



Molecular identification, characterization, and expression analysis of a serine protease inhibitor gene from cotton bollworm, *Helicoverpa armigera* (Lepidoptera: Noctuidae)

Muhammad Shakeel^{a*} 

^aSouth China Agricultural University, College of Agriculture, Key Laboratory of Bio-Pesticide Innovation and Application of Guangdong Province, Guangzhou, China

*e-mail: faizaneabiwaqas@scau.edu.cn

Received: May 6, 2019 – Accepted: February 7, 2020 – Distributed: August 31, 2021

(With 5 figures)

Abstract

Serine protease inhibitors (serpins), a superfamily of protease inhibitors, are known to be involved in several physiological processes, such as development, metamorphosis, and innate immunity. In our study, a full-length serpin cDNA, designated Haserpin1, was isolated from the cotton bollworm *Helicoverpa armigera*. The cDNA sequence of Haserpin1 is 1176 nt long, with an open reading frame encoding 391 amino acids; there is one exon and no intron. The predicted molecular weight of Haserpin1 is 43.53 kDa, with an isoelectric point of 4.98. InterProScan was employed for Haserpin1 functional characterization, which revealed that Haserpin1 contains highly conserved signature motifs, including a reactive center loop (RCL) with a hinge region (E341–N350), the serpin signature, (F367–F375) and a predicted P1–P1' cleavage site (L357–S358), which are useful for identifying serpins. Transcripts of Haserpin1 were constitutively expressed in the fat body, suggesting that it is the major site for serpin synthesis. During the developmental stages, a fluctuation in the expression level of Haserpin1 was observed, with low expression detected at the 5th-instar larval stage. In contrast, relatively high expression was detected at the prepupal stage, suggesting that Haserpin1 might play a critical role at the *H. armigera* wandering stage. Although the detailed function of this serpin (Haserpin1) needs to be elucidated, our study provides a perspective for the functional investigation of serine protease inhibitor genes.

Keywords: cloning, mRNA expression, reactive center loop, serpin, fat body.

Identificação molecular, caracterização e análise de expressão de um gene inibidor de serina protease da lagarta *Helicoverpa armigera* (Lepidoptera: noctuidae) na cultura de algodão

Resumo

Sabe-se que os inibidores de serina protease (serpinas), uma superfamília de inibidores de protease, estão envolvidos em vários processos fisiológicos, como desenvolvimento, metamorfose e imunidade inata. Neste estudo, um cDNA de serpina de comprimento total, denominado Haserpin1, foi isolado da lagarta *Helicoverpa armigera* na cultura de algodão. A sequência de ADNc de Haserpin1 tem 1.176 nt de comprimento, com uma grelha de leitura aberta que codifica 391 aminoácidos; existe um éxon, mas nenhum íntron. O peso molecular previsto de Haserpin1 é de 43,53 kDa, com um ponto isoelétrico de 4,98. O InterProScan foi empregado para a caracterização funcional do Haserpin1, que revelou que o Haserpin1 contém motivos de assinatura altamente conservados, incluindo um loop central reativo (RCL) com uma região de dobradiça (E341–N350), a assinatura da serpina (F367–F375) e um local de clivagem previsto de P1–P1' (L357–S358), que são úteis para identificar serpinas. As transcrições de Haserpin1 foram expressas constitutivamente no corpo gordo, sugerindo que é o principal local para a síntese de serpinas. Durante os estágios de desenvolvimento, observou-se uma flutuação no nível de expressão de Haserpin1, com baixa expressão detectada no estágio larval do 5^o instar. Por outro lado, detectou-se uma expressão relativamente alta no estágio pré-pupal, sugerindo que o Haserpin1 pode desempenhar um papel crítico no estágio errante de *H. armigera*. Embora a função detalhada dessa serpina (Haserpin1) precise ser elucidada, este estudo fornece uma perspectiva para a investigação funcional dos genes inibidores da serina protease.

Palavras-chave: clonagem, expressão de mRNA, loop central reativo, serpina, corpo gordo.

1. Introduction

Serine proteases, ubiquitous enzymes found in eukaryotes, bacteria, and viruses, constitute almost one-third of all proteases and play a pivotal role in the catalysis of intracellular and extracellular hydrolytic reactions (Yang et al., 2017). Serine proteases are known to be involved in a wide range of essential biological processes (Ross et al., 2003; Zou et al., 2006; Zhao et al., 2010). In addition to their vital role in physiological processes, the activity of these proteases, if not controlled properly, might be hazardous to living organisms (Neurath, 1989). Injury due to excessive protease activity includes tissue damage, melanization, and inappropriate coagulation, among others (Gubb et al., 2010; Eleftherianos and Revenis, 2011). Thus, their activity must be properly and strictly controlled by inhibitors (Krowarsch et al., 2003; Rawlings et al., 2004).

Serpins, a super-family of proteins, are the most diverse family of serine protease inhibitors found in animals, plants, and microorganisms (Roberts et al., 2004; Reichhart, 2005). Serpins are composed of approximately 350-500 amino acids, with a reactive center loop (RCL) of approximately 20 residues near the carboxyl terminus (Shakeel et al., 2019). The serpin RCL region, a determining region of serpins that mediates interaction with the target protease, is comprised of a scissile bond between residues designated P1 and P1' (Reichhart et al., 2011). After successful cleavage of the scissile bonds, serpin dramatically undergoes a conformational change that covalently traps the target proteinase (Dissanayake et al., 2006; Ulvila et al., 2011; Yang et al., 2017).

Serpins have been documented to play a critical role in various physiological processes, including reproduction, metamorphosis, inflammation, blood coagulation, complement, and innate immune responses (Bayer et al., 1997; Choo et al., 2012; Davie et al., 1991; Gál et al., 2013; Kim et al., 2013; Shigetomi et al., 2010; Shakeel and Zafar, 2020). All of these processes require the existence of serpins to control cellular homeostasis, inhibit inadmissible proteolytic cascades (Jarasrassamee et al., 2005), or prevent exogenous proteases secreted by pathogens that exploit the protease to invade new tissues and penetrate a host (Wang et al., 2009).

To date, a number of serpins have been identified in several insect species, including *Drosophila melanogaster* (De Gregorio et al., 2002), *Manduca sexta* (Zou and Jiang, 2005), *Bombyx mori* (Zou et al., 2009), *Plutella xylostella* (Lin et al., 2017), *Anopheles gambiae* (Waterhouse et al., 2007), and *Ostrinia furnacalis* (Liu et al., 2014). Previously, it was reported that the Toll pathway and prophenoloxidase pathway activation cascade in *D. melanogaster* and *M. sexta*, respectively, are negatively regulated by serpins (An and Kanost, 2010; Reichhart et al., 2011). In *Aedes aegypti*, Aeserpin1, along with immune melanization proteases (IMP-1 and IMP-2), mediates hemolymph prophenoloxidase cleavage and the immune response against the malarial parasite (Zou et al.,

2010). In addition, serpin-1 is reported to participate in the regulation of cuticle degradation during the molting cycle in *Choristoneura fumiferana* (Zheng et al., 2009). In *Tenebrio molitor*, melanin synthesis is negatively controlled when SPN40, SPN48, and SPN55 inhibit MSP, SPE, and SAE proteases, respectively (Jiang et al., 2009). This type of interaction has also been described in *A. gambiae*, whereby prophenoloxidase-activating protease CLIPB9 is inhibited by SRPN2 to regulate prophenoloxidase activation (An et al., 2011). Similarly, serpin-3 of *O. furnacalis* efficiently inhibits SP105, which acts as a prophenoloxidase activator (Chu et al., 2017).

