Int. J. Indust. Entomol. Vol. 11, No. 2, 2005, pp. 113~117

International Journal of *Industrial Entomology*

A Chymotrypsin Gene Homologue from the Mulberry Longicorn Beetle, *Apriona germari*: cDNA Sequence Characterization and mRNA Expression Pattern

Zong Zheng Gui†, Kwang Sik Lee, Hyung Joo Yoon[1], Iksoo Kim[1], Hung Dae Sohn and Byung Rae Jin*

College of Natural Resources and Life Science, Dong-A University, Busan 604-714, Korea.

[1]Department of Agricultural Biology, National Institute of Agricultural Science and Technology, Suwon 441-100, Korea.

(Received 5 October 2005; Accepted 6 November 2005)

A chymotrpsin gene homologue was cloned from the mulberry longicorn beetle, *Apriona germari*. The *A. germari* chymotrypsin cDNA contains an ORF of 950 nucleotides capable of encoding a 283 amino acid polypeptide with a predicted molecular mass of 29151 Da and pI of 9.38. The *A. germari* chymotrypsin has conserved six cysteine residues and active triad formed by His, Asp and Ser. The deduced amino acid sequence of the *A. germari* chymotrypsin cDNA was closest in structure to the *Anthonomus grandis* chymotrypsin. Northern blot analysis revealed that *A. germari* chymotrypsin showed the midgut-specific expression.

Key words: *Apriona germari*, cDNA cloning, Chymotrypsin, Insect, Mulberry longicorn beetle

Introduction

Serine proteinases are a superfamily and include many different enzymes, e.g. trypsin, chymotrypsin, subtilisin and serine carboxypeptidase (Rawlings and Barrett, 1994). A major characteristic of these serine proteinses is the active triad formed by His, Asp and Ser (Rawlings and Barrett, 1994). Trypsin and chymotrypsin are among the well-characterized serine proteinases in the insect's digestive system. In vertebrates, these enzymes have only three bridges formed by six cysteine residues (Davis *et al.*, 1985). The surface of the substrate-binding pocket for trypsin is conformed by three amino acid residues, Asp, Gly and Gly, and the trypsins preferentially hydrolyze peptide bonds following a lysine or an arginine. On the other hand, the surface of the substrate-binding pocket for chymotrypsin is mainly Ser, Gly and Gly, and chumotrypsins cleave after aromatic amino acid residues (Kraut, 1977; Hedstrom, 2002).

Trypsins and chymotrypsins are involved in protein digestion and are the most abundant digestive proteases in several insect species (Noriega *et al.*, 1996; Gaines *et al.*, 1999; Zhu and Baker, 2000; Yan *et al.*, 2001; Ramalho-Ortigao *et al.*, 2003). However, the molecular biology of coleopteran insect digestive proteases has received somewhat limited attention, apart from the lepidopteran and dipteran species. cDNA clones encoding chymotrypsin-like serine proteases have been isolated from a coleopteran insect species, *Anthonomus grandis* (cotton boll weevil) (Oliveira-Neto *et al.*, 2004). The present paper describes the cDNA cloning and mRNA expression of a chymotrypsin gene homologue from the mulberry longicorn beetle, *Apriona germari* (Coleoptera: Cerambycidae), that feed on mulberry tree, tunneling inside the stem and ingesting the living wood.

Materials and Methods

Insects

The mulberry longicorn beetle, *Apriona germari* (Coleoptera: Cerambycidae), was reared on an artificial diet as described previously (Yoon and Mah, 1999).

cDNA library screening, nucleotide sequencing and data analysis

A cDNA library used in this study was constructed using

†Present address. Sericultural Research Institute, Chinese Academy of Agricultural Sciences, Zhenjiang 212018, China.

*To whom correspondence should be addressed.

