Published in final edited form as:

Gene. 2009 August 1; 442(1-2): 47–54. doi:10.1016/j.gene.2009.04.013.

The serpin gene family in *Anopheles gambiae*

Chansak Suwanchaichinda and Michael R. Kanost*

Department of Biochemistry, Kansas State University, Manhattan, KS 66506, USA

Abstract

Serpins (serine protease inhibitors) regulate some innate immune responses of insects by inhibiting endogenous proteases. In this study, we characterized the serpin (*SRPN*) gene family in the mosquito *Anopheles gambiae*, the major malaria vector in Sub-Saharan Africa. We identified 18 *A. gambiae* *SRPN* genes, all on chromosomes 2 and 3, through searches of genomic DNA and EST databases. In addition to *SRPN10*, previously documented to exhibit alternative splicing, we found three splicing isoforms of *SRPN4*. We completed sequencing of cDNAs for the *A. gambiae* serpins to obtain complete coding sequence information and to verify or improve gene predictions. The predicted *SRPN9* and *15* in the initial genome annotation were determined to be a single gene (*SRPN9*). Sixteen of the serpins contained putative secretion signal sequences. Multiple sequence alignments showing conserved residues important in structural conformation, including the consensus pattern within the hinge region, indicated that most of the *A. gambiae* serpins may be inhibitory. Phylogenetic analyses confirmed that *SRPN1*, 2, 3, 8, 9 and 10 formed phylogenetic clusters with known inhibitory serpins from *Drosophila melanogaster* and *Manduca sexta*. Many of the *A. gambiae* serpins were expressed during all life stages. However, *SRPN7*, 8, 12, and 19 were expressed at very low levels in the adult stage. *SRPN13* was expressed mostly in eggs and young larvae, whereas *SRPN5* and 14 were expressed mostly in adults. Such differences in expression pattern suggest that the serpins are involved in multiple physiological processes. Determining the biological functions of the mosquito serpins will require future work to identify the proteases they inhibit *in vivo*.

Keywords

mosquito; cDNA; protease inhibitor; malaria; expression profile

1. Introduction

Serpins (Serine Protease Inhibitors) are a superfamily of proteins with a unique structure comprising three β -sheets and seven to nine α -helices, folding into a conserved tertiary structure with a reactive center loop (RCL) near the carboxyl-terminal (Pearce et al., 2007). The RCL is exposed at the surface of the serpin and acts as a bait for a target protease. After a protease cleaves the RCL at the scissile bond between amino acid residues designated P1 and P1', the serpin undergoes a large conformational change, and the protease is trapped with its active site distorted, resulting in its inactivation (Gettins, 2007). The metastable structural conformation of serpins is required for their inhibitory activity, and thus facilitates the serpins to act as suicide-substrate inhibitors (Silverman et al., 2001). Most serpins inhibit serine proteases,

*Address for Correspondence: Department of Biochemistry, Kansas State University, 141 Chalmers Hall, Manhattan, KS 66506, USA., Tel.: +1 785 532 6964; Fax: +1 785 532 7278; E-mail address: E-mail: kanost@ksu.edu (M.R. Kanost).

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.

although some can inhibit proteases from other classes. For example, squamous-cell carcinoma antigen-1 inhibits cysteine proteases (Suminami et al., 1998), and *Arabidopsis thaliana* serpin-1 inhibits metacaspase-9, an *A. thaliana* cysteine protease (Vercammen et al., 2006). However, some other serpins lack protease inhibitory properties and carry out other functions, such as, hormone transport or acting as molecular chaperones or storage proteins (Dafforn et al., 2001; Huntington and Stein, 2001; Irving et al., 2000).

Serpins are found in an array of eukaryotes, including animals, plants, and some viruses (Gettins, et al., 2002; Irving et al., 2000). Serpins also occur in bacteria and archaea (Cabrita et al., 2007; Kang et al., 2006; Irving et al., 2002; Irving et al., 2003). In insects, serpins have been identified in several species, but most studies of insect serpins are from the fruit fly *Drosophila melanogaster* and the tobacco hornworm *Manduca sexta*. So far, six *M. sexta* serpins have been characterized (Gan et al., 2001; Jiang and Kanost, 1997, Tong and Kanost, 2005; Wang and Jiang, 2004; Zhu et al., 2003) and shown to be inhibitory. With the exception of *Manduca* serpin-2, they all are involved in regulation of the prophenoloxidase (PPO) activation cascade, an innate immune response, but they are likely to have additional functions not yet discovered. The *D. melanogaster* genome contains at least 29 serpin genes (Reichhart, 2005). However, only a handful of the *Drosophila* serpins have been characterized through genetics or biochemistry and shown to be inhibitory and involved in regulating the Toll and PPO signaling pathways (Gubb et al., 2007).

Research on insect innate immune mechanisms is essential for understanding interactions between insect disease vectors and the pathogens they transmit. Of particular importance is the mosquito *Anopheles gambiae*, the major malaria vector of *Plasmodium falciparum* in Sub-Saharan Africa. Human hosts are infected with the *Plasmodium* parasites through bites of female *Anopheles* mosquitoes. The *A. gambiae* genome was completed (Holt et al., 2002), and 242 genes were identified to be immunity-related, including 14 serpin genes (*SRPNs*) (Christophides et al., 2002). Another serpin gene, *SRPN15*, was annotated as a putative haplotype of *SRPN9*. More recently, the serpin gene family in the *A. gambiae* genome was reexamined and compared with those in the genome of another mosquito, *Aedes aegypti* (Waterhouse et al., 2007), and four additional *A. gambiae* serpin genes (in some cases partial genes) were identified. In order to study functions of the serpins and conduct further bioinformatics analyses, knowing their accurate and complete coding sequences will be crucial. Problems with physical mapping and assembly of the *A. gambiae* genome persist, including unmapped scaffolds, physical gaps between mapped scaffolds, polymorphism, bacterial DNA contamination, and incomplete Y chromosome sequence (Sharakhova et al., 2007). Furthermore, gene predictions may be incomplete or inaccurate due to incorrect intron-exon prediction or regions of poor genomic sequence quality. A pilot study using a full-length enriched cDNA library from adult female *A. gambiae* mosquitoes discovered a number of genes previously unannotated (Gomez et al., 2005). It is likely that there are genes still missing from the genome annotation, particularly genes expressed predominantly in immature life stages, for which EST data are scant. Thus, information obtained from cDNA sequences is essential for fully characterizing gene structure, including intron-exon arrangement, alternative splicing, and complete protein-coding regions. The sequences of transcripts also serve as an important tool to study gene expression and to produce recombinant proteins from cloned cDNAs, which will be vital reagents for biochemical studies of the mosquito serpins.

