J. Biochem. 123, 693-700 (1998)

Functional Expression and Enzymatic Properties of Two *Sitophilus zeamais* Cysteine Proteinases Showing Different Autolytic Processing Profiles *In Vitro*

Ichiro Matsumoto,*[,1] Keiko Abe,* Soichi Arai,* and Yasufumi Emori†

Department of Applied Biological Chemistry, Graduate School of Agricultural and Life Sciences, and †Department of Biophysics and Biochemistry, Faculty of Science, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657

Received for publication, October 30, 1997

To characterize in more detail the cathepsin L-like cysteine proteinases from *Sitophilus zeamais* (SCPs) cloned in our previous study [Matsumoto *et al.* (1997) *J. Biochem.* 121, 464-476], we established a system for their functional expression and purification using a glutathione S-transferase (GST) fusion gene vector from *Escherichia coli*. The proenzyme forms of two representative SCPs, proSCPc1 and proSCPg3, were expressed as GST-fusion proteins and purified on a glutathione Sepharose column. GST-proSCPc1 undergoes autoproteolytic cleavage into the mature form efficiently at acidic pH, and exhibits significant proteolytic activity toward various substrates including hemoglobin and Z-Phe-Arg-MCA. The enzymatic characteristics of the activated form of SCPc1 are similar to those of mammalian cathepsin L, but its pH optimum for the hydrolysis of hemoglobin is significantly lower. The other proSCP, GST-proSCPg3, which has a shorter COOH-terminal domain than SCPc1, undergoes almost no autolytic processing and shows only very slight proteolytic activity, although the other enzymatic characteristics of GST-proSCPg3 are similar to those of GST-proSCPc1.

Key words: cysteine proteinase, recombinant enzyme, *Sitophilus zeamais*.

The papain family is the largest of all cysteine proteinase (CP) superfamilies, comprising the papain, calpain, and bleomycin hydrolase families (*1*). Papain family members exhibit various amino acid sequence homology, but contain a common catalytic triad sequence involving the active site Cys residue. Cathepsin L is a member of the papain family, and related CPs have been widely detected in eukaryotes including plants, invertebrates, and vertebrates (*1*). In mammals, cathepsin L is an abundant CP in lysosomes where it catalyzes the proteolysis of both endocytosed and endogenous proteins (*2*). Cathepsin L is also secreted and participates in extracellular events such as bone resorption (*3*). Until recently, knowledge about cathepsin L and its relatives was mostly restricted to proteins of mammalian origin. However, various CPs, especially those in insects, are now under investigation, marking a new era as to the physiological importance of cathepsin L and/or cathepsin L-like CPs (*4-10*).

Cathepsin L-like CPs have been identified at the cDNA or gene level in four species of insects. These comprise the *Sarcophaga* 50 kDa proteinase (Sar50k) in the dipteran, *Sarcophaga peregrina* (*4*), *Bombyx* CP (BCP) in the lepidopteran, *Bombyx mori* (*5*), *Drosophila* CP1 (DCP1) in the dipteran, *Drosophila melanogaster* (*6*), and three *Sitophilus* CPs (SCPs) in the coleopteran, *Sitophilus zeamais* (*7*). Amino acid sequence comparisons among these CPs showed their primary structures are highly homologous to one another as well as to that of mammalian cathepsin L. The tissue distributions of these CPs have also been investigated; DCP1 and SCPs occur in various tissues, including alimentary organs and oocytes (*6, 7*), BCP occurs in several tissues, such as ovaries and fat bodies (*8, 9*), and Sar50k is secreted from leg imaginal discs (*4*). The physiological functions of these CPs are suggested by their tissue distributions. For example, DCP1 and SCPs may play roles in digestion, germ cell differentiation and embryonic development (*6, 7*), Sar50k may be involved in the differentiation of the imaginal discs (*4*), and BCP is likely to play a role in the programmed degradation of yolk proteins in early embryos (*10*). However, there remains little information concerning the biochemical characteristics of these insect CPs, except that they are stored in part in their proenzyme forms (*7, 10*), and BCP is activated through autolysis to hydrolyze substrates at an acidic pH around 3.5 (*9*). Although the process of the autolytic activation of BCP is similar to that of mammalian procathepsin L (*10, 11*), the pH optimum of BCP activity is significantly lower than that of mammalian cathepsin L (*12*). However, it is not yet known whether the pH dependency of BCP is intrinsic to all insect cathepsin L-like CPs, or whether the catalytic properties of BCP, such as substrate specificity, are common to other insect CPs.

[1] To whom correspondence should be addressed. Tel: +81-3-3812-2111 (Ext. 5129), Fax: +81-3-5802-8897

Abbreviations: PAGE, polyacrylamide gel electrophoresis; PMSF, phenylmethane sulfonyl fluoride; Z-, benzyloxycarbonyl; -MCA, 4-methyl-7-coumarylamide; E-64, *N*-(L-3-carboxy-2,3-*trans*-carboxyoxirane-2-carbonyl)-L-leucyl-(4-guanidino)-butane.

For more detailed determination of the biochemical features of SCPs, the establishment of a recombinant expression system capable of dissecting multiple SCPs encoded by four to six distinct genes (7) is essential. Although SCPs exhibit more than 90% similarity, they can be divided into two groups, one with a COOH-terminal region similar to other cathepsin L-like CPs, and the other with a COOH-terminal domain that is shorter by seven amino acids. Here, we chose SCPc1 and SCPg3 as representatives of the former and latter groups, respectively, and used them in expression experiments. SCPc1 has been shown to be encoded by the major mRNA species, while SCPg3 has been identified only at the gene level (7).

There have been many reported CP expression studies on several mammalian CPs (*13-19*) and one plant CP (papain) (*20-24*). The mature and/or precursor forms of cathepsins B and S and papain have been recovered as active CPs using yeast expression systems (*16, 17, 24*). These systems, however, involve several problems, such as the low yield of recombinant CP (0.1 mg per 1 liter of culture), and hyper- and aberrant glycosylation, that result in molecular variations of the recombinant proteins. On the other hand, bacterial expression systems have also been adopted for CPs. Smith and Gottesman (*14*) succeeded in producing human procathepsin L in *Escherichia coli*. However, proteolytic activity of the recombinant enzyme, which has essentially the same characteristics as native human cathepsin L, was obtained only after long and laborious denaturation and renaturation procedures for the expressed cathepsin L, which was initially recovered as insoluble inclusion bodies. It is thus important to establish an expression system that generally and reproducibly allows one to obtain active enzymes.

In the present study, we established a functional expression and purification system for SCPs produced in *E. coli* as fusion proteins with glutathione S-transferase (GST). We also report the biochemical characterization of two species of SCPs; SCPc1 and SCPg3, which show different autolytic processing activities but similar substrate specificities.

