues) are the smallest among the pepsinogens and pepsins so far sequenced, others having 368-373 and 325-332 residues per molecule, respectively. This is due to the deletions near the NH_2-terminus, and at positions 109 and 291-294 of the pepsin moiety, as can be seen in Fig. 3. The region including the COOH-terminal part of the prosegment linked to the NH_2-terminal part of pepsin is most flexible as well as variable in the pepsinogen molecule (*29-31*), where cleav-

Fig. 3. **Comparison of the amino acid sequences of tuna pepsinogen 2 with those of other pepsinogens and human cathepsin D.** The sequences are aligned so as to maximize the homology. The amino acids are expressed in the single-letter code. The numbering is that of porcine pepsinogen and pepsin (*1, 3*). The residues at least common among higher vertebrate pepsinogens are shadowed, and also shown below the sequences. The hyphens indicate deletions, and asterisks indicate the active site aspartic acid residues. The vertical lines show the boundaries between the propeptides and the pepsin moieties. The subsite residues interacting with an ideal heptapeptide in porcine pepsin (*32*) are denoted by circles. Pg, pepsinogen; Prochy, prochymosin; D, cathepsin D.

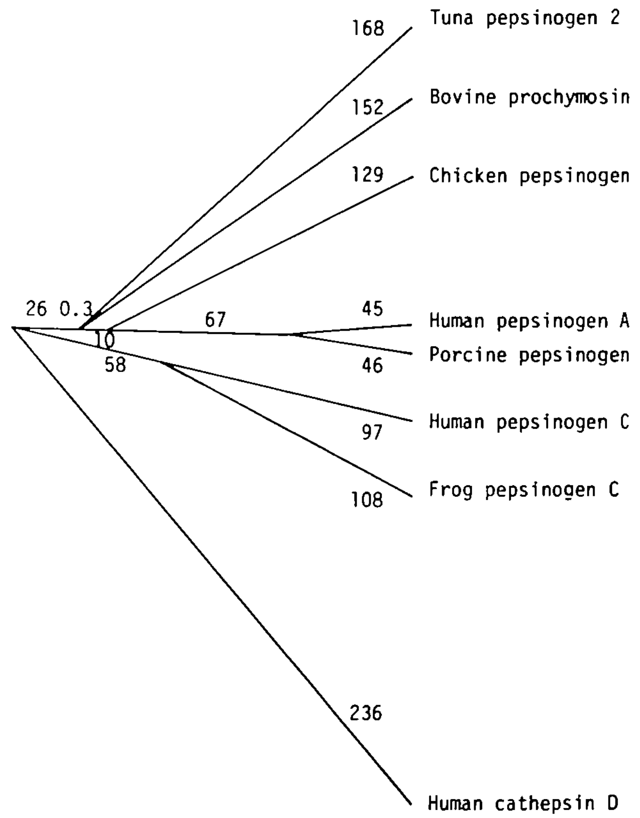

Fig. 5. **A phylogenetic tree of pepsinogens based on the amino acid sequences.** The tree was constructed using the sequence data in Fig. 3 by the method of Fitch and Margoliash (*54*). Human cathepsin D is also included for comparison. The value on each branch indicates the reconstructed mutation distance. The standard deviation of these values is 1.7%.

age occurs eventually upon activation of pepsinogen to pepsin (*34*). This cleavage of pepsinogens usually takes place on the NH_2-side of a residue 4 to 7 residues upstream of the conserved Pro5, depending on the specificity of the pepsin. Because of the deletion in this region, in tuna pepsinogen the length of the peptide chain between the conserved Lys36P and Pro5 residues is shortest among pepsinogens; there are 6 residues between Lys36P and Pro5, which may be the minimum number of residues required for the flap conformation (*29-31*). Thus, the cleavage upon activation occurs at the Phe37P-Tyr38P bond adjacent to Lys36P in tuna pepsinogen (*20*). Lys36P in the prosegment is highly conserved among pepsinogens, and its amino group is thought to interact with the carboxyl groups of Asp32 and Asp215 in the active site of the pepsin moiety and thereby stabilize the zymogen molecule, as shown in the case of porcine pepsinogen (*29-31*). The cleavage of the peptide bond between the prosegment and the pepsin moiety may occur generally through either an intramolecular or intermolecular mechanism (*35-39*). The close location in tuna pepsinogen of the scissile bond to Lys36P, and hence to the catalytic groups, appears to suggest the possibility that this bond in tuna pepsinogen might be partially cleaved through an intramolecular reaction at least in the initial stage of activation, and that this might be correlated with the rapid rate of activation of tuna pepsinogen (*20*). The deletions at position 109 and positions 291-294 are unique to tuna pepsinogen, except that prochymosin has a similar deletion in the latter region.