In this study, we cloned and sequenced cDNAs for *A. gambiae* serpins (with the exception of *SRPN10*, which was previously characterized (Danielli et al., 2003)). We identified 18 serpin genes, with three splicing isoforms of *SRPN4*. The cDNA sequences were used to characterize the serpin gene family of *A. gambiae* through bioinformatics and phylogenetic analyses and to examine expression profiles of the serpins among different developmental stages.

2. Materials and Methods

2.1. Insect rearing

Anopheles gambiae G3 strain was originally obtained from the Malaria Research and Reference Reagent Resource Center (MR4, <http://www.mr4.org/>). The colony was maintained according to Benedict (1997) at 27°C, 80% relative humidity, 16:8h light:dark cycle, with adult females feeding on equine blood through a membrane feeder. Larvae were provided ground VitaPro plus staple power flakes and baker's yeast. Adults were provided cotton balls soaked in 10% sucrose as food.

2.2 Database search and cDNA cloning

Nucleotide sequences of initially predicted *A. gambiae* serpin genes were gathered from Genbank. InterPro domain entry IPR000215 (Serpins) was also used to screen the *A. gambiae* genome database in Ensembl for possible serpin genes that had not been identified previously. Predicted amino acid sequences associated with each gene were used for BLASTP queries against the NCBI database to confirm the presence of conserved serpin domains. The nucleotide sequences of serpin gene candidates obtained from both methods were used for EST database searches through GenBank BLASTN. The closest matches from the EST search results were selected, and the EST clones were ordered from the Malaria Research and Reference Reagent Resource Center (MR4) if available (see supplemental Table S1). These EST clones were from *A. gambiae* Giles adult cDNA libraries, and provided as DNA constructs in the pSPORT1 vector in *Escherichia coli* DH10B. SRPN10 had already been extensively studied (Danielli et al., 2003; 2005; Lycett et al., 2004), and thus it was excluded from this search.

For serpins not obtained from MR4, cDNA clones were produced by RT-PCR. Based on expression profiles of each serpin at different developmental stages, RNA from mosquitoes at a stage that showed a high level of expression was used to make cDNA, which was subsequently used as a template for PCR. cDNAs from adult males were used for *SRPN1*, *16*, *17*, *18*; adult females for *SRPN4A*; day-9 larvae for *SRPN7*; and egg for *SRPN13* (see section 2.6 for detail on cDNA preparation). PCR reactions were performed at 30-32 cycles using high fidelity Platinum Taq DNA polymerase (Invitrogen) and specific forward and reverse primers, corresponding to regions containing the start and stop codons to amplify cDNAs containing the entire open reading frame for each serpin. PCR products were analyzed by agarose gel electrophoresis and detected by ethidium bromide staining. The gel bands were excised, and the cDNAs were then extracted from the gel using QIAquick gel extraction kit (Qiagen). The cDNAs were cloned into pCR4-TOPO plasmid vector (Invitrogen) according to the manufacturer's protocol. The cDNA constructs were used to transform One Shot TOP10 Chemically Competent *E. coli* (Invitrogen), which were subsequently allowed to grow on LB agar plates with ampicillin as a selective agent. Bacterial colonies were harvested individually, and plasmids were purified using QIAprep Spin Miniprep kit (Qiagen). To verify that plasmids contained cDNAs for the genes of interest, they were screened by restriction digestion using *EcoRI* (New England Biolabs).

The nucleotide sequence of a gene, named *SRPN19* in this study, was incomplete at both ends in the gene prediction. Oligo-capping rapid amplification of cDNA ends (RACE) methods were used to obtain full-length 5' and 3' ends of the *SRPN19* transcript. 5'- and 3'-RACE reactions were carried out using a GeneRacer kit (Invitrogen) according to the manufacturer's protocol. Day-9 larval cDNA was used as a template, and high fidelity Platinum Taq DNA polymerase was used for amplification. Both RACE products were then cloned and sequenced. Finally, the entire cDNA from 5' to 3' ends was amplified using RACE-ready cDNA as a template. The product was cloned into pCR4-TOPO plasmid vector. One Shot TOP10 Chemically Competent *E. coli* was used for transformation, and the cDNA insert of this plasmid was sequenced.

2.3 DNA sequencing

Sequencing of all plasmids and PCR products in this study was carried out at either the Iowa State University Sequencing Facility (Ames, IA) or the DNA Sequencing Facility at the Department of Plant Pathology, Kansas State University (Manhattan, KS). The sequence data and chromatograms were thoroughly checked and assembled. Deduced amino acid sequences were produced by using the Translate tool via the ExPASy website (<http://www.expasy.org/tools/dna.html>) (Gasteiger et al., 2003). The final nucleotide and amino acid sequences were deposited in the GenBank database (see Table 1 for GenBank accession numbers).

2.4 Bioinformatic analyses

When cDNAs of the *A. gambiae* serpins from MR4 were incomplete at the 5'-end, corresponding 5'-end cDNA sequences available from GenBank (with overlapping regions containing >95% nucleotide sequence identity) were used to assemble complete coding sequences. The presence of secretion signal sequences and location of their cleavage sites were predicted using the web program SignalP 3.0 (<http://www.cbs.dtu.dk/services/SignalP/>) (Bendtsen et al., 2004). Theoretical molecular weights and isoelectric points of the mature serpin proteins were calculated using the ProtParam tool via the ExPASy website (<http://www.expasy.org/tools/protparam.html>) (Gasteiger et al., 2003).