College of Natural Resources and Life Science, Dong-A University, Busan 604-714, Korea. Tel: +82-51-200-7594; Fax: +82-51-200-7594; E-mail: brjin@dau.ac.kr

whole bodies of *A. germari* larvae (Kim *et al.*, 2001). The clones harboring cDNA inserts were randomly selected and sequenced to generate the expressed sequence tags (ESTs) (Kim *et al.*, 2003). The plasmid DNA from the clones harboring cDNA inserts was extracted by Wizard mini-preparation kit (Promega, Madison, WI). The nucleotide sequence was determined by using a BigDyeTerminator cycle sequencing kit and an automated DNA 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 (http://www.ncbi.nlm.nih.gov/BLAST). GenBank, EMBL and SwissProt databases were searched for sequence homology using a BLAST algorithm program (Altschul *et al.*, 1997). MacVector (ver. 6.5, Oxford Molecular Ltd.) was used to align the amino acid sequences of chymotrypsin. With the eleven GenBank-registered chymotrypsin amino acid sequences, phylogenetic analysis was performed using PAUP (Phylogenetic Analysis Using Parsimony) version 4.0 (Swofford, 2000). The accession numbers of the sequences in the GenBank are as follows: *Apriona germari* (AY771359; this study), *Anthonomus grandis* (AY536261), *Agrotis ipsilon* (AF233729), *Spodoptera frugiperda* (AY251276), *Helicoverpa armigera* (Y12279), *H. zea* (AF233732), *Manduca sexta* (L34168), *Bombyx mori* (AB003670), *Aedes aegypti* (AF187078), *A. polynesiensi* (AY155563), *Ochlerotatus epactius* (AY155562), and *Pediculus humanus* (AY819730).

RNA isolation and Northern blot analysis

Three *A. germari* larvae were dissected on ice under a Stereo-microscope (Zeiss, Jena, Germany), samples of the fat body, midgut and epidermis were collected, and washed twice with PBS (140 mM NaCl, 27 mM KCl, 8 mM Na_2HPO_4, 1.5 mM KH_2PO_4, pH 7.4). Total RNA was isolated from the fat body, midgut and epidermis of *A. germari* larvae using the Total RNA Extraction Kit (Promega). Total RNA (10 μg/lane) from *A. germari* larvae was denatured by glyoxalation (McMaster and Carmichael, 1977), transferred onto a nylon blotting membrane (Schleicher & Schuell, Dassel, Germany) and hybridized at 42°C with a probe in a hybridization buffer containing 5 × SSC, 5 × Denhardt's solution, 0.5% SDS, and 100-μg/ml denatured salmon sperm DNA. The 950 bp *A. germari* chymotrypsin cDNA clone was labeled with [α-^{32}P]dCTP (Amersham, Arlington Heights, IL), using the Prime-It II Random Primer Labeling Kit (Stratagene, La Jolla, CA), and used as a probe for hybridization. After hybridization, the membrane filter was washed three times for 30 min each in 0.1% SDS and 0.2 × SSC (1 × SSC is 0.15 M NaCl and 0.015 M sodium citrate) at 65°C and then exposed to autoradiography film.