The cotton bollworm *Helicoverpa armigera* (Hübner) is an important polyphagous generalist pest species in Asia, Europe, Africa, and Australia that causes serious damage to cotton, corn, sorghum and many other crops (Tay et al., 2013). Over the past decades, primary control strategies for this pest have relied on pesticides and transgenic crops (Tabashnik et al., 2013). However, these methods not only lead to the development of resistance to insecticides but also affect the environment. Thus, there is an urgent need to develop novel biological control strategies.

Considering the importance of serpins in physiological processes, as exhibited by previous reports, the present study aimed to identify and characterize a serpin gene that might also be involved in biological processes in *H. armigera*. Furthermore, we sought to elucidate the phylogenetic relationship of *H. armigera* serpin to serpins from other insects.

2. Material and Methods

2.1. Insect culture

The *H. armigera* population was nurtured on an artificial diet (Abbasi et al., 2007). The rearing conditions were as follows: 70% relative humidity, photoperiod of 16:8-h L:D, and 25±2 °C temperature.

2.2. Collection of tissues

Dissection of healthy *H. armigera* larvae was performed on ice under a stereomicroscope (Zeiss, Jena, Germany). The insect fat body was collected and washed with PBS (140 mM NaCl, 27 mM KCl, 8 mM Na₂HPO₄, and 1.5 mM KH₂PO₄, pH 7.4). The fat body tissues were then stored at -80 °C for RNA extraction.

2.3. RNA isolation and cDNA synthesis

Total RNA was isolated using TRIzol reagent following the manufacturer's protocol. Assessment of the RNA quality was performed by 1% (w/v) agarose gel electrophoresis. A UV spectrophotometer was used to assess RNA concentration. DNA contamination was eliminated by DNaseI (Fermentas). cDNA was synthesized by reverse transcription RevertAid™ Reverse Transcriptase (Fermentas) in a 20-μL reaction. The reaction mix was incubated at 42 °C for 60 min, and the product was stored at -80 °C.

2.4. Identification and primer design of *H. armigera* serpin for PCR

The *H. armigera* serpin (accession number: HQ615869) sequence was retrieved from the GenBank database at NCBI (National Center for Biotechnology Information) and used as a reference sequence for designing the primers. Sequence-specific primers were designed by primer premier 5 in accordance with the length of the cDNA sequence of Haserpin1 to amplify the open reading frame (ORF). The primers were sense, 5'-GGATCCATGTCATCAGATTC CG ATGAACTTCTCA AGCA-3' and antisense, 5'GGTACCTTAATGATG ATGATGATGATGTG ATTGGATGACTCCGCTAAACAG AATGTTATTCGGTT-3', with BamHI and KpnI as restriction enzyme sites (underlined), respectively, and 6×-His-tag at the N terminus of the antisense primer.

2.5. Genomic DNA isolation and PCR

The FGENESH program (Salamov and Solovyev, 2000) was used to predict the number of potential exons and introns in the Haserpin1 genomic DNA. For experimental verification, whole bodies of *H. armigera* (n = 3) were used to isolate DNA (Wizard Genomic DNA Purification Kit, Promega®) for PCR. The oligonucleotide primers used for amplification were the same as those mentioned above. The amplification primers were designed using the cDNA sequence of Haserpin1. All PCR products were verified by DNA sequence analysis.

2.6. *H. armigera* Haserpin1 cloning

To amplify gene-specific primers, cDNA and a set of sense and anti-sense primers were used in PCR. For the digestion reaction, sense primer was incorporated into the BamHI restriction site, whereas to incorporate anti-sense primer, the KpnI restriction site was used. The PCR amplification was performed as follows: denaturation at 95 °C for 3 min, followed by 35 cycles at 95 °C for 30 sec, 55 °C for 30 sec, and 72 °C for 2 min, with a final extension at 72 °C for 5 min. The PCR product was then visually examined on a 1% (w/v) agarose gel stained with ethidium bromide using the BioRad imaging system. A gel extraction kit (Omega) was used to purify the amplified product, which was then ligated to the pGEM-T easy vector (TaKaRa) and transformed into *Escherichia coli* DH5α. A positive clone was selected on LB agar plates that contained 50 µg mL⁻¹ ampicillin after incubation at 37 °C overnight. The resulting clones were sequenced by Shanghai Sunny Biotech Co., Ltd. Triplicates were used for sequencing, and all the sequences were determined in both directions for at least 9-12 clones, and only those cloned PCR products showing high similarity with other serpin genes from insects were selected.

2.7. Biochemical properties and phylogenetic tree construction

NCBI Basic Protein BLAST (Basic Local Alignment Search Tool) was used to identify homologous serpin gene sequences from different insect species with the *H. armigera*

serpin (Hasperin1) sequence used as the query. PEPSTATS was used to compute the primary sequence composition of IPK2 (Rice et al., 2000). The physiochemical parameters of the Haserpin1 protein were predicted using the ProtParam tool (Gasteiger et al., 2005). The SignalP 4.1 webserver was used to identify the possible secretory signal peptide of Haserpin1 (Petersen et al., 2011). To detect the evolutionary location and phylogenetic similarities of Haserpin1 with other insect serpin genes, multiple sequence alignment (MSA) of selected amino acid sequences was carried out to produce quality alignments.

To better understand the evolutionary relationship of Haserpin1 with serpins from other insects, serpin sequences of *O. furnacalis*, *A. gambiae*, *Apis mellifera*, *B. mori*, *D. melanogaster*, *M. sexta*, *T. molitor*, *P. xylostella*, *Mamestra brassicae*, and *Mamestra configurata* were used to construct a phylogenetic tree using the maximum likelihood method in MEGA (Tamura et al., 2013). The maximum likelihood analysis was performed based on the JTT matrix-based model with 1,000 bootstrap replicates. The initial tree for the heuristic search was obtained automatically by applying the neighbor-joining and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model and then selecting the topology with the superior log-likelihood value. The tree is drawn to scale, with branch lengths indicating the number of substitutions per site.

2.8. Fluorescence real-time quantitative PCR analysis of gene expression

Real-time qPCR was performed using a BioRad iQ2 cyclor. The real-time qPCR reaction was carried out in PCR strips with SYBR Green; first-strand cDNA (in triplicate) for each sample served as the template for RT-qPCR using SsoFast™ EvaGreen® Supermix (Bio-Rad, Hercules, CA, USA) with an iQ2 Optical System (Bio-Rad). Every reaction was performed in iQ™ 96-well PCR plates containing a 20-µL volume containing cDNA and 100 nM of each primer and covered with adhesive seals (Bio-Rad). Primers for Haserpin1 (accession number: HQ615869) forward primer 5'-CTTCAGAAGGAGATTCTC-3' and reverse primer 5'-CTTCCTCTACCCATGTAT-3') and a housekeeping gene, the ribosomal protein L28 (accession number: DQ875266.1 forward primer 5'-CATCTGAACTGGATGATC-3' and reverse primer 5'-GTACACTACTGTGAAACC-3'), were designed by the Beacon Designer program (Corbett Robotics). Ribosomal protein L28 was used as the internal control for normalization (Shakeel et al., 2015). The reactions were initiated with denaturation for 30 sec at 95 °C, followed by 40 cycles at 95 °C for 5 sec at 60 °C for 10 s. The amplified product was then subjected to melting curve analysis at 55 to 95 °C to confirm the specificity and consistency. The relative expression of genes was calculated using the 2^{-ΔΔCT} method (Livak and Schmittgen, 2001).

