Int. J. Indust. Entomol. Vol. 4, No. 1, 2002, pp. 63~68

International Journal of *Industrial Entomology*

Molecular Cloning of a cDNA Encoding a Cathepsin B Homologue from the Mulberry Longicorn Beetle, *Apriona germari*

Seong Ryul Kim[1], Hyung Joo Yoon[2], Nam Sook Park[1], Sang Mong Lee[3], Jae Yu Moon[4], Byung Rae Jin[1] and Hung Dae Sohn[1,*]

[1]College of Natural Resources and Life Science, Dong-A University, Busan 604-714, Korea. [2]Department of Sericulture and Entomology, National Institute of Agricultural Science and Technology, RDA, Suwon 441-100, Korea.
[3]Department of Sericultural and Entomological Biology, Miryang National University, Miryang 627-130, Korea.
[4]College of Agriculture and Life Sciences, Seoul National University, Suwon 441-744, Korea.

(Received 12, January 2002; Accepted 7, February 2002)

A cDNA encoding a putative member of cathepsin B of the thiol protease superfamily was cloned from a cDNA library of the mulberry longicorn beetle, *Apriona germari*. Sequence analysis of the cDNA encoding the cathepsin B of *A. germari* (AgCatB) revealed that the 972 bp cDNA has an open reading frame of 324 amino acid residues. The deduced protein sequence of the AgCatB showed high homology with cathepsin B of the insects, *Bombyx mori* (47.3% amino acid identity), *Helicoverpa armigera* (46.6%) and *Sarcophaga peregrina* (45.6%), and the lowest homology with *Aedes aegypti* (33.2%). The AgCatB contains six disulfate bonds typical for cysteine proteases. The three amino acid positions Cys-109, His-267, and Asn-287 which are conserved, active sites characteristic for cathepsin B, were also found. Phylogenetic analysis further confirmed that the AgCatB has a close relationship with that of *B. mori*, *H. armigera* and *S. peregrina*.

Key words : Mulberry longicorn beetle, *Apriona germari*, cDNA cloning, Cathepsin B

Introduction

Cathepsin B is a cysteine protease possessing both endopeptidase and peptidyldipeptidase activities, and plays an important function for intracellular protein catabolism in the lysosomal system. The cysteine proteases are synthesized as inactive precursors and become activated after proteolytic removal of the N-terminal propeptide (Rawlings and Barrett, 1994).

In insects and other arthropods, cathepsin L as well as cathepsin B participate in the key developmental processes. Cathepsin B has been characterized both enzymatically and molecularly. Mosquito cathepsin B (Cho *et al.*, 1999), silkworm cathepsin L (Yamamoto *et al.*, 1994) and cotton bollworm cathepsin L (Zhao *et al.*, 1998) have been implicated in yolk protein degradation during embryonic development. Moreover, *Sarcophaga peregrina* cathepsin B and L (Takahashi *et al.*, 1993), and silkworm cathepsin B (Xu and Kawasaki, 2001) are known to involved in insect metamorphosis.

The Cerambycidae, commonly known as long-horned beetles is one of the largest groups in Coleoptera. The family has about 20,000 species throughout the world and most species of the family are wood-bores (Crowson, 1981; Daly *et al.*, 1998; Yoon *et al.*, 2001). Although the long-horned beetles are known as the largest groups in Coleoptera, little genetic information is available at the molecular level. Of the long-horned beetles, the mulberry longicorn beetle, *Apriona germari*, is an abundant species in Korea (Yoon *et al.*, 1997). Recently, cathepsins D from *A. germari* have been identified (Kim *et al.*, 2001).

In order to obtain genetic information of the mulberry longicorn beetle, we have previously constructed *A. germari* cDNA library from the larval whole body (Kim *et al.*, 2001). In this study we have cloned and characterized a cDNA encoding the cathepsin B homologue from the cDNA library of *A. germari*. The cloning, sequencing and characterization of the *A. germari* cathepsin D homologue gene are described in this paper.

