IMMUNE-RELATED PROTEIN COMPLEXES AND SERPIN-1 ISOFORMS IN *MANDUCA SEXTA* PLASMA

by

EMILY J. RAGAN

A.B., University of California, Berkeley, 2002

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Biochemistry Graduate Group

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2008

Abstract

Manduca sexta is a large insect species well-suited for biochemical analysis of proteins in the hemolymph (blood) that respond to infection. Insects lack adaptive immunity and rely entirely on innate immunity to prevent and manage infection. Immune response proteins include proteins that bind pathogens and activate serine proteases, which function in proteolytic cascades that trigger effector responses, such as antimicrobial peptide production and prophenoloxidase activation. Phenoloxidase catalyzes melanin synthesis, which leads to microbial killing.

I used MALDI-TOF/TOF mass spectrometry and immunoblotting to identify *M. sexta* proteins present in putative immune complexes. From analyses of high molecular weight gel filtration fractions of plasma activated by microbial polysaccharides, I detected hemocytin, prophenoloxidase, and cleaved serine protease homologs, suggesting prophenoloxidase and serine protease homologs form large complexes in plasma. I used *in vitro* bacterial binding assays to identify hemolymph proteins that bind either directly or indirectly to the surface of bacteria or curdlan. Prophenoloxidase, annexin IX, and hemocyte aggregation inhibitor protein were found bound to all the samples tested, indicating they play a role in the early stage of immune response.

Serpins regulate specific active proteases by covalently binding and forming serpin-protease complexes. Serpin-1, an abundant plasma protein, has an alternatively spliced ninth exon encoding 12 serpin-1 isoforms that differ in inhibitory selectivity. RT-PCR showed that all 12 isoforms are expressed in hemocytes, fat body, and midgut. Comparisons of naïve and immune-challenged hemocytes and fat body indicated the immune-related upregulation of serpin-1A but not the other isoforms. Using immunoaffinity chromatography I isolated two serpin-1-protease complexes from plasma after activation with bacterial lipopolysaccharide. MALDI-TOF/TOF analysis of these serpin-1-protease complexes identified the digestive enzyme chymotrypsin as a specific target of serpin-1K. Nine out of the twelve serpin-1 isoforms were identified from control plasma at the protein level using 2D-PAGE. Serpin-1 protease complexes were

identified by 2D-PAGE analysis: serpin-1A, E and J were found to be complexed with hemolymph proteinase-8 and an unidentified isoform of serpin-1 was complexed with hemolymph proteinase-1. Discovering the serpin-1 isoforms that inhibit specific proteases enhances our understanding of the regulation of proteolytic cascades in *M. sexta*.

IMMUNE-RELATED PROTEIN COMPLEXES AND SERPIN-1 ISOFORMS IN *MANDUCA SEXTA* PLASMA

by

EMILY J. RAGAN

A.B., University of California, Berkeley, 2002

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Biochemistry Graduate Group

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2008

Approved by;

Major Professor
Michael R. Kanost
Department of Biochemistry

Copyright

Abstract

Manduca sexta is a large insect species well-suited for biochemical analysis of proteins in the hemolymph (blood) that respond to infection. Insects lack adaptive immunity and rely entirely on innate immunity to prevent and manage infection. Immune response proteins include proteins that bind pathogens and activate serine proteases, which function in proteolytic cascades that trigger effector responses, such as antimicrobial peptide production and prophenoloxidase activation. Phenoloxidase catalyzes melanin synthesis, which leads to microbial killing.

I used MALDI-TOF/TOF mass spectrometry and immunoblotting to identify *M. sexta* proteins present in putative immune complexes. From analyses of high molecular weight gel filtration fractions of plasma activated by microbial polysaccharides, I detected hemocytin, prophenoloxidase, and cleaved serine protease homologs, suggesting prophenoloxidase and serine protease homologs form large complexes in plasma. I used *in vitro* bacterial binding assays to identify hemolymph proteins that bind either directly or indirectly to the surface of bacteria or curdlan. Prophenoloxidase, annexin IX, and hemocyte aggregation inhibitor protein were found bound to all the samples tested, indicating they play a role in the early stage of immune response.

Serpins regulate specific active proteases by covalently binding and forming serpin-protease complexes. Serpin-1, an abundant plasma protein, has an alternatively spliced ninth exon encoding 12 serpin-1 isoforms that differ in inhibitory selectivity. RT-PCR showed that all 12 isoforms are expressed in hemocytes, fat body, and midgut. Comparisons of naïve and immune-challenged hemocytes and fat body indicated the immune-related upregulation of serpin-1A but not the other isoforms. Using immunoaffinity chromatography I isolated two serpin-1-protease complexes from plasma after activation with bacterial lipopolysaccharide. MALDI-TOF/TOF analysis of these serpin-1-protease complexes identified the digestive enzyme chymotrypsin as a specific target of serpin-1K. Nine out of the twelve serpin-1 isoforms were identified from control plasma at the protein level using 2D-PAGE. Serpin-1 protease complexes were

identified by 2D-PAGE analysis: serpin-1A, E and J were found to be complexed with hemolymph proteinase-8 and an unidentified isoform of serpin-1 was complexed with hemolymph proteinase-1. Discovering the serpin-1 isoforms that inhibit specific proteases enhances our understanding of the regulation of proteolytic cascades in *M. sexta*.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

1D	one dimensional
2D	two dimensional
A	absorbance
Arg	arginine
Asp	aspartic acid
bp	base pair
CI%	confidence interval
Glu	glutamic acid
h	hours
HP	hemolymph proteinase
IgG	immunoglobulin G
kDa	kilodaltons
LPS	lipopolysaccharide
LTA	lipotechoic acid
Lys	lysine
MALDI	matrix assisted laser desorption ionization
min	minutes
MS	mass spectrometry
PAGE	polyacrylamide gel electrophoresis
PGN	peptidoglycan
pI	isoelectric point
PO	phenoloxidase
Pro	proline
proPO	prophenoloxidase
RT-PCR	reverse transcriptase polymerize chain reaction
s	seconds
SDS	sodium dodecyl sulfate

Acknowledgements

I thank Dr. Mike Kanost, my major professor, for the opportunity to study and research in his lab and for his calm and steady support.

I would like to recognize all my committee members, Dr. Rollie Clem, Dr. R. Krishnamoorthi, Dr. S. Muthukrishnan, Dr. James Nechols, and Dr. Anna Zolkiewska, for their support, suggestions, and interest in my research.

I would like to thank the many current and recent members of the Kanost lab for their help and camaraderie. Dr. Maureen Gorman, who kept the lab functional and helped keep me sane, Dr. Neal Dittmer, who was a patient teacher and who kept me laughing, Dr. Chunju An, Jayne Christen, Huaien Dai, Stewart Gardner, Sandi Yungeberg, Lucinda Sullivan, Dr. Jeff Fabrick, Dr. Jun Ishibashi, Dr. Dave Levin, Dr. Rich Suderman, Dr. Chansak Suwanchaichinda, Dr. Youren Tong, Dr. Shufei Zhuang, Lisha Breuer, Mary Everhart, Ana Fraire, Yongli Gu, Rose Ochieng, Xuyong Wang, and Celeste Yang.

I learned about biochemistry from every faculty member in the department and thank them for providing a stimulating environment. I thank the many graduate students during my time here for community and the biochemistry staff for their assistance, especially Crystal Sapp, Melinda Bainter, and Dave Manning. I would like to thank Dr. Kristin Michel, from the Division of Biology, for her help translating portions of PhD dissertations from German. I also thank Dr. Haobo Jiang, from Oklahoma State University, for his interest in my work and for access to the *M. sexta* ESTs prior to their publication. I thank Rebekah Woolsey from the Nevada Proteomics Center for performing mass spectrometry.

I thank my friends for their encouragement. I especially would like to mention long time friends Forrest Aubel and Naomi Beeman, as well as Robert Chapman, Dr. Susan Allen and the noontime yoga group, the members of the Monday night book group, members of UUFM, Dr. Jay Ham and Dr. Brice Hobrock for helpful discussions about writing a dissertation, and the late Dr. Dave Hogenkamp, who taught me so much about persistence, determination, and dedication.

Finally, I thank my family for their love and confidence.

Dedication

I dedicate this dissertation to my family: to my husband Eric, whose love, acceptance, and support made it possible for me to finish this work, to my parents, whose confidence and assistance allowed me to undertake such a project, to my daughter Olivia, who has taught me about focus, intensity, and balance, and to my little one on the way, whose presence provided wonderful motivation.

CHAPTER 1 - LITERATURE REVIEW

Introduction

Insects are a phenomenally diverse and successful class of animals. Approximately one million different named insect species are currently living throughout terrestrial earth and a few million more are likely undiscovered (Thomas, 2005). Insects belong to the phylum arthropoda, along with crustaceans, spiders, ticks and other organisms with jointed legs and exoskeletons; there are an estimated 5 million arthropod species in the world (Novotny et al, 2002). While insects are often considered pests, since some are vectors of disease and others consume human crops and food, insects are also an essential part of the ecosystem. From a human centered point of view, we use insects for pollination, honey, silk, and as natural agents to biologically control pests. Insects are also valuable for scientific research. Certain model organisms, including *Drosophila melanogaster*, have helped in the elucidation of many genetic and immune processes that have applications for humans. For example, the Toll pathway was discovered first in *D. melanogaster* embryonic development and then in immunity and these discoveries led to the elucidation of Toll-like receptors in mammals (Sansonetti, 2006).

Insects lack an adaptive immune response, which includes antibody generation, but have a powerful innate immune system that allows them to thrive despite the ubiquitous presence of deleterious microorganisms and other pathogens (Dunn, 1990, Gillespie et al, 1997). Thus, insects provide a simpler system than mammals to study innate immunity (Ferrandon et al, 2007). Researchers hope to apply insights from studies of insect immunity to understand the evolution of innate immunity, and to solving the problem of insect-vectored human diseases, such as the transmission of malaria by *Anopheles gambiae,* an African mosquito (Christophides, 2005). Additionally, the recent mysterious colony collapse of North American honey bees reminds us that the health of beneficial insects can have broad economic implications (Stokstad, 2007). While the cause of colony collapse disorder remains unknown, the Israeli acute paralysis virus was found to be a marker for colony collapse was (Cox-Foster et al, 2007).

