Available online at www.sciencedirect.com
ScienceDirect
Insect Biochemistry and Molecular Biology 37 (2007) 960–968
Insect Biochemistry and Molecular Biology
www.elsevier.com/locate/ibmb

Purification of a cysteine protease inhibitor from larval hemolymph of the tobacco hornworm (*Manduca sexta*) and functional expression of the recombinant protein

Takayuki Miyaji[a], Yoshiaki Kouzuma[a], Jun Yaguchi[a], Rika Matsumoto[a], Michael R. Kanost[b], Karl J. Kramer[c], Masami Yonekura[a,*]

[a]*Laboratory of Food Molecular Functionality, College of Agriculture, Ibaraki University, 3-21-1, Chuo, Ami-machi, Inashiki-gun, Ibaraki 300-0393, Japan*
[b]*Department of Biochemistry, Kansas State University, Manhattan, KS 66506, USA*
[c]*Grain Marketing and Production Research Center, ARS-USDA, 1515 College Avenue, Manhattan, KS 66502, USA*

Received 2 March 2007; received in revised form 1 May 2007; accepted 1 May 2007

Abstract

A cysteine protease inhibitor (CPI) with an apparent molecular mass of 11.5 kDa was purified from larval hemolymph of the tobacco hornworm (*Manduca sexta*) by gel filtration on Sephadex G-50 followed by hydrophobic and ion-exchange column chromatographies. The purified cysteine proteinase inhibitor, denoted as MsCPI, strongly inhibited the plant cysteine protease, papain, with a K_i value of 5.5×10^{-9} M. Nucleotide sequence analysis of a partial cDNA encoding MsCPI indicated that MsCPI consists of 105 amino acid residues in a sequence that is similar to sarcocystatin A from *Sarcophaga peregrina*. However, northern blotting and PCR analyses using the specific primers of MsCPI suggested that the mRNA encoding MsCPI had a size of more than 12 kilobases, which included at least six tandemly repeated MsCPI segments. MsCPI was expressed in *Escherichia coli* and the recombinant protein effectively inhibited cysteine proteases from plants as well as from animals such as cathepsins B (K_i, 6.8 nM), H (3.0 nM), and L (0.87 nM). There was no inhibition exhibited toward trypsin, chymotrypsin, subtilisin, pepsin or themolysin.
© 2007 Elsevier Ltd. All rights reserved.

Keywords: Cysteine protease; Cysteine protease inhibitor; Cystatin; Hemolymph; *Manduca sexta*; Overexpression

1. Introduction

Proteinaceous cysteine protease inhibitors (CPIs) have been found in various animal and plant tissues, and many of them have been isolated and characterized in terms of protein structure and inhibitory activity. Cystatins are the best-characterized CPI group of mammalian and plant origins, and are known to form a superfamily of structurally homologous proteins (Barrett et al., 1986). In mammals, cystatins are believed to be involved in the regulation of endogenous cysteine proteases and/or in the protection against invading organisms such as bacteria, viruses and parasites. They share three conserved sequence motifs including a Gly in the vicinity of the N-terminal region, Q-X-V-X-G in the first hairpin loop, and P-W in the second hairpin loop. Structural analysis revealed that the amino-terminal segment and two loops of cystatins form a tripartite wedge-shaped edge that interacts with the active site cleft of cognate cysteine proteases (Bode et al., 1988; Stubbs et al., 1990). Besides cystatins, several other families of CPIs have been also recognized in the last decade. According to the MEROPS list (http://merops.sanger.ac.uk/) (Rawlings et al., 2004), more than 10 families of CPIs have been classified based on the amino acid sequence of inhibitory domain. The structures, inhibition mechanisms, and physiological roles of several of these families have been described in recent reviews (Dubin, 2005; Rzychon et al., 2004).

Although relatively little is known about CPIs from insects compared with mammals, information is increasing because of the determination of genomic sequences and

*Corresponding author. Tel./fax: +81 29 888 8683.
E-mail address: yonekura@mx.ibaraki.ac.jp (M. Yonekura).

0965-1748/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ibmb.2007.05.003

expressed sequence tag (EST) analysis from several insect species such as *Bombyx mori, Drosophila melanogaster* and *Anopheles gambiae*. In the MEROPS database, only four families of CPIs, that is families I25 (cystatin family), I29 (cytotoxic T-lymphcyte antigen (CTLA) family (Denizot et al., 1989)), I31 (thyropin family (Lenarcic and Bevec, 1998)), and I32 (inhibitor of apoptosis protein (IAP) family (Deveraux and Reed, 1999)), are listed as present in insects. Also, there are only a few reports describing the purification of CPIs from insect tissues, including two structurally different proteins, Sarcocystatin (Suzuki and Natori, 1985) from *Sarcophaga peregrina* (family I25) and a cysteine protease inhibitor, BCPI (Yamamoto et al., 1999), from *B. mori* (family I29).

To obtain a better understanding of structures and functions of other CPIs from insects, we investigated a cystatin from the tobacco hornworm (*Manduca sexta*). Here, we describe the purification, cDNA cloning, functional expression of the recombinant protein in *Escherichia coli* cells, and inhibitory properties of a novel CPI that is found in larval hemolymph.

2. Materials and methods

2.1. Insects

A *M. sexta* colony was reared from eggs originally obtained from Carolina Biological Supply Co. (Burlington, NC). Larvae were reared as described by Dunn and Drake (1983). Hemolymph was collected as previously described (Jiang et al., 2003), freeze dried, and stored at −20 °C until used.

2.2. Enzymes and chemicals

Papain, ficin, bovine cathepsin B, and human cathepsins L were obtained from Sigma Chemical Co. (St. Louis, MO). Human cathepsin H was obtained from Calbiochem (San Diego, USA). L-Pyroglutamyl-L-phenylalanyl-L-leucine *p*-nitroanilide (Pyr-Leu-Phe-pNA), carbobenzoxy-L-phenylalanyl-L-arginine 4-methyl-7-coumarylamide (Z-Phe-Arg-MCA), and L-arginine 4-methyl-7-coumarylamide (Arg-MCA) were obtained from the Peptide Institute (Minoh, Japan). Sephadex G-50 and DEAE-cellulofine (500 m) were from Amersham Biosciences (Piscataway, New Jersey) and Seikagaku Industry (Tokyo, Japan), respectively. Restriction endonucleases and DNA-modifying enzymes were purchased from MBI Fermentas (Burlington, Ontario) or Takara-Bio Co (Shiga, Japan). Plasmid pUC19 and the expression vector, pET-22b, were purchased from Takara-Bio Co. and Novagen (San Diego, CA), respectively. Oligonucleotides used in this study were purchased from Exigen (Tokyo). All other chemicals were of analytical grade for biochemical use.