Among other side chain-side chain interactions between the prosegment and the pepsin moiety, the ionic interaction between Arg13P and Asp11, which are highly conserved among pepsinogens, and are thought to be important for stabilization of the conformation of the prosegment in this region (*29-31*), appears to also be conserved in tuna pepsinogen. On the other hand, however, Tyr37P and Tyr9, which are also highly conserved among pepsinogens and whose hydroxyl groups have been shown to be hydrogen-bonded to the carboxyl groups of Asp215 and Asp32, respectively, in porcine pepsinogen (*29-31*) are substituted by Phe and Asp, respectively, in tuna pepsinogen. Therefore, these hydrogen-bonding interactions are not possible, and, moreover, the latter substitution would even introduce repulsion between the carboxyl groups of Asp9 and Asp32. The Tyr37P→Phe substitution is thought to weaken the interaction between the prosegment and the pepsin moiety in tuna pepsinogen. This may be one of the reasons why tuna pepsinogen can be activated so rapidly.

It is also notable that, among the pepsinogens listed in Fig. 3, the content of basic amino acid residues in the pepsin moiety is higher than that in the prosegment in tuna pepsinogen (prosegment : pepsin = 10 : 13, residues/molecule) as in chicken pepsinogen A (14 : 18), bovine prochymosin (9 : 20), and human cathepsin D (7 : 33), whereas the contents are the reverse in the other four pepsinogens: human pepsinogen A (13 : 4), porcine pepsinogen (13 : 4), human pepsinogen C (12 : 4), and frog pepsinogen C (9 : 2). Furthermore, the net charges of tuna pepsinogen and pepsin are −1 and −6, respectively. These values are rather close to those of chicken pepsinogen and pepsin A (−4/−16), bovine prochymosin and chymosin (−3/−10), and human procathepsin and cathepsin D (+1/−2), whereas the other pepsinogens have much higher negative values: human pepsinogen and pepsin A (−20/−30), porcine pepsinogen and pepsin A (−27/−38), human pepsinogen and pepsin C (−15/−22), and frog pepsinogen C (−8/−14). Thus, tuna pepsinogen and pepsin appear to be least acidic among the pepsinogens and pepsins, and, in this respect, are closest to human procathepsin D and cathepsin D, respectively. The lower acidity of tuna pepsin as well as chicken pepsin A and bovine chymosin may be related with the higher pH optima, which are somewhat closer to that of cathepsin D, of these enzymes (*20, 40, 41*), as compared with those of mammalian pepsins A. These characteristics may also be correlated with the function of undeveloped gastric glands in lower vertebrates (*42*) and mammalian neonates (*43*). Pepsinogen and pepsin are absent in invertebrates, and are thought to have appeared first in fish stomach, differentiation occurring in the digestive tract in the course of evolution. Therefore, tuna pepsinogen may retain some of the characteristics of primitive pepsinogens, which might have evolved from an ancestral cathepsin D or cathepsin D-like tissue proteinase (*44*).

