Gastric Procathepsin E and Progastricsin from Guinea Pig

PURIFICATION, MOLECULAR CLONING OF cDNAs, AND CHARACTERIZATION OF ENZYMATIC PROPERTIES, WITH SPECIAL REFERENCE TO PROCATHEPSIN E*

(Received for publication, January 28, 1992)

Takashi Kageyama†
From the Department of Biochemistry, Primate Research Institute, Kyoto University, Inuyama, Aichi 484, Japan

Masao Ichinose, Shinko Tsukada, Kazumasa Miki, and Kiyoshi Kurokawa
From the First Department of Internal Medicine, Faculty of Medicine, The University of Tokyo, Bunkyo-ku, Tokyo 113, Japan

Osamu Koiwai‡
From the Department of Biochemistry, Institute for Developmental Research, Aichi Prefectural Colony, Kasugai, Aichi 480-03, Japan

Masao Tanji†, Etsuko Yakabe, Senarah B. P. Athauda, and Kenji Takahashi
From the Department of Biophysics and Biochemistry, Faculty of Science, The University of Tokyo, Bunkyo-ku, Tokyo 113, Japan

Procathepsin E and progastricsin were purified from the gastric mucosa of the guinea pig. They were converted to the active form autocatalytically under acidic conditions. Each active form hydrolyzed protein substrates maximally at around pH 2.5. Pepstatin inhibited procathepsin E very strongly at an equimolar concentration, whereas the inhibition was much weaker for gastricsin. Molecular cloning of the respective cDNAs permitted us to deduce the complete amino acid sequences of their pre-proforms; preprocathepsin E and preprogastricsin consisted of 391 and 394 residues, respectively.

Procathepsin E has unique structural and enzymatic features among the aspartic proteinases. Lys at position 37, which is common to various aspartic proteinases and is thought to be important for stabilizing the activation segment, was absent at the corresponding position, as in human procathepsin E. The rate of activation of procathepsin E to cathepsin E is maximal at around pH 4.0. It is very different from the pepstatins and may be correlated with the absence of Lys37. Native procathepsin E is a dimer, consisting of two monomers covalently bound by a disulfide bridge between 2 Cys37. Interconversion between the dimer and the monomer was reversible and regulated by low concentrations of a reducing reagent. Although the properties of the dimeric and monomeric cathepsins E are quite similar, a marked difference was found between them in terms of their stability in weakly alkaline solution: monomeric cathepsin E was unstable at weak alkaline pH whereas the dimeric form was stable. The generation of the monomer was thought to be the process leading to inactivation, hence degradation of cathepsin E in vivo.

The aspartic proteinase family, each member of which has 2 essential aspartyl residues at the active site, includes pepsins (pepsin A, gastricsin, and chymosin), cathepsin E, cathepsin D, and renin in mammals (reviewed in Refs. 1-3). All these enzymes are thought to have diverged from a common ancestor. Significant differences, however, have been observed in their characteristics such as hydrolytic specificity and susceptibility to inhibitors, and this is reflected in the significant variations in primary structure among members of these groups. Therefore, to understand structure-function relationships of aspartic proteinases in greater detail, it was thought to be useful to elucidate the primary structures and enzymatic properties of those aspartic proteinases that have unique characteristics. Cathepsin E represents an important example of such aspartic proteinases.

To date, it has been known that cathepsin E is a non-secretory, intracellular, but non-lysosomal proteinase. Cathepsin E has been isolated from various tissues, such as human (4-10) and rat (11) gastric mucosa, rabbit (12) and rat (13) spleen, human (14) and rat (15) erythrocyte membranes, and rat neutrophils (16). Although various designations were used previously for the enzyme, the name “cathepsin E” is used at present (16-18). Cathepsin E is a dimeric enzyme different from other aspartic proteinases. The enzyme has a molecular mass of about 80 kDa, consisting of two identical 40-kDa subunits (9-13, 16). On the other hand, the other aspartic proteinases are single polypeptides of about 40 kDa (1-3). The enzymatic properties of cathepsin E have been shown to resemble those of pepsin; for example, it has hydrolytic activity at acidic pH, with an optimum at pH 2-3, (7-9, 11, 13-16). Although the physiological role of cathepsin E is still unclear, it has been suggested to play an important role in
intracellular processing of proteins and/or peptides (19, 20), or in immune functions because of its distribution in lymphoid-associated tissue (9, 21).

Structural studies of cathepsin E have not yet progressed to a level comparable with those of pepsinogens mainly because of the difficulty in obtaining a sufficient amount of the native enzyme. Recently, the primary structure of human cathepsin E was deduced from the molecular cloning and analysis of its cDNA (22), and the presence of a pro-peptide was demonstrated by isolation and NH2-terminal sequence analysis of human gastric procathepsin E and cathepsin E (23, 24), indicating that autocatalytic activation of procathepsin E is involved in the generation of active cathepsin E (24).