Manduca sexta is a type of moth (order Lepidoptera) that is a good model organism for biochemical studies due to its relatively large size. For example, about 1 ml of hemolymph (blood) can be obtained from a *M. sexta* larva in its last instar, which is comparable to the amount of blood available from a mouse (Kanost et al, 2004). This is in stark contrast to the 1/10 μl of hemolymph available from an *A. gambiae* adult. Advances in understanding mechanisms of immunity in *M. sexta* contribute to our understanding of immunity in other insects such as disease vectors, honey bees, and silk moths, while discoveries in other insects can enhance our studies of *M. sexta* (Iwanaga & Lee, 2005, Jiravanichpaisal et al, 2006).

At least some processes involved in immune responses, such as phagocytosis, have evolved from more fundamental processes in unicellular and multicellular organisms. These include mechanisms for obtaining nutrition, eliminating defective cells, and cell differentiation during embryonic development (Desjardins et al, 2005, Dunn, 1990). Another important perspective for understanding immunity in insects like *M. sexta* is their life history. Like other holometabolous insects, which undergo complete metamorphosis, *M. sexta* has dramatically different life stages. The larval stage is characterized by active foliage feeding and rapid growth. Growth is limited by the cuticle, necessitating the generation of new cuticle and molting to remove the old one. After cessation of feeding, *M. sexta* larvae wander, find a place in the soil to pupate and undergo the massive transformation to adulthood called metamorphosis. After about three weeks (or longer if in diapause) a mature adult emerges from the pupal case. Adults feed on nectar, can fly long distances, and reproduce. The larvae, pupae, and adults live in different ecological niches; they have different eating habits and are exposed to different potential pathogens. Because of the different environments they inhabit and their overall short lifespan (less than a year), many holometabolous insects can be called ecologically ephemeral (Dunn, 1990). This is likely reflected in their immune systems, which must act quickly but have less need for a long term memory.

Developmental differences can also translate to different immune responses in different stages of life; indeed, different immune responsive proteins are expressed at different developmental stages in *D. melanogaster* (Sowell et al, 2007). Lepidopteran larvae have developmental resistance to baculovirus infection, which is probably related

to midgut sloughing at molts and during metamorphosis (Terenius, 2008). During metamorphosis the midgut undergoes substantial changes which could release commensal bacteria. High levels of antibacterial factors were found in *M. sexta* midgut during metamorphosis (Dunn et al, 1994). Antibacterial factors are also present in *M. sexta* eggs, which can synthesize antimicrobial peptides and deposit melanin in response to bacteria (Gorman et al, 2004).

Hemolymph

Insect physiology provides the setting for the immune response. Barriers to infection include insect cuticle and protective lining in the midgut. An infection that penetrates these outer barriers can reach the interior of the insect. Insects lack veins and arteries and instead have an open circulatory system, where hemolymph, which is analogous to mammalian blood, is pumped through the hemocoel (body cavity) (Chapman, 1998). Hemolymph contains hemocytes (blood cells) and many different chemical and protein components (Kanost et al, 1990, Wyatt, 1961, Wyatt & Pan, 1978). It functions to transport ions, nutrients, and hormones, in addition to being involved in responding to invading pathogens such as bacteria, fungi, and other parasites. Many proteins are secreted into the hemolymph by the fat body, which is analogous to both the mammalian liver and adipose tissue, and other proteins are secreted from hemocytes, epidermis, and midgut epithelium (Chapman, 1998, Dunn, 1990). Lipids are transported in spherical lipophorin particles which contain integral protein components of apolipophoin I and apolipophorin-II and, under certain conditions, contain exchangeable apolipophorin-III (Ryan & van der Horst, 2000). Larval storage proteins arylphorin α and β are very abundant proteins in late larval *M. sexta* hemolymph and are used as a source of amino acids for the synthesis of adult proteins during metamorphosis (Willott et al, 1989). Hemolymph components also help restore homeostasis after injury or infection. When wounding occurs, hemolymph is able to clot; when pathogens invade, the insect factors in the hemolymph are able to mount an immune response.

Immune function

Traditionally, insect immune responses have been characterized as either cellular immune responses, which involve the hemocytes, or humoral immune responses which involve soluble factors such as proteins. Examples of cellular immune response include phagocytosis, nodulation, and encapsulation. Humoral immune responses include melanization, production of reactive oxygen species, and synthesis of antimicrobial peptides. *In vivo*, of course, both types of response can happen simultaneously and even work together, as shown by plasma factors that are important for some cellular immune response (Lavine & Strand, 2001) and "humoral" factors such as prophenoloxidase and antimicrobial peptides may be released from hemocytes (Ferrandon et al, 2007, Jiang et al, 1997, Kanost & Gorman, 2008). Additionally, the many components of the immune system work together and may often times be redundant, as shown by RNAi studies in which a single immune pathway was silenced with no decrease in insect survival (Lesch et al, 2007, Schnitger et al, 2007). Another way to think of an immune response is to focus on the different stages, including the initial step of recognizing a foreign surface, a "signal transduction" phase when information about the recognition is transferred, and finally "effector" responses which are directed toward containing and killing the invader.

Hemocyte immune responses: nodulation and phagocytosis

Circulating within *M. sexta* hemolymph are four types of mature hemocytes: plasmatocytes, granulocytes, oenocytoids, and spherulocytes (Strand, 2008). The most abundant hemocyte type, granulocytes, contains cytoplasmic granules, adhere to foreign surfaces, spread symmetrically, and perform phagocytosis. Plasmatocytes spread asymmetrically and participate in capsule formation around large invaders (Strand, 2008). Spherule cells, which do not spread, aren't as well studied, but were found to secrete an abundant cuticle protein into the hemolymph of the caterpillar *Calpodes ethlius* (Sass et al, 1994). Oenocytoids are fragile cells that contain prophenoloxidase, an important enzyme for immune-related melanization (Jiang et al, 1997, Kanost & Gorman, 2008). Finally, there is a report of immune challenge causing the appearance of a larger, plasmatocyte-like type of hemocyte called hyper-spreading hemocytes in *M. sexta* (Dean et al, 2004).

When *E. coli* is injected into the *M. sexta* hemocoel, the majority of the bacteria are cleared within an hour (Dunn & Drake, 1983), while the same bacteria are able to grow in plasma in the absence of hemocytes. The hemocyte clumping that accompanies bacterial aggregation and clearance is called nodulation (Horohov & Dunn, 1983). This nodulation or hemocyte microaggregation can be mediated by eicosanoids (Miller & Stanley, 2001). Large nodules or other larger targets, such as parasites or certain types of chromatography beads, can be surrounded by layers of plasmatocytes and granulocytes in a process called encapsulation (Jiravanichpaisal et al, 2006, Strand, 2008, Wiegand et al, 2000). After the rapid bacterial clearance seen within an hour in *M. sexta,* another stage of response was observed at 2-8 h after injection when remaining bacteria were removed by phagocytosis but no nodules were observed (Horohov & Dunn, 1983). Phagocytosis of smaller targets, such as bacteria, yeast, and apoptotic cells, is performed by granulocytes in *M. sexta* and by similar cells named plasmatocytes in *D. melanogaster* (Strand, 2008). Phagocytic cells must bind the target, either by directly binding to it or binding another protein, an opsonin, which binds the target surface and flags it for phagocytosis (Stuart & Ezekowitz, 2005). After binding occurs, the actin cytoskeleton is activated and the target is engulfed. Once inside the cell, the early phagosome develops into a phagolysosome where reactive oxygen species and other killing mechanisms are employed (Stuart & Ezekowitz, 2005, Stuart et al, 2007).

While work from the early 1980s in *M. sexta* shows the fast response of nodulation for bacterial clearance (Horohov & Dunn, 1983), phagocytosis is also happening at these early times. In *Aedes aegypti* both phagocytosis and melanization were observed after five minutes (Hillyer et al, 2003). Similarly, in *A. gambiae*, latex beads or *E. coli* underwent phagocytosis within the first five minutes after injection (Hillyer et al, 2007).

Pattern recognition proteins

An essential early step in an immune response is recognition of an invading pathogen as dangerous non-self (Janeway & Medzhitov, 2002, Matzinger, 2002, Sansonetti, 2006). Surface patterns of potential pathogens, such as bacteria and fungi, include β-1,3-glucan, peptidoglycan (PGN), lipopolysaccharide (LPS), and lipoteichoic

acid (LTA), which can be recognized by insect pattern recognition proteins (Kanost et al, 2004, Yu et al, 2002). Binding of these pattern recognition proteins in turn leads to activation of other immune responses. In several cases, pattern recognition proteins have been either discovered or studied by investigating plasma proteins that bind to bacteria *in vitro*.

Hemolin

Sun et al. used an *in vitro* bacterial binding assay to look at proteins from *Hyalophora cecropia* that bind to *E. coli.* They found hemolin and a 125 kDa protein bound to *E. coli* that had been incubated with hemolymph for one minute, and they detected additional polypeptides, including one at 170 kDa, bound to *E. coli* incubated with hemolymph for three minutes (Sun et al, 1990). Hemolin is a 48 kDa soluble protein found in lepidopteran hemolymph. Hemolin contains four immunoglobulin (Ig)-like domains that form a horseshoe-like three-dimensional structure (Su et al, 1998). *M. sexta* hemolin can bind to many different bacterial surface components, including LPS and lipoteichoic acid (Yu & Kanost, 2002), as well as to hemocytes (Ladendorff & Kanost, 1991). A hemocyte membrane associated form of hemolin has also been observed in *H. cecropia* (Bettencourt et al, 1997). Hemolin is present at low levels in naïve hemolymph but is up-regulated in response to bacterial challenge. In naïve insects, hemolin expression increases in both fat body and midgut lumen during the wandering stage (Yu & Kanost, 1999). Recent RNAi studies showed that knockdown of hemolin expression led to decreased phagocytosis and nodulation of *E. coli* (Eleftherianos et al, 2007b). Hemolin is also proposed to be involved in anti-viral defense (Terenius, 2008).