3. Results

3.1. Cloning of *H. armigera* serpin

The present study provides a report on the identification and characterization of a serpin gene (Haserpin1) in *H. armigera*. The Haserpin1 gene sequence (accession number: HQ615869) was obtained from NCBI, and sequence-specific primers were used to amplify the Haserpin1 gene. The amplicon with the appropriate size of 1382 bp fragments containing an ORF of 1174 nucleotides and encoding 391 amino acids was obtained, confirming the presence of the Haserpin1 gene in *H. armigera*.

3.2. Genomic DNA sequence analysis of the *H. armigera* serpin

Genomic DNA sequence analysis was conducted to determine the number of introns and exons present in the Haserpin1 gene of *H. armigera*. For this purpose, FGENESH software was used to predict the Haserpin1 gene sequence. The results of sequence prediction revealed that the full-length genomic DNA sequence of Haserpin1 has one exon and no introns. To further confirm the presence of one exon and no introns in the Haserpin1 sequence, genomic DNA was isolated from *H. armigera* and used as a template for PCR. The amplification results showed an amplicon size of 1174 nucleotides. The PCR products were verified by DNA sequence analysis, which also demonstrated no intron in Haserpin1.

3.3. Biochemical properties and characterization of *H. armigera* serpin

The biochemical properties of the Haserpin1 sequence predicted by ProtParam indicated Leu (11.5%) as the most abundant amino acid, followed by Ala (9.2%), Ser (9.2%), Lys (8.7%) and Val (8.2%); Trp (0.5%) and His (0.8%) are the least abundant amino acids in the Haserpin1 sequence. The molecular weight of Haserpin1 was computed to be 43.53 kDa, with a pI of approximately pH 5. The instability index (Ii) of Haserpin1 was computed to be 37.99, suggesting that Haserpin1 is thermally stable with a long half-life. Moreover, the present study obtained a high aliphatic index (93.79) for the Haserpin1 protein. The extinction coefficient (EC) of the protein at 280 nm in water was 27,390. The gradient average hydrophobicity (GRAVY) index for Haserpin1 was calculated to be -0.203. InterPro scan analysis confirmed the identity of the serpin superfamily. The SignalP 4.1 web server identified the presence of a signal peptide at the N-terminus position of Haserpin1. Furthermore, the PANTHER classification system recognized Haserpin1 as a protease inhibitor belonging to the serpin family (PTHR11461).

When Haserpin1 was compared with other members of the serpin family, multiple sequence alignment showed that Haserpin1 has 65% identity with McSPI1a and MbSPI1a (*M. configurata* and *M. brassicae*), which was greater than the 59% identity with OfSPI (*O. furnacalis*), 54% identity with MsSPI (*M. sexta*), and 51% identity with PxSPI (*Papilio xuthus*) (Figure 1). Our analysis also identified

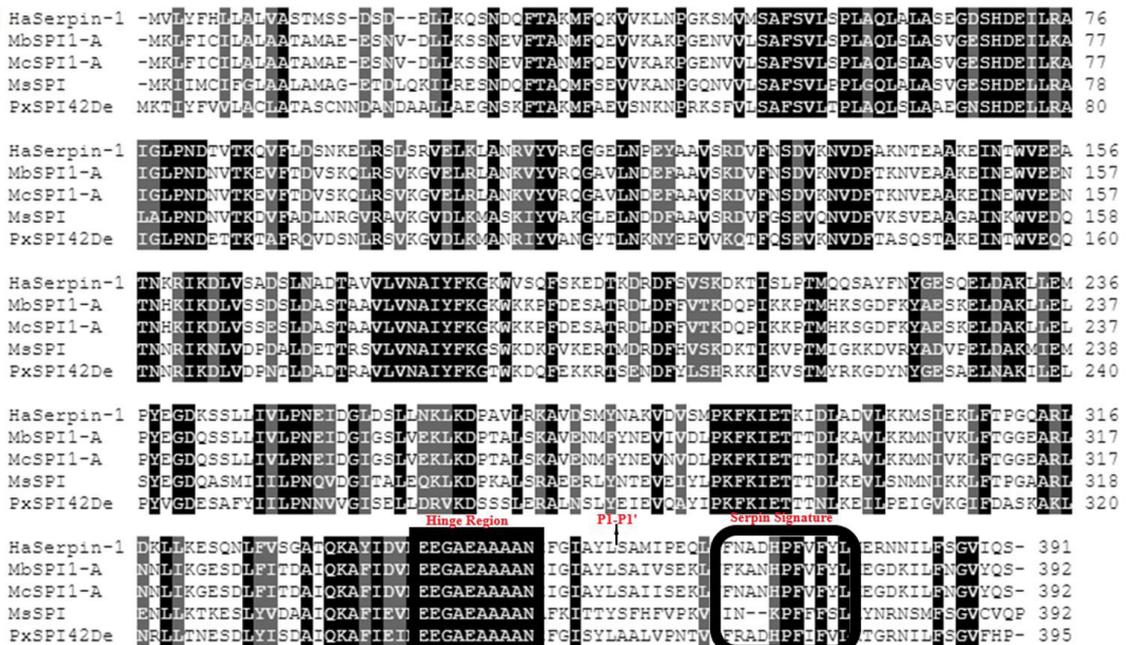


Figure 1. Alignment of the amino acid sequences between Haserpin1 and other known serine protease inhibitors is shown; the characteristic domains of Haserpin1 are boxed and named. The sources of the aligned sequences are *Helicoverpa armigera* Haserpin1 (GenBank accession number HQ615869), *Mamestra brassicae* SPI1A (ACR56864), *Mamestra configurata* SPI1a (AAN71632), *Ostrinia furnacalis* SPI1 (AFV46312), *Manduca sexta* SPI (AAA29334), and *Papilio xuthus* SPI42De (BAM18018).

an RCL, EEGAEAAAANEFGIAYLSAMIPE (341–363), in the C-terminal region of Haserpin1 with characteristic serpin domains, such as the serpin conservative hinge region EEGAEAAAAN (341–350), the serpin signature FNADHPFVF (367–375) and a P1-P1' cleavage site (L357–S358).

3.4. Phylogenetic tree analysis of *H. armigera* serpin

A phylogenetic tree was constructed based on Haserpin1 and other insect serpins using maximum likelihood analysis (Figure 2). The results revealed that insect serpins can be grouped into different clusters, with Haserpin1 being closely related to *M. brassicae* serpin1-A (MbSPI1-A) and *M. configurata* serpin1-A (McSPI1A) along with two serpins from *M. sexta* (MsSPI and MsSerp1-1J) and one from *B. mori* (Bmserpin-1A).

3.5. Prediction of *H. armigera* serpin secondary and tertiary structures

The secondary structure prediction for Haserpin1 conducted by SWISS-MODEL and Phyr2 indicated that Haserpin1 contains 13 α -helices and 10 β -strands (Figure 3). The predicted tertiary structure of Haserpin1 revealed that it has high similarity with *M. sexta* serpin1 (MsSerp1: PDB. 1sek) (Figure 4).

