J. Biochem. 120, 407-414 (1996)

Purification and Characterization of Turtle Pepsinogen and Pepsin[1]

Akira Hirasawa,* Senarath B.P. Athauda,* and Kenji Takahashi*,†,2

*Department of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, Bunkyo-ku, Tokyo 113; and †School of Life Science, Tokyo University of Pharmacy and Life Science, 1432-1 Horinouchi, Hachioji, Tokyo 192-03

Received for publication, April 19, 1996

Pepsinogen was purified from the gastric mucosa of soft-shelled turtle (*Trionyx sinensis*) by a series of chromatographies on DEAE-cellulose, Sephadex G-100, and Q-Sepharose. Upon chromatography on Q-Sepharose, it was separated into nine isoforms. These isoforms showed a relative molecular mass of approximately 43,000 Da on sodium dodecyl sulfate-polyacrylamide gel electrophoresis, and isoforms 4 through 9 contained carbohydrate (approx. 2% each). Insofar as they were examined, their NH_2-terminal sequences differed only in showing substitution at a few positions. At pH 2.0, they were rapidly activated to the corresponding isoforms of pepsin in a stepwise manner. The nine isoforms showed similar specific activity toward hemoglobin and hydrolyzed *N*-acetyl-L-phenylalanyl-L-diiodotyrosine, a good substrate for pepsin A, at somewhat different rates. They were inhibited by pepstatin to various extents, more strongly than human pepsin C but less strongly than human pepsin A. All isoforms appeared to have similar cleavage specificity toward oxidized insulin B chain, which resembled those of both human pepsins A and C. A cDNA clone for one of the zymogen isoforms was isolated and sequenced. The amino acid sequence thus deduced was more homologous with those of mammalian pepsinogens A than those of mammalian pepsinogens C or prochymosin.

Key words: amino acid sequence, cDNA sequence, pepsin, pepsinogen, turtle.

Pepsinogen, the precursor of the major gastric aspartic proteinase pepsin, is secreted from the gastric mucosa into the gastric lumen, where it is autocatalytically activated to pepsin under acidic conditions. It is classified into three major types: pepsinogen A, pepsinogen C (or progastricsin), and prochymosin (or neonatal pepsinogen) (*1*). Further, pepsinogens, especially pepsinogens A, are often composed of several isoforms. To study the structure/function relationship and molecular evolution of pepsinogens and pepsins, we have purified and characterized a number of pepsinogens from various sources, including human (*2-4*), monkey (*5-13*), bear (*14*), rabbit (*15*), guinea-pig (*16*), frog (*17*), and tuna (*18*), and sequenced several of them at the protein and/or DNA level. We have reported the complete amino acid sequences of human (*19*), monkey (*20*), and porcine (*21*) pepsinogens A, and human (*22*), monkey (*23*), rat (*24, 25*), guinea-pig (*16*), and frog (*17*) pepsinogens C. Other workers have reported the complete sequences for porcine (*26-29*), rabbit (*30*), and chicken (*31*) pepsinogens A, chicken embryonic pepsinogen (*32*), and bovine prochymosin (*33*). Thus, for non-mammalian species, the complete sequences of only three pepsinogens have been determined. Furthermore, the biochemical properties of non-mammalian pepsins have been investigated only for chicken (*34*), frog (*17*), and tuna (*18*). To extend this series of studies, it is important to isolate more pepsinogens and pepsins from different sources, especially from non-mammalian species, and to elucidate their properties including amino acid sequences. So far, no studies have been reported on reptilian pepsinogens and pepsins.

In the present study, we have isolated multiple forms of pepsinogen from the gastric mucosa of soft-shelled turtle (*Trionyx sinensis*) as a representative of reptilian species and investigated their properties together with those of the corresponding pepsin isoforms. Further, a cDNA clone for a turtle pepsinogen isoform was isolated and sequenced.

EXPERIMENTAL PROCEDURES

Materials—The stomachs of soft-shelled turtles (*Trionyx sinensis*) were kindly supplied by Murakami Suppon and Hosendo, Tokyo. The following materials were obtained from the sources indicated. Bovine hemoglobin, Worthington Biochem.; DEAE-cellulose (DE-52), Whatman; Sephadex G-100 and Q-Sepharose, Pharmacia; reagents for amino acid analysis, Wako Pure Chem.; *N*-acetyl-L-phenylalanyl-L-diiodotyrosine, Nacalai Tesque; oxidized insulin B chain, Sigma; reagents for molecular cloning, Takara, Nippon gene, Toyobo, Boehringer-Mannheim, and/or BRL; DNA multiprime system and a reagent kit for separation of λgt10 cDNA library, Amersham; a reagent kit for cDNA preparation, Boehringer-Mannheim; reverse

[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] 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: +81-426-76-7146, Fax: +81-426-76-7149, E-mail:kenjitak@ls.toyaku.ac.jp

Abbreviations: SDS, sodium dodecyl sulfate; PAGE, polyacrylamide gel electrophoresis.

transcriptase (MKV-RT), BRL; a DNA ligation kit, Takara; reagents for DNA sequencing, Applied Biosystems; radioactive compounds, Amersham and/or New England Nuclear.

Assay for Proteolytic Activity—Potential pepsin activity of pepsinogen and pepsin activity were determined essentially by the method of Anson (*35*) with slight modification. An enzyme solution (80 μl) was mixed with 400 μl of 2.5% acid-denatured hemoglobin in 0.1 M sodium formate buffer, pH 2.9, and incubated at 37°C for 30 min. To this was added 800 μl of 5% (w/v) trichloroacetic acid and after standing for 30 min, the mixture was centrifuged at 10,000 $\times g$ for 10 min. The absorbance at 280 nm of the supernatant was measured. One unit of activity was defined as the amount causing an increase in absorbance at 280 nm of 1.0 per min. Pepsin activity toward *N*-acetyl-L-phenylalanyl-L-diiodotyrosine was determined essentially by the method of Tang (*36*).

Determination of Protein—Protein concentration was determined by the method of Lowry *et al.* (*37*) with bovine serum albumin as a standard.

Purification of Pepsinogen—All purification procedures were performed at 4°C.

Step 1. Preparation of crude extract: The gastric mucosa (84 g) from four stomachs was homogenized with a Waring blender in 100 ml of 50 mM sodium phosphate buffer, pH 7.5. The homogenate was centrifuged at 3,000 $\times g$ for 30 min. The supernatant was again centrifuged at 100,000 $\times g$ for 120 min to yield a crude extract.