Figure 4 shows a comparison of the amino acid sequences of the regions in pepsins, human (*28*) and porcine (*45, 46*) cathepsins D, human (*47*) and guinea pig (*17*) cathepsins E, and human (*48*) and mouse (*49, 50*) renins, corresponding to the subsite regions of porcine pepsin (*32*). Most residues are highly or well conserved among pepsins. However, tuna pepsin has several unique substitutions. Among the S_1* site residues, residue 112 is Met in tuna pepsin, which is Leu or Phe in other pepsins, and residue 117 is Ala in tuna pepsin, which is Phe in all other pepsins. [*Nomenclature according to Schechter and Berger (*51, 52*).] Residue 287 in the S_2 site is Gly in tuna pepsin, which is a more bulky residue (Gln, Glu, or Met) in other pepsins. Furthermore, residues 291 and 293 at the S'_1 site are apparently deleted in tuna pepsin. It seems worth pointing out that the tetrapeptide sequence of residues 110 to 113, including S_1 site residues 111 and 112, appears to be one of the most variable regions among the subsite regions of pepsins, but is unique to each group of pepsins and related aspartic proteinases (Fig. 4). These differences should affect the nature of the active site, such as catalytic activity, substrate specificity, and susceptibility toward specific inhibitors of the enzyme. Indeed, as compared with porcine and other mammalian pepsins A, tuna pepsin has a higher pH optimum and appears to exhibit higher specific activity toward hemoglobin (*20*). On the other hand, it scarcely hydrolyzes *N*-acetyl-L-phenylalanyl-L-diiodotyrosine, a good substrate for pepsin A (*20*). Furthermore, it is less sensitive to diazoacetyl-D,L-norleucine methyl ester and 1,2-epoxy-3-(*p*-nitrophenoxy)propane, specific pepsin inhibitors (*20*). In this connection, it is interesting to note that the kinetic parameters of bovine chymosin were reported to change markedly on replacement of residues 111 and 112 in the S_1 site by site-directed mutagenesis (*53*).

There are 109 conserved residues in the other pepsinogens shown in Fig. 3, whereas 17 of these residues are different in tuna pepsinogen. The sequence identity was calculated to be 52% with human and chicken pepsinogens A, 51% with porcine pepsinogen A, 48% with bovine prochymosin, 47% with frog pepsinogen C, 45% with human pepsinogen C, and 36% with human procathepsin D. It is interesting to note that tuna pepsinogen shows the lowest identity with human procathepsin D, although fish

pepsins apparently resemble cathepsin D more than mammalian pepsins in certain properties (*44*). A phylogenetic tree of pepsinogens including tuna pepsinogen and human procathepsin D has been constructed based on their amino acid sequences, assuming minimum numbers of nucleotide exchanges among the zymogens (Fig. 5). The tree indicates that the pepsinogen and procathepsin D groups separated first from their common ancestor, and diverged to a greater extent from each other. It also indicates that after separation from procathepsin D, the pepsinogen C (progastricsin) group diverged from the rest of the pepsinogen groups, and that then tuna pepsinogen, prochymosin, and the pepsinogen A group diverged at nearly the same time. This phylogenetic tree is very similar to that constructed previously based on the NH_2-terminal sequences of the propeptides of pepsinogens and procathepsin D (*20*). Tuna pepsinogen, thus, does not appear to belong to either the typical pepsinogen A or prochymosin group although it seems to be closer to the prochymosin and pepsinogen A group than to the pepsinogen C group. The classification of pepsinogens into three major groups, *i.e.*, pepsinogen A, pepsinogen C (progastricsin), and prochymosin, has been established based on comparison of mammalian gastric aspartic proteinases (*1*), and therefore its application to lower vertebrates such as fish may not always be justified. To draw a more definite conclusion regarding the structure/function and evolutionary relationships of fish pepsinogens and pepsins, further studies are necessary, including elucidation of the amino acid sequences of other fish pepsinogens and of their three-dimensional structures.

We wish to thank Dr. Tokiharu Abe (Department of Zoology, University Museum, The University of Tokyo) for his generous help in obtaining fresh tuna stomachs and for the helpful discussion. We also thank Drs. J.J.N. Tang, J.A. Hartsuck, and G. Koelsch (Protein Studies Program, Oklahoma Medical Research Foundation), and Dr. J.C. Waymire and Mr. J. Ainsworth (Department of Neurobiology and Anatomy, University of Texas Medical School at Houston) for their interest in this work and the helpful discussions.