A multiple sequence alignment of amino acid sequences of *A. gambiae* serpins, human alpha-1 antitrypsin (HsA1AT), and *Manduca sexta* serpin-1K (MsSPN1K) was made using TCOFFEE (<http://www.tcoffee.org/>) with default parameters (Notredame et al., 2000). Conserved residues involving structural conformation of serpins were identified based on the amino acid residues of HsA1AT according to Irving et al. (2000). Secondary structures were predicted based on alignment with those identified in the crystal structure of MsSPN1K (Li et al., 1999).

2.5 Phylogenetic analyses

Multiple sequence alignments of *A. gambiae* serpins, *D. melanogaster* serpins (DmSpn4A, 6, 27A, Nec), and *M. sexta* serpins (1K, 2, 3, 4A, 5A, 6) were created using TCOFFEE (default parameters) and MUSCLE (default parameters) (Edgar, 2004). The two alignments were compared, and a region of low similarity at the amino termini of the sequences was manually removed. The remaining portion was re-aligned using CLUSTAL embedded in the Mega 3.1 program (Kumar et al., 2004) to obtain an alignment file, which then was used for creating a neighbor-joining phylogenetic tree and bootstrap test. Gaps were excluded from the analyses, and bootstrapping with 1000 replications was performed.

2.6 Serpin gene expression profiles

Groups of mosquitoes at different developmental stages were homogenized, and total RNA was isolated using the Ultraspec RNA reagent (Biotecx Laboratories), followed by treatment with deoxyribonuclease I (Invitrogen) to remove genomic DNA. Five µg of total RNA was used as template for cDNA synthesis using the SuperScript First-Strand Synthesis System for RT-PCR (Invitrogen), and one µl of the resulting cDNA was used as a template for PCR. PCR reactions were performed using specific primers for each serpin (see Supplemental Table S2). The primers were designed using the web program Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) and searched against the GenBank database of the *A. gambiae* sequences using BLASTN to verify specificity of each primer. Ribosomal protein S7 (RPS7) was used as an RT-PCR internal control. PCR products were analyzed by agarose gel electrophoresis and detected by ethidium bromide staining.

Additional RT-PCR experiments were done to confirm the existence of transcripts containing the alternative exons of *SRPN4* isoforms. A forward primer (5'-AACGAGGACGAAGAAGACGA-3') was located in exon 2, whereas reverse primers (5'-GCAGCACGACCATTACCTTT-3' for *SRPN4A*; 5'-CAGGATGTTGGGAATCTTGC-3' for *SRPN4B*; and 5'-CGACAAAGTCTAGCGGCTTC-3' for *SRPN4C*) were located in the corresponding alternative exon3. The primers spanned at least one intron, depending on the location of exon3 of each isoform. PCR products were amplified with 25 cycles of 94 °C for 30 sec, 56 °C for 30 sec, and 72 °C for 40 sec. Adult female cDNA was used as a template for the PCR reactions.

3. Results and Discussion

3.1 cDNA cloning and sequencing analyses

In this study, we identified 18 serpin genes through BLAST searches of genome and EST databases. They are *SRPN1-14*, *16*, *17*, *18* and *19* (Table 1). We confirmed that previously described *SRPN9* and *SRPN15* are actually the same gene (explained below). Twelve of the serpin genes are present in the genome in four clusters of three genes each (Fig. 1). Two clusters, (*SRPN7*, *14*, *18*) and (*SRPN1*, *2*, *3*), are on chromosome arm 2L, one (*SRPN11*, *12*, *17*) on chromosome arm 2R, and one (*SRPN5*, *6*, *16*) on chromosome arm 3R. No serpin genes were found on 3L or the X chromosome (see Supplemental Table S2 for sequences GenBank accession numbers of the cDNAs and notes on annotation of the genes). Our results are generally consistent with the set of serpin genes identified by Waterhouse et al. (2007), but cDNA sequences we generated have improved gene models and predicted amino acid sequences for some members of the *A. gambiae* serpin family, described below.

We identified possible ESTs for *A. gambiae* serpins in the NCBI database and obtained clones for those available from MR4. However, the sequences of some of those cDNA clones were incomplete. No ESTs were found for *SRPN4A*, *7*, *13* or *16-19*. Some of these were serpins expressed mostly in the egg, larval, and/or pupal stages (see below), and thus were absent from the EST database, which included mainly cDNA clones from adult *A. gambiae*. For our cDNA collection, EST clones for 10 serpin genes were obtained from the MR4 repository. These cDNAs were from libraries made from adult *A. gambiae* Giles mosquitoes. The cDNAs for those serpins not available as ESTs from MR4 were cloned by RT-PCR, using specific primers designed from the gene predictions (see Supplemental Table S1 for details).

Alternative exon splicing has been identified in several insect species (Brandt et al., 2004; Hegedus et al., 2008; Jiang et al., 1996; Krüger et al., 2002;). The alternative splicing leads to serpin isoforms that differ in the carboxyl-terminal region that includes the reactive site loop. In *A. gambiae*, *SRPN10* was previously determined to exhibit alternative splicing with the presence of four isoforms (Danielli et al., 2003). In our study, cDNA and genomic sequence analyses indicated that *SRPN4* consists of three exons with three alternative forms of exon3 (Fig. 2A). This leads to three splicing isoforms of *SRPN4*, namely *SRPN4A*, *4B* and *4C*. In order to confirm the occurrence of alternative splicing of exon3, we conducted RT-PCR using a forward primer located in exon2 and reverse primers located in each exon3, with an expectation of different sizes of PCR products for each isoform. The results clearly verified the presence of the three isoforms (Fig. 2B). The initially predicted *SRPN4* is *SRPN4C*. ESTs for *SRPN4B* and *4C*, but not *4A* were found in sequence databases. Additionally, detection of the *SRPN4A* transcript required a higher number of amplification cycles than the other two isoforms to reach comparable amplification levels (Fig. 5). These data suggest that transcripts for the *SRPN4A* isoform are rare in comparison with the *4B* and *4C* isoforms. In *Ae. aegypti*, four comparable splicing isoforms of the gene orthologous to *SRPN4* were identified (Waterhouse et al., 2007). Functional differences among these isoforms remain unknown.