Step 2. DEAE-cellulose chromatography: The crude extract (110 ml) was applied to a column (6 $\times$ 22 cm) of DE-52 equilibrated with 50 mM sodium phosphate buffer, pH 7.5, and eluted with a gradient of NaCl (0 to 1 M) in a total volume of 2,000 ml of the same buffer at a flow rate of 120 ml/h. Fractions of 15.6 ml were collected.

Step 3. Sephadex G-100 chromatography: The active fractions from step 2 were combined, concentrated by ammonium sulfate precipitation, and dialyzed against 50 mM sodium phosphate buffer, pH 7.5. The dialyzed sample was applied to a column (3.5 $\times$ 150 cm) of Sephadex G-100 equilibrated with the same buffer. The protein was eluted at a flow rate of 18 ml/h and 6.3-ml fractions were collected.

Step 4. Q-Sepharose chromatography: The active fractions from step 3 were combined and one-fifteenth of the combined fraction was applied to a column of Q-Sepharose (HR 10/10, 10 ml) equilibrated with 50 mM piperazine HCl buffer, pH 6.0. The protein was eluted with a gradient of NaCl (0 to 0.45 M) in a total volume of 1,000 ml of the same buffer at a flow rate of 30 ml/h. Fractions of 10 ml were collected. Major active fractions of each peak were pooled and each pooled fraction was purified further by rechromatography under the same conditions.

Polyacrylamide Gel Electrophoresis (PAGE)—SDS-PAGE was performed according to Laemmli (*38*). Protein was stained with Coomassie Brilliant Blue. Native PAGE and the subsequent activity staining were performed essentially as described by Furihata *et al.* (*39*).

Amino Acid Analysis—Each peptide sample was hydrolyzed in 0.3 ml of 6 N HCl containing 0.05% phenol at 110°C for 24 h in an evacuated sealed tube and the amino acids were analyzed in a Hitachi 835 amino acid analyzer. Cysteic acid was determined after performic acid oxidation according to Moore (*40*).

Amino Acid Sequence Determination—The amino acid sequences of the NH_2-terminal regions of pepsinogen and pepsin isoforms were determined by using an automated protein sequencer, model 477A, equipped with an on-line PTH-amino acid analyzer (model 120A) (Applied Biosystems).

Analysis of Carbohydrate—Carbohydrate in each pepsinogen isoform was analyzed by the phenol-sulfuric acid method of Dubois *et al.* (*41*).

Analysis of Phosphate—Phosphate bound to protein was determined according to Bartlett (*42*).

Activation of Pepsinogen to Pepsin—Each pepsinogen isoform solution was adjusted to pH 2.0 with 0.1 N HCl and incubated at 14°C at a protein concentration of 0.13 mg/ml. Aliquots of 25 μl were removed at appropriate intervals, mixed with the SDS-PAGE buffer and submitted to SDS-PAGE. Protein was stained with Coomassie Brilliant Blue.

Hydrolysis of Oxidized Insulin B Chain—Oxidized bovine insulin B chain (100 μg) was digested with one-hundredth by weight of the enzyme in 10 mM Tris glycine buffer, pH 1.5, at 37°C for 3 h. The digest was analyzed by HPLC on a TSKgel ODS-120T column (0.46 $\times$ 25 cm) in a Hitachi 655A-11 HPLC system. Elution was performd with a gradient of acetonitrile (0 to 50% in 50 min) in 0.1% trifluoroacetic acid at a flow rate of 0.8 ml/min and monitored at 215 nm. Isolated peptide fractions were hydrolyzed and submitted to amino acid analysis as described above.

Isolation of a cDNA Clone and Nucleotide Sequence Analysis—Total RNA was obtained from the gastric mucosa essentially according to Chirgwin *et al.* (*43*) and submitted to oligo(dT)-cellulose chromatography. With the isolated poly(A)$^+$ RNA, cDNA was synthesized using a reverse transcriptase and a cDNA synthesis kit, and a λgt10 cDNA library was constructed from this cDNA as described (*44*). The cDNA library was screened by the colony hybridization method of Grunstein *et al.* (*45*) using ^{32}P-labeled human pepsinogen A cDNA as a probe. Hybridization was performed at 65°C overnight, and the filter was washed with 2 $\times$ SSC and 0.2 $\times$ SSC. The isolated cDNA clone was digested with restriction endonucleases and the resulting fragments were inserted into pUC 118 and 119. Single-chain DNAs were isolated according to Kuchino and Takanami (*46*) and sequenced by Sanger's dideoxy-mediated chain-termination method (*47*) using an Applied Biosystems model 370A DNA sequencer.

RESULTS

Purification and Properties of Pepsinogen—Pepsinogen in the mucosal extract was purified as summarized in Table I. It was eluted as one major peak upon chromatographies on DEAE-cellulose and Sephadex G-100 (Fig. 1, a and b) and the Sephadex G-100 pooled fraction gave a single band on sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) corresponding to a molecular mass of approximately 43,000 Da (Fig. 1d). Upon PAGE under non-denaturing conditions, however, both DEAE-cellulose and Sephadex G-100 fractions gave multiple bands of pepsinogen isoforms (data not shown). These isoforms could be separated reproducibly into nine fractions by chromatography on Q-Sepharose, as shown in Fig. 1c. Each

fraction yielded an apparently single band of activity, and each showed different mobility from the others on PAGE under non-denaturing conditions (Fig. 1e), indicating that they were eluted from the Q-Sepharose column in the order of increasing net anionic charge. Pepsinogens 4 through 9 contained carbohydrate (approximately 2% each), whereas none of the nine isoforms contained phosphate. The amino acid compositions of the nine isoforms were similar to one another (data not shown).