3.6. The mRNA expression profile of Haserpin1 in the fat body

In the present study, SYBER Green real-time PCR was employed to analyze the cDNA of *H. armigera* fat body tissue to determine the tissue-specific mRNA expression of Haserpin1. The RPL28 gene was used as an endogenous control. Our results revealed that Haserpin1 mRNA was

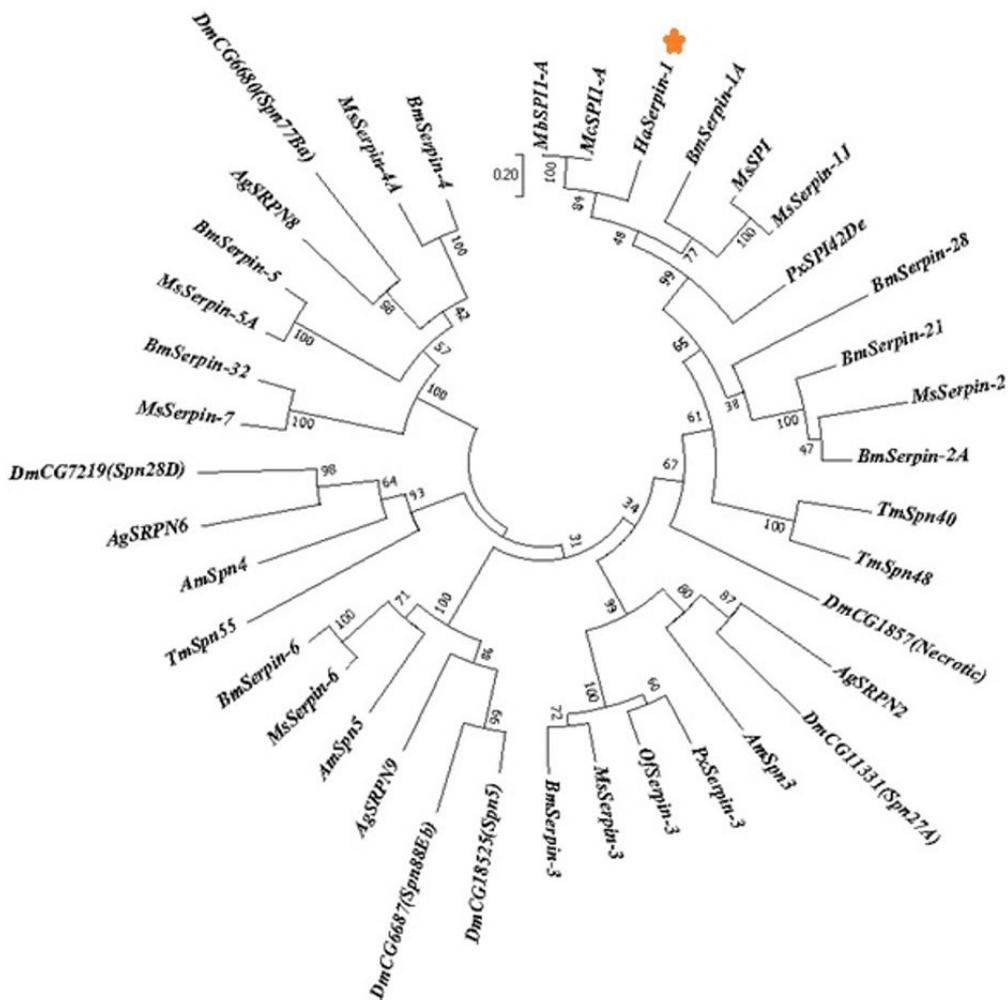


Figure 2. Phylogenetic tree showing the evolutionary relationship of *Helicoverpa armigera* serpin (Haserpin1) with serpins from other insect species: *Ostrinia furnacalis* (Of), *Anopheles gambiae* (Ag), *Apis mellifera* (Am), *Bombyx mori* (Bm), *Drosophila melanogaster* (Dm), *Manduca sexta* (Ms), *Tenebrio molitor* (Tm), *Plutella xylostella* (Px), *Mamestra brassicae* (Mb), and *Mamestra configurata* (Mc). The evolutionary history was inferred using the maximum likelihood method based on the JTT matrix-based model with 1,000 bootstrap replicates. The tree is drawn to scale, with branch lengths indicating the number of substitutions per site. Haserpin1 used in the present study is marked with an asterisk.

constitutively expressed in the fat body at the larval, prepupal, pupal, and adult stages (Figure 5). In addition, the level of Haserpin1 gene expression was relatively

elevated at the prepupal stage and on the 1st day of the pupal stage, whereas very low expression was found on other pupal days and at the adult stage.

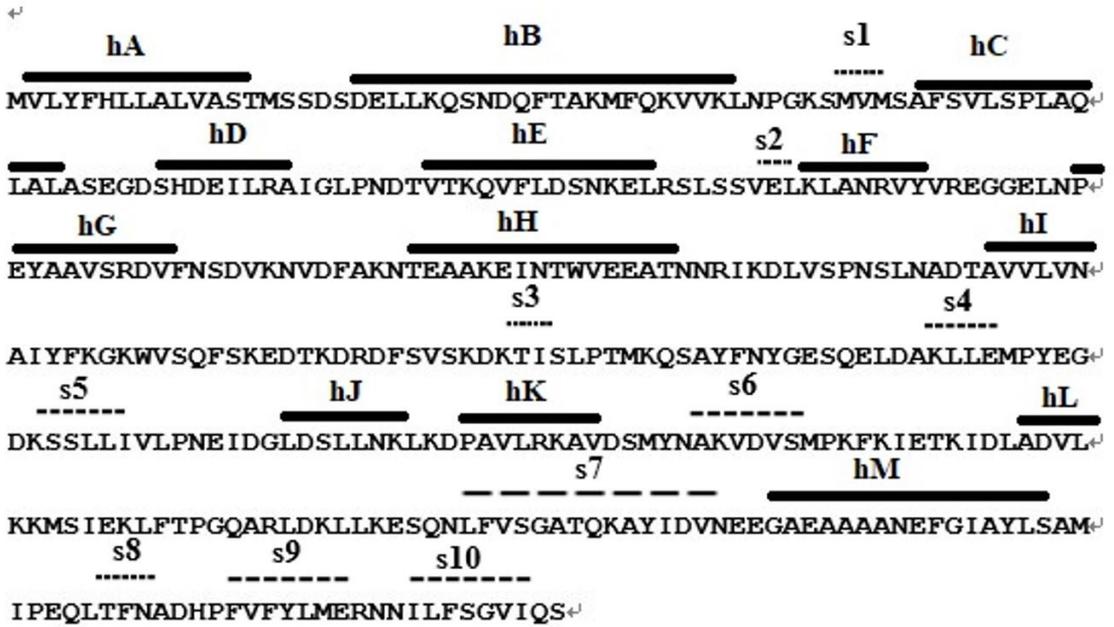


Figure 3. Predicted sequence of Haserpin1 amino acids indicating the location of conserved helices (alpha) and strands (beta). The bold lines represent alpha helices labelled from “hA” to “hM”; the broken lines represent beta-strands labelled from “s1” to “s10”.

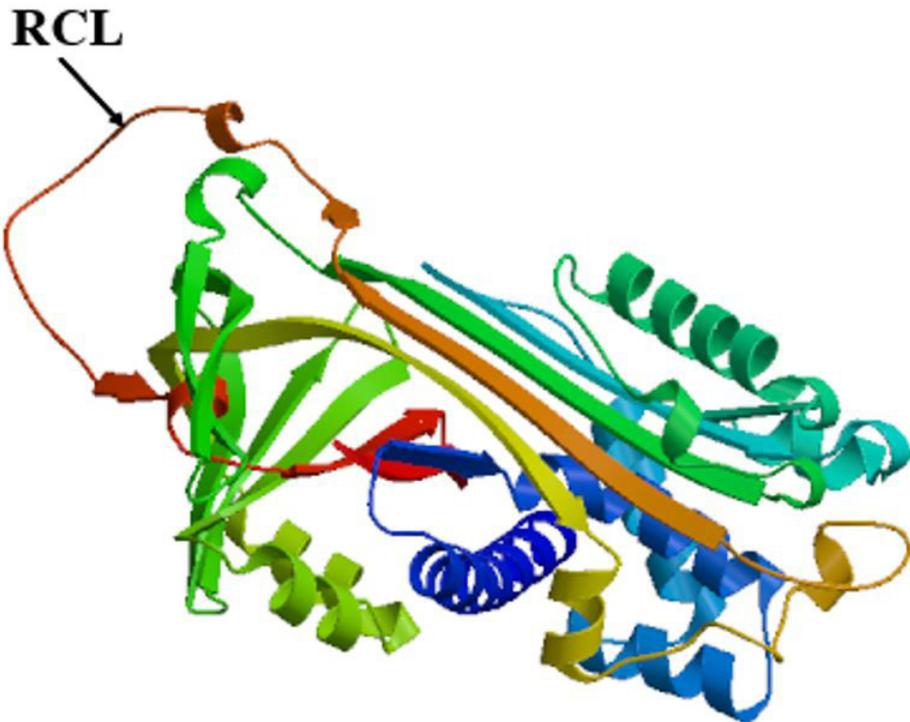


Figure 4. Predicted tertiary structure of Haserpin1 indicating the native conformation of the protein with the reactive center loop (RCL) region indicated by an arrow.