The initial gene set from the genome sequence included *SRPN9* and *SRPN15* (Christophides et al., 2002), whose predicted transcripts were nearly identical except for a 90 bp region present in *SRPN15*, but not in *SRPN9*. *SRPN15* was annotated as a putative haplotype of *SRPN9*, and Waterhouse et al. (2007) eliminated *SRPN15* from the serpin gene set. In our study, this extra sequence region found in the original *SRPN15* was clearly present in the experimental cDNA sequence. In addition, we used PCR to amplify *SRPN9/15* cDNA and *A. gambiae* genomic DNA templates, using primers that flanked the region in question. Both template samples produced one single product (data not shown), and the sequence of the amplified genomic DNA was identical to that of the cDNA. It appeared that a stretch of bases called as “N” in the genomic sequence data led to incorrect exon-intron boundary prediction at the 3' end of the third exon, resulting in an incorrect sequence for the previously annotated *SRPN9*. Our findings confirm that *SRPN9* and *15* are the same gene, and that the sequence of the predicted *SRPN15* transcript contains the correct exon boundaries. We suggest that this gene should now be named *SRPN9*, and that the name *SRPN15* should not be used further, to avoid future confusion.

SRPN13 was present in the original gene set (ENSANG00000013014) but was not found in the current genome assembly. There were no ESTs for *SRPN13* (other than our cDNA clone). This might be because it was expressed primarily in eggs and very young larvae (Fig. 5), for which there were few ESTs. *SRPN14* and *16* were novel among the *A. gambiae* serpin genes in being composed of a single exon. Comparisons of our cDNA sequences with gene models led to identification of incorrect predicted exon borders in *SRPN7*, *SRPN9*, and *SRPN19* (Supplemental Table S3). In *SRPN19*, only one of the five exons was correctly predicted in the gene model.

3.2 Properties of the *A. gambiae* serpin proteins

General features of the *A. gambiae* serpin amino acid sequences are summarized in Table 1. Sixteen of the 18 serpins were predicted to be secreted proteins, based on detection of a putative secretion signal peptide. *SRPN10* and *SRPN12* lack a secretion signal sequence. *SRPN10* was known from experimental results to be intracellular (Danielli et al., 2003;2005). The putative secretion signal sequences ranged from 16 to 35 residues, with little sequence similarity among different serpin genes. For nine of the serpins, the amino-terminal residue of the mature protein was predicted to be glutamine. Serpins are generally 350-500 amino acids long, with molecular masses of 40-60 kDa (Gettins, 2002;Silverman et al., 2001). The length of most mature serpin proteins in *A. gambiae* varies from 373 to 598 amino acid residues, with theoretical molecular masses of about 42-66 kDa with one exception, *SRPN4A*, which is unusually large. The length of the *SRPN4A* mature protein is 808 residues with a molecular mass of 90.8 kDa. This is probably the largest serpin known to date.

Most of the *A. gambiae* serpins have calculated pI between 5 and 6. However, *SRPN4A*, *4B*, *5*, *6*, *13* and *17* are notably more basic, with calculated pIs of 8.4-9.7. The ranges in size and pI in the *A. gambiae* serpins are wider than those of serpins identified in *M. sexta* and *D. melanogaster*. It is interesting to speculate that the differences of serpins among these insects may be due to the variations of their feeding behavior (blood vs. non-blood feeding), ecology (aquatic vs. terrestrial habitats of specific life stages), and the types of pathogenic microbes that these insects encounter in their environment.

3.3 Amino acid sequence comparisons and analysis

Serpins are usually formed from three β sheets (named A, B, C) and nine α -helices (also named alphabetically). Serpins are metastable proteins that undergo a large conformational change as they inhibit a protease. The protease cleaves a peptide bond in the exposed reactive center loop (RCL), and then the RCL inserts as a new strand into the A β -sheet (Silverman et al., 2001). Structural regions essential for conformational changes of serpins are the hinge, at which the

chain bends to permit RCL insertion, the breach at the top of the A β -sheet, and the shutter, near the center of the A β -sheet, both of which must open to allow RCL insertion (Irving et al., 2000; Whisstock et al., 2000). An additional region known as the gate is involved in a structural transition (latency) in which RCL insertion occurs in the absence of RCL cleavage (Stein and Carrell, 1995). In order to examine these conserved regions and potential structural variations, we aligned the *A. gambiae* serpin sequences along with *M. sexta* serpin-1K (Protein Database Bank (PDB) file 1SEK) and human alpha-1 antitrypsin (HsA1AT, PDB file 1QLP_A), which are inhibitory serpins whose structures have been determined by X-ray crystallography (Elliott et al., 1998; Li et al., 1999). As expected, most amino acid residues known to form important structural regions in HsA1AT were also conserved in the mosquito serpins, with the exceptions of SRPN13 and SRPN19, in which these positions were notably less conserved (Supplemental Fig. S1). Residues at several positions were identical across species, but variations were also present, including some particularly large insertions between predicted secondary structural elements in some of the sequences. An insertion of 30-90 residues between helix D and strand 2A occurs exists in SRPN4, 5, 6, 16 and 19. This region includes a very acidic stretch of 6-9 uninterrupted Asp and Glu residues in SRPN4, 5, 6 and 16, which perhaps could be important as a binding site. SRPN12 and 19 have an insertion (36 and 31 residues, respectively) between helix F and strand 3A. The SRPN4 isoforms each have a greatly extended sequence in the loop between helix I and strand 5A (367 residues in SRPN4A). This region, within the alternatively spliced variable region of the SRPN4 variants, may eventually provide a clue to the functions of the SRPN4 isoforms. A much shorter insertion at the same position also occurs in SRPN16. SRPN13 is unique in containing a 39 residue insertion in the RCL region, and thus must have an unusual RCL compared with most other serpins.