NH_2-Terminal Amino Acid Sequences of Pepsinogen and Pepsin—Isoforms 2, 5, 8, and 9, the major isoforms, were sequenced up to residue 30. Minor isoforms 6 and 7 were sequenced up to residues 30 and 18, respectively, and the other minor isoforms, 1, 3, and 4, up to residue 8. The sequences thus obtained were identical in all but three positions. The residue at position 8 was Lys in isoforms 1 through 7 and Gln in isoforms 8 and 9, and that at position 20 was His in isoforms 2, 8, and 9 and Ala in isoforms 5 and 6. The residue at position 27 was indicated, although not conclusively, to be Lys in isoforms 2, 5, and 6 and Asn in isoforms 8 and 9. Thus, all or most of the isoforms are thought to have one of the following NH_2-terminal sequences.

```
1    5         10        15        20        25        30
LVTKVPLQ/KKGKSLRQNLKEH/AGLLE(D)F(K/N)KKH
```

The major pepsin isoforms 2, 5, 8, and 9 were all sequenced up to residue 24 and the minor isoforms 6 and 7 up to residue 20. The results were identical, indicating that all or most of the pepsin isoforms have the same NH_2-terminal sequence as shown below.

```
1    5         10        15    20  24
ATEPLTNYMDAEYFGTISIGTPAQ
```

Activation of Pepsinogen—The pepsinogen isoforms were activated in a similar manner. The activation occurred very rapidly, taking only a few minutes for complete conversion of each pepsinogen isoform to the corresponding pepsin isoform. Figure 2 shows the time courses of activation of pepsinogen isoforms 7, 8, and 9 as typical examples. In each case, apparently two intermediate forms were

TABLE I. **Purification of turtle pepsinogen.**

Purification step	Total activity[a] (units)	Total protein (mg)	Specific activity (units/mg)	Activity yield (%)
Crude extract	3,129	1,623	1.9	100
DE-52	2,402	517	4.7	77
Sephadex G-100	1,273	206	6.2	41
Q-Sepharose[b]	660	57	10.0–14.0	21

[a]Potential pepsin activity was determined with hemoglobin as a substrate. [b]One-fifteenth of the pooled Sephadex G-100 fraction was used; the sums of the activity, protein, and activity yields of the nine isoform fractions calculated for the total sample are shown.

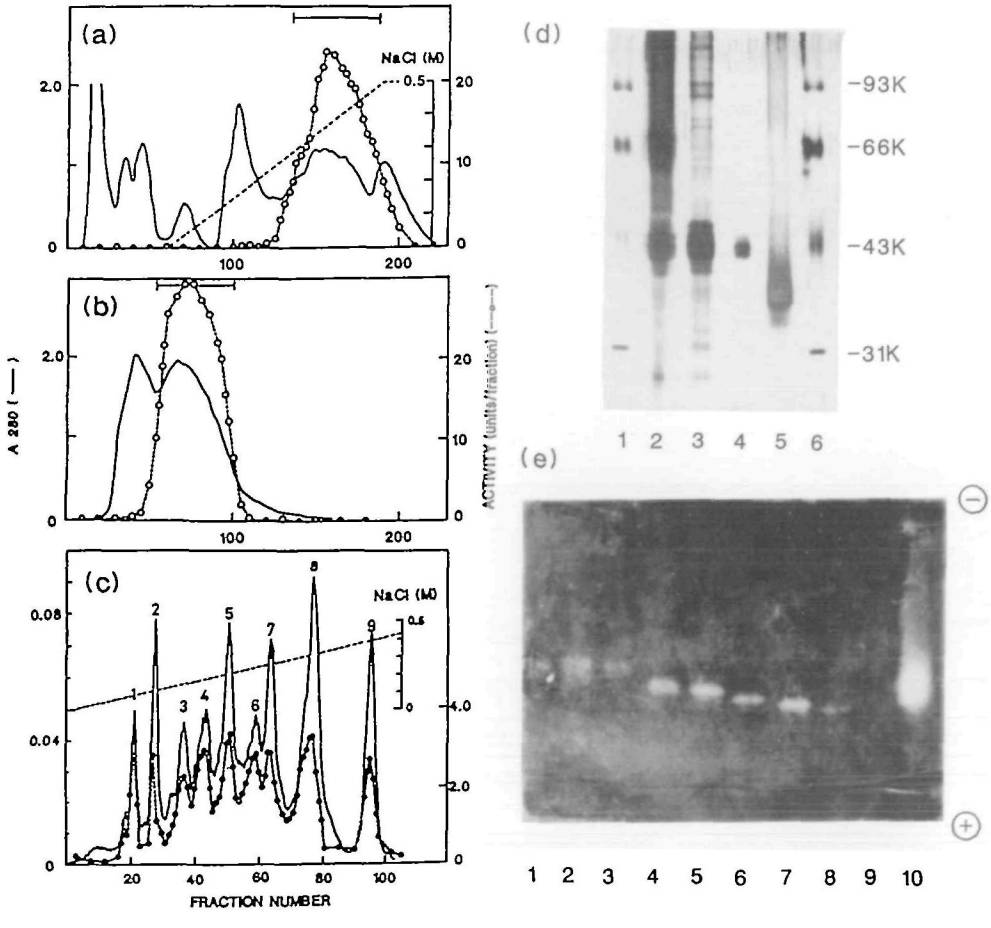

Fig. 1. **Purification of turtle pepsinogen.** (a) DEAE-cellulose. The crude extract of turtle gastric mucosa (84 g) was applied to a column (6×22 cm) of DE-52 equilibrated with 50 mM sodium phosphate buffer, pH 7.5. Fractions under the bar were pooled. (b) Sephadex G-100. The active fraction from DEAE-cellulose chromatography was concentrated and dialyzed against 50 mM sodium phosphate buffer, pH 7.5, and applied to a column (3.5×150 cm) of Sephadex G-100 equilibrated with the same buffer. Fractions under the bar were pooled. (c) Q-Sepharose. About one-fifteenth of the enzyme fraction obtained by Sephadex G-100 chromatography was applied to a column of Q-Sepharose (HR10/10) equilibrated with 50 mM piperazine HCl buffer, pH 6.0. (d) SDS/PAGE. The concentration of acrylamide was 10%. Lanes 1 and 6, marker proteins; lane 2, crude extract; lane 3, DE-52 pooled fraction; lane 4, Sephadex G-100 pooled fraction; lane 5, Sephadex G-100 pooled fraction incubated at pH 2.0 and 37°C for 10 min, indicating conversion of pepsinogen to pepsin. (e) Native PAGE of pepsinogen isoforms separated by Q-Sepharose chromatography. The concentration of acrylamide was 6%. Lanes 1–9, isoforms 1–9, respectively; lane 10, Sephadex G-100 pooled fraction. Isoform 9 band is too faint to be seen on this photograph.

produced during the course of activation, indicating that the activation was a stepwise process, involving at least two intermediate forms.