*To whom correspondence should be addressed.
College of Natural Resources and Life Science, Dong-A University, Busan 604-714, Korea. Tel.: +82-51-200-7553;
E-mail: hdsohn@mail.donga.ac.kr

Materials and Methods

Animals

The larvae of the mulberry longicorn beetle, *Apriona germari* were collected from the mulberry tree branch of the wild mulberry tree field in Korea. *A. germari* was reared on an artificial diet as described previously (Yoon and Mah, 1999).

cDNA library screening, nucleotide sequencing and data analysis

A cDNA library was constructed from the poly(A)+

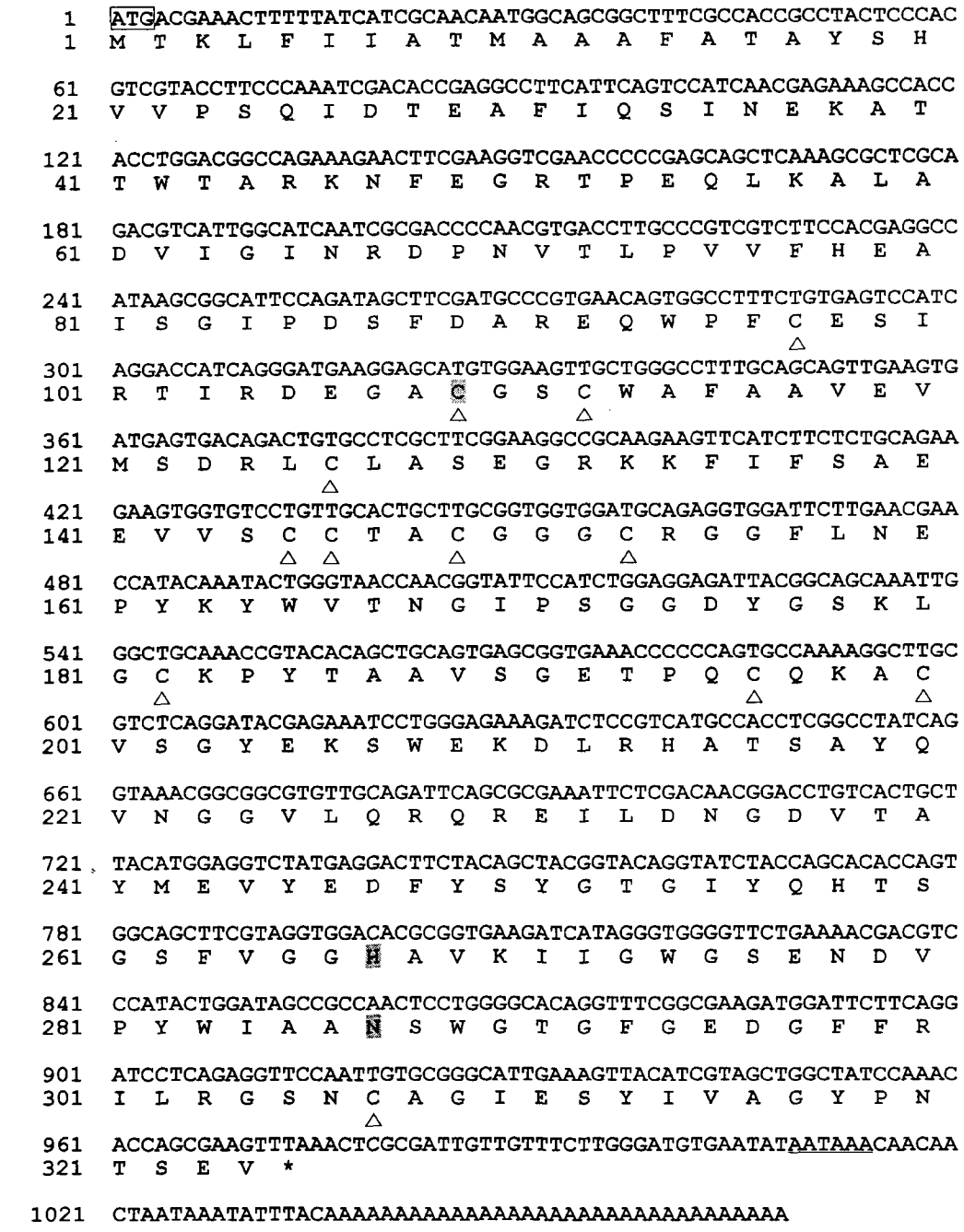

Fig. 1. The nucleotide and deduced amino acid sequences of *A. germari* cathepsin B homologue gene. The start codon of ATG is boxed and the termination codon is asterisk. The polyadenylation signal AATAAA is underlined. Twelve cysteine residues are marked with open triangles. Three active sites characteristic for cathepsin B at positions Cys-109, His-267, and Asn-287 are shaded with squares. The GenBank accession number is AF483623.