Dscam

Another intriguing member of the immunoglobulin superfamily is Dscam, which consists of 10 immunoglobulin-like domains, six type III fibronectin domains, a transmembrane region and a cytoplasmic domain (Boehm, 2007). *D. melanogaster* Dscam has 38,016 possible different isoforms from 95 exons due to mutually exclusive alternative splicing of exons 4, 6, 9, and 17. Dscam is essential for neuronal wiring (Schmucker, 2007). The majority of Dscam isoforms show hemophilic binding: they bind to the same isoform but not to other possible isoforms. Binding triggers cellular

responses but not adhesion, thereby providing a way for sister dendrites to recognize themselves and avoid making connections with self (Schmucker, 2007). In addition to its role in the brain, Dscam has been proposed to have a role in immunity. Expression of alternative exons detected in cDNA from fat body and hemocytes suggested that those tissues could potentially generate 18,000 *D. melanogaster* Dscam isoforms. Secreted forms of Dscam have been observed in S2 cell media and plasma (Watson et al, 2005). Silencing Dscam in hemocytes and Dscam antibodies added to S2 cells both significantly reduced rates of phagocytosis (Watson et al, 2005). *A. gambiae* Dscam has 101 exons and greater than 31,000 potential splice forms. Certain splice forms are upregulated upon immune challenge, and RNAi knock down of a non-spliced exon led to decreased survival rates and a 50% reduction in phagocytosis after exposure to *S. aureus* and *E. coli* (Dong et al, 2006). AgDscam proteins from anophiline cell line Sua 5B bound *in vitro* to *S. aureus* and *E. coli* as well as co-localizing with *E. coli* during immunohistochemistry (Dong et al, 2006). Dscam genes have been found in other insects, including *B. mori* and *Tribolium castaneum* (Watson et al, 2005).

X-ray crystallographic studies of the DmDscam N-terminal four immunoglobulin domains from two isoforms show a horseshoe configuration similar to that of hemolin (Meijers et al, 2007). Alternatively spliced exons 4 and 6 encode part of these first four domains. One side of the horseshoe, epitope I, is required for homophilic dimerization, while epitope II on the other side of the molecule could be involved in immunity-related binding (Boehm, 2007, Meijers et al, 2007). Both exon 4 and exon 6 cover parts of both epitopes. Sequence conservation results are consistent with the proposed dual roles: epitope I, required for neuronal wiring, is more highly conserved among insects while epitope II is more divergent, possibly due to selection pressure in binding to pathogens (Meijers et al, 2007). The role of epitope II in microbial binding still requires experimental testing, as does the role of Dscam in hemocyte interactions.

Immulectins

Another class of proteins that binds both hemocytes and microbial patterns are calcium-dependent (C-type) lectins. In *M. sexta*, four soluble C-type lectins with two carbohydrate-recognition domains have been characterized and named immulectins 1-4

(IML-1-4) (Yu et al, 1999, Yu & Kanost, 2000, Yu et al, 2005, Yu et al, 2006). Similar tandem-domain C-type lectins are found in other lepidopterans and are involved in binding bacteria and hemocyte aggregation (Koizumi et al, 1999, Shin et al, 2000). *A. gambiae* and *D. melanogaster* have about 20 and 32 C-type lectins, respectively, but each of them contain only one carbohydrate recognition domain (Ao et al, 2007, Yu & Kanost, 2003, Yu & Ma, 2006). Expression of all four *M. sexta* IMLs is up-regulated in the fat body upon immune challenge (Yu et al, 2002, Yu et al, 2005, Yu et al, 2006). IML-1 agglutinated bacteria and yeast while IML-2 caused aggregation of only gram-negative bacteria (Yu et al, 1999, Yu & Kanost, 2000). IML-2 has also been found to bind granular cells and oenocytoids, as well as nematodes (Yu & Kanost, 2004). More recently IML-2 was shown to bind to a wide range of microbial patterns, including LTA, laminarin (branched β-1, 3-glucan), mannose, and all three moieties of LPS (lipid A, core carbohydrate, and O-specific antigen) (Yu & Ma, 2006). IML-3 and IML-4 are 56% identical and both can bind to LPS, LTA, and laminarin (Yu et al, 2005, Yu et al, 2006). Coating agarose beads with the carbohydrate recognition domain-2 (CRD-2) of IML-2 or full-length IML-3 or IML-4 increased encapsulation of those beads to almost 100%, compared with close to 0% in control beads coated with BSA or a *M. sexta* cuticle protein, CP36 (Yu & Kanost, 2004, Yu et al, 2005, Yu et al, 2006). However, bead melanization depended on the immulectin used. Few beads coated with IML-3 became melanized, whereas 16% of the IML-4 coated beads were melanized as were the majority of beads with CRD-2 of IML-2 (Yu & Kanost, 2004, Yu et al, 2005, Yu et al, 2006). Immulectins bind microbial surfaces and hemocytes and participate in aggregation and encapsulation responses; this evidence indicates that immulectins are pattern recognition proteins important for cellular responses.

β-1,3 glucan recognition proteins

β-1,3-glucans are components of fungal and some bacterial cell walls and can be recognized by β-1,3-glucan recognition proteins (βGRPs). Two βGRPs have been characterized in *M. sexta* and, like other βGRPs found in arthropods, they bind β-1,3-glucan and have a catalytically inactive glucanase domain at the C-terminus (Kanost et al, 2004). βGRPs also strongly increase prophenoloxidase (proPO) activation in the

presence of laminarin (Fabrick et al, 2003, Jiang et al, 2004, Ma & Kanost, 2000). The two βGRP domains have been studied more in the moth *Plodia interpunctella*. Full-length *P. interpunctella* βGRP is required for agglutination of bacteria, the N-terminal domain alone can bind to β-1,3-glucans (curdlan or laminarin), LPS, and LTA, and the C-terminal domain binds to branched β-1,3-glucans (laminarin) (Fabrick et al, 2004). *M. sexta* βGRP1 and 2 each bind and aggregate yeast, Gram-negative bacteria, and Gram-positive bacteria. Their expression patterns, however, differ. βGRP1 is not induced by immune challenges but is constitutively expressed during the feeding and wandering stages (Jiang et al, 2004, Ma & Kanost, 2000). βGRP2 is only expressed in feeding stage larvae after wounding or immune challenge yet is strongly expressed even in naïve insects starting at the wandering stage (Jiang et al, 2004). βGRP related proteins, called Gram-negative bacteria binding proteins (GNBPs), can bind to LPS (Yu et al, 2002). Lee and coworkers isolated a *Bombyx mori* Gram-negative binding protein (GNBP) from plasma through *in vitro* bacterial binding (1996). *B. mori* plasma 18 h after injection with *Enterobacter cloacae* was incubated Gram-negative bacteria, *E. cloacae,* for 1-30 min. After eluting proteins from the bacteria and analysis by SDS-PAGE, they detected lysozyme and GNBP; other bands remained unidentified. Three GNBP proteins are present in the *D. melanogaster* genome while three other proteins are similar only to the N-terminal domain (Lemaitre & Hoffmann, 2007). The roles of GNBP1 and GNBP3 are discussed in more detail below.

Peptidoglycan recognition proteins

Peptidoglycan (PGN) is an essential component of both Gram-positive and Gram-negative bacterial cell walls (Kurata et al, 2006). PGN is a polymer made up of a repeating disaccharide unit, β-1,4 linked *N*-acetylglucosamine and *N*-acetylmuramic acid. These polymer chains are crosslinked by short peptides in which the third residue is either a lysine, as in most Gram-positive bacteria, or a diaminopimelic acid (DAP) group in Gram-negative bacteria and *Bacillus* species. Lys-type PGN and DAP-type PGN can be recognized by peptidoglycan recognition proteins (PGRP), which are present in invertebrates and vertebrates (Dziarski & Gupta, 2006).

The first PGRP was discovered in *B. mori* and found to activate the prophenoloxidase cascade in the presence of PGN (Yoshida et al, 1996). One PGRP from *M. sexta* has been studied. Its expression is immune-inducible and increases from 2 μg/ml in naïve larval plasma to 60 μg/ml in plasma from larvae 24 h after injection of bacteria (Yu et al, 2002). The two isoforms of *M. sexta* PGRP, A and B, differ only in the signal peptide region and are probably alleles of the same gene. The PGRP protein from *Samia cynthia ricini*, a wild silk worm, has a high similarity to *M. sexta* PGRP and has recently been studied. PGRP binding assays with radiolabeled peptidoglycan indicate that *S. ricini* PGRP binds to DAP-type cross-linked PGN (from *Bacillus* cell wall) and to uncross-linked Lys-type PGN (from *M. luteus*) but not very strongly to cross-linked Lys-type PGN (Onoe et al, 2007).

There are 13 known PGRPs in *D. melanogaster,* and recent work has shown that different PGRPs can recognize different types of PGN (Kurata et al, 2006, Wang & Ligoxygakis, 2006). The long forms of PGRPs (PGRP-Ls) are transmembrane proteins with extracellular PGRP domains, while shorter PGRPs (PGRP-Ss) are soluble and secreted. Over-expression of PGRP-LE in larvae led to ectopic melanization and increased phenoloxidase activity in hemolymph (Takehana et al, 2002) while bacteria-treated PGRP-LE mutants were unable to activate phenoloxidase in hemolymph, unlike wild-type flies (Takehana et al, 2004). PGRP-LC and PGRP-LE together lead to activation of the IMD pathway, which stimulates synthesis of antimicrobial peptides in response to Gram-negative bacteria (Choe et al, 2002, Gottar et al, 2002, Takehana et al, 2004) and PGRP-LC is involved in phagocytosis of Gram-negative bacteria by hemocytes (Ramet et al, 2002). Short PGRPs, PGRP-SA and PGRP-SD, are involved in Toll pathway activation and are discussed in more detail below.

