Published in final edited form as:

Insect Biochem Mol Biol. 2018 November ; 102: 21–30. doi:10.1016/j.ibmb.2018.09.008.

The *Manduca sexta* serpinome: Analysis of serpin genes and proteins in the tobacco hornworm

Miao Li^{#a}, Jayne M. Christen^{#a,d}, Neal T. Dittmer^a, Xiaolong Cao^b, Xiufeng Zhang^b, Haobo Jiang^b, and Michael R. Kanost^{a,e}

^aDepartment of Biochemistry and Molecular Biophysics, Kansas State University, Manhattan, KS 66506, USA

^bDepartment of Entomology and Plant Pathology, Oklahoma State University, Stillwater, OK 74078, USA

^dCurrent address: Leidos Life Sciences, Frederick, MD 21703, USA

[#] These authors contributed equally to this work.

Abstract

Members of the serpin superfamily of proteins occur in animals, plants, bacteria, archaea and some viruses. They adopt a variety of physiological functions, including regulation of immune system, modulation of apoptosis, hormone transport and acting as storage proteins. Most members of the serpin family are inhibitors of serine proteinases. In this study, we searched the genome of *Manduca sexta* and identified 32 serpin genes. We analyzed the structure of these genes and the sequences of their encoded proteins. Three *M. sexta* genes (*serpin-1*, *serpin-15*, and *serpin-28*) have mutually exclusive alternatively spliced exons encoding the carboxyl terminal reactive center loop of the protein, which is the site of interaction with target proteases. We discovered that MsSerp1 has 14 splicing isoforms, including two undiscovered in previous studies. Twenty-eight of the 32 *M. sexta* serpins include a putative secretion signal peptide and are predicted to be extracellular proteins. Phylogenetic analysis of serpins in *M. sexta* and *Bombyx mori* indicates that 17 are orthologous pairs, perhaps carrying out essential physiological functions. Analysis of the reactive center loop and hinge regions of the protein sequences indicates that 16 of the serpin genes encode proteins that may lack proteinase inhibitor activity. Our annotation and analysis of these serpin genes and their transcript profiles should lead to future advances in experimental study of their functions in insect biochemistry.

Graphical Abstract

^eCorresponding author: Michael R. Kanost, Department of Biochemistry and Molecular Biophysics, Kansas State University, Manhattan, KS 66506 USA, Kanost@ksu.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Manduca sexta serpins

Keywords

hemolymph; protease inhibitor; serpin; Lepidoptera; serine protease; insect immunity

1. Introduction

Serpins are a superfamily of proteins, many of which inhibit serine proteinases. They share a conserved tertiary structure composed of three β -sheets, 8–9 α -helices, and an exposed reactive center loop (RCL) (Silverman *et al.*, 2001). The mechanism for proteinase inhibition by serpins includes cleavage by the proteinase of a specific peptide bond (the scissile bond) between residues designated P1-P1' in the serpin RCL. The inhibition reaction begins with the formation of a noncovalent complex between the proteinase and serpin through interaction of the proteinase active site with the serpin RCL (Ye *et al.*, 2001). Once the P1 side chain is correctly positioned in the primary specificity pocket of the proteinase, the active site serine of the proteinase attacks and cleaves the scissile bond, resulting in

formation of a covalent acyl complex of the serpin and proteinase. The serpin undergoes a large conformational change, in which the RCL inserts as the fourth strand in its β -sheet A. The proteinase is translocated by about 70 Å, and its active site catalytic triad becomes distorted, resulting in proteinase inactivation (Gettins and Olson, 2016).

Serpins have been identified in animals, plants, bacteria, archaea, and some viruses (Gettins, 2002). Insect serpins have been a subject of significant research efforts in the last three decades (Meekins *et al.*, 2017). Biochemical studies and genetic analyses have shown that some insect serpins regulate extracellular proteinase cascades in the hemolymph, which stimulate innate immune responses, including the Toll pathway and prophenoloxidase activation (Shin *et al.*, 2006; Jiang *et al.*, 2009; Park *et al.*, 2011; Levashina *et al.*, 1999; Ahmad *et al.*, 2009; Zou and Jiang, 2005; An and Kanost, 2010; An *et al.*, 2011b; Zou *et al.*, 2010; Abraham *et al.*, 2005; Michel *et al.*, 2006; An *et al.*, 2011a; Ligoxygakis *et al.*, 2002; De Gregorio *et al.*, 2002; Scherfer *et al.*, 2008; Tang *et al.*, 2008; Jiang *et al.*, 2003; Zhu *et al.*, 2003; Wang and Jiang, 2004; Tong and Kanost, 2005; Tong *et al.*, 2005; Wang and Jiang, 2006; He *et al.*, 2017; Yang *et al.*, 2018).

Completion of insect genome projects has revealed the presence of many previously unknown serpin genes. Thirty-four serpin genes have been identified in *B. mori* (Zou *et al.*, 2009), thirty-one in *Tribolium castaneum* (Zou *et al.*, 2007), twenty-nine in *Drosophila melanogaster* (Reichhart, 2005; Reichhart *et al.*, 2011), eighteen in *Anopheles gambiae* (Christophides *et al.*, 2002; Suwanchaichinda and Kanost, 2009), and at least five serpin genes in *Apis mellifera* (Zou *et al.*, 2006). However, physiological functions of most insect serpins are unknown.

The first insect serpin with a known sequence was isolated from hemolymph of the tobacco hornworm, *Manduca sexta* (Kanost *et al.*, 1989). Since then, nine additional serpins from *M. sexta* have been identified and investigated through biochemical studies (Jiang *et al.*, 1994; Gan *et al.*, 2001; Zhu *et al.*, 2003; Tong and Kanost, 2005; Zou and Jiang, 2005; Suwanchaichinda *et al.*, 2013; He *et al.*, 2017; Yang *et al.*, 2018). In this study, we used a bioinformatics approach to identify additional serpin genes in the *M. sexta* genome (Kanost *et al.*, 2016). We annotated and analyzed 31 genes encoding serpin proteins and one likely serpin pseudogene. We describe here the analysis of these serpin genes, including comparison with the serpin genes of *B. mori*, detection of alternatively spliced isoforms, structure-function predictions, and mRNA expression profiles.

2. Materials and methods

2.1. Insects

M. sexta eggs originally obtained from Carolina Biological Supply were used to establish a laboratory colony, which has been maintained by feeding larvae on an artificial diet as described by Dunn and Drake (1983).

2.2. Annotation of *M. sexta* serpin genes

Amino acid sequences of previously identified serpins from *M. sexta* (serpins 1–7, 9, and 13) and *B. mori* (serpins 1–34) (Zou *et al.*, 2009) were used as queries to search the *M. sexta*

genome assembly 1.0 (<http://agripestbase.org/manduca/>) using the TBLASTN algorithm with default settings. Amino acid sequences of the newly identified serpins were downloaded from ManducaBase and improved by comparing with *Manduca* RNA-seq Oases and Trinity Assemblies 3.0 (Cao and Jiang, 2015). The corresponding serpin genes were corrected and annotated using WebApollo as part of the development of *M. sexta* Official Gene Set 2.0 (Kanost *et al.*, 2016).