REFERENCES

1. Foltmann, B. (1981) Gastric proteinases—Structure, function, evolution and mechanism of action. *Essays Biochem.* **17**, 52-84
2. Tang, J. and Wong, R.N.S. (1978) Evolution in the structure and function of aspartic proteases. *J. Cell. Biochem.* **33**, 53-63
3. Tang, J., Sepulveda, P., Marciniszyn, J., Jr., Chen, K.C.S., Huang, W.-Y., Tao, N., Liu, D., and Lanier, J.P. (1973) Amino-acid sequence of porcine pepsin. *Proc. Natl. Acad. Sci. USA* **70**, 3437-3439
4. Ong, E.B. and Perlmann, G.E. (1968) The amino-terminal sequence of porcine pepsinogen. *J. Biol. Chem.* **243**, 6104-6109
5. Pederson, V.B. and Foltmann, B. (1973) The amino acid sequence of a hitherto unobserved segment from porcine pepsin preceding the N-terminus of pepsin. *FEBS Lett.* **35**, 255-256
6. Stepanov, V.M., Baratova, L.A., Pugacheva, I.B., Belyanova, L.P., Revina, L.P., and Timokhina, E.A. (1973) N-terminal sequence of swine pepsinogen and pepsin. The site of pepsinogen activation. *Biochem. Biophys. Res. Commun.* **54**, 1164-1170
7. Baudyš, M. and Kostka, V. (1983) Covalent structure of chicken pepsinogen. *Eur. J. Biochem.* **136**, 89-99
8. Kageyama, T. and Takahashi, K. (1986) The complete amino acid sequence of monkey pepsinogen A. *J. Biol. Chem.* **261**, 4395-4405
9. Kageyama, T. and Takahashi, K. (1986) The complete amino acid sequence of monkey progastricsin. *J. Biol. Chem.* **261**, 4406-4419
10. Foltmann, B., Pedersen, V.B., Jacobsen, H., Kauffman, D., and Wybrandt, G. (1977) The complete amino acid sequence of prochymosin. *Proc. Natl. Acad. Sci. USA* **74**, 2321-2324
11. Sogawa, K., Fujii-Kuriyama, Y., Mizukami, Y., Ichihara, Y., and Takahashi, K. (1983) Primary structure of human pepsinogen gene. *J. Biol. Chem.* **258**, 5306-5311
12. Kageyama, T., Tanabe, K., and Koiwai, O. (1990) Structure and development of rabbit pepsinogens. *J. Biol. Chem.* **65**, 17031-17038
13. Hayano, T., Sogawa, K., Ichihara, Y., Fujii-Kuriyama, Y., and Takahashi, K. (1988) Primary structure of human pepsinogen C gene. *J. Biol. Chem.* **263**, 1382-1385
14. Taggert, R.T., Cass, L.G., Mohandas, T.K., Derby, P., Barr, P.J., Pals, G., and Bell, G.I. (1989) Human pepsinogen C (progastricsin). Isolation of cDNA clones, localization to chromosome 6, and sequence homology with pepsinogen A. *J. Biol. Chem.* **264**, 375-379
15. Ichihara, Y., Sogawa, K., Morohashi, K., Fujii-Kuriyama, Y., and Takahashi, K. (1986) Nucleotide sequence of a nearly full-length cDNA coding for pepsinogen of rat gastric mucosa. *Eur. J. Biochem.* **161**, 7-12
16. Ishihara, T., Ichihara, Y., Hayano, T., Katsura, I., Sogawa, K., Fujii-Kuriyama, Y., and Takahashi, K. (1989) Primary structure and transcriptional regulation of rat pepsinogen C. *J. Biol. Chem.* **264**, 10193-10199