The structural analysis also suggested that the dimeric form is produced by a covariant association of two monomers through a disulfide bridge(s) (25, 26). Therefore, to understand structure-function relationships of cathepsin E, it is important to clarify the differences in properties between the proenzyme and the active form and between the dimeric and monomeric forms. Further, it also seems useful to compare the primary structure of procathepsin E and some enzymatic properties of cathepsin E with those of other aspartic proteinases, especially the pepsinogens. Therefore, in the present study, guinea pig procathepsin E and progastricsin (type C pepsinogen) were chosen, as rodents are known to contain both proenzymes at high levels in gastric mucosa (11, 21) to permit simultaneous purification.

Thus, we have carried out a series of studies including purification, molecular cloning of its precursor, and elucidation of enzymatic properties of cathepsin E. The results show that the primary structure and the process of activation of procathepsin E are markedly different from those of progastricsin and other aspartic proteinases. A notable difference in enzymatic properties was also found between the dimeric and monomeric forms of cathepsin E.

**EXPERIMENTAL PROCEDURES AND RESULTS**

**Purification**—The results of the purification are summarized in Table I. Procathepsin E and progastricsin were purified simultaneously from guinea pig gastric mucosa (Fig. 5). The level of procathepsin E in gastric mucosa was the highest among the animals examined so far. On the other hand, progastricsin was the predominant pepsinogen species in guinea pig gastric mucosa. Two progastricsin components were resolved by FPLC, and they had quite similar amino acid compositions. The major component, which was eluted earlier on FPLC, was used for further characterization. Progastricsin became unstable during chromatography on the anion exchanger, in part as a result of its autocatalytic activation.

Each purified proenzyme gave a single protein band upon nondenaturing (Fig. 6) and denaturing (Fig. 3) PAGE. The molecular mass determined by SDS-PAGE under reducing conditions was about 43 kDa for each proenzyme. By contrast, the native procathepsin E was eluted at the position corresponding to a molecular mass of about 80 kDa on gel filtration and gave a band of protein with a similar molecular mass on SDS-PAGE under non-reducing conditions (Fig. 3). Therefore, procathepsin E was deduced to be a dimer. Procathepsin E is a glycoprotein, and the content of carbohydrate was estimated to be about 4% by weight. The amino acid compositions of procathepsin E and progastricsin were rather similar except for notable differences in the content of a few amino acids, such as Asp, Ser, Pro, Met, and Tyr (Table II). The NH2-terminal sequences of about 30 residues of the proenzymes were determined by Edman degradation (Figs. 1 and 8). Only a single residue was identified at each step for both proenzymes. Thus, procathepsin E appears to be composed of identical subunits. Although some common residues were observed, the NH2-terminal sequences of procathepsin E and progastricsin are quite different from those of other aspartic proteinases, especially the pepsinogens.

**FIG. 1.** The nucleotide sequence and the deduced amino acid sequence of the cDNA for guinea pig procathepsin E (pGP477). The 2 active site aspartic acids are shown by shading. The horizontal arrows show the residues determined by Edman degradation of native procathepsin E. The vertical arrowhead indicates the site of cleavage for the activation that generates cathepsin E. Residues in italics, i.e. Asn27 and Asn31, are potential N-glycosylation sites.
procathepsin E and progastricsin from Guinea Pig

Molecular Cloning of cDNAs and Structural Analysis—Among 2,000 recombinant clones of λgt10 prepared from the gastric mucosa of adult guinea pigs, about 50 clones hybridized very strongly with the radiolabeled 45-base oligonucleotide probe. Five clones were chosen at random, and the inserted DNA fragments were subcloned into pUC18 plasmid and definitively identified by sequence analysis. Since the NH₂-terminal sequences of both procathepsin E and progastricsin had been determined at the protein level, identification of clones was rather easy. Thus, these five clones were shown to be those of the cDNA for progastricsin. The restriction map and the nucleotide sequence of a typical clone (pGP461) are shown in Figs. 7 and 8, respectively. Using the cDNA for progastricsin as a probe, we rescreened the 50 clones under high stringency conditions. Two clones that did not hybridize under these conditions with the progastricsin cDNA were isolated and found by sequence analysis to be those of procathepsin E. The restriction map and the nucleotide sequence of one of the clones of procathepsin E (pGP477) are shown in Figs. 7 and 8, respectively.

The deduced amino acid sequences of the two proenzymes consist of three regions, i.e. the pre-peptide (signal peptide), the pro-peptide (activation segment), and the active enzyme. The signal peptides are composed of 19 and 16 residues, the pro-peptides are composed of 32 and 49 residues, and the NH₂-terminal region of the cathepsin E moiety as compared with the sequences of other mammalian aspartic proteinases (Fig. 2). Asn-67 and Asn-311 were found to be the potential N-glycosylation sites (Fig. 1). The molecular masses of procathepsin E and progastricsin were calculated to be 40,086 and 41,150 Da, respectively, based on the amino acid compositions deduced from the cDNAs.