2.2. Identification and cloning of two novel serpin-1 isoforms

Amino acid sequences for the 12 previously known serpin-1 RCL variants (Jiang and Kanost, 1997) were used as TBLASTN queries to search against *M. sexta* Genome Sequence Assembly 1.0 Scaffold database (June, 2011) and as BLASTP queries against *M. sexta* Protein database (December 2011, [www.agripestbase.org](http://agripestbase.org)). Nucleotide sequences for each putative serpin-1 isoform were located in scaffold 00761. This analysis revealed the presence of two putative variants of serpin-1 exon 9 that were not previously detected (named exon 9N and exon 9Y). To verify that these splicing variants are expressed, primers were designed to amplify each putative isoform from larval fat body cDNA by PCR using a forward primer in exon 8 (5' GAAGTCAACGAAGAAGG-3') and a reverse primer specific to each exon 9 (serpin-1N: 5' CGTTTAATCAAATCTCAGG-3' or serpin-1Y: 5'-AGATCACGATGGCATAAAT-3'). PCR was performed as follows using Taq DNA polymerase (Invitrogen): initial denaturation at 94°C for 5 min followed by 30 cycles of denaturing at 94°C for 30 s, annealing at 46°C for 45 s, extension at 72°C for 30 s, and a final extension at 72°C for 5 min. Products were cloned into the pCR4®-TOPO® vector using the TOPO TA Cloning® Kit (Invitrogen), used to transform One Shot® TOP10 *Escherichia coli*, and the inserts in the resulting plasmids were sequenced. The complete coding regions for serpin-1N and -1Y, minus signal peptide, were amplified from fat body cDNA by PCR using a forward primer (5'-CCATGGCCGCGAGACGGATCT-3') in which a *Nco*I site (underlined) was introduced, and a reverse primer (serpin-1N: 5'-GCGGCCGCGAATCAAAATTACCGTTTA-3' or serpin-1Y: 5'-GCGGCCGCGAGATCACGATGGCATAAAT-3'), in which a *Not*I site (underlined) was introduced. PCR conditions were denaturation at 94°C for 2 min, followed by 28 cycles of denaturing at 94°C for 30 s, annealing at 50°C (47°C for serpin-1Y) for 30 s, extension at 68°C for 1.5 min, and a final extension at 68°C for 5 min, using High Fidelity Platinum Taq DNA Polymerase (Invitrogen). PCR products were visualized, expected serpin-1N or serpin-1Y DNA bands were gel purified, and cloned as described above. DNA sequences were obtained on an Applied Biosystems 3730xl DNA Analyzer at the Iowa State University Sequencing Facility (Ames, IA).

2.3. Analysis of predicted protein sequences

Theoretical molecular weight and pI of mature serpins were calculated using ProtParam (<http://us.expasy.org/>) (Gasteiger *et al.*, 2005). Signal peptides were predicted using SignalP 4.1 (<http://www.cbs.dtu.dk/services/SignalP/>) (Petersen *et al.*, 2011). Multiple sequence alignments were performed using ClustalW2 (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>) (Larkin *et al.*, 2007). Prediction of secondary structures in the aligned sequences was based on the crystal structure of *M. sexta* serpin-1K (Li *et al.*, 1999) and human α -1 antitrypsin

(Irving *et al.*, 2000). Conserved residues in the shutter, breach, hinge, gate and other regions were identified as described previously (Irving *et al.*, 2000).

2.4. Phylogenetic analysis

A multiple sequence alignment of all *M. sexta* and *B. mori* serpins was performed using ClustalW2 with the default settings. RCL regions were retrieved from the full-length alignment, followed by manual adjustment using MEGA6 and coloring with ClustalX2 in default parameters (Gulley *et al.*, 2013). Phylogenetic analyses were conducted using MrBayes v3.2.6 (Ronquist *et al.*, 2012) under the default model with the setting “nchains=12”. Markov chain Monte Carlo (MCMC) analyses were terminated after the standard deviations of two independent analyses were <0.01. FigTree 1.4.3 (<http://tree.bio.ed.ac.uk/software/figtree/>) was used to display the phylogenetic trees. The sequences used for analysis were: *M. sexta* MsSerp1-1A (AAC47342.1), MsSerp1-2 (AAB58491.1), MsSerp1-3 (AAO21505.1), MsSerp1-4 (AAS68503.1), MsSerp1-5 (AAS68507.1), MsSerp1-6 (AAV91026.1), MsSerp1-7 (ADM86478.1), MsSerp1-8 (Msex2.01482), MsSerp1-9 (Msex2.09571), MsSerp1-10 (Msex2.11870), MsSerp1-11 (Msex2.11871), MsSerp1-12 (Msex2.11869), MsSerp1-13 (Msex2.06711), MsSerp1-14 (Msex2.09570), MsSerp1-15A (Msex2.06848-PA), MsSerp1-16 (Msex2.09066), MsSerp1-17 (Msex2.15518), MsSerp1-18 (Msex2.10821), MsSerp1-19 (Msex2.10820), MsSerp1-20 (Msex2.15520), MsSerp1-21 (Msex2.09937), MsSerp1-22 (Msex2.06154, Msex2.06167), MsSerp1-23 (Msex2.01480), MsSerp1-24 (Msex2.01481), MsSerp1-25 (Msex2.10827), MsSerp1-26 (Msex2.10822), MsSerp1-27 (Msex2.10826), MsSerp1-28A (Msex2.10818), MsSerp1-29 (Msex2.09069), MsSerp1-30 (Msex2.10824), MsSerp1-31 (Msex2.10823), MsSerp1-32 (Msex2.05826); *B. mori* aBmSerp1-1 (ACT36276.1), BmSerp1-2 (AAF61252.1), BmSerp1-3 (ABD36254.1), BmSerp1-4 (AAS68505.1), BmSerp1-5 (AAS68506.1), BmSerp1-6 (ABV74209.1), BmSerp1-7 (ACG61172.1), BmSerp1-8 (ACG61173.1), BmSerp1-9 (AAK52495.1|AF361483_1), BmSerp1-10 (ACG61174.1), BmSerp1-11 (ACG61175.1), BmSerp1-12 (BAB33293.1), BmSerp1-13 (ACG61176.1), BmSerp1-14 (ACG61177.1), BmSerp1-15 (ACG61178.1), BmSerp1-16 (ACG61179.1), BmSerp1-17 (ACG61180.1), BmSerp1-18 (ACG61181.1), BmSerp1-19 (ACG61182.1), BmSerp1-20 (ACG61183.1), BmSerp1-21 (ACG61184.1), BmSerp1-22 (ACG61185.1), BmSerp1-23 (ACG61186.1), BmSerp1-24 (ACG61187.1), BmSerp1-25 (ACG61188.1), BmSerp1-26 (ACG61189.1), BmSerp1-27 (ACG61190.1), BmSerp1-28 (ACG61191.1), BmSerp1-29 (ACG61192.1), BmSerp1-30 (ACG61193.1), BmSerp1-31 (ACG61194.1), BmSerp1-32 (ACG61195.1), BmSerp1-33 (ACI24664.1), BmSerp1-34 (ACI24665.1).

2.8. Serpin gene expression

The 52 cDNA libraries, representing mRNA samples from whole insects, organs or tissues at various life stages, were constructed and sequenced by Illumina technology (Kanost *et al.*, 2016). Reads from the individual RNA-seq datasets were trimmed to 50 bp and mapped to the updated OGS1.0 using Bowtie (0.12.8) (Langmead *et al.*, 2009). Numbers of the mapped reads were used to calculate FPKM (fragments per kilobase of exon per million fragments mapped) by RSEM (1.2.12) (Li and Dewey, 2011) for interlibrary comparisons. Hierarchical

clustering of the $\log_2(\text{FPKM}+1)$ values was performed using MultiExperiment Viewer (v4.9) with the Pearson correlation-based metric and average linkage clustering method.

3. Results and Discussion

3.1 Overview of *M. sexta* serpin genes

We identified and annotated 32 serpin genes in the *M. sexta* genome (Table 1, S1). We attempted, when possible, to name the *M. sexta* serpins with numbers corresponding to *B. mori* orthologs (Zou *et al.*, 2009), based on the phylogenetic analysis described below. *MsSerp1-29* lacks a significant part of the typically conserved amino-terminal region, perhaps due to incomplete genome assembly. Most of the serpin genes are located on a single genomic scaffold, but *MsSerp1-3* and *MsSerp1-7* were fragmented, with exons on more than one scaffold. Their sequences were confirmed by the published cDNA sequences. *MsSerp1-22* had two apparent haplotypes on different scaffolds. Several genomic scaffolds contain multiple serpin genes within a relatively small region (Fig. 1), including 8 serpin genes located within a 20 kb segment on scaffold 00379. The *M. sexta* serpin genes vary from one exon to ten (Table 1 and Fig. 2), with most composed of 8–10 exons. *MsSerp1-4*, *5*, *7*, *9*, and *14* contain a single exon.

