REGULATION OF PROPHENOLOXIDASE-

ACTIVATING PROTEINASES

IN *MANDUCA SEXTA*

By

ZHEN ZOU

Bachelor of Engineering

East China University of Science and Technology

Shanghai, China

1997

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
July, 2005

REGULATION OF PROPHENOLOXIDASE-

ACTIVATING PROTEINASES

IN *MANDUCA SEXTA*

Dissertation Approved:

Haobo Jiang
_____
Dissertation Adviser
Jack Dillwith

_____
John Sauer

_____
Richard Essenberg

_____
A. Gordon Emslie
_____
Dean of the Graduate College

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

Chapter V

Chapter VI

LIST OF TABLES

LIST OF ABBREVIATIONS

antimicrobial peptide: AMP

Ile-Glu Ala-Arg-p-nitroanilide: IEARpNA

β-1,3-glucan recognition protein: βGRP

carbohydrate recognition domain: CRD

20-hydroxylecdysone: 20E

ecdysone receptor: EcR

ecdysone response element: EcRE

hemolymph proteinase: HP

immulectin: IML

immune deficiency: Imd

interferon stimulus response element: ISRE

juvenile hormone: JH

leucine-rich repeat immune protein: LRIM1

peptidoglycan recognition protein: PGRP

peptidoglycan: PG

prophenoloxidase activating enzyme: PPAE

1-phenyl-2-thiourea: PTU

lipopolysaccharide: LPS

prophenoloxidase: proPO

phenoloxidase: PO

pathogen-associated molecular pattern (PAMP)

pattern recognition receptor: PRR

proPO-activating proteinase: PAP

retinoid X receptor: RXR

reverse transcription-polymerase chain

    reaction: RT-PCR

TNF receptor interacting protein: RIP

Toll interaction receptor: TIR

serine proteinase homolog: SPH

serine proteinase: SP

serine proteinase inhibitor: serpin

signal transducers and activators of

transcription: STAT

thioester-containing protein: TEP

tumor necrosis factor: TNF

Ultraspiracle: USP

Chapter I. Introduction

To survive and prosper in pathogen-rich environments, insects have to rely on their innate immune system to fight against invading pathogens or parasites. Innate immunity is evolutionarily conserved from lower metazoans to mammals. Studies of insect defense mechanisms enhance our understandings of the human immune system and allow us to develop new strategies for controlling agricultural pests and arthropod vectors of human diseases.

Acute immune responses in insects are stimulated by pathogen-associated molecular patterns (PAMPs) including peptidoglycan (PG) from Gram-positive bacteria, lipopolysaccharide (LPS) from Gram-negative bacteria, and β-1,3-glycan from fungi (Janeway and Medzhitov, 2002). A group of hemolymph proteins, designated pattern recognition receptors (PRRs), specifically bind to PAMPs. Molecular interactions between PAMPs and PRRs provoke systemic defense reactions such as phagocytosis, nodule formation, encapsulation, and synthesis of antimicrobial peptides (AMPs).

Prophenoloxidases (proPO) are present in both cuticle and hemolymph of insects. Their active forms, phenoloxidases (POs), catalyze the formation of quinones that are precursors of melanin (Ashida and Brey, 1998). Melanotic encapsulation is an effective mechanism to immobilize parasites. Quinones are toxic to invading microbes. They are also involved in cuticle sclerotization and wound healing. Proteolytic activation of proPO depends on a largely unknown serine proteinase (SP) cascade. The terminal enzyme of the cascade, proPO-activating proteinase (PAP), cleaves proPO. So far, three PAPs have

been identified in the tobacco hornworm, *Manduca sexta* (Jiang et al., 1998; Jiang et al., 2003a; Jiang et al., 2003b). They contain 1 or 2 regulatory clip domains. At least in some insects, PAP requires accessory protein(s) for generating active PO. Two clip-domain serine proteinase homologs (*M. sexta* SPH-1 and SPH-2) function as a "cofactor" for the PAPs (Yu et al., 2003). The expression levels of PAPs and SPHs increase significantly after a bacterial challenge. To date, organization and expression of *M. sexta* PAP genes have not yet been studied.

ProPO activation is negatively regulated by serine proteinase inhibitors of the serpin superfamily. Serpins, 370-450 residues long, share a common tertiary structure. They inhibit SPs through an exposed reactive site loop. In mammals, serpins regulate SPs involved in blood coagulation, fibrinolysis, complement activation, and inflammatory responses (Gettins, 2002). In *M. sexta*, six serpins have been discovered. Serpin-1J and serpin-3 block proPO activation by inhibiting the PAPs in the hemolymph (Jiang et al., 1997; Jiang et al., 2003b; Zhu et al., 2003). *M. sexta* serpin-6 inhibits PAP-3 but not PAP-1 (Wang and Jiang, 2004). Molecular mechanism and physiological function of serpin-6 have not been fully investigated.

In *Drosophila*, the Toll and Imd pathways lead to synthesis of AMPs in response to pathogen infection (Hoffmann and Reichhart, 2002). Recognition of fungi and Gram-positive bacteria triggers an extracellular SP cascade which generates an active ligand (spätzle) for activating the Toll pathway. Persephone, a clip-domain SP, is a component of the cascade. Mutations in its gene shut down the Toll pathway and drosomycin synthesis induced by fungal infection (Ligoxygakis et al., 2002a). Spn43Ac, also known as Necrotic or nec serpin, down-regulates the signal transduction by inhibiting a

proteinase in the pathway (Levashina et al., 1999; Green et al., 2000). Truncated Spn43Ac was expressed in *E. coli* and found to be an inhibitor with a broad spectrum (Robertson et al., 2002). The structure and physiological target of full-length Spn43Ac have not been examined.

My research aims to: 1) study the organization and transcriptional regulation of *M. sexta* PAP-1 and PAP-3 genes; 2) characterize *M. sexta* serpin-6 and its interaction with PAP-3; 3) produce *D. melanogaster* Spn43Ac and analyze its role in regulating immune SPs

Chapter II. Literature review

Immune systems protect multicellular organisms from microbial infection. Compared with higher metazoans, insects lack adaptive immunity and rely solely on their innate immunity, which responds to infectious agents efficiently but less specifically. In the past two decades, *Drosophila* and several other insects have been used as biological models to investigate the role of innate immunity in host-pathogen interactions. Molecular elucidation of immune pathways demonstrates that many defense mechanisms were conserved in vertebrates and invertebrates.

Insect cuticle and midgut peritrophic membrane (containing chitins and proteins mainly) constitute a physical barrier against pathogen infection (Sugumaran, 1996). When their integrity is compromised, a systemic defense reaction occurs at the wounding site to sequester and eliminate the microbes entering the hemoceol. The immune system in insect hemolymph is classified into cellular and humoral responses (Gillespie et al., 1997; Hoffmann and Reichhart, 2002). Figure 1 demonstrates a framework of these responses in the hemocoelof *M. sexta*.

Initiation of immune responses

When microbes infect insects, multiple PRRs recognize their conserved surface determinants (*e.g.* LPS, peptidoglycan, lipoteichoic acid, β-1,3-glucan etc.) and trigger host innate immune reactions (Janeway and Medzhitov, 2002; Yu et al., 2002b). While some PRRs are constitutively expressed and secreted as surveillance sensors in hemolymph, others are synthesized in response to the entry of microorganisms (Gillespie

et al., 1997). Upon binding to their target molecules, PRRs undergo a conformational change required for recruiting plasma factors and hemocytes that eliminate the pathogens. It is unclear how such structural changes activate complex signaling pathways which lead to proPO activation and de novo synthesis of immune proteins, *e.g.* cytokines and antimicrobial peptides (AMPs). Invertebrate PRRs are described in details as follows.

Peptidoglycan recognition proteins (PGRPs) comprise a large group of PRRs identified in most families of Lepidoptera and Diptera. They all contain a 180-residue PGR domain which (supposedly) binds to PG. Upon binding, *B. mori* PGRP triggers an extracellular SP cascade for proPO activation (Yoshida, et al., 1996). The primary structure of *B. mori* PGRP is similar to T7 lysozyme, but it lacks a hydrolytic activity. In the genomes of *D. melanogaster* and *Anopheles gambiae*, there are three types of PGRP genes coding for the secreted short-form, transmembrane and cytoplasmic long-forms ranging from 200 to 600 residues (Gottar et al., 2002). This suggests that PGRPs have multiple functions in the pathogen recognition. Loss-of-function mutant of PGRP-SA is sensitive to Gram-positive bacterial infection, but not to fungal infection (Michel et al., 2001). The membrane-bound PGRP-LC transmits signal of Gram-negative bacterial invasion to the immune deficiency (Imd) pathway (Ramet et al., 2002; Werner et al., 2003). PGRP-SA detects Lys-type PG of Gram-positive bacteria, whereas PGRP-LC (L means long form of PGRP) recognizes DAP type PG of Gram-negative bacteria. Cytoplasmic PGRP-LE, when overexpressed in *Drosophila* larvae, leads to Imd-independent AMP synthesis and spontaneous melanization (Takehana et al., 2002), indicating that PGRP-LE may participate in the epithelial AMP induction and systemic responses. PGR-domains are similar in overall sequence to *N*-acetylmuramoyl *L*-alanine

amidase (NAMLAA). PGRP-SC1B (S means short form of PGRP), containing the catalytic residues of NAMLAA, hydrolyzed PG from *Staphylococcus aureus* (Mellroth et al., 2003). PG treated by PGRP-SC1B lost most of its immunostimulatory properties. The crystal structures of PGRP-SA and PGRP-LB revealed important structure-function relationships of PGRPs (Reiser et al., 2004).

Thioester-containing proteins (TEPs) in nematodes, insects, and mammals have a conserved motif harboring an intrachain β-cysteinyl-γ-glutamyl thioester bond. α2-Macroglobulin acts as an inhibitor of proteinases by trapping the enzymes through the thioester linkage. During immune reactions, complement factors C3, C4, and C5 bind to self/non-self surfaces via the same bond. *Anopheles gambiae* TEP-1 is a 165 kDa hemolymph protein. After a septic injury, it is cleaved to become the 80 kDa active form and attaches to Gram-positive and Gram-negative bacteria through the covalent bond (Levashina et al., 2001). The bacteria are subsequently cleared by phagocytosis. TEP1 is essential for blocking *Plasmodium* oocyst development in the midgut of *A. gambiae* (Blantin et al., 2004). Knocking-out TEP1 expression in the susceptible strain of *A. gambiae* caused a 5-fold increase in the number of oocysts. No oocyst melanization was observed when TEP1 production was blocked in the refractory strain. Therefore, TEPs may function as an immune surveillance against pathogen invasion.

C-type lectins in insects are also involved in pattern recognition and cellular interaction. *Manduca sexta* immulectin-1 and -2 (IML-1 and IML-2) contain two tandem carbohydrate recognition domains. They are highly similar in sequence to such domains in other C-type lectins from insects to mammals. IML-2 is constitutively expressed at a low level in the fat body, and its level significantlyincrease s after injection of Gram-

negative bacteria or LPS (Yu et al., 2000). IML-2 agglutinates *E. coli* in the presence of calcium. It may anchor SPHs, PAPs, and proPO in the vicinity of microbial cells (Yu et al., 2003). IML2 enhances melanization and encapsulation. In *A. gambiae*, two C-type lectins prevent *Plasmodium* from melanization (Osta et al., 2004a).

Members of the β-1,3-glucan recognition protein (βGRP) family include a glucanase-like domain in the carboxyl terminus. They bind to β-1,3-glucan (a fungal cell wall component) but lack the amidase activity, due to amino acid substitutions in the catalytic center. Other members of the family, which bind to LPS of Gram-negative bacteria, are called GNBPs (standing for Gram-negative bacteria-binding proteins). The mRNA level of *B. mori* GNBP is significantly elevated in response to bacterial infection (Lee et al., 1996). *Drosophila* GNBP1 works in concert with PGRP-SA in the activation of Toll pathway (Gobert et al., 2003). Up till now, two βGRPs have been purified and cloned from *M. sexta*. They are similar in sequence to *Bombyx mori* βGRP, which is composed of two functional sites. The N-terminal site comprising 102 residues attaches to β-1,3-glucan, while the C-terminal one resembles a glucanase. The *M. sexta* βGRPs aggregate yeasts and bacteria through β-1,3-glucan and lipoteichoic acids (but not LPS). This binding enhances proPO activation (Ma et al., 2000, Jiang et al., 2004).

As our knowledge on PRRs accumulates rapidly, we still don't know much about how extracellular SP cascades are activated upon microbial infection. Recently, *M. sexta* hemolymph proteinase 14 (HP14) was isolated, which serves as a linker between pathogen recognition and cascade activation. HP14 has a complex domain structure, including five low-density lipoprotein receptor (LDLr) class A repeats, a Sushi domain, an unique Cys-rich region, and a proteinase-catalytic domain (Ji et al., 2004). Analogous

to factor C and factor G in the horseshoe crab blood clotting cascade, HP14 precursor binds PG, autoactivates, and triggers the proPO activation system in the hemolymph. In *A. gambiae*, another modular SP (Sp22D) contains two chitin-binding domains, a mucin-like region, two LDLr class A repeats, and two scavenger receptor cysteine-rich domains (Gorman et al., 2000, Danielli et al., 2000). Sp22D does not bind bacteria, and neither does its interaction with chitin induce autoactivation. The function of Sp22D remains unknown.

Cellular and humoral responses

Insect hemocytes are originated from lymph lobes. In adults, they mostly attach to interior tissues, such as fat body, muscles, and trachea. According to their morphology and biological functions, hemocytes in lepidopteran insects are classified into several groups (Lavine and Strand, 2002). Plasmatocytes, similar to vertebrate macrophages, engulf pathogens through phagocytosis. They secrete soluble plasma factors in response to infection. During cellular encapsulation, granulocytes encapsulate nematodes and filarial worms that are too large to be cleared by phagocytosis. Oenocytoids synthesize proPO for melanogenesis, a process often accompanying encapsulation or wound healing (Ashida and Brey, 1998).

Hemocytes associate with humoral immune responses in many ways. Several plasma proteins induce cellular reactions. For instance, hemolin is composed of four immunoglobulin domains that are most similar to those in neuronal cell adhesion molecules (Yu et al., 2002a). In *M. sexta*, hemolin concentration reaches 1.5 mg/ml at 24 h after a bacterial challenge. Hemolin reversibly binds to and aggregates bacteria and

hemocytes. Plasmatocyte-spreading peptides in lepidopteran insects cause plasmatocytes to spread and aggregate (Wang et al., 2001a). As described earlier, *A. gambiae* TEP1, which is produced by hemocytes, attaches the parasitesurface via a thioester bond.

Series of humoral reactions are also initiated parallel to cellular reactions after infection (Gillespie et al., 1997). Syntheses of AMPs, proPO activation, and hemolymph clotting are known to be key reactions of insect humoral immunity. Hemolymph clotting seals off the wound and prevents entry of pathogens. This reaction involves the interaction between hemocytes and soluble clotting factors. In the crayfish *Pacifastacus leniusculus*, one subunit of clotting protein has been cloned. It is similar to vitellogenin in sequence. The cross-linking of clotting protein is mediated by a calcium-dependent enzyme, transglutaminase (TGase) (Hall et al., 1999). In the human blood clotting cascade, TGase participates in the polymerization of fibrin. In the horseshoe crab *Tachypleus tridentatus*, a branched SP cascade induces the hemolymph clotting (Iwanaga et al., 1998). LPS and β-1,3-glucan trigger autoactivation of clotting factor C and G, respectively. Proclotting enzyme is converted into an active proteinase due to a sequential activation. Then proclotting enzyme cleaves coagulogen to form insoluble coagulin. However, cross-linking of coagulins is not catalyzed by horseshoe crab TGase (Osaki and Kawabata, 2004). Recently, *Drosophila* larval hemolymph coagulation was examined (Karlsson et al., 2004; Scherfer et al., 2004). Putative clotting factors were isolated by comparative proteomic analysis. Yet, substrates of insect TGase are unknown.

ProPO activation system

Insect PO (monophenol, L-dopa: O_2 oxidoreductase; EC 1.14.18.1) is involved in

a series of physiological reactions including pigmentation, cuticle sclerotization, wound healing, and other defense responses. During these processes, PO catalyzes monophenols to o-diphenols and further oxidizes o-diphenols to o-quinones. Quinones are precursors of melanin (Ashida and Brey, 1998). The black pigment (containing polyphenols and proteins) immobilizes pathogens. Quinones may have antimicrobial activities (Nappi and Vass, 2001).

PO genes have been sequenced in many invertebrate species (Kawabata et al., 1995; Kwon et al., 1997; Yasuhara et al., 1995). There are three PO genes in *Drosophila* (Adams et al., 2000). In mosquitoes, melanotic encapsulation is a resistance mechanism against *Plasmodium* infection. Nine PO genes are present in the *A. gambiae* genome, seven located in a gene cluster on the chromosome 2L (Christophides et al., 2002). AgPO2, AgPO3, and AgPO9 expression is induced by blood feeding.

In *M. sexta*, proPO-1 and proPO-2 have been cloned (Hall et al., 1995; Jiang et al., 1997b). The purified proteins migrate at 78 and 80 kDa on SDS-polyacrylamide gels. Depending on ionic strength, proPO exists as monomers, dimers, trimers, or multimers (Jiang et al., 1997b). Phylogenetic analysis shows that POs are related to hemocyanins and hexamerins (Burmester et al., 2002). These proteins consist of two copper-binding domains. Lower arthropods (including horseshoe crabs, spiders, and crustaceans) use hemocyanin to transport oxygen. In insects, delicate tracheal systems serve the purpose of hemocyanin. Horseshoe crab hemocyanin is active as PO in the presence of amphiphilic substances such as SDS, chitin-binding AMPs, or hemocyte components (Adachi et al., 2003; Decker et al., 2001). These results suggest that POs and hemocyanins have evolved from an ancient copper-binding protein.

PO is present in the hemolymph of insect larvae as an inactive proenzyme, proPO. ProPO does not have a signal peptide, and little is known about how it moves out of oenocytoids. One hypothesis is that proPO is released by hemocyte rupture. ProPO is transported to cuticle through cuticular epithelium. Limited proteolysis of proPO is mediated by an extracellular SP cascade and PRRs (pattern recognition receptor) (Ashida and Brey, 1998; Gorman and Paskewitz, 2001). PAP, also known as proPO activating enzyme (PPAE), activates proPO by limited proteolysis. So far, components of the pathway have only been reported in three insect species, *B. mori, M. sexta,* and *Holotrichia diomphalia*.

Historically, PPAE is thought to activate proPO by itself In our group, we found that *M. sexta* PPAE may be composed of PAP and SPHs (serine proteinase homolog). We have isolated PAP-1 from cuticles (Jiang et al., 1998), PAP-2 and PAP-3 from hemolymph of the prepupae (Jiang et al., 2003a and 2003b). Alongwith SPH -1 and SPH-2, these PAPs can activate proPO by cleavage at Arg^{51} (Yu et al., 2003). SPHs have the same domain structure as PAP-1, but a Gly residue replaces the active site Ser in the proteinase-like domain. The domain structure of PAP-2 and PAP-3 are similar: there are two clip domains (instead of one in PAP-1) in the amino terminus followed by one catalytic domain in the carboxyl terminus.

Similar results were reported from *H. diomphalia*, a coleopteran insect. ProPO-activating factor I (PPAF-I) is a single clip domain SP like *M. sexta* PAP-1, and PPAF-II is an SPH similar to *M. sexta* SPH-1 or SPH-2 (Lee et al., 1998a; Lee et al., 1998b). In the presence of PPAF-II, proPO was cut at the Arg^{162} to form highly active 60 kDa PO by PPAH. In the absence of PPAF-II, PPAF-I cleaved proPO only at the first cleavage site

and generated 76 kDa inactive PO. PPAF III was identified as a 40 kDa easter-like SP that cleaves PPAF-II precursor to form an active cofactor (Kim et al., 2002).

Ashida and his colleagues found that silkworm PPAE does not need any cofactor to activate proPO (Satoh and Ashida, 1999). The second cleavage event is not needed for PO activity. The discrepancies in these mechanisms are not fully understood.

The clip-domain SPs and SPHs participate in the immune reactions extensively. In *Drosophila*, 147 SPs and 57 SPHs constitute the second largest gene family. One fifth of them contain at least one clip domain (Ross et al., 2003; Adams et al., 2000). A proteolytic pathway establishes the dorsal-ventral polarity during *Drosophila* embryonic development. Two clip-domain serine proteinases, Snake and easter, are involved in this process. In adults, Persephone is a cascade member that activates the Toll pathway (Ligoxygakis et al., 2002). This extracellular SP pathway leads to induced synthesis of AMPs (*e.g.* drosomycin). In *A. gambiae*, there are at least 41 clip-domain SPs. However, none of them were identified as an ortholog of *Drosophila* Persephone. CLIP14 and CLIP15 were transiently up-regulated during parasite invasion, indicating that both SPs are responsive to *Plasmodium* infection.

Clip domain was initially identified in proclotting enzyme from the horseshoe crab, *T. tridentatus* (Muta et al., 1990). Clotting factor B has the same domain structure (Iwanaga et al., 1998). Most clip-domain SPs consist of two parts and a linker (Jiang and Kanost, 2000). The length of clip domains ranges 37-55 amino acid residues knitted together by three disulfide bonds to form a compact structure. The clip-domain SPs are synthesized as zymogens and are activated by a specific proteolytic cleavage at the end of the linker. A pair of cysteine residues connects the two domains so that the two peptide

chains are still covalently attached. Untill now, structure and function of clip domains are largely unknown. The recombinant clip domain in crayfish PAP showed antibacterial activity to Gram-positive bacteria (Wang et al., 2001c). In *M. sexta*, SPHs and IML-2 may associate with other plasma factors (*e.g.* PAP and proPO) to form a macromolecular complex (Yu et al., 2003). The clip domains may also bind microorganisms to ensure a localized melanization reaction.

Many parasitic wasps inject eggs and venom into their host insects. Their eggs evade the surveillance of host defense system. Development of the wasp larvae *in vivo* consumes the host tissues and finally kills the host. Recently, Vn50 was purified from the venom of *Cotesia rubecula*, a hymenopteran endoparasitoid. This protein blocked the proPO activation system of its host (Asgari et al., 2003). Molecular cloning indicates that Vn50 is a clip domain SPH, yet it is not understood how this venom protein negatively impacts proPO activation.

Serpin structure, function, and evolution

Proteinases and proteinase inhibitors co-exist in biological systems to maintain homeostasis (Bode and Huber, 1992). In insect hemolymph, SP inhibitors are present at high concentrations to prevent excessive proteolysis. So far, several groups of inhibitors have been identified in invertebrates (Kanost, 1999). These include: α-macroglobulin, low M_r SP inhibitors, and serpins. The serpins are a superfamily of proteins consisting of 370~450 amino acid residues. Over 500 serpins are identified in higher eukaryotes and viruses so far. They are divided into 16 clades and 10 highly diverged "orphans" in a phylogenetic analysis (Irving et al, 2000). Albeit low sequence similarities of 25-50%,

serpins share a common tertiary structure (Gettins, 2002).

Most serpins are irreversible inhibitors of SPs. In mammals, they regulate blood coagulation, fibrinolysis, angiogenesis, apoptosis, inflammation, and complement activation (Silverman et al, 2001). Serpin deficiencies cause fatal diseases including emphysema, cirrhosis, and dementia. Over fifty serpin crystal structures are known, including *M*. *sexta* serpin-1K – the first invertebrate serpin structure determined (Li et al., 1999). Serpins are globular proteins, composed of 9-13 α-helices and 3 β-sheets. They inhibit cognate proteinases by a suicide mechanism. The inhibitory selectivity of a serpin is determined primarily by a reactive center loop (RCL) exposed on the surface of the molecule. This loop serves as a pseudo-substrate to interact with its target proteinase and forms a reversible Michaelis complex (Ye et al., 2001). The RCL is then cleaved by the proteinase to form an acyl-enzyme complex. Due to the covalent linkage between the active site Ser residue of the proteinase and the P1 residue of the serpin RCL, the enzyme-inhibitor complex is SDS-stable. The amino-terminal side of the cleaved RCL incorporates into β-sheet A through a thermodynamically favored process, dragging the proteinase to the other end of the serpin. As the proteinase moves, its native structure is distorted and the catalytic activity is lost (Stratikos et al., 1999). One research pointed out that the length of RCL has a major effect on the stability of a serpin-proteinase complex – a long RCL destabilizes the complex (Zhou et al., 2001).

Serpins have been identified in *B. mori*, *M. sexta*, *L. migratoria*, *A. gambiae*, and *D. melanogaster* (Sasaki, 1991; Takagi et al., 1990; Narumi et al., 1993; Kanost et al., 1989; Kanost and Jiang, 1997; Gan et al., 2001; Danielli et al., 2003; Wang and Jiang, 2004). *M. sexta* serpin-1 through -5 have been cloned (Jiang et al., 1994; Gan et al., 2001;

Zhu et al., 2003; Tong and Kanost, 2005). Serpin gene-1 encodes 12 variants through alternative splicing (Jiang et al., 1994; Jiang et al., 1996). These serpins differ only in the RCL near the carboxyl terminus, which are encoded by exons 9A through 9K and 9Z. They were expressed in *E. coli* as recombinant proteins and characterized as inhibitors of different selectivities (Jiang et al., 1994; Jiang and Kanost, 1997a). Cytosolic serpin-2 has a low inhibitory activity against human cathepsin G, suggestive of a function distinct from the serpin-1 variants (Gan et al., 2001).

Mutualy exclusive exons are extensively used to encode serpins from invertebrates, including *D. melanogaster*, *A. gambiae*, *M. sexta*, *B. mori*, *M. configurata*, and *C. elegan*. *Drosophila* serpin-4 gene yields 10 alternatively spliced RNA transcripts and 8 protein variants in embryo (Kruger et al., 2002). Each variant has a different subcellular location or activity, suggesting that the regulation of gene expression is a complex process. *A. gambiae* serpin-10 gene encodes four intracellular inhibitors, expressed mainly in midgut, pericardial cells, and hemocytes (Danielli et al., 2003). Two of them, KRAL and RCM, are dramatically up-regulated in response to *P. berghei* infection. However, their target proteinases are unknown.

Immune pathways regulated by invertebrate serpins

Serpins appear to be major negative regulators of proPO activation systems (Kanost, 1999). *M. sexta* serpin-1J inhibits all the three PAPs by forming SDS-stable complexes (Jiang et al., 2003b). *H. cunea* serpin-3A was up-regulated after a bacterial challenge. The amino acid sequence of its RCL is strikingly similar to that of the proPO activation site (Park et al., 2000), suggesting that it may inhibit PAP. *M. sexta* serpin-3

has been demonstrated to block proPO activation in the hemolymph (Zhu et al., 2003). It has a k_{ass} of 7.5×10^5 $M^{-1}s^{-1}$ for PAP-1 and 6.9×10^5 $M^{-1}s^{-1}$ for PAP-3. The isolation of PAP-serpin-3 complexes from *in vitro* induced hemolymph strongly supports that serpin-3 could be a physiological regulator of proPO activation. *D. melanogaster* serpin27A is orthologous to *M. sexta* serpin-3 and *H. cunea* serpin-3A (Ligoxygakis et al., 2002b). Spontaneous melanization occurs in serpin27A-deficient mutant, suggesting that it inhibits PAP. Recently, *M. sexta* serpin-6 was identified as the third inhibitor that forms a covalent complex with PAP-3 (Wang and Jiang, 2004). It inhibits PAP-3 but not PAP-1. The biochemical mechanism for such selective inhibition is unclear.

During *Drosophila* embyronic development, the Toll pathway establishes dorsal-ventral axis. In the adults, this pathway, triggered by a different extracellular SP cascade, stimulates the synthesis of drosomycin in response to a fungal or Gram-positive bacterial infection (Hoffmann and Reichhart, 2002). A loss-of-function mutation in *spn43Ac* gene leads to necrosis and melanin deposition in the adult fly (Levashina et al., 1999; Green et al., 2000). Meanwhile, the Toll-mediated antifungal response is constitutively active. Spn43Ac seems to be a negative regulator of the proteolytic pathway and its expression is up-regulated after a challenge of Gram-positive bacteria and fungi.

In the *Drosophila* genome, there are fewer serpin genes (27) than SP genes (147) (Ross et al., 2003). It is possible one serpin may interact with several proteinases *in vivo*. On the other hand, more than one serpin may regulate key points of a proteolytic pathway. A thorough characterization of serpins may elucidate mechanisms for regulating proPO activation.

Induced expression of immune proteins

Since the discovery of cecropin (an AMP from *H. cecropia*), over 170 AMPs have been identified from different insect species (Bulet et al., 1999). Generally, cecropins are 35-39 amino acid residue long amphiphilic peptides. They could anchor, plug into, and form pores in the lipid cell membrane of bacteria and cause cell lysis. Insect defensins and drosomycin are Cys-rich polypeptides, too. The latter possesses antifungal activity. Another group of AMPs, including diptericin and drosocin, specifically attacks Gram-negative bacteria. Gambicin, a new type of AMPs identified in *A. gambiae*, contains 61 residues stabilized by four disulfide bridges. It possesses both antimicrobial and antiparasitic activities (Osta et al., 2004b).

Many AMPs are low M_r (<5 kDa molecular weight) molecules. They contain amphiphilic α-helices, hairpin-like β-sheets, and multiple disulfide bonds. AMP gene expression is strongly induced in fat body and epithelia after an immune challenge. Research showed that expression of most AMPs is controlled by the Toll and/or Imd pathway in *Drosophila* (Hoffmann and Reichhart, 2002). The Toll pathway controls drosomycin expression, whereas the Imd pathway specifically induces drosocin and diptericin synthesis. Production of cecropins, attacins and defensins are controlled by both pathways. Microarray analysis provides a global view of gene expression patterns during an immune response (Adams et al., 2000; Irving et al., 2001; De Gregorio et al., 2001; De Gregorio et al., 2002; Boutros et al., 2002).

The *Drosophila* Imd pathway is initiated upon infection by Gram-negative bacteria. Imd is a membrane receptor with a death domain homologous to tumor necrosis factor (TNF) receptor-interacting protein in mammals. Stimulation of Imd leads to the

proteolytic cleavage of a Rel family transcription activator, Relish. The released active domain is translocated into the nucleus to exert its function. death related CED-3/Nedd2-like death domain (DREDD), a caspase-8 homolog, and mitogen-activated protein 3 belong to this pathway (Hoffmann and Reichhart, 2002).

The Toll pathway, triggered by Gram-positive bacteria or fungi, is another major immune pathway (Hoffmann and Reichhart, 2002, Lemaitre et al., 1996). PGRP-SA binds PG of Gram-positive bacteria to initiate a SP pathway that leads to spätzle activation (Michel et al., 2001). The pathway is regulated by Spn43Ac (Levashina et al., 1999; Ligoxygakis et al., 2002a). Binding of active spätzle causes dimerization of the Toll receptor through interaction between leucine-rich repeats. The cytoplasmic domain of Toll interacts with a multimeric protein kinase complex including MyD88, Tube and Pelle. This leads to phosphorylation of Cactus and nuclear translocation of DIF and Dorsal. The Rel family transcription activators induce immune gene expression.

The Toll and Imd pathways are similar in many ways to the mammalian Toll like receptor/interleukin-1 (TLR/IL1) and TNF-R signaling pathways. This suggests innate immunity is evolutionarily conserved throughout the animal kingdom. Besides, JAK (Janus kinase)/STAT (signal transducers and activators of transcription) pathway also participates in the AMP synthesis through Relish in *A. gambiae* and *D. melanogaster* (Osta et al., 2004b). *Drosophila* JAK/STAT pathway is more involved in controlling genes of humoral opsonization factors (Boutros et al., 2002).

Role of hormones in the control of defense response

Juvenile hormones and ecdysteroids are major insect hormones. Coordination of

these molecules regulates molting, embryogenesis, ovary maturation, metamorphosis, and reproduction (Truman and Riddiford, 2002).

Ecdysteroids form complexes with the ecdysone receptor (EcR) dimer in the cell nucleus (Harshman, 1998). The dimer binds to *cis*-elements on DNA to activate target gene transcription. *Drosophila* EcRs differ in the N-terminal region. In *M. sexta*, orthologs of *Drosophila* EcR-B1, EcR-A and Ultraspiracle (USP) have been cloned. USP is a homolog of retinoid X receptor (RXR) in vertebrates. The transcription of EcR-B1 and EcR-A could be detected in the epidermis of *M. sexta*. EcR-B1 mRNA level increases at the beginning of each molt and peaks at the same time as ecdysteroid titer does. This indicates thatEcR -B1 may induce molting. EcR-A controls initiation of cuticle deposition (Jindra et al., 1996; Zhou et al., 1998). The 20-hydroxylecdysone (20E) stimulates expression of a retinoid orphan receptor homolog through interaction with complex EcR-B1-USP1, but not EcR-B1-USP2 (Lan et al., 1999).

Accumulating evidence suggests that insect hormones are also involved in the regulation of immune response genes. Starvation of *Rhodnius prolixius* compromised its immune system as a result of decrease in hormone level (Azambuja et al., 1997). ProPO activation was unaffected, but a reduction in hemocyte, cecropin, and lysozyme levels were observed. In *A. gambiae*, PO1 (AgPO1) gene expression responds to 20E. The promoter region of AgPO1 gene contains an ecdysone responsive element (EcRE), which binds to *Aedes aegypti*EcR/USP heterodimer and its endogenous heterodimer EcR/USP in nuclear extracts of adult *A. gambiae* (Ahmed et al., 1999). Microbial or parasitic challenge could not induce the expression of AgPO1. Muller et al. showed that 20E modulates proPO gene expression in *A. gambiae* cells (Dimarq et al., 1997).

Little is known about gene regulation of components in insect proPO activation system. Neither is it clear whether insect immune genes are controlled by hormonal signals. The expression level of six proPO genes did not significantly change in *A. gambiae* cells after a bacterial challenge, whileGNBP and defensin were up-regulated (Dimopoulos et al., 1997). A variation in proPO gene expression was not observed in the fall webworm, *Hyphantria cunae* after a bacterial challenge (Park et al., 1998). Neither was a change in PO transcription level detected in *Aedes subalbatus* after the mosquito was infected by the microfilaria worm, *Dirofilaria immitis* (Cho et al., 1998). AgproPO1 gene expression in 4a-3B cells did not increase after inoculation with *Wuchereria bancrofti* microfilaria (Dimarq et al., 1997). The irresponsiveness of AgproPO1 expression to microbes and parasites does not rule out its putative immune functions. ProPO is constitutively secreted as a zymogen in the hemolymph or cuticle during different developmental stages. Its transcriptional regulation could be different from other immune inducible genes (Dimopoulos et al., 1997; Dimopoulos et al., 2001).

Recently, there is a report indicating that ecdysteroids affect the cellular immune responses of *Drosophila*. Sorrentino and his colleagues found that loss of ecdysone compromised the encapsulation reaction (Sorrentino et al., 2002). Hemocyte maturation and proliferation were compromised in thelymph lobe of third-instar *ecdysoneless* larvae. These data provide new evidence for the involvement of ecdysone in innate immunity.

Transcription factors and regulatory elements for insect immunity and development

The interactions of *cis*-regulatory elements and specific transcription factors determine the immune gene expression (Engstrom, 1998). The first NF-κB activity was

detected in extracts from the induced *Hyalophora cecropia* pupae (Sun and Faye, 1992). A *H. cecropia* acute-phase transcription factor, CIF, was identified by mobility shift assays. Three NF-κB-like transcription factors were cloned from *Drosophila*, including Dorsal, DIF (dorsal-related immune factor), and Relish. Like other members of the Rel family (such as NF-κB), they form homo- or heterodimers in response to extracellular stimuli and regulate immune gene transcription. Dorsal and DIF belong to the Toll pathway and Relish is terminal component of the Imd pathway. Relish activation is an endoproteolytic process – an inhibitory sequence in Relish was cleaved by an intracellular proteinase. The NF-κB responsive elements are found at the 5' end of AMP genes in many insect species (Reichhart et al., 1992). The κB site was first found as an enhancer of the gene encoding for immunoglobulin κ light chain and many other mammalian genes involved in inflammatory responses. *Drosophila* NF-κB consensus sequence is GGGRAYYYYY. In *M. sexta*, a similar responsive element has been identified in the upstream of hemolin and lysozyme genes (Wang et al., 1995; Mulnix and Dunn, 1994), suggesting that NF-κB-like proteins may exist in lepidopteran insects for immune gene regulation. Similarly, another enhancer was identified in an intron of *H. cecropia* hemolin gene. It activates reporter gene expression through *Drosophila* DIF (Roxstrom-Lindquist et al., 2002).

GATA sequences, or GATA boxes, are commonly found in the 5' region of insect immune genes. The consensus sequence (WGATAA) was first identified in the chorion gene promoter in *B. mori* (Engstrom, 1998). This motif is recognized as a *cis*-regulatory element in many genes from vertebrates and invertebrates. In *Drosophila*, GATA box is important for maintaining the normal expression of *Cec*A1 gene in the presence of an

adjacent NF-κB motif (Kadalayil et al., 1997). *Drosophila* GATA-binding transcriptional factor Serpent is composed by an N- or C-terminal zinc finger domain and interacts with other proteins to induce the systematic expression ofimmune genes (Tingvall et al., 2001, Waltzer et al., 2002).

Cecropin genes in *Aedes albopictus*, *A. gambiae*, and *D. melanogaster* are arranged in a cluster (Ramos-Onsins and Aguade, 1998; Zheng and Zheng, 2002; Sun and Fallon, 2002). Genes in one orientation may share the same regulatory elements. Sequence at the promoter region of AgPO1 gene has NF-κB-like, GATA, and other immune regulatory motifs. Nevertheless, at least under the current experimental conditions, these elements do not seem to be functional (Ahmed et al., 1999).

Are insect immune genes also regulated by transcription factors other than NF-κB? Although some tissue and cell biological analyses suggest that immune gene expression are regulated by hormonal signals, no direct molecular biological evidence is available at present to support this hypothesis. While there are three PAP genes in *M. sexta*, their structures and expression profiles could be different. If so, how are these genes regulated at the transcriptional level? This problem needs to be addressed in order for us to understand function and regulation of the proPO activation system.

Literature cited

Adachi, K., Hirata, T., Nishioka, T., and Sakaguchi, M. (2003) Hemocyte components in crustaceans convert hemocyanin into a phenoloxidase-like enzyme. *Comp Biochem Physiol B Biochem Mol Biol* 134, 135-141.

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Govans, J.D., Amantides, P.G.,

Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F. et al. (2000) The genome sequence of *Drosophila melanogaster*. *Science* 287, 2187-2195.

Asgari, S., Zhang, G., Zareie, R., and Schmidt, O. (2003) A serine proteinase homolog venom protein from an endoparasitoid wasp inhibits melanization of the host hemolymph. *Insect Biochem Mol Biol* 33, 1017-1024.

Ahmed, A., Martin, D., Manetti, A.G., Han, S.J., Lee, W.J., Mathiopoulos, K.D., Muller, H.M., Kafatos, F.C., Raikhel, A., and Brey, P.T. (1999) Genomic structure and ecdysone regulation of the prophenoloxidase 1 gene in the malaria vector *Anopheles gambiae*. *Proc Natl Acad Sci USA* 96, 14795-14800.