Enzymatic Properties—The specific activity toward hemoglobin was similar among the pepsin isoforms (Table I). Each isoform also hydrolyzed *N*-acetyl-L-phenylalanyl-

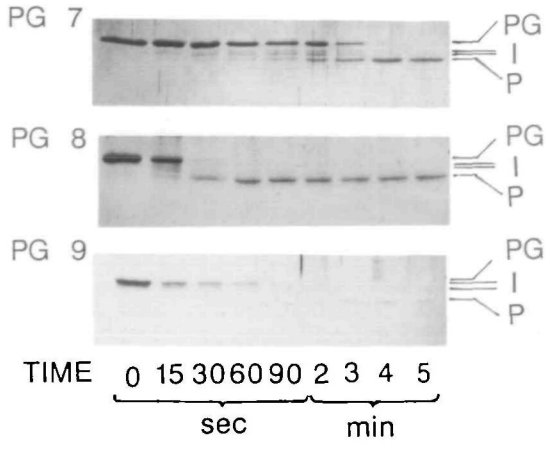

Fig. 2. **Time course of activation of pepsinogen.** The results with isoforms 7, 8, and 9 are shown. They were incubated separately at pH 2.0 and 14°C at a protein concentration of 0.13 mg/ml. Aliquots were removed at appropriate intervals and submitted to SDS-PAGE. The proteins were stained with Coomassie Brilliant Blue. PG, pepsinogen; I, intermediate; P, pepsin.

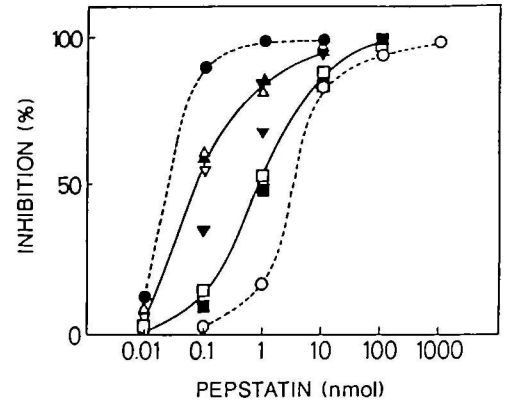

Fig. 3. **Pepstatin inhibition of turtle pepsin activity.** Each pepsinogen isoform (0.06 nmol) in 20 μl of 10 mM piperazine HCl buffer, pH 6.0, was mixed with 20 μl of pepstatin solution in aqueous dimethylsulfoxide and then 400 μl of acid-denatured hemoglobin solution was added to this mixture for assay under the standard conditions. The results with isoforms 1 (□), 4 (■), 5 (△), 6 (▼), 7 (▲), and 9 (▽) are shown. Human pepsinogens A (●) and C (○) were also used for comparison. The results with isoforms 2, 3, 7, and 8 (data not shown) were similar to those with isoforms 1, 4, 6, and 9, respectively.

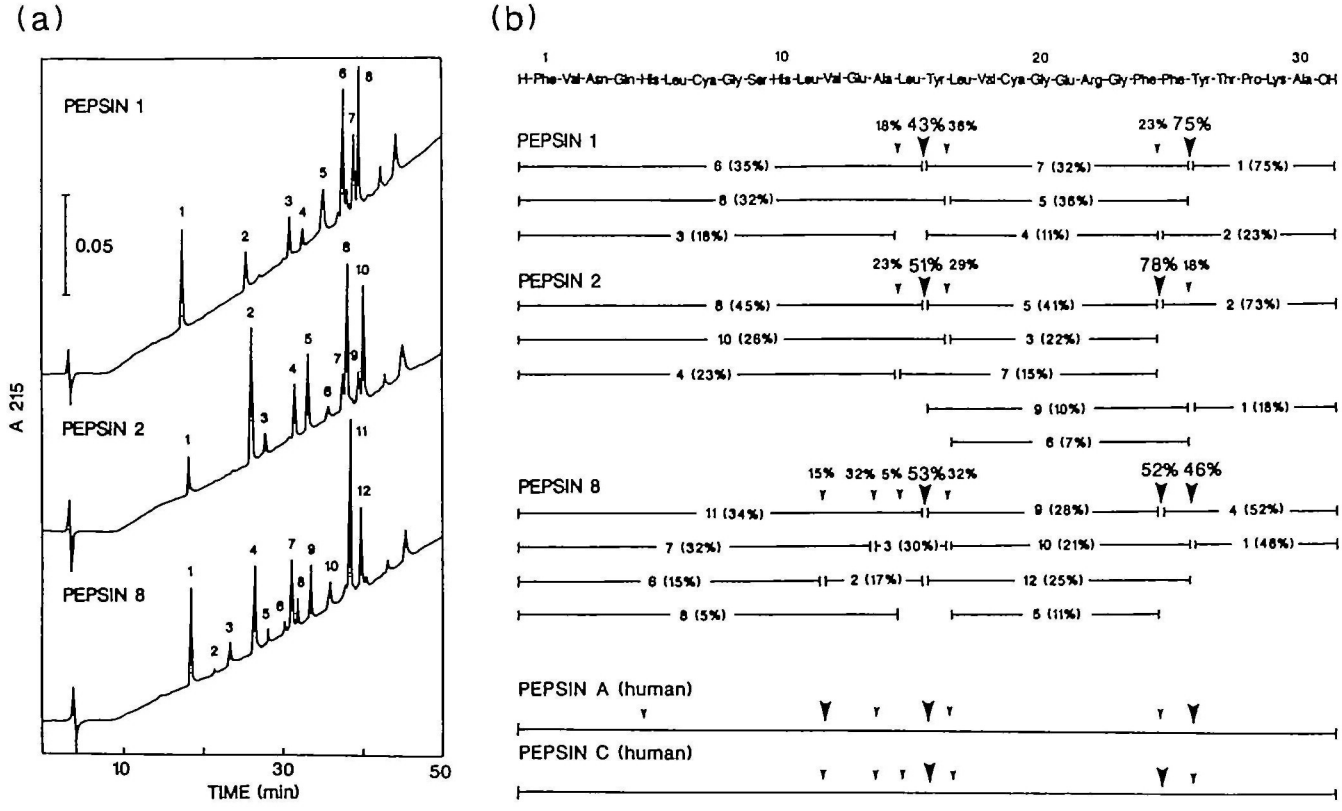

Fig. 4. **Analyses of the cleavage sites of oxidized insulin B chain by pepsin.** The results obtained with pepsin isoforms 1, 2, and 8 are shown. (a) HPLC patterns of 2-h digests of oxidized insulin B chain. An aliquot of each digest was applied to a TSKgel ODS-120T column (0.46×25 cm) and eluted with a gradient of acetonitrile (0–50% in 50 min) in 0.1% trifluoacetic acid at a flow rate of 0.8 ml/min. Elution was monitored at 215 nm. (b) Sites of cleavage and yields of peptides. Large and small arrowheads indicate major and minor cleavage sites, respectively. The value above each arrowhead shows the estimated extent of cleavage (%). The cleavage sites by human pepsins A and C (*49*) are also shown for comparison.