The consensus pattern of specific residues in the hinge region can be used as a theoretical guideline to identify serpins that are inhibitory (Hopkins et al., 1993; Irving et al., 2000). We created an alignment of the hinge region of all *A. gambiae* serpins with known inhibitory serpins from *M. sexta* and *D. melanogaster*. HsA1AT was used as a reference. Indeed, the alignment showed that most of the hinge residues were conserved among the serpins (Fig. 3). SRPN13 and 19 have the least number of conserved residues in the hinge region, supporting a prediction that they may not function as inhibitors. We also predicted the position of the scissile bond of the *A. gambiae* serpins (Fig. 3), based on the alignment with insect serpins with known inhibitory properties. The putative P1 residues in the *A. gambiae* serpins were: Lys (SRPN1, 2, 8, 12, 14, 19); Phe (SRPN3, 10C, 13, 18); Arg (SRPN4A, 4B, 4C, 5, 6, 7, 9, 10A, 10B, 16); and Leu (SRPN11, SRPN17). The P1 residue in SRPN10A was originally predicted to be Lys (Danielli et al., 2003). However, the alignment of the hinge and RSL region showed that SRPN10A exhibited similarity with DmSpn4A, which was experimentally verified to have Arg-Ala as P1-P1' residues (Oley et al., 2004). Therefore, it is likely that the P1 residue in SRPN10A is also Arg. Determining which *A. gambiae* serpins are actually inhibitory and identification of the scissile bond in their RCLs will require further biochemical analysis, which can be accomplished by producing recombinant proteins from the cDNA templates of the mosquito serpins reported here.

3.4 Phylogenetic tree

Phylogenetic trees of subsets of the *A. gambiae* serpins were published previously (Christophides et al., 2002; Michel et al., 2005). In this study, we used all 18 of the *A. gambiae* mosquito serpins for a phylogenetic analysis. A neighbor-joining phylogenetic tree of the *A.* serpins was constructed, and known inhibitory serpins from *M. sexta* and *D. melanogaster* were included in the analysis. The tree showed seven phylogenetic clusters among the serpins, with SRPN12, 13 and 19 not grouped with any serpins (Fig. 4). SRPN1, 2 and 3 were grouped with MsSPN3A and DmSpn27A, which are inhibitors of prophenoloxidase (PPO)-activating proteases (PAPs) (De Gregorio et al., 2002; Zhu et al., 2003). Our previous

experimental data demonstrated that recombinant SRPN1 and 2 can function as inhibitors of *M. sexta* PAP-3 (Michel et al., 2006). SRPN8 was grouped with MsSPN4 and 5, which are inhibitors of PPO activation (Tong and Kanost, 2005). SRPN9 was grouped with MsSPN6, which is also an inhibitor of *M. sexta* PAP-3 (Wang and Jiang, 2004; Zou and Jiang, 2005). SRPN10A was grouped with DmSpn4A and DmSpn6, along with MsSPN1, MsSPN2 and DmNec, all known to be functional serine protease inhibitors (Gan et al., 2001; Han et al., 2000; Jiang and Kanost, 1997; Levashina et al., 1999; Richer et al., 2004). Other phylogenetic clusters were: SRPN11 and 17; SRPN4, 5, 6 and 16; and SRPN7, 14 and 18. With the exception of SRPN12, the mosquito serpins that shared phylogenetic clusters are also located in gene clusters on the chromosomes (Fig. 1), as expected if they are products of fairly recent gene duplications. SRPN8 and 9 are of particular interest because they may be orthologous with *Manduca* serpins of known functions that have been characterized biochemically. Future characterization of recombinant *A. gambiae* serpins will be important for assessing their biochemical properties and biological functions in mosquitoes.

3.5 Expression profiles

We examined the developmental expression profiles of *A. gambiae* serpins using semi-quantitative RT-PCR. This technique allowed us to determine if the serpins were expressed and at what life stages. We found that most of these serpins were expressed in all developmental stages (Fig. 5A). However, SRPN13 was expressed predominantly in eggs (developing embryos) and very young larvae. SRPN13 is similar to the ovalbumin family of clade-B serpins (data not shown), and its high expression in the eggs and very young larvae may point to a role during early development. SRPN7 and 12 transcripts were highly abundant in egg, larvae and tan pupae, whereas SRPN5 transcript was more abundant in black pupae (which are technically pharate adults) and adults. SRPN8 expression was high in eggs and larvae. SRPN14 transcript was more abundant in adults, whereas SRPN19 expression was not detected in adults. We reexamined the expression of serpins whose transcripts were not detected in adults by carrying out RT-PCR analysis of adult and pupal RNA samples at a higher number of amplification cycles. The results revealed that SRPN7 and 12 transcripts were present at a low level in black pupae and adults, whereas SRPN13 expression was not detected at all in these two stages (Fig. 5B). SRPN19 transcript was detected in pupae but not in adults.

In the *A. gambiae* adults, the expression level of *SRPN10B* transcript has been shown to be the most abundant of the four *SRPN10* isoforms (Danielli et al., 2003). All *SRPN10* isoform transcripts were also found to be abundant in the dissected adult midguts, in comparison with the thorax and the gut-free abdomen. The expression levels of SRPN10 proteins were very low in embryos and early instar larvae, very high in last instar larvae, and then decreased in pupae and adults, with the expression level in females higher than in males (Danielli et al., 2003). The differences of expression levels of serpins among life stages may imply different physiological roles in the mosquito.

4. Conclusion

Eighteen serpins have been identified in the mosquito *A. gambiae*. In addition to *SRPN10* gene previously found to exhibit alternative splicing, *SRPN4* with three isoforms as a result of alternative splicing of the last exon has been identified in this study. Some of the *A. gambiae* serpins are expressed fairly uniformly at all developmental stages, whereas others have unique patterns of stage-specific expression. cDNAs of the serpins were cloned, sequenced, and complete coding sequences were generated. The amino acid sequences, were aligned with serpins from other species, showing the conserved residues important in structural conformation required for inhibitory activity. The consensus patterns of the hinge region, along with the phylogenetic analysis data, suggest that most of the *A. gambiae* serpins are inhibitory.