Ashida, M. and Brey, P.T. (1998) Recent advances on the research of the insect prophenoloxidase cascade. In: Brey, P.T. and Hultmark, D. (Eds.), Molecular Mechanisms of Immune Responses in Insects. *Chapman and Hall*, London, pp. 135-172.

Azambuja, P., Garcia, E.S., Mello, C.B., and Feder, D. (1997) Immune responses in *Rhodnius prolixus*: influence of nutrition and ecdysone. *J Insect Physiol* 43, 513-519.

Blandin, S., Shiao, S.H., Moita, L.F., Janse, C.J., Waters, A.P., Kafatos, F.C., and Levashina, E.A. (2004) complement-like protein TEP1 is a determinant of vectorial capacity in the malaria vector *Anopheles gambiae*. *Cell* 116, 661-770.

Bode, W. and Huber, R. (1992) Natural protein proteinase inhibitors and their interaction with proteinases. *Eur J Biochem* 204, 433-451.

Boutros, M., Agaisse, H., and Perrimon, N. (2002) Sequential activation of signaling pathways during innate immune responses in *Drosophila*. *Dev Cell* 3, 711-722.

Bulet, P., Hetru, C., Dimarcq, J.L., and Hoffmann, D. (1999) Antimicrobial peptides in insects; structure and function. *Dev Comp Immunol* 23, 329-344.

Burmester, T. (2002) Origin and evolution of arthropod hemocyanins and related proteins. *J Comp Physiol* 172, 95-107.

Cherbas, L. and Cherbas, P. (1996) Molecular aspects of ecdysteroid hormone action, p. 175-221. In Gillbert, L.I., Tata, J.R., and Atkinson, B.G. (ed.). Metamorphosis: postembryonic reprogramming of gene expression in amphibian and insect cells. *Academic Press*, San Diego, Calif.

Choe, K.M., Werner, T., Stoven, S., Hultmark, D., and Anderson, K.V. (2002) Requirement for a peptidoglycan recognition protein (PGRP) in Relish activation and antibacterial immune responses in *Drosophila*. *Science* 296, 359-362.

Cho, W.L., Liu, H.S., Lee, C.H., Kuo, C.C., Chang, T.Y., Liu, C.T., and Chen, C.C. (1998) Molecular cloning, characterization and tissue expression of prophenoloxidase cDNA from the mosquito *Armigeres subalbatus* inoculated with *Dirofilaria immitis* microfilariae. *Insect Mol Biol* 7, 31-40.

Christophides, G.K., Zdobnov, E., Barillas-Mury, C., Birney, E., Blandin, S., Blass, C., Brey, P.T., Collins, F.H., Danielli, A., Dimopoulos, G., et al. (2002) Immunity-related genes and gene families in *Anopheles gambiae*. 2002. *Science* 298, 159-165.

Danielli, A., Loukeris, T.G., Lagueux, M., Muller, H.M., Richman, A., and Kafatos, F.C. (2000) A modular chitin-binding protease associated with hemocytes and hemolymph in the mosquito *Anopheles gambiae*. *Proc Natl Acad Sci USA* 97, 7136-7141.

Danielli, A., Kafatos, F.C., and Loukeris, T.G. (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.

Decker, H., Ryan, M., Jaenicke, E., and Terwilliger, N. (2001) SDS-induced phenoloxidase activity of hemocyanins from *Limulus polyphemus*, *Eurypelma californicum*, and Cancer magister. *J Biol Chem* 276, 17796-17799.

DeGregorio, E., Spellman, P.T., Rubin., G.M., and Lemaitre, B. (2001) Genome-wide analysis of the *Drosophila* immune response by using oligonucleotide microarrays. *Proc Natl Acad Sci USA* 98, 12590-12595.

DeGregorio, E., Spellman, P.T., Tzou, P., Rubin, G.M., and Lemaitre, B. (2002) The Toll and Imd pathways are the major regulators of the immune response in *Drosophila*. *EMBO J* 21, 2568-2579.

Dimarcq, J.L., Imler, J.L., Lanot, R., Ezekowitz, R.A., Hoffmann, J.A., Janeway, C.A., and Lagueux, M. (1997) Treatment of l(2)mbn *Drosophila* tumorous blood cells with the steroid hormone ecdysone amplifies the inducibility of antimicrobial peptide gene expression. *Insect Biochem Mol Biol* 27, 877-886.

Dimopoulos, G., Richman, A., Muller, H.M., and Kafatos, F.C. (1997) Molecular immune responses of the mosquito *Anopheles gambiae* to bacteria and malaria parasites. *Proc Natl Acad Sci USA* 94, 11508-11513.

Dimopoulos, G., Muller, H.M., Levashina, E.A., and Kafatos, F.C. (2001) Innate immune defense against malaria infection in the mosquito. *Curr Opin Immunol* 13, 79-88.

Dunn, P. and Drake, D. (1983) Fate of bacteria injected into naïve and immunized larvae of the tobacco hornworm, *Manduca sexta*. *J Invert Pathology* 41, 77-85.

Eggleston, P., Lu, W., and Zhao, Y. (2000) Genomic organization and immune regulation of the defensin gene from the mosquito, *Anopheles gambiae*. *Insect Mol Biol* 9, 481-490.

Engstrom, Y. (1998) Recent advances on the research of the insect prophenoloxidase cascade. In: Brey PT, Hultmark D. (Eds.), Insect immune gene regulation. *Chapman and Hall,* London, pp.211-244.

Gan, H., Wang, Y., Jiang, H., Mita, K., and Kanost M.R. (2001) A bacteria-induced, intracellular serpin in granular hemocytes of *Manduca sexta. Insect Biochem Mol Biol* 31, 887-898.

Gettins, P.G. (2002) Serpin structure, mechanism, and function. *Chem Rev* 102, 4751-4804.

Gillespie, J.P., Kanost, M.R., and Trenczek, T. (1997) Biological mediators of insect immunity. *Annu Rev Entomol* 42, 611-643.

Gottar, M., Gobert, V., Michel, T., Belvin, M., Duyk, G., Hoffmann, J.A., Ferrandon, D., and Royet, J. (2002) The *Drosophila* immune response against Gram-negative bacteria is mediated by a peptidoglycan recognition protein. *Nature* 416, 640-644.

Gorman, M.J., Andreeva, O.V., and Paskewitz, S.M. (2000) Sp22D: a multidomain serine protease with a putative role in insect immunity. *Gene* 251, 9-17.

Gorman, M.J. and Paskewitz, S.M. (2001) Serine proteases as mediators of mosquito immune responses. *Insect Biochem Mol Biol* 31, 257-262.

Green, C., Levashina, E., McKimmie, C., Dafforn, T., Reichhart, J.M., and Gubb, D. (2000) The necrotic gene in *Drosophila* corresponds to one of a cluster of three serpin transcripts mapping at 43A1.2. *Genetics* 56, 1117-1127.

Hall, M., Scott, T., Sugumaran, M., Soderhall, K,. and Law, J.H. (1995) Proenzyme of *Manduca sexta* phenol oxidase: purification, activation, substrate specificity of the active enzyme, and molecular cloning *Proc Natl Acad Sci USA* 92, 7764-7768

Hall, M., Wang, R., van Antwerpen, R., Sottrup-Jensen, L., and Soderhall, K. (1999) The crayfish plasma clotting protein: a vitellogenin-related protein responsible for clot formation in crustacean blood. *Proc Natl Acad Sci USA* 96, 1965-1970.

Hoffmann, J.A. and Reichhart, J.M. (2002) *Drosophila* innate immunity: an evolutionary perspective. *Nat Immunol* 3, 121-126.

Irving, J.A., Pike, R.N., Lesk, A.M., and Whisstock, J.C. (2000) Phylogeny of the serpin superfamily: implications of patterns of amino acid conservation for structure and function. *Genome Res* 10, 1845-1864.

Irving, P., Troxler, L., Heuer, T.S., Belvin, M., Kopczynski, C., Reichhart, J.M., Hoffmann, J.A., and Hetru, C. (2001) A genome-wide analysis of immune responses in *Drosophila*. *Proc Natl Acad Sci USA* 98, 15119-15124.

Iwanaga, S., Kawabata, S., and Muta, T. (1998) New types of clotting factors and defense molecules found in horseshoe crab hemolymph: their structures and functions. *J Biochem* (Tokyo). 123, 1-15.

Janeway, C.A. Jr. and Medzhitov, R. (2002) Innate immune recognition. *Annu Rev Immunol* 2002, 197-216.

Iwanaga, S., Kawabata, S., and Muta, T. (1998) New types of clotting factors and defense molecules found in horseshoe crab hemolymph: their structures and functions. *J Biochem (Tokyo)* 123, 1-15.

Ji, C., Wang, Y., Ross, J., and Jiang, H. (2003) Expression and in vitro activation of

Manduca sexta prophenoloxidase-activating proteinase-2 precursor (proPAP-2) from baculovirus-infected insect cells. *Protein Expr Purif* 29, 235-243.

Jiang, H., Wang, Y., and Kanost, M.R. (1994) Mutually exclusive exon use and reactive center diversity in insect serpins. *J Biol Chem* 269, 55-58.

Jiang, H., Wang, Y., Huang, Y., Mulnix, A.B., Kadel, J., Cole, K., and Kanost, M.R. (1996) Organization of serpin gene-1 from *Manduca sexta*. Evolution of a family of alternate exons encoding the reactive site loop. *J Biol Chem* 271, 28017-28023.

Jiang, H. and Kanost, M.R. (1997a) Characterization and functional analysis of 12 naturally occurring reactive site variants of serpin-1 from *Manduca sexta*. *J. Biol. Chem* 272, 1082-1087.

Jiang, H., Wang, Y., Ma, C., and Kanost, M.R. (1997b) Subunit composition of pro-phenol oxidase from *Manduca sexta*: molecular cloning of subunit ProPO-P1. *Insect Biochem Mol Biol* 27, 835-850.

Jiang, H., Wang, Y., and Kanost, M.R. (1998) Pro-phenoloxidase activating proteinase form an insect, *Manduca sexta*: a bacterica-inducible protein similar to Drosophila east. *Proc Natl Acad Sci USA* 95, 12220-12225.

Jiang, H. and Kanost, M.R. (2000) The clip-domain family of serine proteinases in arthropods. *Insect Biochem Mol Biol* 30, 95-105.

Jiang, H., Wang, Y., Yu, X.Q., and Kanost, M.R. (2003a) Prophenoloxidase-activating proteinase-2 from hemolymph of *Manduca sexta*: a bacteria-inducible serine proteinase containing two clip domains. *J Biol Chem* 278, 3552-3561.

Jiang, H., Wang, Y., Yu, X.Q., Zhu, Y., and Kanost, M. (2003b) 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.

Jiang, H., Ma, C., Lu, Z.Q., and Kanost, M. (2004) β-1,3-Glucan recognition protein-2 (βGRP-2) from *Manduca sexta*: an acute-phase protein that binds β-1,3-glucan and lipoteichoic acid to aggregate fungi and bacteria and stimulate prophenoloxidase activation. *Insect Biochem Mol Biol* 34, 89-100.

Jindra, M., Malone, F., Hiruma, K., and Riddiford, L.M. (1996) Developmental profiles and ecdysteroid regulation of the mRNAs for two ecdysone receptor isoforms in the epidermis and wings of the tobacco hornworm, *Manduca sexta*. *Dev Biol* 180, 258-272.

Kadalayil, L., Petersen, U.M., and Engstrom, Y. (1997) Adjacent GATA and kappa B-like motifs regulate the expression of a *Drosophila* immune gene. *Nucleic Acids Res* 25, 1233-1239.

Kanost, M.R., Prasad, S.V., and Wells, M.A. (1989) Primary structure of a member of the serpin superfamily of proteinase inhibitors from an insect, *Manduca sexta*. *J Biol Chem* 264, 965-972.

Kanost, M.R., Prasad, S.V., Huang, Y., and Willott, E. (1995) Regulation of serpin gene-1 in *Manduca sexta*. *Insect Biochem Mol Biol* 25, 285-291.

Kanost, M.R. and Jiang, H. (1997) Serpins from an insect, *Manduca sexta*. *Adv. Exp. Med. Biol* 425, 155-161.

Kanost, M.R. (1999) Serine proteinase inhibitors in arthropod immunity. *Dev. Comp. Immunol* 23, 291-301.

Karlsson, C., Korayem, A.M., Scherfer, C., Loseva, O., Dushay, M.S., and Theopold, U.

(2004) Proteomic analysis of the *Drosophila* larval hemolymph clot. *J Biol Chem* 279, 52033-52041.

Kim, M.S., Baek, M.J., Lee, M.H., Park, J.W., Lee, S.Y., Soderhall, K., and Lee, B.L. (2002) A new easter-type serine protease cleaves a masquerade-like protein during prophenoloxidase activation in *Holotrichia diomphalia* larvae. *J Biol Chem* 277, 39999-40004.

Kruger, O., Ladewig, J., Koster, K., and Ragg, H. (2002) Widespread occurrence of serpin genes with multiple reactive centre-containing exon cassettes in insects and nematodes. *Gene* 293, 97-105.

Kawabata, T., Yasuhara, Y., Ochiai, M., Matsuura, S., and Ashida, M. (1995) Molecular cloning of insect pro-phenol oxidase: a copper-containing protein homologous to arthropod hemocyanin. *Proc Natl Acad Sci USA* 92, 7774-7778.

Kwon, T.H., Lee, S.Y., Lee, J.H., Choi, J.S., Kawabata, S., Iwanaga, S., and Lee, B.L. (1997) Purification and characterization of prophenoloxidase from the hemolymph of coleopteran insect, *Holotrichia diomphalia* larvae. *Mol Cells* 28, 90-97.

Lavine, M.D. and Strand, M.R. (2002) Insect hemocytes and their role in immunity. *Insect Biochem Mol Biol* 32, 1295-1309.

Lan, Q., Hiruma, K., Hu, X., Jindra, M., and Riddiford, L.M. (1999) Activation of a delayed-early gene encoding MHR3 by the ecdysone receptor heterodimer EcR-B1-USP-1 but not by EcR-B1-USP-2. *Mol Cell Biol* 179, 4897-4906.

Lee, S.Y., Kwon, T.H., Hyun, J.H., Choi, J.S., Kawabata, S.I., Iwanaga, S., and Lee, B.L. (1998a) In vitro activation of pro-phenol-oxidase by two kinds of pro-phenol-

oxidase-activating factors isolated from hemolymph of coleopteran, *Holotrichia diomphalia* larvae. *Eur J Biochem* 254, 50-57.

Lee, S.Y., Cho, M.Y., Hyun, J.H., Lee, K.M., Homma, K.I., Natori, S., Kawabata, S.I., Iwanaga, S., and Lee, B.L. (1998b) Molecular cloning of cDNA for pro-phenol-oxidase-activating factor I, a serine protease is induced by lipopolysaccharide or 1,3-beta-glucan in coleopteran insect, *Holotrichia diomphalia* larvae. *Eur J Biochem* 257, 615-621.

Lee, W.J., Lee, J.D., Kravchenko, V.V., Ulevitch, R.J., and Brey, P.T. (1996) Purification and molecular cloning of an inducible gram-negative bacteria-binding protein from the silkworm, *Bombyx mori*. *Pro Natl Acad Sci USA* 93, 7888-7893.

Lemaitre, B., Nicolas, E., Michaut, L., Reichhart, J.M., and Hoffmann, J.A. (1996) The dorsoventral regulatory gene cassette spatzle/Toll/cactus controls the potent antifungal response in *Drosophila* adults. *Cell* 86, 973-983.

Levashina, E.A., Langley, E., Green, C., Gubb, D., Ashburner, M., Hoffmann, J.A., and Reichhart, J.M. (1999) Constitutive activation of toll-mediated antifungal defense in serpin-deficient *Drosophila*. *Science* 285, 1917-1919.

Levashina, E.A., Moita, L.F., Blandin, S., Vriend, G., Lagueux, M., and Kafatos, F.C. (2001) Conserved role of a complement-like protein in phagocytosis revealed by dsRNA knockout in cultured cells of the mosquito, *Anopheles gambiae*. *Cell* 104, 709-718.

Li, J., Wang, Z., Canagarajah, B., Jiang, H., Kanost, M.R., and Goldsmith, E.J. (1999) The structure of active serpin 1K from *Manduca sexta*. *Structure* 7, 103-107.

Ligoxygakis, P., Pelte, N., Hoffmann, J.A., and Reichhart, J.M. (2002a) Activation of

Drosophila Toll during fungal infection by a blood serine protease. *Science* 297, 114-116.

Ligoxygakis, P., Pelte, N., Ji, C., Leclerc, V., Duvic, B., Belvin, M., Jiang, H., Hoffmann, J.A., and Reichhart, J.M. (2002b) A serpin mutant links Toll activation to melanization in the host defence of *Drosophila*. *EMBO J* 21, 6330-6337.

Ma, C. and Kanost, M.R. (2000) A beta1,3-glucan recognition protein from an insect, *Manduca sexta*, agglutinates microorganisms and activates the phenoloxidase cascade. *J Biol Chem* 275, 7505-7514.

Mellroth, P., Karlsson, J., and Steiner, H. (2003) A scavenger function for a *Drosophila* peptidoglycan recognition protein. *J Biol Chem* 278, 7059-7064.

Michel, T., Reichhart, J.M., Hoffmann, J.A., and Royet, J. (2001) *Drosophila* Toll is activated by Gram-positive bacteria through a circulating peptidoglycan recognition protein. *Nature* 414, 756-759.

Muller, H.M., Dimopoulos, G., Blass, C., and Kafatos, F.C. (1999) A hemocyte-like cell line established from the malaria vector *Anopheles gambiae* expresses six prophenoloxidase genes. *J Biol Chem* 274, 11727- 11735.

Mulnix, A.B. and Dunn, P.E. (1994) Structure and induction of a lysozyme gene from the tobacco hornworm, *Manduca sexta*. *Insect Biochem Mol Biol* 24, 271-281.

Muta, T., Hashimoto, R., Miyata, T., Nishimura, H., Toh, Y., and Iwanaga, S. (1990) Proclotting enzyme from horseshoe crab hemocytes. cDNA cloning, disulfide locations, and subcellular localization. *J Biol Chem* 265, 22426-22433.

Narumi, H., Hishida, T., Sasaki, T., Feng, D.F., and Doolittle, R.F. (1993) Molecular cloning of silkworm (*Bombyx mori*) antichymotrypsin. A new member of the

serin superfamily of proteins from insects. *Eur J Biochem* 214, 181-187.

Nappi, A.J. and Vass, E. (2001) Cytotoxic reactions associated with insect immunity. *Adv Exp Med Biol* 484, 329-348.

Osaki, T. and Kawabata, S. (2004) Structure and function of coagulogen, a clottable protein in horseshoe crabs. *Cell Mol Life Sci* 61, 1257-1265.

Osta, M.A., Christophides, G.K., and Kafatos, F.C. (2004a) Effects of mosquito genes on Plasmodium development. *Science* 303, 2030-2032.

Osta, M.A., Christophides, G.K., Vlachou, D., and Kafatos, F.C. (2004b) Innate immunity in the malaria vector *Anopheles gambiae*: comparative and functional genomics. *J Exp Biol* 207, 2551-2563.

Park, D.S., Shin, S.W., Kim, M.G., Park, S.S., Lee, W.J., Brey, P.T., and Park, H.Y. (1998) Isolation and characterization of the cDNA encoding the prophenoloxidase of fall webworm, *hyphantria cunea*. *Insect Biochem Mol Biol* 27, 983-992.

Park, D.S., Shin, S.W., Hong, S.D., and Park, H.Y. (2000) Immunological detection of serpin in the fall webworm, *Hyphantria cunea* and its inhibitory activity on the prophenoloxidase system. *Mol Cells* 10, 186-192.

Ramos-Onsins, S. and Aguade, M. (1998) Molecular evolution of the *cecropin* multigene family in *Drosophila*: functional genes vs. pseudogenes. *Genetics* 150, 157-171.

Ramet, M., Manfruelli, P., Pearson, A., Mathey-Prevot, B., and Ezekowitz, R.A. (2002) Functional genomic analysis of phagocytosis and identification of a *Drosophila* receptor for *E. coli. Nature* 416, 644-648.

Reichhart, J.M., Meister, M., Dimarcq, J.L., Zachary, D., Hoffmann, D., Ruiz, C., Richards, G., and Hoffmann, J.A. (1992) Insect immunity: developmental and

inducible activity of the *Drosophila* diptericin promoter. *EMBO J* 11, 1469-1477.

Reiser J.B., Teyton, L., and Wilson, I.A. (2004) Crystal structure of the *Drosophila* peptidoglycan recognition protein (PGRP)-SA at 1.56 A resolution. *J Mol Biol* 340, 909-917.

Robertson, A.S., Belorgey, D., Lilley, K.S., Lomas, D.A., Gubb, D., and Dafforn, T.R. (2003) Characterization of the necrotic protein that regulates the toll-mediated immune response in *Drosophila*. *J Biol Chem* 278, 6175-6180.

Ross, J., Jiang, H., Kanost, M.R., and Wang, Y. (2003) Serine proteases and their homologs in the *Drosophila melanogaster* genome: an initial analysis of sequence conservation and phylogenetic relationship. *Gene* 304, 117-131.

Roxstrom-Lindquist, K., Lindstrom-Dinnetz, I., Olesen, J., Engstrom, Y., and Faye, I. (2002) An intron enhancer activates the immunoglobulin-related Hemolin gene in *Hyalophora cecropia*. *Insect Mol Biol* 11, 505-515.

Sasaki, T. (1991) Patchwork-structure serpins from silkworm (*Bombyx mori*) larval hemolymph. *Eur J Biochem* 202, 255-261.

Shin, S.W., Park, S.S., Park, D.S., Kim, M.G., Kim, S.C., Brey, P.T., and Park, H.Y. (1998) Isolation and characterization of immune-related genes from the fall webworm, *Hyphantria cunea*, using PCR-based differential display and subtractive cloning. *Insect Biochem Mol Biol* 28, 827-837.

Satoh, D., Horii, A., Ochiai, M., and Ashida, M. (1999) Prophenoloxidase-activating enzyme of the silkworm, *Bombyx mori*. Purification, characterization, and cDNA cloning. *J Biol Chem* 274, 7441-7453.

Scherfer, C., Karlsson, C., Loseva, O., Bidla, G., Goto, A., Havemann, J., Dushay, M.S.,

and Theopold, U. (2004) Isolation and characterization of hemolymph clotting factors in *Drosophila melanogaster* by a pullout method. *Curr Biol* 14, 625-629.

Sorrentino, R.P., Carton, Y., and Govind, S. (2002) Cellular immune response to parasite infection in the *Drosophila* lymph gland is developmentally regulated. *Dev Biol* 243, 65-80.

Stratikos, E. and Gettins, P.G. (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 USA* 96, 4808-4813.

Sugumaran, M. (1996) Roles of the insect cuticle in immunity. In *New Directions in Invertebrate Immunology* (Soderhall K, Iwanaga S, and Vasta GR)

Sun, S.C. and Faye, I. (1992) *Cecropia* immunoresponsive factor, an insect immunoresponsive factor with DNA-binding properties similar to nuclear-factor kappa B. *Eur J Biochem* 204, 885-892.

Sun, D. and Fallon, A.M. (2002) Characterization of genomic DNA encoding cecropins from an *Aedes albopictus* mosquito cell line. *Insect Mol Biol* 11, 21-30.

Takagi, H., Narumi, H., Nakamura, K., and Sasaki, T. (1990) Amino acid sequence of silkworm (*Bombyx mori*) hemolymph antitrypsin deduced from its cDNA nucleotide sequence: confirmation of its homology with serpins. *J Biochem (Tokyo)* 108, 372-378.

Takehana, A., Katsuyama, T., Yano, T., Oshima, Y., Takada, H., Aigaki, T., and Kurata, S. (2002) Overexpression of a pattern-recognition receptor, peptidoglycan-recognition protein-LE, activates imd/relish-mediated antibacterial defense and

the prophenoloxidase cascade in *Drosophila* larvae. *Proc Natl Acad Sci USA* 99, 13705-13710.

Tingvall, T.O., Roos, E., and Engstrom, Y. (2001) The GATA factor Serpent is required for the onset of the humoral immune response in *Drosophila* embryos. *Proc Natl Acad Sci USA* 98, 3884-3888.

Truman, J.W. and Riddiford, L.M. (2002) Endocrine insights into the evolution of metamorphosis in insects. *Annu Rev Entomol* 47, 467-500.

Waltzer, L., Bataille, L., Peyrefitte, S., and Haenlin, M. (2002) Two isoforms of Serpent containing either one or two GATA zinc fingers have different roles in *Drosophila* haematopoiesis. *EMBO J* 21, 5477-5486.

Werner, T., Borge-Renberg, K., Mellroth, P., Steiner, H., and Hultmark, D. (2003) Functional diversity of the Drosophila PGRP-LC gene cluster in the response to lipopolysaccharide and peptidoglycan. *J Biol Chem* 278, 26319-26322.

Wang, Y. and 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.

Wang, Y., Jiang, H., and Kanost, M.R. (2001a). Paralytic and plasmatocyte spreading peptide from *Manduca sexta*. *Adv Exp Med Biol* 484, 373-375.

Wang Y, Jiang H, and Kanost MR. (2001b) Expression and purification of *Manduca sexta* proPO-activating proteinase precursor (proPAP) from baculovirus-infected Insect Cells. *Protein Exp Purif* 23, 328-337.

Wang, R., Lee, S.Y., Cerenius, L., and Soderhall, K. (2001c) Properties of the prophenoloxidase activating enzyme of the freshwater crayfish, *Pacifastacus*

leniusculus. Eur J Biochem 268, 895-902.

Wang, Y., Willott, E., and Kanost, M.R. (1995) Organization and expression of the hemolin gene, a member of the immunoglobulin superfamily in an insect, *Manduca sexta. Insect Mol Biol* 4, 113-123.

Yasuhara, Y., Koizumi, Y., Katagiri, C., and Ashida, M. (1995) Reexamination of properties of prophenoloxidase isolated from larval hemolymph of the silkworm *Bombyx mori. Arch Biochem Biophys* 320, 14-23.

Ye, S., Cech, A.L., Belmares, R., Bergstrom, R.C., Tong, Y., Corey, D.R., Kanost, M.R., and Goldsmith, E.J. (2001) The structure of a Michaelis serpin-protease complex. *Nat Struct Biol* 8, 979-983.

Yoshida, H., Kinoshita, K., Ashida, M., (1996) Purification of a peptidoglycan recognition protein from hemolymph of the silkworm, *Bombyx mori. J Biol Chem* 271, 13854-13860.

Yu, X.Q. and Kanost, M.R. (2000) Immulectin-2, a lipopolysaccharide-specific lectin from an insect, *Manduca sexta*, is induced in response to gram-negative bacteria. *J Biol Chem* 275, 37373-37381.

Yu, X.Q. and Kanost, M.R. (2002a) Binding of hemolin to bacterial lipopolysaccharide and lipoteichoic acid. An immunoglobulin superfamily member from insects as a pattern-recognition receptor. *Eur J Biochem* 269, 1827-1834.

Yu, X.Q., Zhu, Y.F., Ma, C., Fabrick, J.A., and Kanost, M.R. (2002b) Pattern recognition proteins in *Manduca sexta* plasma. *Insect Biochem Mol Biol* 32, 1287-1293.

Yu, X.Q., Jiang, H., Wang, Y., and Kanost, M.R. (2003) Nonproteolytic serine proteinase homologs are involved in prophenoloxidase activation in the tobacco hornworm,

Manduca sexta. Insect Biochem Mol Biol 33, 197-208.

Zhou, A., Carrell, R.W., and Huntington, J.A. (2001) The serpin inhibitory mechanism is critically dependent on the length of the reactive center loop. *J Biol Chem* 276, 27541-27547.

Zhou, B., Hiruma, K., Jindra, M., Shinoda, T., Segraves, W.A., Malone, F., and Riddiford, L.M. (1998) Regulation of the transcription factor E75 by 20-hydroxyecdysone and juvenile hormone in the epidermis of the tobacco hornworm, *Manduca sexta*, during larval molting and metamorphosis. *Dev Biol* 193, 127-138.

Zhu, Y., Wang, Y., Gorman, M.J., Jiang, H., and Kanost, M.R. (2003) *Manduca sexta* serpin-3 regulates prophenoloxidase activation in response to infection by inhibiting prophenoloxidase-activating proteinases. *J Biol Chem* 279, 26605-26611.

Zheng, X-L. and Zheng, L. (2002) Genomic organization and regulation of three *cecropin* genes in *Anopheles gambiae*. *Insect Mol Biol* 11, 517-525.
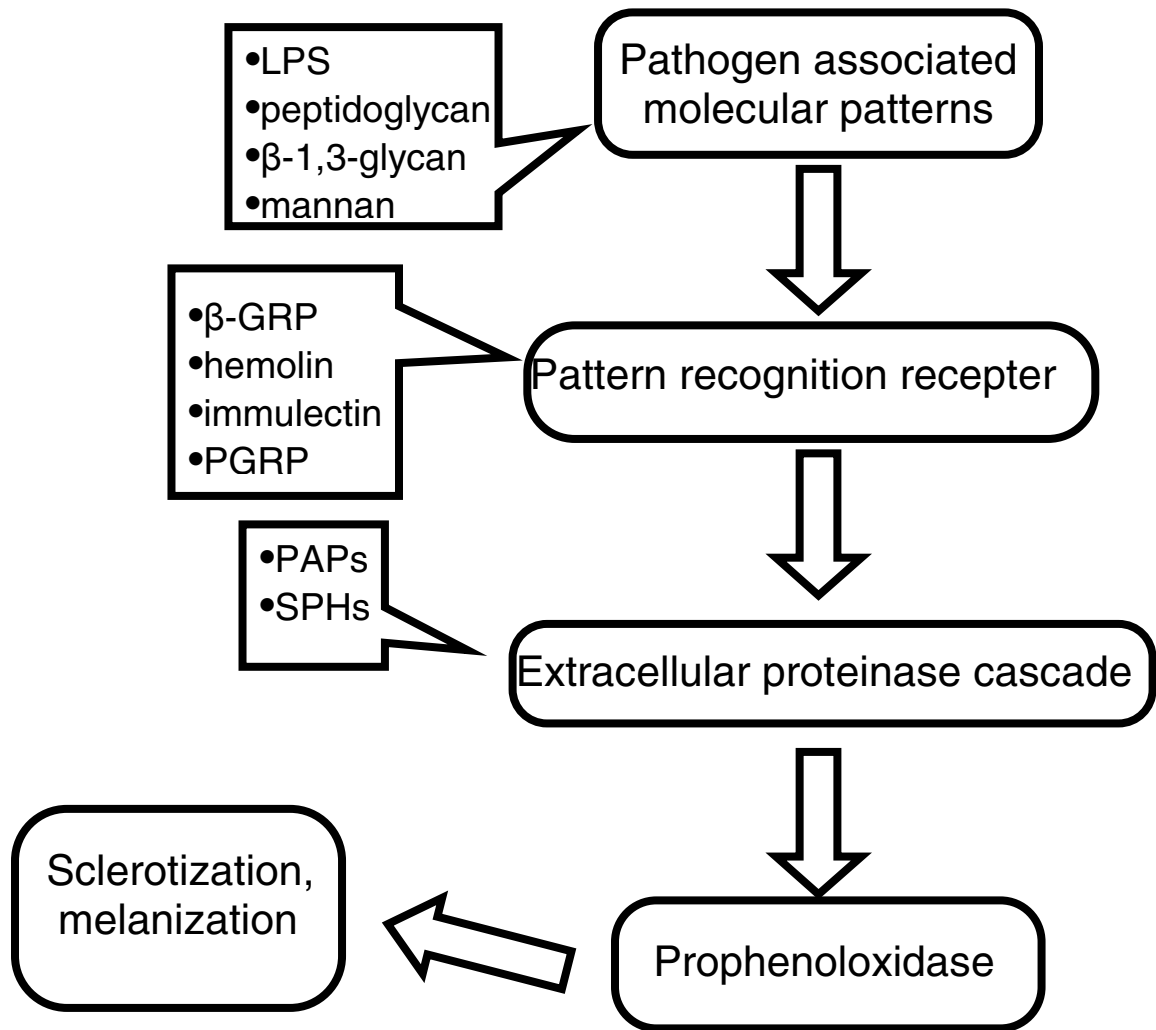
Fig. 1 Innate immunity in *Manduca sexta*

Chapter III. *Manduca sexta* proPO-activating proteinase-1 gene: organization, expression, and regulation by immune and hormonal signals

Abstract

Insect phenoloxidase (PO) participates in melanotic encapsulation, wound healing, and cuticle sclerotization. It is converted from prophenoloxidase (proPO) by a proPO-activating proteinase (PAP). *Manduca sexta* PAP-1, the final component of a serine proteinase cascade, cleaves proPO to generate active PO. In an effort to understand the transcriptional regulation, we isolated a genomic clone of the PAP-1 gene, determined its nucleotide sequence, and elucidated its exon-intron organization. Computer analysis revealed several immune and hormone responsive elements in the upstream region. Southern blot analysis suggested that the *M. sexta* genome contains a single copy of PAP-1 gene. Reverse transcription-polymerase chain reaction showed that PAP-1 was constitutively expressed in fat body, trachea, and nerve tissue of the fifth instar larvae. The mRNA levels in hemocytes and fat body markedly increased in response to a bacterial challenge. We also observed tissue-specific and developmental regulation of the gene's transcription. Treating *M. sexta* fat body culture with 20-hydroxyecdysone reduced the PAP-1 mRNA level. These data indicated that the expression of PAP-1 gene is under the dual control of immune and hormonal signals.

Introduction

Insect phenoloxidase (PO) hydroxylates tyrosine to dopa and oxidizes dopa to dopaquinone. Quinones can undergo cyclization, oxidation, and polymerization to generate eumelanin (Ashida and Brey, 1998). They may also crosslink proteins and polysaccharides to form hardened cuticle, seal off wounds, and encapsulate invading parasites. Reactive oxygen intermediates produced during the process can be cytotoxic to microbes as well as host tissues/cells (Nappi and Vass, 2001). Therefore, insect PO is produced as an inactive zymogen prophenoloxidase (proPO), which requires proPO-activating proteinase (PAP) for proteolytic activation. PAP resides at the end of a largely unknown serine proteinase pathway triggered upon recognition of microbial cell surface molecules, such as lipopolysaccharide, peptidoglycan, and β-1,3-glucan (Ashida and Brey, 1998). Serine proteinase inhibitors in the hemolymph regulate this enzyme system (Kanost, 1999).

PAPs have been purified and cloned from *Bombyx mori* (Satoh et al., 1999), *Holotrichia diomphalia* (Lee et al., 1998), *Manduca sexta* (Jiang et al., 1998, 2003a, and 2003b), and *Pacifastacus leniusculus* (Wang et al., 2001). There is at least one regulatory clip domain at the amino terminus of these enzymes. Clip domains, initially identified in the horseshoe crab proclotting enzyme (Muta et al., 1990), are frequently found in arthropod serine proteinases (SPs) and serine proteinase homologs (SPHs) (Jiang and Kanost, 2000; Ross et al., 2003; Christophides et al., 2003). We have purified three PAPs from *M. sexta* prepupae, which cleave proPO at Arg[51] and require two clip-domain SPHs to generate active PO. Like proPO, these clip-domain SPs and SPHs are produced in a

precursor form and need cleavage for activation. While the proPO activation system is beginning to be understood at the protein level, little is known about transcriptional regulation of the cascade components.

The interactions between *cis*-regulatory elements and transcription factors determine the expressions of insect immune genes (Harshman and James, 1998; Engstrom, 1998). *Cecropia* immune-responsive factor was first identified in nuclear extracts from the *Hyalophora cecropia* pupae challenged with bacteria (Sun and Faye, 1992). Three Rel-family transcription factors (Dorsal, Dif, and Relish) were identified from *Drosophila melanogaster* (Engstrom, 1998). Similar to NF-κB in mammals, these proteins may form homodimers or heterodimers in response to different stimuli and control the differential transcription of immune genes (Hoffmann, 2003). Dorsal and Dif belong to the Toll pathway, whereas Relish is a member of the Imd pathway. NF-κB responsive elements have been identified at the 5' end of antimicrobial peptide genes from many insects (Harshman and James, 1998; Engstrom, 1998). In *Drosophila,* the consensus sequence of these elements is GGGRAYYYYY (Hultmark, 1993). A similar responsive element has been identified upstream of *M. sexta* hemolin and lysozyme genes (Wang et al., 1995; Mulnix and Dunn, 1994), indicating that NF-κB-like proteins may also exist in lepidopteran insects to regulate immune gene expression. Recently, an enhancer was identified in an intron of the *H. cecropia* hemolin gene, which activates a reporter gene through *Drosophila* Dif (Roxstrom-Lindquist et al., 2002). Nevertheless, no NF-κB-like transcription activator has been documented in *M. sexta* so far.

GATA boxes are *cis*-regulatory elements commonly found at the 5' end of insect immune genes. Their consensus sequence, WGATAA, was first identified in the chorion

gene promoter region in *B. mori* (Harshman and James, 1998; Engstrom, 1998). In *Drosophila*, GATA box is important for maintaining the normal expression of cecropin A1 gene in the presence of an adjacent NF-κB motif (Kadalayil et al., 1997). *Drosophila* Serpent, a GATA-binding transcription factor with a zinc finger domain, interacts with other proteins to induce the expression of immune proteins during embryogenesis and hematopoiesis in the adults (Tingvall et al., 2001; Waltzer et al., 2002).

In arthropods, the molting hormone 20-hydroxyecdysone (20E) regulates many biological processes including molting, embryogenesis, metamorphosis, and reproduction (Riddiford et al., 2003). Evidence indicates a linkage between insect development and immune responses. Starvation of *Rhodnius prolixus* reduced hormone levels and compromised its immune system (Azambuja et al., 1997). Lack of ecdysteroids compromised the *Drosophila* cellular immune responses by negatively affecting hemocyte proliferation and encapsulation (Sorrentino et al., 2002). The promoter region of the *Anopheles gambiae* proPO1 gene contains two ecdysone-responsive elements (EcREs) (Ahmed et al., 1999). EcR/USP heterodimer prepared from nuclear extracts of adult *A. gambiae* can bind to one EcRE. Furthermore, Muller and his colleagues (1999) demonstrated 20E modulates proPO gene expressions in *A. gambiae* cell lines, whereas microbial and parasitic challenges did not affect *A. gambiae* proPO1 expression (Ahmed et al., 1999).

To understand the transcriptional regulation of the proPO activation system, we isolated the PAP-1 gene from a *M. sexta* genomic library and elucidated its exon-intron organization. We also identified immune- and ecdysone-responsive elements in the 5' flanking region of the gene. In addition, we examined the PAP-1 expression patterns in

different tissues and at different stages of development. The responsiveness of this immune gene to ecdysteroids was investigated as well.

Materials and methods

Insects

M. sexta larvae were hatched from eggs (Carolina Biological Supply) and reared as previously described (Dunn and Drake, 1983). Day 2 fifth instar larvae were injected with 2×10^8 *E. coli* cells suspended in 100 µl phosphate saline buffer. Hemolymph and fat body samples were collected 24 h after the bacterial challenge.