Interconversion of the Dimeric and Monomeric Forms of Procathepsin E—The conversion of the dimeric procathepsin E to the monomeric form occurred in the presence of a low concentration of a reducing reagent. A typical result is shown in Fig. 9. The dimer was converted to the monomer to the extent of 30-50% by incubation of the former with 1 mM 2-mercaptoethanol, L-cysteine, or reduced glutathione at 37 °C for 20 min. The conversion was complete with any one of these reagents at 10 mM under the same conditions (Fig. 9A). The conversion was reversible, since the dimeric form was regenerated after removing the reducing reagent (Fig. 9B). Proteolytic activity was not affected by interconversion. When the monomer was carboxymethylated, carboxymethyl-Cys was determined to be 0.88 mol/mol of monomeric procathepsin E. This partially modified monomeric procathepsin E retained complete proteolytic activity. Carboxymethyl-Cys was identified at position 4 from the NH₂ terminus of cathepsin E (position 37 of procathepsin E) by Edman degradation. The result thus provided direct evidence that the dimeric form is generated by formation of a disulfide bridge at Cys⁵⁷ between the two monomers.

Activation Profile—The profile of activation of the proenzyme was analyzed by SDS-PAGE (Fig. 3). Activation of procathepsin E proceeded autocatalytically under acidic conditions; the rate of activation was maximal at pH 4.0 and decreased gradually as the pH was lowered to 2.0 (Fig. 3A). In addition, appreciable activation occurred both at pH 5.0 and 6.0 upon prolonged incubation. The rate of activation did not change when monomeric procathepsin E was activated under the same conditions. Procathepsin E appeared to be directly converted to cathepsin E, since the intermediate form(s) was generated at a very low level (Fig. 3, B and C). The bands of procathepsin E and cathepsin E were detected at positions of 82 and 76 kDa, respectively, after SDS-PAGE under non-reducing conditions, whereas the pro- and active forms gave a band of 43 kDa and a band of 39 kDa, respectively, after SDS-PAGE under reducing conditions. Therefore, the dimeric form was maintained throughout the activation. Isolation and structural analysis of the active form revealed that the site of cleavage upon activation was the Leu⁵²-Asn⁵³ bond. Thus, the NH₂ terminus of cathepsin E is located 4 residues before Cys⁶⁷. The cleavage site was the same when monomeric procathepsin E was activated under the same conditions. In addition, the profile of activation of guinea pig progastricsin was also examined (data not shown). The process was largely similar to that observed for other progastricsins (29), and the major cleavage site to generate gastricsin was the Pte⁶⁶-Ser⁶⁷ bond.

Enzymatic Properties of the Active Forms—Cathepsin E and gastricsin are optimally active at around pH 2.5 toward hemoglobin as a substrate (Fig. 10). Cathepsin E has higher specific activity than gastricsin and porcine pepsin A. Both enzymes are inhibited by pepstatin, a specific inhibitor of aspartic proteinases (Fig. 11). Susceptibility of cathepsin E to pepstatin was the same as that of porcine pepsin A, the inhibition profile indicating the strong equimolar binding of pepstatin to the active site. The susceptibility of gastricsin was about 100 times lower than that of cathepsin E and porcine pepsin A. Low susceptibility has commonly been observed with gastricsins of other animals (30, 31).

Cathepsin E is easily converted to monomers in the presence of a low concentration of a reducing reagent, as was procathepsin E as described in the preceding section. Therefore, the difference in enzymatic properties between the dimeric and the monomeric forms of cathepsin E was investigated. Although the hydrolytic activity against hemoglobin at pH 2.0 was the same for both forms, a slight increase in activity was observed with the monomer at pH 5.0 as compared to the dimer. Such an increase was not observed, however, when the enzyme was assayed with other protein substrates (Table III). By contrast, a striking difference between the dimeric and monomeric forms was found in terms of stability at weakly alkaline pH (Fig. 4). While the dimer was stable at weakly alkaline pH, the monomer lost its activity very rapidly above pH 7. On the other hand, gastricsin was very unstable at alkaline pH as reported for gastricsins of other animal sources (data not shown).

Expression of the Genes in Various Tissues—Expression of the genes for procathepsin E and progastricsin was examined in various tissues from adult guinea pigs by Northern analysis (Fig. 12). The mRNAs for both enzymes were expressed at a high level in gastric mucosa only. In addition, procathepsin E mRNA was found to be a low level in spleen. The predominant species of mRNAs of procathepsin E and gastricsin had the same size of around 1.9 kb. The size is very similar to those of pepstatin mRNAs of other mammals (32-34), but is different from that of human procathepsin E mRNA which has been shown to range from 2.2 to 3.6 kb (22).
**DISCUSSION**

Procathepsin E and progastricsin were purified from the gastric mucosa of guinea pigs. The level of procathepsin E was 4-10 times higher than that in human gastric mucosa (8, 10) and was the highest among those reported to date for various animal tissues. The reason for this high level is not clear, but it seems that the gastric mucosa of the guinea pig may serve as a good source of procathepsin E for future studies at the protein level. Progastricsin was found to be the major pepsinogen component rather than pepsinogen A. This result is consistent with the results obtained with rat stomach (32, 35).