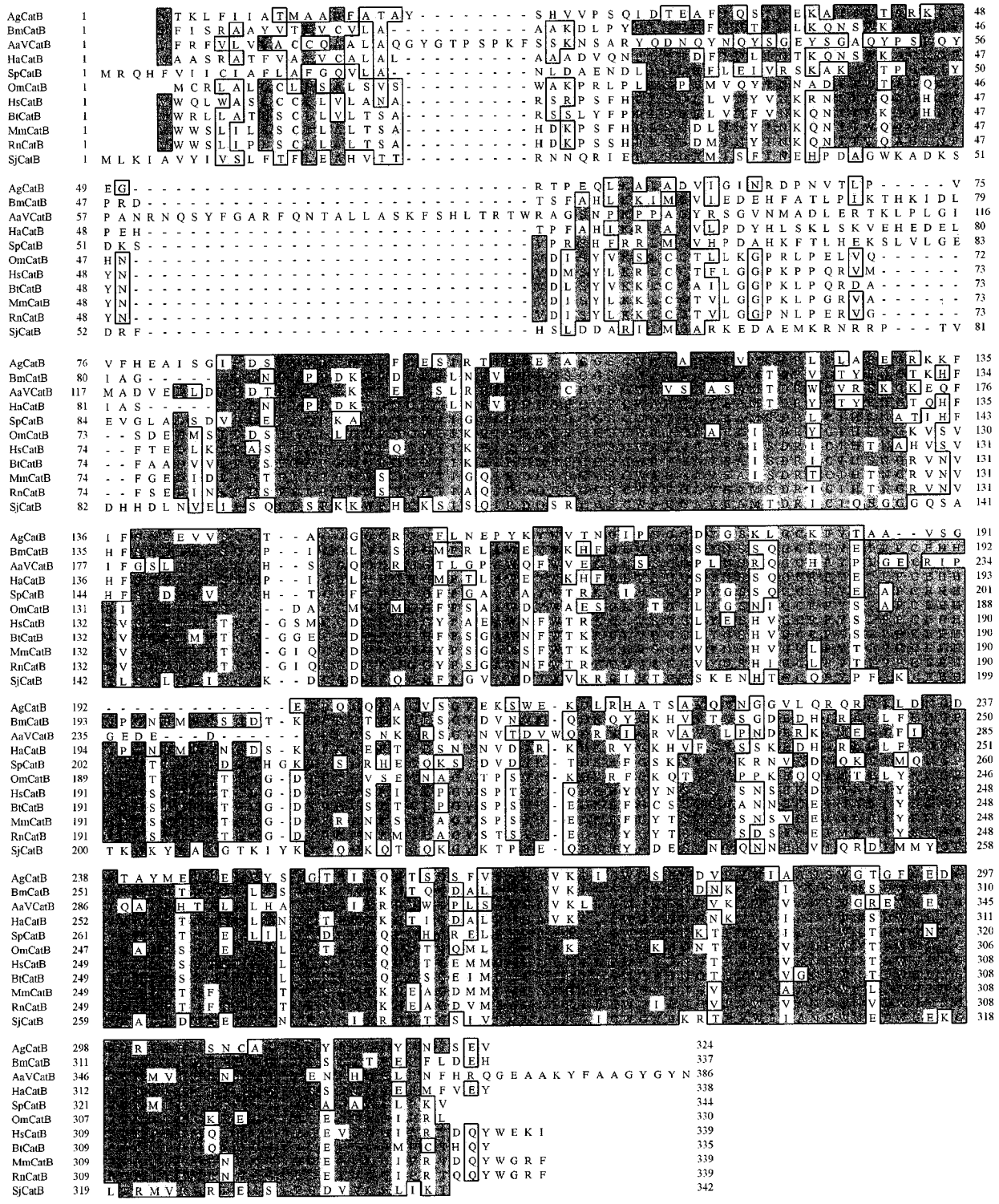

Fig. 2. Multiple sequence alignment of the deduced protein sequence of the *A. germari* cathepsin B homologue gene with other cathepsin B protein sequences. In solid box are the residues that are identical to those in *A. germari* cathepsin B homologue protein (AgCatB). The other cathepsin B protein sequences aligned were: *B. mori* (BmCatB; Xu and Kawasaki, 2001), *A. aegypti* (AaV-CatB; Cho *et al.*, 1999), *H. armigera* (HaCatB; Zhao *et al.*, 1998), *S. peregrina* (SpCatB; Takahashi *et al.*, 1993), *Oncorhynchus mykiss* (OmCatB; Kwon *et al.*, 2001), *Homo sapiens* (HsCatB; Chan *et al.*, 1986), *Bos taurus* (BtCatB; Bechet *et al.*, 1991), *Mus musculus* (MmCatB; Chan *et al.*, 1986), *Rattus norvegicus* (RnCatB; Guenette *et al.*, 1994) and *Schistosoma japonicum* (SjCatB; Merckelbach *et al.*, 1994). Three active sites characteristic for cathepsin B are asterisk.

mRNA isolated from the whole body of *A. germari* larvae by Uni-ZAP XR vector and Gigapack III Gold Packing Extract (Stratagene) (Kim *et al.*, 2001). The sequencing of randomly selected clones harboring cDNA inserts was performed to generate the expressed sequence tags (ESTs).

For DNA sequencing, plasmid DNA was extracted by Wizard mini-preparation kit (Promega). Sequence of each cDNA clone was determined using an automatic sequencer (model 310 Genetic Analyzer; Perkin-Elmer Applied Biosystems, Foster City, CA). The sequences were compared using the DNASIS and BLAST programs provided by the NCBI. GenBank, EMBL and SwissProt databases were searched for sequence homology using a BLAST algorithm program.

Phylogenetic analysis

With the GenBank-registered amino acid sequences of insect cathepsin B genes, phylogenetic analysis among the deduced amino acid sequences was performed with the PAUP (Phylogenetic Analysis using Parsimony) version 3.1 (Swofford, 1990). The accession numbers of the sequences in the GenBank are as follows: mulberry longicorn beetle *A. germari* (AF483623, this study), silkworm *Bombyx mori* (AB045595; Xu and Kawasaki, 2001), cotton bollworm *Helicoverpa armigera* (AF 222788; Zhao *et al.*, 1998), flesh fly *Sarcophaga peregrina* (D16823; Takahashi *et al.*, 1993), and yellow fever mosquito *Aedes aegypti* (AF127592; Cho *et al.*, 1999).