2.3. Purification of M. sexta cysteine protease inhibitor

A cysteine protease inhibitor from *M. sexta* larval hemolymph (MsCPI) was purified as follows. Sixteen grams of hemolymph powder from *M. sexta* was dissolved in 800 ml of 20 mM Tris–HCl buffer (pH 7.5) containing 1 mM phenylthiourea solution and the solution was heated at 80 °C for 10 min. After centrifugation, the supernatant was concentrated by ultrafiltration. The crude inhibitor solution thus obtained was subjected to chromatography on Sephadex G-50 (2.5 × 65 cm) with a mobile phase of 20 mM Tris–HCl buffer (pH 7.5). The fractions exhibiting inhibitory activity toward papain were combined and applied to a butyl-cellulofine column (1.2 × 19 cm) equilibrated with 20 mM Tris–HCl buffer (pH 7.5) containing 1 M ammonium sulfate, and the protein was eluted with a linear gradient of decreasing $(NH_4)_2SO_4$ concentration from 1 to 0 M. The protein that did not adsorb to the butyl-cellulofine was applied to a DEAE-cellulofine column (1.6 × 19 cm) equilibrated with 20 mM Tris–HCl buffer (pH 7.5) and the protein was eluted with a linear gradient of increasing NaCl concentration from 0 to 0.5 M. Fractions exhibiting the inhibitory activity were pooled and rechromatographed using the same column and conditions.

Protein concentration was determined using BCA protein assay reagent (Pierce Chemical, Rockford, IL) with bovine serum albumin as a standard protein. The purity of a protein was judged by SDS-PAGE, which was carried out on a 15% polyacrylamide gel as described by Laemmli (1970). The N-terminal amino acid sequence was determined using a gas-phase protein sequencer ABI494.

2.4. Mass spectrometry

The molecular mass of the protein was determined using a Voyager-RP (Applied Biosystems) matrix-assisted laser desorption and ionization time-of-flight mass spectrometer (MALDI-TOF MS). In this analysis, α-cyano-4-hydroxycinnamic acid was used as matrix.

2.5. Assay of cysteine protease inhibitory activity

Inhibitory activity toward papain was measured by the method of Filippova et al. (1984). Briefly, papain and ficin were dissolved at 0.05 and 0.2 mg/ml, respectively, in 100 mM phosphate buffer (pH 6.5) containing 0.3 M KCl, 0.1 mM EDTA and 1 mM DTT. One hundred microliters of enzyme solution was preincubated at 37 °C for 10 min with 100 μl of various concentrations of the inhibitor solution. One milliliter of substrate solution (Pyr-Phe-Leu-pNA dissolved in the same buffer plus 10% dimethysulfoxide) was added to give a final concentration of 357 mM and incubated at 37 °C for 10 min. After incubation, 200 μl of 30% acetic acid was added and the absorbance was measured at 410 nm. Inhibitory activity was defined as the percentage inhibition of Pyr-Phe-Leu-pNA hydrolyzing

activity of papain in the assay mixture. One unit of papain inhibitory activity was defined as the amount of inhibitor needed to inhibit one microgram of papain.

Cathepsins B, H, and L in 100 mM sodium phosphate buffer (pH 6.0 for cathepsin B, pH 6.8 for cathepsin H, and pH 5.5 for cathepsin L), containing 1 mM DTT and 1 mM EDTA (0.25 ml) were preincubated at 37 °C for 10 min with or without the protease inhibitor (0.25 ml), before addition of 1.0 ml of the respective substrate solutions (Z-Phe-Arg-MCA for cathepsins B and L and Arg-MCA for cathepsin H) to initiate the reaction. The reaction was carried out at 37 °C for 10 min and stopped by addition of 1.0 ml of 50 mM sodium acetate buffer, pH 5.0, containing 10 mM iodoacetic acid. The liberated product, 7-amino-4-methylcoumarin, was estimated by measurement of the fluorescence intensity at 440 nm using a Shimadzu spectrofluorophotometer, model RF-1500. The excitation wavelength was 380 nm. The inhibition constants (K_i) toward all cysteine proteases used in this study were determined by the method of Henderson (1972).

2.6. cDNA cloning and determination of nucleotide sequence

To obtain a partial cDNA fragment of MsCPI for cDNA screening, two oligonucleotides were synthesized for use as forward primers corresponding to the amino acid sequence of MsCPI: F1, 5′-GARACNCCNGTNGGNGGNATHCARGC-3′ (ETPVGGIQA) and F2, 5′-CARGCNCARGAYCCNAAYGAYCC-3′ (QAQDPNDP). The first PCR was performed with the primer set of F1 and an M13 reverse sequence primer using the *M. sexta* fat body cDNA library (Jiang et al., 2003) as the template. After the second PCR with a primer set of F2 and the M13 reverse sequence primer, the PCR product (500 bp) was purified by agarose gel electrophoresis and then subcloned into the pGEM-T-EASY vector (Promega, Tokyo). The nucleotide sequence was determined by the dideoxy chain termination method using a BigDye Terminator v3.1 Cycle Sequencing Ready Reaction Kit and an automated DNA sequencer ABI prism 3100.

2.7. Isolation of total RNA from fat body

For isolating RNA from fat body, day 2 fifth-instar larvae were injected with 50 μl of filter-sterilized saline or *Micrococcus lysodeikticus* suspended in saline (2 μg/μl). Fat body was dissected from larvae 24 h after injection, rinsed briefly in phosphate buffered saline (PBS, 4.3 mM Na_2HPO_4, 1.4 mM KH_2PO_4, 137 mM NaCl, 2.7 mM KCl, pH 7.4). Fat body was homogenized, and isolation of total RNA was performed using an RNeasy Maxi Kit (Qiagen) following the manufacturer's instructions.

2.8. PCR analysis of MsCPI gene expression

Total RNA from fat body of *M. sexta* was reverse-transcribed using a First-Stranded cDNA Synthesis Kit ReverTraAce -α- (Toyobo, Tokyo) following the manufacturer's instructions. MsCPI gene specific primers, F3, 5′-CATATGGAGACACCAGTTGGTGGC-3′ and B1, 5′-GGATCCTCAGGCTCTGTATTTTTGGTTG-3′, were constructed and PCR was performed using the cDNA as a template. Since a ladder of seven bands was observed after amplification, the smallest (324 bp) and larger PCR products (1899 bp) were recovered and the nucleotide sequences determined.