MATERIALS AND METHODS

Construction of Expression Plasmids—An oligonucleotide sense primer corresponding to the NH_2-terminal Ala^{16}-Asp^{21} region of the proenzyme forms of SCPs (7) was synthesized with a DNA synthesizer (Oligo 1000, Beckman). The nucleotide sequence was 5′-AAGGGATCCCTTTCTACGATTTAG-3′, with four nucleotide replacements to generate the underlined *Bam*HI site (Fig. 1A). Specific antisense primers for SCPc1 and SCPg3 corresponding to their 3′-noncoding regions were also synthesized with the sequences, 5′-GAAGAATGAGAATCATAG-3′ and 5′-CGTTAGCATTATTACCCA-3′, respectively. The DNA fragments encoding the proenzyme forms were amplified using these primers and *Pfu* DNA polymerase (Stratagene), SCP cDNA and genomic DNA (7) being used as templates. The amplified DNA fragments were digested with *Bam*HI, and then inserted into the polylinker site downstream of the glutathione S-transferase (GST) gene of pGEX-3X (Pharmacia Biotech) in the same orientation and in-frame with GST. The constructions encoding the GST-proSCP fusion proteins were confirmed by sequencing with a DNA sequencer (Model 373A, Applied Biosystems).

Expression and Purification of SCPc1 and SCPg3 as Glutathione S-Transferase Fusion Proteins—The expression plasmids constructed above were introduced into *E. coli* AD202 kindly provided by Prof. T. Saito (*25*). Bacteria harboring the recombinant plasmids were cultured in LB medium containing 100 μg/ml ampicillin until the turbidity at 560 nm reached 0.5. The culture was then treated with 1 mM isopropyl-β-D-thiogalactopyranoside for 2 to 3 h. The bacterial cells were collected by centrifugation, washed once with 10 mM Tris/HCl buffer, pH 8.0, containing 1 mM EDTA, suspended in the same buffer containing 1% *N*-lauroylsarcosinate, and then sonicated. A one-fifth volume of 10% Triton X-100 was added, followed by mixing with a Vortex mixer and centrifugation at 18,000 × *g* for 5 min at 4°C. The supernatant was loaded onto a glutathione-Sepharose 4B column (Pharmacia Biotech). The column was washed with PBS containing 0.2% Triton X-100 and then with PBS. The GST fusion proteins were eluted with 10 mM glutathione in 50 mM Tris/HCl (pH 8.0). The purified fusion proteins were analyzed by SDS-PAGE (*26*) and stored at 0°C.

Preparation of Crude Extract of Adult S. zeamais—A crude extract of *S. zeamais* was prepared essentially as described (7). In brief, adult insects were homogenized in 10 mM Tris/HCl buffer (pH 7.5) containing 1 mM EDTA and 2 mM β-mercaptoethanol, and then the homogenate was centrifuged to remove debris. To the supernatant, an equal volume of saturated ammonium sulfate in extraction buffer was added. The precipitate was collected by centrifugation, dissolved in a small amount of extraction buffer, and then dialyzed against the same buffer. Pepstatin and PMSF were each added to this crude SCP fraction to a final concentration of 1 mM, and the extract was stored at 4°C.

In Vitro Autolytic Processing of proSCPs—Samples of crude extract or purified GST-proSCPs were incubated for up to 30 min at 37°C in the following buffers containing 0.2% bovine hemoglobin (Sigma): 0.1 M glycine/HCl (pH 2.5-3.0), 0.1 M sodium acetate (pH 3.5-5.5), and 0.1 M MES/NaOH (pH 6.0-7.0). Reactions were terminated by adding an equal volume of 2 × SDS loading buffer composed of 0.1 M Tris/HCl (pH 6.8), 0.5 M β-mercaptoethanol, 4% SDS, 20% glycerol, and 0.1% BPB. The samples were subjected to SDS-PAGE (*26*) and then electroblotted onto a PVDF membrane (Immobilon-P, Millipore) (*27*), which was then immunoreacted with an anti-SCP antibody (1 : 500 dilution) as described previously (7).

NH_2-Terminal Sequencing of the Autolyzed Form of Recombinant proSCPc1—Purified GST-proSCPc1 was incubated in 0.1 M glycine/HCl buffer, pH 3.0, at 37°C for 2 min, and the reaction was terminated by adding an equal volume of 2 × SDS loading buffer. The sample was electrophoresed on a 10% SDS-polyacrylamide gel and then electroblotted onto a PVDF membrane. The band of autoprocessed SCPc1 was visualized by staining with Coomassie Brilliant Blue R-250 and then excised. The NH_2-terminal sequence of the excised protein was determined using an automated protein sequencer (Model 473A, Applied Biosystems).

SCP Assays—The standard enzyme activities of the recombinant SCPs were assayed using 0.2% hemoglobin. An aliquot of GST-proSCPs was incubated with the substrate in 0.1 M sodium acetate buffer (pH 3.5) at 37°C for

30 min, and the reaction was terminated by adding an equal volume of 10% trichloroacetic acid and mixing. The mixture was allowed to stand on ice for 15 min, after which the precipitate was removed by centrifugation. Activities were assayed by measuring the absorbance at 280 nm of the supernatants corrected by subtracting the value at time zero. In some cases, casein and BSA were used as substrates instead of hemoglobin.

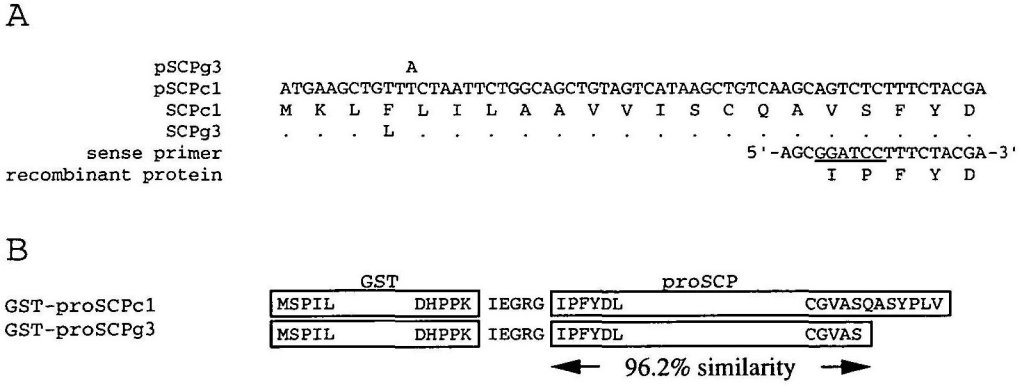
When the fluorescent substrates, Z-Phe-Arg-MCA, Z-Arg-Arg-MCA, and Arg-MCA (Peptide Institute), were used, the reactions were carried out as described above using 4 μM of each peptidyl-MCA at pH 3.5. The reactions were terminated by adding an equal volume of 2% HCl in ethanol. Fluorescence was measured using an FP-777 fluorescent spectrometer (Jasco), with excitation and emission wavelengths of 370 and 460 nm, respectively.

pH Dependent Hydrolytic Activity—The hydrolytic activities of the recombinant SCPs were assayed under various pH conditions using 0.2% hemoglobin or 4 μM Z-Phe-Arg-MCA as the substrate. The buffers used and other reaction conditions were the same as for the *in vitro* autolytic processing experiment.