3.2 Alternative splicing of *M. sexta* serpins

The *MsSerp1-1* gene is composed of 10 exons, with the ninth exon encoding the carboxyl-terminal 35–40 amino acid residues, including the RCL. We previously reported that mutually exclusive alternate splicing of 12 different versions of exon 9 results in production of multiple serpin-1 isoforms, each with a different RCL sequence and inhibitory spectrum (Jiang *et al.*, 1994; 1997; Ragan *et al.*, 2010). In re-examining the new genomic sequence of the *MsSerp1-1* gene, we found two previously unidentified copies of exon 9, which were named 9N and 9Y (Fig. 3). We used PCR to amplify cDNAs for exons 1–9 of the corresponding serpin-1N and serpin-1Y variants, from fat body mRNA. Cloning and sequencing of these cDNAs (*MsSerp1Y*, Genbank accession MG753798, *Serp1N*, accession MG753799) verified the expression of these two alternatively spliced serpin-1 variants, resulting in a total of 14 alternatively spliced serpin-1 isoforms, each with a different RCL sequence. We found two additional *M. sexta* serpin genes, *MsSerp1-15* and *MsSerp1-28*, which have alternative exons encoding the carboxyl terminal ~45 residues, including the RCL (Fig. 3). *MsSerp1-15* encodes two serpin splicing isoforms with different exon 9 sequences encoding the RCL. A closely related *B. mori* gene, *BmSerp1-2*, has a similar gene structure and also encodes two splicing isoforms with different RCL sequences (Zou *et al.*, 2009). The *MsSerp1-28* gene contains three versions of exon 7, each encoding a different RCL sequence.

Lepidopteran orthologs of *M. sexta* serpin-1 also have been found to generate variants by alternative exon splicing at the same position. *BmSerp1-1* encodes 4 variants through alternative exon 9 usage (Zou *et al.*, 2009; Liu *et al.*, 2011), *Mamestra configurata serpin-1* gene encodes 9 alternative variants of exon 9 (Hegedus *et al.*, 2008), and *Choristoneura fumiferana* has at least 4 splicing variants (Zheng *et al.*, 2009). In dipterans, the same type of alternative splicing to produce serpins with different RCLs occurs in *Anopheles gambiae*

SRPN10 (Danielli *et al.*, 2003) and in *spn4* in several *Drosophila* species (Börner and Ragg, 2008).

3.3. Analysis and comparison of serpin amino acid sequences

A majority of the *M. sexta* serpin gene products are predicted to be secreted, with amino terminal secretion signal peptides varying from 15 to 24 residues (Table 1). MsSerpins-2, 15, 16, and 17 lack a detected signal peptide and thus may be cytoplasmic proteins. Alignment of the amino acid sequences (Fig. S1) indicates that although overall sequence identity between some serpins is fairly low (~35%), there are regions of identity between all of the sequences that correspond with secondary structures conserved in insect and mammalian serpin three dimensional structures. The length of the mature serpins varies from 332 to 502 residues with a theoretical mass of 37.8 to 57.1 kDa (excluding MsSerpins-29, 284 residues). Notably, MsSerpins 10 has a nearly 100-amino acid insertion between helix-F and strand-3A (Fig. S1), resulting in its greater than average length of 502 residues. The *B. mori* ortholog, BmSerpins10, has an insertion of a similar length at the same position (Zou *et al.*, 2009). The structure and potential function of this region is not yet known, but perhaps could be a site for interaction with other proteins. The calculated isoelectric point of most of the *M. sexta* serpins varies from 4.4 to 7.3. However, MsSerpins-12, 14, 20, 21, 22, 26, 29 and 31 are more basic, with a calculated pI of approximately 8–9.

3.4 RCL region of *M. sexta* serpins

The RCL in serpins is a solvent-exposed region between strand-4A and strand-1C, consisting of 16–25 amino acid residues (Meekins *et al.*, 2016). In the inhibition of a proteinase by a serpin, the proteinase cleaves the RCL at a specific position called the scissile bond, between two residues designated P1 and P1' (Loebermann *et al.*, 1984). As a result, the amino terminal segment of the RCL inserts into β -sheet A, forming a new strand between strand-3A and strand-5A, with formation of an irreversible serpin-proteinase complex (Stratikos and Gettins, 1999; Huntington *et al.*, 2000; Dunstone and Whisstock, 2011). Some members of the serpin superfamily, such as ovalbumin, Hsp47/SERPINH1, maspin and MENT, lack the capability for protease inhibition (Huntington and Stein, 2001; Nagata, 1996; Zou *et al.*, 1994; Grigoryev *et al.*, 1999). The sequence and length of RCL of such non-inhibitory serpins are factors that distinguish them from inhibitory serpins (Irving *et al.*, 2000). In non-inhibitory serpins, either the length of the RCL limits the complete translocation of the protease, or the critical residues in the hinge region are not favorable for rapid loop insertion (Widmer *et al.*, 2012; Al-Ayyoubi *et al.*, 2004; Stein *et al.*, 1991; McGowan *et al.*, 2006).

In order to predict inhibitory or non-inhibitory *M. sexta* serpins, we analyzed their RCL sequences (Fig. 4). The hinge region (P15-P9) of inhibitory serpins is typically composed of two charged residues, followed by a series of residues with small hydrophobic side chains, which contribute to mobility for the conformational change accompanying RCL insertion into β -sheet A (Hopkins *et al.* 1993). A consensus pattern in the hinge region can be used as a theoretical guideline to recognize inhibitory serpins. P17 is usually glutamate, P16 is glutamate/lysine/arginine, P15 is glycine, P14 is threonine/serine/alanine, and P12-P9 are alanine/glycine/serine (Irving *et al.*, 2000). Additionally, the presence of proline in the

proximal RCL prior to the P2 residue disrupts the formation of a new β strand, leading to failure of RCL incorporation (Hopkins *et al.*, 1993; Hopkins and Stone, 1995; Gettins, 2002). Furthermore, conserved residues in the breach and shutter regions, which facilitate sheet opening to accept the RCL as a new strand in the β -sheet should also be considered (Whisstock *et al.*, 2000a; Whisstock *et al.*, 2000b). Guided by the above criteria, 16 of the 32 serpin genes (*MsSerp*-8, 10, 18–20, 22–27, 29–32) are predicted to encode non-inhibitory serpins. They lack the consensus pattern of inhibitory serpins in the hinge region, and these putative non-inhibitors are relatively rich in proline in the proximal RCL and less conserved in the breach and shutter regions (Fig.S1). The other 16 genes encode proteins more likely to have a mobile hinge region, few proline residues in the proximal RCL, and highly conserved residues in the breach and shutter regions and thus are more likely to function as proteinase inhibitors.

We have previously determined experimentally the scissile bond position for some inhibitory *M. sexta* serpins (Jiang and Kanost, 1997; Zou and Jiang, 2005; He *et al.*, 2017). For the remaining serpins likely to be inhibitory, we have predicted the position of their RCL scissile bond based on the distance from the hinge and previously confirmed cleavage sites of studied serpins (Gettins, 2002; Jiang and Kanost, 1997; Zou and Jiang, 2005; Reichhart *et al.*, 2011). Based on putative P1 residues, *M. sexta* serpins can be categorized into three groups: 1) serpins with K or R at P1, including *MsSerp*-1A, 1E, 1J, 1N, 1Y, 3, 4, 5, 6, 7, 9, 13, 15B, 16, 17, and 28B; 2) serpins with Y or F at P1, including *MsSerp*-1C, 1H, 1K, 1Z, 15A, and 21; and 3) serpins with small hydrophobic residues at P1, including *MsSerp*-1B, 1F, 1G, 1I, 2, 11, 12, 28A, and 28C. Distinguishing inhibitory and non-inhibitory serpins and characterizing inhibitory selectivity will require future biochemical experiments using purified serpins.