17. Kageyama, T., Ichinose, M., Tsukada, S., Miki, K., Kurokawa, K., Koiwai, O., Tanji, M., Yakabe, E., Athauda, S.B.P., and Takahashi, K. (1992) Gastric procathepsin E and progastricsin from guinea pig. Purification, molecular cloning of cDNAs, and characterization of enzymatic properties, with special reference to procathepsin E. *J. Biol. Chem.* **267**, 16450-16459
18. Yakabe, E., Tanji, M., Ichinose, M., Goto, S., Miki, K., Kurokawa, K., Ito, H., Kageyama, T., and Takahashi, K. (1991) Purification, characterization, and amino acid sequences of pepsinogens and pepsins from the esophageal mucosa of bullfrog (*Rana catesbeiana*). *J. Biol. Chem.* **266**, 22436-22443
19. Hayashi, K., Agata, K., Mochii, M., Yasugi, S., Eguchi, G., and Mizuno, T. (1988) Molecular cloning and the nucleotide sequence of cDNA for embryonic chicken pepsinogen: Phylogenetic relationship with prochymosin. *J. Biochem.* **103**, 290-296
20. Tanji, M., Kageyama, T., and Takahashi, K. (1988) Tuna pepsinogens and pepsins. Purification, characterization and amino-terminal sequences. *Eur. J. Biochem.* **177**, 251-259
21. de Bernardo, S., Weigele, M., Toome, V., Manhart, K., Leimgruber, W., Böhln, P., Stein, S., and Udenfriend, S. (1974) Studies on the reaction of fluorescamine with primary amines. *Arch. Biochem. Biophys.* **163**, 390-399
22. Crestfield, A.M., Moore, S., and Stein, W.H. (1963) The preparation and enzymatic hydrolysis of reduced and *S*-carboxymethylated proteins. *J. Biol. Chem.* **238**, 622-627
23. Spackman, D.H., Stein, W.H., and Moore, S. (1958) Automatic recording apparatus for use in the chromatography of amino acids. *Anal. Chem.* **30**, 1190-1206
24. van Eerd, J.-P. and Takahashi, K. (1976) Determination of the complete amino acid sequence of bovine cardiac troponin C. *Biochemistry* **15**, 1171-1180
25. Edman, P. and Henschen, A. (1975) Sequence determination in *Protein Sequence Determination* (Needleman, S.B., ed.), 2nd ed., pp. 232-279, Springer-Verlag, New York
26. Tarr, G.E. (1975) A general procedure for the manual sequencing of small quantities of peptides. *Anal. Biochem.* **63**, 361-370
27. Tsunasawa, S., Kondo, J., and Sakiyama, F. (1985) Isocratic separation of PTH-amino acids at picomole level by reverse-phase HPLC in the presence of sodium dodecylsulfate. *J. Biochem.* **97**, 701-704
28. Faust, P.L., Kornfeld, S., and Chirgwin, J.M. (1985) Cloning and sequence analysis of cDNA for human cathepsin D. *Proc. Natl. Acad. Sci. USA* **82**, 4910-4914
29. James, M.N.G. and Sielecki, A. (1986) Molecular structure of an aspartic proteinase zymogen, porcine pepsinogen, at 1.8 Å resolution. *Nature* **319**, 33-38
30. Sielecki, A.R., Fujinaga, M., Read, R.J., and James, M.N.G.