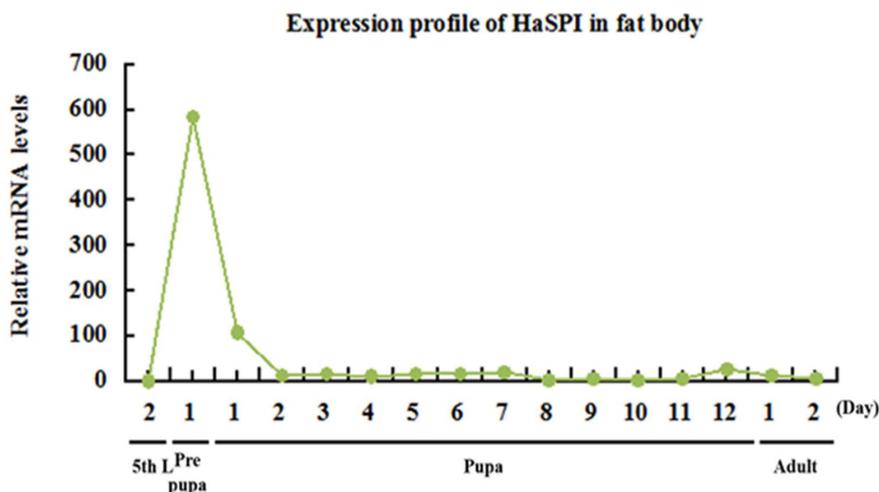


Figure 5. Expression profiles of Haserpin1 in *H. armigera* fat body tissue are shown. The expression profile of Haserpin1 in the fat body during *H. armigera* development was analyzed by RT-qPCR. The ribosomal protein L28 gene was used as the internal control for RT-qPCR. The error bars represent the mean \pm SE values ($n = 3$).

4. Discussion

Serine protease inhibitors, serpins, are the largest and most widely distributed family of protease inhibitors found in all organisms, though rarely in bacteria and fungi (Kanost and Jiang 1997). Serpins have been reported to play important roles in biological processes such as development, metamorphosis, and immunity (Bayer et al., 1997; Choo et al., 2012; Davie et al., 1991; Gál et al., 2013; Kim et al., 2013; Shigetomi et al., 2010). In the present study, a serpin sequence from *H. armigera*, designated Haserpin1, was obtained from NCBI. Currently, there are no other reports on the expression and characterization of *H. armigera* serpin1. Thus, our work presented herein represents the first report of the cloning, characterization, and expression analysis of an *H. armigera* serpin, which we designated Haserpin1. The cDNA sequence of Haserpin1 consists of a 1382-bp open reading frame encoding 391 amino acids with a calculated molecular mass of 43.53 kDa. The molecular weight of our protein is in accordance with other serpins reported in insects (Shukla et al., 2015).

The biochemical analysis of Haserpin1 demonstrated that leucine is the most abundant amino acid, comprising approximately 11.5% of its residues. The isoelectric point (pI) of Haserpin1 is approximately 5, indicating that it might be soluble in an acidic buffer. Instability indices (Ii) are used to determine in vivo half-lives, and a protein with an Ii value lower than 40 is predicted to be stable (Rogers et al., 1986). In the present study, the Ii value of Haserpin1 was less than 40, which indicates that it is thermally stable with a long half-life. The aliphatic index of Haserpin1 was also assessed to evaluate the relative volume occupied by aliphatic side chains (Ikai, 1980). The results indicated that Haserpin1 has a high aliphatic index, which indicates that it may be stable at a high

temperature (Ikai, 1980). GRAVY analysis was conducted to determine the hydrophobicity of Haserpin1, whereby a positive GRAVY value indicates a hydrophobic nature and negative GRAVY value a hydrophilic nature (Kyte and Doolittle, 1982). Our results showed that Haserpin1 is hydrophilic and can interact favorably with water.

Moreover, sequence alignment analysis of Haserpin1 with other insect serpins revealed that it shares the highest (65%) identity with *M. configurata* and *M. brassicae* and the lowest (51%) with *Papilio xuthus*. These sequence similarity results are in accordance with other reports of insect serpin sequence similarity; for example, *O. furnacalis* serpin1 (ofserpin1) shares a 60% identity with *M. configurata* serpin1a and 55% with *P. xylostella* serpin1a (Zhang et al., 2016).

The deduced amino acid sequence of Haserpin1 shows the presence of a signal peptide at the N-terminus, indicating that Haserpin1 is a secretory protein and probably plays an important role in inhibiting extracellular serine proteinase cascades. The potential signal peptide found in Haserpin1 has also been reported in several other secretory serpins (Babin et al., 1984; Bania et al., 1999; Cierpicki et al., 2000; Huang et al., 1994; Kim et al., 2013). The C-terminus of Haserpin1 is composed of characteristic serpin domains, an RCL with a scissile bond between P1-P1', which may be cleaved by the target protease, a serpin signature, and a hinge region, which are also found in all other insect serpins (Chamankhah et al., 2003; Zheng et al., 2009; Zhang et al., 2016). The P1 residue of the RCL region is considered a determinant of substrate specificity. Previously, it has been reported that the presence of Arg or Lys as a P1 residue indicates that the serpin will target trypsin-like proteases (Gulley et al., 2013). Similarly, serpins containing Pro, Tyr, Phe, Leu, or Met at the P1 residue are predicted to inhibit chymotrypsin and chymotrypsin-like proteases (Laskowski

Junior and Kato, 1980). In the present study, a Leu residue at the P1 position of Haserpin1 was observed, suggesting that this serpin might inhibit chymotrypsin-like proteases. A similar Leu residue was found at the P1 position of *O. furnacalis* serpin (ofserpin1), with inhibitory activities against trypsin and chymotrypsin (Zhang et al., 2016).

To illustrate the biological role of Haserpin1 in *H. armigera*, the expression pattern of mRNA was investigated in the fat body at different developmental stages by real-time quantitative PCR. In previous studies, it has been reported that the fat body is the major site for the synthesis of serpins, and constitutive and high expression of serpins has been observed in the fat body (Chamankhah et al., 2003; Li et al., 2012; Liu et al., 2015). Our results also showed that Haserpin1 was constitutively expressed in the fat body, which suggests that Haserpin1 is a serine protease inhibitor derived from the *H. armigera* fat body. Variable expression of Haserpin1 was observed during development. At the 5th-instar larval stage, low expression was detected, which was relatively elevated at the prepupal and on the 1st day of the pupal stage; conversely, very low expression was found on other pupal days and at the adult stage. The variation in the expression of Haserpin1 observed during development in our study is not uncommon, as a similar phenomenon has been observed in previous studies. For example, in *Antheraea pernyi*, expression of Apserspin1 was lowest at the 4th larval instar and highest at the pupal stage (Yu et al., 2017). Similarly, serpin1 of *Choristoneura fumiferana* showed high expression at the intermolt phase compared with the molting phase (Zheng et al., 2009). Further evidence that supports the fluctuation of serpin gene expression at different developmental stages has been obtained from *M. sexta*, in which serpin1 was detected at lower levels during molting and wandering stages (Kanost et al., 1995). These results from different insect species suggest that serpin1 gene expression fluctuates at different developmental stages.