3.5 Phylogenetic analysis of *M. sexta* and *B. mori* serpins

To investigate the evolution of the serpin gene family in *M. sexta* and in *B. mori*, a lepidopteran in the same superfamily (Bombycoidea), we analyzed phylogenetic trees based on alignment of their amino acid sequences. Although these two species contain almost the same number of serpin genes (32 in *M. sexta* and 34 in *B. mori*), there are only 17 orthologous pairs of serpin genes in these species (Fig. 5). These include the serpins (1, 3–9, and 13) that have been characterized as hemolymph proteins in *M. sexta* and shown to regulate protease cascades in immune responses (Jiang and Kanost, 1997; Jiang *et al.*, 2003; Zhu *et al.*, 2003; Tong and Kanost, 2005; Tong *et al.*, 2005; Zou and Jiang, 2005; Suwanchaichinda *et al.*, 2013; He *et al.*, 2017). It appears that these serpin orthologs, existing before divergence of the two species, may have retained essential functions, including regulation of immune responses. There was also apparent expansion of some groups of serpin genes after divergence of the lineages leading to *Manduca* and *Bombyx* (Fig. 5). Most of the *M. sexta* serpins that lack one-to-one orthology with *B. mori* are predicted to be non-inhibitory serpins (Fig. 5, Table 1).

Intron positions are often conserved in closely related genes, including serpin genes (Roy *et al.*, 2003; Kaiserman and Bird, 2005). Based on the number of exons and intron positions, we can categorize *M. sexta* serpins into the following groups (Fig. 2): **I**) serpin genes with a

single exon (*MsSerp*-4, 5, 7, 9 and 14); **II**) serpin genes with 2–3 coding exons and one intron position that is conserved (*MsSerp*-8, 23, 20, 32 and 24); **III**) serpin genes with 6–8 coding exons and 2–3 intron positions conserved (*MsSerp*-3, 10, 11, 12); **IV**) serpins genes with 7–10 coding exons and most intron positions are conserved (*MsSerp*-16, 29, 22, 21, 26, 31, 30, 18, 19, 25, 27, 28, 1, 2, 17, 15). Serpin genes 6 and 13 contain multiple introns but none at conserved positions. Genes in groups **I** and **II** may be derived from a single retrotransposition event followed by several rounds of gene duplication. One series of duplications may have led to *MsSerp*-4, 5, 7, 9, and 14, while a second series of duplications occurring after gaining an intron led to group **II** genes *MsSerp*-8, 20, 23, 24, and 32. A similar process of serpin gene evolution was also described in *B. mori* (Zou *et al.*, 2009). It is interesting to note that 13 of the 16 genes in groups **I**, **II**, and **III** have orthologs in *B. mori*, thus these genes evolved before the divergence of these two lineages. In contrast, 12 of the 16 genes in group **IV** do not have an ortholog in *B. mori*, indicating that either those genes arose after the divergence or that they were lost in *B. mori*. It is intriguing that of the 13 serpin genes that appear to have evolved in *M. sexta* after its divergence from the common ancestor with *B. mori*, 10 encode proteins that are predicted to be non-inhibitory. Of the 17 orthologous gene pairs between *M. sexta* and *B. mori*, 14 are predicted to encode protease inhibitors, while only three are non-inhibitory. Thus, it appears that most of the serpin proteins present in *M. sexta* but not in *B. mori* have a function other than protease inhibition.

3.6 Expression profiles of *M. sexta* serpin transcripts

The transcription patterns of *M. sexta* serpin genes were examined by analyzing the RNA-seq data from 52 tissue samples (Kanost *et al.*, 2016) at different life stages (Table S2). Four gene groups were recognized by cluster analysis based on distinct expression profiles (Fig. 6). Group A includes *MsSerp*-1 and 15, whose transcripts are generally abundant in every stage and tissue. Both of these genes include splicing variants encoding 14 (*MsSerp*-1) or 2 (*MsSerp*-15) versions of the RCL (expression of individual splicing variants is not analyzed here). *MsSerp*-1 was previously known to be highly expressed in larval fat body and to a lower degree in hemocytes, with all splicing isoforms present (Kanost *et al.*, 1995; Ragan *et al.*, 2010). *MsSerp*-15 transcript was particularly abundant in midgut.

Serpin genes of group B, including *MsSerp*-8, 18, 19, 21–28, 30, 31 and 32, are barely transcribed in most stages and tissues examined. These genes are mostly predicted non-inhibitory serpins (exceptions are *MsSerp*-21 and 28) (Table 1). However, non-inhibitory *MsSerp*-25, 26, 27 and 31 are notably expressed in late fifth instar and wandering stage fat body, Malpighian tubules, and muscle, suggesting a potential function in these tissues in preparation for pupation. Transcripts of *MsSerp*-18 were most abundant in feeding stage larvae. *MsSerp*-30 had highest expression in larval muscle. Transcripts of *MsSerp*-8 were quite low in most tissues at most life stages, but were present at a high level in 3h eggs (Fig. 6), indicating a potential role in early development. Transcripts of *A. gambiae* SRPN13 are abundant in eggs and very young larvae, but hardly present in subsequent life stages (Suwanchaichinda and Kanost, 2009). In birds, the non-inhibitory serpin ovalbumin and related serpins are major components of egg white (Dombre *et al.*, 2017), thus there may be some conserved function for non-inhibitory serpins in eggs.

Group C consists of 7 *M. sexta* serpin genes, including *MsSerp**in-3*, 4, 5, 6, 9, 12 and 13, whose transcripts reach a peak during the wandering and early pupal stages in fat body. *MsSerp**in-4* and 5 are also highly transcribed in midgut of day 1 pupae. Additionally, mRNA levels of group C serpin genes are moderately abundant in last larval instar in muscle. Group D is comprised of *MsSerp**in-2*, 7, 10, 11, 14, 16, 17, 20, and 29. Head and muscle are principal tissues where expression of group D serpin genes was observed, and some with high expression in fat body of early pupae.

We re-examined the data from our previous transcriptomic studies to examine expression of the serpin genes in larval hemocytes and fat body in response to an immune challenge (Zhang *et al.*, 2011; Gunaratna and Jiang, 2013). Transcripts of *MsSerp**in-1-7*, 9, 11–15, 17–19, and 22 were detected in these studies, and the change in their mRNA level in response to injection of a mixture of *E. coli*, *Micrococcus luteus*, and curdolan was evaluated (Table S3). *MsSerp**in-1* and 17 exhibited >2-fold increase in mRNA level in hemocytes, while not changing in fat body. In contrast, *MsSerp**in-4*, 7, 14, 18 and 22 displayed significant up-regulation of transcript levels in fat body, but not hemocytes after immune induction. mRNA levels of *MsSerp**in-2*, 3, 5, 6, 11, 12, 15 and 17 increased more than 2-fold in both fat body and hemocytes after injection of bacteria. These data, are consistent with previous northern blot and RT-PCR results for *MsSerp**in-3*, 4, 5, 6 and 7 (Zhu *et al.*, 2003; Tong and Kanost, 2005; Zou and Jiang, 2005; Suwanchaichinda *et al.*, 2013). *MsSerp**in-19* is the only serpin gene whose transcript level decreased after injection of immune elicitors. A decrease in serpin expression has been reported in multiple lepidopteran insects in response to parasitoid wasps, which interfere with serpin levels by targeting transcripts with miRNA, such as miR-8 (Beckage and Kanost, 1993; Song *et al.*, 2008; Etebari and Asgari, 2013; Lee and Hyun, 2014).

Most previous studies have focused on serpins that are expressed in larval fat body and secreted into hemolymph, with functions in regulating immune responses. These new observations indicate more diverse tissue and developmental patterns for members of the serpin family in *M. sexta*, suggesting that serpins are likely to have additional undiscovered functions in insect development and physiology.

4. Conclusion

We identified 32 serpin genes in the *M. sexta* genome. The new sequence data and analysis revealed two previously undiscovered splicing isoforms of *MsSerp**in-1*, for a total of 14 serpin variants from one gene, with potential to inhibit different proteases. Similar alternatively spliced variants were also found in *MsSerp**in-15* and 28, with 2 and 3 isoforms respectively. Phylogenetic analysis and comparison with *B. mori* serpins revealed 17 pairs of orthologous genes in these two species, with likely conserved biological functions. An analysis of conserved amino acid residues in the gate, shutter and breach regions of serpins and the pattern of residues in the RCL hinge region suggested that approximately three quarters of *M. sexta* serpin proteins (including splicing isoforms) are functional proteinase inhibitors. These predictions require future experimental testing. The biological functions of non-inhibitory insect serpins are unknown and an area in need of substantial research. Analysis of RNAseq data from many tissues and developmental stages detected stage-

specific and tissue-specific expression patterns for most serpins, which may help to guide future experiments to investigate their biological functions. We hope these studies will promote further research regarding the evolution, structure and function of serpins in *M. sexta* and other insects.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Lisa Brummett for maintaining the *M. sexta* colony. This work was supported by NIH grants GM041247 (to MK) and GM58634 and AI112662 (to H. Jiang). This is publication 18–274-J from the Kansas Agricultural Experiment Station.