Inhibitory Effects of Various Proteinase Inhibitors against SCPc1—The inhibitory effects of various chemical and proteinaceous inhibitors against SCPc1 were measured essentially under the same conditions as employed for the standard assay. The chemically synthesized inhibitors used were E-64 (10 μM) (Peptide Institute), leupeptin (10 μM) (Peptide Institute), antipain (0.1 mM) (Sigma), aprotinin (2 μg/ml) (Sigma), PMSF (0.1 mM) (Nacalai Tesque), and pepstatin (0.1 mM) (Peptide Institute). Cystatins were also used at appropriate concentrations as proteinaceous CP inhibitors, and included oryzacystatin-I (5×10^{-11}-1×10^{-7} M) encoded by pOCE26-5′-1 (*28*), soyacystatin (5×10^{-11}-1×10^{-7} M) encoded by pSCΔNΔC (*29*), and egg white cystatin (5×10^{-11}-1×10^{-7} M) purchased from Takara Shuzo. The inhibitory activities of the cystatins were assayed in advance using papain as the target CP (*28, 29*). The effects of inhibitors are presented as residual activities compared to the activity in the absence of an inhibitor.

RESULTS

Establishment of a Functional Expression System for the Production of Recombinant SCPs in E. coli—In order to characterize *S. zeamais* cysteine proteinases (SCPs) by means of biochemical methods, we had to obtain sufficient amounts of active SCPs in an *E. coli* expression system, because the amount of biological materials (live insects) are limited, and because any SCPs obtained from biological sources might contain plural species of related CPs originating from multiple SCP genes (*7*) or other genes encoding similar proteinases. In this study, two SCP DNAs, SCPc1, and SCPg3 DNA (Fig. 1B), were selected for expression. When SCP DNA fragments encoding putative proenzyme and mature enzyme forms were expressed as fusion and non-fusion proteins in *E. coli* by means of several methods, all the expressed SCP proteins were recovered in the insoluble materials, probably as inclusion bodies (data not shown). Therefore, the following approach was adopted for



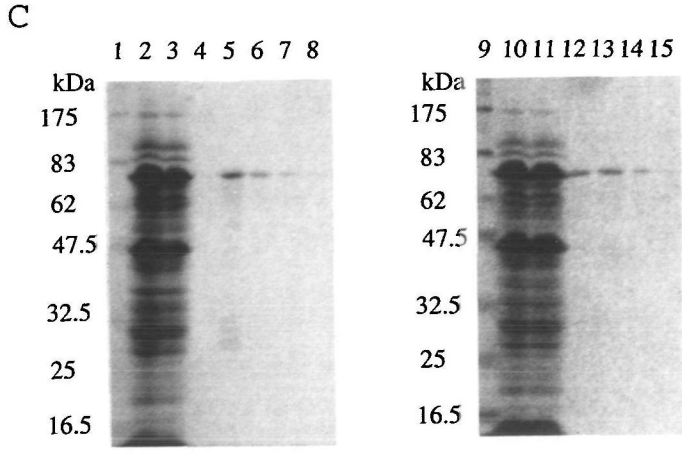


Fig. 1. **Expression and purification of recombinant *Sitophilus* cysteine proteinases fused with glutathione S-transferase.** A: Nucleotide sequence of the sense primer used to amplify the cDNA fragments of *Sitophilus zeamais* cysteine proteinases c1 and g3 (SCPc1 and SCPg3). The underlined *Bam*HI site was introduced into the primer to generate a cohesive end in-frame with the GST encoded by the vector. B: Schematic structures of the proenzyme forms of SCPc1 and SCPg3 fused with glutathione S-transferase (GST-proSCPc1 and GST-proSCPg3). C: SDS-PAGE of fractions obtained on glutathione-Sepharose column chromatography. The left panel shows the profile of GST-proSCPc1 and right panel that of GST-proSCPg3. Lanes 1 and 9, prestained standard molecular weight markers (BioRad); lanes 2 and 10, detergent-solubilized bacterial extract containing GST-proSCPc1 (lane 2) or GST-proSCPg3 (lane 10); lanes 3 and 11, flow-through fractions from the glutathione Sepharose 4B column; lanes 4-8, GST-proSCPc1 fractions eluted with 10 mM glutathione; lanes 12-15, GST-proSCPg3 fractions eluted with 10 mM glutathione.

purifying SCP proteins. First, the SCPs were expressed as fusion proteins with glutathione S-transferase (GST) as a purification tag. An expressed protein was then solubilized in an ionic detergent such as *N*-lauroylsarcosinate, and then the solubilized GST-fusion protein was bound to an affinity column of glutathione-Sepharose. The detergent was then washed off and the bound GST-fusion proteins were eluted with glutathione. By this method, proenzymes of SCPc1 and SCPg3 fused to the COOH terminus of GST are obtained in apparently purified forms in a solution free from a detergent (Fig. 1, B and C). Approximately 0.1 mg of purified SCP can be recovered from 100 ml culture in the final eluate. This yield is almost the same as those by reported for bacterial (*14, 23, 30*), yeast (*16*), and baculovirus (*31*) expression systems.

Next, we conducted preliminary examination of the activities of these fusion proteins at acidic pH (pH 3.5) using hemoglobin as a substrate, because we previously observed that the probable proenzyme form(s) of SCP(s) present in crude extracts were converted into the mature form(s) at acidic pH (*7*), and because BCP, an SCP-related protein in silkworms, has been shown to hydrolyze hemoglobin effectively (*5, 32*). As a result, proteinase activities of GST-proSCPc1 and GST-proSCPg3 were observed (details are given below). We then characterized the autolytic activation and biochemical natures of these enzymes.