The structures and some enzymatic properties of procathepsin E and progastricsin were determined and compared between the two proenzymes and also with those of other

---

**FIG. 2.** Comparison of the amino acid sequences of guinea pig procathepsin E and progastricsin with those of typical mammalian aspartic proteinases. A, F, Ch, C, E, D, and R stand for pepsinogen A, pepsinogen F, prochymosin, progastricsin, procathepsin E, procathepsin D, and prorenin, respectively. Amino acid sequences were deduced from the cDNA or genomic DNA sequences of monkey pepsinogen A (34), rabbit pepsinogen F (33), bovine prochymosin (49), rat progastricsin (32), human procathepsin E (22), human procathepsin D (50), and human prorenin (51). The numbering is based on the sequence of monkey pepsinogen A. Bars represent deleted residues. The italic letters around positions 1 and 50 show the NH₂-terminal amino acids of the proenzymes and the active forms, respectively. Human procathepsin E has been shown to be cleaved at two adjacent sites (Met-Ile and Phe-Thr bonds) upon activation (23, 24, and S. B. P. Athauda et al. manuscript submitted). The NH₂-termini of rabbit pepsinogen F and pepsin F have not been determined and are not indicated. Residues in filled boxes are conserved in mammalian aspartic proteinases and indicated under the aligned sequences. Common residues between guinea pig and human procathepsins E and between guinea pig and rat progastricsins are shown by deep and faint shading, respectively. Asterisks show the aspartic acids at the active sites.

---

**Procathepsin E and Progastricsin from Guinea Pig**

16453
**Procathepsin E and Progastricsin from Guinea Pig**

**A**

![Image of SDS-PAGE gel showing pH variation for procathepsin E](image)

**B**

![Image of SDS-PAGE gel showing time course of activation](image)

**C**

![Image of SDS-PAGE gel showing reducing conditions](image)

**Fig. 3. Analysis of the time course of activation of procathepsin E under various conditions.** Activation was carried out at an initial concentration of procathepsin E of 0.05 mg/ml at 14 °C. A, activation at various pH values was analyzed by SDS-PAGE under non-reducing (−) and reducing (+) conditions. The time of activation reaction was 15 min. The right half of the band of dimeric procathepsin E in lane 5 (from the left) migrated to the position of monomeric procathepsin E as a result of the effect of 2-mercaptoethanol in lane 6. B and C, time course of activation at pH 2.0 analyzed by SDS-PAGE under non-reducing conditions (B) and reducing conditions (C). PCE, CE, and Int stand for procathepsin E, cathepsin E, and the intermediate form(s), respectively.

**Fig. 4. Stability of procathepsin E and cathepsin E at various pH values.** The native dimeric forms of procathepsin E (○) and cathepsin E (△) were incubated at 28 °C for 2 h in 0.1 M buffer at various pH values (4.0, sodium formate; 5.0, sodium acetate; 6.0 and 7.0, sodium phosphate; 8.0 and 8.8, Tris-HCl). The residual activity was determined by the hemoglobin-digestion method. The details of the method are shown in the legend to Table I. The monomeric forms of procathepsin E (□) and cathepsin E (▲), generated by treatment with 2-mercaptoethanol, were subjected to the same analysis. The assay mixture of the monomeric form contained 0.2 mM 2-mercaptoethanol to prevent regeneration of the dimer.

aspartic proteinases. The unique characteristics of procathepsin E were clarified by these studies. It is noteworthy that lysine 37 (numbering based on the sequence of monkey pepsinogen A, see Fig. 2) of the propeptide (activation segment), which is present in other mammalian aspartic proteinase zymogens, was not found in guinea pig nor in human procathepsin E. The positive charge of the lysine residue has been shown to provide electrostatic stabilization via hydrogen bonding to one of the net negative charges of the two aspartic acids at the active site (27, 28). Therefore, the lysine residue has been suggested to be essential for maintaining the proenzyme in an inactive form, thereby playing an important role in the activation of these aspartic proteinases. The activation of procathepsin E proceeded most rapidly at pH 4.0, and appreciable activation occurred at even higher pH. This phenomenon was markedly different from the pepsinogens which are activated most rapidly at pH 2.0 and below (36). The maximal activation at weakly acidic pH may be correlated with the absence of Lys37 in procathepsin E, since electrostatic stabilization is thought to be weak in procathepsin E. Since procathepsin E is a non-secretory intracellular proteinase and since its activation would occur at physiological pH, the maximum rate of activation at weakly acidic pH seems to be well adapted to the physiology of the proenzyme. Lys37 was conserved in guinea pig progastricin.