2.9. Northern hybridization

Twenty micrograms of total RNA per lane were loaded on a 1% agarose gel containing 2.2 M formaldehyde and electrophoresis was performed. It was confirmed that equal RNA was loaded in all lanes by ethidium bromide staining of ribosomal RNA. RNA was transferred to nylon membrane Hybond-N+ (Amersham Biosciences) and the membrane was heated at 80 °C for 2 h. A 324-bp MsCPI gene fragment was labeled with alkaline phosphatase (AlkPhos Direct Labeling and Detection System with CDP-Star, Amersham Bioscience) for use as a probe. Hybridization was performed overnight at 55 °C. After washing for 10 min twice at 55 °C with 50 mM Na-phosphate buffer (pH 7.0) containing 0.2% blocking reagent, 0.1% SDS, 1 mM $MgCl_2$, 150 mM NaCl, and 2 M urea followed by 5 min twice at room temperature with 50 mM Tris–HCl buffer (pH 10) containing 2 mM $MgCl_2$, and 100 mM NaCl, the membrane was incubated with CDP-Star detection reagent for 5 min and visualized using a LAS3000mini (Fuji Film, Tokyo).

2.10. Expression of MsCPI protein in E. coli

The 324 bp MsCPI cDNA fragment was ligated into plasmid T-EASY vector (Promega). After confirmation of the nucleotide sequence, the DNA fragment encoding the mature MsCPI protein was excised by digestion with *Nde*I and *Bam*HI, and ligated into the expression vector pET-22b previously digested with the same enzymes. The resulting construct pET-MsCPI was introduced into *E. coli* BL21(DE3) codonplus RIL strain and the recombinant protein was induced with 1 mM IPTG according to the supplier's instructions.

2.11. Purification of recombinant MsCPI protein

After induction with 1 mM IPTG, the culture was incubated for an additional 4 h at 37 °C, then the cells were harvested and lysed by sonication in 30 mM Tris–HCl buffer (pH 7.5) containing 30 mM NaCl. The lysate was centrifuged at 10,000*g* for 10 min and the pellet was collected and washed with 1 M sucrose, followed by 1% Triton X-100/ 10 mM EDTA. The inclusion bodies thus obtained were dissolved in 50 mM Tris–HCl buffer (pH 8.0) containing 150 mM NaCl, 1 mM DTT, 1 mM EDTA and 8 M urea. The protein solution was dialyzed against

50 mM Tris–HCl buffer (pH 8.0) containing 150 mM NaCl, 1 mM DTT, 1 mM EDTA and 4 M urea for 12 h followed by 20 mM Tris–HCl buffer (pH 7.5) twice. The refolded MsCPI protein was purified by gel filtration on a Sephadex G-50 column (2.5 × 65 cm) equilibrated with 20 mM Tris–HCl buffer (pH 7.5).

3. Results

3.1. Purification of MsCPI protein

MsCPI protein was purified from *M. sexta* hemolymph by gel-filtration column chromatography on Sephadex G-50, hydrophobic interaction chromatography on butyl-cellulofine and ion-exchange chromatography on a DEAE-cellulofine column. Fig. 1 shows the elution profiles from Sephadex G-50 and the second DEAE-cellulofine column. The flow through fraction of the DEAE-cellulofine column exhibited strong papain inhibitory potency and a single band was detected by SDS-PAGE. The apparent molecular mass was calculated to be approximately 14 kDa (Fig. 1). On the other hand, this MsCPI fraction showed relatively broad peaks in MALDI TOF-MS analysis and had a smaller molecular mass of approximately 11,400–11,600 Da (data not shown). A typical purification run of MsCPI is summarized in Table 1.

MsCPI was subjected to N-terminal amino acid sequence analysis. The amino acid sequence was determined up to and including the 27th residue, ETPVGGIQAQ-(D/E)PNDPIFQSLAEESMQK. In addition to the presence of an aspartic acid at residue 11, a smaller amount of glutamic acid was also detected.

The inhibitory activity of MsCPI was investigated using papain as the proteolytic enzyme. Fig. 2 shows the titration