Processing of Recombinant SCP Proenzymes Fused to Glutathione S-Transferase—Since the pro-forms of CPs are known to be converted into their mature forms at acidic pH (*11, 16, 17, 19, 33*) through proteolytic cleavage before or concomitantly with the substrate interaction, we examined whether or not such conversion also takes place with the recombinant SCP proenzymes obtained as GST-fusion proteins. As shown in Fig. 2A, an immunoblot profile of GST-proSCPc1 after incubation at various pHs showed that GST-proSCPc1 is converted into a 37 kDa form on 30 min incubation between pH 2.5 and 4.5. However, this processing gradually decreases as the pH is raised, and conversely the amount of the original 75 kDa form increases together with the appearance of several intermediate bands. The time course of this conversion at pH 3.5 showed that the 37 kDa form as well as probable intermediate bands are detected immediately, but most of the GST-proSCPc1 remains in the original 75 kDa form at 30 s (Fig. 2B). When the incubation time was prolonged to 4 min, GST-proSCPc1 was completely converted to the 37 kDa form, while no such form was detected in the presence of E-64 even after 20 min incubation. Inhibitors against serine, metallo- and aspartic proteinases do not inhibit this conversion (data not shown). These results indicate that recombinant proSCPc1 undergoes autolytic conversion, either intermolecularly or intramolecularly, into the 37 kDa form, which corresponds roughly to the apparent

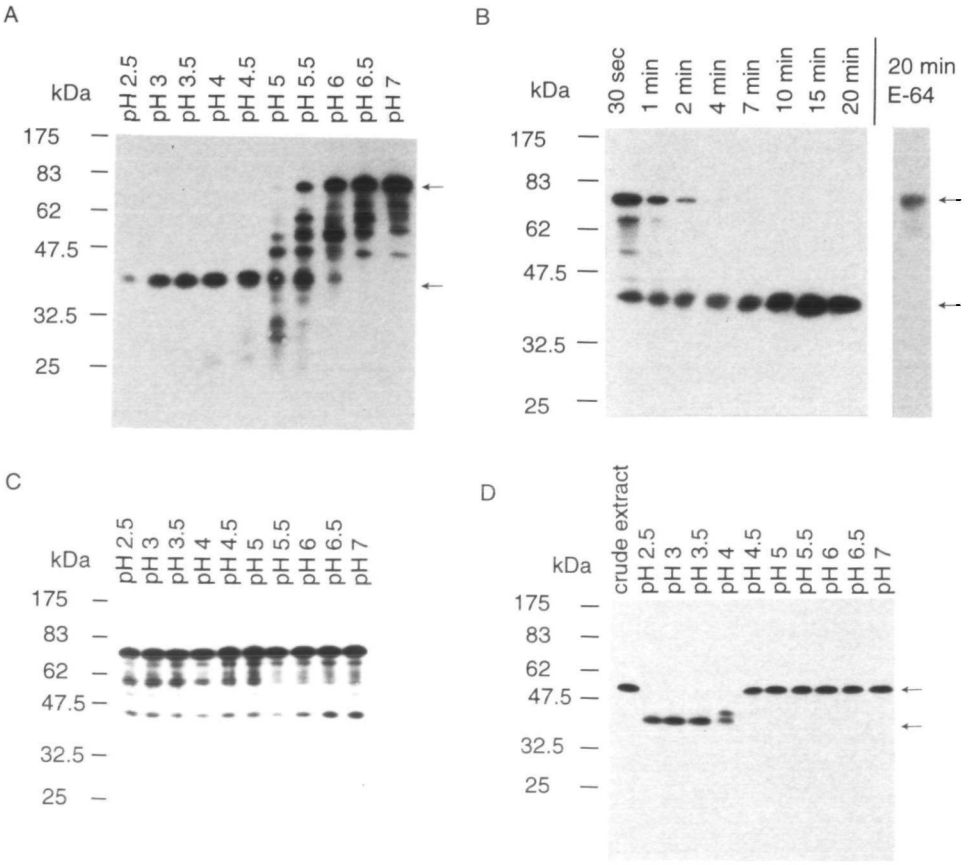


Fig. 2. ***In vitro* processing of recombinant and native SCP proenzymes.** The pH dependency (A, C, and D) and time course (B) of SCP processing were visualized by immunoblot analysis using an anti-SCP antibody raised against the mature protein region of SCP (*7*) and an ECL kit (Amersham). The positions of prestained standard molecular weight markers (BioRad) are indicated to the left of each panel. A: pH dependency of the processing of recombinant proSCPc1. GST-proSCPc1 was incubated in the presence of 0.2% hemoglobin at 37°C for 30 min in the following buffers; 0.1 M glycine/HCl (pH 2.5-3.0), 0.1 M sodium acetate (pH 3.5-5.5), and 0.1 M MES/NaOH (pH 6.0-7.0). The bands at 75 and 37 kDa indicated by the arrows are GST-proSCPc1 and the putative mature SCPc1, respectively. B: Time course of the processing at pH 3.5. GST-proSCPc1 was incubated with 0.2% hemoglobin in 0.1 M sodium acetate buffer, pH 3.5, at 37°C for the indicated periods, and then similarly incubated for 20 min in the presence of 0.1 mM E-64. The bands at 75 and 37 kDa, indicated by arrows, are GST-proSCPc1 and the putative mature SCPc1, respectively. C: pH dependence of the processing of recombinant proSCPg3. GST-proSCPg3 was incubated under the same conditions as in A. D: pH dependency of the processing of native proSCP(s) contained in a crude extract of adult *S. zeamais*. The crude extract was incubated with 0.2% hemoglobin at 37°C for 30 min in the same buffers as in A. The bands at 49 and 37 kDa, indicated by arrows, are proSCP(s) and mature SCP(s), respectively.

molecular size of the mature protein region (7).

On the contrary, recombinant proSCPg3 fused to GST does not seem to be converted into the mature form in the pH range of 2.5 to 7.0 (Fig. 2C), since most of the GST-proSCPg3 remains in its original 74 kDa form, although 55, 50, and 42 kDa bands, probably corresponding to partial degradation products that arise during storage, can be detected.

We also investigated the processing of endogenous SCP(s) mainly encoded by SCPc1 mRNA (7) in order to determine whether or not the above pH dependent reaction is intrinsic to naturally occurring proSCP(s). A crude extract of *S. zeamais* was prepared in the presence of several inhibitors against metallo- (EDTA), serine (PMSF), and aspartic (pepstatin) proteinases to inactivate non-CP proteinases that could degrade CPs in the extract. The extract was incubated at various pHs, and the molecular forms of SCP after the incubation were analyzed by immunoblotting. As shown in Fig. 2D, the 49 kDa form of SCP was completely converted into the 37 kDa form below pH 3.5, no 37 kDa form being detected above pH 4.5. This result is essentially the same as the profile for recombinant SCPc1 (Fig. 2B), although the pH range for the autolysis of SCP(s) in the crude extract is somewhat narrower than that for recombinant proSCPc1. The apparent molecular weight of the processed form of endogenous SCP(s) is the same as that of recombinant SCPc1 (Fig. 3A, lanes 1 and 2). In addition, the data suggest that this conversion occurs autolytically, because the reaction mixture contained various proteinase inhibitors other than those against CPs, although the possibility remains that CP(s) other than SCPs are involved in the reaction. This observation strongly suggests that the autolytic processing of recombinant SCPc1 that occurs in our functional expression system reflects the intrinsic biochemical characteristics applicable to endogenous SCP(s).