Library screening, subcloning, and DNA sequencing

The 5' *Pst*I-*Sma*I fragment of PAP-1 cDNA (Jiang et al., 1998) was labelled with ^{32}P-dCTP (3000 Ci/mmol) using Multiprime DNA Labelling System (Amersham Pharmacia Biotech). A *M. sexta* genomic DNA library in λGEM-11, kindly provided by Dr. Yucheng Zhu at Southern Insect Management Research Unit (USDA-ARS), was screened according to Sambrook and Russell (2001). After plaque purification and amplification, phage DNA was prepared using Wizard Lambda Preps DNA Purification System (Promega). The restriction enzyme map was determined by single and double digestions with *Xho*I, *Nco*I, *Xba*I, and *Sac*I. Digested DNA fragments were separated by agarose gel electrophoresis, transferred to a nitrocellulose membrane, and hybridized with the full-length cDNA probe. Fragments of the PAP-1 gene were subcloned into pBluescript-(KS) (Stratagene). Inserts from the resulting transformants were sequenced using BigDye v2.0 Terminator Cycle Sequencing Ready Reaction Kit (PE Applied

Biosystems). Sequences were edited and analyzed using MacVector Sequence Analysis Software (Version 6.5, Oxford Molecular Ltd.).

Primer extension

Primer extension was performed by following a standard protocol (Sambrook and Russell, 1999). Primer j394 (5'-CCA ATA AAC TGC AAA CAC AAT GAA CAC-3'), corresponding to the reverse complement of nucleotides 16-42 of the PAP-1 coding region, was terminally labeled with γ-^{32}P-ATP (3000 Ci/mmol) using T4 polynucleotide kinase. The primer (1.0×10^5 cpm) was added to 15 µg of total RNA, incubated at 60°C for 15 min, and then slowly cooled to 30ºC for 45 min. Annealed primer-RNA complexes were extended with MMLV reverse transcriptase (200 U/µl) at 42 °C for 1 h. The extension products were analyzed on 8% polyacrylamide gels containing 7 M urea, along with DNA size standards.

Southern blot Analysis

M. sexta genomic DNA was extracted from a single fifth instar larva. Aliquots of the DNA sample (15 µg each) were incubated with restriction enzymes (*Sal*I, *Sma*I, *Hind*III, and *Xba*I) at 37ºC for 6 h. The digested DNA samples were separated by electrophoresis on a 1% agarose gel and transferred to a GeneScreen Plus nitrocellulose membrane (NEN Life Science Products). The DNA blot was hybridized with ^{32}P-labelled PAP-1 cDNA as described by Sambrook and Russell (2001).

PAP-1 mRNA level in the cultured fat body

The fat body tissue was prepared according to Kanost et al (1995). Briefly, fat body tissues were dissected from day 2, fifth instar larvae, rinsed twice in SFM900II insect cell medium (Invitrogen Life Technologies), and transferred to separate wells of a tissue culture plate (Falcon). Each well contained 1 ml of the same medium supplemented with 10,000 unit/ml penicillin G and streptomycin (Sigma). 20E (1 mg/ml in ethanol) was added to the culture to a final concentration of 1.0 or 5.0 µg/ml. The control culture was treated with the same volume of the solvent. All cultures were incubated at 27ºC with shaking at 120 rpm. Total RNA samples, extracted from the cultured fat body after 48 h, were analyzed by RT-PCR.

RNA extraction and RT-PCR analysis

Total RNA samples were prepared from different tissues of *M. sexta* at various developmental stages using Micro-to-Midi Total RNA Purification System (Invitrogen Life Technologies). Similarly, fat body and hemocyte RNA samples were isolated from naive and bacteria-injected *M. sexta* larvae. First-strand cDNA was synthesized using total RNA (2-4 µg), oligo(dT)$_{17}$ (10 pmol), and MMLV reverse transcriptase (200 U, Invitrogen Life Technologies) at 37°C for 1 h. *M. sexta* ribosomal protein S3 (rpS3) transcripts were used as an internal standard to normalize the cDNA templates in a preliminary PCR experiment (Jiang et al., 2003b). Relative levels of PAP-1 cDNA in the samples (adjusted to contain an equal amount of rpS3 cDNA) were measured by semi-quantitative PCR using k628 (5'-GTC AAT ACA TAT CGC TGG TTG-3') and j386 (5'-ATC TCC CTT TAC GAG TGC CC-3'). The thermal cycling conditions were 94°C, 30 s; 50°C, 30 s; 72°C, 60 s, and the cycle numbers were chosen empirically to produce

comparable band intensities while avoiding saturation. After separation on a 1.3%
agarose gel electrophoresis, intensities of the PCR products were quantified and
compared using Kodak Digital Science 1D Gel Analysis Software.

Results

Structure of M. sexta PAP-1 gene

Using the 5' *Pst*I-*Sma*I fragment of PAP-1 cDNA as a probe, we screened
approximately 1 x 10^5 plaques and isolated one positive clone from the *M. sexta* genomic
library. Restriction mapping and Southern blot analysis revealed that a 17.3 kb *Sac*I-*Xho*I
fragment included the entire PAP-1 gene (Fig. 1, upper panel). We subcloned seven
restriction enzyme fragments into a plasmid vector and determined the complete
nucleotide sequence of *M. sexta* PAP-1 gene.

A comparison between the gene and cDNA sequences indicated that the PAP-1
gene is composed of 7 exons and 6 introns (Fig. 1, lower panel). Exon 1 encodes the first
17 amino acid residues of the signal peptide. Exon 2 codes for the last two residues of the
signal peptide and a complete clip domain. The subsequent 52 amino acid residues,
linking the clip and proteinase domains, are encoded by 3' end of exon 2, complete exon
3, and 5' end of exon 4. The remaining portion of exon 4, the entire exons 5 and 6, and
the 5' coding region of exon 7 encode the SP catalytic domain. The 3' untranslated region
(226 bp) of PAP-1 cDNA is represented by the 3' end of exon 7. We compared the 5' and
3' ends of the intron sequences and identified the consensus: 5'-GTRW(/A)R(/A)R(/G)T
W(/A)NWV(/A) and W(/T)NW(/T)W(/A)W(/T)HHHW(/T)Y(/T)W(T)B(/T)Y(/C)Y(/C)
AG-3', where the nucleotide in each parenthesis appears in 4~5 of the six sequences at

that position (Table 1). These two consensus sequences agree well with those deduced from the *M. sexta* serpin-1 gene (Jiang et al., 1996).

To determine the transcription initiation site of PAP-1 gene, we carried out a primer extension experiment and detected a 167 bp product (Fig. 2). Since the primer corresponds to the reverse complement of nucleotides 16-42 of the PAP-1 coding region (Fig. 3), transcription of the gene starts at the first C of CCATG – twenty nucleotides before the 5' end of the reported cDNA (Jiang et al, 1998). Another CCAGT is present between nucleotides -7 and -3. These two sequences closely resemble the 5-nucleotide consensus sequence (TCAGT), which typically resides within 10 nucleotides before or after the transcription start site in arthropod genes (Cherbas and Cherbas, 1993). We have identified a sequence between positions -31 and -26 (TATTTA), reminiscent of the TATA or Goldberg-Hogness box (TATAAA or TATATA). There is a perfect match further upstream (TATAAA, nucleotides -115 to -110), but it is too distant from the transcription start site (Fig. 3).

The exons are 98% identical in sequence to the PAP-1 cDNA isolated from a *M. sexta* larval hemocyte library (Jiang et al., 1998). Within the open reading frame, we have identified 20 nucleotide differences, most of which are synonymous substitutions. Two amino acid changes (R^2 to K^2, I^5 to T^5) occur in the signal peptide, and three others (A^{28} to V^{28}, A^{286} to V^{286}, H^{343} to N^{343}) are present in the mature protein.

To examine whether there is more than one copy of the *M. sexta* PAP-1 gene, we isolated the genomic DNA and carried out a Southern blot analysis (Fig. 4). The number, size and intensity of the hybridizing restriction fragments were consistent with those predicted from the restriction map (Fig. 1). The *Sal*I-digested genomic DNA generated at

least 4 fragments at 0.6, 0.9, 2.4, and 7 kb. *Hin*dII digestion gave rise to 5 hybridized fragments at about 0.7, 2.5, 3.4, 8 and >12 kb. *Sma*I cut the PAP-1 gene once but did not result in hybridizing bands smaller than 6 kb, whereas *Xba*I yielded five hybridizing bands at 0.5, 0.6, 1.2, 3.4 and >12 kb positions. No additional bands were observed when we used low-stringency conditions for hybridization and washing. These results indicate that the *M. sexta* genome may contain a single copy of the PAP-1 gene and that the nucleotide differences in the genomic sequences (Fig. 3) are most likely caused by allelic variations.

Putative regulatory elements in the PAP-1 gene

We discovered an inverted repeat in the first intron (nucleotides 3327-3404) (Fig. 3). This region, with 3 mismatches in the middle, may form a 78 bp stem-loop in the PAP-1 pre-mRNA or a cruciform in the double-stranded DNA of its gene. Conceivably, such a structure may influence RNA splicing of intron 1 (4.0 kb). The lack of mismatches in this long inverted repeat suggests that there is some selective pressure to preserve its unknown structure and function.

Computer analysis of the 5' flanking sequence allows us to locate potential regulatory elements in the PAP-1 gene (Table 2). Three NF-κB motifs are present at positions -432, -138, and -76, two of which are located on the plus strand. In comparison to the insect NF-κB consensus (Hultmark, 1993), there are 1 or 2 mismatches in these ten-nucleotide-long sequences. We have also identified three perfect GATA boxes at -618, -266, and -138. Like the NF-κB motifs, these conserved sequences frequently exist in the promoter regions of insect immune genes. There are two interferon-stimulated

response elements (ISREs) at positions -58 and -120. These two motifs, both located on the plus strand, contain 1 or 2 nucleotides that mismatch the 13-base mammalian ISRE consensus. Moreover, we have identified a putative EcRE at +1098 in the first intron. Like the 15-nucleotide consensus, this sequence (GGGT<u>A</u>CAGT<u>G</u>T<u>A</u>CCC) is a perfect inverted repeat – the two mismatches (double underlined) can still form a base pair.

Assuming that immune-responsive genes are regulated similarly in *M. sexta*, we compared the 5' flanking regions of PAP-1, lysozyme and hemolin genes to identify species-specific elements. The PAP-1 and hemolin sequences are 33% identical whereas the identity between PAP-1 and lysozyme sequences is 47%. We have identified 24 NF-κB, ISRE, and GATA elements in these sequences, nine of which have 0~1 mismatch as compared with the consensus motifs (Fig. 5A). However, we did not detect any regulatory elements in the corresponding regions of these sequences. The only exception was the TATA box, which is located at -29 to -31 of the three *M. sexta* immune genes.

When Genbank was searched with the PAP-1 intron sequences, we found that two *M. sexta* sequences contain a region closely similar to a 156-bp region at the end of intron 3 (Fig. 5B). These include juvenile hormone-binding protein and eclosion hormone genes. The sequence identities among these regions range from 80% to 86%. It is unclear whether these sequences have a regulatory function or not.

Transcription of the PAP-1 gene.

We employed RT-PCR to examine the PAP-1 expression profile in different tissues at development stages. In the control reactions, we amplified an expected 0.42 kb PCR product from normalized total RNA samples (Fig. 6C). Relative band intensities

indicated that PAP-1 mRNA levels in fat body and hemocytes of bacteria-injected larvae were significantly higher than those of the naïve insects. This result, consistent with our previous data from Northern blot analysis (Jiang et al., 1998) and prediction of the immune-responsive elements validated the RT-PCR method for estimating the transcript levels.

PAP-1 mRNA was detected in trachea, fat body, and the nerve cord from day 3, 5th instar naïve larvae (Fig. 6A). It was also abundantly present in integument and hemocytes from day 3, wandering larvae. Lower levels of the transcripts were identified in the other tissues at the same developmental time. In the early pupae (day 2), we detected the PAP-1 mRNA in integument, as well as midgut and fat body. Since the PAP-1 mRNA was detected in fat body samples at all these stages, we decided to closely monitor its level in fat body of *M. sexta* at different developmental stages.

We found that PAP-1 gene was transcribed in fat body throughout the 4th, 5th, and wandering stages (Fig. 6B). The fat body mRNA levels were higher in days 1-3 of the 5th instar and days 1-4 of the wandering larvae. In the fat body samples collected from the pupae, PAP-1 mRNA significantly decreased with time – it totally disappeared in the late pupae (day 30). The transcripts resurged in the adult fat body tissue.

The identification of a putative EcRE in intron 1 led us to test whether transcription of this immune gene is also regulated by ecdysteroids. We treated the cultured larval fat body with 20E at 0, 1, and 5 μg/ml and detected a concentration-dependent reduction in PAP-1 transcript level (Fig. 6D). While there was a significant decrease in the mRNA level at 1 μg/ml 20E, the transcripts almost completely disappeared at 5 μg/ml.

Discussion

So far, our knowledge of the transcriptional regulation of the insect proPO activation system is in a large part limited to proPO genes. Dimopoulos et al. (1997) found that the mRNA levels of six *A. gambiae* proPO did not significantly change after a bacterial challenge, while the production of GNBP and defensin was up-regulated. An infection of the microfilaria worm *D. immitis* failed to induce proPO gene expression in the mosquito *A. subalbatus* (Cho et al., 1998). Neither was a proPO level change observed in *A. gambiae* 4a-3B cells after an inoculation of *W. bancrofti* microfilaria (Dimarcq et al., 1997). This unresponsiveness to pathogen or parasite infection suggests that the regulation of proPO transcription is different from those of other acute-phase genes (Dimopoulos et al., 1997; Dimopoulos et al., 2001). We propose that several factors may contribute to such an anomaly: 1) proPO concentration in insect hemolymph is sufficiently high so further increase is unnecessary for an immune response against secondary infection; 2) proPO activation is regulated as a local reaction so only a small amount of active PO is generated near the site of invasion; 3) other PO-mediated physiological processes (*i.e.* molting) play more important roles in the control of the proPO expression.

The regulation of *M. sexta* PAP-1 transcription is different from that of the proPO expression. PAP-1 is an immune-responsive gene, whose transcripts and translation products increase significantly after a bacterial challenge (Fig. 5C and Jiang et al., 1998). Unlike *M. sexta* proPO, whose synthesis only occurs in oenocytoids (Jiang et al., 1997), PAP-1 transcripts were detected in several tissues/cells at different developmental stages

(Fig. 5A and 5B). Perhaps there is a more complex mechanism for regulating the PAP-1 expression.

PAP-1 and the proPO expression also appear to be related – ecdysteroids affect the transcription of both genes. In an *A. gambiae* cell-line, 20E treatment up-regulates the gene transcription of proPO1-proPO4 and proPO6 (Ahmed et al., 1999; Muller et al., 1999). In contrast, the proPO5 expression was suppressed in the cells after an exposure to 20E. RT-PCR analysis of their transcripts in the mosquito yielded a similar result: proPO1-proPO4 and proPO6 mRNA became more abundant after a blood meal whereas proPO5 mRNA decreased. Analogous to the mosquito proPOs, multiple PAPs are present in the tobacco hornworm. *M. sexta* PAP-1 gene is likewise under the control of 20E. It is known that 20E and juvenile hormone regulate the *M. sexta* metamorphosis and ovary maturation (Riddiford et al., 2003). There are two ecdysteroid peaks in the plasma of the late 5th instar and late wandering larvae. Consistent with the result from RT-PCR analysis of the cultured fat body (Fig. 5D), PAP-1 mRNA levels in these two periods were much lower than those in the early 5th instar and early wandering larvae (when the 20E titers were low) (Fig. 5B). The major ecdysone peak in the middle pupal stage also coincided with the drastic reduction of PAP-1 mRNA level in the fat body. Certainly, further experiments are needed to validate if the putative EcRE in intron 1 is indeed involved in the down-regulation of PAP-1 expression. Furthermore, it would be interesting to examine if the highly similar intron sequences in the PAP-1, juvenile hormone-binding protein, and eclosion hormone genes (Fig. 5B) play a role in their transcriptional regulation. In addition, immune and hormonal responsiveness and transcriptional profiles of *M. sexta* PAP-2 and PAP-3 genes are worth investigating.

To identify DNA sequences that might account for its immune responsiveness, we searched the 5' flanking sequence of the *M. sexta* PAP-1 gene. We found three sequences with similarity to the NF-κB consensus (Table 2 and Fig. 3). Three GATA boxes, two ISREs, and one 78-bp inverted repeat were identified. Further analyses are needed to test whether any of these putative elements are involved in the up-regulation of PAP-1 expression upon microbial infection.

PAP-1 gene, spanning nearly 13,000 nucleotides, consists of seven exons. Compared to the clip-domain SPs from *D. melanogaster* and *A. gambiae*, the exon number and gene length are significantly larger (data not shown). The clip domain in PAP-1 is entirely encoded by exon 2. In the putative *Drosophila* PAP genes (De Gregorio et al., 2001), the clip domain is also encoded by a single exon in SP4, SP7 and SP10 genes (Ross et al., 2003). These results suggest that the evolution of clip domains could result from exon shuffling. In contrast, the clip domain in *Drosophila* SP25 (another predicted PAP) is encoded by two exons. The clip domains in *Drosophila* SPs and SPHs are dissimilar in their sequences, except for the conserved Cys residues. These observations indicate that the evolutionary history of clip domains is complex and has to be examined individually.

Table 1   Intron sequences at the 5'and 3' splicing junctions in *M. sexta* PAP-1 gene

Intron	5' end	3' end
1	GTAAGTT	TCTTTCTTTTTGCAG
2	GTAAATA	TTAAAACATTTTTAG
3	GTATGTA	TATTATTCTCATCAG
4	GTAAGTA	ACTATCATTTTCCAG
5	GTAAGTA	TTTATATTTTTTCAG
6	GTGAGTA	TGAATAACATTTCAG
consensus*	GTRWRTW	WNWWWHHHWYWBYAG
predominant**	GTAAGTA	T-TAT---TTTTCAG

*: consensus sequences were determined at each position when a particular type of
nucleotide residue was present in all six positions. *R*: A,G; *Y* : C,T; *W*: A,T; *H*:
A,C,T; *B*: C,G,T; *V*: A,C,G; *N*: A,C,G,T.

**: predominant sequences were determined by the nucleotide residue present by
the most frequency at the specific position

Table 2   Sequence analysis of the 5' flanking region of *M. sexta* PAP -1 gene

Motif name and consensus*	Sequence found**	Location	Matched
NF-κB: GGGRAYYYYY	aGGtATTCTT(+)	-432	8/10
	aGGAATTTTT(-)	-138	9/10
	GcGAACgCCT(+)	-76	8/10
GATA: WGATAA	TGATAA(+)	-618	6/6
	AGATAA(+)	-266	6/6
	TGATAA(+)	-138	6/6
ISRE: GGAAANNGAAANN	GGAAATAtAAAAA(+)	-120	12/13
	GGAAcAAcAAAAC(+)	-58	11/13
EcRE: RRGKTCANTGAMCYY	GGGTaCAGTGtACCC(+)	1098	13/15

*: *R*: A,G; *Y*: C,T; *W*: A,T; *H*: A,C,T; *K*: G,T; *M*: A,C. *S*: C,G; *W*: A,T; *N*:
A,C,G,T.

**: nucleotides in the lower case do not match with the consensus.

References

Ahmed, A., Martin, D., Manetti, A.G., Han, S.J., Lee, W.J., Mathiopoulos, K.D., Muller, H.M., Kafatos, F.C., Raikhel, A., and Brey, P.T. (1999) Genomic structure and ecdysone regulation of the prophenoloxidase 1 gene in the malaria vector *Anopheles gambiae*. *Proc Natl Acad of Sci USA* 96, 14795-14800.

Ashida, M. and Brey, P.T., (1998) Recent advances on the research of the insect prophenoloxidase cascade. In: Brey, P.T., Hultmark, D. (Eds.), *Molecular Mechanisms of Immune Responses in Insects.* Chapman & Hall, London, pp. 135-172.

Azambuja, P., Garcia, E.S., Mello, C.B., and Feder, D. (1997) Immune responses in *Rhodnius prolixus*: influence of nutrition and ecdysone. *J Insect Physiol* 43, 513-519.

Cherbas L and Cherbas, P. (1993) The arthropod initiator: the capsite consensus plays an important role in transcription. *Insect Biochem Mol Biol* 23, 81-90.

Cherbas, L. and Cherbas, P. (1996) Molecular aspects of ecdysteroid hormone action. In: Gilbert, L.I., Tata, J.R., Atkinson, B.G. (Eds.), *Metamorphosis: postembryonic reprogramming of gene expression in amphibian and insect cells*. Academic Press, San Diego, pp. 175-221.

Cho, W.L., Liu, H.S., Lee, C.H., Kuo, C.C., Chang, T.Y., Liu, C.T., and Chen, C.C. (1998) Molecular cloning, characterization and tissue expression of prophenoloxidase cDNA from the mosquito *Armigeres subalbatus* inoculated with *Dirofilaria immitis* microfilariae. *Insect Mol Biol* 7, 31-40.

Christophides, G.K., Zdobnov, E., Barillas-Mury, C., Birney, E., Blandin, S., et al. (2002)

Immunity-related genes and gene families in *Anopheles gambiae*. *Science* 298, 159-165.

De Gregorio, E., Spellman, P.T., Rubin., G.M., and Lemaitre, B. (2001) Genome-wide analysis of the *Drosophila* immune response by using oligonucleotide microarrays. *Proc Natl Acad Sci USA* 98, 12590-12605.

Dimarcq, J.L., Imler, J.L., Lanot, R., Ezekowitz, R.A., Hoffmann, J.A., Janeway, C.A., and Lagueux, M. (1997) Treatment of l(2)mbn *Drosophila* tumorous blood cells with the steroid hormone ecdysone amplifies the inducibility of antimicrobial peptide gene expression. *Insect Biochem Mol Biol* 27, 877-886.

Dimopoulos, G., Richman, A., Muller, H.M., and Kafatos, F.C. (1997) Molecular immune responses of the mosquito *Anopheles gambiae* to bacteria and malaria parasites. *Proc Natl Acad Sci USA* 94, 11508-11513.

Dimopoulos, G., Muller, H.M., Levashina, E.A., and Kafatos, F.C. (2001) Innate immune defense against malaria infection in the mosquito. *Curr Opin Immunol* 13, 79-88.

Dunn, P. and Drake, D. (1983) Fate of bacteria injected into naïve and immunized larvae of the tobacco hornworm, *Manduca sexta*. *J Invert Pathol* 41, 77-85.

Engstrom, Y. (1998) Insect immune gene regulation. In: Brey, P.T., Hultmark, D. (Eds.), *Molecular Mechanisms of Immune Responses in Insects*. Chapman & Hall, London, pp. 211-244.

Harshman, L.G. and James, A.A., (1998) Differential gene expression in insects: transcriptional control. *Annu Rev Entomol* 43, 671-700.

Hoffmann, J.A. (2003). The immune response of *Drosophila*. *Nature* 426, 33-38.

Hultmark, D. (1993) Immune reactions in *Drosophila* and other insects: a model for

innate immunity. *Trends Genet* 9, 178-183.

Jiang, H., Wang, Y., Huang, Y., Mulnix, A.B., Kadel, J., Cole, K., and Kanost, M.R. (1996) Organization of serpin gene-1 from *Manduca sexta*: evolution of a family of alternate exons encoding the reactive site loop. *J Biol Chem* 271, 28017-28023.

Jiang, H., Wang, Y., and Kanost, M.R. (1998) Pro-phenoloxidase activating proteinase from an insect, *Manduca sexta*: a bacteria-inducible protein similar to *Drosophila* easter. *Proc Natl Acad Sci USA* 95, 12220-12225.

Jiang, H. And Kanost, M.R. (2000) The clip-domain family of serine proteinases in arthropods. *Insect Biochem Mol Biol* 30, 95-105.

Jiang, H., Wang, Y., Yu, X.Q., and Kanost, M.R. (2003a) Prophenoloxidase-activating proteinase-2 from hemolymph of *Manduca sexta*: a bacteria-inducible serine proteinase containing two clip domains. *J Biol Chem* 278, 3552-3561.

Jiang, H., Wang, Y., Yu, X.Q., Zhu, Y., and Kanost, M. (2003b) 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.

Kadalayil, L., Petersen, U.M., and Engstrom, Y. (1997) Adjacent GATA and kappa B-like motifs regulate the expression of a *Drosophila* immune gene. *Nucleic Acids Res* 25, 1233-1239.

Kanost, M.R., Prasad, S.V., Huang, Y., and Willott, E. (1995) Regulation of serpin gene-1 in *Manduca sexta*. *Insect Biochem Mol Biol* 25, 285-291.

Kanost, M.R. (1999) Serine proteinase inhibitors in arthropod immunity. *Dev Comp Immunol* 23, 291-301.

Lee, S.Y., Kwon, T.H., Hyun, J.H., Choi, J.S., Kawabata, S.I., Iwanaga, S., and Lee, B.L. (1998) In vitro activation of pro-phenol-oxidase by two kinds of pro-phenol-oxidase-activating factors isolated from hemolymph of coleopteran, *Holotrichia diomphalia* larvae. *Eur J Biochem* 254, 50-57.

Muller, H.M., Dimopoulos, G., Blass, C., and Kafatos, F.C. (1999) A hemocyte-like cell line established from the malaria vector *Anopheles gambiae* expresses six prophenoloxidase genes. *J Biol Chem* 274, 11727-11735.

Mulnix, A.B. and Dunn, P.E. (1994) Structure and induction of a lysozyme gene from the tobacco hornworm, *Manduca sexta*. *Insect Biochem Mol Biol* 24, 271-281.

Muta, T., Hashimoto, R., Miyata, T., Nishimura, H., Toh, Y., and Iwanaga, S. (1990) Proclotting enzyme from horseshoe crab hemocytes: cDNA cloning, disulfide locations, and subcellular localization. *J Biol Chem* 265, 22426-22433.

Nappi, A.J. and Vass, E. (2001) Cytotoxic reactions associated with insect immunity. *Adv Exp Med Biol* 484, 329-348.

Riddiford, L.M., Hiruma, K., Zhou, X., Nelson, C.A. (2003) Insights into the molecular basis of the hormonal control of molting and metamorphosis from *Manduca sexta* and *Drosophila melanogaster Insect Biochem Mol Biol* 33, 1327-1338.

Ross, J., Jiang, H., Kanost, M.R., and Wang, Y. (2003) Serine proteases and their homologs in the *Drosophila melanogaster* genome: an initial analysis of sequence conservation and phylogenetic relationship. *Gene* 304, 117-131.

Roxstrom-Lindquist, K., Lindstrom-Dinnetz, I., Olesen, J., Engstrom, Y., and Faye, I. (2002) An intron enhancer activates the immunoglobulin-related hemolin gene in *Hyalophora cecropia*. *Insect Mol Biol* 11, 505-515.

Satoh, D., Horii, A., Ochiai, M., and Ashida, M. (1999) Prophenoloxidase-activating enzyme of the silkworm, *Bombyx mori*: purification, characterization, and cDNA cloning. *J Biol Chem* 274, 7441-7453.

Sambrook, J. and Russell, D.W. (2001) Molecular cloning, a laboratory manual, third ed. Cold Spring Harbor Laboratory, New York.

Sorrentino, R.P., Carton, Y., Govind, S. (2002) Cellular immune response to parasite infection in the *Drosophila* lymph gland is developmentally regulated. *Dev Biol* 243, 65-80.

Sun, S.C., Faye, I. (1992) *Cecropia* immunoresponsive factor, an insect immuno-responsive factor with DNA-binding properties similar to nuclear-factor kappa B. *Eur J Biochem* 204, 885-892.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22, 4673-4680.

Tingvall, T.O., Ross, E., and Engstrom, Y. (2001) The GATA factor Serpent is required for the onset of the humoral immune response in *Drosophila* embryos. *Proc Natl Acad Sci USA* 98, 3884-2888.

Waltzer, L., Bataille, L., Peyrefitte, S., and Haenlin, M. (2002) Two isoforms of Serpent containing either one or two GATA zinc fingers have different roles in *Drosophila* hematopoiesis. *EMBO J* 21, 5477-5486.

Wang, R., Lee, S.Y., Cerenius, L., and Söderhäll, K. (2001) Properties of the prophenoloxidase activating enzyme of the freshwater crayfish, *Pacifastacus*

leniusculus. *Eur J Biochem* 268, 895-902.

Wang, Y., Willott, E., and Kanost, M.R. (1995) Organization and expression of the hemolin gene, a member of the immunoglobulin superfamily in an insect, *Manduca sexta*. *Insect Mol Biol* 4, 113-123.

Fig. 1 Structure of *M. sexta* PAP-1 gene and subcloning strategy. Upper panel: restriction map of the PAP-1 genomic insert in a positive λ bacteriophage. H, *Hind*III; N, *Nco*I; P, PstI; Sa, *Sal*I; Sc, *Sac*I; Sm, *Sma*I; X, *Xba*I; Xh, *Xho*I. The underlined recognition sites denote their presence in the cDNA. Horizontal bars mark the PAP-1 genomic fragments for subcloning. The dashed line indicates the region which is not sequenced. Lower panel: exon-intron organization of the PAP-1 gene. Numbered vertical bars: exons with the noncoding regions shown as open boxes. S, signal peptide; C, clip domain; L, linker region; P, proteinase domain. The same scale and starting site are used for the restriction map and exon-intron structure.

IF   A C G T
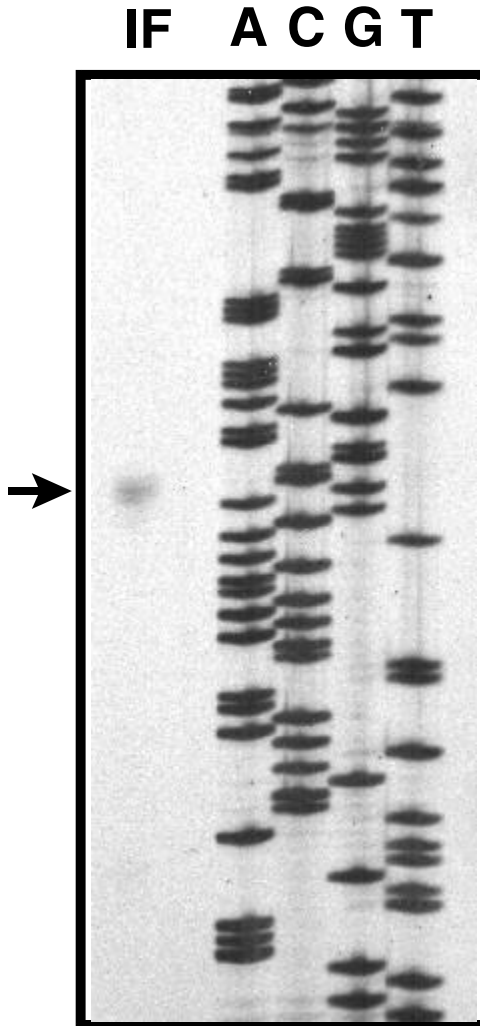
Fig. 2 Determination of the transcription initiation site in *M. sexta* PAP-1 gene. A primer, complementary to nucleotides 16-42 of the PAP-1 coding region at the 3' end of exon 1, was terminally labeled with γ-^{32}P-dATP and annealed to total RNA from fat body from bacteria-injected larvae (15 μg). After annealing to RNA, the primer was extended with reverse transcriptase. The set of sequencing reactions (ACGT) on the right of the primer extension lane for use as a sizing ladder were from dideoxynucleotide sequencing of single-stranded M13 mp18 DNA using -40 primer. The arrow indicates the 167 bp extension product from fat body RNA isolated from the induced larvae.

Fig. 3 Nucleotide sequence and structural features of *M. sexta* PAP-1 gene.
Nucleotides in the 5'-flanking region are assigned negative numbers. Nucleotide 1 is assigned based on the primer extension results (Fig. 2). Exon sequences are underlined with the encoded amino acid sequences listed below translated exons, using the one-letter code under the 2nd nucleotide of each codon. While some regions of the intron sequences (marked "---") are not shown, their sizes and positions are indicated. GATA boxes (6-nucleotide) and ISRE sites (13-nucleotide), bold and double underlined; NF-κB motifs (10-nucleotide) and EcRE (15-nucleotide), bold and single underlined; TATA boxes (6-nucleotide) and an imperfect inverted repeat (78 nucleotides), bold italic and double underlined. Mismatches in the repeat are marked with "▲". Single nucleotide polymorphic sites are in bold italic on the DNA sequence. Among them, nonsynonymous substitutions are further indicated on the affected amino acid residues (bold and underlined). The Cys residues in the clip domain and the catalytic residues in the SP domain are indicated with "+" and "#", respectively.

```
 -667   GATCACCTTCGCATTACATACATCATATTTATGTCATCATAGTTGTTAATTGATAACACTGTAATACGTATAATTTATTTAAGGATATTGTGAAGACCCGG

 -567   GTAAATCCATGATTCAGAGGGCGAGCTAAATCAATGCTATATACTAAGAAACTAGTTATACACAGTTATGTATGACGTAAGTACCGATAGCAAAACAGTT

 -467   TGCATAATTTGAGATTAAATTTAAATTTATTTTTAAGGTATTCTTATAATATTACGATGGTTCTTTTAGTTGATATTTACATTTATTTTTGTTATATTTC

 -367   TAAACATCATAAGTATTTATGTTGTAAAAATAACAAAACCGTTAAATCTTTTTTATATGGTAAATTGTTTCTATTAAAGTAAAAGTGTTCAAATATCGAG

 -267   GAGATAAAATAATATTTTAATACCTGGCACACTTTTAGACTAAAATATAATAAGTTCTGTTCGTAAAACAGGTAATTATAAATGTTACGTGTTAAAAACC

 -167   TCAACCCAAAAAAACATACTAAAAATTCCTGATAAGATTTCATCAACGGAAATATAAAAAATAAGAACACACAAAGGTATTAGGTTTTTATGCGAACGCC

                                                                -1 ┌ transcription initiation site
  -67   TACTTGATAGGAACAACAAAACGTATTTGTAAGAAGTATTTAACGTACCGACCTTGACCTCCAGTATCCAGTGTAGCAAAGCGGACGAGCGAGTCCAGCG

   34   CTCTTTAATTTATTTATATAAAAATATTTAAAAAATATTGCATAGAAAGTTGTTATTTATTAAGTTATGATATAGTGTTGAAGTGAACAAAATGAAGAA

                                                                                    Exon 1:  M  K  N
  134   TCACACTGTGTTCATTGTGTTTGCAGTTTATTGGACATGTGTCTGTAAGTTGTTGAGCCAGTTTTGTTTCAACTATTTTTTATGTTTTGAGGCGCCATAGC
           H  T  V  F  I  V  F  A  V  Y  W  T  C  V
  ---                                                          234-1033 (0.8 kb)
 1034   GTCTACTTAAACATCAAATATTCCATGGGAAAGAATAACCGATAAACAGCCTATATGTAATCCAGGGTACAGTGTACCCGTGTGAAAAATTTCATCCAAA
  ---                                                          1134-3233 (2.1 kb)            EcRE
 3234   AAATTATATTAGAACTATGAGGTTCTACAACAAATATAATGTCACATAGTATTTAATAACTGTTCTTAGTGGCTTTTTACTGTGTATGTGAAGTCTTGTC
 3334   TATATATATTGGACCTCACTTGCTTATATCTTCAGAGGTATAAGCAAGTGAGGTCCAATATATAGACAAGATATAGCCAAGCCTTATCTTACATCGACTC
  ---                   ▲  ▲▲▲▲  ▲                              3434-4133 (0.7 kb)
 4134   TCTACAATAACTCTTTCTTTTTGCAGTTTCACAGTTCATGTACAACACCACAAGGGGTCGATAGTAACTGTATCTCTCTTTACGAGTGCCCTCAATTACTG
                             Exon 2:  F  S ‖ Q  S  C +  T  T  P  Q  G  V  D  S  N  C +  I  S  L  Y  E  C +  P  Q  L  L
 4234   TCGGCTTTTTGAGCAACGGCCTCTGCCTAGCCCCGTTGTAAACTACTTGCGCAAGTCACAGTGTGGTTTCGATGGATATACTCCTCGTGTGTGCTGCGGAC
            S  A  F  E  Q  R  P  L  P  S  P  V  V  N  Y  L  R  K  S  Q  C +  G  F  D  G  Y  T  P  R  V  C +  C +  G
 4334   CCTTACCTCAACAGGCCAGCAGGCCGCAACCCACGGTAAATAAAAATGTTAATGAAAAATTTTCCTCACAATAAAAATACAATCTAGATGCAGCTGGTCA
            P  L  P  Q  Q  A  S  R  P  Q  P  T
  ---                                                          4434-5233 (0.8 kb)
 5234   TCATCAATCTAATGCTATCAAGATGATTCTCCTAATATACCCTTATTAAGGATTCTACTTCTGAAGTCGTATTGTAGTATACTAACAAGTTTAAAACATTT
                                                                                                      Exon
 5334   TTAGCCGGCCCCTGTACCAACACGCGCTCCACCTGTTAACCCGGGTGGTGTAGACCCAACTTATGATGAAGATTCATCACCAGCACCAAGGAATCAGTGC
         3:  P  A  P  V  P  T  R  A  P  P  V  N  P  G  V  D  P  T  Y  D  E  D  S  S  P  A  P  R  N  Q  C
 5434   GGTGTGGATATGAATGGAGACCGCATATACGGGGGACAAATCACAGACCTCGACGAGTTTCCATGGATGGCGTTGCTTGGATATTTGACTCGTATGTATT
            G  V  D  M  N  G  D  R ‖ I  Y  G  G  Q  I  T  D  L  D  E  F  P  W  M  A  L  L  G  Y  L  T
  ---                                                          5534-7633 (2.1 kb)
 7634   AGACTACGGTGTATAGTTCGCACTGGAAAACAACCTTGGGTTTTAATTAAGTTTGGTTGATACTAATTTTATTATTCTCATCAGGTACTGGCTCTACTAC
                                                                                    Exon 4:  R  T  G  S  T  T
 7734   GTATCAATGCGGAGGTGTGTTGATCAACCAGCGATATGTATTGACCGCCGCTCATTGTACTATTGGCGCTGTGGAGAGAGAGGTCGGGAAACTGTAAGTA
            Y  Q  C  G  G  V  L  I  N  Q  R  Y  V  L  T  A  A  H #  C  T  I  G  A  V  E  R  E  V  G  K  L
  ---                                                          7834-9433 (1.6 kb)
 9434   ATTTATTTACGATTATATCTTTTCATTTCAATTACTATCATTTTCCAGAATAACAGTGCGACTTGGAGAGTACGACACCCAAAACTCTGTGGACTGCGTA
                                  Exon 5:  I  T  V  R  L  G  E  Y  D  T  Q  N  S  V  D  C  V
 9534   GACGATGTTTGTGCTGACCCTCCACAAAACATTCCAATTGAAGTAGCCTACCCTCATTCAGGGTACAGTGACAATAACAAGAACAGGAAAGATGACATTG
            D  D  V  C  A  D  P  P  Q  N  I  P  I  E  V  A  Y  P  H  S  G  Y  S  D  N  N  K  N  R  K  D  D #  I
 9634   CCCTTGTTAGGCTTACGAGACGAGCTCAGTACACATGTAAGTATTCAATTGTAGCAACAGACTTAATGGCATAACTTGTTACCTGTTCGACCTTGAAATT
            A  L  V  R  L  T  R  R  A  Q  Y  T
  ---                                                          9734-11033 (0.3 kb)
11034   TGTTAACGCAGTCATTCGGTCGATAAACTCGAAACTCAATATTACCTTTTTTATATTTTTTCAGATTATGTTAAACCTATATGTCTGGCCAACAACAACG
                                            Exon 6:  Y  Y  V  K  P  I  C  L  A  N  N  N
11134   AGCGGCTAGCTACTGGCAATGACGTTTTCGTGGCCGGTTGGGGAAAAACCTTGAGCGGTGAGTATTATCGATTTTATTGCATTCCACGTTCACAGTTTTA
            E  R  L  A  T  G  N  D  V  F  V  A  G  W  G  K  T  L  S
  ---                                                          11234-11533 (0.3 kb)
11534   TCATTGCCTTGAACAAGATAAACATACACATTATTTTTAATTTCTGTGAACTGATTATATTTTTGTTGAATAACATTTCAGGCAAAAGCAGCCCAATTAA
                                                                       Exon 7:  G  K  S  S  P  I  K
11634   GTTAAGTTAGGTATGCCAATTTTCGACAAAAGCGACTGCGCATCAAAATACAGAAATTTAGGAGCAGAATTGACTGACAAACAGATATGCGCTGGTGGA
            L  K  L  G  M  P  I  F  D  K  S  D  C  A  S  K  Y  R  N  L  G  A  E  L  T  D  K  Q  I  C  A  G  G
11734   GTATTCGCCAAAGATACTTGCAGAGGTGACTCCGGCGGTCCTCTCATGCAGAGAAGACCAGAGGGAATCTGGGAGGTCGTAGGCATAGTATCCTTTGGCA
            V  F  A  K  D  T  C  R  G  D  S #  G  G  P  L  M  Q  R  R  P  E  G  I  W  E  V  V  G  I  V  S  F  G
11834   ATAGGTGTGGACTCGATGGATGGCCAGGCGTTTACTCATCAGTTGCGGGCTATTCTGACTGGATTCTATCTACACTCAGATCAACCAATGTGTGAATAGT
            N  R  C  G  L  D  G  W  P  G  V  Y  S  S  V  A  G  Y  S  D  W  I  L  S  T  L  R  S  T  N  V  *
11934   GTTAGACATTTGTTCTTCTTGTGTTGTGTCGTAAGTAGATTAATTAATGGTTGCTCTCTGACACGCGATAACTCAGTGATTGTGATAATCTTTGACGCTT
12034   TCTATGTTTACATGTTAGTTTTACGTTCTGTCTACTAGTTTGTTTTGTAAATAATTGAACTGATATGTAAATGCTCAATTACATCTTTCATCACGTGTAT
12134   TAAAAAACGTTATTTTTTATTATTTTTTTTGTTTCATTCATAACACTATTTCTTCTTTTATTAAGACGGAATGTAGTTCTTTCCAGAAAATCGTATAATAT
12234   ATTTCTTAATCAAATAAAGCCATAATTATGACTGCTCTCCATAATCGATGTCAAGGTGATTTTTTAATGTAAGAACTTAGATTTGTATAATTTTTTTAAG
12334   TAAAAGGTTTTTAAATTTAATTGTGCCAATGTAAAAATTTGTTTATTTACAAATAAAATA
```
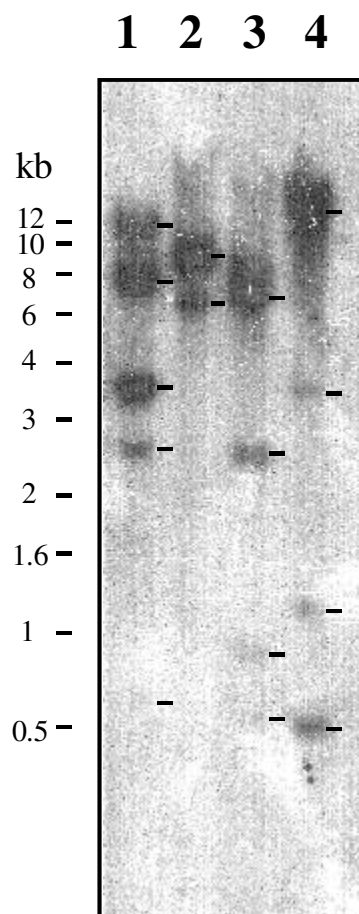
Fig. 4 Southern blot analysis of *M. sexta* genomic DNA using ³²P-labeled PAP-1 cDNA. Samples of the genomic DNA (15 µg) were digested with *Hid* III (lane 1), *Sma*I (lane 2), *Sal*I (lane 3), or *Xba*I (lane 4). After separation by agarose gel electrophoresis and transfer to a nitrocellulose membrane, the DNA fragments were hybridized with the probe. The positions and sizes of the molecular markers are indicated on the left side.

