Insect Biochemistry and Molecular Biology 40 (2010) 835–846

Contents lists available at ScienceDirect

Insect Biochemistry and Molecular Biology

journal homepage: www.elsevier.com/locate/ibmb

Molecular cloning of a multidomain cysteine protease and protease inhibitor precursor gene from the tobacco hornworm (*Manduca sexta*) and functional expression of the cathepsin F-like cysteine protease domain

Takayuki Miyaji [a], Satoshi Murayama [a], Yoshiaki Kouzuma [a], Nobutada Kimura [b], Michael R. Kanost [c], Karl J. Kramer [d], 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] Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki 305-8566, Japan
[c] Department of Biochemistry, Kansas State University, Manhattan, KS 66506, USA
[d] Center for Grain and Animal Health Research, Agricultural Research Service, United States Department of Agriculture, 1515 College Avenue, Manhattan, KS 66502, USA

ARTICLE INFO

Article history:
Received 3 June 2010
Received in revised form
10 August 2010
Accepted 10 August 2010

Keywords:
Cysteine protease inhibitor
Cystatin
Cathepsin F-like protease
Manduca sexta
Multifunctional domains

ABSTRACT

A *Manduca sexta* (tobacco hornworm) cysteine protease inhibitor, MsCPI, purified from larval hemolymph has an apparent molecular mass of 11.5 kDa, whereas the size of the mRNA is very large (~9 kilobases). MsCPI cDNA consists of a 9,273 nucleotides that encode a polypeptide of 2,676 amino acids, which includes nine tandemly repeated MsCPI domains, four cystatin-like domains and one procathepsin F-like domain. The procathepsin F-like domain protein was expressed in *Escherichia coli* and processed to its active mature form by incubation with pepsin. The mature enzyme hydrolyzed Z-Leu–Arg–MCA, Z-Phe–Arg–MCA and Boc–Val–Leu–Lys–MCA rapidly, whereas hydrolysis of Suc–Leu–Tyr–MCA and Z-Arg–Arg–MCA was very slow. The protease was strongly inhibited by MsCPI, egg-white cystatin and sunflower cystatin with K_i values in the nanomolar range. When the MsCPI tandem protein linked to two MsCPI domains was treated with proteases, it was degraded by the cathepsin F-like protease. However, tryptic digestion converted the MsCPI tandem protein to an active inhibitory form. These data support the hypothesis that the mature MsCPI protein is produced from the MsCPI precursor protein by trypsin-like proteases. The resulting mature MsCPI protein probably plays a role in the regulation of the activity of endogenous cysteine proteases.

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 their protein structure and inhibitory activity. According to the MEROPS listing (Rawlings et al., 2004), more than 10 families of CPIs have been classified based on the amino acid sequences of inhibitory domains. Cystatins (family I25) are the best characterized CPI group of mammalian and plant origins. On the basis of sequence homology, the cystatin superfamily is divided into three subfamilies, I25A (cystatin A and sarcocystatin), I25B (cystatin C and phytocystatin) and I25C (kininogens). Cystatins are believed to be involved in the regulation of endogenous cysteine protease activity and/or in defensive mechanisms against invading organisms such as bacteria, viruses and parasites (Dubin, 2005). Cystatins from animal origin target lysosomal cysteine proteases, generally known as cathepsins. The cysteine proteases are involved both in the physiological protein breakdown (cathepsins B, C, F, H, L, O and X) and specific functions correlated with their tissue distribution (cathepsins K, S and V). The cysteine proteases are optimally active in the slightly acidic, reducing milieu found in lysosomes. They comprise a group of papain-related enzymes, sharing similar amino acid sequences and folds (Turk et al., 2001).

Generally, cystatins have 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 structure that interacts with the active site cleft of the cognate cysteine proteases, papain and cathepsin H (Rzychon et al., 2004; Jenko et al., 2003). Among the three regions, the Q–X–V–X–G motif is critical for

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

0965-1748/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ibmb.2010.08.003

inhibition of cysteine proteases as revealed from the mutational analysis of egg-white cystatin (Auerswald et al., 1992).

In insects, cystatins include sarcocystatin from *Sarcophaga peregrina* and a cystatin-like protein from *Drosophila melanogaster* (Suzuki and Natori, 1985; Delbridge and Kelly, 1990). Saito et al. (1989) suggested that sarcocystatin A from *S. peregrina* is involved in morphogenesis of larval and adult structures of *Sarcophaga*. Besides cystatins, two other types of CPIs, the propeptide region of a cysteine protease homologous protein from *Bombyx mori* (family I29) and an inhibitor of apoptosis protein from *D. melanogaster* (family I32) also are found in insects (Yamamoto et al., 1999; Hay, 2000). However, relatively little information is known about CPIs from insects compared with mammals and plants. To obtain a better understanding about structures and functions of novel CPIs from insects, we investigated a CPI from the tobacco hornworm (*Manduca sexta*).

Previously, a CPI from *M. sexta*, MsCPI, was purified from larval hemolymph of the tobacco hornworm and the partial nucleotide sequence was determined (Miyaji et al., 2007). Although the MsCPI protein purified from the hemolymph had a small molecular mass of 11,400–11,600 Da, northern blot analysis indicated that the message was extremely large (~9 kb). Molecular cloning of the MsCPI cDNA revealed that the encoded protein belonged to the cystatin family and was composed of at least six tandemly repeated MsCPI segments.