Results and Discussion

Cloning, sequencing and analysis of *A. germari* chymotrypsin cDNA

In search of *A. germari* ESTs, a cDNA was identified that had a high homology with previously reported chymotrypsin genes. The cDNA clone, including the full-length open reading frame (ORF), was sequenced and characterized. The nucleotide and deduced amino acid sequences of the cDNA encoding chymotrypsin homologue are presented in Fig. 1. The *A. germari* chymotrypsin cDNA is 950 bp long and contains an ORF of 849 nucleotides capable of encoding a 283 amino acid polypeptide with a predicted molecular mass of 29151 Da and pI of 9.38. The ORF had both a start (ATG) and stop codon (TAA), indicating that the sequences contain the complete coding region. A putative polyadenylation signal, AATAAA, is located at nucleotide position 899-904. This *A. germari* chymotrypsin cDNA sequence has been deposited in GenBank under accession number

```
-13                                                    GGTATCATCGAAG
  1 ATGAAGTTCGCGGTGTGTTTCTTGGCAGTCATCTCCTGCACTGTGGCTCTGCCCGCAGAA
  1 M  K  F  A  V  C  F  L  A  V  I  S  C  T  V  A  L  P  A  E
 61 GTCCCTTTCGAGACGCCTGAGGAGTTTGACCTGTCGAAAATAACCCCAAACAACATCTAT
 21 V  P  F  E  T  P  E  E  F  D  L  S  K  I  T  P  N  N  I  Y
121 CTTGAACCCATTAATCCCGTAAATGAAGCTGATCTTCCCGAGGGCAGAATTATCGGTGGT
 41 L  E  P  I  N  P  V  N  E  A  D  L  P  E  G  R  I  I  G  G
181 AGCGAAGTAAGTCGTAATTCCGTCCCATACCAGGCTGCCCTCCTCATCAATGGTGGAAGC
 61 S  E  V  S  R  N  S  V  P  Y  Q  A  A  L  L  I  N  G  G  S
241 TTGTGCGGTGGTTCCCTTATATCAACCACTCGTGTTCTGACAGCTGCTCATTGTACCGTC
 81 L  C  G  G  S  L  I  S  T  T  R  V  L  T  A  A  H  C  T  V
301 AGTGCCTCCAATGTCCAAGTCAGACTTGGTGCCCACAACTTCCACCAGAACGAAGCCACC
101 S  A  S  N  V  Q  V  R  L  G  A  H  N  F  H  Q  N  E  A  T
361 CAGGTCCGCATCACCAGTACCAGCATCAGGAACCACGCTTCTTACAACGTTGGCAGCTTC
121 Q  V  R  I  T  S  T  S  I  R  N  H  A  S  Y  N  V  G  S  F
421 GCCGCCAACGACATCGCCGTCATCCTCTTGCCTTCTGCTGTGTCCACTAACAACAACATC
141 A  A  N  D  I  A  V  I  L  L  P  S  A  V  S  T  N  N  N  I
481 AGGACTATTTCTTTGGCCCCAGCTAACGCCGGCAACTTTGCCGGCTCCAGGGCCTTCCTA
161 R  T  I  S  L  A  P  A  N  A  G  N  F  A  G  S  R  A  F  L
541 AGTGGATGGGGAAGAACTTCCGATGGGTCCAACGCCGTATCTGCTGTTCTCCGTGGCGTG
181 S  G  W  G  R  T  S  D  G  S  N  A  V  S  A  V  L  R  G  V
601 AACCTCAACATCATCACCAACGCCGTCTGCAGAAACGCGTTTGGCAACGCGATTTTGGCT
201 N  L  N  I  I  T  N  A  V  C  R  N  A  F  G  N  A  I  L  A
661 TCCACAATATGTACCTCTGGAGCCGGTACTGTCGGTGCTTGTAACGGCGACTCCGGTGGT
221 S  T  I  C  T  S  G  A  G  T  V  G  A  C  N  G  D  S  G  G
721 CCATTGGTGGTTAATGGAGTCCAAGTCGGTGTCGTGTCTTTCGGCGCTGCCAGGTGCCAA
241 P  L  V  V  N  G  V  Q  V  G  V  V  S  F  G  A  A  R  C  Q
781 AGCGGAAATCCAACTGCCTTCGCCAGGATCAGTTCCTTCAGAAACTGGATTACTCAGAAC
261 S  G  N  P  T  A  F  A  R  I  S  S  F  R  N  W  I  T  Q  N
841 GCCGGTGTCTAAATTTTCGCGAAATGGTCCTTGTTACGAAAACTGTACAGCAATATCAAA
281 A  G  V  *
901 TAAAGAGCATTAAAAATAAAAAAAAAAAAAAAAAAA
```

Fig. 1. The nucleotide and deduced protein sequence of the *A. germari* chymotrypsin. The start codon is boxed and the termination codon is indicated by an asterisk. In the cDNA sequence, the polyadenylation sequence is underlined. The GenBank accession number is AY771359.

AY771359.