Fig. 5 Multiple sequence comparison. Panel A: 5'-flanking regions of *M. sexta* immune genes**.** The promoter regions of *M. sexta* lysozyme and hemolin genes (Mulnix and Dunn, 1994; Wang et al., 1995) were retrieved from Genbank and compared with the same region in the PAP-1 gene using ClustalW program (Tompson et al., 1994). Positions identical in all four sequences are marked with "*". GATA boxes (6-nucleotide, marked G) and ISRE sites (13-nucleotide, marked I), bold and double underlined; NF-κB motifs (10-nucleotide, marked N), bold and single underlined; TATA boxes (6-nucleotide), bold italic and double underlined. Mismatches are in lower case and the motifs with 0-1 mismatch are indicated. Panel B: a highly similar intron sequence in *M. sexta* PAP-1, juvenile hormone-binding protein (JHBP), and eclosion hormone (EH) genes. Based on a BLASTN search, similar sequences in *M. sexta* PAP-1 (nucleotides 6621-6777), JHBP (nucleotides 6096-6252, AF527636), and EH (nucleotides 3468-3322, M27808) genes were retrieved from Genbank and aligned. Positions identical in all three sequences are marked with "*".

A

```
                     -N                        N                    G(6/6)                              -I      -I(12/13)
lysozyme ATTATTTCC--ATTATAAAATGCATATTACCTATTTATATGTTATTTTTGGCAAATTATATTTGAAACAACTTTtTATTTtCAGCTCTTAGTAAGGTAT  98
hemolin  ATTAGtGGTCCgCACTAAATAAAAGTTTGcGAAaCCCTGT-TCTAGTAAGTACATCTTCTATCATAGACCGAA-GTTGTTTTTTTCTATTTtCAACATCA  98
PAP-1    ATCACCTTCGCATTACATACATCATATTATGTCATCATAGTTGTTAATTGATAACACTGTAATACGTATAA---TTTATTTAAGGATATTGTGAAGACCC  97
         **  *       *   *   *  **   *    *  *        *  * **    *      ** ***   *  *** *  **

lysozyme T---ATAGGTATTTTAGGTTTTATGTGTACATCGTTATTTTTTTAGAAAATACTATATTGAC-TAGTTACTTCTGGC--TCACTATTAGGTACAATATATG  193
hemolin  TTTCTATTAAAATCTACAATTAAACA--AAGAAAATAAGTTAAAATAGCaGAAAAAGcAATTTTAGTTTTTGTCACACCAGCTATGAGGCGGGGTATCAC  196
PAP-1    GTAAATCCATGATTCAGAGGGCGAGC-TAAATCAATGCTATATACTAAGAACATAGTTATACACAGTTATGTATGACGTAAGTACCGATAGCAAAACATT  196
              *  *          *     *    *       *   * * *      ****        *   **        *

                          -N            N                -N                    -I
lysozyme TTATGCATTCTTAAAAT-TATATATATCACATGATTCTATGCTATATCTTAAGATGAATAAAGGGGTaaCCATGGATTATTGTTCTTTCTATTTttA-TA  291
hemolin  GTACCCGAACTCAGTTTGCAAGCGTGTAACTCGAATGcAGGTTgC--CgCGACACTGTGATCACCCCACTATTACCTTCCACAACCCTGTAACTCCAGCT  294
PAP-1    TTGCATAAT-TTGAGAT-TAAATTTAA-ATTTATTTTTAaGGtAT-TCTTATAAT-ATTACGATGGTTCTTTTAGTTGATATTTACATTTATTTTTG-TT  290
              *          *        *    *      *  * * *    *        *     *     *  *           * ** *

lysozyme AATATGAGAATTTAAAATGtTCATTTTtCCAATATAAATAAAAAtTaCCATAGAACCTTGTGCTAGTAAGCCTACGATATTTAAAACATAATTGATTTAA  391
hemolin  ACTCAAAGAGACTATAAGTTTGAAAGTCAGGCCCAAAAAATATACGCCAGTAATTAACATGGGGGTATGACATACCCGCCACACGCTCGCCGGGTAAAA  394
PAP-1    ATATTTCTAAACATCATAAGTATTTATGTTGTAAAAATAACAAAACCGTTAAATCTTTTTTATATGGTAAATTGTTTCTATTAAAGTAAAAGTGTTCAAA  390
          *     *     *            *        *** ** *  *       *       *                *          * *  **

               G(6/6)                                                          N  -G(6/6)
lysozyme -TCATTAGCATG--TATTTATTTTTTCAGTGTAAGTTAGGTACTAATTAAAAGATTATTAATTAATAAGTCACTAGTCCATAGCCACAGTATTTAAAAACAA  488
hemolin  GTCTATAATATGGTTCTTTCGCTGTTGTAGTTTCTTGTCGTTTAATTGAACTCAGTAGCCCAATCaGGCAATTaTCTAATACAAAATGCTATTTTCGTAT  494
PAP-1    -TATCGAGGAGA--TAAAATAATATTTTA-ATACCTGGCAC--ACTTTTAGACTAAAATATAATAAGTTCT--GTTCGTAAAACAGGTAATTATAAATGT  482
          *   *   *  *        *    *       * *  *        * *    ***  **   *         *

                             I        -N(9/10) G(6/6)               I(12/13)            I
lysozyme ACAAAGACATTAATGTAATCTGGATAATACGCCTTCTGTGtAAAACcAAATCCCAACACTAATGTTATTAACAAAGCGCATCACtGAAACAtAAACAAAC  588
hemolin  TAGATTAAAGTGACTACAATAACCTTGTACGCAGC------AATTGTGAAGCGTAAGAC-ATTGCTAGCTACGGGACTCGGTTCATTACCTGAAGCCTGT  586
PAP-1    TACGTGTTAAAAACCTCAACCCAAAAAAACATACT------AA---AAATTCCtGATAA-GATTTCATCAACGGAA-ATAtAA--AAAATAAGAACACAC  569
              *   *          **           *     *  *      *  *   *  *    *  *         *

              N(9/10)           N                  I                                  N(9/10)
lysozyme AGATACGCaGGAACCCCCTTTTACAATACACTG--CAATACGAGGACAATAATCTATCTATGAGC-GCGTATAAAAGCGATCCATTGCGTTCAATTACTC  687
hemolin  TGACGCGTACAGCGTGGAACGAGCGGCGCGGCGCATGCGAGTCGCGACTCACTGTGTTGTGCAGGCGTATAAAAAGCCGCGCCTCGGGGACCTaCC  684
PAP-1    AAAGGTATTAGGTTTTTATGcGAACgCCTACT---TGATaGGAAcAAcAAAACGTATTTGTAAGAAGTATTTAACGTACCGACCTTGACCTCCAGTAT  666
          *           *            *    *     *  * *    ** *  **     *    *     * *
```

B

```
PAP-1  GGTATTCCACTGCGCTCGCCATCCTAAGACATGATATGTTAAGTCTTATTATGTATAGTAGTTTCACTGGCTACAATGT
JHBP   GGTATTCTACTGCGCTTGCCATTCTGAGTCATGAGATGTTAAGTCTTATTATGTCCTGTAGTTACATTGGCTACAGTGT
EH     GGTAATCCACTGCAATCGCCATCTCGAGATATGAGATGTTAAGTCTTATTACGTCCAGTAGTTACACTGGCTACAA---
       **** ** ***** * ***** ** **** ***************** **  ****** ** ** ********

PAP-1  TCTTCAAACCATAACACAACATTGACTACACACTGCTGCTTGGCGGCAGAAATAGACATTGCGGTGGTACCAACCCAG
JHBP   CCTTCAAACTGGAACACAACAGTGACTACACACTGCTACTTGGCGGCAGAAATAAACATTGCGGTGGTATCTACCCAG
EH     ------ACCGGAACACATCAGTGACTACACACTGTTGCTTGGCGGCAGAAATACACATTGCGGTGGTACCCACCCAG
       **  ****** ** ************* * **************** ** ************** * *****
```
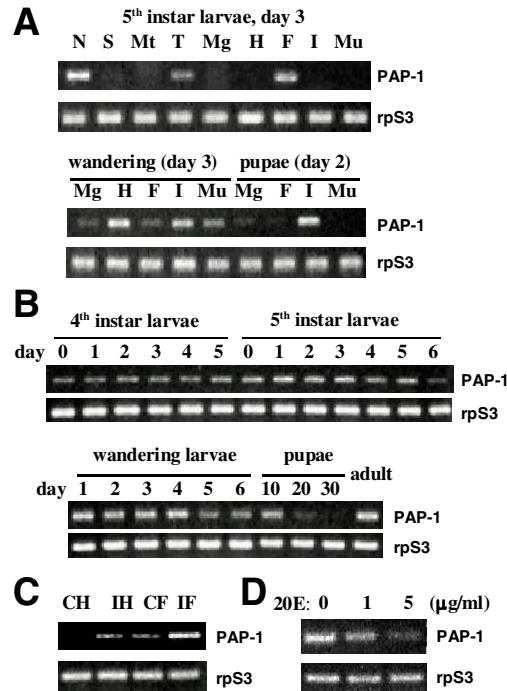
Fig. 6 RT-PCR analysis of *M. sexta* PAP-1 mRNA levels. Panel A: PAP expression in nerve tissues (N), salivary glands (S), Malpighian tubules (Mt), trachea (T), midgut (Mg), hemocytes (H), fat body (F), integument (I), and muscle (Mu) of the 5th instar larvae (day 3), wandering larvae (day 3) or early pupae (day 2). Panel B: changes of PAP-1 mRNA level in fat body from *M. sexta* at different developmental stages. Panel C: induced transcription of PAP-1 gene in hemocytes and fat body upon bacterial infection. CH and CF: hemocytes and fat body from the naïve larvae; IH and IF: hemocytes and fat body collected from the larvae at 24 h after injection of *E. coli*; Panel D: transcriptional suppression of PAP-1 gene in the cultured fat body by 20-hydroxyl ecdysone at different concentrations. *M. sexta* ribosomal protein S3 (rpS3) transcripts were normalized in all these analyses.

Chapter IV. Gene structure and expression profile of *Manduca sexta* prophenoloxidase-activating proteinase-3 (PAP-3), an immune protein containing two clip domains

Abstract

Prophenoloxidase-activating proteinase-3 (PAP-3) is a component of the defense system in *Manduca sexta*. We have isolated genomic clones and elucidated organization of this immune gene. The 3' end of exon 2, entire exon 3, and 5' end of exon 4 encode the two amino-terminal clip domains. Southern blot analysis suggested a single copy of PAP-3 gene in the genome. We identified several putative immune-responsive elements in the upstream region. PAP-3 gene is not highly expressed in fat body during larval development until the wandering stage begins. The mRNA level is high in the epithelium, fat body and hemocytes. Tissue-specific alternative splicing occurs in fat body and trachea. A bacterial injection markedly induced the gene expression in fat body and hemocytes.

Introduction

To survive and prosper in pathogen-rich environments, insects must possess an efficient defense system to prevent and control microbial infection (Gillespie *et al*., 1997). Sclerotized cuticle is effective in fending off infectious agents in the environment.

Plasma proteins, hemocytes and fat body initiate a series of humoral and cellular responses to immobilize and kill the invading microorganisms. Quinone production plays an important role in many of these immune mechanisms, including cuticle sclerotization, melanin synthesis, wound healing, encapsulation, and free radical-mediated pathogen killing (Nappi & Vass, 2001). Phenoloxidase (PO), which catalyzes quinone formation, is synthesized as an inactive zymogen prophenoloxidase (proPO) and activated by a cascade of serine proteinases (SPs), known as the proPO activation system (Ashida & Brey, 1998).

Analogous to the blood clotting and complement systems in vertebrates, the proPO activation cascade is triggered upon recognition of physiological or pathological signals. In the presence of microbial cell wall components (*e.g.* peptidoglycan, lipopolysaccride, and β-1,3-glucan), circulating or cell surface receptors specifically bind to these pathogen-associated molecular patterns and trigger autoactivation of the first proteinase precursor. This, in turn, leads to sequential cleavage activation of other cascade members. To date, our understanding of this pathway is largely limited to the final step, namely proPO activation by its activating enzyme PAP. PAP has been isolated and cloned from *Manduca sexta*, *Bombyx mori*, *Holotrichia diomphalia,* and *Pacifastacus leniusculus* (Jiang *et al*., 1998, 2003a, and 2003b; Satoh *et al*., 1999; Lee *et al*., 1998; Wang *et al*., 2001). In *M. sexta* and *H. diomphalia*, PAP converts proPO to active PO only when clip-domain serine proteinase homologs (SPHs) are present.

The clip domain is a structural unit stabilized by three disulfide bonds. It is thought to have regulatory functions (Jiang & Kanost, 2000). While other PAPs contain one clip domain at the amino terminus, *B. mori* proPO activating enzyme, *M. sexta* PAP-

2 and PAP-3 have two – the role of an extra clip domain is not understood. Many arthropod SPs and SPHs contain clip domains. However, little is known about their structure, function, and evolution. There is no report on the exon-intron organization and expression regulation of PAP genes.

With three terminal proteinases, the proPO activation system in *M. sexta* appears to be quite complex. A bacterial challenge significantly increased the mRNA and protein levels of PAP-1, PAP-2, and PAP-3 (Jiang *et al.*, 1998, 2003a, and 2003b). Their tissue localization may also change at different developmental stages. For instance, PAP-1 is present in the prepupal cuticle whereas PAP-2 and PAP-3 accumulate in the hemolymph of the same stage insect. Profiling these PAP mRNAs in various tissues and cells should allow us to better understand the transcriptional regulation of these genes in a developmental context. In this paper, we report the exon-intron organization and expression profile of *M. sexta* PAP-3 gene.

Materials and methods

Insect rearing and bacterial injection

M. sexta larvae were hatched from eggs (Carolina Biological Supply) and reared according to Dunn & Drake (1983). For bacterial induction, day 2 fifth instar larvae were injected with formalin-killed *Escherichia coli* XL1-blue (2×10^8 cells/larva). Hemolymph and fat body samples were collected 24 h later for RNA preparation.

Library screening, subcloning, sequencing and computer analysis

To screen the genomic library, full-length or PCR-derived PAP-3 cDNA was

labelled with α-^{32}P-dCTP using Multiprime DNA Labelling System (Amersham Pharmacia Biotech). The intact cDNA was retrieved from the PAP-3 clone by restriction enzyme digestion, whereas the 5' PCR fragment (115 bp) was amplified using primers j662 (5'-CGA CGC TGA GGT AAC ACG T-3', nucleotides 2-20) and j663 (5'-TTG TCC ACT AAC AAA GCA G-3', reverse complement of nucleotides 98-116). A *M. sexta* genomic library, generously provided by Dr. Yucheng Zhu at Southern Insect Management Research Unit (USDA-ARS), was screened with the cDNA probes at 5 x 10^5 cpm/ml by a standard protocol (Sambrook & Russell, 2001). Positive plaques were purified to homogeneity by repeated screening, and λ DNA was prepared from the amplified bacteriophages using Wizard Lambda Preps DNA Purification System (Promega). Restriction maps were determined by single and double digestions of the DNA samples with *EcoR*I, *Sac*I, and *Xba*I. Following agarose gel electrophoresis and capillary transfer, the separated digestion products on the nitrocellulose membrane were hybridized with the cDNA probes and visualized by autoradiography. PAP-3 gene fragments were subcloned into the same sites in pBluescript-(KS). The resulting recombinant plasmids were sequenced using BigDye v2.0 Terminator Cycle Sequencing Ready Reaction Kit (PE Applied Biosystems). Sequence editing, assembly, and analysis were performed using MacVector Sequence Analysis Software (Version 6.5, Oxford Molecular Ltd.).

Determination of PAP-3 transcription initiation site

As previously described (Sambrook & Russell, 2001), primer extension was carried out to determine the transcription starting position. An oligonucleotide (j396, 5'-

CCA CTA ACA AAG CAG AAA TAC GAA GCT AA-3'), corresponding to the reverse complement of nucleotides 28-56 of the PAP-3 coding region, was terminally labeled with γ-^{32}P-ATP using T4 polynucleotide kinase (Promega). The primer (1.0×10^5 cpm/ml) was added to 15 μg of total RNA, incubated at 60°C for 15 min, and then slowly cooled down below 30ºC in 45 min. Annealed primer-RNA complexes were extended with MMLV reverse transcriptase (200 U/μl, 1 μl) for 1 h at 42°C. The extension products as well as the ^{35}S-labeled DNA size standards were separated on 8% polyacrylamide gels containing 7 M urea and visualized by autoradiography.

Southern blot analysis

   M. sexta genomic DNA was extracted from a single fifth instar larva. Aliquots of the DNA sample (15 μg each) were incubated with restriction enzymes (*BamH*I, *EcoR*I, *Hind*III, *Sac*I, *Xba*I, *Sac*I-*Xba*I) at 37ºC for 6 h. After electrophoretic separation on a 1% agarose gel and capillary transfer onto GeneScreen-Plus nitrocellulose membrane (NEN Life Science Products), the digested genomic fragments on the membrane were hybridized with α-^{32}P-labelled full-length PAP-3 cDNA.

Immunocytochemistry

   At 1, 8, and 24 h after day 2 fifth instar larvae were injected with *E. coli* XL1-blue (1 x 10^7/larva), hemocytes were collected, washed, and fixed on multi-well slides as described by Willott *et al* (1995). These cells were incubated with 1:1000 diluted PAP-3 antiserum overnight at 4ºC. After removing excess antibodies with Tris-buffered saline, the hemocytes were reacted with 1:500 diluted FITC-conjugated goat-anti rabbit IgG

(Bio-Rad) at room temperature for 1 h. Following a washing step, cells were observed and photographed with phase contrast or fluorescence optics under an Olympus BH-2 microscope.

RNA extraction and RT-PCR analysis

RNA samples were extracted from various tissues of *M. sexta* at different developmental stages (see Fig. 5 legend for details), using Micro-to-Midi Total RNA purification system (Invitrogen Life Technologies). Similarly, fat body and hemocyte RNA samples were isolated from naive and bacteria-challenged *M. sexta* larvae. First-strand cDNA synthesis was performed using 2-4 μg RNA, 10 pmol oligo(dT)$_{17}$, and 200 U MMLV reverse transcriptase (Invitrogen Life Technologies) at 37°C for 1 h. *M. sexta* ribosomal protein S3 transcripts were used as an internal standard to control the template amount in a preliminary PCR experiment. Relative levels of PAP-3 cDNA in the normalized samples were determined by semi-quantitative PCR using j699 (5'-CGT GTT GTT ATT AGC TTC GTA TTT CT-3') and j700 (5'-CAT CCC CCC AGC CTC TAC-3' for PAP-3. The thermal cycling conditions were: 94°C, 30 s; 50°C, 30 s; 72°C, 60 s. PCR cycle numbers were empirically chosen to show comparable band intensity and avoid reaction saturation. After electrophoretic separation on a 1.3% agarose gel, intensities of the PCR products were quantified and compared using Digital Science 1D Gel Analysis Software (Kodak).

Results

PAP-3 genomic clones

A *M. sexta* genomic library was screened with the full-length PAP-3 cDNA and two positive clones, λ2 and λ3, were obtained (Fig. 1A). Sequence analysis indicated that a 5' cDNA fragment (115 bp) was absent in the genomic clones. Using the corresponding fragment derived from a PCR, we screened the genomic library again and isolated another positive clone, λ7. From these three bacteriophages, we isolated DNA, mapped the restriction sites, and subcloned 10 genomic fragments into a vector. The inserts in the resulting recombinant plasmids, 24.4 kb long in total, were completely sequenced.

Clone λ7, containing the 5' end of the PAP-3 gene, did not overlap with clone λ3 (Fig. 1A). Long-distance PCR using primers located at the 3' end of λ7 and 5' end of λ3 did not yield any PCR product (data not shown), indicating that the gap between the genomic clones is long. After the *BamH*I site near the 3' end of λ3 (Fig. 1A), there was a 167 bp sequence identical to the 5' end of λ2 (Fig. 1C). While this result strongly suggested these two genomic fragments overlap, we identified an extra 633 bp sequence at the 3' end of λ3 insert (Fig. 1C, clone 3-3). This fragment did not match λ2 sequence. To test whether or not it represents a cloning artifact, we amplified the genomic DNA using primers flanking the 167 bp overlap (Fig. 1A) and obtained a single 954 bp product. Cloning and sequence analysis indicated that the PCR product was identical to the overlap and flanking genomic sequences in λ2 and λ3, but does not contain the 633 bp fragment.

Exon-intron organization

A comparison between the genomic and cDNA sequences revealed that *M. sexta* PAP-3 gene is composed of 8 exons and 7 introns (Fig. 1B). Exon 1 includes a 5'

untranslated region and a coding region for the 19-residue signal peptide of PAP-3 (Fig. 2). Exon 2 and 5' end of exon 3 code for the first clip domain, whereas the rest of exon 3 and 5' end of exon 4 encode the second clip domain. The remaining part of exon 4 covers the linker sequence and the first two residues of the SP domain. Most part of the catalytic domain is encoded by exon 5, exon 6, and 5' end of exon 7. The other part of exon 7 (1040 bp) corresponds to the 3' untranslated region in the PAP-3 cDNA, which contains the polyadenylation signal (AATAAA) at ten nucleotides before the poly (A) tail.

Introns #1, #2 and #3 are >10, 2.6, and 0.8 kb in length, respectively, whereas the average size of introns #4-#7 is only 0.4 kb. We compared the 5' and 3' ends of these introns and identified the consensus: 5'-GTR(/G)W(/A)D(/G)K(/T) and B(/T)TY(/T)BC AG-3', where the nucleotide in each parenthesis appears in 5-6 out the seven sequences at that position (Table 1).

Transcription initiation, sequence variations and copy number

We determined the transcriptional initiation site of PAP-3 gene by primer extension. After annealing with RNA from fat body of bacteria-induced larvae, the primer (derived from nucleotides 28-56 of the PAP-3 coding region) was extended 96 nucleotides by reverse transcriptase (Fig. 3). Therefore, the RNA synthesis started at an A (nucleotide +1), and there was a TCAGT sequence at nucleotides -2 to +3. This motif is typically present within 10 nucleotides before or after the transcription initiation site in arthropod genes (Cherbas & Cherbas, 1993). We did not identify a TATA or Goldberg-Hogness box around the "-30 region". Nevertheless, a perfect TATA sequence (TATAAA) was present at nucleotides -94 to -89 (Fig. 2).

The exons were nearly identical in sequence to the PAP-3 cDNA clone isolated from the bacteria-induced fat body library (Jiang *et al*., 2003b). There are 9 nucleotide differences in the coding region, most likely caused by allelic variations. Two of these give rise to amino acid residue changes (T^{137} to S^{137}, Y^{383} to F^{383}).

Southern blot analysis was used to estimate the copy number of PAP-3 gene in the *M. sexta* genome (Fig. 4). The *BamH*I digest (Fig. 4, lane 1) contained a 6 kb fragment that hybridized with the cDNA probe before exon 4 (Fig. 1A). The >12 kb band with higher radioactivity probably included exons 4-8. A 2.3 kb *Hind*III fragment (Fig. 4, lane 2) matched the sequence between exon 4 and 5' end of exon 7 (Fig. 1) and produced a strong hybridization signal. The other three bands (> 10 kb) probably included exon 1, exons 2-3, and exons 7-8. With two recognition sites in the PAP-3 gene (Fig. 1A), *Sac*I yielded three high molecular weight bands that did not separate well (Fig. 4, lane 3). In contrast, *Xba*I generated a much smaller band at 3.1 kb (Fig. 4, lane 4), probably corresponding to exons 7 and 8. The other part of PAP-3 gene migrated to the ~11 kb position and hybridized with the probe much stronger. *EcoR*I digestion generated two strong bands at 2.4 and >10 kb and two weak bands at 2.7 and ~8 kb (Fig. 4, lane 5). Finally, *Sac*I and *Xba*I double digestion generated at least two bands (Fig. 4, lane 6). The lower one has the same mobility as the 3.1 kb *Xba*I fragment. The upper one (3.6 kb) strongly reacted with the cDNA probe in the region corresponding to exons 3 through 6. Longer exposure revealed three other bands at 5.5, 7 and 12 kb (data not shown). In summary, the number, size, and intensity of these radioactive restriction fragments agree well with those predicted from the restriction enzyme map of the cloned gene (Fig. 1A). There is only one copy of the PAP-3 gene in the *M. sexta* genome and the sequence

variations (Fig. 2) are likely due to allelic differences in the insects used for library constructions.

Putative regulatory elements

In order to identify putative regulatory elements, we searched the 1.8 kb upstream sequence of the PAP-3 gene by computer analysis (Table 2). Nine NF-κB motifs are present in this region, four on the plus strand and five on the minus strand. The motifs starting at nucleotides -1018 and -427 match 9 of 10 positions in the NF-κB consensus, originally derived from *Drosophila melanogaster* (Hultmark, 1993). A perfect match is located on the minus strand from -1756 to -1765. There are five interferon-stimulated response elements (ISREs) beginning at nucleotide -1724, -944, -292, -213, and -193. As compared with the 13-base mammalian ISRE consensus, all these sequences contain two mismatches. An ISRE was identified in *Anopheles gambiae* prophenoloxidase gene-1 (Ahmed et al., 1999)

Four GATA boxes exist at positions -1713, -1137, -536, and -145 on the plus strand. We have also identified three ecdysone response elements (EcREs) on the same strand at position -804, -291 and -6. While the GATA boxes perfectly match the consensus (WGATAA), three of the 15 nucleotides in the putative EcREs are not consistent with the motif (RRGKTCANTGAM CY) (Cherbas & Cherbas, 1996).

A 151 bp segment in the 5' upstream region is 85% identical in nucleotide sequence to an intron region in a *Bombyx mori* ABC transporter gene (Fig. 5A). It is unclear whether or not these sequences represent a lepidopteran repetitive sequence with a regulatory function. Using BLASTN, we searched Genbank database with the PAP-3

intron sequences and identified several high-score matches from *M. sexta* and other insects. For instance, a 68 bp fragment between exons 3 and 4 is 85-93% similar to intron sequences from *M. sexta* juvenile hormone binding protein, eclosion hormone, serpin-1, and arylphorin genes (Fig. 5B). The significance of these similar sequences is not understood.

Temporal and spatial regulation of transcription

To understand the regulation of PAP-3 transcription, we examined its mRNA levels in different tissues and development stages (Fig. 6). Semi-quantitative RT-PCR using PAP-3 specific primers resulted in PCR products at the expected sizes. Due to a 75 bp insert (Jiang *et al*., 2003b and data not shown), most of the PAP-3 transcripts in midgut gave rise to a 438 bp PCR product migrating slower than the 363 bp one from fat body, muscle, integument, salivary gland, and nerve tissue (Fig. 6A). Both transcripts were detected in hemocytes, trachea, and Malpighian tubules. There was no major change in relative abundances of the mRNAs between feeding and wandering larvae. Relative intensities indicated that PAP-3 transcripts were present at a low level in fat body and hemocytes of the naive larvae and, after a bacterial challenge, the mRNA levels increased considerably (data not shown). This result, consistent with our previous RT-PCR data (Jiang *et al*., 2003b) and the existence of immune-responsive elements in the PAP-3 gene, validated our methods for estimating the mRNA levels. It also agreed well with the data from immunoblot analysis of the hemolymph samples (Jiang *et al*., 2003b) and immunohistochemical analysis of hemocytes (Fig. 7).

Considering its roles as a center for intermediary metabolism and hemolymph

protein synthesis in insects, we closely monitored the PAP-3 mRNA level in fat body at different developmental stages. The PAP-3 gene was transcribed at a low level in this tissue throughout the last two larval feeding stages. There was a small, daily increase in the first five days in the 4th instars and a decrease thereafter (Fig. 6B). The mRNA was almost undetectable after day 3 of the 5th instar. However, after the insect entered the wandering stage, the PAP-3 mRNA level largely increased and then gradually decreased. The transcripts became abundant again in the early pupal stages and was reduced subsequently. The mRNA was still detectable in the fat body from late wandering larvae, late pupae, and adults.

Discussion

Although it is not well studied, transcriptional regulation of immune genes in lepidopteran insects is generally considered to be similar to that in *Drosophila*. In *Drosophila*, transcription factors of the Rel family (*i.e.* Dorsal, Dif, and Relish) control the expression of many immune genes (Engstrom, 1998; Harshman & James, 1998). They form homodimers or heterodimers, bind to their recognition sequences (NF-κB motifs), and induce the synthesis of antimicrobial peptides (*e.g.* diptericin, cecropin, and drosomycin) and other acute-phase proteins (Hoffmann, 2003). The nuclear translocation of these transcription factors are regulated by the Toll and Imd pathways. Pathogen recognition directly or indirectly stimulates these intracellular signaling pathways. For example, *Drosophila* PGRP-SA, Persephone (a clip-domain SP), and cleaved spätzle are components of an extracellular signaling pathway that activates the Toll receptor (Michel *et al*., 2001; Ramet *et al*., 2002; Ligoxygakis *et al*., 2002; Choe *et al*., 2002; Takehana *et*

al., 2002). NF-κB motifs and their binding proteins exist in other insects. *Cecropia* immune-responsive factor was identified in nuclear extracts from the induced *Hyalophora cecropia* pupae (Sun & Fye, 1992). Responsive elements similar to *Drosophila* NF-κB motif (GGGRAYYYYY) are present in the 5' flanking region of *M. sexta* hemolin and lysozyme genes (Wang *et al*., 1995; Mulnix & Dunn, 1994). In this work, we elucidated the structure of *M. sexta* PAP-3, an immune-responsive SP involved in melanization and other defense mechanisms. In particular, nine NF-κB motifs were identified, three of which closely resemble the *Drosophila* consensus (Table 2). Like other immune responsive elements (*e.g.* ISREs, GATA boxes), the role of these putative recognition sequences has to be further investigated.

Many arthropod SPs and SPHs involved in immunity and development contain clip domains (Jiang & Kanost, 2000). One of five *Drosophila* SPs/SPHs has such a regulatory unit (Ross *et al*., 2003). So far, the structure and function of clip domains are poorly understood and their evolutionary history has not been closely examined. We found that the *Drosophila* SPs/SPHs with multiple clip domains are different in domain subgroups and linker sizes from *B. mori* PPAE, *M. sexta* PAP-2 and PAP-3 (Satoh *et al*., 1999; Jiang *et al*., 2003a and 2003b). Perhaps, PPAE-like SPs evolved after Lepidoptera and Diptera diverged from their common ancestor ~330 million years ago (Gaunt & Miles, 2002), and there was a subsequent expansion of dual clip-domain SPs in lepidopteran insects. In this study, we discovered that each clip domain in PAP-3 did not evolve as complete exon unit(s). While such an exon-domain relationship impedes domain shuffling/duplication, mutations at the splicing junctions may have contributed to the sequence diversity within the two domains. Exon duplication is responsible for the

domain organization of *Drosophila* masquerade, whose clip domains 1, 2, 3-4, and 5 are separately encoded by four exons (http://www.flybase.org/).

The PAP-3 mRNA levels varied in different tissues of *M. sexta*. In the 5th instar larvae, PAP-3 transcripts were more abundant in hemocytes and trachea than in fat body. After the larvae entered the wandering stage, however, fat body became a major source of PAP-3 expression. Additionally, the size of the PCR product was 75 bp smaller in fat body than that in some other tissues (Fig. 6A, Jiang *et al*., 2003b). At this moment, we do not understand the significance or mechanism of such tissue-specific regulation of PAP-3 transcription/splicing. In fact, the 75 bp sequence was not identified in the sequenced regions of λ7 and λ3.

The transcription profiles of PAP-3 and PAP-1 (Zou, unpublished studies) are somewhat similar in the fat body at different developmental stages. Their up-regulation during the wandering stage (days 1-4) and down-regulation in the middle-to-late pupae stage suggest that hormonal signals might play a role in the transcriptional control of PAP-1 and PAP-3 expression. Indeed, upon treatment with 20-hydroxyecdysone, PAP-1 mRNA level in the cultured larval fat body was reduced in a concentration-dependent manner (Zou, unpublished data). We identified three ecdysone responsive elements in both genes. Nonetheless, we did not detect a significant change in PAP-3 mRNA level after the cultured fat body was incubated with the molting hormone (Zou and Jiang, data not shown). The hormonal regulation of immune proteins is worth exploring in the future.

Table 1 *M. sexta* PAP -3 intron sequences at the 5'and 3' splicing junctions

intron	5' end	3' end
1	GTGAGT	TTTCCAG
2	GTAAGT	TTTCCAG
3	GTGAGG	TTTTCAG
4	GTGAGT	TTTTCAG
5	GTGAGT	GTTGCAG
6	GTATAG	CTTGCAG
7	GTGATT	TTCGCAG
consensus*	GTRWDK	BTYBCAG
predominant**	GTGAGT	TTT-CAG

*: consensus sequences were determined at each position when a particular type of nucleotide
was present in all six positions. *R*: A,G; *Y*: C,T; *W*: A,T; *B*: C,G,T; *D*: A,G,T.

**: a nucleotide present in 5-7 of the sequences at the same position.

Table 2 Sequence analysis of the 5' flanking region of *M. sexta* PAP-3 gene

motif name and consensus*	Sequence found**	Location	Matched
	GGGGACTTCT -	-1765	10/10
	GGcAATTCCC +	-1018	9/10
	GcGGAgTCTC +	-922	8/10
	aGGAATTCaT -	-760	8/10
NF-κB: GGGRAYYYYY	GGtAACTTgT -	-694	8/10
	GaGGAaTCTT +	-683	8/10
	GGaGATTCTC +	-427	9/10
	tGaAACTTTT -	-196	8/10
	GGtcACTCTT -	-35	8/10
	TGATAA +	-1713	6/6
GATA: WGATAA	AGATAA +	-1137	6/6
	TGATAA +	-536	6/6
	AGATAA +	-145	6/6
	GGtAAAAaAAATG +	-1724	11/13
	GtAAACTGAtACA -	-944	11/13
ISRE: GGAAANNGAAANN	tGAAtTTGAAATT -	-292	11/13
	GGgcACGGAAAAG +	-213	11/13
	aGAAgGAGAAACT +	-193	11/13
	AGtTTgAATGcACTT +	-804	12/15
EcRE: RRGKTCANTGAMCYY	AAGaTtATTGcACTC +	-291	12/15
	AGGTTCAGTGcgCgT +	-6	12/15

*: *R*: A,G; *Y*: C,T; *W*: A,T; *K*: G,T; *M*: A,C. *W*: A,T; *N*: A,C,G,T.