References

- Abraham EG, Pinto SB, Ghosh A, Vanlandingham DL, Budd A, Higgs S, Kafatos FC, Jacobs-Lorena M, Michel K, 2005 An immune-responsive serpin, SRPN6, mediates mosquito defense against malaria parasites. *Proc. Natl. Acad. Sci. U. S. A* 102, 16327–16332. [PubMed: 16260729]
- Ahmad ST, Sweeney ST, Lee J, Sweeney NT, Gao F, 2009 Genetic screen identifies serpin5 as a regulator of the toll pathway and CHMP2B toxicity associated with frontotemporal dementia. *Proc. Natl. Acad. Sci. U. S. A* 106, 12168–12173. [PubMed: 19581577]
- Al-Ayyoubi M, Gettins PG, Volz K, 2004 Crystal structure of human maspin, a serpin with antitumor properties: reactive center loop of maspin is exposed but constrained. *J. Biol. Chem* 279, 55540–55544. [PubMed: 15501821]
- An C, Budd A, Kanost MR, Michel K, 2011 Characterization of a regulatory unit that controls melanization and affects longevity of mosquitoes. *Cellular and Molecular Life Sciences* 68, 1929–1939. [PubMed: 20953892]
- An C, Kanost MR, 2010 *Manduca sexta* serpin-5 regulates prophenoloxidase activation and the Toll signaling pathway by inhibiting hemolymph proteinase HP6. *Insect Biochem. Mol. Biol* 40, 683–689. [PubMed: 20624461]
- An C, Ragan EJ, Kanost MR, 2011 Serpin-1 splicing isoform J inhibits the proSpatzle activating proteinase HP8 to regulate expression of antimicrobial hemolymph proteins in *Manduca sexta*. *Dev. Comp. Immunol* 35, 135–141. [PubMed: 20851714]
- Beckage NE, Kanost MR, 1993 Effects of parasitism by the braconid wasp *Cotesia congregata* on host hemolymph proteins of the tobacco hornworm, *Manduca sexta*. *Insect Biochem. Mol. Biol* 23, 643–653. [PubMed: 8353522]
- Börner S, Ragg H, 2008 Functional diversification of a protease inhibitor gene in the genus *Drosophila* and its molecular basis. *Gene* 415, 23–31. [PubMed: 18395367]
- Cao X, Jiang H, 2015 Integrated modeling of protein-coding genes in the *Manduca sexta* genome using RNA-Seq data from the biochemical model insect. *Insect Biochem. Mol. Biol* 62, 2–10. [PubMed: 25612938]
- Christophides GK, Zdobnov E, Barillas-Mury C, Birney E, Blandin S, Blass C, Brey PT, Collins FH, Danielli A, Dimopoulos G, Hetru C, Hoa NT, Hoffmann JA, Kanzok SM, Letunic I, Levashina EA, Loukeris TG, Lycett G, Meister S, Michel K, Moita LF, Muller HM, Osta MA, Paskewitz SM, Reichhart JM, Rzhetsky A, Troxler L, Vernick KD, Vlachou D, Volz J, von Mering C, Xu J, Zheng L, Bork P, Kafatos FC, 2002 Immunity-related genes and gene families in *Anopheles gambiae*. *Science* 298, 159–165. [PubMed: 12364793]
- Danielli A, Kafatos FC, Loukeris TG, 2003 Cloning and characterization of four *Anopheles gambiae* serpin isoforms, differentially induced in the midgut by *Plasmodium berghei* invasion. *J. Biol. Chem* 278, 4184–4193. [PubMed: 12456678]

- De Gregorio E, Han S, Lee W, Baek M, Osaki T, Kawabata S, Lee B, Iwanaga S, Lemaitre B, Brey PT, 2002 An immune-responsive Serpin regulates the melanization cascade in *Drosophila*. *Developmental cell* 3, 581–592. [PubMed: 12408809]
- Dombre C, Guyot N, Moreau T, Monget P, Da Silva M, Gautron J, Réhault-Godbert S 2017 Egg serpins: The chicken and/or the egg dilemma. *Semin Cell Dev Biol.* 62,120–132. [PubMed: 27565683]
- Dunn PE, Drake DR, 1983 Fate of bacteria injected into naive and immunized larvae of the tobacco hornworm *Manduca sexta*. *J. Invertebr. Pathol* 41, 77–85.
- Dunstone MA, Whisstock JC, 2011 Crystallography of Serpins and Serpin Complexes. *Methods in Enzymology*, Vol 501: Serpin Structure and Evolution 501, 63–87.
- Etebari K, Asgari S, 2013 Conserved microRNA miR-8 blocks activation of the Toll pathway by upregulating Serpin 27 transcripts. *RNA Biology* 10, 1356–1364. [PubMed: 23806890]
- Gan H, Wang Y, Jiang H, Mita K, Kanost MR, 2001 A bacteria-induced, intracellular serpin in granular hemocytes of *Manduca sexta*. *Insect Biochem. Mol. Biol* 31, 887–898. [PubMed: 11439248]
- Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A, 2005 Protein identification and analysis tools on the ExPASy server In: Walker JM (Ed), *The Proteomics Protocols Handbook*. Humana Press, pp571–607.
- Gettins PG, 2002 Serpin structure, mechanism, and function. *Chem. Rev* 102, 4751–4804. [PubMed: 12475206]
- Gettins PG, Olson ST, 2016 Inhibitory serpins. New insights into their folding, polymerization, regulation and clearance. *Biochem. J* 473, 2273–2293. [PubMed: 27470592]
- Grigoryev SA, Bednar J, Woodcock CL, 1999 MENT, a heterochromatin protein that mediates higher order chromatin folding, is a new serpin family member. *J. Biol. Chem* 274, 5626–5636. [PubMed: 10026180]
- Gulley MM, Zhang X, Michel K, 2013 The roles of serpins in mosquito immunology and physiology. *J. Insect Physiol* 59, 138–147. [PubMed: 22960307]
- Gunaratna RT, Jiang H, 2013 A comprehensive analysis of the *Manduca sexta* immunotranscriptome. *Dev. Comp. Immunol* 39, 388–398. [PubMed: 23178408]
- He Y, Wang Y, Zhao P, Rayaprolu S, Wang X, Cao X, Jiang H, 2017 Serpin-9 and -13 regulate hemolymph proteases during immune responses of *Manduca sexta*. *Insect Biochem. Mol. Biol* 90, 71–81. [PubMed: 28987647]
- Hegedus DD, Erlandson M, Baldwin D, Hou X, Chamankhah M, 2008 Differential expansion and evolution of the exon family encoding the Serpin-1 reactive centre loop has resulted in divergent serpin repertoires among the Lepidoptera. *Gene* 418, 15–21. [PubMed: 18495381]
- Hopkins PC, Carrell RW, Stone SR, 1993 Effects of mutations in the hinge region of serpins. *Biochemistry (N. Y.)* 32, 7650–7657.
- Hopkins PC, Stone SR, 1995 The Contribution of the Conserved Hinge Region Residues of α_1 -Antitrypsin to Its Reaction with Elastase. *Biochemistry* 34, 15872–15879. [PubMed: 7495819]
- Huntington JA, Stein PE, 2001 Structure and properties of ovalbumin. *Journal of Chromatography B: Biomedical Sciences and Applications* 756, 189–198. [PubMed: 11419711]
- Huntington JA, McCoy A, Belzar KJ, Pei XY, Gettins PG, Carrell RW, 2000 The conformational activation of antithrombin. A 2.85-Å structure of a fluorescein derivative reveals an electrostatic link between the hinge and heparin binding regions. *J. Biol. Chem* 275, 15377–15383. [PubMed: 10809774]
- Irving JA, Pike RN, Lesk AM, Whisstock JC, 2000 Phylogeny of the serpin superfamily: implications of patterns of amino acid conservation for structure and function. *Genome Res.* 10, 1845–1864. [PubMed: 11116082]
- Jiang H, Kanost MR, 1997 Characterization and functional analysis of 12 naturally occurring reactive site variants of serpin-1 from *Manduca sexta*. *J. Biol. Chem* 272, 1082–1087. [PubMed: 8995406]
- Jiang H, Wang Y, Yu X, Zhu Y, Kanost M, 2003 Prophenoloxidase-activating proteinase 3 (PAP-3) from *Manduca sexta* hemolymph: a clip-domain serine proteinase regulated by serpin 1J and serine proteinase homologs. *Insect Biochem. Mol. Biol* 33, 1049–1060. [PubMed: 14505699]