When the sequences of the connecting region of the pro-
peptide and the cathepsin E moiety of guinea pig and human procathepsin E were compared with those of other aspartic proteinases, deletions of several residues around Cys appears to be significant (Fig. 2). The cleavage sites associated with activation in other aspartic proteinases, in particular in peptidinogens A and progastricsins, are located in this area (29) and indicate a high degree of conformational lability (28). Therefore, if the deleted positions of procathepsin E were actually occupied by amino acids, cleavage might occur at these sites after CyP, resulting in the generation of monomeric cathepsin E after activation. Therefore, the deletions may be essential for the cleavage before Cys and, thus, for maintaining the dimeric form via a disulfide bond during activation. In progastricsins, the activation segment is composed of 49 residues, the longest among known sequences of mammalian aspartic proteinases. The role of this extended segment, however, remains to be clarified, since the cleavage site for activation is the same as that of rat progastricsin, which has a shorter segment of 46 residues.

With respect to the structure of the cathepsin E moiety, the common residues among other aspartic proteinases, including those around the 2 aspartic acid residues of the active site, are well conserved (Fig. 2). One notable point is that the higher level of homology associated with cathepsin D, the other intracellular aspartic proteinase (Table II). The level is comparable to that in pepsins and gastricsins and this similarity may be correlated with the optimal activity at lower pH.

The structure of guinea pig procathepsin E is most similar to that of human procathepsin E with 86 and 84% identity at the nucleotide and the amino acid levels, respectively (Fig. 2). The identity with other aspartic proteinases is less than 60%. The evolutionary relationships among various gastric aspartic proteinase zymogens, including pepsinogens A, prochymosins, and progastricsins, have been deduced (33, 37, 38). However, the relationship between procathepsin E and these gastric aspartic proteinases and other non-pepsin-type aspartic proteinases, such as cathepsin D and renin, have not been elucidated. Therefore, we constructed a phylogenetic tree to examine the relationships among various aspartic proteinases including procathepsin E (Fig. 13). The tree shows clearly that procathepsin E is closer to pepsinogens than are procathepsin D and prorenin.

The generation of a dimer is characteristic of (pro)cathepsin E. The present results provide direct evidence that a disulfide bridge involving Cys between the two monomers is responsible for generating the dimer. The interconversion between the dimer and the monomer is reversible; the dimer is easily converted to the monomer in the presence of a low concentration of a reducing agent (Refs. 25 and 26, Fig. 9A), and the monomer is forced to regenerate the dimer in the absence of a reducing agent (Fig. 9B). Such high susceptibility to a reducing agent is thought to be due to the tertiary structure of procathepsin E, in which the region around Cys is presumed to be on the surface of the protein as expected from the tertiary structures of other aspartic proteinases (27, 28). Therefore, it may be reasonable to consider that a reducing agent, such as glutathione, may regulate the interconversion between the two forms in vivo. Indeed, the occurrence of the monomeric form of procathepsin E has been detected in human gastric mucosa. The interconversion between the two forms may have little significance in the case of procathepsin E, since no difference in properties was found between the two forms. On the other hand, the conversion of the dimer to the monomer seems critical in the case of cathepsin E.
Proteasomes E and Procathepsin from Guinea Pig