5. Conclusion

In conclusion, a serpin designated Haserpin1 was identified and characterized for the first time in the fat body at different developmental stages of *H. armigera*. Constitutive but fluctuating expression of Haserpin1 was observed in the fat body at different stages of development. Our results indicate that the fat body is the major site of serpin synthesis in *H. armigera*. Furthermore, higher expression at the prepupal stage suggests that Haserpin1 might play a critical role at the wandering stage of *H. armigera*. Although the detailed function of this serpin (Haserpin1) still needs to be investigated, our study provides a perspective for the functional investigation of serine protease inhibitor genes.

Acknowledgments

We are thankful to the anonymous reviewers for their useful comments on the initial draft of the manuscript. This study was supported by a grant from the Postdoctoral Science Foundation of China (2018M640788).

References

- ABBASI, B.H., AHMED, K., KHALIQUE, F., AYUB, N., LIU, H.J., RAZA KAZMI, S.A. and AFTAB, M.N., 2007. Rearing the cotton bollworm, *Helicoverpa armigera*, on a tapioca-based artificial diet. *Journal of Insect Science*, vol. 7, no. 1, pp. 35-35. <http://dx.doi.org/10.1673/031.007.3501>. PMID:20302464.
- AN, C., BUDD, A., KANOST, M. R., and MICHEL, K., 2011. Characterization of a regulatory unit that controls melanization and affects longevity of mosquitoes. *Cellular and Molecular Life Sciences: CMLS*, vol. 68, no. 11, pp. 1929-1939.
- AN, C. and KANOST, M.R., 2010. *Manduca sexta* serpin-5 regulates prophenoloxidase activation and the Toll signaling pathway by inhibiting hemolymph proteinase HP6. *Insect Biochemistry and Molecular Biology*, vol. 40, no. 9, pp. 683-689. <http://dx.doi.org/10.1016/j.ibmb.2010.07.001>. PMID:20624461.
- BABIN, D.R., PEANASKY, R.J. and GOOS, S.M., 1984. The iso-inhibitors of chymotrypsin/elastase from *Ascaris lumbricoides*: the primary structure. *Archives of Biochemistry and Biophysics*, vol. 232, no. 1, pp. 143-161. [http://dx.doi.org/10.1016/0003-9861\(84\)90530-7](http://dx.doi.org/10.1016/0003-9861(84)90530-7). PMID:6564898.
- BANIA, J., STACHOWIAK, D. and POLANOWSKI, A., 1999. Primary structure and properties of the cathepsin G/chymotrypsin inhibitor from the larval hemolymph of *Apis mellifera*. *European Journal of Biochemistry*, vol. 262, no. 3, pp. 680-687. <http://dx.doi.org/10.1046/j.1432-1327.1999.00406.x>. PMID:10411628.
- BAYER, C.A., KALM, L. and FRISTROM, J.W., 1997. Relationships between protein isoforms and genetic functions demonstrate functional redundancy at the broad-complex during drosophila metamorphosis. *Developmental Biology*, vol. 187, no. 2, pp. 267-282. <http://dx.doi.org/10.1006/dbio.1997.8620>. PMID:9242423.
- CHAMANKHAH, M., BRAUN, L., VISAL-SHAH, S., O'GRADY, M., BALDWIN, D., SHI, X., HEMMINGSEN, S., ALTING-MEES, M. and HEGEDUS, D., 2003. Mamestra configurata serpin-1 homologues: cloning, localization and developmental regulation. *Insect Biochemistry and Molecular Biology*, vol. 33, no. 3, pp. 355-369. [http://dx.doi.org/10.1016/S0965-1748\(02\)00263-1](http://dx.doi.org/10.1016/S0965-1748(02)00263-1). PMID:12609520.
- CHOO, Y.M., LEE, K.S., YOON, H.J., QIU, Y., WAN, H., SOHN, M.R., SOHN, H.D. and JIN, B.R., 2012. Antifibrinolytic role of a bee venom serine protease inhibitor that acts as a plasmin inhibitor. *PLoS One*, vol. 7, no. 2, pp. e32269. <http://dx.doi.org/10.1371/journal.pone.0032269>. PMID:22359676.
- CHU, Y., HONG, F., LIU, Q. and AN, C., 2017. Serine protease SP105 activates prophenoloxidase in Asian corn borer melanization, and is regulated by serpin-3. *Scientific Reports*, vol. 7, no. 1, pp. 45256. <http://dx.doi.org/10.1038/srep45256>. PMID:28358031.
- CIERPICKI, T., BANIA, J., and OTLEWSKI, J., 2000. NMR solution structure of *Apis mellifera* chymotrypsin/cathepsin G inhibitor-1 (AMCI-1): structural similarity with *Ascaris* protease inhibitors. *Protein Science*, vol. 9, no. 5, pp. 976-984. <http://dx.doi.org/10.1110/ps.9.5.976>.
- DAVIE, E.W., FUJIKAWA, K. and KISIEL, W., 1991. The coagulation cascade: initiation, maintenance, and regulation. *Biochemistry*, vol. 30, no. 43, pp. 10363-10370. <http://dx.doi.org/10.1021/bi00107a001>. PMID:1931959.
- DISSANAYAKE, S.N., MARINOTTI, O., RIBEIRO, J.M.C. and JAMES, A.A., 2006. *Anopheles gambiae* gene expression database with integrated comparative algorithms for identifying