Results and Discussion

Construction of cDNA library was prepared from the whole body of *A. germari* larvae (Kim *et al.*, 2001). The sequencing of randomly selected clones harboring cDNA inserts was performed to generate the *A. germari* ESTs. Of these ESTs, one exhibited similarity to the reported cathepsins B. The complete DNA sequence of a cDNA encoding a putative member of the insect cathepsin B gene family designated AgCatB revealed that the 972 bp cDNA has an open reading frame of 324 amino acid residues (GenBank accession number AF483623) (Fig. 1). In the AgCatB gene sequence, a polyadenylation signal AATAAA was found at nucleotides 1,008, twenty-seven bp upstream of the poly(A) tail. The AgCatB indicates an enzyme consisted of a prepro-protein of 324 amino acid residues with a predicted molecular mass of approximately XX kDa. The AgCatB contains six disulfate bonds typical for thiol (cysteine) proteases (Musil *et al.*, 1991).

A multiple sequence alignment of the deduced protein sequence of AgCatB gene with other cathepsin B sequences is shown in Fig. 2. Alignment of the AgCatB sequence with those for cathepsin B from several other species indicates the extent of the identity that exists. Three active sites characteristic for cathepsin B (Cho *et al.*, 1999; Xu and Kawasaki, 2001) were conserved in the deduced amino acid sequence of AgCatB at positions Cys-109, His-267, and Asn-287. The three active sites characteristic for cathepsin B (Cho *et al.*, 1999; Xu and Kawasaki, 2001) were conserved at the same position among the species aligned (Cho *et al.*, 1999; Xu and Kawasaki, 2001; Takahashi *et al.*, 1993; Zhao *et al.*, 1998; Kwon *et al.*, 2001; Chan *et al.*, 1986; Bechet *et al.*, 1991; Guenette *et al.*, 1994; Merckelbach *et al.*, 1994).

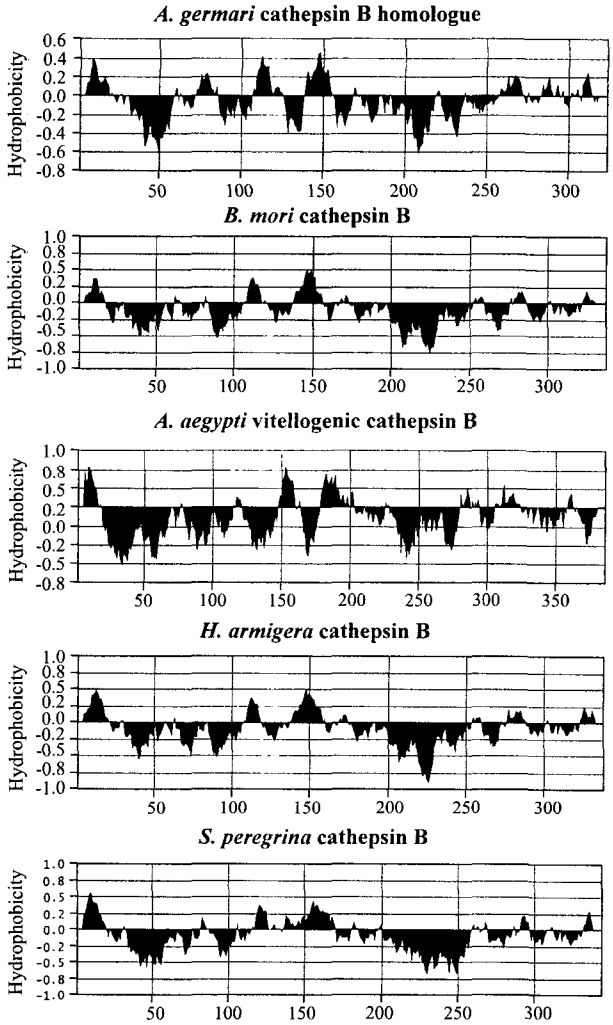

Fig. 3. The hydropathy profile of *A. germari* cathepsin B homologue and known insect cathepsins B. Hydropathic analysis was done as described by Kyte and Doolittle (1982). The known insect cathepsins B were: *B. mori* (Xu and Kawasaki, 2001), *A. aegypti* (Cho *et al.*, 1999), *H. armigera* (Zhao *et al.*, 1998) and *S. peregrina* (Takahashi *et al.*, 1993).