Comparison of the deduced amino acid sequence of the *A. germari* chymotrypsin with that of other chymotrypsin sequences is shown in Fig. 2. The deduced pre-proenzyme and the mature chymotrypsin contain 283 and 227 amino acid residues. The positions of amino acid residues forming the active triad (His, Asp and Ser) and the substrate-binding pocket and those of the six cysteine residues known to form three disulphide bridges (Kraut, 1977; Davis *et al.*, 1985; Rawlings and Barrett, 1994; Hedstrom, 2002; Waniek *et al.*, 2005) are conserved in the sequence of *A. germari* chymotrypsin. On the basis of these characteristics, it is proposed that *A. germari* chymotrypsin is one of the expanding family of chymotrysins.

Based on sequence and structure similarity, *A. germari* chymotrypsin was closest to that of *Anthonomus grandis*. In the phylogram of several chymotrypsins, *A. germari* chymotrypsin clusters together with that of *A. grandis* (Fig. 3A). When compared with homologues in the GenBank database, *A. germari* chymotrypsin has an identity between 37% and 42% with other insect chymotrypsins (Fig. 3B).

Expression of *A. germari* chymotrypsin mRNA

To characterize the expression of the *A. germari* chymotrypsin gene at the transcriptional level, Northern blot analysis was performed using total RNA obtained from fat body, midgut and epidermis, respectively. Northern blot analysis showed that a hybridization signal was detected as a single band of mRNA from the midgut, indicating that the *A. germari* chymotrypsin is expressed only in the midgut (Fig. 4). The result is consistent with the previous reports that the chymotrypsin gene is tissue-specifically expressed in midgut (Mazumdar-Leighton and Broadway, 2001; Oliveira-Neto *et al.*, 2004; Waniek *et al.*, 2005)

In conclusion, this study was undertaken to search for a homologue of chymotrypsin in *A. germari* and to elucidate its sequence and structure of chymotrypsin in beetles. The data presented suggest that *A. germari* chymotrypsin is a member of chymotrypsin family, which is expressed only in the midgut and functions as a digestive protease. We hope that the molecular characterization of chymotrypsin in a beetle in this study will expand the understanding of insect chymotrypsins.

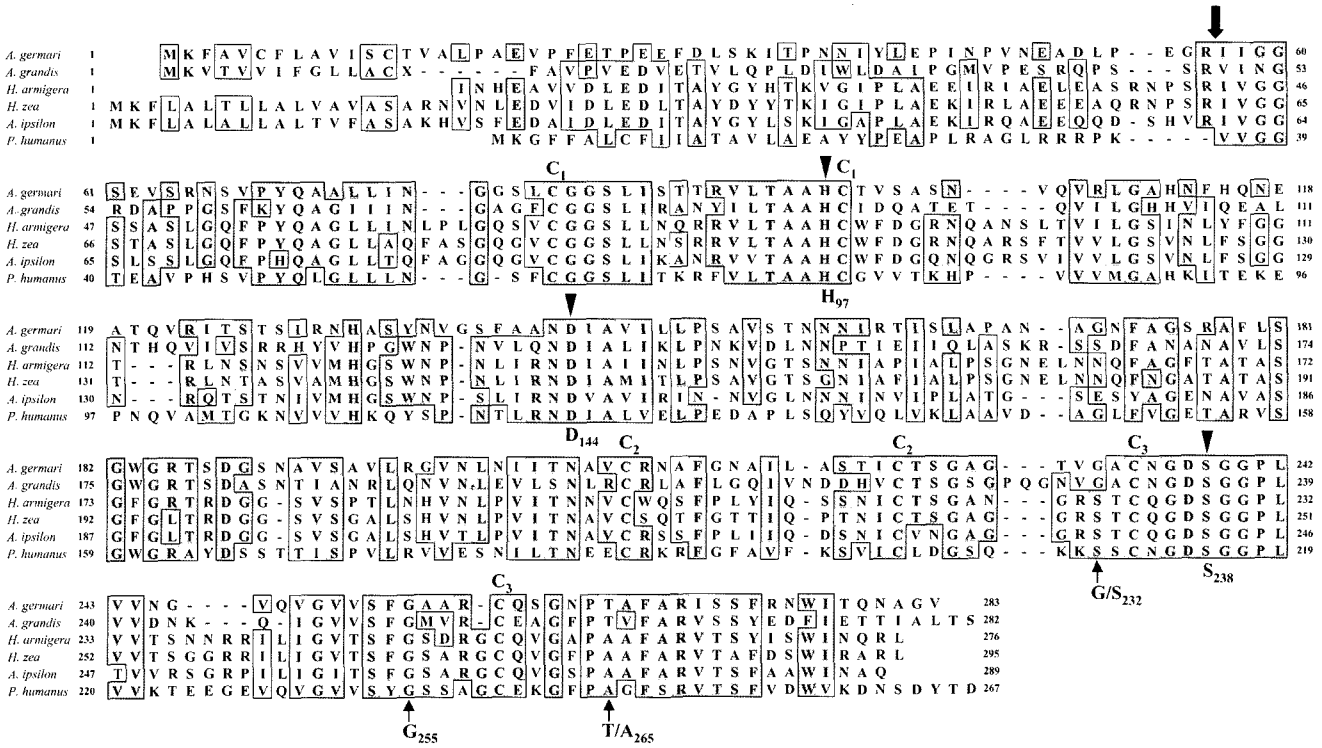
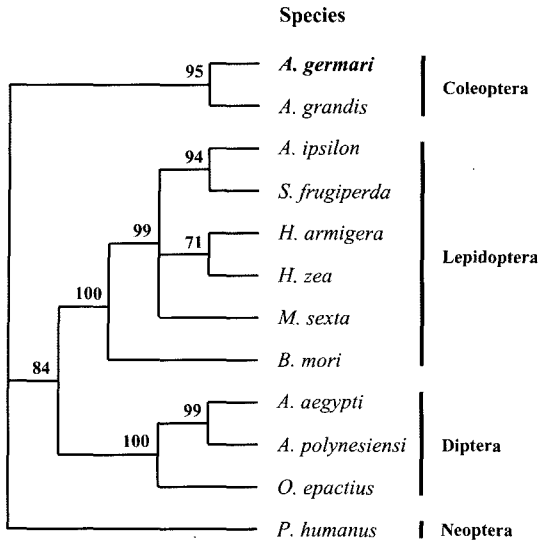