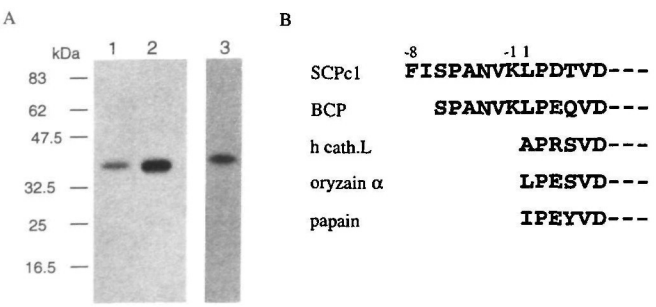


Fig. 3. **NH_2-terminal sequence of the processed form of GST-proSCPc1.** A: Comparison of the electrophoretic mobilities of differently processed SCPs. The positions of prestained standard molecular weight markers (BioRad) are indicated to the left of the panel. Lane 1, 37 kDa form of SCP from a crude *S. zeamais* extract after *in vitro* processing at pH 3.5; lane 2, processed form derived from GST-proSCPc1 after incubation at pH 3.5 in the presence of 0.2% hemoglobin; lane 3, processed form of GST-proSCPc1 after incubation at pH 3.0 in the absence of hemoglobin. Processed SCPs were detected by immunoblotting as described in the legend to Fig. 2. B: Alignment of the NH_2-terminal sequences of cysteine proteinases. The NH_2-terminal sequence of the *in vitro* autolytically processed product of GST-proSCPc1 (lane 3 in A) was compared with those of other cysteine proteinases [*Bombyx mori* cysteine proteinase (BCP), human cathepsin L (h cath.L), oryzain α, and papain]. Only the NH_2-terminal sequences are shown. The amino acid residues are numbered from the Leu of SCPc1 aligned with the other four CPs, and the extended amino acids are negatively numbered.

Autolytic Processing Site of proSCPc1—As described above, GST-proSCPc1 and the proenzyme form of SCP in insect extracts are each converted into a processed form with an apparent molecular weight of 37,000. In order to determine the structure of this form, we first tried to determine the NH_2-terminal sequence of the 37 kDa form of recombinant SCPc1 derived from GST-proSCPc1 obtained after the standard reaction, *i.e.*, incubation at pH 3.5 with 0.2% hemoglobin at 37°C for 30 min. However, we could not determine any distinct sequence(s) due to contamination by other protein sequences probably derived from the substrate. Therefore, GST-proSCPc1 autolysis was carried out in the absence of other proteinaceous materials. The processed protein, which showed slightly slower mobility than that in the presence of hemoglobin (Fig. 3A), was subjected to NH_2-terminal sequencing. The sequence obtained was Phe-Ile-Ser-Pro-Ala-Asn-Val-Lys-Leu-Pro-Asp-Thr-Val-Asp, in complete accordance with the sequence of Phe^{113} to Asp^{126} deduced from SCPc1 cDNA (7). As shown in Fig. 3B, this converted form has an extended octapeptide when compared with other CPs in the papain family (*5, 34-36*). The difference in mobility on SDS-PAGE between the forms of proSCPc1 processed with and without hemoglobin is equivalent to about 1 kDa in molecular mass, which is comparable to the calculated molecular mass of 875 for the extended octapeptide.

Substrate Specificity and pH Dependency of GST-proSCPc1—As a first step for elucidating the enzymatic characteristics of SCPs, we evaluated the substrate specificity with several proteins and peptide derivatives as substrates. As shown in Table I, SCPc1 shows a broad specificity toward protein substrates; it efficiently hydrolyzes bovine hemoglobin, casein and serum albumin, of which hemoglobin is the best substrate. This profile is essentially the same as that reported for BCP, which also exhibits a broad substrate specificity for protein substrates (*32*). On the other hand, SCPc1 shows strict selectivity for peptidyl-MCA substrates; Z-Phe-Arg-MCA was effectively hydrolyzed, while Z-Arg-Arg-MCA and Arg-MCA were hardly hydrolyzed. The specificity of SCPc1 for peptidyl-

TABLE I. **Substrate specificities of GST-proSCPs.** Standard enzyme activities were assayed using 0.2% of each protein substrate and 4 μM of each peptide substrate at 37°C in 0.1 M sodium acetate buffer, pH 3.5. The incubation periods for proteinaceous substrates and fluorescent substrates were 30 min and 10 min, respectively. The hydrolytic activities measured as described under "MATERIALS AND METHODS" are shown as percentages of the highest activity in each case.

Substrate	Relative activity (%)	
	GST-proSCPc1	GST-proSCPg3[a]
Hemoglobin	100	n.d.[b]
Casein	85.1	n.d.
Serum albumin	57.9	n.d.
Z-Phe-Arg-MCA	100[c]	100[d]
Z-Arg-Arg-MCA	<0.1	2.3
Arg-MCA	0.13	<0.1

[a]Assayed at pH 4.5. [b]Not determined. [c]Corresponding to 1.18 units, where 1 unit stands for the activity when 1 pmol 4-amino-7-methylcoumarin is released in 10 min when 1 ng enzyme is used. At pH 4.5, the optimum pH for the hydrolysis of Z-Phe-Arg-MCA, this enzyme exhibits 3.91 units. [d]Corresponding to 8.43×10^{-2} units.

TABLE II. **Effects of proteinase inhibitors on SCPc1 activity.** The effects of proteinase inhibitors on the hydrolytic activity of GST-proSCPc1 are shown as residual activities relative to that detected in the absence of inhibitors. The assay was performed as described in Table I using 0.2% hemoglobin and 20 μM Z-Phe-Arg-MCA as substrates.

Inhibitor	Concentration	Relative activity (%)	
		Hemoglobin	Z-Phe-Arg-MCA
None	—	100	100
E-64	10 μM	5.2	2.24
Leupeptin	10 μM	7.1	1.30
Antipain	0.1 mM (10 μM)[a]	13.1	1.01
Aprotinin	0.1 mM	85.1	n.d.[b]
PMSF	0.1 mM	97.3	90.2
Pepstatin	0.1 mM	102.9	125.3

[a]Inhibition concentration when Z-Phe-Arg-MCA was used as the substrate. [b]Not determined.