- conserved DNA motifs in promoter sequences. *BMC Genomics*, vol. 7, no. 1, pp. 116-116. <http://dx.doi.org/10.1186/1471-2164-7-116>. PMID:16707020.
- ELEFThERIANOS, I. and REVENIS, C., 2011. Role and importance of phenoloxidase in insect hemostasis. *Journal of Innate Immunity*, vol. 3, no. 1, pp. 28-33. <http://dx.doi.org/10.1159/000321931>. PMID:21051882.
- GÁL, P., DOBÓ, J., BEINROHR, L., PÁL, G. and ZÁVODSZKY, P., 2013. Inhibition of the serine proteases of the complement system. In: J. D. LAMBRIS, V. M. HOLERS and D. RICKLIN, eds. *Complement therapeutics*. Boston, MA: Springer. http://dx.doi.org/10.1007/978-1-4614-4118-2_2.
- GASTEIGER, E., HOOGLAND, C., GATTIKER, A., DUVAUD, S. E., WILKINS, M. R., APPEL, R. D., and BAIROCH, A., 2005. Protein identification and analysis tools on the ExPASy server. In: J. M. WALKER, eds. *The proteomics protocols handbook*. Totowa: Humana Press. Springer Protocols Handbooks. <https://doi.org/10.1385/1-59259-890-0:571>.
- GUBB, D., SANZ-PARRA, A., BARCENA, L., TROXLER, L. and FULLAONDO, A., 2010. Protease inhibitors and proteolytic signaling cascades in insects. *Biochimie*, vol. 92, no. 12, pp. 1749-1759. <http://dx.doi.org/10.1016/j.biochi.2010.09.004>. PMID:20850496.
- GULLEY, M.M., ZHANG, X. and MICHEL, K., 2013. The roles of serpins in mosquito immunology and physiology. *Journal of Insect Physiology*, vol. 59, no. 2, pp. 138-147. <http://dx.doi.org/10.1016/j.jinsphys.2012.08.015>. PMID:22960307.
- HUANG, K., STRYNADKA, N.C.J., BERNARD, V.D., PEANASKY, R.J. and JAMES, M.N.G., 1994. The molecular structure of the complex of *Ascaris* chymotrypsin/elastase inhibitor with porcine elastase. *Structure (London, England : 1993)*, vol. 2, no. 7, pp. 679-689. [http://dx.doi.org/10.1016/S0969-2126\(00\)00068-X](http://dx.doi.org/10.1016/S0969-2126(00)00068-X). PMID:7922044.
- IKAI, A., 1980. Thermostability and aliphatic index of globular proteins. *Journal of Biochemistry*, vol. 88, no. 6, pp. 1895-1898. PMID:7462208.
- JARASSRAMEE, B., SUPUNGUL, P., PANYIM, S., KLINBUNGA, S., RIMPHANICHAYAKIT, V. and TASSANAKAJON, A., 2005. Recombinant expression and characterization of five-domain Kazal-type serine proteinase inhibitor of black tiger shrimp (*Penaeus monodon*). *Marine Biotechnology (New York, N.Y.)*, vol. 7, no. 1, pp. 46-52. <http://dx.doi.org/10.1007/s10126-004-0100-6>.
- JIANG, R., KIM, E.-H., GONG, J.-H., KWON, H.-M., KIM, C.-H., RYU, K.-H., PARK, J.-W., KUROKAWA, K., ZHANG, J., GUBB, D. and LEE, B.-L., 2009. Three pairs of protease-serpin complexes cooperatively regulate the insect innate immune responses. *The Journal of Biological Chemistry*, vol. 284, no. 51, pp. 35652-35658. <http://dx.doi.org/10.1074/jbc.M109.071001>. PMID:19858208.
- KANOST, M.R. and JIANG, H., 1997. Serpins from an insect, *Manduca sexta*. In: C. F. CHURCH, D. D. CUNNINGHAM, D. GINSBURG, M. HOFFMAN, S. R. STONE and D. M. TOLLEFSEN, eds. *Chemistry and biology of serpins*. Boston, MA: Springer, pp. 155-161. http://dx.doi.org/10.1007/978-1-4615-5391-5_15.
- KANOST, M.R., PRASAD, S.V., HUANG, Y. and WILLOTT, E., 1995. Regulation of serpin gene-1 in *Manduca sexta*. *Insect Biochemistry and Molecular Biology*, vol. 25, no. 2, pp. 285-291. [http://dx.doi.org/10.1016/0965-1748\(94\)00067-R](http://dx.doi.org/10.1016/0965-1748(94)00067-R). PMID:7711756.
- KIM, B.Y., LEE, K.S., WAN, H., ZOU, F.M., CHOI, Y.S., YOON, H.J., KWON, H.W., JE, Y.H. and JIN, B.R., 2013. Anti-elastolytic activity of a honeybee (*Apis cerana*) chymotrypsin inhibitor. *Biochemical and Biophysical Research Communications*, vol. 430, no. 1, pp. 144-149. <http://dx.doi.org/10.1016/j.bbrc.2012.11.056>. PMID:23200835.
- KROWARSCHE, D., CIERPICKI, T., JELEN, F. and OTLEWSKI, J., 2003. Canonical protein inhibitors of serine proteases. *Cellular and Molecular Life Sciences CMLS*, vol. 60, no. 11, pp. 2427-2444. <http://dx.doi.org/10.1007/s00018-003-3120-x>. PMID:14625687.
- KYTE, J. and DOOLITTLE, R.F., 1982. A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*, vol. 157, no. 1, pp. 105-132. [http://dx.doi.org/10.1016/0022-2836\(82\)90515-0](http://dx.doi.org/10.1016/0022-2836(82)90515-0). PMID:7108955.
- LASKOWSKI JUNIOR, M. and KATO, I., 1980. Protein inhibitors of proteinases. *Annual Review of Biochemistry*, vol. 49, no. 1, pp. 593-626. <http://dx.doi.org/10.1146/annurev.bi.49.070180.003113>. PMID:6996568.
- LI, Y., ZHAO, P., LIU, S., DONG, Z., CHEN, J., XIANG, Z. and XIA, Q., 2012. A novel protease inhibitor in *Bombyx mori* is involved in defense against *Beauveria bassiana*. *Insect Biochemistry and Molecular Biology*, vol. 42, no. 10, pp. 766-775. <http://dx.doi.org/10.1016/j.ibmb.2012.07.004>. PMID:22841512.
- LIU, D., WANG, L., YANG, L., QIAN, C., WEI, G., DAI, L., LI, J., ZHU, B. and LIU, C., 2015. Serpin-15 from *Bombyx mori* inhibits prophenoloxidase activation and expression of antimicrobial peptides. *Developmental and Comparative Immunology*, vol. 51, no. 1, pp. 22-28. <http://dx.doi.org/10.1016/j.dci.2015.02.013>. PMID:25720980.
- LIVAK, K.J. and SCHMITTGEN, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻ $\Delta\Delta$ CT method. *Methods (San Diego, Calif.)*, vol. 24, no. 4, pp. 402-408. <http://dx.doi.org/10.1006/meth.2001.1262>.
- NEURATH, H., 1989. Proteolytic processing and physiological regulation. *Trends in Biochemical Sciences*, vol. 14, no. 7, pp. 268-271. [http://dx.doi.org/10.1016/0968-0004\(89\)90061-3](http://dx.doi.org/10.1016/0968-0004(89)90061-3). PMID:2672446.
- PETERSEN, T.N., BRUNAK, S., VON HEIJNE, G. and NIELSEN, H., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, vol. 8, no. 10, pp. 785-786. <http://dx.doi.org/10.1038/nmeth.1701>. PMID:21959131.
- RAWLINGS, N.D., TOLLE, D.P. and BARRETT, A.J., 2004. Evolutionary families of peptidase inhibitors. *The Biochemical Journal*, vol. 378, no. Pt 3, pp. 705-716. <http://dx.doi.org/10.1042/bj20031825>. PMID:14705960.
- REICHHART, J.-M., 2005. Tip of another iceberg: drosophila serpins. *Trends in Cell Biology*, vol. 15, no. 12, pp. 659-665. <http://dx.doi.org/10.1016/j.tcb.2005.10.001>. PMID:16260136.
- REICHHART, J.M., GUBB, D. and LECLERC, V., 2011. The *Drosophila* serpins: multiple functions in immunity and morphogenesis. In: *Methods in Enzymology*, vol. 499, pp. 205-225. <http://dx.doi.org/10.1016/B978-0-12-386471-0.00011-0>. PMID:21683256.
- RICE, P., LONGDEN, I. and BLEASBY, A., 2000. EMBOSS: the European molecular biology open software suite. *Trends in Genetics*, vol. 16, no. 6, pp. 276-277. [http://dx.doi.org/10.1016/S0168-9525\(00\)02024-2](http://dx.doi.org/10.1016/S0168-9525(00)02024-2). PMID:10827456.
- ROBERTS, T.H., HEJGAARD, J., SAUNDERS, N.F., CAVICCHIOLI, R. and CURMI, P.M., 2004. Serpins in unicellular Eukarya, Archaea, and Bacteria: sequence analysis and evolution. *Journal of Molecular Evolution*, vol. 59, no. 4, pp. 437-447. <http://dx.doi.org/10.1007/s00239-004-2635-6>. PMID:15638455.