Toll pathway

First discovered in *D. melanogaster*, the Toll pathway triggers synthesis of drosomycin and other antimicrobial peptides in response to fungi and many Gram-positive bacteria (Lemaitre & Hoffmann, 2007). The response is initiated by recognition proteins that bind pathogens or microbial patterns. That binding triggers an extracellular serine protease cascade involving multiple serine proteases, the first of which to be

discovered was Persephone (Ligoxygakis et al, 2002). The serine protease cascade also includes spätzle processing enzyme (SPE) which cleaves and activates the cytokine spätzle (Jang et al, 2006). Activated spätzle binds to the Toll membrane receptor in fat body and hemocytes and triggers an intracellular signal transduction cascade that releases the transcription factors dorsal and dorsal-related immune factor (DIF) from their inhibitor, cactus. Dorsal and DIF then translocate to the nucleus and induce synthesis of drosomycin and other antimicrobial peptides (Imler & Hoffmann, 2001). The Toll pathway was initially discovered for its role in dorsal-ventral patterning during embryogenesis (Anderson et al, 1985). The Toll pathway required for dorsal ventral patterning is similar to the immune related Toll pathway, but it involves different serine proteases leading to spätzle processing, and stimulates transcription of different genes (LeMosy et al, 2001). Other Toll family members in *D. melanogaster* include 18 wheeler and Toll-3 through 9. They share domain structure with Toll (extracellular leucine-rich repeats and cytoplasmic TIR domains) but their roles are still poorly defined (Imler & Hoffmann, 2001).

In *D. melanogaster* multiple PGRPs and other pattern recognition proteins are involved in activation of the Toll pathway. PGRP-SA and PGRP-SD activate the Toll pathway in response to Lys-type PG, found in many Gram-positive bacteria (Bischoff et al, 2004, Michel et al, 2001). PGRP-SA forms a complex with Gram-negative binding protein 1 (GNBP1), and both proteins are required for Toll activation in response to Gram-negative bacteria (Gobert et al, 2003, Pili-Floury et al, 2004). GNBP1 can enzymatically cleave PGN into monomeric muropeptides, an activity that is also enhanced by the presence of PGRP-SA (Wang et al, 2006). PGRP-SA is proposed to bind the released PG fragments and activate an unknown serine protease to start a serine protease cascade leading to cleavage of Spätzle and activation of the Toll receptor (Wang et al, 2006).

The Toll pathway is activated in response to fungi via two pathways. GNBP3, which is ~65% similar in its N-terminal domain to the N-terminal domains of *M. sexta* βGRPs, binds to β-1,3 glucan and induces expression of drosomycin in response to *Candida albicans* (Gottar et al, 2006). Interestingly, certain entomopathogenic fungi including *Beauveria bassiana* and *Metarhizium anisopliae,* which secrete serine proteases

and chitinases to penetrate the cuticle, activate the Toll pathway in the absence of GNBP3. The fungal PR1A protease activates the serine protease Persephone and thereby initiates the Toll pathway (Gottar et al, 2006).

Toll-like receptors have been identified in numerous insect species (Ao et al, 2008). *M. sexta* has a Toll pathway but it is significantly less well understood than the pathway in *D. melanogaster*. Recently *B. mori* spätzle-1 has been shown to up-regulate antimicrobial gene expression in both *B. mori* and *M. sexta* (Wang et al, 2007). A *M. sexta* Toll receptor is constitutively expressed in several tissues and is present on the surface of hemocytes. This Toll receptor's mRNA is up-regulated in hemocytes by 2, 12, or 27 fold in response to yeast, gram-positive bacteria, or gram-negative bacteria, respectively (Ao et al, 2008). Some features of this *M. sexta* Toll are more similar to mammalian Toll-like receptors, which directly bind to pathogens and other signals, than to other insect Toll-like receptors (Ao et al, 2008). More research is required before we can understand the role of Toll-like receptors in *M. sexta*.

Phenoloxidase activation cascade

In contrast to the Toll pathway, the *M. sexta* pathways for prophenoloxidase (proPO) activation are relatively well understood. PO is a key enzyme involved in melanin synthesis, a component of some immune responses that has been implicated in microbial killing through the generation of toxic compounds. If all the necessary components are present, melanization can occur hours faster than antimicrobial peptide synthesis. In addition to immune defense reactions, melanin may be important for wound healing (Sugumaran, 2002). Synthesis of melanin leads to formation of reactive quinone intermediates, cytotoxic molecules like 5,6-dihydroxyindole (DHI), and reactive oxygen or nitrogen intermediates that may contribute to the killing of invading pathogens (Cerenius & Soderhall, 2004, Nappi & Christensen, 2005, Zhao et al, 2007a). Melanization can also occur in hemocyte nodules or on the surface of encapsulated objects or parasites (Cerenius & Soderhall, 2004). It may also stabilize clots at wound sites (Agianian et al, 2007, Galko & Krasnow, 2004).

PO hydroxylates monophenols such as tyrosine to *o*-diphenols like 3,4-dihydroxyphenylalanine (DOPA) using its monophenol monooxygenase (tyrosinase)

activity (Gorman et al, 2007). PO also has activity as an *o*-diphenol oxidase and oxidizes *o*-diphenols such as DOPA to their corresponding quinones (Gorman et al, 2007, Hall et al, 1995, Sugumaran, 2002). The orthoquinones resulting from this process polymerize to form melanin (Nappi & Christensen, 2005). Specifically, dopaquinone can non-enzymatically convert into dopachrome (Sugumaran, 2002). Dopachrome converting enzyme catalyzes the decarboxylation of dopachrome to DHI, which can be converted by PO to indole-5,6-quinone, which then polymerizes to form DHI eumelanin (Nappi & Christensen, 2005).

Tyrosine hydroxylase is another enzyme with monophenoloxidase activity that can convert tyrosine to DOPA (Gorman et al, 2007). Tyrosine hydroxylase expression is upregulated in *M. sexta* upon immune challenge and has recently been proposed to have a role in melanin synthesis in immune response (Gorman et al, 2007). Tyrosine hydroxylase is intracellular, unlike PO which is found in the hemolymph. Another intracellular enzyme important in the early stages of melanin synthesis is dopa decarboxylase (Huang et al, 2005b). Dopa decarboxylase converts DOPA to dopamine, which is a better PO substrate than DOPA and, after oxidation by PO, non-enzymatically forms DHI eumelanin (Sugumaran, 2002). RNAi studies have shown that reduced dopa decarboxylase expression drastically decreases the melanization rate in the mosquito *Armigeres subalbatus* (Huang et al, 2005a). Thus, even though PO is capable of all the reactions required for oxidation of tyrosine to dopaquinone, other enzymes appear important for a rapid melanization response *in vivo*.

PO is expressed as an inactive precursor, proPO. There are at least two proPO genes in *M. sexta* (Hall et al, 1995, Jiang et al, 1997). *M. sexta* proPO2 contains 695 amino acid residues, runs on SDS-PAGE at about 80 kDa, and is 77% similar to *B. mori* proPO-p2 (Hall et al, 1995). *M. sexta* proPO1 contains 685 residues, runs at about 78 kDa, and is 78% similar to *B. mori* proPO-p1(Jiang et al, 1997). The two *M. sexta* proPOs share almost 50% identity and have been shown to form heterodimers. Both *M. sexta* proPOs lack signal peptides, which is common among arthropod phenoloxidases; they are produced in oenocytoids and, unlike many other plasma proteins, not in fat body (Jiang et al, 1997). In *B. mori,* another study used a sensitive but less specific DOPA assay and suggested that PO is present in a wide variety of hemocytes, including

oenocytoids, plasmatocytes, about half of the prohemocytes and young granulocytes (Ling et al, 2005). Also, PPO has been detected at the cell surface of spherule cells and granulocytes in *M. sexta* (Ling & Yu, 2005).

In lepidopterans, proPO circulates in the hemolymph (Kanost & Gorman, 2008). Controlled activation is essential for confining PO activity to the times and places it is needed. How, then, is proPO activated? *In vivo* proPO is activated through specific proteolytic cleavage. In *M. sexta* three proPO activating proteases (PAPs) have been discovered: PAP1, PAP2, and PAP3 (Jiang et al, 1998, Jiang et al, 2003a, Jiang et al, 2003b). Each of the three PAPs is synthesized as a zymogen and must be activated by specific protease cleavage before being able to cleave and activate proPO. PAP1 contains a single amino terminal clip domain and is activated by cleavage after Arg-127 (Jiang et al, 1998). PAP2 and PAP3 each contain two N-terminal clip domains and are activated by cleavage after Lys-153 and Lys-146, respectively (Jiang et al, 2003a, Jiang et al, 2003b). In addition, all three PAPs also require the presence of serine protease homologs 1 and 2 (SPH-1 and -2) to efficiently activate PO (Jiang et al, 2003a). *M. sexta* proPO can also be artificially activated by the cationic detergent cetylpyridinium chloride (CPC) (Hall et al, 1995, Jiang et al, 1997).

Recently serine protease cascades have been elucidated that lead to proPO activation in *M. sexta* (Gorman et al, 2007, Wang & Jiang, 2007). These discoveries relied on earlier efforts that identified serine proteinases in *M. sexta* by screening cDNA libraries (Jiang et al, 2005). At the top of the identified pathways is HP14, a complex protein that contains five low-density lipoprotein receptor class A domains, a sushi domain, a cysteine rich region, and a C-terminal protease domain (Ji et al, 2004). HP14 can auto-activate in the presence of PGN or β-1,3-glucan and βGRP2 (Wang & Jiang, 2006). HP14 then activates HP21, although not necessarily cleaving it at the expected cleavage site (Gorman et al, 2007). Finally, HP21 cleaves and activates either proPAP2 or proPAP3, which can then activate PPO in the presence of cleaved SPHs (Gorman et al, 2007, Wang & Jiang, 2007). Preliminary experiments show proPAP1 is not activated by HP21 (Wang & Jiang, 2007).