L-diiodotyrosine, a good substrate for pepsin A, although the rate of hydrolysis varied among them; isoforms 1 and 2 had a specific activity somewhat lower than human pepsin A but much higher than that of human pepsin C, whereas isoforms 4 to 9 had a specific activity comparable to or somewhat higher than that of human pepsin A (data not shown). On the other hand, they were less sensitive to pepstatin than human pepsin A but more sensitive to pepstatin than human pepsin C as shown in Fig. 3. The HPLC patterns of the digests of oxidized insulin B chain were nearly identical among the isoforms except for isoforms 1 and 2. Therefore, most of the isoforms are thought to have essentially the same specificity. Figure 4 shows the HPLC pattern and the results of analysis of the peptides for isoform 8 together with those for isoforms 1 and 2, which gave slightly different HPLC patterns. Although these isoforms showed similar specificity, their relative extents of cleavage of the Phe^{24}-Phe^{25} and Phe^{25}-Tyr^{26} bonds varied significantly. These specificities, however, generally resembled those of both human pepsins A and C (*48*) as can be seen from Fig. 4.

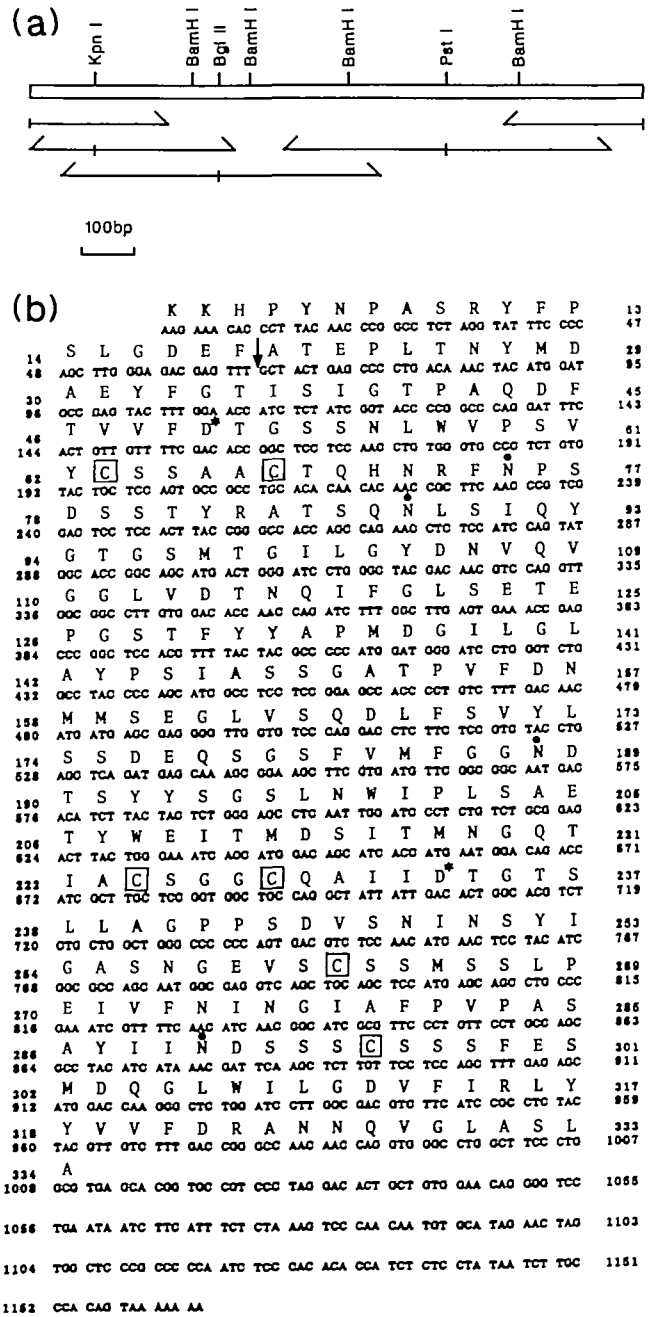

Fig. 5. **Sequencing of a turtle pepsinogen isoform cDNA and the deduced amino acid sequence.** (a) Restriction map and sequencing strategy. The coding region is indicated by an open frame. The arrows indicate the directions and extents of sequence determination. (b) The nucleotide sequence (lower line) and the deduced amino acid sequence (in one-letter code) (upper line) are shown. The vertical arrow indicates the boundary between the propart and the pepsin moiety. The putative active site aspartic acid residues are shown with asterisks and the potential *N*-glycosylation sites with closed circles. The half-cystine residues are boxed.

Molecular Cloning of Pepsinogen cDNA and the Deduced Amino Acid Sequence—By screening 4,000 clones of the λgt10 library using a human pepsinogen A cDNA as a probe, 22 positive clones were isolated. A restriction nuclease map of one of these clones is shown in Fig. 5a, and its determined nucleotide sequence and deduced amino acid sequence are shown in Fig. 5b. The cDNA clone encoded part (19 residues) of the propeptide and the complete amino acid sequence (315 residues) of the pepsin part of a pepsinogen isoform. The sequene starting from position 20 in the deduced sequence (Fig. 5b) agreed with the NH_2-terminal sequence of the pepsin isoforms determined at the protein level.

DISCUSSION

Multiple isoforms of pepsinogen have been found in various animal species. Notably, five pepsinogen A isoforms were found in human (*3, 48*) and monkey (*6*), and six pepsinogen A isoforms in rabbit (*15*) stomachs. The nine isoforms in turtle stomach thus exceed these values. The same number of isoforms was also obtained from each stomach when three individual turtle stomachs were examined separately. Therefore, this multiplicity does not seem to be due to individual differences. Since these isoforms were separated from one another fairly well by Q-Sepharose chromatography and native PAGE, they should differ from each other in net charge. The fact that none of them contains phosphate suggests that such differences would be due primarily to amino acid substitutions. The NH_2-terminal sequences of pepsinogen and pepsin isoforms so far determined do not account for the charge differences, so additional substitutions may be present elsewhere in the zymogen molecules. Further, isoforms 4 through 9 contain carbohydrate, which may also be partly responsible for generation of the isoforms. Carbohydrate-containing pepsinogens are rather rare. We previously reported occurrence of carbohydrate-containing pepsinogens A in Japanese monkey stomach as minor components (*5*) and carbohydrate-containing pepsinogen C in frog esophageal gland (*17*). In addition, only chicken pepsinogen A was reported to be a glycoprotein (*31*). It is also interesting to note that the turtle pepsinogen generally appeared to be activated much faster than pepsinogens from other sources (*3, 17, 50*). The resulting pepsin was also fairly susceptible to autolysis (data not shown).