- Jiang H, Wang Y, Kanost MR, 1994 Mutually exclusive exon use and reactive center diversity in insect serpins. *J. Biol. Chem* 269, 55–58. [PubMed: 8276850]
- Jiang R, Kim E, Gong J, Kwon H, Kim C, Ryu K, Park J, Kurokawa K, Zhang J, Gubb D, Lee B, 2009 Three Pairs of Protease-Serpin Complexes Cooperatively Regulate the Insect Innate Immune Responses. *J. Biol. Chem* 284, 35652–35658. [PubMed: 19858208]
- Kaiserman D, Bird PI, 2005 Analysis of vertebrate genomes suggests a new model for clade B serpin evolution. *BMC Genomics* 6, 167. [PubMed: 16305753]
- Kanost MR, Prasad SV, Wells MA, 1989 Primary structure of a member of the serpin superfamily of proteinase inhibitors from an insect, *Manduca sexta*. *J. Biol. Chem* 264, 965–972. [PubMed: 2463253]
- Kanost MR, Arrese EL, Cao X, Chen Y, Chellapilla S, Goldsmith MR, Grosse-Wilde E, Heckel DG, Herndon N, Jiang H, Papanicolaou A, Qu J, Soulages JL, Vogel H, Walters J, Waterhouse RM, Ahn S, Almeida FC, An C, Aqrawi P, Bretschneider A, Bryant WB, Bucks S, Chao H, Chevignon G, Christen JM, Clarke DF, Dittmer NT, Ferguson LCF, Garavelou S, Gordon KHJ, Gunaratna RT, Han Y, Hauser F, He Y, Heidel-Fischer H, Hirsh A, Hu Y, Jiang H, Kalra D, Klinner C, Konig C, Kovar C, Kroll AR, Kuwar SS, Lee SL, Lehman R, Li K, Li Z, Liang H, Lovelace S, Lu Z, Mansfield JH, McCulloch KJ, Mathew T, Morton B, Muzny DM, McCulloch KJ, Mathew T, Morton B, Muzny DM, Neunemann D, Onger F, Pauchet Y, Pu L, Pyrousis I, Rao X, Redding A, Roesel C, Sanchez-Gracia A, Schaack S, Shukla A, Tetreau G, Wang Y, Xiong G, Traut W, Walsh TK, Worley KC, Wu D, Wu W, Wu Y, Zhang X, Zou Z, Zucker H, Briscoe AD, Burmester T, Clem RJ, Feyereisen R, Grimmelikhuijzen CJP, Hamodrakas SJ, Hansson BS, Huguet E, Jermini LS, Lan Q, Lehman HK, Lorenzen M, Merzendorfer H, Michalopoulos I, Morton DB, Muthukrishnan S, Oakeshott JG, Palmer W, Park Y, Passarelli AL, Rozas J, Schwartz LM, Smith W, Southgate A, Vilcinskis A, Vogt R, Wang P, Werren J, Yu X, Zhou J, Brown SJ, Scherer SE, Richards S, Blissard GW, 2016 Multifaceted biological insights from a draft genome sequence of the tobacco hornworm moth, *Manduca sexta*. *Insect Biochem. Mol. Biol* 76, 118–147. [PubMed: 27522922]
- Kanost MR, Prasad SV, Huang Y, Willott E, 1995 Regulation of serpin gene-1 in *Manduca sexta*. *Insect Biochem Mol Biol*. 25, 285–291. [PubMed: 7711756]
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG, 2007 Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948. [PubMed: 17846036]
- Langmead B, Trapnell C, Pop M, Salzberg SL, 2009 Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 10, R25. [PubMed: 19261174]
- Lee GJ, Hyun S, 2014 Multiple targets of the microRNA miR-8 contribute to immune homeostasis in *Drosophila*. *Dev. Comp. Immunol* 45, 245–251. [PubMed: 24694685]
- Levashina EA, Langley E, Green C, Gubb D, Ashburner M, Hoffmann JA, Reichhart JM, 1999 Constitutive activation of toll-mediated antifungal defense in serpin-deficient *Drosophila*. *Science* 285, 1917–1919. [PubMed: 10489372]
- Li J, Wang Z, Canagarajah B, Jiang H, Kanost M, Goldsmith EJ, 1999 The structure of active serpin 1K from *Manduca sexta*. *Structure* 7, 103–109. [PubMed: 10368276]
- Li B, Dewey CN, 2011 RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12, 323. [PubMed: 21816040]
- Ligoxygakis P, Pelte N, Ji C, Leclerc V, Duvic B, Belvin M, Jiang H, Hoffmann JA, Reichhart JM, 2002 A serpin mutant links Toll activation to melanization in the host defence of *Drosophila*. *EMBO J*. 21, 6330–6337. [PubMed: 12456640]
- Liu H, Li Y, Jia R, Cui W, Mu Z, Zhang Z, 2011 Alternative splicing of the antitrypsin gene in the silkworm, *Bombyx mori*. *Mol. Biol. Rep* 38, 2793–2799. [PubMed: 21104446]
- Loebermann H, Tokunaka R, Deisenhofer J, Huber R, 1984 Human α_1 -proteinase inhibitor: crystal structure analysis of two crystal modifications, molecular model and preliminary analysis of the implications for function. *J. Mol. Biol* 177, 531–557. [PubMed: 6332197]
- McGowan S, Buckle AM, Irving JA, Ong PC, Bashtannyk-Puhlovich TA, Kan W, Henderson KN, Bulynko YA, Popova EY, Smith AI, Bottomley SP, Rossjohn J, Grigoryev SA, Pike RN, Whisstock JC, 2006 X-ray crystal structure of MENT: evidence for functional loop-sheet polymers in chromatin condensation. *EMBO J*. 25, 3144–3155. [PubMed: 16810322]