To determine the complete structure of the MsCPI gene, we performed genomic and cDNA cloning of the full-length gene, and functionally expressed the procathepsin F-like domain that is a part of the MsCPI precursor. In addition, we determined the conceptual amino acid sequence of MsCPI precursor protein and the interaction that occurs between MsCPI and the cathepsin F-like protease.

2. Materials and methods

2.1. Screening of MsCPI cDNA from M. sexta fat body cDNA library

An *M. sexta* fat body cDNA library that was previously constructed (Jiang et al., 2003) was screened by plaque hybridization using a partial MsCPI cDNA (324 bp) labeled with digoxigenin as a probe (Miyaji et al., 2007). Six positive clones were isolated and phage DNAs were prepared from them. The DNA with the longest insert was digested with *Eco*RI and *Xho*I, and the DNA fragment was subcloned into the plasmid vector pUC18 or pBluescript SK^-. The nucleotide sequence was determined by the dideoxy chain termination method using the BigDye Terminator v3.1 cycle sequencing Ready Reaction Kit and automated DNA sequencer ABI prism 3130xl (Applied Biosystems Inc.).

2.2. Isolation of genomic DNA from an M. sexta larva

Genomic DNA was prepared from a fifth instar *M. sexta* larva using the method of Blin and Stafford (1976), as follows. The larva was frozen in liquid nitrogen and ground to a powder. The powder was then dissolved in extraction buffer (20 μg/ml RNase A, 0.5% SDS, 100 mM EDTA and 10 mM Tris–HCl, pH 8.0). After incubation at 37 °C for 1 h, the solution was gently mixed with proteinase K at a final concentration of 100 μg/ml at 50 °C for 3 h, and then incubated for 10 min with an equal volume of phenol equilibrated with TE (1 mM EDTA, 10 mM Tris-HCl, pH 8.0). After three extractions with phenol, the supernatant was mixed thoroughly with 0.1 volume of 3 M sodium acetate, pH 4.6 and 2 volumes of ethanol. The precipitated DNA was recovered and washed using 70% ethanol twice. The purity and concentration of the genomic DNA thus obtained were confirmed by agarose gel electrophoresis and by measuring the absorbance at 260 nm.

2.3. M. sexta genomic library construction and screening

An *M. sexta* genomic library was constructed using the Copy-Control™ Fosmid Library Production Kit (EPICENTRE) according to the manufacturer's instructions. *M. sexta* genomic DNA was incubated with the End-Repair Enzyme Mix at room temperature for 45 min. Approximately 40 kb end-repaired DNAs were resolved by 1% low melting point agarose gel electrophoresis. The 40 kb fragments were recovered from the gel using GELase (45 °C for 1 h). The size-fractionated DNA was ligated with the CopyControl pCC1FOS vector at room temperature for 2 h and then the mixture was incubated with the MaxPlax Lambda Packaging Extracts at 37 °C for 180 min. The packaged CopyControl fosmid clones were incubated with *Escherichia coli* EPI300 cells at 37 °C for 20 min after which the infected EPI300 cells were spread on LB plate containing 12.5 μg/ml chloramphenicol and incubated at 37 °C overnight. The *M. sexta* fosmid library thus constructed was screened by colony hybridization using a 324-bp MsCPI gene fragment labeled with digoxigenin as a probe. After the second and third screening, one positive clone was isolated and fosmid DNA was prepared from it. To confirm the insert position of MsCPI precursor gene, the fosmid DNA was digested with various restriction enzymes and Southern hybridization was performed. The DNA fragments obtained by *Eco*RI, *Hin*dIII or *Pst*I digestions were subcloned into the plasmid vector pUC18 and the nucleotide sequences were determined.

2.4. Southern hybridization

The fosmid DNA from the positive clone was digested with *Bgl*II, *Eco*RI, *Hin*dIII, *Pst*I, *Sac*I, *Sal*I, *Sph*I, *Xba*I, or *Xho*I. The DNA fragments were resolved by agarose gel electrophoresis and transferred to a Hybond-N^+ nylon membrane (GE Healthcare), after which the membrane was baked 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, GE Healthcare) for use as a probe. Hybridization was performed in hybridization buffer (4% blocking reagent/0.5 M NaCl) at 55 °C overnight. 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 an LAS 3000 Mini Imaging System (Fuji Film).

2.5. 5′-RACE and RT-PCR of MsCPI gene

Isolation of total RNA from fat body of *M. sexta* was performed as previously reported (Miyaji et al., 2007). 5′-RACE and *RT-PCR* were conducted using the GeneRacer™ Kit (Invitrogen) according to the manufacturer's instructions. Briefly, full-length mRNA with a 5′-cap structure was decapped by tobacco acid pyrophosphatase and then ligated with RNA Oligo. First-strand cDNA was synthesized by SuperScript® III Reverse Transcriptase using the mRNA and MsCPI specific primers. PCR was performed using the first-strand cDNA as a template and gene-specific primers, and the PCR products were purified by using the Wizard® SV Gel and PCR Clean-Up System (Promega). The purified PCR products were cloned into the pGEM T-Easy vector (Promega) and then sequenced.

2.6. Expression of procathepsin F-like protein in E. coli

MsCPI cDNA prepared from a positive clone from *M. sexta* fat body cDNA library was digested with *Bam* HI and the DNA fragment including the region encoding the procathepsin F-like domain