Some enzymatic properties of the pepsin isoforms were intermediate between those of human pepsins A and C.

This seems to indicate that pepsin(ogen)s A and C may not always be clearly distinguishable by enzymatic properties alone. The results of NH_2-terminal sequencing appear to indicate that all or most isoforms have the same NH_2-terminal sequence except for substitutions at a few positions. In mammalian pepsinogens A and C and prochymosin, the NH_2-terminal sequences of the three groups are considerably different. Taken together, these findings suggest that all or most of the turtle pepsinogen isoforms belong to the same group of pepsinogen, and that the variations in several enzymatic properties among them are due to differences in sequence in the internal part of the pepsin moiety.

The NH_2-terminal Lys-His sequence deduced from the cDNA sequence overlaps the same sequence at position 29 to 30 in the NH_2-terminal sequence of the major turtle pepsinogen isoforms determined at the protein level and this is supported by comparison with the known NH_2-terminal sequences of various pepsinogens. Therefore, we tentatively deduced a complete sequence of turtle pepsinogen by overlapping these sequences, as shown in Fig. 6, where this sequence is compared with the sequences of other typical pepsinogens. The sequences around the two active-site aspartic acid residues and the location of the three disulfide bonds are also conserved in turtle pepsinogen. All other pepsinogens so far sequenced have a conserved Lys-Tyr sequence near the carboxyl-terminal region of the propeptide; this Lys residue has been shown by X-ray crystallographic analysis to interact electrostatically with the active-site aspartyl residues in porcine pepsinogen (*51-53*) and human progastricsin (*54*) and is assumed to be important for correct folding and stabilization of the pepsinogen molecule. Interestingly, this lysine residue is replaced with an Arg residue in turtle pepsinogen, indicating that the conserved Lys residue may be replaced with another basic residue, Arg. The identity of the deduced amino acid sequence of turtle pepsinogen was approximately 70, 69, and 63% with human (*19*), porcine (*21, 26-29*), and chicken (*31*) pepsinogens A, respectively, 59% with chicken embryonic pepsinogen (*32*), 55% with bovine prochymosin (*33*), and 50% with human pepsinogen C (*22*). Thus, the deduced sequence is thought to belong to A-type pepsinogens rather than C-type or prochymosin-type pepsinogens. It is interesting to note that turtle pepsinogen shows somewhat higher homology with mammalian (porcine and human) pepsinogens A than with chicken pepsinogen A, indicating that the evolutionary distance is somewhat greater between turtle and chicken pepsinogens A than between turtle and mammalian pepsinogens A. On the other hand, in turtle pepsinogen, 5-residue deletions were observed at positions between 283 and 287 and between 330 and 331 (Fig. 6). A similar deletion in the first region was observed only with chicken pepsinogen A, and that in the second region was observed only with chicken embryonic pepsinogen and bovine prochymosin. The turtle pepsinogen sequence contains four potential *N*-glycosylation sites, whereas chicken embryonic pepsinogen contains five potential *N*-glycosylation sites, chicken pepsinogen A and bovine prochymosin have two, human pepsinogen C has one, and human and porcine

```
                                  1       K   11        H  21       (K)     31          41          51          61
Turtle pepsinogen                 ·LVTKVPLQKG KSLRQNLKEA GLLE(D)F(N)KKH PYNPASRYFP SLGD··EFATEP LTNYMDAEYF GTISIGTPAQ
Human pepsinogen A                ·IMY----IRK ----RT-S-R ---K - - L --- NL---RK--- QWEAPTLVDEQ- -E--L-M--- ---G------
Porcine pepsinogen A              ··LV----VRK -------IKD ---K - - L -T- KH----K--- ··EAAALIGD-- -E--L-T--- ---G------
Chicken pepsinogen A              ·SIHR---K-- ----KQ--DH ---- - - L --- ------K-H- V-TA··TESY-- M------S-Y --------Q-
Chicken embryonic pepsinogen      ··I-RL--ER- -K--EI-R-K ---H H - L QH- R-DIGTKFPH AFP-VLTVV--- -L-TL-M--Y --------P-
Bovine prochymosin                AEI-RI--Y-- ----KA---H ---- - - L Q-Q Q-GIS-K-SG ·F-···-V-SV- ----L-SQ-- -K-YL---P-
Human pepsinogen C                ·A-V----K-F --I-ETM--K ---G E - L RT- K-D--WK-R· ·F--L-SVTY-- MA·----A-- -E------P-

71    *   81          91          101         111         121         131         141         151         161
DFTVVFDTGS SNLWVPSVYC SSAACTQHNR FNPSDSSTYR ATSQNLSIQY GTGSMTGILG YDNVQVGGLV DTNQIFGLSE TEPGSTFYYA PMDGILGLAY
---------- ---------- --L---N--- ---E-----Q S--ETV--T- ---------- --T-----IS ---------- -----FL--- -F--------
----I----- ---------- --L---D--Q ---D----FE ---E---T-- ---------- --T-----IS ---------- -----FL--- -F--------
--S-I----- -------I-- K-S--SN-K- -D--K----V S-NETVY-A- -----S---- --T-A-SSID VQ-------- -----F---C NF-------F
---------- --------S- T-P--QS-QM ---Q----K S-G---H-- ---D-E-TV- C-T-T-AS-M ----L----T S---QF-V-V KF------G-
E---L----- -DF----I-- K-N--KN-Q- -D-RK---FQ NLGKP---H- -----Q---- --T-T-SNI- -IQ-TV---T Q---DV-TT- EF-----M--
N-L-L----- ---------- Q-Q---S-S- ---E----S TNG-TF-L-- ----L--FF- --TLT-QSIQ VP--E----- N---TN-V-- QF---M----

171         181         191         201         211         221         231         241         251    *    261
PSIASSGATP VFDNMMSEGL VSQDLFSVYL SSDEQSGS·FV MFGGNDTSYY SGSLNWIPLS AETYWEITMD SITMNGQTIA· CSGGCQAIID TGTSLLAGPP
---S------ ----IWNQ-- ---------- -A-D----·V- I---I-S--- T-----V-VT V-G--Q--V- ------EA-·· -AE-----V- ------T--T
---SA----- ----LWDQ-- ---------- --NDD---·V- LL--I-S--- T-----V-V- V-G--Q--L- ----D-E---· --------V- ------T--T
---S------ -------QH- -A-------- -K-GET--·-- L---I-PN-T TKGIY-V--- -----Q---- RV-VGNKYV-· -FFT----V- ------VM-Q
--L-AD-I-- -----VN-S- LE-N------ -R·-PM--·M- V---I-E--F T--I----V- YQG--Q-S-- --IV-K-E--· --S------- -----V---A
--L--EYSI- ------NRH- -A-------M DR-G-ESM·LT L·-AI-P--- T---H-V-VT VQQ--QFYV- -V-IS-VVV-· -E------L- ----K-V--S
-ALSVDE--T AMQG-VQ--A LTSPV----- -NQQG-SGGA- V---V-S-L- T-QIY-A-VT Q-L--Q-GIE EFLIG--ASGW --E-----V- ------TV-Q

271         281         291         301         311         321         331         341         351         361
SDVSNINSYI GAS···NGE··VSCS SMSSLPEIVF NINGIAFPVP ASAYIINDS·S SCSSS·FESMD ·····QGLWILGDVF IRLYYVVFDR ANNQVGLASLA
-PIA--Q-D- ---ENSD-DMV---- AI----D--- T---VQY--- P----LQSE·G --I-G·-QG-N LPTESGE-------- --Q-FT---- --------PV-
-AIA--Q-D- ---ENSD--MVI--- -ID---D--- T-D-VQY-LS P----LQ-D·D --T-G·--G-- VPTSSGE-------- --Q--T---- ---K----PV-
GAYNR-IKDL -V-··SD---·I--D DI-K--DVT- H---H--TL- ----VL-ED·G --MLG·--N-G TPTELGEQ------- --E---I--- ---K---SP-S
--IND-Q-AV --NQNTY--YS-N-- HILAM-DV-- V-G--QY--- -L--TEQNGQG T-M---QNSS ·····AD-------- --V--SI--- ---R----KAI
--IL--QQA- --TQNQY--FDID-D NL-YM-TV-- E---KMY-LT P---TSQ-Q·G F-T-G·-Q-EN H····SQK------- --E--S---- ---L----KAI
QYM-ALLQAT ---EDEY-QFL-N-N -IQN--SLT- I---VE--L- P-S--LSNN·G Y-TVGVEPTYL SSQNG-P-------- L-S--S-Y-L G--R--F-TA-
```