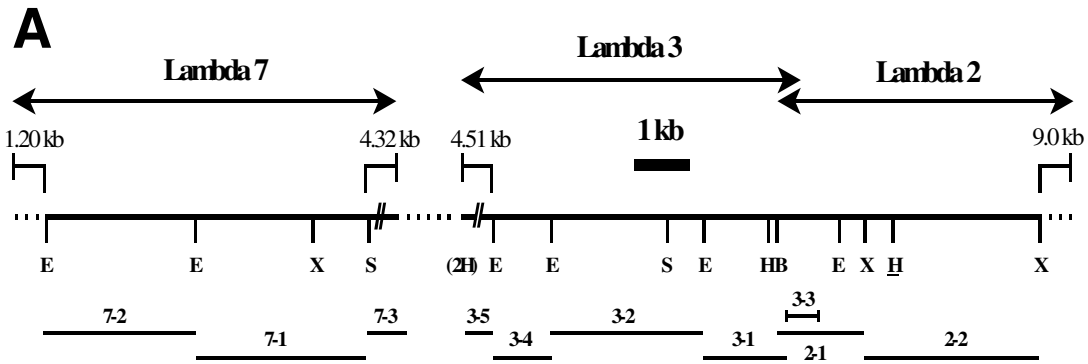**: nucleotides in the lower case do not match with the consensus.

References

Ashida, M. and Brey, P.T. (1998) Recent advances on the research of the insect prophenoloxidase cascade. In: Brey, P.T. and Hultmark, D. (Eds.), Molecular Mechanisms of Immune Responses in Insects. *Chapman and Hall*, London, pp. 135-172.

Ahmed, A., Martin, D., Manetti, A.G., Han, S.J., Lee, W.J., Mathiopoulos, K.D., Muller, H.M., Kafatos, F.C., Raikhel, A., and Brey, P.T. (1999) Genomic structure and ecdysone regulation of the prophenoloxidase 1 gene in the malaria vector *Anopheles gambiae*. *Proc Natl Acad Sci USA* 96, 14795-14800.

Cherbas, L. and Cherbas, P. (1993) The arthropod initiator: the capsite consensus plays an important role in transcription. *Insect Biochem Mol Biol* 23, 81-90.

Cherbas, L. and Cherbas, P. (1996) Molecular aspects of ecdysteroid hormone action, In: Gillbert, L.I., Tata, J.R. and Atkinson, B.G. (Eds.). Metamorphosis: postembryonic reprogramming of gene expression in amphibian and insect cells. *Academic Press*, San Diego, California. pp. 175-221.

Choe, K.M., Werner, T., Stoven, S. Hultmark, D., and Anderson, K.V. (2002) Requirement for a peptidoglycan recognition protein (PGRP) in Relish activation and antibacterial immune responses in *Drosophila*. *Science* 296, 359-362.

Dunn, P. and Drake, D. (1983) Fate of bacteria injected into naïve and immunized larvae of the tobacco hornworm, *Manduca sexta*. *J Invert Pathol* 41, 77-85.

Engstrom, Y (1998) Insect immune gene regulation. In: Brey, P.T. and Hultmark, D. (Eds.), Molecular Mechanisms of Immune Responses in Insects. *Chapman and Hall*, London, pp. 211-244.

Gaunt, M. W. and Miles, M. A. (2002) An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks. *Mol Biol Evol* 19, 748–761.

Gillespie, J.P., Kanost, M.R., and Trenczek, T. (1997) Biological mediators of insect immunity. *Ann Rev Entomol* 42, 611-643.

Harshman, L.G. and James, A.A. (1998) Differential gene expression in insects: transcriptional control. *Ann Rev Entomol* 43, 671-700.

Hoffmann, J.A. (2003) The immune response of *Drosophila*. *Nature* 426, 33-38.

Hultmark, D. (1993) Immune reactions in *Drosophila* and other insects: a model for innate immunity. *Trends Genet* 9, 178-183.

Jiang, H., Wang, Y., and Kanost, M.R. (1998) Pro-phenoloxidase activating proteinase form an insect, *Manduca sexta*: a bacteria-inducible protein similar to *Drosophila* easter. *Proc Natl Acad Sci USA* 95, 12220-12225.

Jiang, H. and Kanost, M.R. (2000) The clip-domain family of serine proteinases in arthropods. *Insect Biochem Mol Biol* 30, 95-105.

Jiang, H., Wang, Y., Yu, X.Q., and Kanost, M.R. (2003a) Prophenoloxidase-activating proteinase-2 from hemolymph of *Manduca sexta*: a bacteria-inducible serine proteinase containing two clip domains. *J Biol Chem* 278, 3552-3561.

Jiang, H., Wang, Y., Yu, X.Q., Zhu, Y., and Kanost, M. (2003b) 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.

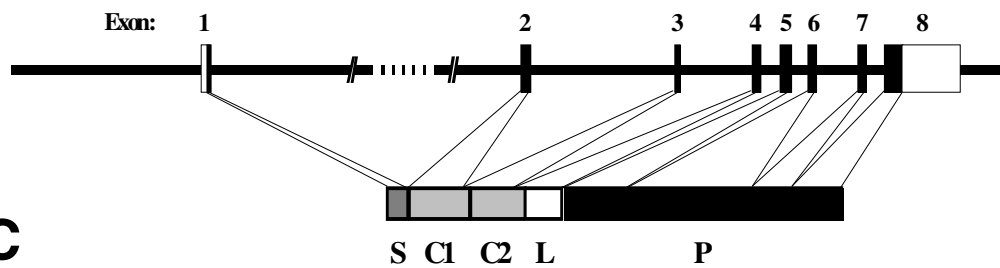Ligoxygakis, P., Pelte, N., Hoffmann, J.A., and Reichhart, J.M. (2002) Activation of

Drosophila Toll during fungal infection by a blood serine protease. *Science* 297, 114-116.

Michel, T., Reichhart, J.M., Hoffmann, J.A., and Royet, J. (2001) *Drosophila* Toll is activated by Gram-positive bacteria through a circulating peptidoglycan recognition protein. *Nature* 414, 756-759.

Mulnix, A.B. and Dunn, P.E. (1994) Structure and induction of a lysozyme gene from the tobacco hornworm, *Manduca sexta*. *Insect Biochem Mol Biol* 24, 271-281.

Nappi, A.J. and Vass, E. (2001) Cytotoxic reactions associated with insect immunity. *Adv Exp Med Biol* 484, 329-348.

Ramet, M., Manfruelli, P., Pearson, A., Mathey-Prevot, B., and Ezekowitz, R.A. (2002) Functional genomic analysis of phagocytosis and identification of a *Drosophila* receptor for *E. coli. Nature* 416, 644-648.

Ross, J., Jiang, H., Kanost, M.R., and Wang, Y. (2003) Serine proteases and their homologs in the *Drosophila melanogaster* genome: an initial analysis of sequence conservation and phylogenetic relationship. *Gene* 304, 117-131.

Sambrook, J. and Russell, D.W. (2001) Molecular cloning, a laboratory manual, 3rd edn. *CSHL Press*, New York.

Satoh, D., Horii, A., Ochiai, M., and Ashida, M. (1999) Prophenoloxidase-activating enzyme of the silkworm, *Bombyx mori*. Purification, characterization, and cDNA cloning. *J Biol Chem* 274, 7441-7453.

Sun, S.C. and Faye, I. (1992) *Cecropia* immunoresponsive factor, an insect immunoresponsive factor with DNA-binding properties similar to nuclear-factor kappa B. *Eur J Biochem* 204, 885-892.

Takehana, A., Katsuyama, T., Yano, T., Oshima, Y., Takada, H., Aigaki, T., and Kurata, S. (2002) Overexpression of a pattern-recognition receptor, peptidoglycan-recognition protein-LE, activates imd/relish-mediated antibacterial defense and the prophenoloxidase cascade in *Drosophila* larvae. *Proc Natl Acad Sci USA* 99, 13705-13710.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22, 4673-4680.

Wang, R., Lee, S.Y., Cerenius, L., and Söderhäll, K. (2001) Properties of the prophenoloxidase activating enzyme of the freshwater crayfish, *Pacifastacus leniusculus*. *Eur J Biochem* 268, 895-902.

Wang, Y., Willott, E., and Kanost, M.R. (1995) Organization and expression of the hemolin gene, a member of the immunoglobulin superfamily in an insect, *Manduca sexta*. *Insect Mol Biol* 4, 113-123.

Willott, E., Lowenberger, C., Christensen, B.M., and Kanost, M.R. (1995) Monoclonal antibodies against *Manduca sexta* hemocytes bind *Aedes aegypti* hemocytes: characterization of six monoclonal antibodies that bind hemocytes from both species. *Dev Comp Immunol* 19, 451-461.

Fig. 1 Structure of *M. sexta* PAP-3 gene. A) Restriction map of PAP-3 genomic inserts in the positive clones, λ2, λ3, λ7. Arrow heads and vertical bars indicate the start and end points of the clones. B, *BamH*I, E, *EcoR*I; H, *Hind*III; S, *Sac*I; X, *Xba*I. The underlined *Hind*III site denotes its presence in the cDNA. The 4.51 kb fragment at the 5' end of λ3 contains two *Hind*III sites (in parenthesis) at positions 1110 and 1500. Horizontal bars mark the PAP-3 genomic fragments for subcloning. The segment "3-3" represents the 3' end of λ3 whose sequence does not match λ2. Dotted lines, regions not sequenced; broken lines, 3' end of λ7 and 5' end of λ3, sequenced. **B)** Exon-intron organization of the PAP-3 gene. Numbered vertical bars denote the exons, with the noncoding regions shown as open boxes. S, signal peptide; C1 and C2, clip domains 1 and 2; L, linker region; P, proteinase domain. The same scale and starting site are used for the restriction map and exon-intron structure. **C)** Sequence of the *BamH*I-*Bgl*II fragment at the 3' end of λ3. The mismatching region, starting 2 nucleotides before the *EcoR*I site, is shown in italics. The 5 nucleotides AGATC (underlined) resemble the recognition sequence (AGATCT) of *Bgl*II, a restriction enzyme used for the library construction.

A

Lambda 7   Lambda 3   Lambda 2

1 kb

1.20 kb   4.32 kb   4.51 kb   9.0 kb

E   E   X   S   (2H) E   E   S   E   HB   E X   H   X

7-2   7-3   3-5   3-2   3-3
7-1   3-4   3-1   2-2
2-1

B

Exon: 1   2   3   4 5 6   7   8

S C1 C2 L   P

C

*Bam*HI
GGATCCAATAAAATATCGAATGAACGAGGACATGATTACCCCTCGGCAGTCGACACATTTATGCCGGCTTATTGAAACCG
GATGTTCTTTTTTGTCCCATTCCCCTACTAGAATTTAAACCATCGACCTAATCATTGGACCACACATTTTCAGATGCCAA

*Bgl*II   ‖ *Eco*RI
GGGGCAGATCAGTGAGAATTCTCAGTCTTTAATCAATGTTCGAGACGGAATTGCAGTACAGAGAGTTAATGTAACAGGTA
ATTTAGTAGAATAATTCTGTAGATGTTCTCGTCCACGCTAAGAATGAGTTGGCAAAGGACACTGCTTTACAGAAATCATG
TACGTACTGCTTTGCTTGTATTTCGTACATACGTATAGTAAAATTGCTGATGCTTTCAAAATCTCTTTTTTCTGCATAGAA
CATATTTACTCCTTCAACATCAGCTAACAAGAACAGTTTATCTAACAAGAATTGTTTATTTGCGGTGACTTGTCCGCCTT
CTTTTAATCTAAATTCCTTTTTAAAAGTTCTCAATTGATAATAATATGAGAATGAGTGGAGATCTTGAAACCATGCAAGG
TGGACGATTTAGTTTATTGGTCTCTGAACAGAACGATAACTGAAGCCAGAGACGATCGATAAGTCTCAATCTGTGTTCCA
TTTTTTTTCTCTAACATCCCCAAACTATCCATTTATTATTGTGGGTGTTTATGGGCTGCGATATTTAGTTTAAAACCTGG
*Bgl*II
ATACTTACTGCTTTTACAAAAAAATTAATGCTTTTTTTTTTAATAGTCACGGGACTCATAACTCGCAAGATACCTTAGATC

90

Fig. 2 Nucleotide sequence and structural features of *M. sexta* PAP-3 gene. Nucleotides in the 5'-flanking region are assigned negative numbers. Nucleotide 1 is assigned based on the primer extension results (Fig. 3). Exon sequences are underlined with the encoded amino acid sequences listed below translated exons, using the one-letter code under the 2nd nucleotide of each codon. While some regions of the intron sequences (marked "---") are not shown, their sizes and positions are indicated. GATA boxes (6-nucleotide) and ISRE sites (13-nucleotide), bold and double underlined; NF-κB motifs (10-nucleotide) and EcRE (15-nucleotide), bold and single underlined; TATA boxes (6-nucleotide), bold italic and double underlined. The mismatches in these motifs are in the lower case. Sequences of 3' end of λ7 (4.32 kb), gap, and 5' end of λ3 (4.51 kb) are not shown. The *Sac*I and *EcoR*I recognition sites (in bold) are numbered consecutively as 2336-2341 and 2342-2347, respectively. Single nucleotide polymorphic sites are in bold italic on the DNA sequence. Among them, nonsynonymous substitutions are further indicated on the affected amino acid residues (bold and underlined). The Cys residues in the clip domain and the catalytic residues in the SP domain are indicated with "+" and "#", respectively.

```
-1804  GGTTGAAATTCCTTCTATTCTCTCGACCTAAGAAGTCCCCAAGCAAAGTGTCATACATCCGCAGTGATTTCATACAGTATGGtAAAAaAAATGATAATCA
----                                           0.5 kb (-1205 to -1704)
-1204  TCACGAAACAATAATTGTTAAAAATCTATTTGGTCAAACATATCCAATATCAGCTTTATTACTACGTAGATAAGACTAACAGCAAACATTTGATCCAACT
-1104  ATACAAGCTATAGTATCGCCACTCGCCAAATATCTAAATGAATGACGAACGCATTGTAAACAAAACTAAGCGACTATCAACGTATGGGcAATTCCCAGAA
-1004  GATATTGAAGTCACACCGGTTTGATACTTGCGAATTAAAATTATGGACTGTaTCAGTTTaCATTAAAGTTGTAAATCACGTGGCGGGAgTCTCACTGTATC
-904   GTGTCTACAAAGATTCTGGTGTTTGTATAGAATAATAAATAGACGCGGCAATTAAATGTATTAAATAATTAATATTGTCGATGTAGCTGTGAAGTCCATA
-804   AGtTTgAATGcACTTTGTTCCAGGATTTCTGTTGTAtGAATTCCtGATACAGCTCGAGGGCTTTGCGATGTACCGAGATGGTGTGACTAGAAAAAACCAA
-704   AAcAAGTTaCCAGATTTCATGGaGGAaTCTTGAACTTTGGACAGCATTTACTTAATCTGCACTAACAGAACACATCTAACCTAACATCCTAACATCTGCA
-604   TATAGTGCTTATAGTCGAATCATCTTATAACTTTATACGAATTTTGCTATTGATTGCCCACTTGTCTGTGATAAAAATAATTATTTATTTTAATTCAGAA
-504   TTTTGTATATTTGATCATTTCTAACAAACATCAAAACACGAACTTTCGCATTTATAAATTTTAATTAAAGCAACACCGGaGATTCTCGCCCGTTCTTTTC
-404   TTATGAGAACTGGTTTCTGAACTGGCGGTAGAGTTAATATTGTCGGAATCTAAATAATGTATAACATTTTACAGGTTGAAATATATTCATTCATAAATTT
-304   AATTTCAAaTTCaAAGaTtATTGcACTCCACAATCATTTTATACATAAGATGTTACATTTTAATACGCGTGACATAATTGTGGACCGTGTCGGgcACGGA
-204   AAAGTTtCaAGaGAAgGAGAAACTATACAAAGTACCTTATAATAGGTATATTTAATTTCAGATAAAAACTCTGAGACGATACGTCATCTTAAATGCCGTA
-104   AACACGTAGGTATAAATTTGCATCGACACGTCCTATCTGCAAACTACGAACCTTGTCCATAAGAGTgaCCAGCGAGCCGGTAGACCGTGGCCACTGGCAG
-4     GTTCAGTGcgCgTCGAACCGACGCTGAGGTAACACGTATTGTGATTAAATGTAAATGTGAATACGAAAAAAAAATGGCAAATTTAGTTTACGTGTTGTTA
-1                                                               Exon 1:     M  A  N  L  V  Y  V  L  L
97     TTAGCTTCGTATTTCTGCTTTGTTAGTGGACGTGAGTATTTCATCGTTTCCTCTAAACTATTTTAAGTTTTTCGCTAATTATCATTGTGTTTCATACTTA
        L  A  S  Y  F  C  F  V  S  G   (signal peptide)
----                                           2.1 kb (+ 197 to +2296)
2297   ACTATGCTTCCCAGAATATAAGACATAAGACGAGTCTGCGAGCTC 2341 (4.32 kb) - - - - - - Unknown gap - - - - - -
                                              SacI EcoRI
----    - - - Unknown gap - - - - - - (4.51 kb) 2342 GAATTCCATGAAAACGACCTGTTTTTTTTTTTCATATGGCTCAAAAGTAAAAAAAA
----                                           1.0 kb (+2397 to +3396)
3397   TTACACCCATGAATTCTTGACTGATGAATGCTTATTAAATTATATGCTCATGATATTTATTGGAACTTTCCAGAATCATGCACGACGCCCAACAACGGTA
                                                               Exon 2:     Q  S  C+ T  T  P  N  N  G
3497   AAGGCACGTGTAAGTCGATATACGAGTGCGAAGAGCTTCTCAAACTGGTGTACAAGAAGGACAGGACGCAGCAGGACACGGATTATTTGAAGAAGTCCCA
        K  G  T  C+ K  S  I  Y  E  C+ E  E  L  L  K  L  V  Y  K  K  D  R  T  Q  Q  D  T  D  Y  L  K  K  S  Q
3597   GTGCGGCTTCATGGGCAACACACCAACGGTAAGTACTAAACACATACCCCCCTTATTCATAGACTTTTTTACCTAAGGACAGAGCAAAGCTGTGATAACAA
        C+ G  F  M  G  N  T  P  T
----                                           2.5 kb (+3697 to +6196)
6197   TTTCCAGCTAAACATTCAACATGCATTATCAATTTCCAGGTTTGCTGCCCGAACCCTTGCATCACGCCACAGGGAGAGCCAGGGCAATGCGTCAGCATTT
                    Exon 3:     V  C+ C+ P  N  P  C+ I  T  P  Q  G  E  P  G  Q  C+ V  S  I
        - - - - - - clip domain 1 - - - - - -||  ||- - - - - - clip domain 2 - - - - - -
6297   ACGAATGCACTAATCTCGCCAATTTGCTGAAGCCACCCATTACCGCCGATACGTATAACTATGTGCAGAAGTCTAGGTGAGGTTTAGTTTTTATACGGAC
        Y  E  C+ T  N  L  A  N  L  L  K  P  P  I  T  A  D  T  Y  N  Y  V  Q  K  S  R
----                                           1.2 kb (+6397 to +7596)
7597   CTAATCATTGGACCACACATTTTCAGATGCCAAGGGGCAGATCAGTACAGTGTCTGCTGCGGATCAGCGCCAAACTTCCCTTCCAGTGGCGACTGCAAGG
                    Exon 4:     C+ Q  G  A  D  Q  Y  S  V  C+ C+ G  S  A  P  N  F  P  S  S  G  D  C  K
                 - - - - - - clip domain 2 - - - - - -||  - - - - - - linker region - - - -
7697   CTTCCGTTTCCGCCTTCCCGCCAGATCCCAAGTCCAAGTGCTGCGGTGTAGACAGCAAGATCGGGAACAAAATCATTGGTGAGTGATTGTAAACTTTCTA
        A  S  V  S  A  F  P  P  D  P  K  S  K  C  C  G  V  D  S  R  V  G  N  K || I  I
                             - - - - - - linker region - - - - - - ||  - - - catalytic domain - -
----                                           0.3 kb (+7797 to +8096)
8097   TTCTAAGCGATTGACGCCCAAATCTTTTCAGGTGGAAATGCGACGACGTAGACCAATACCCATGGTTAACCATCATCGAGTACGTGAAAACCGGCCCCAT
                    Exon 5:     G  G  N  A  T  D  V  D  Q  Y  P  W  L  T  I  I  E  Y  V  K  T  G  P  I
8197   TAAGCTGTTGTGTGGTGGCGTCCTCATCAGTGGAAAATACGTGCTGACAGCAGGACACCTGTTCACGGGACCCGTTTTGCAAATTGGCACACCGTGAGTG
        K  L  L  C  G  G  V  L  I  S  G  K  Y  V  L  T  A  G  H# C  L  T  G  P  V  L  Q  I  G  T  P
----                                           0.3 kb (+8297 to +8596)
8597   TCCATATACACCGAACACGAGCAGTATGCTAAATCTTTCTTTATACCTAAATTGTTGCAGAACAAACGTCCGCCTGGGAGAATACAATACGAAAAACGAC
                                                         Exon 6:     T  N  V  R  L  G  E  Y  N  T  K  N  D
8697   GGCGCGGACTGCGTGACGGTAGAGGCTGGGGGGATGGACTGCACAGAGGGAGCGGTCATCGTGCCCATAGAGAAGACAATACCGCACCCGGAGTACAACC
        G  A  D  C  V  T  V  E  A  G  G  M  D  C  T  E  G  A  V  I  V  P  I  E  K  T  I  P  H  P  E  Y  N
8797   CCATCAGCAGCGAGGAGGAACGACATAGGACTCATCAGGCTGAAGGAGATGGCGCCGTTTACTGGTATAGTTTTGTTTATTAACACTTTAATTAGATT
        P  I  S  R  T  R  R  N  D# I  G  L  I  R  L  K  E  M  A  P  F  T
----                                           0.6 kb (+8897 to +9496)
9497   TTACCGTGATTATTTATTCATTAATAAATTGCCACTTGCAGATTTCATCCGTCCCATCTGCCTGCCAAGCTTGGACCTGACGCAGGCTCCGCCGGTGAAC
                                                Exon 7:     D  F  I  R  P  I  C  L  S  L  D  L  T  Q  A  P  V  N
9597   TTCACGCTGTACGCCGCCGGTTGGGGAGCTGTCAGCACCTCACAGCCCAGCAGCAACGTCAAACTGCACGTCCAGCTGCCGTTCATCAGTTACGAGGTGA
        F  T  L  Y  A  A  G  W  G  A  V  S  T  S  Q  P  S  S  N  V  K  L  H  V  Q  L  P  F  I  S  Y  E
----                                           0.3 kb (+9697 to +9996)
9997   CAATGAAATAATCCATCCATTCATTCGCAGCGATGTCAGCCGTCGTACGCCGTACAGAACAGACAGATCGAGTTGTGGGAGAAGCAAGTGTGCGCGGGCG
                         Exon 8:     R  C  Q  P  S  Y  A  V  Q  N  R  Q  I  E  L  W  E  K  Q  V  C  A  G
10097  GCGAGGCTGGCAAAGACTCCTGCAAGGGTGACTCCGGAGGTCCCCTCATGTTCGAAAACGGGCAGACTTACGAAGTCATCGGCATAGTCTCTTTCGGACC
        G  E  A  G  K  D  S  C  K  G  D  S# G  G  P  L  M  F  E  N  G  Q  T  Y  E  V  I  G  I  V  S  F  G  P
10197  GACGCCGTGTGGAATGCAAGATATACCAGCCGTGTACACTAAAGTACACTCTTACAAAGATTGGATAATAAGTAACATAAAGCCGTAGGTTTTGATAATG
        T  P  C  G  M  Q  D  I  P  G  V  Y  T  K  V  H  S  Y  K  D  W  I  I  S  N  I  K  P  *
10297  TGGTCATTGTAAATGAGATTACGTAAAACTATTCAGTTCGAAGTGTTCTAATTATAATCCATATTTACGTCGCTATAAGGTAAAAATCTTACGAGTCAAA
10397  CTTATTTATGAGCAATAAATATTAATGAAGATAATATTAATGATCAAGAAATAATAAAGTAAGTTATCGTTATGATTTCTTATGTTTACGCGAATGTTAG
10497  ACATTGAAATAAAACTATTTCGCATTCAACAAAGTCTTCGAAAACGTCGGGAGAAAATGTAATACAAAAACGCGATTAATTCCGAAAAAATAGTTTTAT
10597  TTCAATATAGGTAATATACGACATAAAGGATGAACTCTCTTAAATATAAATAGTGTAGTGTTTTTATATTTTTTACAATTTACATAAGTTTATTCAAAG
10697  TAAAACAAAACATTCTTCACTATAGATAGCATTATAACTCAATATACTCATTATTTGTTTAAAATGCTTACTAAAGTCAATTCGAATGTGTTTTATATTT
10797  TAACGCTATTATAGTCGAATCATCTTAATAATCACACGTTTAAGTTATCTTCAAAGAAAACACTTATTTAATCTAATCATAGCTTAAAATGGTCGGTTTGCT
10897  GATGTTAAAACCCGCCATAGTGGCTCACGTTAGTGTGTCGCGTTCCAGGATCAGCCTGTGTATAATCCGGTTCCAACTGGCCGGGATAATTGTATCGACT
10997  GTCGAAGGGCAATCATCTCTCGTCTCTAAGTTTTTAATATAAATTTCAACTTTGCTAGCTCCACGCGTTCTTTAAAAAAATGGCTCTTGACAGACAGAG
11097  AGACTGACAAAGGTGATCCTGTAAGGGTTCCTTTTGTAGTACGGAATAATATTTTTGAAACTTATTAGCAATGATTAAAAGTTAATTAAATTTATCAATA
11197  AAGTTACACTATTACGATATTTTAAAATAATGTTGGGATGGGATGATCTCAATAGCCCTCCTTGATAATACTATAATTTTTTATTAAGATTCTAAATAAG
11297  ATTATGTGGTAATAAACTATTTTTATAAAGTGTTCTCTTTTTATTGATTGACTCAACTGGCCCGCTAGATGTCGTCCTTTCCTAACTATGAATATTC
```
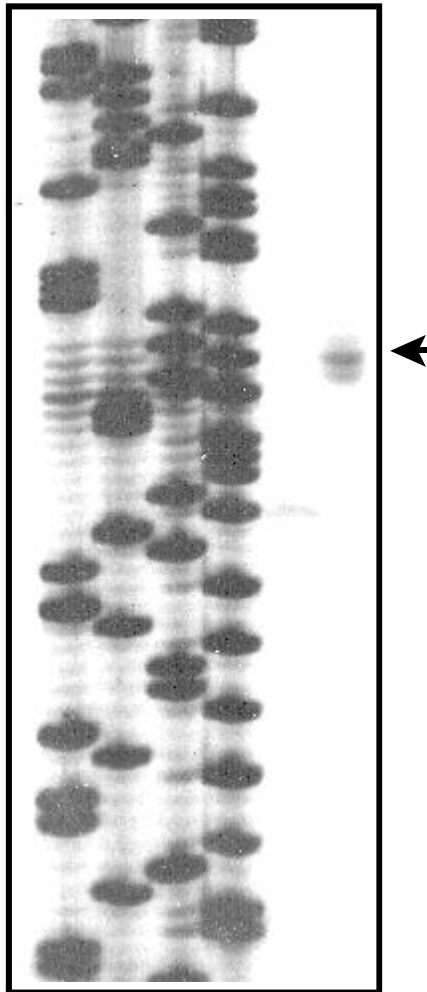
Fig. 3 Determination of the transcription initiation site in *M. sexta* PAP-3 gene. A primer, complementary to nucleotides 26-55 of the PAP-3 coding region near the 3' end of exon 1, was terminally labeled with γ-^{32}P-dATP and annealed to total RNA from fat body from bacteria-injected larvae (15 μg). After annealing to RNA, the primer was extended with MMLV reverse transcriptase. The set of sequencing reactions (ACGT) on the left of the primer extension lane for use as a sizing ladder was from dideoxynucleotide sequencing of single-stranded M13 mp18 DNA using -40 primer. The arrow indicates the 125 bp extension product from fat body RNA isolated from the induced larvae.
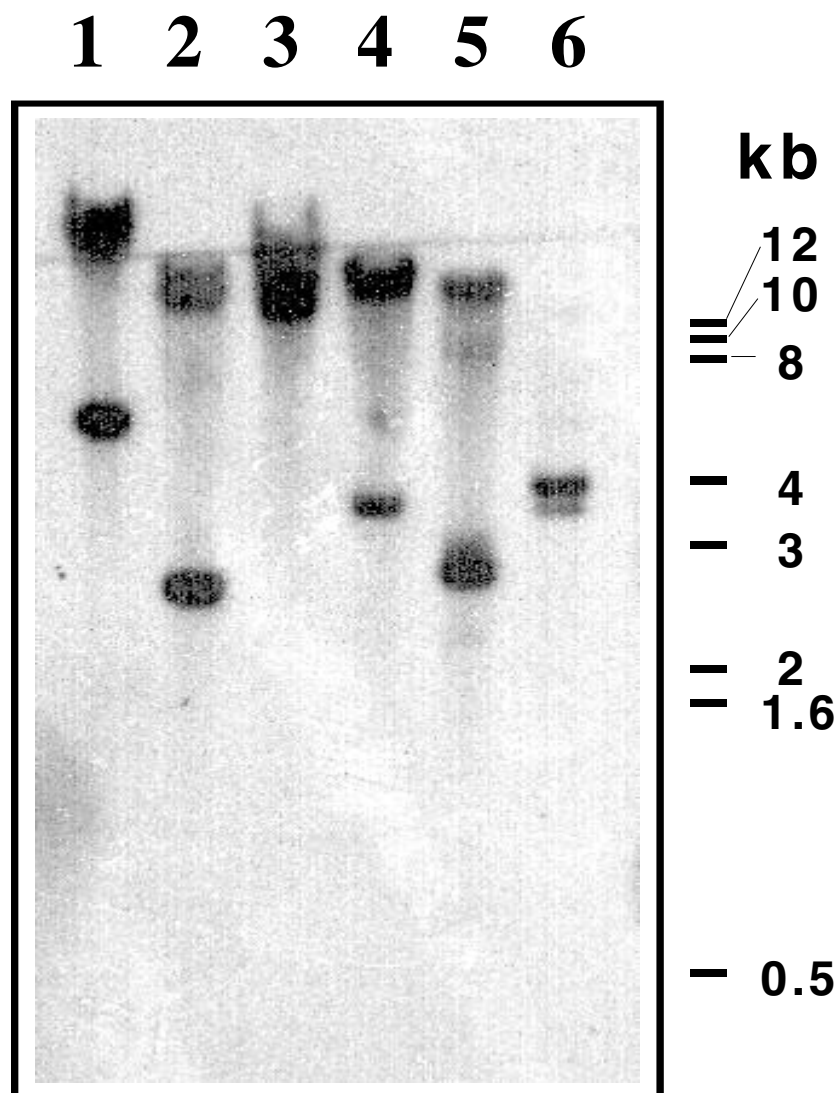
Fig. 4 Southern blot analysis of *M. sexta* genomic DNA using ^{32}P-labeled PAP-3 cDNA. Samples of the genomic DNA (15 µg) were digested with *BamH*I (lane 1), *Hind*III (lane 2), *Sac*I (lane 3); *Xba*I (lane 4), *EcoR*I (lane 5), *Sac*I and *Xba*I (lane 6). After separation by agarose gel electrophoresis and transfer to nitrocellulose membrane, the DNA fragments were hybridized with the full-length cDNA probe. The positions and sizes of the molecular markers are indicated on the right.

A

```
PAP-3  -2262 AAATTTTACTTACTATGTTTTGATAAAAAATTTCATAGTGATTTCTTCCTAAAA-CT--TATTTATATTTCTTAAGGGA -2188
ABC-T  10442 AAATATTACTTACTGTGTTTTGATAACAATTTCATAGTGATTTCTTTCTAAAATCTTATATTCATGTTTCTTAAGTGA 10519
             **** ********* *********** ******************* ****** **  **** ** ********* **

PAP-3  -2187 ACAAAGCAATTTTTAAGTGTTAAAACTT-ATATATGGGGTGCATAAAAATTTTAGAAATGTTAAGGTGGGGCATGGC -2112
ABC-T  10520 ACAAATCATTTTTTA--TGTTAAAAATTAATGTATGGGTGGCATCAAAA-TTTAGAAATGTTACAGTGGGGCATGGC 10593
             ***** ** ******  ******** ** ** ******  **** **** ************* ************
```

B

```
JHBP        6150  GACATAATAAGACTTAACATCTCATGACTCAGAATGGCAAGCGCAGTAGAATACCGAACAATACTTT  6084
serpin-1    6117  GACATAATTAGACTTAACATCTCATGTCTCAGGATGGCGAGCGCAGT-GAATACTAAACAATACTTT  6052
PAP-3       7932  GACATAATAAGACTTAATATCTCATGTCTCAGGATGACGAGCGCAGTGGAATACCAAACAATACTTT  7999
arylphorin  2900  GAGATAATACGACTTAACACCTAATGTCTCAGGATGGCGAGCGCAGTGGAATACCAAACAATACTTT  2835
EH          3414  GACGTAATAAGACTTAACATCTCATATCTCGAGATGGCGATTGCAGTGGATTACCAAACAATATTTT  3480
                  **+*+++***++********++*+**++*+*+*++++++*+*++++++ **+++++++++++++**
```

Fig. 5 **Multiple sequence comparison. A)** A 151 bp sequence in the PAP-3 gene aligned with *Bombyx mori* ABC transporter (ABC-T, AF245662) gene segment. Identical positions are indicated with "*". **B)** A highly similar intron sequence in *M. sexta* PAP-3, juvenile hormone-binding protein (JHBP), eclosion hormone (EH), serpin-1, and arylphorin genes. Based on a BLASTN search, similar sequences in the PAP-3 (nucleotides 6150-6084), JHBP (nucleotides 6084-6150, AF527636), EH (nucleotides 3414-3480, M27808), serpin-1 (nucleotides 6052-6117, U58361), and arylphorin (nucleotides 2835-2900, M28394) genes were retrieved from Genbank and aligned using ClustalW program (Thompson, 1994). The JHBP, serpin-1, and arylphorin sequences are shown as reverse complement of the plus strand DNA sequence. *, position identical in all five sequences; +, position identical in four of the five sequences.

Fig. 6 Profiling of PAP-3 gene transcription by RT-PCR. A) Expression of PAP-3 gene in different tissues from day 3 5th instar larvae or day 3 wandering larvae. N, nerve tissue; S, salivary gland; Mt, Malpighian tubule; T, trachea; Mg, midgut; H, hemocytes; F, fat body; I, integument; Mu, muscle; l, 438 bp, **l**ong PCR product with the 75 bp insert; s, 363 bp, **s**hort PCR product; l/s, a mixture of l and s. Two forms of the PAP-3 cDNA clones (Jiang et al., 2003b) were used as templates to amplify l and s. **B)** changes of PAP-3 mRNA level in fat body from *M. sexta* at different developmental stages. *M. sexta* ribosomal protein S3 (rpS3) transcripts were normalized in all these samples.
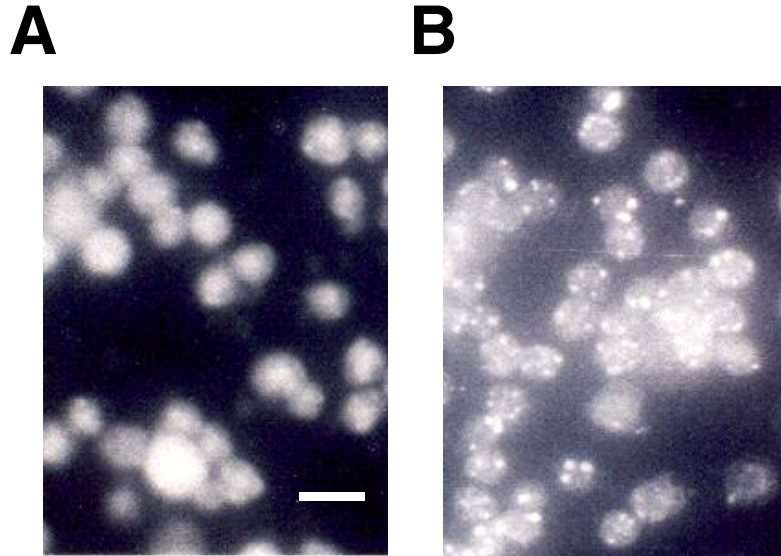
Fig. 7 Immunolocalization of PAP-3 protein in granular cells from the bacteria-injected larvae. The same microscopic field of fixed hemocytes was photographed using fluorescence microscopy. The hemocytes were reacted with a polyclonal antiserum to PAP-3 and FITC-labeled secondary antibodies. **A)** hemocytes from uninjected control larvae. **B)** hemocytes collected at 24 h after injection of *M. luteus.* Scale bar = 50 μm. The bright spots in the granular hemocytes represent PAP-3 synthesized and stored after the immune challenge.

Chapter V. *Manduca sexta* serpin-6 regulates immune serine proteinases PAP-3 and HP8: cDNA cloning, protein expression, inhibition kinetics, and function elucidation

Abstract

Analogous to blood coagulation and complement activation in mammals, some insect defense responses (*e.g.* prophenoloxidase (proPO) activation and Toll pathway initiation) are mediated by serine proteinase cascades and regulated by serpins in hemolymph. We recently isolated *Manduca sexta* serpin-6 from hemolymph of the bacteria-challenged larvae, which selectively inhibited proPO-activating proteinase-3 (PAP-3) (Wang and Jiang, 2004). To further characterize its structure and function, we cloned serpin-6 from an induced fat body cDNA library using a PCR-derived probe. *M. sexta* serpin-6 is 55% similar in amino acid sequence to *D. melanogaster* serpin-5, an immune-responsive protein. We produced serpin-6 in an *E. coli* expression system and purified the soluble protein by nickel affinity and hydrophobic interaction chromatography. The recombinant protein specifically inhibited PAP-3 and blocked proPO activation *in vitro* in a concentration-dependent manner. MALDI-TOF mass spectrometry indicated that the cleavage site of serpin-6 is between R^{373} and S^{374}. Serpin-6 is constitutively present in hemolymph of naïve larvae, and its mRNA and protein levels significantly increase after a bacterial injection. The association rate constant of

serpin-6 and PAP-3 is 2.6 x 10^4 $M^{-1}s^{-1}$, indicating that serpin-6 may contribute to the inhibitory regulation of PAP-3 in the hemolymph. We also identified the covalent complex of serpin-6 and PAP-3 in induced hemolymph by immunoaffinity chromatography and mass spectrometry. Furthermore, immulectin-2, serine proteinase homologs, proPO, PO, attacin-2 and a complex of serpin-6 and hemolymph proteinase-8 were also detected in the proteins eluted from the immunoaffinity column using serpin-6 antibody. These results suggest that serpin-6 plays important roles in the regulation of immune proteinases in the hemolymph.