- ROGERS, S., WELLS, R. and RECHSTEINER, M., 1986. Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. *Science*, vol. 234, no. 4774, pp. 364-368. <http://dx.doi.org/10.1126/science.2876518>. PMID:2876518.
- ROSS, J., JIANG, H., KANOST, M.R. and WANG, Y., 2003. Serine proteases and their homologs in the *Drosophila melanogaster* genome: an initial analysis of sequence conservation and phylogenetic relationships. *Gene*, vol. 304, pp. 117-131.
- SALAMOV, A.A. and SOLOVYEV, V.V., 2000. Ab initio gene finding in *Drosophila* genomic DNA. *Genome Research*, vol. 10, no. 4, pp. 516-522. <http://dx.doi.org/10.1101/gr.10.4.516>. PMID:10779491.
- SHAKEEL, M. and ZAFAR, J., 2020. Molecular identification, characterization, and expression analysis of a novel trypsin inhibitor-like cysteine-rich peptide from the cotton bollworm, *Helicoverpa armigera* (hübner) (Lepidoptera: noctuidae). *Egyptian Journal of Biological Pest Control*, vol. 30, no. 1, pp. 10. <http://dx.doi.org/10.1186/s41938-020-0208-7>.
- SHAKEEL, M., XU, X., DE MANDAL, S. and JIN, F., 2019. Role of serine protease inhibitors in insect-host-pathogen interactions. *Archives of Insect Biochemistry and Physiology*, vol. 102, no. 3, pp. e21556. <http://dx.doi.org/10.1002/arch.21556>.
- SHAKEEL, M., ZHU, X., KANG, T., WAN, H. and LI, J., 2015. Selection and evaluation of reference genes for quantitative expression studies in cotton bollworm, *Helicoverpa armigera* (Lepidoptera: noctuidae). *Journal of Asia-Pacific Entomology*, vol. 18, no. 2, pp. 123-130. <http://dx.doi.org/10.1016/j.aspen.2015.01.001>.
- SHIGETOMI, H., ONOGI, A., KAJIWARA, H., YOSHIDA, S., FURUKAWA, N., HARUTA, S., TANASE, Y., KANAYAMA, S., NOGUCHI, T., YAMADA, Y., OI, H. and KOBAYASHI, H., 2010. Anti-inflammatory actions of serine protease inhibitors containing the Kunitz domain. *Inflammation Research*, vol. 59, no. 9, pp. 679-687. <http://dx.doi.org/10.1007/s00011-010-0205-5>. PMID:20454830.
- SHUKLA, E., THORAT, L.J., NATH, B.B. and GAIKWAD, S.M., 2015. Insect trehalase: physiological significance and potential applications. *Glycobiology*, vol. 25, no. 4, pp. 357-367. <http://dx.doi.org/10.1093/glycob/cwu125>. PMID:25429048.
- TABASHNIK, B.E., BRÉVAULT, T. and CARRIÈRE, Y., 2013. Insect resistance to Bt crops: lessons from the first billion acres. *Nature Biotechnology*, vol. 31, no. 6, pp. 510-521. <http://dx.doi.org/10.1038/nbt.2597>. PMID:23752438.
- TAMURA, K., STECHER, G., PETERSON, D., FILIPSKI, A. and KUMAR, S., 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution*, vol. 30, no. 12, pp. 2725-2729. <http://dx.doi.org/10.1093/molbev/mst197>. PMID:24132122.
- TAY, W.T., SORIA, M.F., WALSH, T., THOMAZONI, D., SILVIE, P., BEHERE, G.T., ANDERSON, C. and DOWNES, S., 2013. A brave new world for an old world pest: *Helicoverpa armigera* (Lepidoptera: Noctuidae) in Brazil. *PLoS One*, vol. 8, no. 11, pp. e80134. <http://dx.doi.org/10.1371/journal.pone.0080134>. PMID:24260345.
- ULVILA, J., VANHA-AHO, L.-M. and RÄMET, M., 2011. *Drosophila* phagocytosis – still many unknowns under the surface. *APMIS*, vol. 119, no. 10, pp. 651-662. <http://dx.doi.org/10.1111/j.1600-0463.2011.02792.x>. PMID:21917002.
- WANG, Z.-H., ZHAO, X.-F. and WANG, J.-X., 2009. Characterization, kinetics, and possible function of Kazal-type proteinase inhibitors of Chinese white shrimp, *Fenneropenaeus chinensis*. *Fish & Shellfish Immunology*, vol. 26, no. 6, pp. 885-897. <http://dx.doi.org/10.1016/j.fsi.2009.03.024>. PMID:19379816.
- YANG, L., MEI, Y., FANG, Q., WANG, J., YAN, Z., SONG, Q., LIN, Z. and YE, G., 2017. Identification and characterization of serine protease inhibitors in a parasitic wasp, *Pteromalus puparum*. *Scientific Reports*, vol. 7, no. 1, pp. 15755. <http://dx.doi.org/10.1038/s41598-017-16000-5>. PMID:29147019.
- YU, H.M., ZHU, B.J., SUN, Y., WEI, G.Q., WANG, L., QIAN, C., NADEEM ABBAS, M. and LIU, C.L., 2017. Characterization and functional analysis of serpin-1 like gene from oak silkworm *Antheraea pernyi*. *Bulletin of Entomological Research*, vol. 107, no. 5, pp. 620-626. <http://dx.doi.org/10.1017/S000748531700013X>. PMID:28228181.
- ZHANG, B., WU, T., TANG, X., ZHANG, S., XU, Q., ZHAO, Y., WANG, Y. and FENG, C., 2016. Cloning, expression and characterization of *Ostrinia furnacalis* serpin1, a regulator of the prophenoloxidase activation system. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, vol. 192, pp. 9-20. <https://doi.org/10.1016/j.cbpb.2015.11.004>.
- ZHAO, P., WANG, G.-H., DONG, Z.-M., DUAN, J., XU, P.-Z., CHENG, T.-C., XIANG, Z.-H. and XIA, Q.-Y., 2010. Genome-wide identification and expression analysis of serine proteases and homologs in the silkworm *Bombyx mori*. *BMC Genomics*, vol. 11, no. 1, pp. 405. <http://dx.doi.org/10.1186/1471-2164-11-405>. PMID:20576138.
- ZHENG, Y.-P., HE, W.-Y., BÉLIVEAU, C., NISOLE, A., STEWART, D., ZHENG, S.-C., DOUCET, D., CUSSON, M. and FENG, Q.-L., 2009. Cloning, expression and characterization of four serpin-1 cDNA variants from the spruce budworm, *Choristoneura fumiferana*. *Comparative Biochemistry and Physiology Part B, Biochemistry & Molecular Biology*, vol. 154, no. 2, pp. 165-173. <http://dx.doi.org/10.1016/j.cbpb.2009.05.016>. PMID:19524698.
- ZOU, Z., LOPEZ, D.L., KANOST, M.R., EVANS, J.D. and JIANG, H., 2006. Comparative analysis of serine protease-related genes in the honey bee genome: possible involvement in embryonic development and innate immunity. *Insect Molecular Biology*, vol. 15, no. 5, pp. 603-614. <http://dx.doi.org/10.1111/j.1365-2583.2006.00684.x>. PMID:17069636.
- ZOU, Z., SHIN, S.W., ALVAREZ, K.S., KOKOZA, V. and RAIKHEL, A.S., 2010. Distinct melanization pathways in the mosquito *Aedes aegypti*. *Immunity*, vol. 32, no. 1, pp. 41-53. <http://dx.doi.org/10.1016/j.immuni.2009.11.011>. PMID:20152169.