MCA substrates is similar to those observed for human and rat cathepsins L (*37*).

Next, the pH dependency of SCPc1 was investigated using hemoglobin and Z-Phe-Arg-MCA as substrates. The pH optima for hemoglobin and Z-Phe-Arg-MCA are 3.5 and 4.5, respectively (data not shown). When hemoglobin was used as a substrate, proteolytic activity of SCPc1 was distinctly detected in the pH range of 2.5-4.5, with no appreciable hydrolytic activity above pH 5.0. The pH dependency of SCPc1 is essentially the same as that of BCP purified from *Bombyx* eggs (*32*). For the hydrolysis of hemoglobin, the pH optimum of SCPc1 was significantly lower than that of human cathepsin L, while they show almost the same pH optima when Z-Phe-Arg-MCA is used (*14*).

Proteolytic Activity of GST-proSCPg3—In contrast to GST-proSCPc1, GST-proSCPg3, which is highly similar in sequence to SCPc1 but lacks the COOH-terminal 7 amino acid residues (*7*), is converted poorly into the mature form (Fig. 2C). Furthermore, only slight activity can be observed when the proteolytic activity of SCPg3 is measured in the pH range of 2.5-7.0 using protein and peptidyl-MCA substrates. However, the pH optimum for the hydrolysis of Z-Phe-Arg-MCA is 4.5, in agreement with the pH optimum of SCPc1. In addition, the substrate specificity of SCPg3 is also similar to that of SCPc1 as far as peptidyl-MCA substrates are concerned (Table I). These results suggest that SCPg3 has essentially the same enzymatic profile as SCPc1, despite exhibiting much lower activity. The lower activity may arise from the fact that GST-proSCPg3 undergoes almost no autolytic processing.

Inhibitors of SCPc1—The effects of various proteinase inhibitors on the hydrolytic activity of recombinant SCPc1 were measured. As shown in Table II, the hydrolytic activity of SCPc1 is inhibited by synthetic CP inhibitors such as E-64, leupeptin and antipain.

We finally investigated the inhibition of SCPc1 by cystatins including oryzacystatin-I of rice origin (*38*). This investigation was carried out from the biodefensive point of view that insect CPs in the digestive fluid would be the target of plant cystatins contained in cereals upon which insect pests feed (*7*). As shown in Fig. 4, all cystatins of both animal and plant origin are very effective inhibitors of SCPc1. The concentrations of cystatins required for 50% inhibition (IC_{50}), as estimated from Fig. 4, are 6.8×10^{-9} M for oryzacystatin-I, 1.1×10^{-8} M for soyacystatin, and

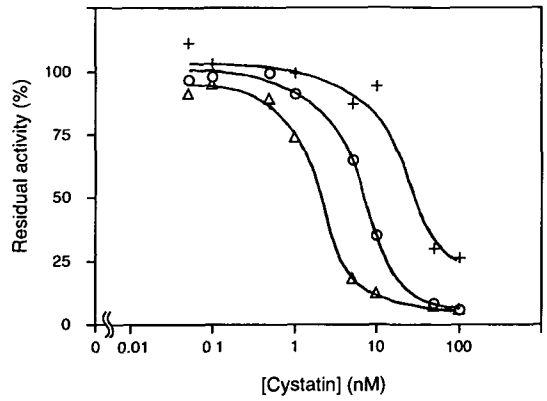


Fig. 4. **Inhibition profile of animal and plant cystatins against SCPc1.** The residual activities of GST-proSCPc1 were measured using Z-Phe-Arg-MCA as a substrate in the presence of cystatins at the indicated concentrations ranging from 50 pM to 0.1 μM. The hydrolytic activity of *ca.* 1 nM GST-proSCPc1 against 20 μM Z-Phe-Arg-MCA at pH 4.5 in the absence of cystatin was taken as 100%, and the residual activities are represented as relative values. △, egg white cystatin; ○, oryzacystatin-I; +, soyacystatin.

2.0×10^{-9} M for egg white cystatin. This strongly suggests that SCP(s) in the digestive fluid is a potential target of plant cystatins.

DISCUSSION

In this study, we established a functional bacterial expression and purification system for the production of SCPs, *i.e.* cathepsin L-like CPs in *Sitophilus zeamais*, and characterized their enzymatic features. The method for the preparation of functional recombinant CPs used in our experiments simplifies the laborous and time-consuming processes involved in the solubilization, renaturation and purification of proteins expressed in *E. coli*.

In many cases, recombinant proteins expressed in *E. coli* are recovered as insoluble aggregates. Such observations have been reported for papain (*20, 23*), cathepsins B (*13, 18*), L (*14*), and S (*19*), and cruzain (*30*); they are also insoluble when expressed in *E. coli*. Because of this insolubility, proper solubilization and renaturation, which are difficult and laborious, are needed to obtain functional recombinant CPs (*18, 19, 23*). In our experiments too, recombinant SCPs fused to GST are recovered in the insoluble fractions. However, the insoluble SCPs can be solubilized with an ionic detergent, *N*-lauroylsarcosinate, and subsequently renatured on an affinity column. This solubilization method and the purification tag, GST, make it easy to purify recombinant SCPs because renaturation and purification can be performed on a single column; all the procedures needed to obtain active SCPs can be completed in about two hours. Therefore, compared with procedures reported up to now, our method is simple and reproducible. Petanceska and Devi (*39*) reported a method similar to ours involving cathepsin S expressed as a GST-fusion protein in *E. coli*. Unfortunately, it is difficult to compare their expression and purification system with ours, because no details were given, and because their construct lacked a pro-region that might have a significant effect on the autolytic and catalytic activities. However, the methods for

cathepsin S and SCPc1 share the common feature that the presence of GST did not severely inhibit the enzyme activity. In addition, our study on SCPc1 indicates that GST does not interfere with the activation process in which the proenzyme form is autolyzed to the mature enzyme form. It will thus be very interesting to apply our method to other CPs, and we are now producing active CPs of plant origin successfully (Kiyosaki, T., Matsumoto, I., Arai, S., and Abe, K., manuscript in preparation). The functional expression and purification method we established in this study is a general method applicable to the preparation of active CPs by means of DNA technology.

Using recombinant GST-proSCPs, we investigated the *in vitro* activation process. In the profile of the autolytic cleavage of SCPc1, several intermediate bands were observed when autolysis occurred at pH 5-6 (Fig. 2A) or when short incubation periods were used (Fig. 2B). This suggests that the reaction occurs in multiple steps as observed for cathepsin L (*11, 40*). The occurrence of such a multistep reaction is also supported by the finding that under certain conditions, *e.g.*, in the absence of a substrate protein, a change occurs in the pattern of proteolytic cleavage (Fig. 3).