Microbial patterns are not the only stimuli that trigger proPO activation. One example is the melanization that occurs during wound healing in *D. melanogaster* (Bidla

et al, 2007). Bidla and coworkers observed that clot fragments that became melanized often contained crystal cell fragments and plasmatocytes (which are analogous to lepidopteran oenocytoid fragments and granulocytes, respectively). An analysis of *D. melanogaster* mutants indicates that crystal cell lysis is triggered through a signal transduction pathway that includes the membrane-bound TNF homolog Eiger, small GTPases including RhoA, and Jun N-terminal kinase Basket. Additionally, melanization was activated when the pro-apoptotic protein Grim was over-expressed in hemocytes but inhibited by co-expression of the caspase inhibitor p35 (Bidla et al, 2007).

Inhibition of PO

PO activation is a dangerous process, involving active proteases and leading to the production of cytotoxic quinones and reactive oxygen and nitrogen species (Kanost & Gorman, 2008). Overactive melanization can lead to abnormal systemic melanization and damage to the host tissues (Kanost & Gorman, 2008). Inhibition of active phenoloxidase can occur through multiple mechanisms. The first phenoloxidase inhibitor (POI) to be discovered was a 4.2 kDa peptide in pupae of the house fly, *Musca domestica* (Daquinag et al, 1995, Tsukamoto et al, 1992). This 38 amino acid peptide contains a DOPA at position 32, which has been modified from tyrosine, and has 6 cysteines linked together in 3 disulfide bridges. *M. domestica* POI functions as a competitive inhibitor of PO but the exact nature of the interaction is unknown.

Based on the *M. domestica* sequence, a 335 amino acid protein in *A. gambiae* was found that contains 5 tandem POI inhibitor motifs, the fifth of which contains a tyrosine in the same position as in the *M. domestica* sequence (Shi et al, 2006). RNAi of the *A. gambiae* POI increased the amount of melanization at wound sites but not of injected beads. This group also reported finding 2 ESTs for a *M. sexta* POI with a single inhibitor motif. Lu and Jiang expressed and purified the *M. sexta* POI from one of the identified ESTs but it did not show significant PO inhibition (Lu & Jiang, 2007). They did, however, partially isolate a low molecular weight compound from larval plasma that strongly inhibited PO. This inhibitor is suspected to contain a phenyl ring based on its strong absorption at 250-60 nm.

Another small molecule inhibitor of phenoloxidase made by a pathogenic bacterium, *Photorhabdus luminescens*, has been shown to inhibit phenoloxidase activity in *M. sexta* (Eleftherianos et al, 2007a). This inhibitor, (E)-1,3-dihydroxy-2-(isopropyl)-5-(2-phenylethenyl)benzene (ST), is a major secreted product of *P. luminescens* and is made from cinnamic acid. Strains of *P. luminescens* unable to produce ST due to disrupted production of cinnamic acid were slower to proliferate and took significantly longer to kill infected *M. sexta* than the wild type strains. However, the same attenuated strain was effective at killing insects in which PPO had been knocked down by RNAi, leading Eleftherianos et al. to conclude that the effect of ST occurs through its inhibition of PO (Eleftherianos et al, 2007a).

Inhibition of PO itself is only part of the down-regulation that needs to occur to limit the melanization response. If not inhibited, the active proteases involved in proPO activation could cleave other host proteins and cause damage or trigger excessive activation of proPO. Thus, serine protease inhibitors are an important part of controlling the responses that rely on serine proteases. Melanization and activation of the Toll pathway both use a cascade of serine proteases to generate an effecter response. Another well-known example of a serine protease cascade in arthropod hemolymph is coagulation in the horse shoe crab (Iwanaga, 2007). Vertebrate systems have many similarly organized serine protease dependent processes, including complement activation, blood clotting, and fibrinolysis (Krem & Di Cera, 2002).

Serine protease inhibitors in insect hemolymph

Serine protease inhibitors are important for regulation of serine proteases and the physiological processes in which serine proteases are involved. Small inhibitors in hemolymph include both Kunitz-type inhibitors (Kanost et al, 1990, Kanost, 1999, Zhao et al, 2007b), and ~35 residue pacifastin inhibitors, best studied in locust species but identified in other insects including *A. gambiae* (based on genomic sequence) and *M. sexta* (based on ESTs) (Gaspari et al, 2004). Larger pacifastin inhibitors have been found in other arthropods, such as crayfish, and shown to inhibit proPO activation as do certain smaller pacifastin inhibitors from locust (Franssens et al, 2007, Kanost, 1999). Serine protease inhibitors from the serpin superfamily are about 45 kDa and have been more

extensively studied in insects than the other types of inhibitors (Kanost et al, 2004). Serpins are known to inhibit serine proteases involved in the Toll pathway (Levashina et al, 1999) and the melanization pathways in several insects, including *D. melanogaster* (De Gregorio et al, 2002), *M. sexta* (Jiang & Kanost, 1997, Tong & Kanost, 2005, Tong et al, 2005, Wang & Jiang, 2004, Zhu et al, 2003), and *A. gambiae* (Michel et al, 2006). Human plasma also contains serpins that regulate serine protease cascades for processes such as blood coagulation and complement activation (Kanost, 1999, Pike et al, 2005, Silverman et al, 2001).

Serpin mode of action

Serpins form covalent complexes with target proteases through a dramatic mechanism. The carboxyl-terminal region of the serpin is an extended loop that serves as bait for the target protease(s). Protease selectivity is based on the sequence of this serpin reactive site loop, especially the P1 residue. The protease binds, with the P1 residue of the serpin in the protease substrate specificity pocket, and then cleaves the serpin. Upon cleavage, the serpin undergoes a massive rearrangement, inserting its reactive site loop into one of its beta sheets and moving the protease with it about 70 angstroms, distorting the protease active site (Silverman et al, 2001, Whisstock & Bottomley, 2006). The covalent ester linkage between the protease and serpin remains intact because it is not available to water (Dementiev et al, 2006).

Insect serpins

Genome analysis of certain insect species shows a range in the number of serpin genes from 7 in *Apis mellifera* to 31 in *T. castaneum* (Zou et al, 2007). Phylogenetic analysis demonstrates that certain serpins involved in regulating melanization are related: *D. melanogaster* Spn27a, *M. sexta* serpin-3, and *A. gambiae* serpins-1, -2, and -3. All of these serpins, except AgSerpin-3, can inhibit PPO activation when added to plasma *in vitro* (De Gregorio et al, 2002, Michel et al, 2006, Zhu et al, 2003). In mosquitoes, an expansion cluster of serpins includes serpins-4, -5, and -6. Serpin-6 is expressed in *A. gambiae* midguts and salivary glands during invasion by *Plasmodium berghei* (Abraham et al, 2005, Pinto et al, 2007).

M. sexta serpins

M. sexta has at least seven serpin genes but the overall number of functional serpins is higher due to alternative splicing of the ninth exon of serpin-1 which yields 12 isoforms with different reactive site loops and therefore different protease selectivities (Jiang et al, 1994, Jiang et al, 1996, Jiang & Kanost, 1997, C. Suwanchaichinda, M. Kanost, and R. Ochieng, personal communication). Serpin-2 is located within hemocytes and its expression is up-regulated upon bacterial challenge, but not much is known about its functional role in the insect (Gan et al, 2001). The other six serpins are secreted into hemolymph. Serpin-1 isoforms, examined as an aggregate, are expressed during larval feeding, but not during larval or pupal molts, and overall expression level is unchanged after immune challenge (Kanost et al, 1995). In response to bacterial challenge, serpins-3, -4, -5, and -6 are up-regulated in fat body; serpins-4, -5, and -6 are also up-regulated in hemocytes (Tong & Kanost, 2005, Zhu et al, 2003, Zou & Jiang, 2005). Serpin-3 inhibits all three PAPs while serpin-6 specifically inhibits PAP-3 (Wang & Jiang, 2004). Serpin-1 isoform J can form a complex with PAP-3 *in vitro* (Jiang & Kanost, 1997). Serpin-4 forms serpin-protease complexes with hemolymph proteinases HP-1, HP-6 and HP-21 while serpin-5 complexes with HP-1 and HP-6 (Tong et al, 2005). Addition of recombinant serpins-1J, 3, 4, 5, and 6 can inhibit proPO activation in plasma, indicating that a protease they inhibit is involved in proPO activation. Because serpin-4 and serpin-5 inhibit proPO activation but do not inhibit the PAPs, at least one target of serpin-4 and serpin-5 (HP-1 and/or HP-6) is likely to be involved in a serine protease cascade upstream of proPO. The other target of serpin-4, HP21, has a known function in proPO activation through activation of PAP-2 and PAP-3, as discussed above. HP-1 and/or HP6 may be part of an additional PAP activation pathway or a pathway involved in SPH cleavage. Functions for the remaining serpin-1 isoforms & alternative functions for the other extracellular serpins remain unknown.

Proteomics

Many proteins from insect hemolymph are of interest because of their roles in immunity, wound healing, and the regulation of those processes, leading to a need for more thorough studies of insect proteomes. Proteomics is derived from the term

proteome, which describes all of the expressed proteins in a sample (such as a tissue, fluid, or cell) and was first used in 1994 (Shi & Paskewitz, 2006). Mass spectrometry is a powerful technique for studying proteins in biological systems and has been used to identify proteins involved complexes and pathways (Aebersold & Mann, 2003, Cravatt et al, 2007). Advances in protein identification by mass spectrometry (MS) have facilitated massive growth of the proteomics field.

Identification of proteins using peptide mass fingerprinting

A protein can be identified by its amino acid sequence and, if cleaved after a few specific amino acids by a protease such as trypsin, from the masses of the resulting peptides if the sequence is known or can be predicted from cDNA or gene sequences. This is the foundation of peptide mass fingerprinting (PMF), a technique first conceived in 1989 and now in widespread use (Henzel et al, 2003). Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry (MS) is one method that is now routinely used to generate experimental data for PMF and thereby allows fast identification of proteins (Webster & Oxley, 2005). For identification by PMF, proteins are often separated by SDS-PAGE, either by molecular weight (1D) or both isoelectric point and molecular weight (2D). Bands or spots of proteins on the gel are visualized by staining, cut out of the gel, and digested with a protease such as trypsin, which cleaves after Arg and Lys residues. The masses of the resulting peptides are determined by MALDI-TOF and then compared against the calculated peptide masses of known proteins in the protein databases using software such as Mascot or Aldente (Biron et al, 2006). The PMF results include a score to indicate whether the match with the database is statistically significant. Three or four different proteins can be identified from a single band or spot (Henzel et al, 2003).