Fig. 2. Comparison of the deduced amino acid sequence of *A. germari* chymotrypsin with that of other insect chymotrpsins. Invariant residues are shaded black. Vertical arrow indicates the end of the predicted activation peptide. The conserved catalytic site residues (H, D and S) are indicated by arrow heads above the alignment. Amino acid residues involved in determining enzyme substrate specificity and conserved cysteine residues are indicated by arrows and C_n, respectively, where *n* matches pairwise the residues predicting to be involved in disulphide bridge formation. GenBank accession numbers are: *A. grandis* (AY536261), *H. armigera* (Y12279), *H. zea* (AF233732), *A. ipsilon* (AF233729) and *P. humanus* (AY819730).

(A)



(B)

Percent similarity

Species	GenBank No.	1	2	3	4	5	6
1. ***A. germari***	AY771359		54	56	56	53	51
2. *A. grandis*	AY536261	38		49	50	51	48
3. *H. armigera*	Y12279	42	34		80	71	50
4. *H. zea*	AF233732	39	31	73		77	45
5. *A. ipsilon*	AF233729	39	32	60	67		46
6. *P. humanus*	AY819730	37	32	34	30	30	

Percent identity

Fig. 3. Relationships among amino acid sequences of the *A. germari* chymotrypsin and the known chymotrypsins. (A) A maximum parsimony analysis for the amino acid sequences of *A. germari* chymotrypsin and known chymotrypsins derived from insects. The sequence sources are described in Materials and methods. 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 and (B) Pairwise similarities and identities of the deduced amino acid sequence of *A. germari* chymotrypsin among chymotrypsin sequences.

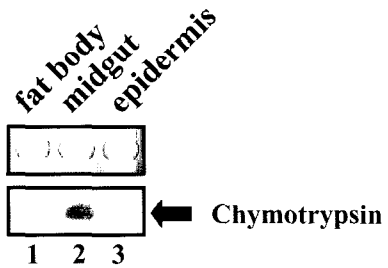


Fig. 4. *A. germari* chymotrypsin mRNA expression in *A. germari* tissues. Total RNA was isolated from the fat body (lane 1), midgut (lane 2) and epidermis (lane 3), respectively (upper panel). The RNA was separated by 1.0% formaldehyde agarose gel electrophoresis, transferred on to a nylon membrane and hybridized with radiolabelled 950 bp *A. germari* tissues cDNA (lower panel). Transcripts are indicated on the right side of the panel by an arrow.