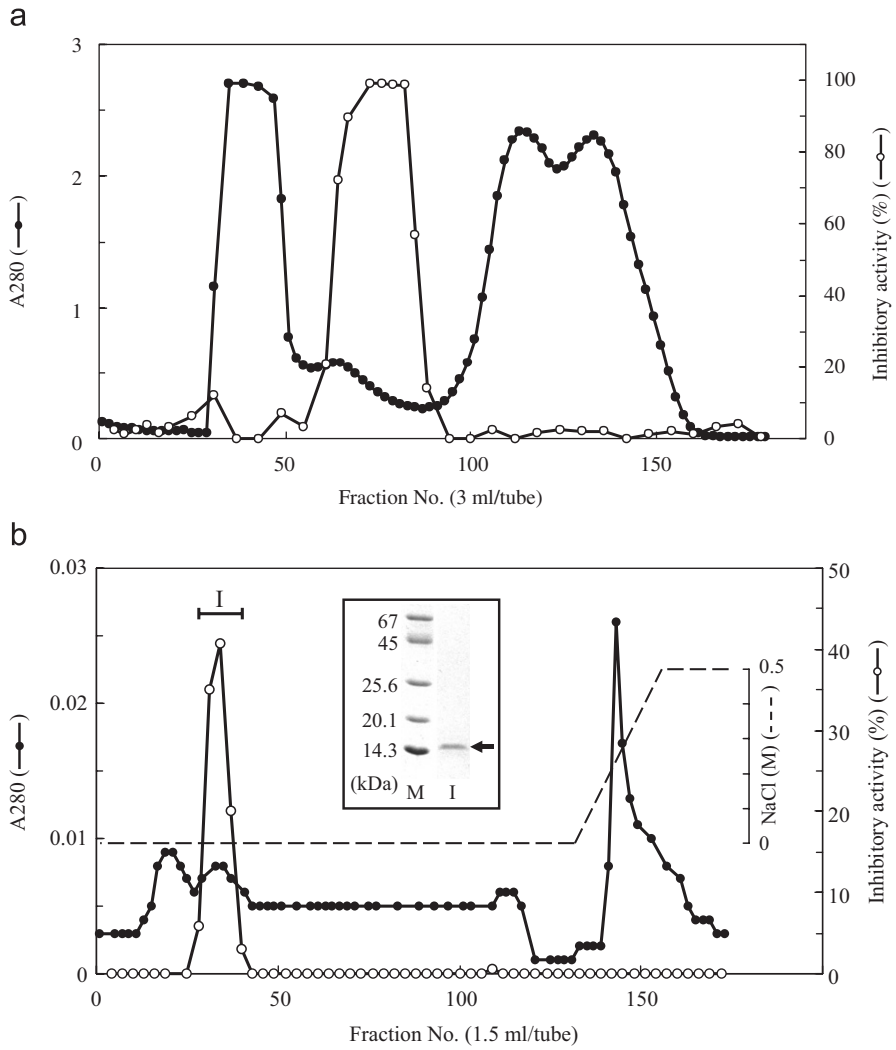

Fig. 1. Purification of cysteine protease inhbitor from the hemolymph of *M. sexta* larvae. After heat-treatment of hemolymph of *M. sexta*, MsCPI protein was purified by gel-filtration column chromatography on Sephadex G-50, hydrophobic interaction chromatography on butyl-cellulofine and ion-exchange chromatography on a DEAE-cellulofine column. (a) and (b) show the elution patterns of the Sephadex G-50 and the second DEAE-cellulofine purification steps, respectively. The flow through protein fraction I (black bar) in (b) was recovered and subjected to SDS-PAGE (in inset). The arrow indicates the position of the MsCPI protein. M, molecular weight markers; I, the flow through protein fraction I.

Table 1
Summary of purification of MsCPI protein from the hemolymph of *Manduca sexta* Larvae

	Total protein (mg)	Total inhibitory activity (IU)	Specific inhibitory activity (IU/mg)	Yield (%)	Purification (fold)
Hemolymph	6801	80,000	12	100	1
Heat treatment	3921	63,000	16	79	1.4
Sephadex G-50	132	48,000	370	60	31
Butyl-cellulofine	8.5	6200	730	7.8	62
DEAE-cellulofine (1st)	0.80	430	530	0.54	45
DEAE-cellulofine (2nd)	0.090	83	930	0.10	78

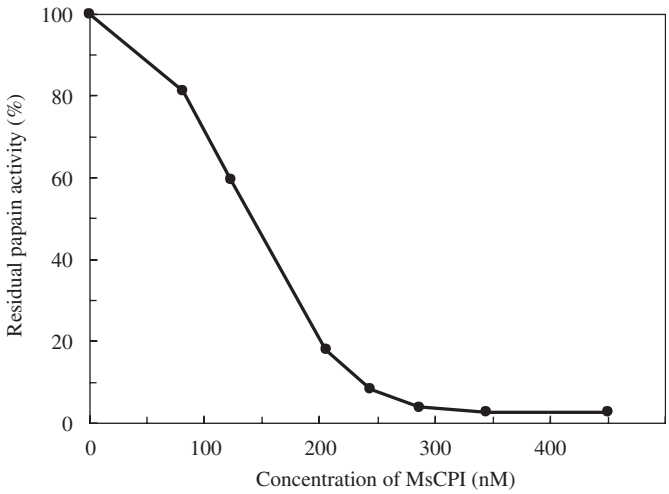

Fig. 2. Inhibition of papain by increasing amounts of MsCPI. A fixed amount of papain was mixed with increasing amounts of MsCPI and residual enzyme activity was assayed. Pyr-Phe-Leu-pNA was used as the substrate.

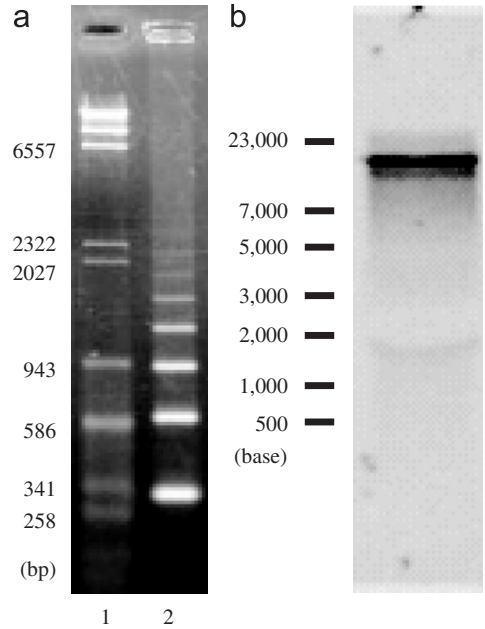

Fig. 3. Analysis of MsCPI gene expression. (a) RT-PCR analysis of the number of MsCP gene segments. PCR was performed using the MsCPI specific primers and the first-strand cDNA reverse transcribed using total RNA from fat body of *M. sexta* as the template. Lane 1, Marker, a mixture of λ/*Hin*d III and pUC 18/*Sau* 3AI; Lane 2, PCR products amplified by MsCPI-specific primers. (b) Northern blotting of MsCPI gene. Twenty micrograms of total RNA from *M. sexta* were used for northern blotting. A single segment of the MsCPI gene was used as probe.

curve for papain with MsCPI. From these data, the inhibitory constant (K_i) of MsCPI for papain was calculated to be 5.5 nM.