- Meekins DA, Kanost MR, Michel K, 2017 Serpins in arthropod biology. 62, 105–119.
- Meekins DA, Zhang X, Battaile KP, Lovell S, Michel K, 2016 1.45 angstrom resolution structure of SRPN18 from the malaria vector *Anopheles gambiae*. *Acta Crystallographica Section F-Structural Biology Communications* 72, 853–862.
- Michel K, Suwanchaichinda C, Morlais I, Lambrechts L, Cohuet A, Awono-Ambene PH, Simard F, Fontenille D, Kanost MR, Kafatos FC, 2006 Increased melanizing activity in *Anopheles gambiae* does not affect development of *Plasmodium falciparum*. *Proc. Natl. Acad. Sci. U. S. A* 103, 16858–16863. [PubMed: 17065316]
- Nagata K, 1996 Hsp47: A collagen-specific molecular chaperone. *Trends Biochem. Sci* 21, 23–26.
- Park SH, Jiang R, Piao S, Zhang B, Kim E, Kwon H, Jin XL, Lee BL, Ha N, 2011 Structural and Functional Characterization of a Highly Specific Serpin in the Insect Innate Immunity. *J. Biol. Chem* 286, 1567–1575. [PubMed: 21047786]
- Petersen TN, Brunak S, von Heijne G, Nielsen H, 2011 SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* 8, 785–786. [PubMed: 21959131]
- Ragan EJ, An C, Yang CT, Kanost MR, 2010 Analysis of mutually exclusive alternatively spliced serpin-1 isoforms and identification of serpin-1 proteinase complexes in *Manduca sexta* hemolymph. *J. Biol. Chem* 285, 29642–29650. [PubMed: 20624920]
- Reichhart J, 2005 Tip of another iceberg: *Drosophila* serpins. *Trends Cell Biol.* 15, 659–665. [PubMed: 16260136]
- Reichhart JM, Gubb D, Leclerc V, 2011 The *Drosophila* Serpins: Multiple Functions in Immunity and Morphogenesis. *Methods in Enzymology: Biology of Serpins* 499, 205–225.
- Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP, 2012 MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol* 61, 539–542. [PubMed: 22357727]
- Roy SW, Fedorov A, Gilbert W, 2003 Large-scale comparison of intron positions in mammalian genes shows intron loss but no gain. *Proc. Natl. Acad. Sci. U. S. A* 100, 7158–7162. [PubMed: 12777620]
- Scherfer C, Tang H, Kambris Z, Lhocine N, Hashimoto C, Lemaitre B, 2008 *Drosophila* Serpin-28D regulates hemolymph phenoloxidase activity and adult pigmentation. *Dev. Biol* 323, 189–196. [PubMed: 18801354]
- Shin SW, Bian G, Raikhel AS, 2006 A toll receptor and a cytokine, Toll5A and Spz1C, are involved in toll antifungal immune signaling in the mosquito *Aedes aegypti*. *J. Biol. Chem* 281, 39388–39395. [PubMed: 17068331]
- Silverman GA, Bird PI, Carrell RW, Church FC, Coughlin PB, Gettins PG, Irving JA, Lomas DA, Luke CJ, Moyer RW, Pemberton PA, Remold-O'Donnell E, Salvesen GS, Travis J, Whisstock JC, 2001 The serpins are an expanding superfamily of structurally similar but functionally diverse proteins. Evolution, mechanism of inhibition, novel functions, and a revised nomenclature. *J. Biol. Chem* 276, 33293–33296. [PubMed: 11435447]
- Song K, Jung M, Eum J, Hwang I, Han SS, 2008 Proteomic analysis of parasitized *Plutella xylostella* larvae plasma. *J. Insect Physiol* 54, 1271–1280.
- Stein PE, Leslie AG, Finch JT, Carrell RW, 1991 Crystal structure of uncleaved ovalbumin at 1.95 Å resolution. *J. Mol. Biol* 221, 941–959. [PubMed: 1942038]
- Stratikos E, Gettins PG, 1999 Formation of the covalent serpin-proteinase complex involves translocation of the proteinase by more than 70 Å and full insertion of the reactive center loop into beta-sheet A. *Proc. Natl. Acad. Sci. U. S. A* 96, 4808–4813. [PubMed: 10220375]
- Suwanchaichinda C, Kanost MR, 2009 The serpin gene family in *Anopheles gambiae*. *Gene* 442, 47–54. [PubMed: 19394412]
- Suwanchaichinda C, Ochieng R, Zhuang S, Kanost MR, 2013 *Manduca sexta* serpin-7, a putative regulator of hemolymph prophenoloxidase activation. *Insect Biochem. Mol. Biol* 43, 555–561. [PubMed: 23567587]
- Tang H, Kambris Z, Lemaitre B, Hashimoto C, 2008 A Serpin that Regulates Immune Melanization in the Respiratory System of *Drosophila*. *Developmental Cell* 15, 617–626. [PubMed: 18854145]

- Tong Y, Jiang H, Kanost MR, 2005 Identification of plasma proteases inhibited by *Manduca sexta* serpin-4 and serpin-5 and their association with components of the prophenol oxidase activation pathway. *J. Biol. Chem* 280, 14932–14942. [PubMed: 15695806]
- Tong Y, Kanost MR, 2005 *Manduca sexta* serpin-4 and serpin-5 inhibit the prophenol oxidase activation pathway: cDNA cloning, protein expression, and characterization. *J. Biol. Chem* 280, 14923–14931. [PubMed: 15695807]
- Wang Y, Jiang H, 2004 Purification and characterization of *Manduca sexta* serpin-6: a serine proteinase inhibitor that selectively inhibits prophenoloxidase-activating proteinase-3. *Insect Biochem. Mol. Biol* 34, 387–395. [PubMed: 15041022]
- Wang Y, Jiang H, 2006 Interaction of beta-1,3-glucan with its recognition protein activates hemolymph proteinase 14, an initiation enzyme of the prophenoloxidase activation system in *Manduca sexta*. *J. Biol. Chem* 281, 9271–9278. [PubMed: 16461344]
- Whisstock JC, Skinner R, Carrell RW, Lesk AM, 2000 Conformational changes in serpins: I. The native and cleaved conformations of α 1-antitrypsin. *J. Mol. Biol* 295, 651–665. [PubMed: 10623554]
- Whisstock JC, Skinner R, Carrell RW, Lesk AM, 2000 Conformational changes in serpins: I. the native and cleaved conformations of α 1-antitrypsin. *J. Mol. Biol* 296, 685–699. [PubMed: 10669617]
- Widmer C, Gebauer JM, Brunstein E, Rosenbaum S, Zaucke F, Droegemueller C, Leeb T, Baumann U, 2012 Molecular basis for the action of the collagen-specific chaperone Hsp47/SERPINH1 and its structure-specific client recognition. *Proc. Natl. Acad. Sci. U. S. A* 109, 13243–13247. [PubMed: 22847422]
- Yang F, Wang Y, Sumathipala N, Cao X, Kanost MR, Jiang H, 2018 *Manduca sexta* serpin-12 controls the prophenoloxidase activation system in larval hemolymph. *Insect Biochem. Mol. Biol* 99, 27–36. [PubMed: 29800677]
- Ye S, Cech AL, Belmares R, Bergstrom RC, Tong Y, Corey DR, Kanost MR, Goldsmith EJ, 2001 The structure of a Michaelis serpin–protease complex. *Nature Structural & Molecular Biology* 8, 979–983.
- Zhang S, Gunaratna RT, Zhang X, Najar F, Wang Y, Roe B, Jiang H, 2011 Pyrosequencing-based expression profiling and identification of differentially regulated genes from *Manduca sexta*, a lepidopteran model insect. *Insect Biochem. Mol. Biol* 41, 733–746. [PubMed: 21641996]
- Zheng Y, He W, Béliveau C, Nisole A, Stewart D, Zheng S, Doucet D, Cusson M, Feng Q, 2009 Cloning, expression and characterization of four serpin-1 cDNA variants from the spruce budworm, *Choristoneura fumiferana*. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* 154, 165–173.
- Zhu Y, Wang Y, Gorman MJ, Jiang H, Kanost MR, 2003 *Manduca sexta* serpin-3 regulates prophenoloxidase activation in response to infection by inhibiting prophenoloxidase activating proteinases. *J. Biol. Chem* 278, 46556–46564. [PubMed: 12966082]
- Zou Z, Jiang H, 2005 *Manduca sexta* serpin-6 regulates immune serine proteinases PAP-3 and HP8. cDNA cloning, protein expression, inhibition kinetics, and function elucidation. *J. Biol. Chem* 280, 14341–14348. [PubMed: 15691825]
- Zou Z, Lopez DL, Kanost MR, Evans JD, Jiang H, 2006 Comparative analysis of serine protease-related genes in the honey bee genome: possible involvement in embryonic development and innate immunity. *Insect Mol. Biol* 15, 603–614. [PubMed: 17069636]
- Zou Z, Evans JD, Lu Z, Zhao P, Williams M, Sumathipala N, Hetru C, Hultmark D, Jiang H, 2007 Comparative genomic analysis of the *Tribolium* immune system. *Genome Biol.* 8, R177. [PubMed: 17727709]
- Zou Z, Picheng Z, Weng H, Mita K, Jiang H, 2009 A comparative analysis of serpin genes in the silkworm genome. *Genomics* 93, 367–375. [PubMed: 19150649]
- Zou Z, Shin SW, Alvarez KS, Kokoza V, Raikhell AS, 2010 Distinct Melanization Pathways in the Mosquito *Aedes aegypti*. *Immunity* 32, 41–53. [PubMed: 20152169]
- Zou Z, Anisowicz A, Hendrix M, Thor A, Neveu M, Sheng S, Rafidi K, Seftor E, Sager R, 1994 Maspin, a Serpin with Tumor-Suppressing Activity in Human Mammary Epithelial-Cells. *Science* 263, 526–529. [PubMed: 8290962]

Highlights

- 32 serpin genes were annotated in the *Manduca sexta* genome.
- 17 of the *M. sexta* serpins have orthologs in *Bombyx mori*.
- Alternative splicing generates serpins with different reactive center sequences.
- Expression profiles of the serpins in different tissues and developmental stages were analyzed.