Acknowledgements

This work was funded by the Dong-A University Research Fund.

References

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.* **25**, 3389-3402.

Davis, C. A., D. C. Riddell, M. J. Higgins, J. J. A. Holden and B. N. White (1985) A gene family in *Drosophila melanogaster* coding for trypsin-like enzymes. *Nucl. Acids Res.* **13**, 6605-6619.

Hedstrom, L. (2002) Serine protease mechanism and specificity. *Chem. Rev.* **102**, 4501-4523.

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.

Kim, S. R., H. J. Yoon, K. S. Lee, I. Kim, Y. H. Je, H. D. Sohn and B. R. Jin (2003) Molecular cloning of three cDNAs encoding putative larval cuticle protein expressed differentially after larval ecdysis from the mulberry longicorn beetle, *Apriona germari*. *Comp. Biochem. Physiol. B* **136**, 803-811.

Kraut, J. (1977) Serine proteases: structure and mechanism of catalysis. *Annu. Rev. Biochem.* **46**, 331-358.

Mazumdar-Leighton, S. and R. M. Broadway (2001) Identification of six chymotrypsin cDNAs from larval midguts of *Helicoverpa* zea and *Agrotis ipsilon* feeding on the soybean (Kunitz) trypsin inhibitor. *Insect Biochem. Mol. Biol.* **31**, 633-644.

McMaster, G. K. and G. G. Carmichael (1977) Analysis of single- and double-stranded nucleic acids on polyacrylamide and agarose gels by using glyoxal and acridine orange. *Proc. Natl. Acad. Sci. USA* **74**, 4835-4838.

Noriega, F. C., X. Y. Wang, J. E. Pennington, C. V. Barillas-Mury and M. A. Wells (1996) Early trypsin, a female-specific midgut protease in *Aedes aegypti*: isolation, aminoterminal sequence determination, and cloning and sequencing of the gene. *Insect Biochem. Mol. Biol.* **26**, 119-126.

Oliveira-Neto, O. B., J. A. N. Batista, D. J. Rigden, R. R. Fragoso, R. O. Silva, E. A. Gomes, O. L. Franco, S. C. Dias, C. M. T. Cordeiro, R. G. Monnerat and M. F. Grossi-de-Sa (2004) A diverse family of serine proteinase expressed in cotton boll weevil (*Anthonomus grandis*): implications for the design of pest-resistant transgenic cotton plants. *Insect Biochem. Mol. Biol.* **34**, 903-918.

Ramalho-Ortigao, J. M., S. Kamhawi, E. D. Rowton, J. M. C. Ribeiro and J. G. Valenzuela (2003) Cloning and characterization of trypsin- and chymotrypsin-like proteases from the midgut of the sand fly vector *Phlebotomus papatasi*. *Insect*

Biochem. Mol. Biol. **33**, 163-171.

Waniek, P. J., U. B. hendgen-Cotta, P. Stock, C. Mayer, A. H. Kollien and G. A. Schaub (2005) Serine proteinase of the human body louse (*Pediculus humanus*): sequence characterization and expression patterns. *Parasitol. Res.* **97**, 486-500.

Yan, J. Q. Cheng, C. B. Li and S. Aksoy (2001) Molecular characterization of two serine proteases expressed in gut tissue of the African trypanosome vector, *Glossina morsitans morsitans*. *Insect Mol. Biol.* **10**, 47-56.

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.

Zhu, Y. C., F. Zeng and B. Oppert (2003) Molecular cloning of trypsin-like cDNAs and comparison of proteinase activities in the salivary glands and gut of the tarnished plant bug *Lygus lineolaris* (Heteroptera: Miridae). *Insect Biochem. Mol. Biol.* **33**, 889-899.