Additional information about individual peptides can be obtained by tandem mass spectrometry or MS/MS. One method is MALDI-TOF/TOF, where a peptide of a particular mass is selected and further fragmented. The masses of the fragmented peptide give additional information about its sequence (Coon et al, 2005, Vestal & Campbell, 2005).

Overcoming challenges with PMF

One obvious limitation in identification of a protein by PMF is that the protein sequence must be in a database. Complete databases exist if the protein is from an organism that has its entire genome sequenced, but a search of Genbank in February 2008 shows that only 37 species from 13 insect genera have been sequenced, and not all of them are completely assembled. If a protein happens to be ~70% or more identical with protein(s) from another species they will most likely have some tryptic peptides with the same masses and cross-species identification can occur (Biron et al, 2006). However, even one amino acid change in a peptide (with the exception of leucine to or from isoleucine) will change the mass of that peptide and prevent its identification by PMF.

To identify proteins in *M. sexta* by using PMF, we can use the hundreds of *M. sexta* proteins currently in the protein databases and, for highly identical proteins, the known insect proteins from other insect species. We can also search against *M. sexta* EST sequences translated in all 6 frames against the MS and MS/MS data (Zou et al, in press).

Another challenge in PMF is identification of specific protein isoforms. A protein identification is probably significant if the experimentally determined masses cover 20% or more of the protein sequence (Biron et al, 2006). Isoform specific identification cannot be made if the region of the protein that differs between isoforms is not included in the masses detected by MALDI-TOF. Even very good PMF data seldom give more than 40% coverage of a 30-50 kDa protein because MALDI-TOF best identifies masses of peptides between 800-3,000 Da, and, usually, some tryptic peptides are smaller or larger (Biron et al, 2006). Higher sequence coverages are possible by digesting samples with different proteases. For example, 100% sequence coverage of a 20 kDa protein was obtained by using four separate enzymes and combining the sequence coverage from each digestion and MALDI-TOF analysis (Myung et al, 2007).

Use of PMF in insect studies

Proteomic profiling studies in insects have often used 2D SDS-PAGE followed by MS to identify the proteins present in different tissues or those differentially regulated under different treatment conditions. There have been some studies of the proteins

expressed in *D. melanogaster* larval plasma and hemolymph (de Morais Guedes et al, 2003, Karlsson et al, 2004, Vierstraete et al, 2003), including examination of specific changes in protein expression after immune challenge (de Morais Guedes et al, 2005, Levy et al, 2004, Vierstraete et al, 2004a, Vierstraete et al, 2004b). A similar study has also been conducted in *A. gambiae* (Paskewitz & Shi, 2005). *Apis mellifera* hemolymph proteins were compared across life stages; these comparisons revealed that adult workers had higher levels of immune-related proteins than larvae (Chan et al, 2006). Methodological differences between proteomic profiling studies, including the age of the insects, the tissue studied, the method of tissue collection, and method and timing of immune challenge (if used) help explain the differences between studies within the same insect species (Shi & Paskewitz, 2006).

In *B. mori*, studies have looked at the proteome of the silk gland (Hou et al, 2007, Zhang et al, 2006) and at the changes in the silk gland during programmed cell death (Jia et al, 2007). Larval hemolymph in *B. mori* has been more modestly studied. One study identified eight hemolymph proteins from day seven fifth instar larvae (Li et al, 2006). Another *B. mori* larval hemolymph study identified four proteins induced 24 hours after LPS injection: serpin-2, two novel proteins, and the antimicrobial peptide attacin (Wang et al, 2004). A third study compared hemocyte lysates and plasma before and 6 hours after immune challenge with heat-killed *Bacillus megaterium*. Heat shock protein 70 was upregulated in hemocytes, whereas in plasma, a PGRP, a protein related to antennal pheromone binding protein, a serpin (antichymotrypsin precursor), and four gloverin-like proteins were upregulated (Song et al, 2006).

Recently, proteins from *M. sexta* hemolymph were studied by 2D-PAGE followed by mass spectrometry and 1D-PAGE followed by mass spectrometry or Edman degradation (Furusawa et al 2008). Immune-related proteins accounted for half of the proteins identified, which is not surprising because the *M. sexta* genome is not yet sequenced and many of the *M. sexta* proteins in the database are there because they have been studied for their role in immune responses. Eleven spots on the 2D gel were identified as serpin-1, and serpin-1 and serpin-3 were also identified by analysis of 1D gel slices. The authors reported 58 non-redundant proteins were identified by at least one

of their methods, although they counted certain proteins, like serpin-1 and an imaginal disc growth factor-like protein, multiple times.

In contrast to profiling studies, which generally try to identify as many of the proteins in a tissue or fluid as possible, functional proteomics studies focus on subsets of proteins. For example, functional studies may identify proteins involved in a certain process or identify post-translational modifications on particular proteins (Shi & Paskewitz, 2006). Proteomics has been used to study hemolymph clotting in *D. melanogaster* and *A. gambiae* (Agianian et al, 2007, Karlsson et al, 2004, Scherfer et al, 2004). Karlsson and coworkers used 2D SDS-PAGE to investigate *D. melanogaster* larval plasma proteins before and after clotting. Proteins depleted from plasma after clotting included fondue, which lacks strong sequence homology to any other known proteins, and a cuticle-like protein with a mucin domain (CG8502) (Karlsson et al, 2004, Scherfer et al, 2006). *D. melanogaster* clotting was studied by mixing paramagnetic beads with hemolymph, isolating and washing the beads, and analyzing the proteins that bound to the beads (Scherfer et al, 2004). Nine proteins were enriched on the beads and proposed to be involved in clotting, including hemolectin, larval serum protein γ, fondue, and phenoloxidase (Scherfer et al, 2004, Scherfer et al, 2006). A similar study looking at *A. gambiae* proteins bound to paramagnetic beads detected lipophorin and prophenoloxidase 3 (Agianian et al, 2007). Another functional proteomics study isolated midgut proteins from *A. gambiae* and *D. melanogaster*, bound them to filters, and incubated them with purified *P. berghei* ookinetes. Annexin IX from both species was found to bind ookinetes *in vitro* (Kotsyfakis et al, 2005).

Goals of current research

I brought a functional proteomics approach to three different projects involving plasma proteins in *M. sexta*. The aims of one project were to identify the serpin-1 isoforms present in plasma and to discover some of the endogenous protease targets of serpin-1. To accomplish these goals, serpin-1 isoforms were immunoaffinity purified from plasma, separated by 2D SDS-PAGE, and then individual isoforms and putative serpin-1 protease complexes were identified by MALDI-TOF/TOF. I also investigated complexes of proteins that form during an immune response. I performed gel filtration of

plasma from naïve and immune challenged larvae. Proteins in high molecular weight complexes isolated from the plasma were identified by MALDI-TOF/TOF and immunoblotting. I also used a similar strategy to identify hemolymph proteins that bound to the β-1, 3 glucan curdlan and a variety of bacteria. In both types of immune-related complexes I found proteins known to be involved in melanization, including prophenoloxidase. I also found additional proteins that are candidates for participating in the early stages of immune responses and hemolymph clotting. These studies provide insight into the functions of specific proteins in *M. sexta* plasma.

References

Abraham EG, Pinto SB, Ghosh A, Vanlandingham DL, Budd A, Higgs S, Kafatos FC, Jacobs-Lorena M, & Michel K (2005) An immune-responsive serpin, SRPN6, mediates mosquito defense against malaria parasites. *Proc Natl Acad Sci U S A* **102**: 16327-16332

Aebersold R & Mann M (2003) Mass spectrometry-based proteomics. *Nature* **422**: 198-207

Agianian B, Lesch C, Loseva O, & Dushay MS (2007) Preliminary characterization of hemolymph coagulation in *Anopheles gambiae* larvae. *Dev Comp Immunol* **31**: 879-888

Anderson KV, Jurgens G, & Nusslein-Volhard C (1985) Establishment of dorsal-ventral polarity in the *Drosophila* embryo: genetic studies on the role of the Toll gene product. *Cell* **42**: 779-789

Ao J, Ling E, & Yu XQ (2007) *Drosophila* C-type lectins enhance cellular encapsulation. *Mol Immunol* **44**: 2541-2548

Ao JQ, Ling E, & Yu XQ (2008) A Toll receptor from *Manduca sexta* is in response to *Escherichia coli* infection. *Mol Immunol* **45**: 543-552

Bettencourt R, Lanz-Mendoza H, Lindquist KR, & Faye I (1997) Cell adhesion properties of hemolin, an insect immune protein in the Ig superfamily. *Eur J Biochem* **250**: 630-637

Bidla G, Dushay MS, & Theopold U (2007) Crystal cell rupture after injury in *Drosophila* requires the JNK pathway, small GTPases and the TNF homolog Eiger. *J Cell Sci* **120**: 1209-1215

Biron DG, Brun C, Lefevre T, Lebarbenchon C, Loxdale HD, Chevenet F, Brizard JP, & Thomas F (2006) The pitfalls of proteomics experiments without the correct use of bioinformatics tools. *Proteomics* **6**: 5577-5596

Bischoff V, Vignal C, Boneca IG, Michel T, Hoffmann JA, & Royet J (2004) Function of the drosophila pattern-recognition receptor PGRP-SD in the detection of Gram-positive bacteria. *Nat Immunol* **5**: 1175-1180

Boehm T (2007) Two in one: dual function of an invertebrate antigen receptor. *Nat Immunol* **8**: 1031-1033

Cerenius L & Soderhall K (2004) The prophenoloxidase-activating system in invertebrates. *Immunol Rev* **198**: 116-126