As for the enzymatic characteristics exhibited by the autolytically derived mature enzyme, SCPc1 is essentially the same as *Bombyx mori* CP (BCP) in terms of substrate specificity and pH dependency. Furthermore, these insect cathepsin L-like CPs are similar to mammalian cathepsin L in the same criteria. However, some differences are observed in terms of pH dependence between insect cathepsin L-like CPs and mammalian cathepsin L; the pH optima of the insect CPs are lower than those of the mammalian enzymes. This suggests that the physiological conditions are slightly different in the tissues in which these CPs function. Mammalian cathepsin L is known to be localized primarily in the lysosomes, where the pH is around 5.5, which is consistent with the pH optimum for this enzyme (*33*). On the other hand, insect CPs have been shown to exist in several distinct tissues (*7, 9*), but their microenvironments have not been definitely described. Thus, it is not clear at present whether the pH optima observed for SCPc1 and BCP are consistent with their environments. However, since other hydrolytic enzymes such as α-amylases and α-glucosidases that occur in the digestive fluid exhibit pH optima of around 5.0 (*41*), the pH of the interspaces of the digestive organs, where SCPc1 and *Drosophila* CP1 (DCP1) are secreted, may be as low as pH 5.0, the marginal value for SCPc1 activity. Taken together, it is postulated that these insect cathepsin L-like CPs are active in target tissues after the proenzymes have been activated, that might result from a temporal pH change or from proteolytic cleavage by some other proteinases.

In this study we examined the activities of two SCPs selected from among the four to six CPs in *S. zeamais* (*7*), and describe an unexpected difference between the two related enzymes, SCPc1 and SCPg3; even though their amino acid sequences are highly homologous (96.1%), their susceptibility to autoproteolytic conversion differs significantly (Fig. 2). In terms of structural features, the most noticeable difference is in the length of the COOH termini (Fig. 1B); the COOH terminus of SCPg3 is shorter by 7 amino acids than that of SCPc1, whose sequence is similar to those of other CPs (*7*). Also, mammalian CPs are known to undergo autolysis and thereby become activated (*11, 16, 17, 19, 33*). Thus, it is highly probable that the lack of 7 amino acids is the reason why SCPg3 fails to undergo autoproteolytic activation. Consistent with the above hypothesis, the deletion of 7 amino acid residues from the COOH terminus of SCPc1 significantly impaired its autolytic processing (Matsumoto, I., Abe, K., Arai, S., and Emori, Y., manuscript in preparation). However, the simple addition of the 7 amino acid sequence to the COOH terminus of SCPg3 failed to increase its activity, indicating that other structural elements are needed to elicit the full activity of SCPc1. In other words, SCPg3 lacks other unknown element(s) required for autolytic processing. The crucial residues responsible for the autolytic activity of SCPs can probably be elucidated by mutagenesis analysis of 13 different residues (*7*) one by one.

In view of the physiological regulation and function of SCPs, the biochemical differences between SCPc1 and SCPg3 may be very important, because the former can be activated at acidic pH, while the latter may need post-translational modification for activation, for example, *N*-glycosylation or proteolytic cleavage(s) by other proteinase(s). Thus, it is speculated that these two similar enzymes function under different physiological conditions and/or play different physiological roles due to the difference in their activation processes, not to other enzymatic characteristics such as substrate specificity. This possibility awaits further analysis using molecular probes such as anti-peptide antibodies and oligonucleotides that distinguish between these two SCPs.

REFERENCES

1. Berti, P.J. and Storer, A.C. (1995) Alignment/phylogeny of the papain superfamily of cysteine proteases. *J. Mol. Biol.* **246**, 273-283
2. Kominami, E., Ueno, T., Muno, D., and Katunuma, N. (1991) The selective role of cathepsins B and D in the lysosomal degradation of endogenous and exogenous proteins. *FEBS Lett.* **287**, 189-192
3. Kakegawa, H., Nikawa, T., Tagami, K., Kamioka, H., Sumitani, K., Kawata, T., Drobnic-Kosorok, M., Lenarcic, B., Turk, V., and Katunuma, N. (1993) Participation of cathepsin L on bone resorption. *FEBS Lett.* **321**, 247-250
4. Homma, K., Kurata, S., and Natori, S. (1994) Purification, characterization, and cDNA cloning of procathepsin L from the culture medium of NIH-Sape-4, an embryonic cell line of *Sarcophaga peregrina* (flesh fly), and its involvement in the differentiation of imaginal discs. *J. Biol. Chem.* **269**, 15258-15264
5. Yamamoto, Y., Takimoto, K., Izumi, S., Toriyama-Sakurai, M., Kageyama, T., and Takahashi, S.Y. (1994) Molecular cloning and sequencing of cDNA that encodes cysteine proteinase in the eggs of the silkmoth, *Bombyx mori*. *J. Biochem.* **116**, 1330-1335
6. Matsumoto, I., Watanabe, H., Abe, K., Arai, S., and Emori, Y. (1995) A putative digestive cysteine proteinase from *Drosophila melanogaster* is predominantly expressed in the embryonic and larval midgut. *Eur. J. Biochem.* **227**, 582-587
7. Matsumoto, I., Emori, Y., Abe, K., and Arai, S. (1997) Characterization of a gene family encoding cysteine proteinases of *Sitophilus zeamais* (maize weevil), and analysis of the protein distribution in various tissues including alimentary tracts and germ cells. *J. Biochem.* **121**, 464-476
8. Takahashi, S.Y., Zhao, X., Kageyama, T., and Yamamoto, Y. (1992) Acid cysteine proteinase from the eggs of silkmoth, *Bombyx mori*: tissue distribution, developmental changes and the site of synthesis for the enzyme. *Insect Biochem. Mol. Biol.* **22**, 369-377