3.2. Cloning and nucleotide sequence of MsCPI cDNA

To obtain partial cDNA fragments encoding MsCPI, we performed PCR using two sets of degenerate primers based on the N-terminal amino acid sequence of MsCPI. One fragment (375 bp) was reproducibly amplified by PCR (data not shown). Nucleotide sequence analysis of the PCR product revealed that it was only a partial MsCPI cDNA, and the deduced amino acid sequence implied, judging from a BLAST search of the sequence, that MsCPI belonged to a member of the cystatin family of proteolytic enzyme inhibitors. However, the N-terminal sequence of MsCPI was directly followed by a possible C-terminal sequence of another MsCPI. This result indicated that the MsCPI gene might encode tandem-repeats of cystatin domains. When RT-PCR analysis was performed using the oligonucleotide primers corresponding to potential N- and C-terminal regions of MsCPI, a ladder of products with sizes in multiple of about 300 bp was amplified (Fig. 3a). This result indicated that the MsCPI mRNA included at least six cystatin repeats. The nucleotide sequence of the largest RT-PCR product (1899 bp) was determined (Genebank accession no. AB264728), and the deduced amino acid sequence is shown in Fig. 4. The product encoded six-repeated MsCPI segments, each composed of 105 amino acid residues, which were homologous with each other, having 95–100% identical residues. The N-terminal amino acid sequences of segments 1–6 were almost identical to that of the purified MsCPI protein. The 11th amino acid of these segments is either Asp (segments 1, 3 and 6) or Glu (segments 2, 4 and 5), a result consistent with the N-terminal amino acid sequence analysis of the MsCPI protein. The essential structural motifs of the cystatin domain are conserved in the all segments including a G in the N-terminal region, Q-X-V-X-G in the middle region, and P-W in the C-terminal region (Barrett et al., 1986), and two conserved cysteine residues indicating that these

```
                    10        20        30        40        50
Segment 1   ETPVGGIQAQDPNDPIFQSLAEESMQKYLQSIGSTKPHKVVRVVKATTQVVSGSM
Segment 2   ETPVGGIQAQEPNDPIFQSLAEESMQKYLQSIGSTKPHKVVRVVKASTQVVSGSM
Segment 3   ETPVGGIQAQDPNDPIFQSLAEESMQKYLQSIGSTKPHKVVRVVKATTQVVSGSM
Segment 4   ETPVGGIQAQEPNDPIFQSLAEESMQKYLQSIGSTKPHKVVRVVKASTQVVSGSM
Segment 5   ETPVGGIQAQEPNDPIFQSLAVESMQKYLQSIGSTKPHKIVRVVKASTQVVSGSM
Segment 6   ETPVGGIQAQDPNDPIFQSLADESMQKYLQSIGSTKPHKVVRVVKASTQVVSGSM

              60        70        80        90       100
Segment 1   TRIEFVISPSDGNSGDVISCYSEVWEQPWRHKKEITVDCKINNQKYRAKR    105
Segment 2   TRIEFVISPSDRNSGDVISCYSEVWEQPWRHKKEITVDCKINNQKYRAKR    210
Segment 3   TRIEFVISPSDGNSGDVISCYSEVWEQPWRHKKEITVDCKINNQKYRAKR    315
Segment 4   TRIEFVISPSDGNSGDVISCYSEVWEQPWRHKKEITVDCKINNQKYRAKR    420
Segment 5   TRIEFVISPSDGNSGDVISCYSEVWEQPWMHKKEITVDCKINNQKYRAKR    535
Segment 6   TRIEFVISPSDGNSGDVISCYSEVWEQPWMHKKEITVDCKINNQKYRA      628
```

Fig. 4. Amino acid sequence of MsCPI segments deduced from the nucleotide sequence of PCR products. The six MsCPI segments were aligned using ClustalW. Different amino acids within the segments are indicated as black boxes. The three conserved regions of cystatins are underlined. Two basic amino acid motifs, KR and RXXR, are double underlined. A probable disulfide bridge is shown by a bracket.

domains are members of the insect cystatin family. In Fig. 5, the amino acid sequence of segment 4 was compared with typical animal and plant cystatins including sarcocystatin A from *Sarcophaga peregrina* (Saito et al., 1989). It shared 16–35% identical residues with other cystatins, and showed the highest similarity (35% identity) with a cystatin from the saturnid moth, *Lonomia oblique* (Veiga et al., 2005).

3.3. Northern blotting analysis

Results from PCR analysis of the MsCPI gene indicated that the full-length MsCPI mRNA had a large number of cystatin repeats in the molecule. Therefore, attempts at cDNA cloning of the full-length cDNA by screening an *M. sexta* fat body cDNA library using 5′-RACE and 3′-RACE with specific primers were so far unsuccessful. To obtain the information about the length of the MsCPI mRNA, northern blotting was performed (Fig. 3b). Hybridization with the MsCPI probe produced a strong mRNA signal with a size of 12 kb. No minor bands of a smaller size were observed.

3.4. Expression of recombinant MsCPI protein

For further studies on inhibitory activity of MsCPI it was produced as a recombinant protein in *E. coli*. SDS-PAGE analysis of the total proteins from the IPTG-induced *E. coli* cells indicated that the recombinant protein corresponded to a major band with mobility on SDS-PAGE identical to that of mature MsCPI (14 kDa). After extraction of the insoluble rMsCPI in 8 M urea and refolding, it was purified by gel filtration on Sephadex G-50, and the purity was confirmed by SDS-PAGE (Fig. 6). A single band was obtained and rMsCPI was homogeneous, as judged from SDS-PAGE. Amino-terminal sequencing of rMsCPI yielded the sequence Met-Glu-Thr-Pro. This result indicated that the N-terminus of rMsCPI has an extra methionine residue. The yield of rMsCPI was about 5 mg/liter of induced culture.

3.5. Inhibitory activities of recombinant MsCPI toward various proteases

The inhibitory activity of rMsCPI was first investigated using papain and ficin, cysteine proteases derived from plants. Similar inhibition curves were obtained as in Fig. 2 and the inhibition constants for papain and ficin were calculated to be 0.71 and 9.8 nM, respectively.

The effect of rMsCPI on the activities of cathepsins B, H, and L, cysteine proteases of animal origin, were also examined. All cathepsins were strongly inhibited by rMsCPI and the inhibition constants for cathepsins B, H, and L were calculated to be 6.8, 3.0 and 0.87 nM, respectively. However, rMsCPI did not inhibit the serine, aspartic and metallo proteases, such as trypsin, chymotrypsin, subtilisin, pepsin, and thermolysin (data not shown).

4. Discussion

Proteinaceous cysteine protease inhibitors inactivate cysteine proteases from various organisms. In mammals, physiological roles of CPIs are recognized as regulatory proteins for endogenous cysteine proteases and also as defense proteins that inhibit exogenous cysteine proteases that facilitate infections caused by bacteria and viruses. Although it is thought that CPIs from insects have also similar functions, little is known about insect CPI proteins. In the present study, a novel CPI from *M. sexta* (MsCPI) was discovered and its cDNA nucleotide and amino acid sequences were investigated as well as its inhibitory potency to obtain a better understanding of insect CPIs.