Fig. 6. **Comparison of the amino acid sequence of turtle pepsinogen with those of some other typical pepsinogens.** For turtle pepsinogen, the tentative amino acid sequence as described in the text is used. Amino acids are shown in one-letter code. The sequences are aligned so as to maximize the homology. The numbering is that of turtle pepsinogen. Each vertical line indicates the boundary between the propart and the pepsin moiety. The boundary in chicken embryonic pepsinogen has not been determined. The positions of the active-site aspartic acids are shown with asterisks and the potential *N*-glycosylation sites in each zymogen with closed circles. Dots indicate deletions, and dashes indicate residues identical with those of turtle pepsinogen.

pepsinogens A have none. Among these sites, two sites (Asn^{102} and Asn^{115}) in chicken embryonic pepsinogen and one site (Asn^{102}) in human pepsinogen C are common to turtle pepsinogen, and two sites (Asn^{113} and Asn^{218}) are in the vicinity of Asn^{115} and Asn^{215}, respectively, of turtle pepsinogen. In these properties, turtle pepsinogen more closely resembles chicken pepsinogen A and/or chicken embryonic pepsinogen than mammalian pepsinogens A. Thus, turtle pepsinogen appears to retain certain structural features resembling those of chicken pepsinogen A and/or chicken embryonic pepsinogen, although it more closely resembles mammalian pepsinogens A in its amino acid sequence. These results are consistent with the phylogenic relationship of turtle, chicken, and mammals.

We are most grateful to Murakami Suppon Co., and Hosendo Co., Ltd., Tokyo for generous supply of the stomachs of soft-shelled turtles.