Introduction

Phenoloxidase (PO) participates in several insect physiological processes, including melanogenesis, cuticle sclerotization, wound healing, and other defense responses (Ashida and Yamazaki, 1990; Ashida and Brey, 1998). It catalyzes the formation of quinones that are precursors of melanin. Melanin and proteins may crosslink to form a capsule around invading parasites. Additionally, quinones might have antimicrobial effects (Nappi and Vass, 2001). Proteolytic activation of prophenoloxidase (proPO) in insects is mediated by a largely unknown serine proteinase pathway triggered by microbial surface molecules, such as lipopolysaccharide, peptidoglycan, and β-1,3-glucan (Ashida and Brey, 1998; Cerenius and Soderhall, 2004). ProPO-activating proteinase (PAP), also known as proPO-activating enzyme (PPAE), cleaves proPO and causes its activation.

So far, we have isolated three PAPs from the tobacco hornworm, *Manduca sexta* (Jiang et al., 1998; Jiang et al., 2003a; Jiang et al., 2003b). They all cleave proPO at Arg[51]

but require an auxiliary factor to generate active PO. We have purified and characterized the "cofactor" as a high M_r complex of serine proteinase homolog-1 and -2 (SPH-1 and -2), but do not yet understand its mechanism (Yu et al., 2003; Wang et al., 2004). A similar phenomenon was reported in the beetle *Holotrichia diomphalia* (Lee et al., 1998). In contrast, *Bombyx mori* PPAE does not appear to need any cofactor for proPO activation (Satoh et al., 1999).

Proteinase inhibitors in human plasma remove excessive proteinases (Bode and Huber, 1992). These include serine proteinase inhibitors of the serpin superfamily, whose typical sizes range from 370 to 450 residues (Gettins, 2002; Silverman et al, 2001). To date, over 500 serpins have been identified in eukaryotes, bacteria, and viruses (Irving et al., 2000). They share a common tertiary structure albeit low sequence similarities of 25-50%. Many serpins have been purified from arthropods, including *B. mori*, *M. sexta*, *A. aegypti*, *M. unipuncta*, and *T. tridentatus* (Sasaki and Kobayashi, 1984; Sasaki, 1991; Kanost et al., 1989; Gan et al., 2001; Cherqui et al., 2001; Stark and James, 1998; Iwanaga et al., 1998). They regulate a number of biological processes including hemolymph coagulation, proPO activation, and induced synthesis of antimicrobial peptides. There are nearly 30 and 15 serpin genes in the *D. melanogaster* and *A. gambiae* genomes (Adams et al., 2002; Christophides et al., 2002), respectively.

In *M. sexta*, the serpin-1 gene encodes 12 variants through alternative exon usage (Jiang et al., 1994; Jiang et al., 1996; Jiang and Kanost, 1997). These serpins differ in their reactive center loops near the carboxyl terminus, which are encoded by variants of exons 9. The serpin-1J variant inhibits all three PAPs by forming SDS-stable complexes (Jiang et al., 2003b). *M. sexta* serpin-3 also blocks proPO activation by inhibiting PAP-1

and PAP-3 in the hemolymph (Zhu et al., 2003). The detection of PAP-serpin-3 complexes in the induced hemolymph provided strong evidence supporting that serpin-3 is a physiological regulator of the two PAPs. *D. melanogaster* serpin-27A, orthologous to *M. sexta* serpin-3, regulates melanization *in vivo* (De Gregorio et al., 2002; Ligoxygakis et al., 2002). *D. melanogaster* Spn43Ac, which controls a putative serine proteinase cascade for the Toll pathway activation, also affects melanin deposition (Levashina et al., 1999). Recently, we partially purified a new serpin (serpin-6) from *M. sexta* hemolymph (Wang and Jiang, 2004). It regulated proPO activation by forming high M_r, SDS-stable complex with PAP-3 *in vitro*. However, *M. sexta* serpin-6 did not strongly inhibit PAP-1 or PAP-2.

Here, we report the cDNA cloning of *M. sexta* serpin-6 and its recombinant expression in *E. coli*. Biochemical analysis demonstrated that recombinant serpin-6 inhibited PAP-3 efficiently. The mRNA and protein levels of serpin-6 were up-regulated in hemocytes and fat body after a bacterial challenge. Moreover, we characterized the hemolymph proteins bound to serpin-6.

Materials and methods

Insects and collection of hemolymph, hemocytes, and fat body

M. sexta eggs were purchased from Carolina Biological Supply, and the larvae were reared as described previously (Dunn and Drake, 1983). Day 2 fifth instar larvae were injected with 1×10^7 *E. coli* cells, or 0.1 mg *Micrococcus luteus* suspended in 50 μl phosphate buffered saline. Hemolymph and fat body samples were collected at 12 and 24 h after the microbial challenge.

Cloning of serpin-6 cDNA fragments by PCR

To isolate *M. sexta* serpin-6 cDNA, degenerate primers (j801-j810) were designed based on its internal peptide sequences (Wang and Jiang, 2004) for PCR amplification of the serpin cDNA fragments. Primers j801 (5'-GGT GTC ACW AAY ATH ATH GC-3') and j802 (5'-GTT CAG GTT IGC DAT DAT RTT-3') encode peptide 1 (EGVTNIIANL NTER); primers j803 (5'-GAC TCC TAY GAR TTY CAN AC-3') and j804 (5'-GTC GGA GTT GTT RTT NGT CAT-3') encode peptide 2 (MTNXNSDSYEFTTANK); primers j805 (5'-GTC ATG GAG GAA WSN TAY ATG-3') and j806 (5'-CAT GTA GGA YTC YTC CAT NAC-3') encode peptide 3 (LAAVMEESYMSR); primers j807 (5'-AAG AAG GAA CCI TTY TTY GT-3') and j808 (5'-ACG AAG AAG GGY TCY TTY TT-3') encode peptide 4 (TKKEPFVSSETR); primers j809 (5'-GAC TGT GTC TCI GAR TTY CAN AA-3') and j810 (5'-GGT GAA TTC GGA NAC RCA RTC-3') encode peptide 5 (LLPADCVSEFT K). The PCR reactions contained 5 μl buffer, 5 μl dNTPs (2.5 mM each), 2.5 μl MgCl$_2$ (25 mM), a pair of degenerate primers (25 pmol each), 10 ng λDNA from the induced *M. sexta* larval fat body cDNA library (Jiang et al., 2003a), Taq DNA polymerase (5 U, Promega), and H$_2$O to bring the total volume up to 50 μl. Thermal cycling conditions were: 94°C, 3 min and 35 cycles of 94°C, 30s, 50°C, 40s, and 72°C, 60s. After electrophoresis, PCR products were excised and recovered from the agarose gel using QIAquick Gel Extraction Kit (Qiagen). The fragments were individually ligated with pGem-T vector and transformed into high efficiency *E. coli* JM109 competent cells (Promega). Plasmid DNAs were isolated from resulting transformants and sequenced using BigDye Terminator Cycle Sequencing Ready

Reaction Kit (PE Applied Biosystems) and vector-specific primers.

Isolation of full-length serpin-6 cDNA by library screening

Once confirmed by BLAST search to be serpin cDNA fragments, the inserts in the PCR-derived clones were labeled with [α-^{32}P]-dCTP (Multiprime DNA Labeling System, Amersham Pharmacia Biotech) and utilized as probes to screen the cDNA library in λZAP2 (Stratagene). Positive plaques were purified to homogeneity and subcloned by *in vivo* excision of pBluescript phagemids. Complete nucleotide sequences were determined as described above. Sequence comparison was performed using ClustalW 1.7 program from MacVector 6.5 software (Genetics Computer Group, 1998).

Expression and purification of soluble serpin-6 from E. coli cells

To amplify a cDNA fragment encoding the mature serpin-6, a PCR reaction was performed using forward primer (j815: 5'-TAA CCA TGG AAT GTT TCT CC-3') and reverse primer (j816: 5'-TAT GCA TGC TTA TTT CTT AGG GT-3'). The PCR product was digested with *Nco*I-*Sph*I and cloned into the same sites in plasmid H6pQE60 (Wang et al., 2001). The resulting recombinant plasmid, serpin-6/H6pQE60, was sequenced to confirm correct insertion and sequence. The expression, extraction, and affinity purification of serpin-6 were performed under nondenaturing conditions as previously described (Jiang and Kanost, 1997; Wang et al., 2001). The recombinant protein eluted from nickel-nitrilotriacetic acid-agarose (Qiagen) was adjusted with ammonium sulfate to a final concentration of 1.0 M. After centrifugation at 15,000×g for 30 min, the cleared protein solution was separated by hydrophobic interaction chromatography on a 5-ml

Phenyl Sepharose column (Amersham Pharmacia Biotech) and eluted at 1.0 ml/min for 40 min with a linear, descending gradient of 1.0 – 0 M ammonium sulfate in 10 mM potassium phosphate, pH 6.4. Concentration and buffer exchange of the serpin-6 fractions were performed in a Centriprep-30 (Millipore). The purified protein was stored at -70ºC in 20 mM Tris-HCl, pH 7.5. A rabbit polyclonal antiserum was prepared against the recombinant serpin (Cocalico Biologicals Inc.).

Inhibition of PAP-3 amidase activity by M. sexta serpin-6

PAP-3 and its cofactor were purified from the prepupal hemolymph (Jiang et al., 2003b; Wang and Jiang, 2004). Purified PAP-3 (100 ng/μl, 1 μl) and serpin-6 (150 ng/μl) were individually mixed in microplate wells. The total reaction volume was adjusted to 12 μl with 20 mM Tris-HCl, pH 8.0. After incubation at room temperature for 10 min, the residual amidase activity was measured using 150 μl, 50 μM acetyl-Ile-Glu Ala-Arg-*p*-nitroanilide (IEAR*p*NA) (Jiang et al., 2003a).

Inhibition of proPO activation by PAP-3 and SPHs

ProPO was purified from *M. sexta* larval hemolymph as described before (Jiang et al., 1997). Purified PAP-3 (100 ng/μl, 1 μl), SPHs (50 ng/μl, 1 μl), serpin-6 (150 ng/μl, 10 μl), and buffer (7.4 μl, 20 mM Tris-HCl, pH 8.0) were preincubated at room temperature for 10 min. In a control, serpin-6 was replaced with the buffer (10 μl). ProPO (10 μl, 10 μg/ml) was separately added to the wells, and the reaction mixtures were then placed on ice for 1 h. A mixture of proPO (10 μl) and buffer (20 μl) was used as another control. PO activity was determined using dopamine as a substrate (Jiang et al., 2003a).

Inhibition of proPO activation by serpin-6 in a fraction of the larval hemolymph

The induced hemolymph was fractionated with 0-50% saturation of ammonium sulfate to obtain the proPO activation system (Jiang and Kanost, 1997). Recombinant serpin-6 (10 μl, at final concentrations of 0-150 μg/ml) was incubated with the plasma fraction (10 μl, 1:5 diluted in H_2O) in the presence of *M. luteus* (1 μl, 1 μg/μl). As controls, 10 μl buffer or the fraction without *M. luteus* were added individually and incubated at room temperature for 10 min prior to the PO activity assay.

Kinetic inhibition assay

A progressive curve method was employed to determine the association rate constant (Askew et al., 2004). Purified PAP-3 (1.8 μM) was mixed with IEAR*p*NA solutions containing serpin-6 at different molar ratios. Absorbance at 405 nm was monitored in the kinetic mode on a tunable VERSAmax microplate reader (Molecular Devices). The progressive curves were first analyzed according to Equation 1: $p = V_z/k(1 - e^{-kt})$, where k is the pseudo-first-order rate constant of inhibition, and V_z is the initial velocity. The experimental values were fit into the equation using Prism 3.0 (GraphPad Software, Inc.). From the plot of k *versus* [I], k_a was derived from a linear regression analysis of the data according to Equation 2: $k = k_a[I]/(1 + [S]/K_m)$. For calculating *in vivo* half-lives of PAP-3, $t_{1/2} = 1/(k_a[I])$, in naïve and induced larval hemolymph, the plasma concentrations of serpin-6 were estimated by SDS-PAGE, immunoblot analysis, and densitometry using the purified recombinant serpin-6 as a standard (Wang et al., 2001).

Detection of serpin-enzyme complexes by immunoblot analysis

PAP-3 (1 μl, 100 μg/ml) and recombinant serpin-6 (2 μl, 100 μg/ml) were incubated at room temperature for 10 min, and then treated with SDS reducing sample buffer at 95°C for 5 min. SDS-PAGE and immunoblot analysis were performed as described previously (Wang and Jiang, 2004).

Determination of cleavage site in serpin-6

PAP-3 (12 μl, 100 μg/ml) was mixed with serpin-6 (12 μl, 280 μg/ml) for 10 min at room temperature. The reaction mixture was subjected to MALDI-TOF mass spectrometry (Jiang et al., 2003b). The molecular mass of a peak that was absent in the control spectra of serpin-6 and PAP-3 was compared with calculated values of carboxyl terminal peptides to deduce the cleavage site in serpin-6.

RNA extraction and reverse transcription (RT)-PCR analysis

Total RNA samples were extracted from fat body and hemocytes of naïve and induced *M. sexta* larvae using Micro-to-Midi Total RNA Purification System (Invitrogen Life Technologies). First-strand cDNA synthesis was performed using 2-4 μg total RNA, 10 pmol oligo(dT)$_{17}$, and 200 U MMLV reverse transcriptase (Invitrogen Life Technologies) at 37°C for 1 h. *M. sexta* ribosomal protein S3 (rpS3) cDNA was used as an internal standard to normalize the templates in a preliminary PCR experiment. After template adjustment, PCRs were performed to detect relative levels of serpin-6 cDNA. The primers used were: j812 (5'-TGA TGA CTG CGT ATA AGG TA-3') and j820 (5'-CTT CTC GAC TCA TGT AGC TCT CT-3') for serpin-6; k504 (5'-CGC GAG TTG

ACT TCG GT-3') and k501 (5' GCC GTT CTT GCC CTG TT-3') for rpS3. The thermal cycling conditions were: 94°C, 30s; 50°C, 30s; 72°C, 60s. PCR cycle numbers were empirically chosen to show comparable band intensity and avoid saturation. After separation by 1.3% agarose gel electrophoresis, intensities of the PCR products were quantified using Kodak Digital Science 1D gel analysis software (Wang et al., 2001).

Immunoaffinity chromatography and MALDI-TOF mass fingerprint analysis

To couple serpin-6 antibodies to protein A-Sepharose, 4.8 ml of the rabbit antiserum, 2.4 ml resin (Sigma) and 19.2 ml phosphate-buffered saline were incubated at room temperature for 1 h with gentle shaking. The resin was washed with 10 volumes of 0.2 M sodium borate, pH 9.0 and reacted with dimethylpielimidate at a final concentration to 20 mM. After 2 h, 0.2 M ethanolamine (pH 8.0) was added to terminate the coupling reaction for another 2 h. For isolating serpin-6-proteinase complexes, 15 ml of cell-free hemolymph was collected from naïve or bacteria-injected larvae (5th instar, day 3). *M. luteus*, diethylthiocarbonate and 1-phenyl-2-thiourea were added to the plasma at final concentrations of 1 μg/μl, 10 mM and 1 mM, respectively, to activate the serine proteinase system yet prevent hemolymph melanization. After incubation at room temperature for 30 min, 1 mM PMSF and 0.5 ml proteinase inhibitor cocktail (Sigma, P-8849) were added to block the proteinases in the activated hemolymph for 10 min. The plasma samples were centrifuged at 5000 x *g* for 15 min at 4ºC to separate debris from the supernatant. After binding with the affinity matrix (1.0 ml) for 8 h at 4ºC with gentle agitation, the unbound proteins were removed by loading the suspension into an empty Poly-Prep column (Bio-Rad). The resin was sequentially washed with 20 ml, 1 M NaCl

and 20 ml, pH 6.8, 10 mM sodium phosphate to reduce nonspecific binding. Bound proteins were eluted with 50 mM glycine-HCl, pH 2.5, and the fractions (0.5 ml) were neutralized with 50 μl, 1.0 M sodium phosphate, pH 8.0. The affinity-purified proteins were subjected to SDS-PAGE, immunoblot, and peptide mass fingerprint analyses (Wang and Jiang, 2004). The mass profiles were analyzed using Mascot (http://matrixscience.com/) to identify significant matches in GenBank. The monoisotopic peptide masses were also compared with the theoretical masses of *M. sexta* hemolymph proteinases (Jiang et al., unpublished data), predicted by MS-digest (http://prospector.ucsf.edu/).

Results

Molecular cloning and structural features of M. sexta serpin-6

We designed degenerate primers encoding five internal peptides of *M. sexta* serpin-6 (Wang and Jiang, 2004). Because the order of these peptides was unknown, we used different combinations of the primers to amplify cDNA fragments by PCR and cloned the products. Sequence analysis confirmed the cDNA inserts encoded different regions of a serpin. Using the longest one (642 bp, amplified by j802 and j803) as a probe, we screened approximately 6×10^5 plaques in an induced *M. sexta* fat body cDNA library and identified 78 positives. We carried out plaque purification and *in vivo* excision for five of the positive clones. Complete sequence analysis indicated that their cDNA inserts were 1.5 or 2.1 kb. The long form contained an additional 0.6 kb sequence in the 3' untranslated region (Fig. 1). A polyadenylation signal at the end of the coding region is probably responsible for the short transcript. We identified eight nucleotide substitutions in the open reading frames, one of which was nonsynonymous – Thr[109] is replaced by

Asn[109]. Such nucleotide differences may represent allelic variations in the serpin-6 gene.

The longest cDNA contains a complete open reading frame spanning nucleotides 113-1531. Its corresponding protein is 412 residues long and includes the five peptide sequences of *M. sexta* serpin-6 (Fig. 1). Following a predicted 17-residue secretion peptide, the mature protein starts with a Gln, consistent with the observation that serpin-6 is blocked at the amino terminus (Wang and Jiang, 2004), presumably by pyroGlu. The predicted mature protein (395 residues) has a calculated molecular mass of 44,957 Da, smaller than the experimental value (46,710 ± 10 Da). The difference (~1,153 Da) may result from *N*-linked glycosylation at Asn[100] and/or Asn[312] – serpin-6 binds to concanavalin A, and Asn[100] is covalently modified (Wang and Jiang, 2004). The calculated isoelectric point of serpin-6 is 6.1, higher than the experimental value of 5.4.

BLAST search of GenBank revealed that *M. sexta* serpin-6 is most similar in amino acid sequence to *D. melanogaster* serpin-5 and *A. gambiae* serpin-9 (Fig. 2). It is 39% identical and 55% similar to *Drosophila* serpin-5, which may regulate the Toll and Imd pathways (De Gregorio et al., 2001; De Gregorio et al., 2002). *M. sexta* serpin-6 is 36% identical and 53% similar to *A. gambiae* serpin-9, but only 25% and 27% identical to *M. sexta* serpin-1J and sepin-3, respectively. While the reactive center loops of serpins are hypervariable, our multiple sequence alignment showed that the loop sequences of *M. sexta* serpin-6, *D. melanogaster* serpin-5, and *A. gambiae* serpin-9 are strikingly similar (Fig. 2). The predicted scissile bond is located between Arg and Ser residues, suggesting that these three serpins inhibit proteinases with a trypsin-like specificity (*e.g.* PAP-3). The loop sequences are less similar to the corresponding regions in *M. sexta* serpin-1J and serpin-3. *M. sexta* serpin-3 contains a 20-residue extension at the amino terminus,

which is absent in the other serpins.

Increases in serpin-6 mRNA and protein levels after a bacterial challenge

In order to understand the transcriptional regulation of serpin-6 gene in response to a microbial infection, we analyzed the relative mRNA levels in the normalized total RNA samples by RT-PCR (Fig. 3A). While the faint bands in naïve hemocytes and fat body represented a low, constitutive level of serpin-6 transcripts, a significant up-regulation was observed in both tissues after a bacterial challenge. Additionally, we examined the serpin-6 concentration in hemolymph by immunoblot analysis (Fig. 3B). A low level of serpin-6 was present in naïve *M. sexta* larval hemolymph. After injecting the larvae with killed *E. coli* or *M. luteus*, we detected a small decrease of the 46 kDa protein after 6 h, and the band intensity was higher than the control at 24 h post injection. The observed protein level changes probably reflected consumption and replenishing of serpin-6 after the bacterial challenge. The concentrations of serpin-6 in naïve and induced hemolymph are estimated to be 15 and 30 µg/ml, respectively.

Purification and characterization of recombinant serpin-6

To explore its biochemical functions, we produced serpin-6 in an *E. coli* expression system. The soluble serpin-6 (15 µg/ml, ~50% of the total serpin-6 produced) was purified to near homogeneity by nickel affinity and hydrophobic interaction chromatography. Serpin-6 migrated as a 44 kDa single band on a 10% SDS-PAGE under the reducing condition (Fig. 4). Trypsinolytic peptide mass fingerprint analysis of the recombinant protein revealed 39 matching peptides, representing 77% of the overall

sequence including the carboxyl-terminal fragment (data not shown).

A characteristic feature of a serpin-proteinase reaction is formation of a high M_r, SDS-stable complex of the serpin and its cognate enzyme. We conducted an experiment to test whether serpin-6 can form such a complex with PAP-3. In the control of serpin-6 or PAP-3 only, PAP-3 antibodies recognized the two polypeptide chains of the proteinase but not serpin-6 (Fig. 5A). After the two proteins were incubated together, the light chain of PAP-3 remained at the 21 kDa position, but the 37 kDa catalytic domain completely disappeared. Instead, a new immunoreactive band migrated to the 75 kDa position anticipated for a complex of serpin-6 and PAP-3 catalytic domain. This 75 kDa band was also recognized by serpin-6 antibodies (Fig. 5B). Therefore, the recombinant serpin-6 was an active inhibitor which formed a covalent complex with PAP-3.

After PAP-3 was incubated with the recombinant serpin at different molar ratios, the residual amidase activity decreased linearly as serpin-6 concentration increased (Fig. 6, A and B). Complete inhibition occurred at a molar ratio of 2:1 (serpin-6:PAP-3). Moreover, recombinant serpin-6 also blocked proPO activation by PAP-3 and SPHs. To test if serpin-6 could inhibit proPO activation in hemolymph, we obtained the 0~50% ammonium sulfate fraction of *M. sexta* plasma, which contained all the components for proPO activation. Added serpin-6 at 45 and 140 ng/μl blocked proPO activation by 50 and 90%, respectively (Fig. 6C). This result suggested that serpin-6 inhibits one or more of the serine proteinases in the proPO activation system.

To further characterize the serpin-proteinase reaction, we determined the second-order association rate constant (k_{ass}) for the inhibition (Fig. 7). The k_{ass} ($2.6 \times 10^4 \, \mathrm{M^{-1} s^{-1}}$) indicated that serpin-6 may contribute to the inhibitory regulation of PAP-3 in

hemolymph.

We determined the cleavage site of serpin-6 upon reacting with PAP-3 by MALDI-TOF mass spectrometry (Fig. 8). A major peak of 4,639 Da was detected in the proteinase-inhibitor mixture, but not in the control spectra of serpin-6 and PAP-3 only. This peak had exactly the same mass as the carboxyl-terminal peptide released from a cleavage of serpin-6 between R^{373} and S^{374}.

Isolation and characterization of hemolymph proteins associated with serpin-6 – We activated the induced larval hemolymph with *M. luteus* and examined whether or not serpin-6 and PAP-3 form a complex in plasma. Immunoaffinity chromatography using serpin-6 antibodies allowed us to isolate serpin-6 and its associating proteins from the plasma sample. We analyzed the proteins eluted from the affinity column by immunoblot (Fig. 9) and trypsinolytic peptide mass fingerprint analyses. While most serpin-6 remained intact, part of it was either cleaved or covalently linked with PAP-3 or hemolymph proteinase 8 (HP8) (Fig. 9, B-D). Formation of the serpin-proteinase complexes markedly increased in the induced hemolymph after a bacterial elicitation. The peptide mass fingerprint of the upper band had 12 mass fits with serpin-6 and 8 with the PAP-3 catalytic domain (Fig. 9A). The lower band represents the complex of serpin-6 and HP8: 11 and 5 mass peaks match those calculated from sepin-6 and HP8 proteinase domain sequences, respectively. These results indicate that serpin-6 is a physiological regulator of PAP-3 and HP8 during immune responses.

Immunoblot analysis indicated that several other immune proteins also tightly associated with the antibody column. These include: *M. sexta* immulectin-2, proPO, PO, SPH-1, SPH-2, and HP14 (Fig. 9, E-I). Peptide mass fingerprint analysis confirmed the

presence of these proteins. For instance, peptide map of the 37 kDa immunoreactive band (Fig. 9E) covers 40% of immulectin-2 sequence. Additionally, thirteen trypsinolytic peptides of the 15 kDa band (Fig. 9A) match 58% of attacin-2 sequence. While similar amounts of serpin-6 and immunlectin-2 were present in the naïve and induced hemolymph, we isolated more attacin-2, proPO, PO, SPHs, HP14 and serpin-6-proteinase complexes from the induced hemolymph (Fig. 9A). This result suggests that the secondary defense reactions involve more molecules in hemolymph from the bacteria-challenged larvae.

Discussion

It is common in vertebrates and invertebrates that extracellular serine proteinase cascades mediate acute-phase responses upon microbial infection. These pathways are often regulated by irreversible inhibitors of the serpin superfamily (Silverman et al., 2001; Jiang and Kanost, 2000; Kanost, 1999). In this study, we have cloned and characterized the newly discovered serpin-6 from *M. sexta* hemolymph. It is constitutively synthesized in the larval fat body and hemocytes. After a microbial challenge, induced transcription and translation of serpin-6 gene lead to a higher protein level in the plasma. Recombinant serpin-6 formed covalent complex with PAP-3 and inhibited proPO activation by the purified enzyme and in the hemolymph. The *in vivo* half-life of PAP-3, calculated as $1/(k_a[I])$, is 2.4 min in naïve larval hemolymph and 1.2 min in induced hemolymph. Endogenous serpin-6 was identified in an SDS-stable complex with PAP-3 in the hemolymph after *M. luteus* stimulation. These results provide good evidence that serpin-6 contributes to the regulation of PAP-3 under physiological

conditions. Along with our previous work on *M. sexta* serpin-1J and serpin-3 (Jiang et al.,
2003b; Jiang and Kanost., 1997; Zhu et al., 2003), we demonstrate in this report that
multiple serpins inhibit a single proteinase in the insect hemolymph. Comparative kinetic
analysis should allow us to evaluate relative importance of these three serpins in the
regulation of PAP-1, PAP-2, and PAP-3.

Considering the low sequence conservation of the serpin family, we believe the
high similarity among *M. sexta* serpin-6, *D. melanogaster* serpin-5 and *A. gambiae*
serpin-9 is significant. In particular, their P4-P4' residues are nearly identical in this
typically hypervariable region (Fig. 2). Since this is the site where serpins specifically
interact with their target enzymes, we anticipate that *M. sexta* serpin-6 and the dipteran
serpins may perform similar physiological functions. The expression profile of serpin-6
appears to be consistent with that of *Drosophila* serpin-5, a potential regulator of the Toll
and Imd pathways (De Gregorio et al., 2002; Irving et al., 2001).

In contrast, serpin-6 is only 25% and 27% identical in sequence to serpin-1J and
serpin-3. Its reactive center loop is 2 residues shorter. The P4-P4' region of serpin-6
(Phe-Gly-Phe-Arg-Ser-Ser-Arg-Pro) is remarkably different from those of serpin-1J
(Leu-Thr-Asp-Arg-Cys-Cys-Ser-Asp) and serpin-3 (Ile-Gln Asn-Lys-Phe-Gly-Glu Asp),
except for the positively charged residue at the P1 site, which fits the negatively charged
primary specificity pocket of PAPs. The P4-P4' regions in *M. sexta* serpin-3 and its
orthologs from *H. cunea* and *D. melanogaster* best resemble the proPO sequences around
the cleavage activation site (De Gregorio et al., 2002; Ligoxygakis et al., 2002; Park et
al., 2002).

Since serine proteinase genes greatly outnumber serpin genes in the *D.*

melanogaster and *A. gambiae* genomes, we postulated that a single insect serpin may regulate multiple serine proteinases in one or more pathways (Wang and Jiang, 2004). Genetic analyses provide support for such hypothesis: loss-of-function mutations in a serpin gene can lead to pleiotropic phenotypes (De Gregorio et al., 2002; Ligoxygakis et al., 2002; Levashina et al., 1999). In this paper, we present direct biochemical evidence that serpin-6 regulates two proteinases (*i.e.* PAP-3 and HP8) in the hemolymph. While the involvement of PAP-3 in proPO activation is established, we have not yet elucidated the biological function of HP8. Of the 18 hemolymph proteinases we recently cloned, HP8 is most similar in sequence to *H. diomphalia* PPAF-III, a Ca^{2+}-dependent activating enzyme for PPAF-II precursor (Kim et al., 2002). Further analysis is required to test the effect of serpin-6 on blocking the proteolytic processing of *M. sexta* proSPH-1 and proSPH-2, as a step toward understanding the physiological roles of serpin-6 in regulating proPO activation. A high M_r complex of SPH-1 and SPH-2 is structurally similar and functionally equivalent to *H. diomphalia* PPAF II.

Association of hemolymph proteins to the serpin antibody column provides us a unique opportunity to examine a macromolecular complex formed in the insect defense response. This complex associated with serpin-6 specifically, since we did not detect any of these proteins bound to the control column coupled with the preimmune antibodies (Zou and Jiang, data not shown). Due to the stringent washing conditions and absence of abundant hemolymph proteins in the eluate, direct or indirect associations of these proteins with serpin-6 appear to be strong and specific. Additionally, the composition of such complex appears to be quite stable. Many of these proteins were also identified in the bound fractions from the immunoaffinity columns for *M. sexta* serpin-4 and serpin-5,

negative regulators of the proPO activation system (Tong et al., unpublished results).

Among the co-purified proteins, immulectin-2 bound to lipopolysaccharide of Gram-negative bacteria and stimulated the proPO activation (Yu and Kanost, 2000). SPH-1 and SPH-2, which associated with immulectin-2 (Yu et al., 2003), were also identified as major proteins in this experiment (Fig. 9, E, G and H). These two proteins are required by PAP to generate active PO. ProPO and PO were detected in the complex and so was HP14, an initiating proteinase of the proPO activation system (Fig. 9, F and I) (Ji et al., 2004). While these proteins all participate in melanization, we have also co-eluted several minor components that may be also involved in proPO activation and/or other defense reactions. These include HP17 and HP21 (Zou and Jiang, data not shown).

Much to our surprise, *M. sexta* attacin-2 in the induced hemolymph also associated with the antibody column (Fig. 9A), we expected antimicrobial peptides to directly and independently interact with bacterial surfaces. The detection of attacin-2 suggests that recognition specificity and binding strength of the immune protein complex may be enhanced by associating with a bacteria-killing protein.

Taken together, our results strongly suggest that a bacterial entry may elicit a subset of hemolymph proteins to associate and form a non-covalent protein complex. The molecular interactions among pattern recognition proteins, HPs, SPHs, proPO and other proteins in the complex ensure a localized defense reaction against the invading organisms. After exerting their functions, active serine proteinases are inhibited by multiple serpins to limit potential damage to the host tissues and cells. Serpin affinity chromatography has allowed us to observe the immune complex for the first time. Further analyses of this complex may lead to the elucidation of proPO activation cascade and

other immune proteinase pathway in *M. sexta*.

References

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., et al. (2000) The genome sequence of *Drosophila melanogaster. Science* 287, 2185-2195.

Ashida, M. and Brey, P.T. (1998) Recent advances on the research of the insect prophenoloxidase cascade. In: Molecular Mechanisms of Immune Responses in Insects (Brey, P.T. and Hultmark, D. Eds.) pp. 135-172, *Chapman and Hall*, London.

Ashida, M. and Yamazaki, H.I. (1990) Biochemistry of phenoloxidase system in insects: with special reference to its activation. In: Molting and Metamorphosis (Ohnishi, E. and Ishizake, H. Eds.) pp. 237-263, *Jpn. Sci. Soc. Press/Springer,* Berlin.

Askew, D. J., Askew, Y. S., Kato, Y., Luke, C. J., Pak, S. C., Brömme, D., and Silverman, G. A. (2004) The amplified mouse squamous cell carcinoma antigen gene locus contains a serpin (Serpinb3b) that inhibits both papain-like cysteine and trypsin-like serine proteinases. *Genomics* 84**,** 166-175

Bode, W. and Huber, R. (1992) Natural protein proteinase inhibitors and their interaction with proteinases. *Eur J Biochem* 204, 433-451.

Cerenius L. and Söderhäll K. (2004) The prophenoloxidase-activating system in invertebrates. *Immunol Rev* 198, 116-126.

Cherqui, A., Cruz, N., and Simoes, N. (2001) Purification and characterization of two serine protease inhibitors from the hemolymph of Mythimna unipuncta. *Insect Biochem Mol Biol* 31, 761-769.

Christophides, G.K., Zdobnov, E., Barillas-Mury, C., Birney, E., Blandin, S., et al. Immunity-related genes and gene families in *Anopheles gambiae*. (2002) *Science* 298, 159-165.

De Gregorio, E., Han, S.J., Lee, W.J., Baek, M.J., Osaki, T., Kawabata, S., Lee, B.L., Iwanaga, S., Lemaitre, B., and Brey, P.T. (2002) An immune-responsive Serpin regulates the melanization cascade in *Drosophila*. *Dev Cell* 3, 581-592.

De Gregorio, E., Spellman, P.T., Rubin., G.M., and Lemaitre, B. (2001) Genome-wide analysis of the *Drosophila* immune response by using oligonucleotide microarrays. *Proc Natl Acad Sci USA* 98, 12590-12595.

De Gregorio, E., Spellman, P.T., Tzou, P., Rubin, G.M., and Lemaitre, B. (2002)) The Toll and Imd pathways are the major regulators of the immune response in *Drosophila*. *EMBO J* 21, 2568-2579.

Dunn, P. and Drake, D. (1983) Fate of bacteria injected into naïve and immunized larvae of the tobacco hornworm, Manduca sexta. *J Invert Pathol* 41, 77-85.

Gan, H., Wang, Y., Jiang, H., Mita, K., and Kanost, M.R. (2001) A bacteria-induced, intracellular serpin in granular hemocytes of *Manduca sexta. Insect Biochem Mol Biol* 31, 887-898.

Gettins, P.G. (2002) Serpin structure, mechanism, and function. *Chem Rev* 102, 4751-4804.

Irving, J.A., Pike, R.N., Lesk, A.M., and Whisstock, J.C. (2000) Phylogeny of the serpin superfamily: implications of patterns of amino acid conservation for structure and function. *Genome Res* 10, 1845-1864.

Irving, P., Troxler, L., Heuer, T.S., Belvin, M., Kopczynski, C., Reichhart, J.M.,

Hoffmann, J.A., and Hetru, C. (2001) A genome-wide analysis of immune responses in *Drosophila*. *Proc Natl Acad Sci USA* 98, 15119-15124.

Iwanaga, S., Kawabata, S., and Muta, T. (1998) New types of clotting factors and defense molecules found in horseshoe crab hemolymph: their structures and functions. *J Biochem* 123, 1-15

Ji, C., Wang, Y., Guo, X., Hartson, S., and Jiang, H. (2004) A pattern recognition serine proteinase triggers the prophenoloxidase activation cascade in the tobacco hornworm, Manduca sexta. *J Biol Chem* 279, 34101-34106.

Jiang, H. and Kanost, M.R. (1997) Characterization and functional analysis of 12 naturally occurring reactive site variants of serpin-1 from *Manduca sexta*. *J Biol Chem* 272, 1082-1087.

Jiang, H. and Kanost, M.R. (2000) The clip-domain family of serine proteinases in arthropods. *Insect Biochem Mol Biol* 30, 95-105.

Jiang, H., Wang, Y., and Kanost, M.R. (1994) Mutually exclusive exon use and reactive center diversity in insect serpins. *J Biol Chem* 269, 55-58.

Jiang, H., Wang, Y., and Kanost, M.R. (1998) Pro-phenoloxidase activating proteinase form an insect, *Manduca sexta*: a bacterica-inducible protein similar to Drosophila east. *Proc Natl Acad Sci USA* 95, 12220-12225.

Jiang, H., Wang, Y., Huang, Y., Mulnix, A.B., Kadel, J., Cole, K., and Kanost, M.R. (1996) Organization of serpin gene-1 from *Manduca sexta*. Evolution of a family of alternate exons encoding the reactive site loop. *J Biol Chem* 271, 28017-28023.

Jiang, H., Wang, Y., Ma, C., and Kanost, M.R. (1997) Subunit composition of pro-phenol oxidase from *Manduca sexta*: molecular cloning of subunit ProPO-P1. *Insect*

Biochem Mol Biol 27, 835-50.

Jiang, H., Wang, Y., Yu, X.Q., and Kanost, M.R. (2003a) Prophenoloxidase-activating proteinase-2 from hemolymph of *Manduca sexta*: a bacteria-inducible serine proteinase containing two clip domains. *J Biol Chem* 278, 3552-3561.

Jiang, H., Wang, Y., Yu, X.Q., Zhu, Y., and Kanost, M. (2003b) 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.

Kanost, M.R. (1990) Isolation and characterization of four serine proteinase inhibitors (serpins) from hemolymph of *Manduca sexta*. *Insect Biochem* 20, 141-147.

Kanost, M.R. (1999) Serine proteinase inhibitors in arthropod immunity. *Dev Comp Immunol* 23, 291-301.

Kanost, M.R., Prasad, S.V., and Wells, M.A. (1989) Primary structure of a member of the serpin superfamily of proteinase inhibitors from an insect, *Manduca sexta*. *J Biol Chem* 264, 965-972.

Kim, M.S., Baek, M.J., Lee, M.H., Park, J.W., Lee, S.Y., Söderhäll, K., and Lee, B.L. (2002) A new easter-type serine protease cleaves a masquerade-like protein during prophenoloxidase activation in *Holotrichia diomphalia* larvae. *J Biol Chem* 18, 39999-40004.

Lee, S.Y., Kwon, T.H., Hyun, J.H., Choi, J.S., Kawabata, S.I., Iwanaga, S., and Lee, B.L. (1998) In vitro activation of pro-phenol-oxidase by two kinds of pro-phenol-oxidase-activating factors isolated from hemolymph of coleopteran, *Holotrichia diomphalia* larvae. *Eur J Biochem* 254, 50-57.

Levashina, E.A., Langley, E., Green, C., Gubb, D., Ashburner, M., Hoffmann, J.A., and Reichhart, J.M. (1999) Constitutive activation of toll-mediated antifungal defense in serpin-deficient *Drosophila*. *Science* 285, 1917-1919.

Ligoxygakis, P., Pelte, N., Ji, C., Leclerc, V., Duvic, B., Belvin, M., Jiang, H., Hoffmann, J.A., and Reichhart, J.M. (2002) A serpin mutant links Toll activation to melanization in the host defence of *Drosophila*. *EMBO J* 21, 6330-6337.