MALDI TOF-MS analysis indicated that the MsCPI protein purified from the hemolymph had a molecular mass of 11,400–11,600 Da, whereas northern blot analysis

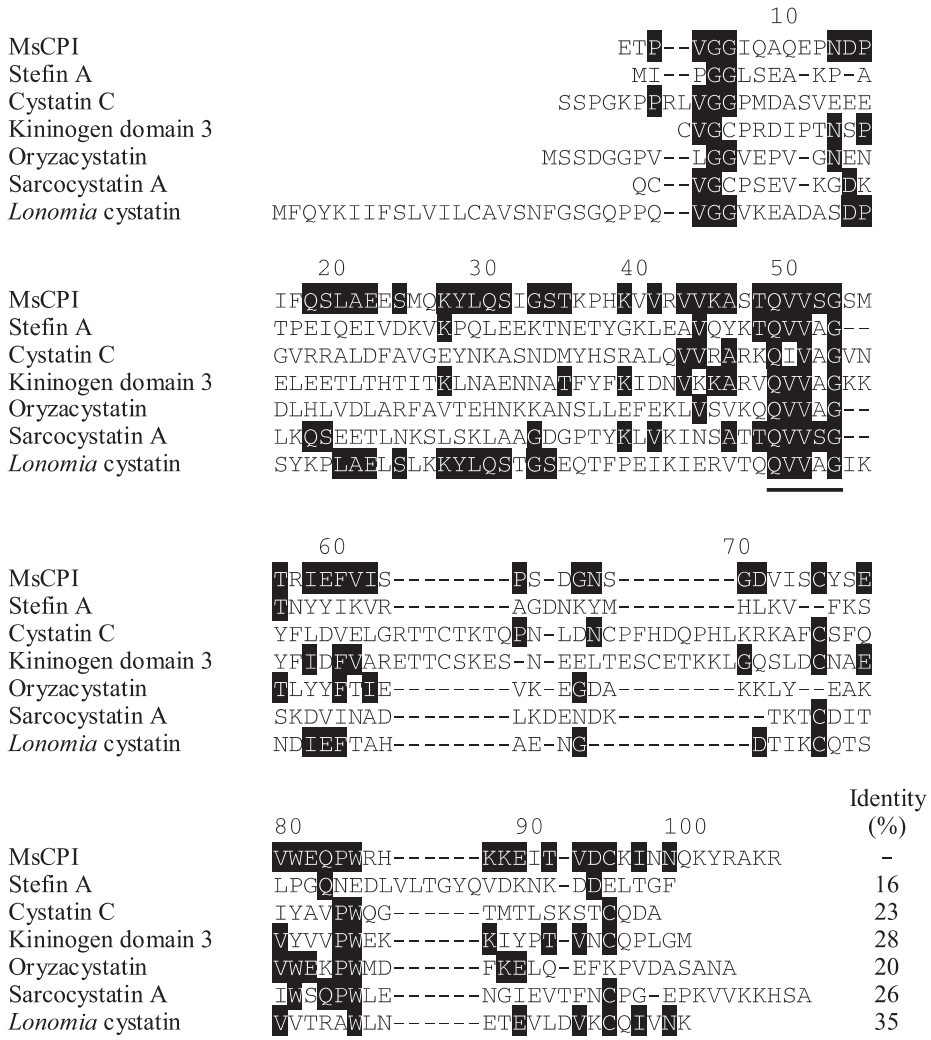

Fig. 5. Comparison of the amino acid sequences between MsCPI segment 4 and various cystatins. MsCPI segment 4 was aligned together with various cystatins using ClustalW and adjusted manually. Identical amino acids with MsCPI segment 4 are enclosed in black boxes. The conserved Q-X-V-X-G sequence was underlined. Stefin A, human stefin A; Cystatin C, human cystatin C; Kininogen domain 3, human low molecular weight kininogen domain 3; Oryzacyatatin I, oryzacystatin I from rice; Sarcocystatin A, cystatin from *Sarcophaga peregrina*; and *Lonomia* cystatin, putative cystatin from *Lonomia obliqua*.

(Fig. 3b) indicated that the message was extremely large (12 kb). Molecular cloning of MsCPI gene indicated that MsCPI belongs to one of the cystatin family of proteinaceous inhibitors and also the presence of a large number of tandem-repeat MsCPI gene segments each encoding a domain that contains 105 amino acid residues (Fig. 4). One type of mammalian cysteine protease inhibitor with tandem repeats of a cystatin domain is known as the kininogens, which are a member of the mammalian cystatin family (Ohkubo et al., 1984). Kininogens, glycoproteins with molecular masses of 50–114 kDa, comprise three cystatin domains in the N-terminal region and an additional extension in the C-terminal region, from which a peptide called kinin is derived (Higashiyama et al., 1986; Ohkubo et al., 1988). In the phytocystatin family, a potato multicystatin (PMC) has been described (Rodis and Hoff, 1984; Waldron et al., 1993; Walsh and Strickland, 1993). PMC found in potato tubers in the form of proteinaceous crystals consists of a polypeptide of about 85 kDa and has eight repeats of a cystatin domain. Therefore, PMC potentially can inhibit eight papain molecules simultaneously. There is also a multicystatin from sunflower seeds, which has three repeats of a cystatin domain (Kouzuma et al., 2000). Some tandem-repeat types of cystatins occur in a multiple-domain form. However, the MsCPI protein was purified in a single-domain form, suggesting that the large MsCPI precursor protein was post-translationally processed to produce multiple copies of the mature single-domain protein. Each copy contains only a few amino acid substitutions. In the C-terminal regions of all of the MsCPI segments, there are two basic amino acid motifs, KR and RXXR (Fig. 4). In the processing of typical prohormones, active peptides flanked by paired basic residues undergo proteolytic cleavage at the N- or C-terminal side of those residues or between the two basic residues by cathepsin L or

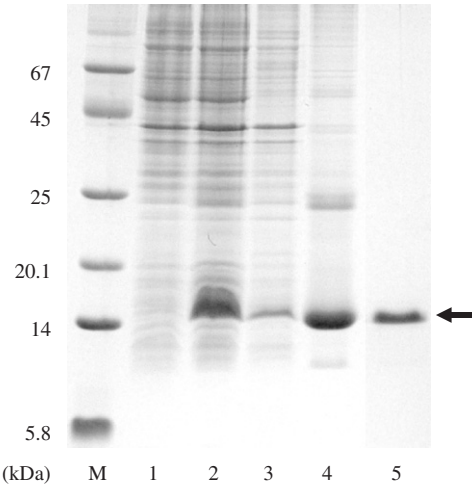

Fig. 6. Overproduction of the recombinant MsCPI in *E. coli* cells. Arrows indicate the position of expressed rMsCPI protein. Lane 1, total cell extract of *E. coli* without induction; lane 2, total cell extract after induction by IPTG; Lane 3, soluble fraction of *E. coli* after sonication; Lane 4, insoluble fraction after sonication; Lane 5, purified rMsCPI; M, molecular weight markers: bovine serum albumin (67 kDa), ovalbumin (45 kDa), α-chymotrypsinogen (25 kDa), soybean trypsin inhibitor (20.1 kDa), lysozyme (14 kDa) and insulin (5.8 kDa).

prohormone convertases (Hook et al., 2004). In another case, general precursor proteins with the RXXR motif are cleaved at the C-terminal side of the second Arg residue by furin-like enzymes during processing (Nakayama, 1997). The theoretical molecular masses of MsCPI segments 1–5 (Fig. 4) were from 11,769 to 11,909 Da. These sizes were on average about 300 Da larger than that of the purified MsCPI protein. This difference indicated that two amino acids are eliminated during the processing of the MsCPI precursor protein. It is possible that the MsCPI precursor is cleaved by processing enzymes during maturation and two basic residues KR might be removed by exopeptidases such as Arg/Lys aminopeptidases (Hook and Eiden, 1984) or carboxypeptidases E/H (Fricker, 1988).