Biochim. Biophys. Acta 884, 607-609
21. Muto, N., Yamamoto, M., Tanii, S., and Yonezawa, S. (1988) Biochim. Biophys. Acta 925, 161-168.
22. Azuma, T., Tsai, H., Mahdas, T. K., Couvreur, J. M., and Taggart, R. T. (1989) Biochim. Biophys. Acta 924, 162-163.
23. Atsumi, S. B. F., Matsuoka, S., Kageyama, R., and Takahashi, K. (1990) Biochim. Biophys. Acta 1019, 9-19.
24. Yamamoto, M., Tanii, S., and Yonezawa, S. (1990) Biochim. Biophys. Acta 1019, 9-19.
25. Yonezawa, S., Fujii, K., Maejima, Y., Yamato, K., Mori, Y., and Muto, N. (1990) Arch. Biochem. Biophys. 267, 176-185.
26. Yamamoto, K., Kageyama, H., Takesue, M., and Kato, Y. (1988) Bioclim. Biophys. Acta 924, 162-163.
27. James, M. N. G., and Sherratt, R. A. (1980) Nature 19, 33-38.
28. Sieskii, A. R., Fujimura, M., Rend, R. J., and James, M. N. G. (1991) J. Biol. Chem. 241, 671-692.
29. Takemaki, K., and Kageyama, K. (1985) in Aspartic Proteinases and Their Inhibitors (Kotek, V., ed) pp. 265-282, Walter de Gruyter, Berlin.
30. Kageyama, T., and Takahashi, K. (1976) Biochem. (Tokyo) 80, 893-992.
31. Tang, J. (1970) in Enzyme Microassay (1970) Academic Press, New York.
32. Ichihara, Y., Sogawa, K., Moroshki, K., Fujii-Kurya, Y., and Takahashi, K. (1986) Eur. J. Biochem. 161, 7-12.
33. Kageyama, T., Tanabe, K., and Kawai, O. (1990) J. Biochem. 199, 703-7058.
34. Kageyama, T., Tanabe, K., and Kawai, O. (1991) Eur. J. Biochem. 202, 205-215.
35. Furnerits, C., Saio, D., Fujii, H., Kashi, Y., Matsushita, T., and Sugimura, T. (1990) Eur. J. Biochem. 195, 45-50.
36. Al-Janabi, J., Harstuck, A. J., and Tang, J. (1992) J. Biochem. 247, 4659-4662.
37. Kageyama, T., and Takahashi, K. (1986) J. Biol. Chem. 261, 400-419.
38. Tanii, M., Kageyama, T., and Takahashi, K. (1986) Eur. J. Biochem. 177, 261-268.
39. Horii, R. (1992) J. Gen. Physiol. 45, 57-76.
40. McPhie, P. (1989) Biochem. Biophys. Res. Commun. 158, 115-119.
41. Atsumi, S. B. F., Matsuoka, S., Inzue, H., Ichimura, M., and Kageyama, K. (1991) FEBS Lett. 292, 53-56.
42. Lei, K.-H., Wyckoff, J. B., and Samloff, I. M. (1988) Gastroenterology 95, 2005-2012.
43. Hirsch-Marie, H., Lesiiffier, F., Toublout, J. P., and Burtin, P. (1984) Lab. Invest. 43, 623-632.
44. Sakai, T., Sakai, H., Shibata, K., Kato, Y., and Yamamoto, K. (1991) J. Biochem. (Tokyo) 110, 956-964.
45. Aruma, T., Liu, W., Vanslye, D., Bowcock, A. M., and Taggart, R. T. (1993) J. Biol. Chem. 268, 1609-1614.
46. Furnerits, C., Kawachi, T., and Sugimura, T. (1972) Biochim. Biophys. Acta 267, 705-711.
47. Ichihara, T., Ichihara, M., Hayano, T., Katsu, K., Sogawa, K., Fuji-Kurya, Y., and Takahashi, K. (1989) J. Biol. Chem. 264, 1095-1100.
48. Ichihara, M., Kikai, M., Ichihara, C., Terasaw, K., Kikai, Y., Ichihara, M., Ishihara, A., Katsu, K., Sogawa, K., Fuji-Kurya, Y., and Takahashi, K. (1989) Cancer Res. 49, 1603-1608.
49. Harada, M., Sasaki, K., Tomii, T., and Beppu, T. (1989) Gen. Genet. 143, 197-203.
50. Furne, L., Kornfield, S., and Chigwin, J. M. (1985) Proc. Natl. Acad. Sci. U.S.A. 82, 4910-4914.
51. Imat, T., Miyaski, H., Hirose, S., Hori, H., Hayashi, K., Kageyama, R., Okabe, T., Nakamura, S., and Kurokawa, K. (1985) Proc. Natl. Acad. Sci. U.S.A. 80, 7405-7409.
52. Azuma, T., I. M. (1978) J. Genet. Physiol. 22, 79-89.
53. de Bernardis, S., Weigmann, T., Meurer, Y., Manhart, K., Leunberger, W., Bohlen, P., Stein, S., and Udenfried, S. (1974) Arch. Biochem. Biophys. 183, 280-292.
54. Orzstein, L. (1984) Ann. N. Y. Acad. Sci. 121, 321-349.
55. Davis, B. J. (1984) Ann. N. Y. Acad. Sci. 411, 404-437.
56. Laskin, K. L. (1970) Nature 227, 890-892.
57. Speckman, D. H., Stein, W. H., and Moore, S. (1958) Anal. Chem. 30, 1190-1200.
58. Edman, P., and Henschel, A. (1975) in Protein Sequence Determination (Needleman, S. B., ed) 2nd ed., pp. 237-250, Springer-Verlag, Berlin.
59. Dubois, M., Gilles, A. K., Hamiton, J. K., Rebers, P. A., and Smith, F. (1956) Anal. Chem. 28, 350-356.
60. Kageyama, T., and Takahashi, K. (1982) J. Biochem. (Tokyo) 92, 1179-1187.
61. Kageyama, T., and Takahashi, K. (1982) J. Biochem. 1019, 483-490.
62. Chirgwin, J. M., Frabry, A. E., Macdonald, R. J., and Rutter, W. J. (1979) Biochemistry 18, 5294-5299.
63. Aviv, H., and Leder, P. (1972) Proc. Natl. Acad. Sci. U.S.A. 69, 1408-1410.
64. Gubler, U., and Hoffman, J. B. (1983) Gene (Amst.) 25, 283-299.
65. Aviv, H., Leder, P. (1972) Proc. Natl. Acad. Sci. U.S.A. 69, 1408-1410.
66. Gubler, U., and Hoffman, J. B. (1983) Gene (Amst.) 25, 283-299.
67. Chirgwin, J. M., Frabry, A. E., Macdonald, R. J., and Rutter, W. J. (1979) Biochemistry 18, 5294-5299.
68. Gubler, U., and Hoffman, J. B. (1983) Gene (Amst.) 25, 283-299.
69. Gubler, U., and Hoffman, J. B. (1983) Gene (Amst.) 25, 283-299.
70. Gubler, U., and Hoffman, J. B. (1983) Gene (Amst.) 25, 283-299.
71. Gubler, U., and Hoffman, J. B. (1983) Gene (Amst.) 25, 283-299.
72. Gubler, U., and Hoffman, J. B. (1983) Gene (Amst.) 25, 283-299.
73. Gubler, U., and Hoffman, J. B. (1983) Gene (Amst.) 25, 283-299.
74. Gubler, U., and Hoffman, J. B. (1983) Gene (Amst.) 25, 283-299.
75. Gubler, U., and Hoffman, J. B. (1983) Gene (Amst.) 25, 283-299.
complementary DNA was prepared by the procedure of Gubler and Hoffman (64) using a cDNA-synthesis kit. After methylation of the internal EcoRI sites of Fc0R1 linkers, the cDNA was fractionated according to size by sedimentation through a 20 Ng ml of glycerol. One aliquot was digested with 1 kb to 3 kb in length and ligated into the EcoRI site of pBR322. The phages were screened, and the recombinants were selected by plating on E. coli cells. The recombinant plasmids were isolated from cultures of the E. coli cells and digested with EcoRI. After gel electrophoresis and Southern hybridization, the sequence of the oligonucleotide probe was identified by the method of Sambrook et al. (65). The hybridization was carried out at high-stringency conditions. The mobility of fragments of a DNA generated by digestion with HindIII. After the RNA had been transferred to nitrocellulose paper, the paper was exposed to X-ray film for 24 h. The absorbance of the filtrate was measured at 280 nm. One unit of activity was defined as the amount of enzyme that caused an increase in absorbance of 1% per min.