Nappi, A.J. and Vass, E. (2001) Cytotoxic reactions associated with insect immunity. *Adv. Exp. Med. Biol.* 484, 329-348.

Park, D.-S., Shin, S. W., Hong, S.-D., and Park, H.-Y. (2000) Immunological detection of serpin in the fall webworm, *Hyphantria cunea* and its inhibitory activity on the prophenoloxidase system. *Mol Cell* 10, 186-192.

Sasaki, T. (1991) Patchwork-structure serpins from silkworm (*Bombyx mori*) larval hemolymph. *Eur J Biochem.* 202, 255-261.

Sasaki, T. and Kobayashi, K. (1984) Isolation of two novel proteinase inhibitors from hemolymph of silkworm larva, *Bombyx mori*. Comparison with human serum proteinase inhibitors. *J. Biochem* 95, 1009-1117.

Satoh, D., Horii, A., Ochiai, M., and Ashida, M. (1999) Prophenoloxidase-activating enzyme of the silkworm, *Bombyx mori*. Purification, characterization, and cDNA cloning. *J Biol Chem* 274, 7441-7453.

Silverman, G.A., Bird, P.I., Carrell, R.W., Church, F.C., Coughlin, P.B., et al. (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.

Stark, K.R. and James, A.A. (1998) Isolation and characterization of the gene encoding a novel factor Xa-directed anticoagulant from the yellow fever mosquito, *Aedes aegypti*. *J Biol Chem* 273, 20802-20809.

Wang Y., Jiang, H., and Kanost, M.R. (2001) Expression and Purification of *Manduca sexta* Prophenoloxidase-Activating Proteinase Precursor (proPAP) from Baculovirus-Infected Insect Cells. *Protein Exp Purif* 23, 328-337.

Wang, Y. and Jiang, H. (2004) Prophenoloxidase (proPO) activation in *Manduca sexta*: an analysis of molecular interactions among proPO, proPO-activating proteinase-3, and a cofactor. *Insect Biochem Mol Biol* 34, 731-742.

Wang, Y. and 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.

Yu, X-Q and Kanost, M.R. (2000) Immulectin-2, a lipopolysaccharide-specific lectin from an insect, *Manduca sexta*, is induced in response to gram-negative bacteria. *J Biol Chem* 275, 37373-37381.

Yu, X-Q, Jiang, H., Wang, Y., and Kanost, M.R. (2003) Nonproteolytic serine proteinase homologs are involved in prophenoloxidase activation in the tobacco hornworm, *Manduca sexta*. *Insect Biochem Mol Biol* 33, 197-208.

Zhu, Y., Wang, Y., Gorman, M.J., Jiang, H., and Kanost, M.R. (2003) *Manduca sexta* serpin-3 regulates prophenoloxidase activation in response to infection by inhibiting prophenoloxidase-activating proteinases. *J Biol Chem* 278, 46556-46564.

```
                     GCGCTACGTCGTGCGTGTCGCCGCAATCTTTTAATACAACAAATCACGTATAAATTATACTAA   63
GAATACGCTTGTATTTAAAAAAAATATATTTTAGTGACACATACAGCTACCTAATTTGCATTCAAAACTTAGCGTGTTACTTTCGAAGAAA  153
CTTTACCTGCAGTCGCGTACGGACATCTGCCGATGCAATATGTTGAAGTCAGCGGCTCTGGTACTGTTAGTAGCGACGTGTGTAAGCAGT  243
                                                 M  L  K  S  A  A  L  V  L  L  V  A  T  C  V  S  S    -1
CAATGTTTCTCCAAAGATGATTCATCTAAGAAGTTGGATCCGGGGGCGAGGACATCCTTGTATTCAGGGCAATTGGCGTTCACACTGAAC   333
 Q  C  F  S  K  D  D  S  S  K  K  L  D  P  G  A  R  T  S  L  Y  S  G  Q  L  A  F  T  L  N    30
CTGTTTCAGACAATTAATAGCGCGGTTCCCGACGATAATATATTCTTCTCACCCTTTTCTGTATATCAATCCCTATTACTCGCCTACTTT  423
 L  F  Q  T  I  N  S  A  V  P  D  D  N  I  F  F  S  P  F  S  V  Y  Q  S  L  L  L  A  Y  F    60
TCTACCGGTGGCCGAACTGAAGAATCTCTAAAGAAATCGTTGGAAATTGAAGATAACATGGACAAGATGAACTTGATGACTGCGTATAAG  513
 S  T  G  G  R  T  E  E  S  L  K  K  S  L  E  I  E  D  N  M  D  K  M  N  L  M  T  A  Y  K    90
GTAGACAAGAGGTCTCGCATGACGAACAACAACAGCGATAGCTATGAGTTCACCACCGCCGCCAACAAACTGTTCGTCGCAAACGAGCTGCAA  603
 V  D  K  R  S  R  M  T  N  N♦ N  S  D  S  Y  E  F  T  T  A  N  K  L  F  V  A  N  E  L  Q    120
GTGCGCCAGTGCATGTTTGACCTCTTCGGAGAAGAAATTGAAGCGTTGAACTTCCGAGAAAATCCCGAGGTCAGTCGTGAATACATCAAC  693
 V  R  Q  C  M  F  D  L  F  G  E  E  I  E  A  L  N  F  R  E  N  P  E  V  S  R  E  Y  I  N    150
AATTGGGTTGAACGAATTACTAAAAACCACATCAAGAAGTTGTTGCCCGCTGACGGAGTGAGCGAATTTACGAAATTGGTGCTTGCGAAT  783
 N  W  V  E  R  I  T  K  N  H  I  K  K  L  L  P  A  D  G  V  S  E  F  T  K  L  V  L  A  N    180
GCCGCTTACTTCAAAGGAGTATGGGCATCAAAGTTTTCCCCTGAAAGAACCAAAAAGGAACCTTTCTTTGTTTCTGAAACCCGACAGACG  873
 A  A  Y  F  K  G  V  W  A  S  K  F  S  P  E  R  T  K  K  E  P  F  F  V  S  E  T  R  Q  T    210
CTTGTGCCTTTCATGAAGCAGAAGGGAACTTTCCACTACGGTGTAAGCGAAGAGCTGGGCGCCCAAGTCTTGGAACTGCCTTATAAGGGT  963
 L  V  P  F  M  K  Q  K  G  T  F  H  Y  G  V  S  E  E  L  G  A  Q  V  L  E  L  P  Y  K  G    240
AACGACATTAGCATGTTCATCCTCCTGCCACCCTACTCCATGAAAGAAGGTGTTACCAACATCATCGCCAACTTGAATACTGAACGCCTC  1053
 N  D  I  S  M  F  I  L  L  P  P  Y  S  M  K  E  G  V  T  N  I  I  A  N  L  N  T  E  R  L    270
GCTGCAGTTATGGAAGAGAGCTACATGAGTCGAGAAGTTATTGTCGAAATCCCCAAATTCACCATTGAAGAACTCTATCACTTAGACCT  1143
 A  A  V  M  E  E  S  Y  M  S  R  E  V  I  V  E  I  P  K  F  T  I  E  R  T  L  S  L  R  P    300
ATTCTGGATAGACTTGGTGTAGGCGATCTCTTCAACGTTAGTGCAGATTTCAGCACTCTTACAGAGGACTCCGGCATTCGATTCGACGAC  1233
 I  L  D  R  L  G  V  G  D  L  F  N♦ V  S  A  D  F  S  T  L  T  E  D  S  G  I  R  F  D  D    330
GCCGTGCACAAAGCGAAAATACAAATCGACGAAGAAGGTACTGTAGCGGCTGCGGCCACAGCTCTGTTCGGGTTCAGGTCGTCTAGGCCG  1323
 A  V  H  K  A  K  I  Q  I  D  E  E  G  T  V  A  A  A  T  A  L  F  G  F  R  S  S  R  P    360
                                                                          P1 P1'
GCGGAGCCCACTCGCTTCATCGCCAACTTTCCGTTCGTCTATCTTATTTACGAACGTCCCACCAATTCTATTCTTTCTTCGGAGTCTAC  1413
 A  E  P  T  R  F  I  A  N  F  P  F  V  Y  L  I  Y  E  R  P  T  N  S  I  L  F  F  G  V  Y    390
                                                              AAAAAAAAAA
AGGGACCCTAAGAAATAAATACTTATATGTACATAAATAATTATAATTTATCTAACCGAACTCTTACTCGTTTTCAATATGACAACACTC  1503
 R  D  P  K  K  *                                                                            395
TCTGTGATGCTTACGCAATATCTTTACACATTGTGATAAGTGATCTTTTAATATTCGATTGTAATTTTGATAGTGTATATTTTAGTTTTG  1593
AAGCTATTTAACCAACTTCGTTGTACGTTATTTCTAGCGCCTTTTTTCTGAGGTATTTATTCATAATTTATTGTCCAATAATAAAAATAG  1683
TAGGTCCAATTCAAAATAACTTTTGCTAATAGACTTGAGAATTAATTTATTAGTATCTAATATAAGCTGGGAGGTCTTTCGTATTCCAAA  1773
GTCCGCAGATACATGTGGGTATCATTGCTTTGTCTATCACTGCGGCCAAGCAACAGTGTGCGAAAGCGGTGCCGTGCAGTATTTTACAGT  1863
CGGTAGTGATACTGTGCGACCAGTTCCTGTGTCTCATCATGAAGTTGTATAAAGATATTTGTTTATTTTTAATGATTACTAAAGTACTTC  1953
GTATCGTTTTGTGTCATCAGCTATATTGATGTTTATCACTTTGTGTTATTATTGTTTCGTAGGACTTAATACCACATATATTTAATTTAA  2043
ATCTATTATAATTAGGGAAGTAACATTAAGTAATTTGTAAAATTATTTTACTGAAATAAGTTTAATCAATAAACCTATTAATACTGTAAA  2133
AAAAAAAAAAAAAAAAAAAAA                                                                        2154
```

Fig. 1 Nucleotide and deduced amino acid sequence of *M. sexta* serpin-6. The deduced amino acid sequence from long form of cDNA is shown below the nucleotide sequence. The one-letter codes are aligned with the second nucleotide of each codon. The 17-residue signal peptide is double underlined. Two putative *N*-linked glycosylation sites are marked with ◆ after the Asn residues. Underlined sequences perfectly match the five trypsinolytic peptides from purified serpin-6 (1). The predicted cleavage site is indicated as P1. Polyadenylation signals (AATAAA) near the 3' terminal of the cDNA are bold faced. A polyA sequence in italics, above the nucleotide sequence, indicates the termination of short form of serpin-6 cDNA. Substitutions of nucleotides in the open reading frame are underlined. A nonsynonymous substitution (A**C**C to A**A**C) changes T[109] to N[109].

```
                                                                            .
Ms serpin-6    MLKSAALVLLVATCVSS-------------------------QCFSKDDS------SKKLDPGARTSLYSGQLAFTLN
Ag serpin-9    MKLSTHLGTVAPLLLLTLLIVPHHGTNG---------------QCSPGDEKKAEK--ATDLEQ-SRSRLYKGESIFTLR
Dm serpin-5    MHILSISLMAVLPAIALA-----------------------GLCGVEPDA------G-LLDQ--RLNLYKGQQNFAVS
Ms serpin-3    MTSTIYFAFFVAPLLLCSLADDVDPNTLRAVFGYSALDQAALVGAESNKQAATVTPDNGTLVDPDYWDTEEFQPSTADYD
Ms serpin-1J   MKIIMCIFGLAALAMA-----------------------GETDLQKILRES---------NDQ-----------FTAQ


               ..  .              *...* .       *          .   *    *.
Ms serpin-6    LFQTIN----SAVPDDNIFFSPFSVYQSLLLAYFSTGGRTEESLKKSLEIEDNMDKMNLMTAYKVDKRSRMTNNN-SDSY
Ag serpin-9    LLDAIN----TATPNENLFFSPYSLYNVLLMMYFGARDTTEKLLRTSLNLQWADSKTTVYEAYDTARKSLRGRFSESNAV
Dm serpin-5    MLNVIR----QSTPNENVFFSPYSTYHALLLAYFGSSGDTEKELAKVLHLDWADSKEVVRSAYILEKMNRKERQS-KMPL
Ms serpin-3    IFDWVLTKRVASTSNANFLLSPLGLKLALAILTEAATGSTRSELASALGFGLDRTEVRRKFSTIIESLKRESPDY-ILNL
Ms serpin-1J   MFSEVV----KANPGQNVVLSAFSVLPPLGQLALASVGESHDELLRALALPNDNVTKDVFADLNRGVRAVKGVDL-KMAS


                        .                  .     .              ** *.   *  .*  ..          . *. *.*
Ms serpin-6    EFTTANKLFVANELQVRQCMFDLFGEEIEALNFRENPEVSREYINNWVERITKNHIKKLLPADGVSEFTKLVLANAAYFK
Ag serpin-9    GFSSADKLFFGRQIPVSTCVQQKFADTIELLDYQTQPDEQRAYINRWVENATRGQIKDLLEPGAITRNTKLAVANAAYFK
Dm serpin-5    EFSSADRIFFANDLHVTECARNRLAEEVQQIDFKSQTEESRKQINDWIAKQTHDQIRNMLSADEITPRTRLVLANAAYLK
Ms serpin-3    GSRIYMGEGVQPRQRFAAIAQEFYKTELKTTNFFKPEVAAR-EINNWVSNATQGKIPNLVEADDVAD-VILLILNTLYFK
Ms serpin-1J   KIYVAKGLELNDDFAAVS--RDVFGSEVQNVDFVKSVEAAG-AINKWVEDQTNNRIKNLVDPDALDETTRSVLVNAIYFK


               * *  .*      *      *    *      * *  .   .    . ..... *                         .*
Ms serpin-6    GVWASKFSPERTKKEPFFVSETRQTLVPFMKQKGTFHYGVSEELGAQVLELPYK-------------------GNDISM
Ag serpin-9    GTWQTKFKAAETNKEIFYVSADQQKFVDMMHVEGTFSHAANEKLGCHILELPYSAGPGADNADDGPYQQGAANPDNQVSM
Dm serpin-5    GQWLSQFKTEKTVPMPFYTSPSNYSLVSMMQQKGTFLLNVDEQLRAHVLQLPYRTVFESQEKEDS-----SPDENSDISM
Ms serpin-3    GTWRHQFAPNATKPGPFYVSPQLQKTVPFMNVKDNFYYVDSSRFDAKILRLPYR-------------------GNKYAM
Ms serpin-1J   GSWKDKFVKERTMDRDFHVSKDKTIKVPTMIGKKDVRYADVPELDAKMIEMSYE-------------------GDQASM


               ....*      .  .      *              . .  .*** .*    .     .*
Ms serpin-6    FILLPPYSMKEGVTNIIANLNTER--LAAVMEESYMSREVIVEIPKFTIERTLSLRPILDRLGVGDLFNVS-ADFSTLTE
Ag serpin-9    FVFLPPAEPNA-LSKLLSRLAAETDILHEVVNEG-ISRKVDVKLPKFSIEKTVGMKPVLQKS-----------------
Dm serpin-5    VLILPPFNSNS-LEDVLSRLNADS--LDDSLKQA-MPREIEVSLPKFEFEQRLELNPILAKMGVSKMFDESVATFDDLTS
Ms serpin-3    YIVVPNSLTG--LPRVLNNLSELR--TEMIYLQE---RLVDVILPKFQFEYMSRLEGVLREMGVREAFEDT-ASFPGIAR
Ms serpin-1J   IIILPNQVDG--ITALEQKLKDPK--ALSRAEERLYNTEVEIYLPKFKIETTTDLKEVLSNMNIKKLFTPGAARLENLLK


                           *  .  * *.  *    *         P1‖P1'         .  *  **..  .         ** *    *
Ms serpin-6    ---DSG-IRFDDAVHKAKIQIDEEGTVAAAATALGF-R‖-SSRPAEPTRFIANFPFVYLIYERPTNSILFFGVYRDPKK
Ag serpin-9    ----------------KIEVNEEGSVAASATVAFSF-R‖-SSRPADPAMFHCNHPFVFLIYDYGTRSVLFNGVYRQPE
Dm serpin-5    ---ET--ISIGDSKHVAKIKVDEEGSTAAAATVLFTY-R‖-SARPVEPAKFECNHPFLFVIYDRTSRSILFTGIYRDPKTIKQ
Ms serpin-3    GQLLYQRLRVSKVFQRTGIEVNELGSVAFSATQIGIQNK‖FGEDSDINYEVVANKPFMFFIQEESTRQTLFTGRVSDPALVDGAFKA
Ms serpin-1J   ---TKESLYVDAAIQKAFIEVNEEGAEAAAANAFILTDR‖CCSDYDDNIEFDVNRPFYLNL—RTNE-HLLFSGICIQPEI
```

Fig. 2 Sequence alignment of five insect serpins. The amino acid sequences of *M. sexta* serpin-6, serpin-1J, serpin-3, *D. melanogaster* serpin-5, and *A. gambiae* serpin-9 are aligned. Amino-terminal secretion sequences are double-underlined. The predicted P1 residue before the scissile bond is double underlined. The reactive site loop extends from P16 to P5'. *, identical; ·, conservative substituted.
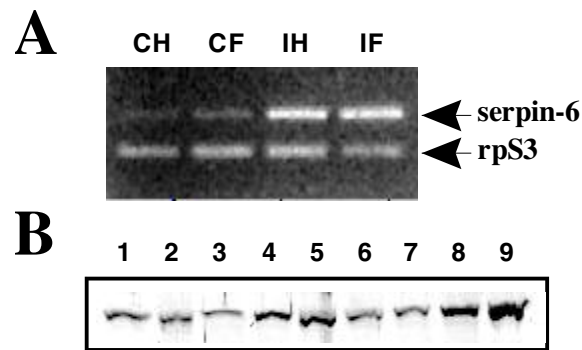
Fig. 3 Changes in PAP-3 mRNA and protein levels in bacteria-challenged *M. sexta* larvae. (**A**) Total RNA samples from hemocytes (CH, IH) or fat body (CF, IF) of control or induced insects (24 h after injection with H_2O or *E. coli*, respectively) were an alyzed by RT-PCR as described under *Experimental Procedures*. The *M. sexta* ribosomal S3 mRNA was used as an internal standard to demonstrate equal amount of RNA templates. (**B**) Cell-free hemolymph samples (2 μl) collected at 1, 6, 12, and 24 h after fifth instar day 2 larvae had been injected with 1×10^7 killed *E. coli* cells (lanes 2-5) or 100 μg *M. luteus* (lane 6-9) were subjected to 10% SDS-PAGE analysis under the reducing condition. Lane 1 was the sample at 0 h. Immunoblot analysis was performed using 1:2000 diluted serpin-6 antiserum as the first antibody.
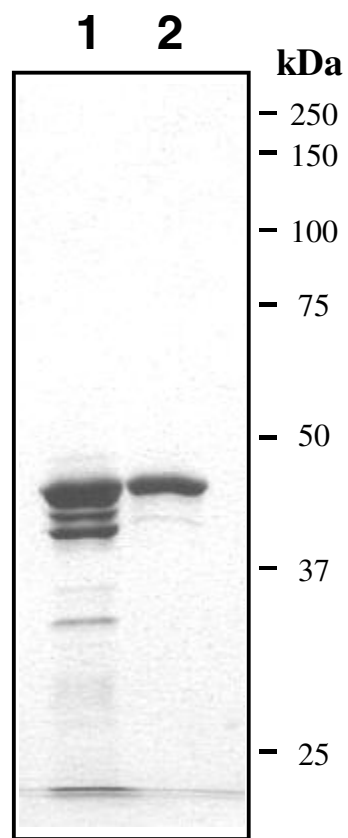
Fig. 4 Isolation of recombinant serpin-6 from *E. coli*. Purified serpin-6 is analyzed on 10% SDS-PAGE by sliver staining. Lane 1, 6 µg protein from the nickel affinity column. Lane 2, 4 µg protein eluted from the Phenyl-Sepharose column. Positions and sizes of marker proteins are indicated on the right.
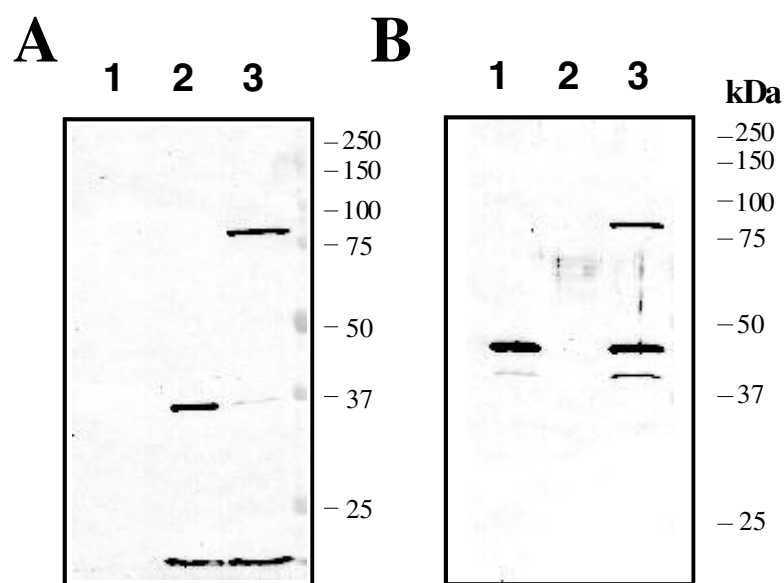
Fig. 5 Detection of high M_r, SDS-stable complex of serpin-6 and PAP-3. Purified recombinant serpin-6 was incubated with PAP-3 at room temperature for 10 min. The reaction mixtures and controls were subjected to 10% SDS-PAGE and immunoblot analysis using the PAP-3 (**A**) or serpin-6 (**B**) antibodies. Lane 1, serpin-6 only; lane 2, PAP-3 only; lane 3, serpin-6 and PAP-3. Sizes and positions of molecular mass standards are indicated on the right.

Fig. 6 Concentration-dependent inhibition of *M. sexta* PAP-3 and proPO activation by serpin-6. Purified recombinant serpin-6 was incubated with purified PAP-3 at different molar ratios for 10 min at room temperature. The residual activities (%) were plotted against the corresponding molar ratios of *M. sexta* serpin-6 and PAP-3 (**A**) amidase activity: the activity was then measured using IEAR*p*NA substrate at 405 nm. (**B**) proPO activation reaction: the above reaction mixtures were further incubated with proPO and SPHs on ice for 30 min prior to PO activity assay. (**C**) proPO activation system: The plasma fraction containing different amounts of serpin-6 was incubated with *M. luteus* for 10 min at room temperature before PO activity assay.
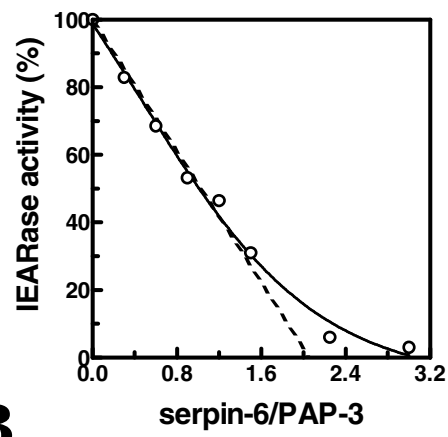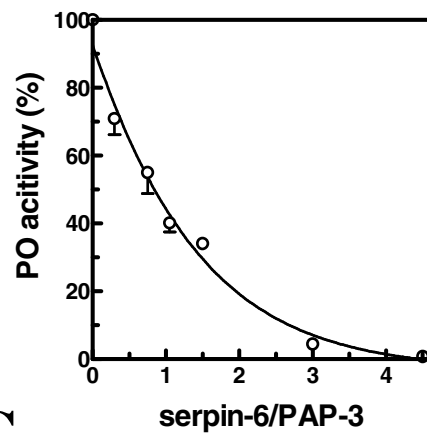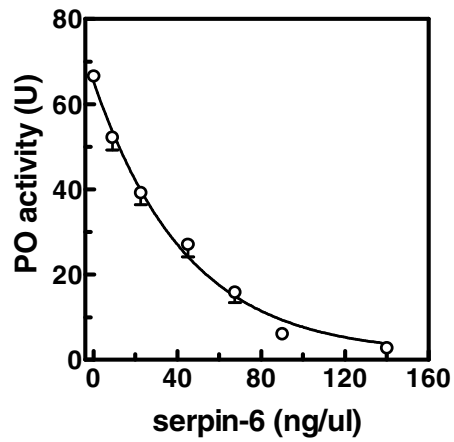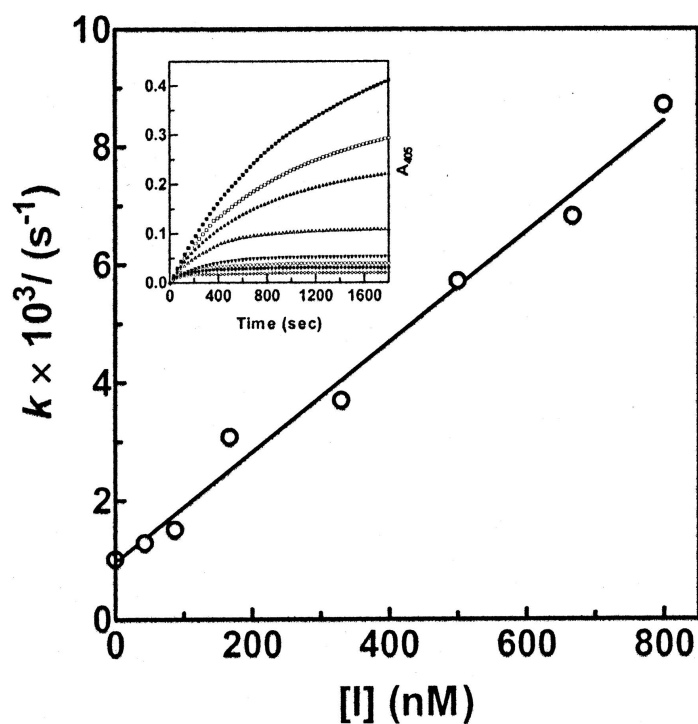
Fig. 7 Determination of association rate constant between serpin-6 and PAP-3. PAP-3 (3.6 pmol) was added to a mixture of 150 μM IEAR*p*NA and serpin-6 at 0 (■), 43 (□), 87 (▲), 167 (△), 333 (▼), 500 (○), 667 (•), and 800 (◇) nM. The progress of enzyme inactivation (*inset*) was followed by measuring absorbance at 405 nm on a tunable VERSAmax microplate reader. Pseudo-first-order rate constants of inhibition of PAP-3 (*k*) were plotted as a function of serpin-6 concentration ([I]). The association rate constant k_a was calculated by linear regression analysis and correction with [S], K_m, and SI.
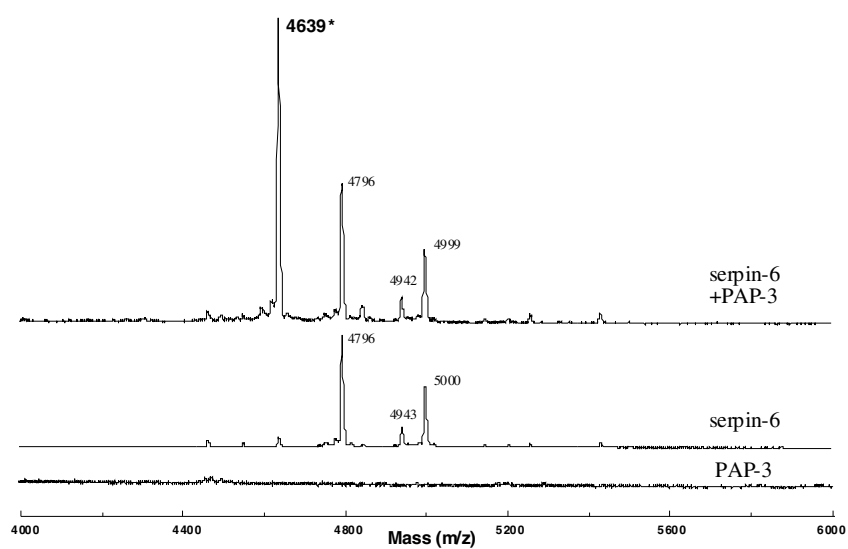
Fig. 8 Cleavage site determination of serpin-6. The reaction mixture of serpin-6 and PAP-3 was directly analyzed by MALDI-TOF mass spectrometry. A representative strong single-accumulation spectrum is presented with the mass values on top of the MH$^+$ peaks. The spectrum was subjected to noise removal and calibrated with an external standard of horse apomyoglobulin. The control spectra of serpin-6 and PAP-3 alone do not contain corresponding mass peak.
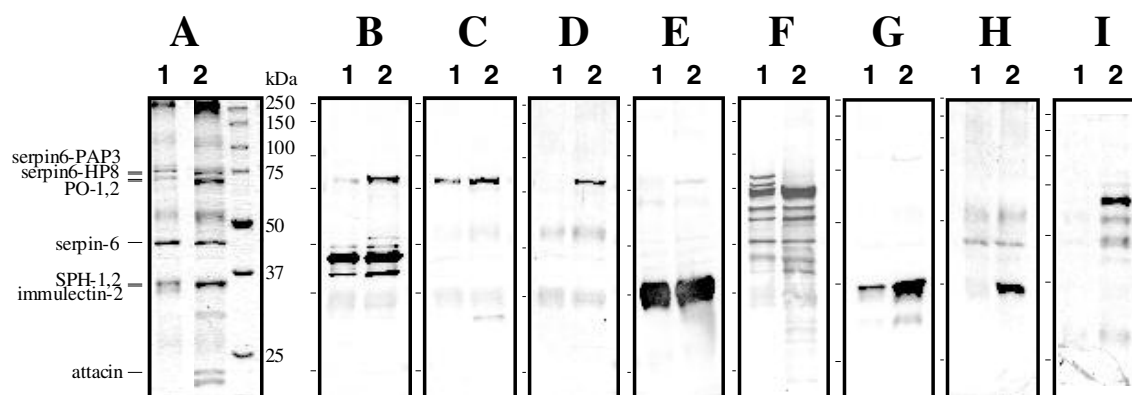
Fig. 9 Proteins eluted from the immunoaffinity column. Lane 1: naïve hemolymph samples from day 2, fifth instar larvae were incubated with 1 mM PMSF and a proteinase inhibitor cocktail (Sigma, P-8849) at 1/30 (v/v); lane 2: induced hemolymph samples activated with 1 μg/ml *M. luteus* for 0.5 h and treated with 1 mM PMSF and the proteinase inhibitor cocktail at 1:30 (v/v). The proteins were analyzed by 10% PAGE gel followed by Coomassie blue staining (A) or immunoblot analysis using serpin-6 (B), HP8 (C), PAP-3 (D), immulectin-2 (E), proPO (F), SPH-1 (G), SPH-2 (H), and HP14 (I) antibodies. *M. sexta* attacin-2 (A) was identified by in-gel trypsin digestion and MALDI-TOF mass fingerprint analysis.

Chapter VI. Recombinant expression and characterization of full-length Necrotic, a *Drosophila* serpin that efficiently inhibits chymotrypsin- and elastase-like enzymes

Abstract

Drosophila Necrotic, or Spn43Ac, is a serine proteinase inhibitor of the serpin superfamily. This serpin is thought to down-regulate an extracellular serine proteinase cascade that activates the Toll signaling pathway. To characterize the inhibitor and its interaction with a cognate enzyme, we expressed and purified the intact serpin in a baculovirus-insect cell system. The purified protein, present in a monomeric form, strongly inhibited chymotrypsin- and elastase-like serine proteinases. MALDI mass spectrometry indicated that the cleavage site of Necrotic is located between Leu438 and Ser439. Spn43Ac also inefficiently inhibited human thrombin, suggesting that its physiological target may not be Persephone, a Toll-pathway serine proteinase predicted to have a thrombin-like specificity.

Introduction

The serpins are a superfamily of proteins, most of which function as inhibitors of serine proteinases (Potempa et al, 1994; Silverman et al, 2001). They are known in mammals to regulate blood coagulation, complement activation, and proteinases released

during inflammatory responses. Serpins are also involved in controlling defense responses in invertebrates, such as hemolymph clotting, prophenoloxidase activation, and induced synthesis of antimicrobial peptides (Jiang and Kanost, 2001). The selectivity of a serpin for inhibiting a specific type of proteinase depends on the amino acid sequence of a reactive center region near the carboxyl terminus. Upon binding to the target proteinase, this region is cleaved by the enzyme to release the carboxyl terminal fragment and the rest of the serpin molecule forms a covalent complex with the proteinase (Ye and Goldsmith, 2001). While some of these acyl-enzyme intermediates rapidly dissociate, others remain stable over a prolonged period of time even in the presence of SDS. The inhibitory selectivity and substrate-inhibitor partition are primarily determined by the residues close to the scissile bond, particularly the one preceding it.

In the *Drosophila melanogaster* genome, serine proteinases and their homologs constitute the second largest gene family (Adams et al., 2000). About sixty of the proteinases contain regulatory sequences and possibly function in enzyme cascades to mediate various physiological signals during development and immune responses (Ross et al, 2003). Some of them are likely regulated by serpins. So far, twenty-seven serpin genes have been identified in the *Drosophila* genome (http://www.fruitfly.org). The existence of so many proteinases and serpins has made it extremely difficult to elucidate their biological functions through genomic, genetic and biochemical studies. Nevertheless, activation and regulation of serine proteinase systems during immune response in *Drosophila* as well as other insects has now become an exciting research focus.

Similar to other insects, *Drosophila* possesses an efficient defense system against pathogen infection. The Toll and imd pathways lead to the induced synthesis of immune

factors including antimicrobial peptides (Hoffmann and Reichhart, 2002). Limited proteolysis is required to generate the active ligand (spätzle) for triggering the Toll pathway. An extracellular serine proteinase cascade, different from the one which activates spätzle for determining the dorsal-ventral polarity of embryo, appears to be responsible for the processing of spätzle in immunity. Activation of this cascade must somehow link to the recognition of fungi and Gram-positive bacteria. Persephone, a clip-domain serine proteinase, is a component of the cascade. Mutation of its gene shuts down the Toll pathway and synthesis of drosomycin (Ligoxygakis et al., 2002a). Spn43Ac, also called Necrotic or nec serpin, down-regulates the signal transduction by inhibiting a proteinase in the pathway (Levashina et al., 1999). Recently, a truncated form of Spn43Ac was expressed in *E. coli* and found to be an inhibitor with a broad spectrum (Robertson et al., 2002). It was suggested that Spn43Ac may directly inhibit Persephone. Since the recombinant nec serpin previously tested lacked one fifth of the mature protein sequence, we wanted to test whether the amino-terminal extension could affect the inhibitory selectivity or association stoichiometry of the inhibitor.

In this work, we report the expression and purification of full-length nec serpin in a baculovirus-insect cell system. The purified protein was present in a monomeric form and strongly inhibited chymotrypsin- or elastase-like serine proteinases. The cleavage site was determined by MALDI mass spectrometry to be located between Leu^{438} and Ser^{439}. Spn43Ac didn't inhibit trypsin and inefficiently inhibited thrombin, suggesting that its physiological target may not be Persephone, a serine proteinase predicted to have a trypsin-like specificity.

Materials and methods

Construction of recombinant baculovirus for expression of Spn43Ac

The Spn43Ac cDNA, located between *EcoR*I and *Xho*I sites of pBluescript-SK (Levashina et al., 1999), was removed from the vector by *Spe*I-*Kpn*I digestion and directionally inserted to the same sites of pFastBac1 to generate a recombinant plasmid. *In vivo* transposition of the expression cassette, selection of colonies carrying the recombinant bacmid, and isolation of the bacmid DNA were performed by following the manufacturer's instructions (Invitrogen Life Technology). High molecular-weight bacmid DNA was isolated and used to transfect *Spodoptera frugiperda* Sf9 cells in the presence of CellFECTION. Titer of the initial virus stock was improved to the highest level through serial infections of Sf9 cells (Wang et al, 2001).

Expression and purification of recombinant Spn43Ac

Sf9 cells were inoculated at 2 x 10^6 cells/ml in 500 ml Sf-900 II serum-free medium (Invitrogen Life Technology) and infected with the recombinant viruses at a multiplicity of infection of 2. The culture was incubated at 27ºC for 96 h with gentle agitation (100 rpm). After the cells were removed by centrifugation at 500 x *g* for 20 min, the cleared culture supernatant was harvested and stored at -70ºC in the presence of 0.5 mM benzamidine. For isolating the recombinant protein, conditioned medium (200 ml) was thawed and incubated with 6.0 ml dextran sulfate-Sepharose (Nakamura et al, 1985) equilibrated in 10 mM potassium phosphate, 0.5 mM benzamidine, 0.5 mM phenylmethylsulfonyl fluoride, pH 6.4. The mixture was gently stirred at 4°C for 1 h and then loaded into an empty column (2.5 cm i.d. × 6.5 cm). Following a washing step,

bound proteins were eluted from the column at 0.5 ml/min for 80 min with a linear gradient of 0 - 1.0 M NaCl in the same buffer. Western blot analysis of the fractions was carried out using 1:2,000 diluted anti-Spn43Ac polyclonal antiserum and 1:3,000 diluted goat-anti rabbit IgG conjugated with alkaline phosphatase (Wang et al., 2001). Fractions containing Spn43Ac were pooled and supplemented with ammonium sulfate to a final concentration of 1.0 M. After centrifugation at 15,000 x g for 30 min, the cleared protein solution was applied to an Econo-Pac methyl HIC column (5.0 ml, Bio-Rad) and eluted with a linear gradient of 1.0 – 0 M ammonium sulfate in 10 mM potassium phosphate, pH 6.4 at 1.0 ml/min for 40 min. Concentration and buffer exchange of the Spn43Ac fractions were performed in a Centricon-30 (Millipore). The purified protein was stored at -70ºC in pH 8.5, 20 mM Tris-HCl.

Characterization of recombinant Spn43Ac

Molecular mass of Spn43Ac was determined by MALDI mass spectrometry as described previously (Jiang et al., 2003). For studying the association state of Spn43Ac, the purified protein (20 µg) was separated on an HPLC gel filtration column (Bio-Silect SEC 250, Bio-Rad) equilibrated with 0.1 M sodium phosphate, 0.1 M NaCl, pH 6.9, and absorbance was monitored at 280 nm. The column was calibrated with a mixture of molecular weight standards (Bio-Rad) including thyroglobulin (670 kDa), bovine γ-globulin (158 kDa), chicken ovalbumin (44 kDa), equine myoglobin (17 kDa), and vitamin B_{12} (1.35 kDa).