The yield of MsCPI protein purified from hemolymph is very low (Table 1). Therefore, a functional expression system for recombinant MsCPI protein (rMsCPI) in *E. coli* cells was constructed to obtain sufficient protein for characterization of the properties of MsCPI. The recombinant protein obtained inhibited papain effectively with the K_i value of 0.71 nM. This result indicated that the inhibitory activity of rMsCPI for papain was 7.7-fold stronger than that of the native MsCPI protein ($K_i = 5.5$ nM). The main reason for the difference of inhibitory potency between authentic and recombinant MsCPI is that the rMsCPI protein was composed of a single component with the amino acid sequence of segment 4 (Fig. 4); on the other hand, authentic MsCPI was composed of a mixture of isoinhibitors that had various amino acid sequences and different inhibitory potencies. Therefore, the value of the inhibition constant of authentic MsCPI (5.5 nM) is probably a mean value of those of the isoinhibitors. rMsCPI also inhibited mammalian cathepsins B, H, and L very effectively. Cathepsins are lysosomal proteases involved in the degradation of intracellular proteins. In insects, one of the cysteine proteases, cathepsin L, acts as a regulator of metamorphosis, and is essential for the differentiation of imaginal discs (Homma et al., 1994). Furthermore, cathepsin B is involved in dissociation of the fat body during metamorphosis (Takahashi et al., 1993). One of roles of MsCPI might be the regulation of proteases that are involved in metamorphosis.

In a few other insect species, the presence of cystatin genes has been confirmed. Among them are sarcocystatin A from *S. peregrina* (Saito et al., 1989) and cystatins from *D. melanogaster* (Delbridge and Kelly, 1990) and *S. crassipalpis* (Goto and Denlinger, 2002). These all have the three conserved regions of a cystatin, a Gly in the N-terminal region, Q-X-V-X-G in the middle region, and P-W in the C-terminal region (Rawlings et al., 2004). Furthermore, two cysteine residues in the C-terminal region are characteristic. These structural motifs are also conserved in all MsCPI segments (Fig. 4). However, the length of the cystatin cDNAs from *S. crassipalpis* and *D. melanogaster* are only 0.7–0.8 kb and larger signals are not detected in northern blotting analysis (Delbridge and Kelly, 1990; Goto and Denlinger, 2002). This result shows that *S. crassipalpis* and *D. melanogaster* have genes encoding only a single-cystatin domain, which is in contrast to the multi-domain-encoding gene present in *M. sexta*.

Recently, a new cystatin with six tandemly repeated domains was reported (Khaznadji et al., 2005). A multi-domain cystatin gene was found in the parasitic trematode *Fasciola hepatica*, whose infection causes fasciolosis. The cDNA (2359 bp) encodes a putative protein of 690 amino acids with six repeated cystatin domains of approximately 100 amino acids. The inhibitory potency of one cystatin domain toward cathepsin L was confirmed by using the recombinant protein. However, the potency of the other cystatin domains, which do not have all three of the consensus regions, G, Q-X-V-X-G, and P-W, was undetermined. In contrast, all of the consensus regions are conserved in the individual MsCPI segements (Fig. 4). We have not obtained an MsCPI gene segment that lacks any of the consensus regions.

Results from this and unpublished studies of MsCPI cDNA indicate that the gene for MsCPI encodes more than 10 cystatin domains, several functionally unknown domains, and one putative protease domain (unpublished data). We conclude that MsCPI is a novel type of cystatin that is synthesized as part of a multi-domain precursor protein. Although the complete structure of the precursor and its processing mechanism are not yet understood, nevertheless our results provide some new insights for insect biological research on proteolytic enzyme inhibitors. Studies on the structure of the MsCPI precursor, structure-function relationships of the each domain, and the precursor processing mechanism are in progress.

Acknowledgements

We thank Prof. Makoto Kimura, Kyushu University, for obtaining the MALDI-TOF mass spectra and Dr. Neal Dittmer, Kansas State University, for preparing RNA from tobacco hornworm fat body.

Mention of trade names or commercial products in this publication is solely for providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

References

Barrett, A.J., Rawlings, N.D., Davis, M.E., Machleidt, W., Salvesen, G., Turk, V., 1986. Cysteine proteinase inhibitors of the cystatin superfamily. In: Barrett, A.J., Salvesen, G. (Eds.), Proteinase Inhibitors. Elsevier Science Publishers BV, New York, pp. 515–569.

Bode, W., Engh, R., Musil, D., Thiele, U., Huber, R., Karshikov, A., Brzin, J., Kos, J., Turk, V., 1988. The 2.0 Å X-ray crystal structure of chicken egg white cystatin and its possible mode of interaction with cysteine proteinases. EMBO J. 7, 2593–2599.

Delbridge, M.L., Kelly, L.E., 1990. Sequence analysis, and chromosomal localization of a gene encoding a cystatin-like protein from *Drosophila melanogaster*. FEBS Lett. 274, 141–145.

Denizot, F., Brunet, J.F., Roustan, P., Harper, K., Suzan, M., Luciani, M.F., Mattei, M.G., Golstein, P., 1989. Novel structures CTLA-2 alpha and CTLA-2 beta expressed in mouse activated T cells and mast cells and homologous to cysteine proteinase proregions. Eur. J. Immunol. 19, 631–635.

Deveraux, Q.L., Reed, J.C., 1999. IAP family proteins—suppressors of apoptosis. Genes. Dev. 13, 239–252.

Dubin, G., 2005. Proteinaceous cysteine protease inhibitors. Cell Mol. Life Sci. 62, 653–669.

Dunn, P.E., Drake, D., 1983. Fate of bacterial infected into naïve and immunized larvae of the tobacco hornworm, *Manduca sexta*. J. Invertebr. Pathol. 41, 77–85.