However, additional biochemical and biological studies will be required to determine their functions. The available cDNA information will permit production of recombinant proteins, allowing exploration of their functions *in vitro* and *in vivo*. The information will also be useful in further molecular genetic analyses in this medically important mosquito.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Dr. Maureen Gorman for providing some of the RNAs and cDNA pools that were used in this study and for technical advice, and we thank Ms. Sandi Yungeberg for maintaining the mosquito colony. We also thank MR4 for providing us with *Anopheles gambiae* cDNA clones originally contributed by Robert A. Holt. This work was supported by NIH grants AI31084 and GM41247, and contribution 07-124-J from the Kansas Agricultural Experiment Station.

References

- Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 2004;340:783–795. [PubMed: 15223320]
- Benedict, MQ. Care and maintenance of anopheline mosquito colonies. In: Crampton, JM.; Beard, CB.; Louis, C., et al., editors. *The molecular biology of insect disease vectors: A methods manual*. Chapman & Hall; London: 1997. p. 3-12.
- Brandt KS, Silver GM, Becher AM, Gaines PJ, Maddux JD, Jarvis EE, Wisneski N. Isolation, characterization, and recombinant expression of multiple serpins from the cat flea, *Ctenocephalides felis*. *Arch Insect Biochem Physiol* 2004;55:200–214. [PubMed: 15027074]
- Cabrita LD, Irving JA, Pearce MC, Whisstock JC, Bottomley SP. Aeropin from the extremophile *Pyrobaculum aerophilum* bypasses the serpin misfolding trap. *J Biol Chem* 2007;282:26802–26809. [PubMed: 17635906]
- Christophides GK, Zdobnov E, Barillas-Mury C, et al. Immunity-related genes and gene families in *Anopheles gambiae*. *Science* 2002;298:159–165. [PubMed: 12364793]
- Dafforn TR, Della M, Miller AD. The molecular interactions of heat shock protein 47 (Hsp47) and their implications for collagen biosynthesis. *J Biol Chem* 2001;276:49310–49319. [PubMed: 11592970]
- Danielli A, Barillas-Mury C, Kumar S, Kafatos FC, Loukeris TG. Overexpression and altered nucleocytoplasmic distribution of *Anopheles* ovalbumin-like SRPN10 serpins in *Plasmodium*-infected midgut cells. *Cell Microbiol* 2005;7:181–190. [PubMed: 15659062]
- Danielli A, Kafatos FC, Loukeris TG. Cloning and characterization of four *Anopheles gambiae* serpin isoforms, differentially induced in the midgut by *Plasmodium berghei* invasion. *J Biol Chem* 2003;278:4184–4193. [PubMed: 12456678]
- De Gregorio E, Han S, Lee W, et al. An immune-responsive serpin regulates the melanization cascade in *Drosophila*. *Dev Cell* 2002;3:581–592. [PubMed: 12408809]
- Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004;32:1792–1797. [PubMed: 15034147]
- Elliott PR, Pei XY, Dafforn TR, Lomas DA. Wild type α 1-antitrypsin is in the canonical inhibitory conformation. *J Mol Biol* 1998;275:419–425. [PubMed: 9466920]
- Gan H, Wang Y, Jiang H, Mita K, Kanost MR. A bacterial-induced, intracellular serpin in granular hemocytes of *Manduca sexta*. *Insect Biochem Mol Biol* 2001;31:887–898. [PubMed: 11439248]
- Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* 2003;31:3784–3788. [PubMed: 12824418]
- Gettins PGW. Serpin structure, mechanism, and function. *Chem Rev* 2002;102:4751–4803. [PubMed: 12475206]