Fig. 1. *M. sexta* serpin gene clusters. Seven clusters of serpin genes identified on individual genomic scaffolds are shown, with length and orientations identified by red arrows. An overlapping region of oppositely directed MsSerp24 and MsSerp8 is indicated by brown color.

Fig. 2.

Intron positions of *M. sexta* serpin genes. Intron positions are marked on linear maps of the serpin polypeptides, based on alignment of their amino acid sequences (Fig. S1).

Arrowheads of the same color indicate identical intron position. Unconserved intron positions are shown with black arrowheads. Serpins are organized based on the topology of a phylogenetic tree based on their amino acid sequence alignment (Fig. 5). Serpins lacking an intron are in group I, those with one conserved intron position are in group II, with 2–3 conserved intron positions are in group III, with four or more conserved intron positions are in group IV. Serpins 6 and 13 contain multiple introns, but the intron positions are not shared with other serpin genes.

Fig. 3.

Simplified splicing variant diagrams of three *M. sexta* serpin genes. (A) Fourteen spliced isoforms of MsSerpin1 result from the use of alternative versions of exon 9. (B) Alternative exon 9 of MsSerpin15 has two variants, 9A and 9B. (C) Alternate splicing to include exon 7A, 7B or 7C leads to the expression of MsSerpin28A, 28B and 28C respectively. Alternative exons are in color, exons which are always expressed are in grey, and untranslated regions of the last exon of each gene are in white.

Fig. 4.

Alignment of the RCL region of *M. sexta* serpins. The consensus sequence observed in the hinge region of inhibitory serpins is indicated above the alignment. Predicted and known P1 residues are in red. Serpins whose scissile bond has been experimentally determined are marked with an asterisk. Predicted non-inhibitory serpins are underlined.

Fig 5.
Phylogenetic tree from MrBayes analysis of serpin amino acid sequences from *M. sexta* (Ms) and *B. mori* (Bm). Percentage of Bayesian posterior probabilities are shown at the nodes (* represents 100). Branch lengths represents the expected substitution rates per site. One-to-one orthologs of *M. sexta* and *B. mori* serpins are labeled with red lines.

Fig 6.

Temporal and spatial transcript profiles of *M. sexta* serpin genes. The level of mRNA for each serpin gene from different RNAseq libraries (columns) was presented as $\log_2(\text{FPKM}+1)$ values, which are exhibited in the gradient heat map as bins labeled from blue (0) to red (10). The $\log_2(\text{FPKM}+1)$ values were binned as groups 0 (0–0.49), 1 (0.50–1.49), 2 (1.50–2.49), 3 (2.50–3.49), 4 (3.50–4.49), 5 (4.50–5.49), 6 (5.50–6.49), 7 (6.50–7.49), 8 (7.50–8.49), 9 (8.50–9.49), A (9.50–10.49), B (10.50–11.49), C (11.50–12.49), and D (12.50–13.49). Four clusters of genes with similar transcript patterns could be distinguished (A–D, labeled on the right side of the heat map). Underlined libraries are from single-end sequencing, and the others are from pair-end sequencing. MT: Malpighian tubules, L: larvae, D: day, preW: pre-wandering, P: pupae, A: adult, W: wandering, E: embryos. The original data for this figure are shown in Table S2.

Table 1

Summary of *Manduca sexta* serpins

<i>M. sexta</i> serpin name	Gene ID in <i>M. sexta</i> OGS2 ^a	Exons ^b	Predicted inhibitory	Protein Length (residues) ^c	Signal peptide length ^d	MW (kDa) ^e	pI ^e	<i>B. mori</i> ortholog
serpin-1 ^f	Msex2.12851	10	+	375–382	16	41.8–42.7	4.88–5.70	BmSerp11
serpin-2	Msex2.04874	9	+	381	none	42.2	4.99	
serpin-3	Msex2.11872, 15169, 15173	7	+	435	20	48.9	5.27	BmSerp3
serpin-4	Msex2.04237	1	+	391	16	44.7	5.90	BmSerp4
serpin-5	Msex2.12952	1	+	379	17	42.7	6.12	BmSerp5
serpin-6	Msex2.06874	9	+	395	17	45.0	5.57	BmSerp6
serpin-7	Msex2.15425, 15342	1	+	385	15	42.8	5.29	BmSerp32
serpin-8	Msex2.01482	3	–	401	21	44.2	4.45	BmSerp8
serpin-9	Msex2.09571	1	+	382	20	43.3	4.89	BmSerp7
serpin-10	Msex2.11870	8	–	502	17	57.1	5.11	BmSerp10
serpin-11	Msex2.11871	9	+	391	18	44.5	6.23	BmSerp11
serpin-12	Msex2.11869	6	+	467	18	51.4	8.52	BmSerp12
serpin-13	Msex2.06711	10	+	423	21	48.3	5.27	BmSerp13
serpin-14	Msex2.09570	1	–	386	17	45.0	9.52	BmSerp14
serpin-15 ^f	Msex2.06848	9	+	373–377	none	41.0–41.3	4.60–4.92	
serpin-16	Msex2.09066	9	+	389	none	43.2	5.80	BmSerp9
serpin-17	Msex2.15518	8	+	381	none	42.5	4.93	
serpin-18	Msex2.10821	10	–	408	19	45.5	4.86	
serpin-19	Msex2.10820	9	–	390	19	44.2	4.87	
serpin-20	Msex2.15520	2	–	404	16	45.7	8.80	
serpin-21	Msex2.09937	9	+	386	19	43.1	9.64	
serpin-22 ^g	Msex2.06154, 06167	9	–	387	22	46.8	9.27	BmSerp33
serpin-23	Msex2.01480	3	–	375	16	42.1	4.86	BmSerp31
serpin-24	Msex2.01481	3	–	411	22	46.7	5.63	
serpin-25	Msex2.10827	10	–	382	22	43.9	5.96	

<i>M. sexta</i> serpin name	Gene ID in <i>M. sexta</i> OGS2 ^a	Exons ^b	Predicted inhibitory	Protein Length (residues) ^c	Signal peptide length ^d	MW (kDa) ^e	pI ^e	<i>B. mori</i> ortholog
serpin-26	Msex2.10822	10	–	381	20	43.8	9.53	
serpin-27	Msex2.10826	10	–	386	19	44.2	7.26	
serpin-28 ^f	Msex2.10818	8	+	332–333	24	37.8–38.0	5.04–6.06	BmSerp28
serpin-29 ^h	Msex2.09069	7	–	284	none	31.8	8.22	
serpin-30	Msex2.10824	9	–	379	23	44.0	6.29	
serpin-31	Msex2.10823	10	–	382	20	43.8	9.05	
serpin-32	Msex2.05826	2	–	406	17	46.5	6.84	

^a *M. sexta* Official Gene Set-2 (OGS2) is available at https://isik.nal.usda.gov/Manduca_sexta

^b Number of exons included in the mRNA after splicing.

^c Length of mature serpin was calculated after removing the putative signal peptide.

^d SignalP 4.1 was used for signal peptides prediction.

^e Theoretical molecular weight and isoelectric point were calculated by ProParam tool on the ExPASy website.

^f Serpin-1, 15 and 28 produce multiple polypeptides via alternative splicing.

^g Serpin-22 has a haplotype without signal peptide.

^h Serpin-29 gene may be a pseudogene or an incomplete assembly.