Inhibitory activity assays

Proteinase samples were individually mixed with different concentrations of Spn43Ac in the presence of 0.1 M Tris-HCl, 0.1 M NaCl, 5 mM $CaCl_2$, pH 7.8. After incubation at room temperature for 10 min, a chromogenic substrate solution (150 μl) was added for measuring the residual enzyme activities. Absorbance change at 405 nm was determined in the kinetic mode on a tunable VERSAmax microplate reader (Molecular Devices). The enzymes and their substrates used in the assays were: 0.2 μg human neutrophil cathepsin G (Athens Research and Technology, Inc.) with 100 μM N-succinyl-Ala-Ala-Pro-Phe-*p*Na (Sigma); 0.02 μg bovine pancreatic α-chymotrypsin (Sigma) with 50 μM N-succinyl-Ala-Ala-Pro-Phe-*p*Na (Sigma); 0.2 μg of human neutrophil elastase (HNE) (Athens Research and Technology, Inc.) with 100 μM N-succinyl-Ala-Ala-Pro-Leu-*p*Na (Sigma); 0.2 μg of porcine pancreatic elastase (Worthington) with 50 μM N-succinyl-Ala-Ala-Pro-Leu-*p*Na; 0.02 μg bovine pancreatic trypsin (Sigma) with 50 μM H-D-Phe-pipecolyl-Arg-*p*Na (S-2238, Amersham Pharmacia Biotech); 0.02 μg partially purified *Metarhizium anisopliae* PR-2 (a gift from Dr. J. Gillespie at Kansas State University) with 50 μM S-2238; 5 ng human thrombin (Sigma) with 50 μM S-2238; 0.01 μg of human plasmin (Sigma) with 50 μM S-2238.

Detection of serpin-proteinase complexes by immunoblot analyses

As described in the legend to Fig. 5, Spn43Ac was incubated with cathepsin G, HNE, or thrombin for 10 min at room temperature. The reaction mixtures as well as the negative controls of Spn43Ac and proteinases only were treated with SDS sample buffer for 5 min at 100°C and subjected to SDS-polyacrylamide gel electrophoresis and immunoblot analysis as described above.

Determination of the cleavage site in Spn43Ac – To determine the scissile bond position in Spn43Ac, 1.2 µg purified inhibitor was reacted with 0.4 µg human neutrophil cathepsin G at room temperature for 30 min, and the reaction mixture was subjected to the MALDI-TOF mass spectrometry (Jiang et al., 2003). The molecular mass of the peak which was absent in the control spectra of Spn43Ac and cathepsin G only was compared with calculated values of carboxyl terminal peptides to deduce the cleavage site in Spn43Ac.

Results

To produce *Drosophila* Spn43Ac for functional analysis, we constructed a recombinant virus encoding the intact serpin. *Spodoptera frugiperda* Sf9 cells infected with the virus synthesized and secreted the protein. The recombinant serpin accumulated at a concentration of 8.64 µg/ml in the medium, accounting for 10% of the total protein (Table 1). While the serpin was detected by immunoblot analysis as a doublet at about 52 kDa, small molecular weight species also existed (Fig. 1, lane 3) and probably represented the degradation products of Spn43Ac. Addition of serine proteinase inhibitors significantly reduced the unwanted proteolysis during purification. Ion exchange chromatography on dextran sulfate-Sepharose enriched the recombinant serpin 13.9-fold and removed 86% of the other proteins (Table 1). We further purified Spn43Ac by hydrophobic interaction chromatography on a methyl HIC column, which efficiently separated the contaminantts from the prominent serpin proteins (Fig. 1, lane 3). Approximately 60% of the total serpin was lost during concentration and buffer exchange

on the centrifugal filter device (Table 1). The recovered Spn43Ac migrates as two close bands at about the same intensity on the SDS-polyacrylamide gel, and the overall purification was 9.4-fold.

MALDI-TOF mass spectrometry indicated that the Spn43A preparation contains two major proteins corresponding to the double bands on SDS-polyacrylamide gel (Fig. 2A). Their molecular masses are 51,792 Da and 52,838 Da, significantly larger than the calculated value of the mature protein (51,417 Da). The mass difference between the theoretical and observed values may result from posttranslational modifications such as glycosylation. Microheterogeneity of the proteins is apparent in the MH^{2+} region: two groups of peaks were detected with m/z ranging from 25,804 to 26,604 Da (Fig. 2B). Individual peaks in one group are closely similar to the corresponding ones in the other, suggesting that removal of a small peptide from the intact protein may account for the lower molecular mass species. From the most prominent MH^{2+} peaks in each group, we calculated the mass difference to be 1,042 Da, which is in good agreement with the molecular mass of the N-terminal first eight residues (1,043 Da). The insect cells may secret a trace amount of proteinase that cleaves after Arg^8.

The purified Spn43Ac eluted as a major peak from the HPLC gel filtration column at a retention time of 8.42 min (Fig. 3). This corresponds to a molecular mass of 49 kDa, indicating that the serpin is present as a monomer. Minor peaks at 7.45 and 9.78 min represent impurities in the preparation which account for less than 10% of the total protein.

To test if Spn43Ac is active as a serine proteinase inhibitor, we incubated the serpin at various concentrations with eight proteinases with different specificities and

then measured the remaining amidase activities (Fig. 4, panels A and B). Spn43Ac strongly inhibited chymotrypsin-like proteinases including human neutrophil cathepsin G and bovine pancreatic chymotrypsin. At a molar ratio of 1.2:1, the amidolytic activity of cathepsin G was completely blocked after incubation with the inhibitor at room temperature for 10 min. The serpin inhibited α-chymotrypsin less efficiently: the enzyme activity reduced to 50% when the molar ratio of Spn43Ac:chymotrypsin was 3:1.

The inhibition of elastase-like proteinases by Spn43Ac was comparable but less efficient (Fig. 4, panel C). Fifty percent inhibition of porcine pancreatic elastase occurred at about 10:1 (data not shown). A less robust enzyme, human neutrophil elastase (HNE), was completely inhibited at a molar ratio of 3.6:1. Similar to the Spn43Ac-cathepsin G pair, inhibition curve of the Spn43Ac-HNE reaction was a straight line, indicating a fast association between the proteinase and inhibitor.

Spn43Ac is a poor inhibitor of trypsin-like enzymes. It did not inhibit bovine trypsin, human plasmin, or *M. anisopliae* PR-2 even when the serpin was in large excess (>1000:1) (data not shown). However, we did observed weak inhibition of human thrombin by Spn43Ac (Fig. 4, panel D). At a molar ratio of about 120:1, 50% of the thrombin activity was blocked by the serpin. Complete inhibition of thrombin did not occur even when 360-fold excess of Spn43Ac was present.

To further characterize the serpin-proteinase interactions, we tested whether Spn43Ac could form complexes with the enzymes it inhibits. In the control of Spn43Ac only, the majority of the serpin migrated as two stalk bands at about 52 kDa, while a small amount of the protein appears as 47, 48 kDa bands (Fig. 5A). After incubation with cathepsin G, the major serpin bands were significantly reduced in intensity and several

new bands were detected at around 70 kDa. Absent in the controls of Spn43Ac and cathepsin G only, these high molecular weight bands represent the SDS-stable complexes formed between the inhibitor and proteinase. These complexes as well as the original serpin were susceptible to proteolysis by residual cathepsin G, which yielded products in the range of 30-45 kDa.

Serpin-proteinase complexes were also detected in the Spn43Ac-HNE reaction, although turnover of Spn43Ac was more severe (Fig. 5B). This is consistent with the observation that a higher molar ratio was required for complete inhibition of HNE by Spn43Ac. In contrast, the Spn43Ac-thrombin complex was hardly detected under the experimental conditions, because the partition ratio between substrate and inhibitor is over 360 for the Spn43Ac-thrombin pair (Fig. 5C). Unlike cathepsin G and HNE, human thrombin cleaved Spn43Ac specifically and generated two major bands at about 48 and 49 kDa.

Since *Drosophila* Spn43Ac is a good inhibitor of chymotrypsin- or elastase-like serine proteinases, we further characterized this reaction by determining their cleavage site in Spn43Ac using MALDI mass spectrometry (Fig. 6). In the control spectra of Spn43Ac and cathepsin G only, there was no mass peak in the range of 4,000 to 5,000 Da (data not shown). A new peak of 4,316 Da was detected in the proteinase-inhibitor mixture, representing the carboxyl-terminal peptide released from a proteolytic cleavage of Spn43Ac between Leu[438]-Ser[439]. The calculated molecular mass of the peptide (4,316 Da) perfectly matches the experimental value. This identification of Leu[438] as P1 residue is in good agreement with the observed inhibitory selectivity of Spn43Ac.

Discussion

Similar to their homologs in mammals, serpins from invertebrates are important regulators of defense responses (Jiang and Kanost, 2001). They may prevent excessive activation of the insect phenoloxidase system, horseshoe crab clotting pathway, and *Drosophila* spätzle-activating cascade. During embryonic development, cleaved spätzle binds to the Toll receptor and leads to the establishment of the dorsal-ventral axis (Hoffmann and Reichhart, 2002). The proteolytic activation of spätzle, mediated by different serine proteinases, stimulates defense responses against fungi and Gram-positive bacteria in *Drosophila* adults (De Gregorio et al., 2001, Irving et al., 2001). A loss-of-function mutation in the *Drosophila* Spn43Ac gene causes constitutive expression of the antifungal peptide drosomycin in the absence of elicitor (Levashina et al., 1999, Green et al., 1999). Spn43Ac is also named Necrotic, or Nec serpin, as the mutation also leads to spontaneous melanization, cellular necrosis, and death in early adulthood. These phenotypes are suppressed by a mutation in a serine proteinase gene named *persephone* (Ligoxygakis et al, 2002a). Truncated Spn43Ac from an *E. coli* expression system inhibited elastase, chymotrypsin, and thrombin (Robertson et al., 2003). It was suggested that Necrotic may act as the physiological inhibitor of Persephone. Melanization is also affected by another serpin, Spn27A, which may directly inhibit prophenoloxidase-activating enzyme (De Gregorio et al., 2002; Ligoxygakis et al, 2002b). Mutation in Spn27Ac did not seem to affect the Toll signaling pathway. Therefore, the proteinase cascade that leads to spätzle and proPO activation may be a split one, with Persephone located before the branching point.

Because the first 77 amino acid residues of mature Spn43Ac were absent from the

previously examined recombinant Spn43Ac (Robertson et al., 2002). We attempted to express the intact serpin in a baculovirus-insect cell expression system and study whether the amino terminal truncation altered selectivity or stability of the inhibitor. The purified protein inhibited cathepsin G and HNE most efficiently with stoichiometry of inhibition (SI) values of 1.2 and 3.6, respectively. Half of the chymotrypsin and elastase activities were inhibited when the molar ratio of serpin to enzyme reached 3 and 10. Spn43Ac inhibits thrombin, a trypsin-like serine proteinase, but only when the serpin was in large excess. Half of the amidase activity was blocked by Spn43Ac at a molar ratio of 120:1. While these data are in general agreement with the results from the earlier study using truncated Necrotic (Robertson et al., 2003), much higher levels of the intact serpin are required for complete inhibition of thrombin, α-chymotrypsin, and pancreatic elastase. The gel filtration experiment indicated that the high SI values are not caused by polymer formation of the serpin. Rather, the amino-terminal truncated part seems to selectively interfere with the association between Spn43Ac and some proteinases: the stoichiometry of the Spn43Ac-cathepsin G reaction was close to 1.2 whereas that of the Spn43Ac-thrombin reaction increased from 4.4 to more than 1000.

The first 80 residues of mature Spn43Ac are rich in Gln (22), Pro (11), and Leu (10). The polyglutamine and leucine stretches are located near the amino terminus whereas proline residues are mostly at the carboxyl end of this region. The extension, not found in most other serpins, modifies the stability and inhibitory selectivity of Spn43Ac. MALDI-TOF mass spectrometry identified nine peptides in the Spn43Ac-cathepsin G reaction that matches the calculated peptide mass fingerprints. While the most prominent one came from cleavage of the scissile bond, seven others resulted from the amino-

terminal extension (data not shown). These data indicates that this part of the molecule is exposed on the surface and easily accessible by proteolytic enzymes. Further studies are needed to elucidate the structure and function of the amino terminal extension. One interesting possibility is that this region may serve as a sensor for detecting the activation status of the proteinase cascade: when it is cleaved from the intact molecule, the serpin becomes a better inhibitor of its target proteinase (Robertson et al, 2002).

Microarray experiments indicated that the Toll and Imd pathways are the major regulators of the humoral immune responses in adult *Drosophila* (De Gregorio et al., 2002). The Toll pathway is stimulated by infection by fungi and Gram-positive bacteria. Since Spn43Ac affects activation of both spätzle and prophenoloxidase, its target enzyme must be located upstream of the branching point, before, at or after the position of Persephone. Our data suggest Persephone or other trypsin-like serine proteinases might not be regulated by Necrotic. Because all known invertebrate serine proteinase cascades employ a multi-domain enzyme at the top and trypsin-like, clip-domain enzymes at or near the bottom (Jiang and Kanost, 2001), we suggest that Spn43Ac may target a large serine proteinase with chymotrypsin- or elastase-like specificity, such as horseshoe crab Factor C. Perhaps, transcription of the target enzyme is induced upon microbial infection to replenish what was used up during initial response. Based on these premises, we predict that SP55 and/or SP60 could be directly regulated by Spn43Ac. The corresponding genes for these elastase-like enzymes are located in the same cluster probably resulting from recent gene duplication (Ross et al., 2003). When the microarray data were examined to see if the transcription of SP55 and SP60 is up-regulated by the Toll pathway but not by the imd pathway, we found this is true indeed.

Table 1 Purification of Recombinant Spn43Ac from the baculovirus infected insect cells

Puification step	Volume (ml)	Protein[a]		Spn43Ac[b]		Yield (%)	Purification fold
		mg/ml	mg	µg/ml	µg		
medium	195	0.081	15.8	8.64	1685	100	1
dextran sulfate	14	0.162	2.68	113.6	1590	94.3	5.56
HIC, Centricon	1.95	0.219	0.427	219	427.1	25	9.38

[a] Protein concentrations were determined by a modified Bradford method using a commercial kit (Pierce). Bovine serum albumin was used as the standard.
[b] Concentration of Spn43Ac was determined by ELISA. Purified Spn43Ac of known concentration was used as the standard.

References

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Govans, J.D., Amantides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., and Galle, R.F. (2000) The genome sequence of *Drosophila melanogater*. *Science* 287, 2187-2195.

DeGregorio, E., Spellman, P.T., Rubin, G.M., and Lemaitre, B. (2001) Genome-wide analysis of the Drosophila immune response by using oligonucleotide microarrays. *Proc Natl Acad Sci USA* 98, 12590-12595.

DeGregorio, E., Spellman, P.T., Tzou, P., Rubin, G.M., and Lemaitre, B. (2002) The Toll and Imd pathways are the major regulators of the immune response in *Drosophila*. *EMBO J* 21, 2568-2579.

Green, C., Levashina, E., McKimmie, C., Dafforn, T., Reichhart, J.M., and Gubb, D. (2000) The necrotic gene in *Drosophila* corresponds to one of a cluster of three serpin transcripts mapping at 43A1.2. *Genetics* 56, 1117-1127.

Hoffmann, J.A. and Reichhart, J.M. (2002) *Drosophila* innate immunity: an evolutionary

perspective. *Nat Immunol* 3, 121-126.

Irving, J.A., Pike, R.N., Lesk, A.M., and Whisstock, J.C. (2000) Phylogeny of the serpin superfamily: implications of patterns of amino acid conservation for structure and function. *Genome Res* 10, 1845-1864.

Irving, P., Troxler, L., Heuer, T.S., Belvin, M., Kopczynski, C., Reichhart, J.M., Hoffmann, J.A., and Hetru, C. (2001) A genome-wide analysis of immune responses in *Drosophila*. *Proc Natl Acad Sci USA* 98, 15119-15124.

Jiang H. and Kanost M.R. (2001) Proteolytic activation of prophenoloxidase in an insect *Manduca sexta*. *Adv Exp Med Biol* 484, 313-7.

Levashina, E.A., Langley, E., Green, C., Gubb, D., Ashburner, M., Hoffmann, J.A., and Reichhart, J.M. (1999) Constitutive activation of toll-mediated antifungal defense in serpin-deficient *Drosophila*. *Science* 285, 1917-1919.

Ligoxygakis, P., Pelte, N., Hoffmann, J.A., and Reichhart, J.M. (2002) Activation of *Drosophila* Toll during fungal infection by a blood serine protease. *Science* 297, 114-116.

Ligoxygakis, P., Pelte, N., Ji, C., Leclerc, V., Duvic, B., Belvin, M., Jiang, H., Hoffmann, J.A., and Reichhart, J.M. (2002) A serpin mutant links Toll activation to melanization in the host defence of *Drosophila*. *EMBO J* 21, 6330-6337.

Potempa, J., Korzus, E., and Travis, J. (1994) The serpin superfamily of proteinase inhibitors: structure, function, and regulation. *J Biol Chem* 269, 15957-15960.

Robertson, A.S., Belorgey, D., Lilley, K.S., Lomas, D.A., Gubb, D., Dafforn, T.R. (2002) Characterization of the necrotic protein that regulates the toll-mediated immune response in *Drosophila*. *J Biol Chem* 278, 6175-6180.

Ross, J., Jiang, H., Kanost, M.R., and Wang, Y. (2003) Serine proteases and their homologs in the *Drosophila melanogaster* genome: an initial analysis of sequence conservation and phylogenetic relationship. *Gene* 304, 117-131.

Silverman, G.A., Bird, P.I., Carrell, R.W., Church, F.C., Coughlin, P.B., Gettins, P.G., Irving, J.A., Lomas, D.A., Luke, C.J., Moyer, R.W., Pemberton, P.A., Remold-O'Donnell, E., Salvesen, G.S., Travis, J., and Whisstock, J.C. (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.

Wang, Y., Jiang, H., and Kanost, M.R. (2001) Expression and Purification of *Manduca sexta* Prophenoloxidase-Activating Proteinase Precursor (proPAP) from Baculovirus-Infected Insect Cells. *Protein Expression Purification* 23, 328-337.

Ye, S., Cech, A.L., Belmares, R., Bergstrom, R.C., Tong, Y., Corey, D.R., Kanost, M.R., and Goldsmith, E.J. (2001) The structure of a Michaelis serpin-protease complex. *Nat Struct Biol* 8, 979-983.
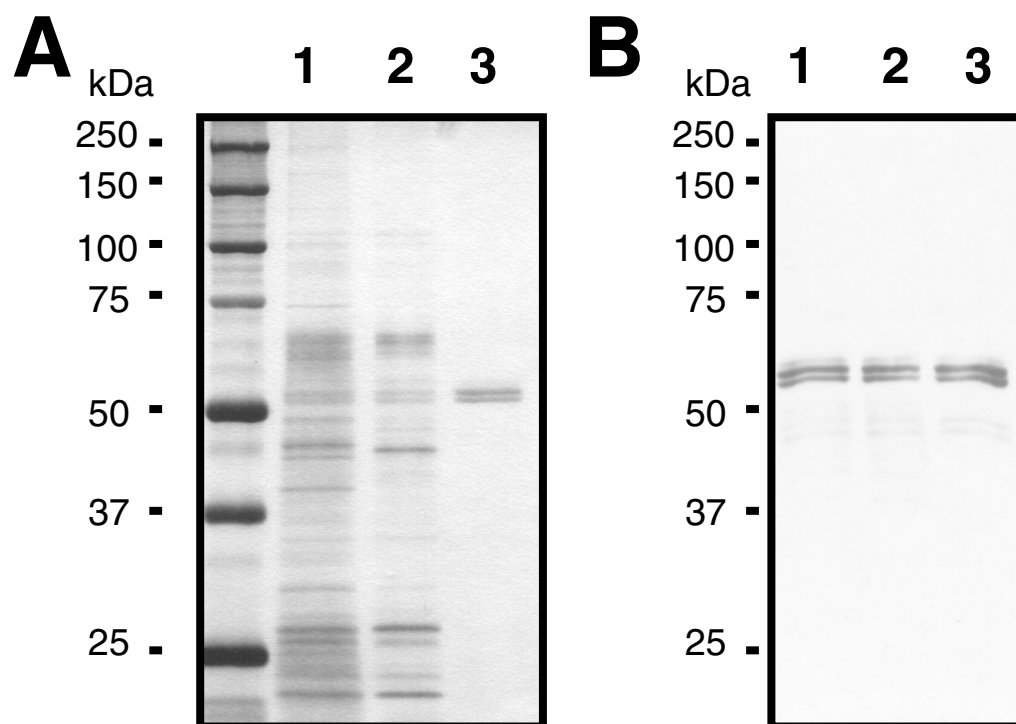
Fig. 1 Purification of recombinant Spn43Ac from the baculovirus-insect cell expression system. The Spn43Ac fractions, equivalent to 10 μl culture medium, were treated with SDS sample buffer containing β-mercaptoethanol at 95°C for 5 min, separated on 10% SDS-PAGE gel by electrophoresis, and detected silver staining (A) and immunoblot analysis (B). Lane 1, culture supernatant of Sf9 cells infected by the baculovirus, lane 2, protein eluted from dextran sulfate-Sepharose column, lane 3, protein eluted from the methyl-HIC column and concentrated by Centricon-30. Molecular masses of the standard proteins are shown on the left side.
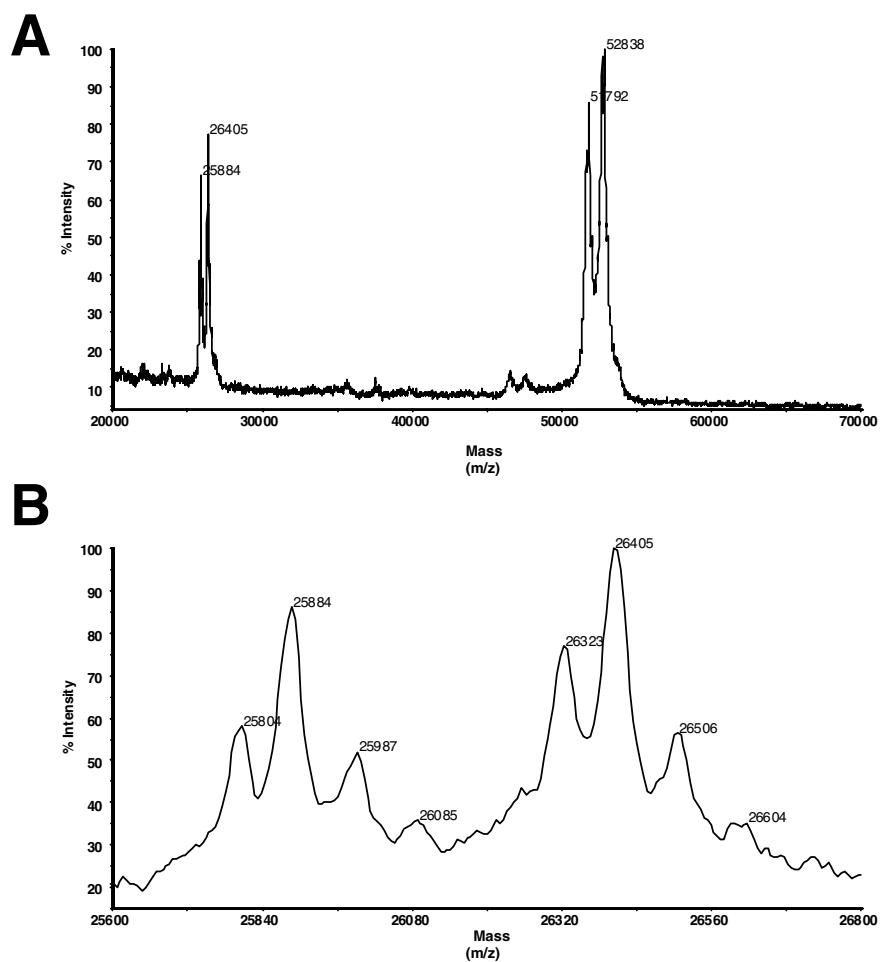
Fig. 2 MALDI-TOF mass spectrometry of Spn43Ac. A) A representative spectrum for Spn43Ac is shown. The spectrum was calibrated with an external standard of rabbit muscle aldolase (MW: 39,212). The mass values are indicated above the MH[+] and MH[+2] peaks. B) MH[2+] region of the spectrum.
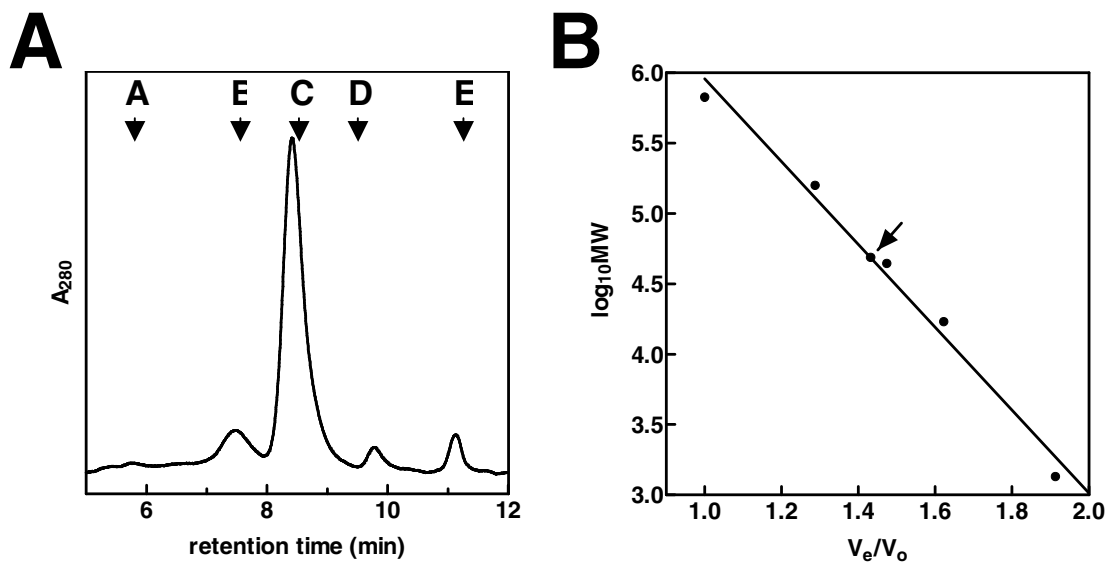
Fig. 3 Association status of Spn43Ac determined by gel filtration chromatography.
A) Chromatogram of purified Spn43Ac on Bio-Silect SEC 250. A major peak at 8.42 min indicates that Spn43Ac exists as a monomer. Arrows mark the positions of molecular weight standards: A, thyroglobulin; B, bovine γ-globulin; C: chicken ovalbumin; D, equine myoglobin; E, vitamin B_{12}. B) Standard curve. Known molecular weights of the protein standards (in logarithm) were plotted against their Ve/Vo's. From the linear regression curve, molecular weight of Spn43Ac (marked by an arrow) was calculated from its Ve/Vo.
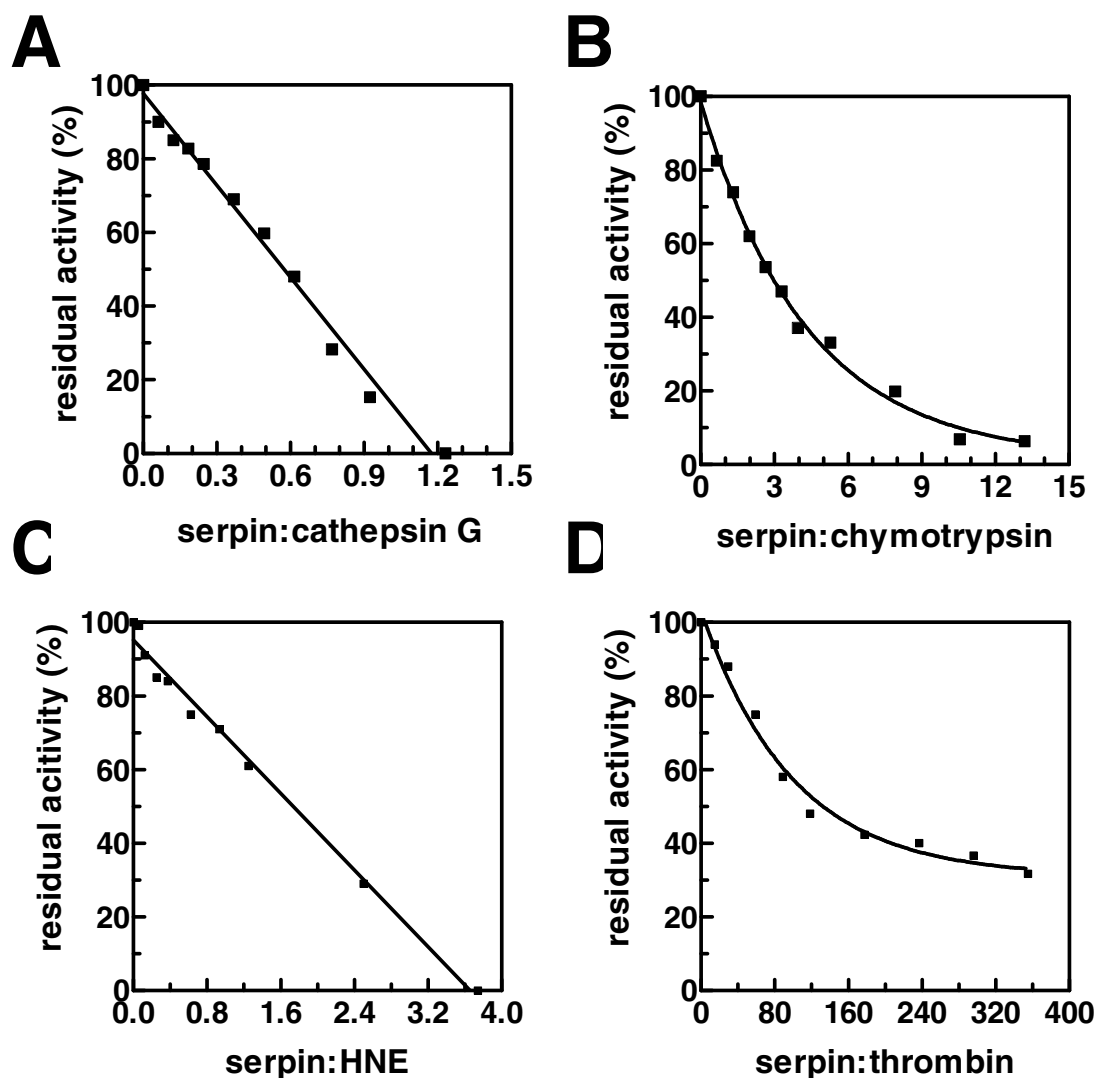
Fig. 4 Concentration-dependent inhibition of serine proteinases by Spn43Ac. As described in Experimental Procedures, aliquots of proteinases were mixed with the recombinant serpins at different molar ratio for 10 min at room temperature. Residual proteinase activities were determined, and the relative activities (%) were plotted against the corresponding molar ratios of Spn43Ac and the enzyme. A) Spn43Ac and human neutrophil cathepsin G; B) Spn43Ac and bovine α-chymotrypsin; C) Spn43Ac and human neutrophil elastase (HNE); D) Spn43Ac and human thrombin.
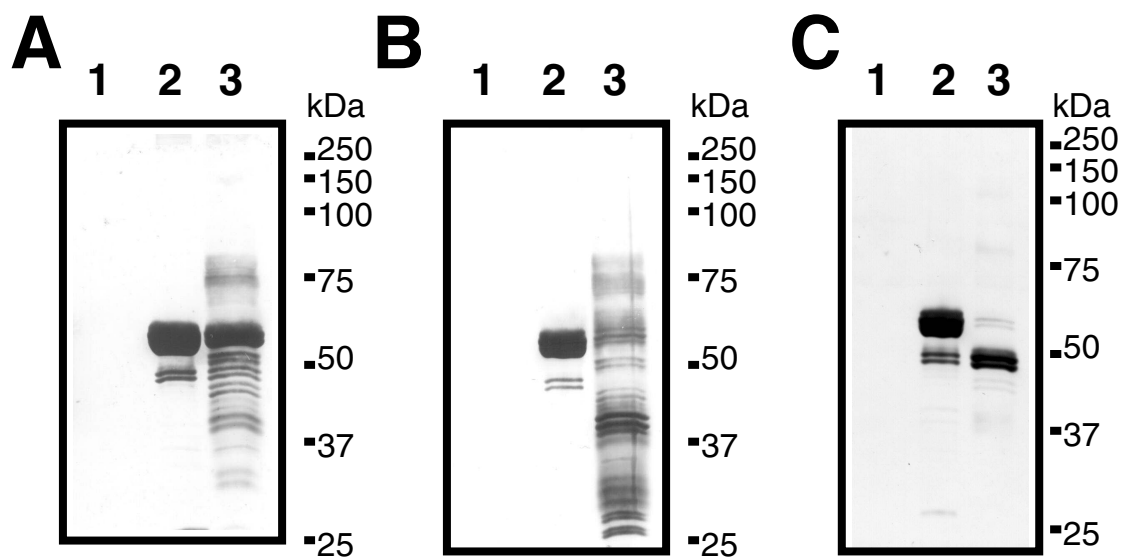
Fig. 5 Detection of SDS-stable serpin-proteinase complexes by immunoblot analysis.
A proteinase (lane 1), Spn43Ac (lane 2), and a mixture of both (lane 3) were incubated at room temperature for 10 min. The reaction mixtures were boiled in the presence of SDS-sample buffer and subjected to SDS-PAGE on a 10% polyacrylamide gel. Immunoblot was performed using 1:2000 diluted Spn43Ac antiserum as the first antibody.   A) Spn43Ac (1.2 µg) and human neutrophil cathepsin G (400 ng); B) Spn43Ac (0.6 µg) and human nutrophil elastase (100 ng); C) Spn43Ac (0.6 µg) and human thrombin (62.5 ng).
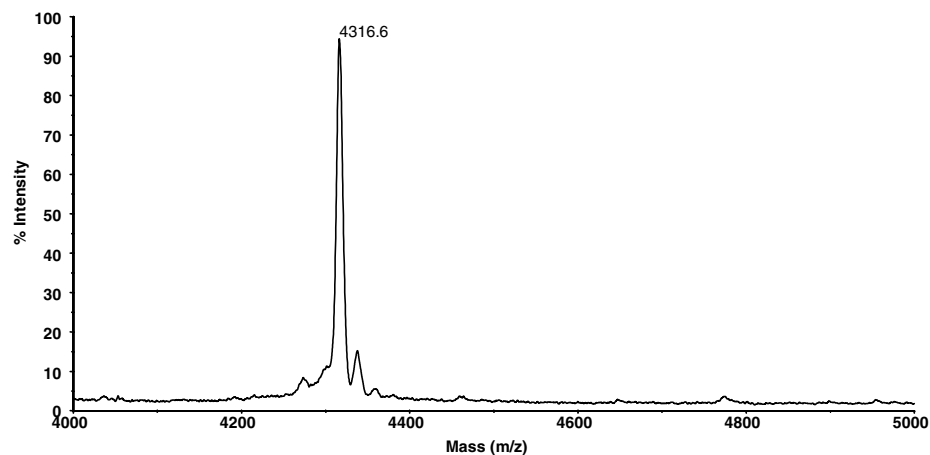
Fig 6. Determination of the cleavage site in Spn43Ac by MALDI-TOF mass spectrometry. As described in Experimental Procedures, the reaction mixture of Spn43Ac and cathepsin G was directly analyzed by MALDI-TOF mass spectrometry. A representative strong single-accumulation spectrum is presented with the mass value marked on top of the MH[+] peak. The spectrum was subjected to noise removal and calibrated with an external standard of insulin. The control spectra of Spn43Ac and cathepsin G alone do not contain any mass peak in the range of 4,000 – 5,000 Da (data not shown).

Chapter VII. Summary

Analogous to blood coagulation in mammals, serine proteinases play important roles in insect innate immunity. One of the insect defense responses, proPO activation, is mediated by a cascade of serine proteinases. In the final step of the pathway, proPO is cleaved by its activating proteinase PAP to form PO. PO participates in the formation of quinones that are precursors of melanin. Melanin entraps and immobilizes pathogens and parasites. The cytotoxic substances generated during melanization may kill microbes. While the proteinase cascade is largely unknown, three PAPs have been discovered in *M. sexta*. They all belong to clip-domain serine proteinase family. In the hemolymph, serpins regulate serine proteinases to prevent over-stimulation of the defense reaction.

Chapters III and IV report the elucidatation of PAP-1 and PAP-3 gene structure and expression patterns. PAP-1 gene contains seven exons and six introns whereas PAP-3 gene consists of eight exons and seven introns. With different exon-intron organizations, PAP-1 and PAP-3 genes may have arisen from two distinct evolutionary processes. Various putative regulatory elements are present in the 5' flanking region. These data, as well as tissue-specific and developmental expression profiles, lead to the conclusion that PAP-1 expression is under the dual control of immune and hormonal signals whereas PAP-3 transcription is regulated in a different manner.

ProPO activation is controlled by serpins. Serpin-6 inhibits PAP-3, but not PAP-2. As discussed in Chapter V, I attempted to understand how serpin-6 regulates proPO activation. I cloned and characterized *M. sexta* serpin-6 and found its amino acid sequence most similar to *D. melanogaster* serpin-5 and *A. gambiae* serpin-9. The recombinant serpin-6 specifically inhibited PAP-3 by forming an SDS-stable acyl-

enzyme complex. I determined the cleavage site of serpin-6, which is identical to our prediction. The association rate constant of serpin-6 and PAP-3 is 2.6 x 10^4 $M^{-1}s^{-1}$, indicating that serpin-6 may contribute to the inhibition of PAP-3 in the hemolymph. Hemolymph proteins eluted off serpin-6 immunoaffinity column include serpin-6-PAP-3 and serpin-6-HP8 complexes, as well as immulectin-2, serine proteinase homologs, proPO, PO, and attacin-2. Therefore, serpin-6 regulatesproPO activation by inhibiting PAP-3 and HP8, components of a macromolecular complex formed around invading pathogens.

Drosophila Spn43Ac down-regulates an extracellular serine proteinase cascade that activates the Toll signaling pathway. Chapter VI reports the expression, purification, and characterization of the intact serpin. The purified protein, present in a soluble, monomeric form, inhibited Cathepsin G and human neutrophil elastase at SI's of 1.2 and 4.0, respectively. MALDI mass spectrometry indicated that the cleavage site of Necrotic is between Leu[438] and Ser[439]. These results indicate that Spn43Ac may regulate chymotrypsin- and elastase-like enzymes. Spn43Ac inefficiently inhibited thrombin, suggesting that Spn43Ac's physiological target may not be Persephone, a thrombin-like proteinase.

VITA

Zhen Zou

Candidate for the Degree of

Doctor of Philosophy

Dissertation: REGULATION OF PROPHENOLOXIDASE-ACTIVATING PROTEINASES IN *MANDUCA SEXTA*

Major Field: Entomology

Biographical:

Personal Data: Born in Shanghai, People's Republic of China, May 24, 1975, son of Renyuan Zou and Yuyi Wang.

Education: Graduated from Hongkou High School, Shanghai, P. R. China, in July 1993. Received Bachelor of Engineering Degree in Biochemical Engineering from East China University of Science and Technology, Shanghai, P. R. China, in July, 1997. Completed Requirements for the Doctor of Philosophy Degree at Oklahoma State University in July, 2005.

Professional Experience: Research Assistant, Plant Molecular Biology Laboratory, Shanghai Institute of Biochemistry, Chinese Academy of Sciences, P. R. China from August 1997 to Jan., 2001. Graduate Research Assistant, Insect Molecular Biology Laboratory, Department of Entomology and Plant Pathology, Oklahoma State University, Stillwater, Oklahoma, USA, from January 2001 to July 2005.