Filippova, I.Y., Lysogorskaya, E.N., Oksenoit, E.S., Rudenskaya, G.N., Stepanov, V.M., 1984. L-Pyroglutamyl-L-phenylalanyl-L-leucine-p-nitroanilide—a chromogenic substrate for thiol proteinase assay. Anal. Biochem. 143, 293–297.

Fricker, L.D., 1988. Carboxypeptidase E. Annu. Rev. Physiol. 50, 309–321.

Goto, S.G., Denlinger, D.L., 2002. Genes encoding two cystatins in the flesh fly *Sarcophaga crassipalpis* and their distinct expression patterns in relation to pupal diapause. Gene 292, 121–127.

Henderson, P.J., 1972. A linear equation that describes the steady-state kinetics of enzymes and subcellular particles interacting with tightly bound inhibitors. Biochem. J. 127, 321–333.

Higashiyama, S., Ohkubo, I., Ishiguro, H., Kunimatsu, M., Sawaki, K., Sasaki, M., 1986. Human high molecular weight kininogen as a thiol proteinase inhibitor: presence of the entire inhibition capacity in the native form of heavy chain. Biochemistry 25, 1669–1675.

Homma, K., Kurata, S., 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.

Hook, V., Yasothornsrikul, S., Greenbaum, D., Medzihradszky, K.F., Troutner, K., Toneff, T., Bundey, R., Logrinova, A., Reinheckel, T., Peters, C., Bogyo, M., 2004. Cathepsin L and Arg/Lys aminopeptidase: a distinct prohormone processing pathway for the biosynthesis of peptide neurotransmitters and hormones. Biol. Chem. 385, 473–480.

Hook, V.Y., Eiden, L.E., 1984. Two peptidases that convert 125I-Lys-Arg-(Met)enkephalin and 125I-(Met)enkephalin-Arg6, respectively, to 125I-(Met)enkephalin in bovine adrenal medullary chromaffin granules. FEBS Lett. 172, 212–218.

Jiang, H., Wang, Y., Yu, X.Q., Kanost, M.R., 2003. Prophenoloxidase-activating proteinase-2 from hemolymph of *Manduca sexta*. A bacteria-inducible serine proteinase containing two clip domains. J. Biol. Chem. 278, 3552–3561.

Khaznadji, E., Collins, P., Dalton, J.P., Bigot, Y., Moire, N., 2005. A new multi-domain member of the cystatin superfamily expressed by *Fasciola hepatica*. Int. J. Parasitol. 35, 1115–1125.

Kouzuma, Y., Inanaga, H., Doi-Kawano, K., Yamasaki, N., Kimura, M., 2000. Molecular cloning and functional expression of cDNA encoding the cysteine proteinase inhibitor with three cystatin domains from sunflower seeds. J. Biochem. (Tokyo) 128, 161–166.

Laemmli, U.K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227, 680–685.

Lenarcic, B., Bevec, T., 1998. Thyropins—new structurally related proteinase inhibitors. Biol. Chem. 379, 105–111.

Nakayama, K., 1997. Furin: a mammalian subtilisin/Kex2p-like endoprotease involved in processing of a wide variety of precursor proteins. Biochem. J. 327, 625–635.

Ohkubo, I., Kurachi, K., Takasawa, T., Shiokawa, H., Sasaki, M., 1984. Isolation of a human cDNA for alpha 2-thiol proteinase inhibitor and its identity with low molecular weight kininogen. Biochemistry 23, 5691–5697.

Ohkubo, I., Namikawa, C., Higashiyama, S., Sasaki, M., Minowa, O., Mizuno, Y., Shiokawa, H., 1988. Purification and characterization of alpha 1-thiol proteinase inhibitor and its identity with kinin- and fragment 1.2-free high molecular weight kininogen. Int. J. Biochem. 20, 243–258.

Rawlings, N.D., Tolle, D.P., Barrett, A.J., 2004. Evolutionary families of peptidase inhibitors. Biochem. J. 378, 705–716.

Rodis, P., Hoff, J.E., 1984. Naturally occurring protein crystals in the potato. Plant Physiol. 74, 907–911.

Rzychon, M., Chmiel, D., Stec-Niemczyk, J., 2004. Modes of inhibition of cysteine proteases. Acta Biochim. Pol. 51, 861–873.

Saito, H., Suzuki, T., Ueno, K., Kubo, T., Natori, S., 1989. Molecular cloning of cDNA for sarcocystatin A and analysis of the expression of the sarcocystatin A gene during development of *Sarcophaga peregrina*. Biochemistry 28, 1749–1755.

Stubbs, M.T., Laber, B., Bode, W., Huber, R., Jerala, R., Lenarcie, B., Turk, V., 1990. The refined 2.4 Å X-ray crystal structure of recombinant human stefin B in complex with the cysteine proteinase papain: a novel type of proteinase inhibitor interaction. EMBO J. 9, 1939–1947.

Suzuki, T., Natori, S., 1985. Purification and characterization of an inhibitor of the cysteine protease from the hemolymph of *Sarcophaga peregrina* larvae. J. Biol. Chem. 260, 5115–5120.

Takahashi, N., Kurata, S., Natori, S., 1993. Molecular cloning of cDNA for the 29 kDa proteinase participating in decomposition of the larval fat body during metamorphosis of *Sarcophaga peregrina* (flesh fly). FEBS Lett. 334, 153–157.

Veiga, A.B., Ribeiro, J.M., Guimaraes, J.A., Francischetti, I.M., 2005. A catalog for the transcripts from the venomous structures of the caterpillar *Lonomia obliqua*: identification of the proteins potentially involved in the coagulation disorder and hemorrhagic syndrome. Gene 355, 11–27.

Waldron, C., Wegrich, L.M., Merlo, P.A.O., Walsh, T.A., 1993. Characterization of a genomic sequence coding for potato multicystatin, an eight-domain cysteine proteinase inhibitor. Plant Mol. Biol. 23, 801–812.

Walsh, T.A., Strickland, J.A., 1993. Proteolysis of the 85-kilodalton crystalline cysteine proteinase inhibitor from potato releases functional cystatin domains. Plant Physiol. 103, 1227–1234.

Yamamoto, Y., Watabe, S., Kageyama, T., Takahashi, S.Y., 1999. Purification and characterization of *Bombyx* cysteine proteinase specific inhibitors from the hemolymph of *Bombyx mori*. Arch. Insect. Biochem. Physiol. 42, 119–129.