Chan QW, Howes CG, & Foster LJ (2006) Quantitative comparison of caste differences in honeybee hemolymph. *Mol Cell Proteomics* **5**: 2252-2262

Chapman RF (1998) *Insects: Structure and function.* Cambridge University Press: Cambridge UK

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

Christophides GK (2005) Transgenic mosquitoes and malaria transmission. *Cell Microbiol* **7**: 325-333

Coon JJ, Syka JE, Shabanowitz J, & Hunt DF (2005) Tandem mass spectrometry for peptide and protein sequence analysis. *BioTechniques* **38**: 519, 521, 523

Cox-Foster DL, Conlan S, Holmes EC, Palacios G, Evans JD, Moran NA, Quan PL, Briese T, Hornig M, Geiser DM, Martinson V, vanEngelsdorp D, Kalkstein AL, Drysdale A, Hui J, Zhai J, Cui L, Hutchison SK, Simons JF, Egholm M et al (2007) A metagenomic survey of microbes in honey bee colony collapse disorder. *Science* **318**: 283-287

Cravatt BF, Simon GM, & Yates JR,3rd (2007) The biological impact of mass-spectrometry-based proteomics. *Nature* **450**: 991-1000

Daquinag AC, Nakamura S, Takao T, Shimonishi Y, & Tsukamoto T (1995) Primary structure of a potent endogenous dopa-containing inhibitor of phenol oxidase from *Musca domestica*. *Proc Natl Acad Sci U S A* **92**: 2964-2968

De Gregorio E, Han SJ, Lee WJ, Baek MJ, Osaki T, Kawabata S, Lee BL, Iwanaga S, Lemaitre B, & Brey PT (2002) An immune-responsive Serpin regulates the melanization cascade in *Drosophila*. *Dev Cell* **3**: 581-592

de Morais Guedes S, Vitorino R, Domingues R, Tomer K, Correia AJ, Amado F, & Domingues P (2005) Proteomics of immune-challenged *Drosophila* melanogaster larvae hemolymph. *Biochem Biophys Res Commun* **328**: 106-115

de Morais Guedes S, Vitorino R, Tomer K, Domingues MR, Correia AJ, Amado F, & Domingues P (2003) *Drosophila melanogaster* larval hemolymph protein mapping. *Biochem Biophys Res Commun* **312**: 545-554

Dean P, Richards EH, Edwards JP, Reynolds SE, & Charnley K (2004) Microbial infection causes the appearance of hemocytes with extreme spreading ability in monolayers of the tobacco hornworm *Manduca sexta*. *Dev Comp Immunol* **28**: 689-700

Dementiev A, Dobo J, & Gettins PG (2006) Active site distortion is sufficient for proteinase inhibition by serpins: structure of the covalent complex of alpha1-proteinase inhibitor with porcine pancreatic elastase. *J Biol Chem* **281**: 3452-3457

Desjardins M, Houde M, & Gagnon E (2005) Phagocytosis: the convoluted way from nutrition to adaptive immunity. *Immunol Rev* **207**: 158-165

Dong Y, Taylor HE, & Dimopoulos G (2006) AgDscam, a hypervariable immunoglobulin domain-containing receptor of the *Anopheles gambiae* innate immune system. *PLoS Biol* **4**: e229

Dunn PE (1990) Humoral immunity in insects. *BioScience* **40**

Dunn PE, Bohnert TJ, & Russell V (1994) Regulation of antibacterial protein synthesis following infection and during metamorphosis of *Manduca sexta*. *Ann N Y Acad Sci* **712**: 117-130

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

Dziarski R & Gupta D (2006) The peptidoglycan recognition proteins (PGRPs). *Genome Biol* **7**: 232

Eleftherianos I, Boundy S, Joyce SA, Aslam S, Marshall JW, Cox RJ, Simpson TJ, Clarke DJ, ffrench-Constant RH, & Reynolds SE (2007a) An antibiotic produced by an insect-pathogenic bacterium suppresses host defenses through phenoloxidase inhibition. *Proc Natl Acad Sci U S A* **104**: 2419-2424

Eleftherianos I, Gokcen F, Felfoldi G, Millichap PJ, Trenczek TE, ffrench-Constant RH, & Reynolds SE (2007b) The immunoglobulin family protein Hemolin mediates cellular immune responses to bacteria in the insect *Manduca sexta*. *Cell Microbiol* **9**: 1137-1147

Fabrick JA, Baker JE, & Kanost MR (2004) Innate immunity in a pyralid moth: functional evaluation of domains from a beta-1,3-glucan recognition protein. *J Biol Chem* **279**: 26605-26611

Fabrick JA, Baker JE, & Kanost MR (2003) cDNA cloning, purification, properties, and function of a beta-1,3-glucan recognition protein from a pyralid moth, *Plodia interpunctella*. *Insect Biochem Mol Biol* **33**: 579-594

Ferrandon D, Imler JL, Hetru C, & Hoffmann JA (2007) The *Drosophila* systemic immune response: sensing and signalling during bacterial and fungal infections. *Nat Rev Immunol* **7**: 862-874

Franssens V, Simonet G, Breugelmans B, Van Soest S, Van Hoef V, & Vanden Broeck J (2008) The role of hemocytes, serine protease inhibitors and pathogen-associated patterns in prophenoloxidase activation in the desert locust, *Schistocerca gregaria*. *Peptides* **29**: 235-41.

Furusawa T, Rakwal R, Nam HW, Hirano M, Shibato J, Kim YS, Ogawa Y, Yoshida Y, Kramer KJ, Kouzuma Y, Agrawal GK, & Yonekura M (2008) Systematic investigation of the hemolymph proteome of *Manduca sexta* at the fifth instar larvae stage using one- and two-dimensional proteomics platforms. *J. Proteome Res.* **7**: 938-959

Galko MJ & Krasnow MA (2004) Cellular and genetic analysis of wound healing in *Drosophila* larvae. *PLoS Biol* **2**: E239

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

Gaspari Z, Ortutay C, & Perczel A (2004) A simple fold with variations: the pacifastin inhibitor family. *Bioinformatics* **20**: 448-451

Gillespie JP, Kanost MR, & Trenczek T (1997) Biological mediators of insect immunity. *Annu Rev Entomol* **42**: 611-643

Gobert V, Gottar M, Matskevich AA, Rutschmann S, Royet J, Belvin M, Hoffmann JA, & Ferrandon D (2003) Dual activation of the *Drosophila* toll pathway by two pattern recognition receptors. *Science* **302**: 2126-2130

Gorman MJ, An C, & Kanost MR (2007) Characterization of tyrosine hydroxylase from *Manduca sexta*. *Insect Biochem Mol Biol* **37**: 1327-1337

Gorman MJ, Kankanala P, & Kanost MR (2004) Bacterial challenge stimulates innate immune responses in extra-embryonic tissues of tobacco hornworm eggs. *Insect Mol Biol* **13**: 19-24

Gorman MJ, Wang Y, Jiang H, & Kanost MR (2007) *Manduca sexta* hemolymph proteinase 21 activates prophenoloxidase-activating proteinase 3 in an insect innate immune response proteinase cascade. *J Biol Chem* **282**: 11742-11749

Gottar M, Gobert V, Matskevich AA, Reichhart JM, Wang C, Butt TM, Belvin M, Hoffmann JA, & Ferrandon D (2006) Dual detection of fungal infections in *Drosophila* via recognition of glucans and sensing of virulence factors. *Cell* **127**: 1425-1437

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

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

Henzel WJ, Watanabe C, & Stults JT (2003) Protein identification: the origins of peptide mass fingerprinting. *J Am Soc Mass Spectrom* **14**: 931-942

Hillyer JF, Barreau C, & Vernick KD (2007) Efficiency of salivary gland invasion by malaria sporozoites is controlled by rapid sporozoite destruction in the mosquito haemocoel. *Int J Parasitol* **37**: 673-681

Hillyer JF, Schmidt SL, & Christensen BM (2003) Rapid phagocytosis and melanization of bacteria and *Plasmodium sporozoites* by hemocytes of the mosquito *Aedes aegypti*. *J Parasitol* **89**: 62-69

Horohov DW & Dunn PE (1983) Phagocytosis and nodule formation by hemocytes of *Manduca sexta* larvae following injection of *Pseudomonas aeruginosa*. *J Invertebr Pathol* **41**: 203-213

Hou Y, Xia Q, Zhao P, Zou Y, Liu H, Guan J, Gong J, & Xiang Z (2007) Studies on middle and posterior silk glands of silkworm (*Bombyx mori*) using two-dimensional electrophoresis and mass spectrometry. *Insect Biochem Mol Biol* **37**: 486-496

Huang CY, Chou SY, Bartholomay LC, Christensen BM, & Chen CC (2005a) The use of gene silencing to study the role of dopa decarboxylase in mosquito melanization reactions. *Insect Mol Biol* **14**: 237-244

Huang CY, Christensen BM, & Chen CC (2005b) Role of dopachrome conversion enzyme in the melanization of filarial worms in mosquitoes. *Insect Mol Biol* **14**: 675-682

Imler JL & Hoffmann JA (2001) Toll receptors in innate immunity. *Trends Cell Biol* **11**: 304-311

Iwanaga S (2007) Biochemical principle of *Limulus* test for detecting bacterial endotoxins. *Proc Jpn Acad , Ser B* **83**: 110-119

Iwanaga S & Lee BL (2005) Recent advances in the innate immunity of invertebrate animals. *J Biochem Mol Biol* **38**: 128-150

Janeway CA,Jr & Medzhitov R (2002) Innate immune recognition. *Annu Rev Immunol* **20**: 197-216

Jang IH, Chosa N, Kim SH, Nam HJ, Lemaitre B, Ochiai M, Kambris Z, Brun S, Hashimoto C, Ashida M, Brey PT, & Lee WJ (2006) A Spätzle-processing enzyme required for toll signaling activation in *Drosophila* innate immunity. *Dev Cell* **10**: 45-55

Ji C, Wang Y, Guo X, Hartson S, & 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