**TABLE I**

| Step | Protein | Activity (units) | Specific activity (units/mg protein) | Yield (%) |
|------|---------|-----------------|-------------------------------------|-----------|
| 1. Supernatant of crude homogenate | 176 | 161 | 0.91 | 100 |
| 2. DEAE-Sepharose | | | | |
| Procathepin E | 6.5 | 31 | 4.6 | 19 |
| Progastxin | 1.1 | 85 | 7.7 | 53 |
| 3. Purification of procathepin E | | | | |
| Sephadex 0-100 | 2.0 | 20 | 10 | 12 |
| FPLC | 0.67 | 15 | 22 | 9.3 |
| 4. Purification of progastxin | | | | |
| Sephadex 0-100 | 4.5 | 49 | 11 | 30 |
| FPLC | 1.4 | 27 | 19 | 17 |
| 2 | 0.92 | 16 | 17 | 9.5 |

**TABLE II**

Amino acid compositions of guinea-pig procathepin E and progastxin, as compared with those of other proteins. The homology of the amino acid sequences of the guinea-pig syngenes is based on the amino acid sequences deduced from nucleotide sequences. The numbers in parentheses are those determined by amino acid analysis of the purified syngenes. nd, not detectable due to decomposition during acid hydrolysis.

| Amino acid | Number of residues per molecule |
|------------|-------------------------------|
| Procathepin E | Progastxin | HE | MA | HD | GR |
| Asp | 15 (35) | 19 (28) | 20 | 23 | 21 | 18 |
| Asn | 17 | 8 | 15 | 17 | 11 | 12 |
| Thr | 14 (28) | 27 (27) | 22 | 26 | 23 | 25 |
| Ser | 33 (32) | 41 (41) | 40 | 46 | 34 | 35 |
| Gly | 23 (44) | 19 (46) | 24 | 19 | 15 | 13 |
| Pro | 24 (27) | 17 (17) | 25 | 21 | 31 | 17 |
| Gln | 40 (37) | 40 (40) | 39 | 38 | 39 | 37 |
| Ala | 21 (31) | 12 (16) | 22 | 20 | 17 | 17 |
| Val | 31 (31) | 25 (25) | 30 | 27 | 34 | 24 |
| Met | 18 (2) | 5 (2) | 3 | 4 | 9 | 11 |
| Ile | 19 (19) | 19 (19) | 20 | 29 | 23 | 24 |
| Leu | 33 (35) | 39 (40) | 29 | 30 | 23 | 35 |
| Tyr | 14 (14) | 22 (22) | 13 | 18 | 19 | 17 |
| Phe | 10 (19) | 10 (19) | 10 | 17 | 23 | 19 |
| His | 5 (6) | 5 (6) | 5 | 3 | 5 | 3 |
| Arg | 9 (13) | 9 (13) | 10 | 9 | 5 | 5 |
| Trp | 5 (38) | 5 (38) | 5 | 4 | 5 | 4 |
| Total | 372 | 376 | 379 | 373 | 392 | 383 |

**Fig. 5. Chromatography of a sample of crude extract of guinea-pig gastric mucosa on a column of DEAE-Sepharose.** The column (1.2 x 25 cm) was equilibrated with 0.1 M sodium acetate buffer, pH 5.0. The proteins were resolved by elution with NaCl at pH 5.0. The absorbance at 280 nm. Fractions under the bars were pooled. E, procathepin E; C, progastxin.