9. Yamamoto, Y. and Takahashi, S.Y. (1993) Cysteine proteinase from *Bombyx* eggs: role in programmed degradation of yolk proteins during embryogenesis. *Comp. Biochem. Physiol.* **106B**, 35-45
10. Takahashi, S.Y., Yamamoto, Y., Shionoya, Y., and Kageyama, T. (1993) Cysteine proteinase from the eggs of the silkmoth, *Bombyx mori*: identification of a latent enzyme and characterization of activation proteolytic processing *in vivo* and *in vitro*. *J. Biochem.* **114**, 267-272
11. Gal, S. and Gottesman, M.M. (1986) The major excreted protein of transformed fibroblasts is an activable acid-protease. *J. Biol. Chem.* **261**, 1760-1765
12. Mason, R.W., Gal, S., and Gottesman, M.M. (1987) The identification of the major excreted protein (MEP) from a transformed mouse fibroblast cell line as a catalytically active precursor form of cathepsin L. *Biochem. J.* **248**, 449-454
13. Mort, J.S., Tam, A., Steiner, D.F., and Chan, S.J. (1988) Expression of rat and mouse cathepsin B precursors in *Escherichia coli*. *Biol. Chem. Hoppe-Seyler* **369**, 163-167
14. Smith, S.M. and Gottesman, M.M. (1989) Activity and deletion analysis of recombinant human cathepsin L expressed in *Escherichia coli*. *J. Biol. Chem.* **264**, 20487-20495
15. Rowan, A.D., Mason, P., Mach, L., and Mort, J.S. (1992) Rat procathepsin B: proteolytic processing to the mature form *in vitro*. *J. Biol. Chem.* **267**, 15993-15999
16. Brömme, D., Bonneau, P.R., Lachance, P., Wiederanders, B., Kirschke, H., Peters, C., Thomas, D.Y., Storer, A.C., and Vernet, T. (1993) Functional expression of human cathepsin S in *Saccharomyces cerevisiae*: purification and characterization of the recombinant enzyme. *J. Biol. Chem.* **268**, 4832-4838
17. Mach, L., Mort, J.S., and Glössl, J. (1994) Maturation of human procathepsin B: proenzyme activation and proteolytic processing of the precursor to the mature proteinase, *in vitro*, are primarily unimolecular processes. *J. Biol. Chem.* **269**, 13030-13035
18. Kuhelj, R., Dolinar, M., Pungercar, J., and Turk, V. (1995) The preparation of catalytically active human cathepsin B from its precursor expressed in *Escherichia coli* in the form of inclusion bodies. *Eur. J. Biochem.* **229**, 533-539
19. Kopitar, G., Dolinar, M., Strukelj, B., Pungercat, J., and Turk, V. (1996) Folding and activation of human procathepsin S from inclusion bodies produced in *Escherichia coli*. *Eur. J. Biochem.* **236**, 558-562
20. Vernet, T., Tessier, D.C., Laliberté, F., Dignard, D., and Thomas, D.Y. (1989) The expression in *Escherichia coli* of a synthetic gene coding for the precursor of papain is prevented by its own putative signal sequence. *Gene* **77**, 229-236
21. Cohen, L.W., Fluharty, C., and Dihel, L.C. (1990) Synthesis of papain in *Escherichia coli*. *Gene* **88**, 263-267
22. Vernet, T., Tessier, D.C., Richardson, C., Laliberté, F., Khouri, H.E., Bell, A.W., Storer, A.C., and Thomas, D.Y. (1990) Secretion of functional papain precursor from insect cells. *J. Biol. Chem.* **265**, 16661-16666
23. Taylor, M.A.J., Pratt, K.A., Revell, D.F., Baker, K.C., Sumnar, I.G., and Goodenough, P.W. (1992) Active papain renatured and processed from insoluble recombinant propapain expressed in *Escherichia coli*. *Protein Eng.* **5**, 455-459
24. Vernet, T., Chatellier, J., Tessier, D.C., and Thomas, D.Y. (1993) Expression of functional papain precursor in *Saccharomyces cerevisiae*: rapid screening of mutants. *Protein Eng.* **6**, 213-219
25. Nakano, H., Yamazaki, T., Ikeda, M., Masai, H., Miyatake, S., and Saito, T. (1994) Purification of glutathione S-transferase fusion proteins as a non-degraded form by using a protease-negative *E. coli* strain, AD202. *Nucleic Acids Res.* **22**, 543-544
26. Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685
27. Harlow, E. and Lane, D. (1988) *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, New York
28. Abe, K., Emori, Y., Kondo, H., Arai, S., and Suzuki, K. (1988) The NH_2-terminal 21 amino acid residues are not essential for the papain-inhibitory activity of oryzacystatin, a member of the cystatin superfamily. *J. Biol. Chem.* **263**, 7655-7659
29. Misaka, T., Kuroda, M., Iwabuchi, K., Abe, K., and Arai, S. (1996) Soyacystatin, a novel cysteine proteinase inhibitor in soybean, is distinct in protein structure and gene organization from other cystatins of animal and plant origins. *Eur. J. Biochem.* **240**, 609-614
30. Eakin, A.E., Mills, A.A., Harth, G., McKerrow, J.H., and Craik, C.S. (1992) The sequence, organization, and expression of the major cysteine protease (cruzain) from *Trypanosoma cruzi*. *J. Biol. Chem.* **267**, 7411-7420
31. Vernet, T., Khouri, H.E., Laflamme, P., Tessier, D.C., Musil, R., Gour-Salin, B.J., Storer, A.C., and Thomas, D.Y. (1991) Processing of the papain precursor: purification of the zymogen and characterization of its mechanism of processing. *J. Biol. Chem.* **266**, 21451-21457
32. Kageyama, T. and Takahashi, S.Y. (1990) Purification and characterization of a cysteine proteinase from silkworm eggs. *Eur. J. Biochem.* **193**, 203-210
33. Mason, R.W. and Massey, S.D. (1992) Surface activation of pro-cathepsin L. *Biochem. Biophys. Res. Commun.* **189**, 1659-1666
34. Mason, R.W., Walker, J.E., and Northrop, F.D. (1986) The *N*-terminal amino acid sequences of the heavy and light chains of human cathepsin L. *Biochem. J.* **240**, 373-377
35. Watanabe, H., Abe, K., Emori, Y., Hosoyama, H., and Arai, S. (1991) Molecular cloning and gibberellin-induced expression of multiple cysteine proteinases of rice seeds (Oryzains). *J. Biol. Chem.* **266**, 16897-16902
36. Cohen, L.W., Coghlan, V.M., and Dihel, L.C. (1986) Cloning and sequencing of papain-encoding cDNA. *Gene* **48**, 219-227
37. Barrett, A.J. and Kirschke, H. (1981) *Methods Enzymol.* **80**, 535-561
38. Abe, K., Emori, Y., Kondo, H., Suzuki, K., and Arai, S. (1987) Molecular cloning of a cysteine proteinase inhibitor of rice (oryzacystatin). *J. Biol. Chem.* **262**, 16793-16797
39. Petanceska, S. and Devi, L. (1992) Sequence analysis, tissue distribution, and expression of rat cathepsin S. *J. Biol. Chem.* **267**, 26038-26043
40. Ishidoh, K. and Kominami, E. (1994) Multi-step processing of procathepsin L *in vitro*. *FEBS Lett.* **352**, 281-284
41. Terra, W.R. and Ferreira, C. (1994) Insect digestive enzymes: properties, compartmentalization and function. *Comp. Biochem. Physiol.* **109B**, 1-62