REFERENCES

1. Foltmann, B. (1981) Gastric proteinases—Structure, function, evolution and mechanism of action. *Essays Biochem.* **17**, 52-84
2. Kageyama, T. and Takahashi, K. (1980) Isolation of an activation intermediate and determination of the amino acid sequence of human pepsinogen A. *J. Biochem.* **88**, 571-582
3. Athauda, S.B.P., Tanji, M., Kageyama, T., and Takahashi, K. (1989) A comparative study on the NH_2-terminal amino acid sequences and some other properties of six isozymic forms of human pepsinogens and pepsins. *J. Biochem.* **106**, 920-927
4. Kageyama,T., Ichinose, M., Miki, K., Athauda, S.B.P., Tanji, M., and Takahashi, K. (1989) Difference of activation processes and structure of activation peptides in human pepsinogen A and progastricsin. *J. Biochem.* **105**, 12-15
5. Kageyama, T. and Takahashi, K. (1976) Pepsinogens and pepsins from gastric mucosa of Japanese monkey. Purification and characterization. *J. Biochem.* **79**, 455-468
6. Kageyama, T. and Takahashi, K. (1976) Pepsinogen C and pepsin C from gastric mucosa of Japanese monkey. *J. Biochem.* **80**, 983-992
7. Kageyama, T. and Takahashi, K. (1977) The carbohydrate moiety of Japanese monkey pepsinogen—Its composition and site of attachment to protein. *Biochem. Biophys. Res. Commun.* **74**, 789-795
8. Kageyama, T. and Takahashi, K. (1978) Monkey pepsinogens and pepsins. III. Carbohydrate moiety of Japanese monkey pepsinogens and amino acid sequence around the site of its attachment to protein. *J. Biochem.* **84**, 771-778
9. Kageyama, T. and Takahashi, K. (1980) Monkey pepsinogens and pepsins. IV. The amino acid sequence of the activation segment of Japanese monkey pepsinogen. *J. Biochem.* **88**, 9-16
10. Kageyama, T. and Takahashi, K. (1980) Monkey pepsisnogens and pepsins. V. Purification, characterization, and amino acid sequence determination of crab-eating monkey pepsinogens and pepsins. *J. Biochem.* **88**, 635-645
11. Kageyama, T. and Takahashi, K. (1982) Monkey pepsinogens and pepsins. VI. One-step activation of Japanese monkey pepsinogen to pepsin. *J. Biochem.* **92**, 1179-1182
12. Kageyama, T. and Takahashi, K. (1985) Monkey pepsinogens and pepsins. VII. Analysis of the activation process and determination of the NH_2-terminal 60-residue sequence of Japanese monkey progastricsin, and molecular evolution of pepsinogens. *J. Biochem.* **97**, 1235-1246
13. Kageyama, T. and Takahashi, K. (1987) Activation mechanism of monkey and porcine pepsinogens A. *Eur. J. Biochem.* **165**, 483-490
14. Kageyama, T. and Takahashi, K. (1983) Purification and characterization of pepsinogens and pepsins from Asiatic black bear, and amino acid sequence determination of the NH_2-terminal 60 residues of the major pepsinogen. *J. Biochem.* **94**, 1557-1567
15. Kageyama, T. and Takahashi, K. (1984) Rabbit pepsinogens. Purification, characterization, analysis of the conversion process to pepsin, and determination of the NH_2-terminal amino acid equences. *Eur. J. Biochem.* **141**, 261-269
16. 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
17. 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
18. Tanji, M., Kageyama, T., and Takahashi, K. (1988) Tuna pepsinogens and pepsins. Purification, characterization and amino-terminal sequences. *Eur. J. Biochem.* **177**, 251-259
19. 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
20. Kageyama, T. and Takahashi, K. (1986) The complete amino acid sequence of monkey pepsinogen A. *J. Biol. Chem.* **261**, 4395-4405
21. Tsukagoshi, N., Ando, Y., Tomita, Y., Uchida, R., Takemura, T., Sasaki, T., Yamagata, H., Udaka, S., Ichihara, Y., and Takahashi, K. (1988) Nucleotide sequence and expression in *Escherichia coli* of cDNA of swine pepsinogen: Involvement of the amino-terminal portion of the activation peptide segment in restoration of the functional protein. *Gene* **65**, 285-292
22. 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
23. Kageyama, T. and Takahashi, K. (1986) The complete amino acid sequence of monkey progastricsin. *J. Biol. Chem.* **261**, 4406-4419
24. 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
25. 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 gene. *J. Biol. Chem.* **264**, 10193-10199
26. 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
27. Ong, E.B. and Perlmann, G.E. (1968) The amino-terminal sequence of porcine pepsinogen. *J. Biol. Chem.* **243**, 6104-6109
28. Pederson, V.B. and Foltmann, B. (1973) The amino acid sequence of a hitherto unobserved segment from porcine pepsinogen preceding the N-terminus of pepsin. *FEBS Lett.* **35**, 255-256
29. 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
30. Kageyama, T., Tanabe, K., and Koiwai, O. (1990) Structure and development of rabbit pepsinogens. Stage-specific zymogens, nucleotide sequences of cDNAs, molecular evolution, and gene expression during development. *J. Biol. Chem.* **265**, 17031-17038
31. Baudyš, M. and Kostka, V. (1983) Covalent structure of chicken pepsinogen. *Eur. J. Biochem.* **136**, 89-99
32. 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
33. 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
34. Bohak, Z. (1969) Purification and characterization of chicken

pepsinogen and chicken pepsin. *J. Biol. Chem.* **244**, 4638-4648
35. Anson, M.L. (1938) The estimation of pepsin, trypsin, papain, and cathepsin with hemoglobin. *J. Gen. Physiol.* **22**, 79-89
36. Tang, J. (1970) Gastricsin and pepsin. *Methods Enzymol.* **19**, 406-421
37. Lowry, O.H., Rosebrough, N.J., Farr, A.L., and Randall, R.J. (1951) Protein measurement with the Folin phenol reagent. *J. Biol. Chem.* **193**, 265-275
38. Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685
39. Furihata, C., Kawachi, T., and Sugimura, T. (1972) Premature induction of pepsinogen in developing rat gastric mucosa by hormones. *Biochem. Biophys. Res. Commun.* **47**, 705-711
40. Moore, S. (1963) On the determination of cystine as cysteic acid. *J. Biol. Chem.* **238**, 235-237
41. Dubois, M., Gilles, K.A., Hamilton, J.K., Rebers, P.A., and Smith, F. (1956) Colorimetric method for determination of sugars and related substances. *Anal. Chem.* **28**, 350-356
42. Bartlett, G.R. (1959) Phosphorus assay in column chromatography. *J. Biol. Chem.* **234**, 466-468
43. Chirgwin, J.M., Przybyla, A.E., MacDonald, R.H., and Rutter, W.J. (1979) Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* **18**, 5294-5299
44. Huynh, T.V., Young, R.A., and Davies, R.W. (1985) Constructing and screening cDNA libraries in λgt10 and λgt11 in *DNA Cloning* (Glover, D.M., ed.) Vol. 1, pp. 49-78, IRL Press, Oxford
45. Grunstein, M. and Hogness, D.S. (1975) Colony hybridization: A method for the isolation of cloned DNAs that contain a specific gene. *Proc. Natl. Acad. Sci. USA* **72**, 3961-3965
46. Kuchino, Y. and Takanami, M. (1984) Methods of M 13 shotgun DNA sequencing (in Japanese). *Protein, Nucleic Acid and Enzyme* **29**, 294-306
47. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467
48. Samloff, I.M. (1969) Slow moving protease and the seven pepsinogens. Electrophoretic demonstration of the existence of eight proteolytic fractions in human gastric mucosa. *Gastroenterology* **57**, 659-669
49. Matsuzaki, O. and Takahashi, K. (1988) Improved purification of slow moving protease from human gastric mucosa and its action on the B chain of oxidized bovine insulin. *Biomed. Res.* **9**, 515-523
50. Takahashi, K. and Kageyama, T. (1985) Muliplicity 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 & New York
51. James, M.N.G. and Sielecki, A.R. (1986) Molecular structure of an aspartic proteinase zymogen, porcine pepsinogen at 1.8 Å resolution. *Nature* **319**, 33-38
52. 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
53. Hartsuck, J.A., Koelsh, G., and Remington, S.J. (1992) The high resolution crystal structure of porcine pepsinogen. *Proteins* **13**, 1-25
54. Moore, S.A., Sielecki, A.R., Chernaia, M.M., Tarasova, N.I., and James, M.N.G. (1995) Crystal and molecular structures of human progastricsin at 1.62 Å resolution. *J. Mol. Biol.* **247**, 466-485