**Fig. 6.** Monitoring PAGE of purified procathepin E and progastxin. The concentration of acrylamide was 10%. The bottom of each gel corresponds to the position of bromphenol blue. Proteins were stained with Coomassie brilliant blue R-250. E, Procathepin E; C, progastxin (first peak on FPLC).
Procathepsin E and Progastricsin from Guinea Pig

Fig. 7. Restriction maps of the cDNA inserts of the two plasmids and the strategy for nucleotide sequencing. The open-reading frame encoding the primary structure of each proenzyme is shown by an open bar. A filled bar indicates the signal sequence. Lines represent the 5'- and 3'-untranslated regions. The sequencing strategy is shown for pGP477. Arrows indicate the direction and extent of sequencing of fragments subcloned in M13mp18 or M13mp19.

Fig. 8. The nucleotide and deduced amino acid sequences of cDNA for progastricsin (pGP461). The symbols are the same as those used in Fig. 1. The position of the polyadenylation signal sequence, AATAAA, in the 3'-untranslated region is underlined.

Fig. 9. Interconnection between the diurnal and the monomeric procathepsin E. D and M indicate the positions of the diurnal and the monomeric lines, respectively. (A) FPC analysis of the conversion of the diurnal to the monomer. A column of hydroxylapatite 12 × 10−2 M was equilibrated with 0.01 M sodium phosphate buffer, pH 6.9, that contained 0.1 M NaCl. The flow rate was 0.4 ml/min. (a) Native procathepsin E (B), (c), and (d) procathepsin E dissolved in 25 mM sodium phosphate buffer, pH 7.0, was incubated with 2-mercaptoethanol at concentrations of 1 mM, 5 mM, and 10 mM, respectively, at 37°C for 0.5 h. The same elution profile when procathepsin E was treated with 2-mercaptoethanol or reduced gluathionine. (B) Separation of the diurnal and the monomeric procathepsin E by 2-mercaptoethanol treatment was precipitated by adding solid ammonium sulfate to 0.7 saturation. The precipitate was washed three times with 100 mM of 0.7 saturated solution of ammonium sulfate, and then dissolved in 20 μl of 25 mM Tris-HEPES buffer, pH 6.1. The solution was kept at 4°C. Aliquots were withdrawn just after the precipitate was dissolved in Tris-HEPES buffer (a) and at 24 h (b), and subjected to SDS-PAGE under non-reducing conditions. PCE, procathepsin E before 2-mercaptoethanol treatment.

Fig. 10. The effect of pH on the hemoglobin-digesting activity of cathepsin E (A), gastricsin (B), and porcine pepsin (C). Hemoglobin (4 μl of a 5 mg/ml solution) was incubated at various pH values with 0.15 μg of enzyme at 37°C for 1.5 h. Cathepsin E was used in the native diurnal form.
Fig. 11. Inhibition of the activities of cathepsin E (E), gastric cathepsin (C), and porcine pepsin (P) by pepstatin. Cathepsin E and gastricsin were incubated in 50 mM sodium acetate buffer, pH 5.5, with pepstatin at various concentrations at 30°C for 20 min. The residual activity was determined by the hemoglobin-digestion method. Although cathepsin E was used in the native dimeric forms, molar concentration was calculated based on the amount of the monomeric form.

Fig. 12. Northern blot hybridization of total RNA from various guinea-pig tissues. The amount of total RNA applied was 5 µg in each lane. The 32P-labeled cDNAs for procathepsin E (pGP4771A) and progastricsin (pGP4611B) were used for detection of mRNA under high-stringency conditions. Lanes: 1, small intestine; 2, spleen; 3, liver; 4, stomach; 5, large intestine.

Fig. 13: A phylogenetic tree of mammalian aspartic proteinases based on their nucleotide substitutions. The nucleotide sequences of the protein-encoding regions were compared. A complete comparison was achieved by introducing deletions at appropriate positions, as shown in Fig. 2. The matrix of the mutation distances for all possible pairings was calculated by the method of Fitch and Margoliash (1971). The tree was constructed from the matrix by the method of Fitch and Margoliash (1971) with the aid of a computer. The lengths of the branches indicate the mutation distances. The standard deviation between two matrices of the original mutation distances and the reconstructed mutation distances is 1.7%. Abbreviations of aspartic proteinases are the same as those in Fig. 2.