- Gettins, PGW. Mechanisms of serpins inhibition. In: Silverman, GA.; Lomas, DA., editors. Molecular and Cellular Aspects of the Serpinopathies and Disorders in Serpin Activity. World Scientific; Hackensack, NJ: 2007. p. 67-100.
- Gomez SM, Eiglmeier K, Segurens B, Dehoux P, Couloux A, Scarpelli C, Wincker P, Weissenbach J, Brey PT, Roth CW. Pilot *Anopheles gambiae* full-length cDNA study: sequencing and initial characterization of 35,575 clones. *Genome Biol* 2005;6:R39. [PubMed: 15833126]
- Gubb, D.; Robertson, A.; Dafforn, T.; Troxler, L.; Reichhart, J. *Drosophila* serpins: regulatory cascades in innate immunity and morphogenesis. In: Silverman, GA.; Lomas, DA., editors. Molecular and cellular aspects of the serpinopathies and disorders in serpin activity. World Scientific Publishing; Singapore: 2007. p. 207-227.
- Hegedus DD, Erlandson M, Baldwin D, Hou X, Chamankhah M. Differential expansion and evolution of the exon family encoding the Serpin-I reactive centre loop has resulted in divergent serpin repertoires among the Lepidoptera. *Gene* 2008;418:15–21. [PubMed: 18495381]
- Han J, Zhang H, Min G, Kemler D, Hashimoto C. A novel *Drosophila* serpin that inhibits serine proteases. *FEBS Letters* 2000;468:194–198. [PubMed: 10692585]
- Holt RA, Subramanian GM, Halpern A, et al. The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* 2002;298:129–149. [PubMed: 12364791]
- Hopkins PCR, Carrell RW, Stone SR. Effects of mutations in the hinge region of serpins. *Biochemistry* 1993;32:7650–7657. [PubMed: 8347575]
- Huntington JA, Stein PE. Structure and properties of ovalbumin. *J Chromatogr B Biomed Sci Appl* 2001;756:189–198. [PubMed: 11419711]
- Irving JA, Pike RN, Lesk AM, Whisstock JC. Phylogeny of the serpin superfamily: Implications of patterns of amino acid conservation for structure and function. *Genome Res* 2000;10:1845–1864. [PubMed: 11116082]
- Irving JA, Steenbakkens PJM, Lesk AM, et al. Serpins in prokaryotes. *Mol Biol Evol* 2002;19:1881–1890. [PubMed: 12411597]
- Irving JA, Cabrita LD, Rossjohn J, Pike RN, Bottomley SP, Whisstock JC. The 1.5 Å crystal structure of a prokaryote serpin: Controlling conformational change in a heated environment. *Structure* 2003;11:387–397. [PubMed: 12679017]
- Jiang H, Wang Y, Huang Y, Mulnix AB, Kadel J, Cole K, Kanost MR. Organization of serpin gene-1 from *Manduca sexta*. Evolution of a family of alternative exons encoding the reactive site loop. *J Biol Chem* 1996;271:28017–28023. [PubMed: 8910411]
- Jiang H, Kanost MR. Characterization and functional analysis of 12 naturally occurring reactive site variants of serpin-1 from *Manduca sexta*. *J Biol Chem* 1997;272:1082–1087. [PubMed: 8995406]
- Kang S, Barak Y, Lamed R, Bayer EA, Morrison M. The functional repertoire of prokaryote cellulosomes includes the serpin superfamily of serine proteinase inhibitors. *Mol Microbiol* 2006;60:1344–1354. [PubMed: 16796673]
- Krüger O, Ladewig J, Köster K, Ragg H. Widespread occurrence of serpin genes with multiple reactive centre-containing exon cassettes in insects and nematodes. *Gene* 2002;293:97–105. [PubMed: 12137947]
- Kumar S, Tamura K, Nei M. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* 2004;5:150–163. [PubMed: 15260895]
- Levashina EA, Langley E, Green C, Gubb D, Ashburner M, Hoffmann JA, Reichhart J. Constitutive activation of Toll-mediated antifungal defense in serpin-deficient *Drosophila*. *Science* 1999;285:1917–1919. [PubMed: 10489372]
- Li J, Wang Z, Canagarajah B, Jiang H, Kanost M, Goldsmith EJ. The structure of active serpin 1K from *Manduca sexta*. *Structure* 1999;7:103–109. [PubMed: 10368276]
- Lycett GJ, Kafatos FC, Loukeris TG. Conditional Expression in the Malaria Mosquito *Anopheles stephensi* With Tet-On and Tet-Off Systems. *Genetics* 2004;167:1781–1790. [PubMed: 15342516]
- Michel K, Budd A, Pinto S, Gibson TJ, Kafatos FC. *Anopheles gambiae* SRPN2 facilitates midgut invasion by the malaria parasite *Plasmodium berghei*. *EMBO Reports* 2005;6:891–897. [PubMed: 16113656]

- Michel K, Suwanchaichinda C, Morlais I, et al. Increased melanizing activity in *Anopheles gambiae* does not affect development of *Plasmodium falciparum*. PNAS 2006;103:16858–16863. [PubMed: 17065316]
- Notredame C, Higgins D, Heringa J. T-Coffee: A novel method for multiple sequence alignments. J Mol Biol 2000;302:205–217. [PubMed: 10964570]
- Oley M, Letzel MC, Ragg H. Inhibition of furin by serpin Spn4A from *Drosophila melanogaster*. FEBS Letters 2004;577:165–169. [PubMed: 15527779]
- Pearce, MC.; Pike, RN.; Lesk, AM.; Bottomley, SP. Serpin conformations. In: Silverman, GA.; Lomas, DA., editors. Molecular and Cellular Aspects of the Serpinopathies and Disorders in Serpin Activity. World Scientific; Hackensack, NJ: 2007. p. 35–66.
- Reichhart J. Tip of another iceberg: *Drosophila* serpins. Trends Cell Biol 2005;15:659–665. [PubMed: 16260136]
- Richer MJ, Keays CA, Waterhouse J, Minhas J, Hashimoto C, Jean F. The Spn4 gene of *Drosophila* encodes a potent furin-directed secretory pathway serpin. PNAS 2004;101:10560–10565. [PubMed: 15247425]
- Sharaknova MV, Hammond MP, Lobo NF, et al. Update of the *Anopheles gambiae* PEST genome assembly. Genome Biol 2007;8:R5. [PubMed: 17210077]
- Silverman GA, Bird PI, Carrell RW, et al. The serpins are an expanding superfamily of structurally similar but functionally diverse proteins. J Biol Chem 2001;276:33293–33296. [PubMed: 11435447]
- Stein PE, Carrell RW. What do dysfunctional serpins tell us about molecular mobility and disease? Nat Struct Biol 1995;2:96–113. [PubMed: 7749926]
- Suminami Y, Nawata S, Kato H. Biological role of SCC antigen. Tumour Biol 1998;19:488–493. [PubMed: 9817978]
- Tong Y, Kanost MR. *Manduca sexta* serpin-4 and serpin-5 inhibit the prophenol oxidase activation pathway. J Biol Chem 2005;280:14923–14931. [PubMed: 15695807]
- Vercammen D, Belenghi B, van de Cotte B, et al. Serpin1 of *Arabidopsis thaliana* is a suicide inhibitor for metacaspase 9. J Mol Biol 2006;364:625–636. [PubMed: 17028019]
- Wang Y, Jiang H. Purification and characterization of *Manduca sexta* serpin-6: a serine proteinase inhibitor that selectively inhibits prophenoloxidase-activating proteinase-3. Insect Biochem Mol Biol 2004;34:387–395. [PubMed: 15041022]
- Whisstock JC, Skinner R, Carrell RW, Lesk AM. Conformational changes in serpins I. The native and cleaved conformations of α 1-antitrypsin. J Mol Biol 2000;296:685–699. [PubMed: 10669617]
- Waterhouse RM, Xi ZY, Kriventseva E, Meister S, Alvarez KS, Bartholomay LC, Barillas-Mury C, Blandin S, Christensen BM, Dong Y, Jiang H, Kanost MR, Koutsos AC, Levashina EA, Li J, Ligoxygakis P, MacCallum R, Mayhew GF, Mendes A, Michel K, Osta M, Paskewitz S, Shin SW, Vlachou D, Wang L, Wei W, Zheng L, Zou A, Severson DW, Raikhel AS, Kafatos FC, Dimopoulos G, Zdobnov E, Christophides GK. Evolutionary dynamics of immune-related genes and pathways in disease vector mosquitoes. Science 2007;316:1738–1743. [PubMed: 17588928]
- Zhu Y, Wang Y, Gorman MJ, Jiang H, Kanost MR. *Manduca sexta* serpin-3 regulates prophenoloxidase activation in response to infection by inhibiting prophenoloxidase-activating proteinases. J Biol Chem 2003;278:46556–46564. [PubMed: 12966082]
- Zou Z, Jiang H. *Manduca sexta* serpin-6 regulates immune serine proteases PAP-3 and HP8. J Biol Chem 2005;280:14341–14348. [PubMed: 15691825]