(1991) Refined structure of porcine pepsinogen at 1.8 Å resolution. *J. Mol. Biol.* **219**, 671-692
31. Hartsuck, J.A., Koelsch, G., and Remington, S.J. (1992) The high resolution crystal structure of porcine pepsinogen. *Proteins Struct. Funct. Genet.* **13**, 1-25
32. Sielecki, A.R., Fedorov, A.A., Boodhoo, A., Andreeva, N.S., and James, M.N.G. (1990) Molecular and crystal structures of monoclinic porcine pepsin refined at 1.8 Å resolution. *J. Mol. Biol.* **214**, 143-170
33. Cooper, J.B., Khan, G., Taylor, G., Tickle, I.J., and Blundell, T.L. (1990) X-ray analyses of aspartic proteinases. II. Three-dimensional structure of the hexagonal crystal form of porcine pepsin at 2.3 Å resolution. *J. Mol. Biol.* **214**, 199-222
34. Takahashi, K. and Kageyama, T. (1985) Multiplicity and intermediates of the activation mechanism of zymogens of gastric aspartic proteinases in *Aspartic Proteinases and Their Inhibitors* (Kostka, V., ed.) pp. 265-282, Walter de Gruyter, Berlin
35. Bustin, M. and Conway-Jacobs, A. (1971) Intramolecular activation of porcine pepsinogen. *J. Biol. Chem.* **246**, 615-620
36. Al-Janabi, J., Hartsuck, J.A., and Tang, J. (1972) Kinetics and mechanism of pepsinogen activation. *J. Biol. Chem.* **247**, 4628-4632
37. McPhie, P. (1974) Pepsinogen: Activation by a unimolecular mechanism. *Biochem. Biophys. Res. Commun.* **56**, 789-792
38. Marciniszyn, J., Jr., Huang, J.S., Hartsuck, J.A., and Tang, J. (1976) Mechanism of intramolecular activation of pepsinogen. Evidence for an intermediate delta and the involvement of the active site of pepsin in the intramolecular activation of pepsinogen. *J. Biol. Chem.* **254**, 7095-7102
39. Kageyama, T. and Takahashi, K. (1987) Activation mechanism of monkey and porcine pepsinogens A. *Eur. J. Biochem.* **165**, 483-490
40. Bahak, Z. (1969) Purification and characterization of chicken pepsinogen and chicken pepsin. *J. Biol. Chem.* **244**, 4638-4648
41. Foltmann, B. (1966) A review on prorennin and rennin. *C.R. Trav. Lab. Carlsberg* **35**, 143-231
42. Kapoor, B.G., Smit, H., and Verighina, I.A. (1975) The alimentary canal and digestion in teleosts. *Adv. Mar. Biol.* **13**, 109-239
43. Deren, J.S. (1971) Development of structure and function in the fetal and newborn stomach. *Am. J. Clin. Nutr.* **24**, 144-159
44. Gildberg, A. (1988) Aspartic proteases in fishes and aquatic invertebrates. *Comp. Biochem. Physiol.* **91B**, 425-435
45. Takahashi, T. and Tang, J. (1983) Amino acid sequence of porcine spleen cathepsin D light chain. *J. Biol. Chem.* **258**, 6435-6443
46. Shewale, J.G. and Tang, J. (1984) Amino acid sequence of porcine spleen cathepsin D. *Proc. Natl. Acad. Sci. USA* **81**, 3703-3707
47. Azuma, T., Pals, G., Mohandas, T.K., Couvreur, J.M., and Taggart, R.T. (1989) Human gastric cathepsin E. Predicted sequence, localization to chromosome 1, and sequence homology with other aspartic proteinases. *J. Biol. Chem.* **264**, 16748-16753
48. Imai, T., Miyazaki, H., Hirose, S., Hori, H., Hayashi, T., Kageyama, R., Ohkubo, H., Nakanishi, S., and Murakami, K. (1983) Cloning and sequence analysis of cDNA for human renin precursor. *Proc. Natl. Acad. Sci. USA* **80**, 7405-7409
49. Misono, K.S., Chang, J.-J., and Inagami, T. (1982) Amino acid sequence of mouse submaxillary gland renin. *Proc. Natl. Acad. Sci. USA* **79**, 4858-4862
50. Panthier, J.-J., Foote, S., Chambraud, B., Strosberg, A.D., Corvol, P., and Rougeon, F. (1982) Complete amino acid sequence and maturation of the mouse submaxillary gland renin precursor. *Nature* **298**, 90-92
51. Schechter, I. and Berger, A. (1967) On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **27**, 157-162
52. Schechter, I. and Berger, A. (1968) On the active site of proteases. 3. Mapping of the active site of papain; specific peptide inhibitors of papain. *Biochem. Biophys. Res. Commun.* **32**, 898-902
53. Suzuki, J., Sasaki, K., Sasao, Y., Hamu, A., Kawasaki, H., Nishiyama, M., Horinouchi, S., and Beppu, T. (1989) Alteration of catalytic properties of chymosin by site-directed mutagenesis. *Protein Eng.* **2**, 563-569
54. Fitch, W.M. and Margoliash, E. (1967) Construction of phylogenetic trees. *Science* **155**, 279-284