Table 1. Pairwise comparison among amino acid sequences of the *A. germari* cathepsin B gene and the known cathepsin B genes

Species	GenBank No.	1	2	3	4	5
1. *A. germari*	AF483623	-	0.527	0.534	0.544	0.668
2. *B. mori*	AB045595	208	-	0.159	0.397	0.618
3. *H. armigera*	AF222788	211	63	-	0.400	0.618
4. *S. peregrina*	D16823	215	157	158	-	0.628
5. *A. aegypti*	AF127592	264	244	244	248	-

Numbers above the diagonal are mean distance values; numbers below the diagonal are absolute distance values.

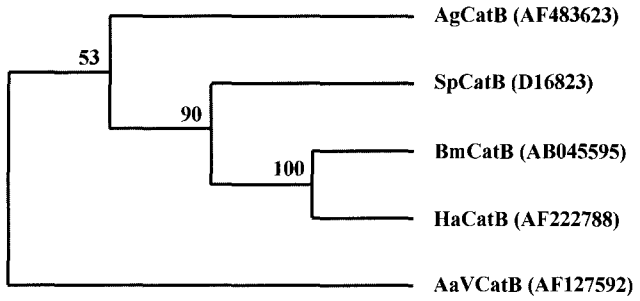

Fig. 4. Phylogenetic tree for aligned amino acid sequences of the *A. germari* cathepsin B homologue protein and the known cathepsin D proteins. The sequences were extracted from; *A. germari* (this study), *B. mori* (Xu and Kawasaki, 2001), *A. aegypti* (Cho *et al.*, 1999), *H. armigera* (Zhao *et al.*, 1998) and *S. peregrina* (Takahashi *et al.*, 1993). The tree was obtained by bootstrap analysis with the option of heuristic search and the numbers on the branches represent bootstrap values for 1,000 replicates.

The hydropathy plot of AgCatB was analyzed by the Kyte and Doolittle method (Kyte and Doolittle, 1982), and compared with other insect cathepsins B showing a high homology. As shown in Fig. 2, hydropathic analysis suggests the profile of AgCatB is similar to known insect cathepsins B such as *B. mori* (Xu and Kawasaki, 2001), *A. aegypti* (Cho *et al.*, 1999), *H. armigera* (Zhao *et al.*, 1998) and *S. peregrina* (Takahashi *et al.*, 1993). Hydropathy analysis revealed that the AgCatB protein was hydrophilic, possessing a strongly hydrophobic putative signal peptide of 18 residues.

The deduced amino acid sequence of the AgCatB had a homology with that of cathepsins B from other insect species (Table 1). This alignment illustrates that AgCatB is 47.3% identical to the *B. mori*, 46.6% to the *H. armigera*, and 45.6% to the *S. peregrina*, but lowest with *A. aegypti* (33.2%).

A phylogenetic tree was constructed using the protein sequences of cathepsins B (Fig. 4). The phylogenic analysis revealed that AgCatB is closer to cathepsins B of the *S. peregrina*, *B. mori* and *H. armigera* than *A. aegypti* cathepsin B.

In conclusion, we showed gene nucleotide sequence of the cDNA encoding a cathepsin B homologue from the mulberry longicorn beetle, *A. germari* for the first time.

References

Bechet, D. M., M. J. Ferrara, S. B. Mordier, M. P. Roux, C. D. Deval and A. Obled (1991) Expression of lysosomal cathepsin B during calf myoblast-myotube differentiation. Characterization of a cDNA encoding bovine cathepsin B. *J. Biol. Chem.* **266**, 14104-14112.