Abbreviations

SRPN

serine protease inhibitor

Ag

Anopheles gambiae

Ms

Manduca sexta

Dm	<i>Drosophila melanogaster</i>
PPO	prophenoloxidase
PAP	prophenoloxidase-activating protease

Fig. 1.
Chromosomal location of the *A. gambiae* serpin genes.

A **B** **Fig. 2.**

Intron-exon arrangement of the three isoforms of the *A. gambiae SRPN4*. (A) The *SRPN4* gene consists of exon1, exon2 and the alternative exon3. Uncolored areas of the exons indicate untranslated regions. The numbers indicate the sizes of introns and exons. (B) Coding regions of isoform transcripts are shown with arrows to indicate the locations of primers used for RT-PCR, and numbers to indicate the expected sizes of the amplification products (see text for detail). The PCR products were electrophoresed and gel stained with ethidium bromide. These results confirm the existence of transcripts for the predicted *SRPN4A*, *4B* and *4C* isoforms.

Fig. 3.

Alignment of the hinge and reactive site loop region of *A. gambiae* serpins. Sequences for human alpha-1 antitrypsin (HsA1AT) and several serpins of known function from *Drosophila melanogaster* (DM) and *Manduca sexta* (MS) are included for comparison. *A. gambiae* serpins are presented in numerical order and grouped with similar serpins from other species as determined by phylogenetic analysis (see Fig. 4). Predicted P1 residues are highlighted in yellow. Serpins for which the scissile bond has been determined experimentally are marked with an asterisk.

Fig. 4.

A neighbor-joining phylogenetic tree of the *A. gambiae* serpins (AgSRPN) and known inhibitory serpins from *Drosophila melanogaster* (DrSpn) and *Manduca sexta* (MsSPN). Gaps were excluded, and bootstrap support values higher than 60% are shown at the nodes ($n = 1000$). The branch length represents the accumulation of amino acid changes. The scale indicates 0.1 amino acid change per site. Vertical bars indicate phylogenetic clusters.

Fig. 5.

Expression profile of the *A. gambiae* serpins among developmental stages. Total RNA was isolated from mosquitoes, and deoxyribonuclease-I was used to remove genomic DNA. An equal amount of RNA was used to make cDNA, using the SuperScript first-strand cDNA synthesis System. PCR reactions were performed at 25 cycles (unless otherwise indicated) using specific primers for each serpin. Ribosomal protein S7 (RPS7) was used as an internal control. SRPN10 was excluded from this experiment (see text for detail). E, eggs; L2, day-two larvae; L6, day-six larvae; L9, day-nine larvae; TP, tan pupae; BP, black pupae; AF, adult females; AM, adult males.

Table 1

Summary of *Anopheles gambiae* serpins.

Name	Source of coding sequences ^a	Mature protein length (aa)	Predicted Signal peptide and N-terminal residue ^c	MW (kDa) ^d	pI ^d
SRPN1	cDNA EF117200	395	1-26 (F)	44.95	5.42
SRPN2	cDNA DQ974161	388	1-21 (Q)	44.25	5.20
SRPN3	cDNA DQ974162	396	1-22 (A)	44.71	5.55
SRPN4A	cDNA EF034031	808	1-25 (A)	90.76	8.38
SRPN4B	cDNA DQ974163 + 5'end BX027825	598	1-25 (A)	66.23	8.55
SRPN4C	cDNA DQ974164 + 5'end BX027825	522	1-25 (A)	59.04	5.95
SRPN5	cDNA DQ974165	450	1-32 (E)	50.41	8.68
SRPN6	cDNA DQ974166	473	1-21 (Q)	52.86	9.59
SRPN7	cDNA EF989011	374	1-25 (Q)	41.86	5.02
SRPN8	cDNA DQ974167	414	1-20 (Q)	46.61	4.74
SRPN9(15)	cDNA DQ974168	419	1-28 (Q)	46.92	5.49
SRPN10A, 10B, 10C, 10D	AJ420785 (protein ID no. CAD12784, CAD 12783, CAD12782, CAD12781) ^b	395, 380, 382, 379	None	43.99, 42.58, 42.47, 42.16	5.56, 5.33, 5.14, 5.13
SRPN11	cDNA DQ974169	492	1-16 (Q)	55.48	6.40
SRPN12	cDNA DQ974170	511	None	58.04	4.64
SRPN13	cDNA DQ974172	379	1-30 (E)	42.25	9.67
SRPN14	cDNA DQ974171	386	1-17 (G)	43.93	7.39
SRPN16	cDNA DQ974173	525	1-28 (Q)	58.08	5.99
SRPN17	cDNA EF117201	446	1-35 (Q)	49.74	9.01
SRPN18	cDNA DQ974174	373	1-18 (D)	42.06	5.03
SRPN19	cDNA EU124656	568	1-28 (Q)	64.06	5.05

^a GenBank accession no.

^b See Reference [Danielli et al., 2003].

^c Signal peptides were predicted using the SignalP 3.0 program.

^d Molecular weight (MW) and theoretical pI were calculated using the ProtParam tool via ExPASy website.