Chan, S. J., B. San Segundo, M. B. McCormick and D. F. Steiner (1986) Nucleotide and predicted amino acid sequences of cloned human and mouse preprocathepsin B cDNAs. *Proc. Natl. Acad. Sci. USA* **83**, 7721-7725.

Cho, W. L., S. M. Tsao, A. R. Hays, R. Walter, J. S. Chen, E. S. Snigirevskaya and A. S. Raikhel (1999) Mosquito cathepsin B-like protease involved in embryonic degradation of vitellin is produced as a latent extraovarian precursor. *J. Biol. Chem.* **274**, 13311-13321.

Crowson, R. A. (1981) The biology of the Coleoptera. Academic Press. London.

Daly, H. V., J. T. Doyen and A. H. Purcell III (1998) Introduction to insect biology and diversity. 2nd edition, Oxford University Press, London.

Guenette, R. S., M. Mooibroek, K. Wong, P. Wong and M. Tenniswood (1994) Cathepsin B, a cysteine protease implicated in metastatic progression, is also expressed during regression of the rat prostate and mammary glands. *Eur. J. Biochem.* **226**, 311-321.

Kim, S. R., H. J. Yoon, N. S. Park, S. M. Lee, J. Y. Moon, B. R. Jin and H. D. Sohn (2001) Molecular cloning of a cDNA encoding a cathepsin D homologue from the mulberry longicorn beetle, Apriona germari. *Int. J. Indust. Entomol.* **3**, 121-126.

Kwon, J. Y., F. Prat, C. Randall and C. R. Tyler (2001) Molecular characterization of putative yolk processing enzymes and their expression during oogenesis and embryogenesis in rainbow trout (Oncorhynchus mykiss). *Biol. Reprod.* **65**, 1701-1709.

Kyte, J. and R. F. Doolittle (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105-132.

Merckelbach, A., S. Hasse, R. Dell, A. Eschlbeck and A. Ruppel (1994) cDNA sequences of Schistosoma japonicum coding for two cathepsin B-like proteins and Sj32. Trop. *Med. Parasitol.* **45**, 193-198.

Musil, D., D. Zucic, D. Turk, R. A. Engh, I. Mayr, R. Huber, T. Popovic, V. Turk, T. Towatari, N. Katunuma and W. Bode (1991) The refined 2.15 A X-ray crystal structure of human liver cathepsin B: the structural basis for its specificity.

EMBO J. **10**, 2321-2330.

Swofford, D. L. (1990) PAUP: phylogenetic analysis using parsimony, ver. 3.0. Illinois Natural History Survey, Champaign (on disk).

Takahashi, N., S. Kurata and S. Natori (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.

Xu, Y. and H. Kawasaki (2001) Isolation and expression of cathepsin B cDNA in hemocytes during metamorphosis of *Bombyx mori*. *Comp. Biochem. Physiol. B.* **130**, 393-399.

Yoon, H. J., J. S. Bae, I. Kim, B. R. Jin, Y. I. Mah, J. Y. Moon and H. D. Sohn (2001) A phylogenetic study in some long-horned beetle (Coleoptera: Cerambycidae) using mitochondrial COI gene and 16S rRNA sequences. *Int. J. Indust. Entomol.* **2**, 37-53.

Yoon, H. J. and Y. I. Mah (1999) Life cycle of the mulberry longicorn beetle, *Apriona germari* Hope on an artificial diet. *J. Asia-Pacific Entomol.* **2**, 169-173.

Yoon, H. J., Y. I. Mah, I. G. Park, S. B. Lee and S. Y. Yang (1997) The mode of hibernation of mulberry longicorn beetle, Apriona germari Hope in Korea. *J. Seric. Sci. Jpn.* **66**, 128-131.

Zhao, X. F., J. X. Wang and Y. C. Wang (1998) Purification and characterization of a cysteine proteinase from eggs of the cotton boll worm, *Helicoverpa armigera*. *J. Insect Biochem. Mol. Biol.* **28**, 259-264.