Jia SH, Li MW, Zhou B, Liu WB, Zhang Y, Miao XX, Zeng R, & Huang YP (2007) Proteomic analysis of silk gland programmed cell death during metamorphosis of the silkworm *Bombyx mori*. *J Proteome Res* **6**: 3003-3010

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

Jiang H, Ma C, Lu ZQ, & Kanost MR (2004) Beta-1,3-glucan recognition protein-2 (betaGRP-2)from *Manduca sexta*; an acute-phase protein that binds beta-1,3-glucan and lipoteichoic acid to aggregate fungi and bacteria and stimulate prophenoloxidase activation. *Insect Biochem Mol Biol* **34**: 89-100

Jiang H, Wang Y, Gu Y, Guo X, Zou Z, Scholz F, Trenczek TE, & Kanost MR (2005) Molecular identification of a bevy of serine proteinases in *Manduca sexta* hemolymph. *Insect Biochem Mol Biol* **35**: 931-943

Jiang H, Wang Y, Huang Y, Mulnix AB, Kadel J, Cole K, & Kanost MR (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, & Kanost MR (1998) Pro-phenol oxidase activating proteinase from an insect, *Manduca sexta*: a bacteria-inducible protein similar to *Drosophila* easter. *Proc Natl Acad Sci U S A* **95**: 12220-12225

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

Jiang H, Wang Y, Ma C, & Kanost MR (1997) 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, Yu XQ, & Kanost MR (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 XQ, Zhu Y, & 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

Jiravanichpaisal P, Lee BL, & Soderhall K (2006) Cell-mediated immunity in arthropods: hematopoiesis, coagulation, melanization and opsonization. *Immunobiology* **211**: 213-236

Kanost M & Gorman MJ (2008) Phenoloxidases in insect immunity. In *Insect Immunity,* Beckage NE (ed) pp 69-96. Elsevier

Kanost MR (1999) Serine proteinase inhibitors in arthropod immunity. *Dev Comp Immunol* **23**: 291-301

Kanost MR, Jiang H, & Yu XQ (2004) Innate immune responses of a lepidopteran insect, *Manduca sexta*. *Immunol Rev* **198**: 97-105

Kanost MR, Kawooya JK, Law JH, Ryan RO, Van Heusden MC, Ziegler R (1990) Insect haemolymph proteins. In *Advances in Insect Physiology,* Evans PD & Wigglesworth VB (eds) pp 299-396. Academic Press: San Diego

Kanost MR, Prasad SV, Huang Y, & Willott E (1995) Regulation of serpin gene-1 in *Manduca sexta*. *Insect Biochem Mol Biol* **25**: 285-291

Karlsson C, Korayem AM, Scherfer C, Loseva O, Dushay MS, & Theopold U (2004) Proteomic analysis of the *Drosophila* larval hemolymph clot. *J Biol Chem* **279**: 52033-52041

Koizumi N, Imai Y, Morozumi A, Imamura M, Kadotani T, Yaoi K, Iwahana H, & Sato R (1999) Lipopolysaccharide-binding protein of *Bombyx mori* participates in a hemocyte-mediated defense reaction against gram-negative bacteria. *J Insect Physiol* **45**: 853-859

Kotsyfakis M, Ehret-Sabatier L, Siden-Kiamos I, Mendoza J, Sinden RE, & Louis C (2005) *Plasmodium berghei* ookinetes bind to *Anopheles gambiae* and *Drosophila melanogaster* annexins. *Mol Microbiol* **57**: 171-179

Krem MM & Di Cera E (2002) Evolution of enzyme cascades from embryonic development to blood coagulation. *Trends Biochem Sci* **27**: 67-74

Kurata S, Ariki S, & Kawabata S (2006) Recognition of pathogens and activation of immune responses in *Drosophila* and horseshoe crab innate immunity. *Immunobiology* **211**: 237-249

Ladendorff NE & Kanost MR (1991) Bacteria-induced protein P4 (hemolin) from *Manduca sexta*: a member of the immunoglobulin superfamily which can inhibit hemocyte aggregation. *Arch Insect Biochem Physiol* **18**: 285-300

Lavine MD & Strand MR (2001) Surface characteristics of foreign targets that elicit an encapsulation response by the moth *Pseudoplusia includens*. *J Insect Physiol* **47**: 965-974

Lee WJ, Lee JD, Kravchenko VV, Ulevitch RJ, & Brey PT (1996) Purification and molecular cloning of an inducible gram-negative bacteria-binding protein from the silkworm, *Bombyx mori*. *Proc Natl Acad Sci U S A* **93**: 7888-7893

Lemaitre B & Hoffmann J (2007) The host defense of *Drosophila melanogaster*. *Annu Rev Immunol* **25**: 697-743

LeMosy EK, Tan YQ, & Hashimoto C (2001) Activation of a protease cascade involved in patterning the *Drosophila* embryo. *Proc Natl Acad Sci U S A* **98**: 5055-5060

Lesch C, Goto A, Lindgren M, Bidla G, Dushay MS, & Theopold U (2007) A role for Hemolectin in coagulation and immunity in *Drosophila melanogaster*. *Dev Comp Immunol*

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

Levy F, Bulet P, & Ehret-Sabatier L (2004) Proteomic analysis of the systemic immune response of *Drosophila*. *Mol Cell Proteomics* **3**: 156-166

Li XH, Wu XF, Yue WF, Liu JM, Li GL, & Miao YG (2006) Proteomic analysis of the silkworm (*Bombyx mori* L.) hemolymph during developmental stage. *J Proteome Res* **5**: 2809-2814

Ligoxygakis P, Pelte N, Hoffmann JA, & Reichhart JM (2002) Activation of *Drosophila* Toll during fungal infection by a blood serine protease. *Science* **297**: 114-116

Ling E, Shirai K, Kanehatsu R, & Kiguchi K (2005) Reexamination of phenoloxidase in larval circulating hemocytes of the silkworm, *Bombyx mori*. *Tissue Cell* **37**: 101-107

Ling E & Yu XQ (2005) Prophenoloxidase binds to the surface of hemocytes and is involved in hemocyte melanization in *Manduca sexta*. *Insect Biochem Mol Biol* **35**: 1356-1366

Lu Z & Jiang H (2007) Regulation of phenoloxidase activity by high- and low-molecular-weight inhibitors from the larval hemolymph of *Manduca sexta*. *Insect Biochem Mol Biol* **37**: 478-485

Ma C & Kanost MR (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

Matzinger P (2002) The danger model: a renewed sense of self. *Science* **296**: 301-305

Meijers R, Puettmann-Holgado R, Skiniotis G, Liu JH, Walz T, Wang JH, & Schmucker D (2007) Structural basis of Dscam isoform specificity. *Nature* **449**: 487-491

Michel K, Suwanchaichinda C, Morlais I, Lambrechts L, Cohuet A, Awono-Ambene PH, Simard F, Fontenille D, Kanost MR, & Kafatos FC (2006) Increased melanizing activity in *Anopheles gambiae* does not affect development of *Plasmodium falciparum*. *Proc Natl Acad Sci U S A* **103**: 16858-16863

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

Miller JS & Stanley DW (2001) Eicosanoids mediate microaggregation reactions to bacterial challenge in isolated insect hemocyte preparations. *J Insect Physiol* **47**: 1409-1417

Myung JK, Frischer T, Afjehi-Sadat L, Pollak A, & Lubec G (2007) Mass spectrometrical analysis of the processed metastasis-inducing anterior gradient protein 2 homolog reveals 100% sequence coverage. *Amino Acids*

Nappi AJ & Christensen BM (2005) Melanogenesis and associated cytotoxic reactions: applications to insect innate immunity. *Insect Biochem Mol Biol* **35**: 443-459

Novotny V, Basset Y, Miller SE, Weiblen GD, Bremer B, Cizek L, & Drozd P (2002) Low host specificity of herbivorous insects in a tropical forest. *Nature* **416**: 841-844

Onoe H, Matsumoto A, Hashimoto K, Yamano Y, & Morishima I (2007) Peptidoglycan recognition protein (PGRP) from eri-silkworm, *Samia cynthia ricini*; protein purification and induction of the gene expression. *Comp Biochem Physiol B Biochem Mol Biol* **147**: 512-519

Paskewitz SM & Shi L (2005) The hemolymph proteome of *Anopheles gambiae*. *Insect Biochem Mol Biol* **35**: 815-824

Pike RN, Buckle AM, le Bonniec BF, & Church FC (2005) Control of the coagulation system by serpins. Getting by with a little help from glycosaminoglycans. *FEBS J* **272**: 4842-4851

Pili-Floury S, Leulier F, Takahashi K, Saigo K, Samain E, Ueda R, & Lemaitre B (2004) *In vivo* RNA interference analysis reveals an unexpected role for GNBP1 in the defense against Gram-positive bacterial infection in *Drosophila* adults. *J Biol Chem* **279**: 12848-12853

Pinto SB, Kafatos FC, & Michel K (2007) The parasite invasion marker SRPN6 reduces sporozoite numbers in salivary glands of *Anopheles gambiae*. *Cell Microbiol*

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

Ryan RO & van der Horst DJ (2000) Lipid transport biochemistry and its role in energy production. *Annu Rev Entomol* **45**: 233-260

Sansonetti PJ (2006) The innate signaling of dangers and the dangers of innate signaling. *Nat Immunol* **7**: 1237-1242

Sass M, Kiss A, & Locke M (1994) Integument and hemocyte peptides. *Journal of Insect Physiology* **40**: 407-421

Scherfer C, Karlsson C, Loseva O, Bidla G, Goto A, Havemann J, Dushay MS, & Theopold U (2004) Isolation and characterization of hemolymph clotting factors in *Drosophila melanogaster* by a pullout method. *Curr Biol* **14**: 625-629

Scherfer C, Qazi MR, Takahashi K, Ueda R, Dushay MS, Theopold U, & Lemaitre B (2006) The Toll immune-regulated *Drosophila* protein Fondue is involved in hemolymph clotting and puparium formation. *Dev Biol* **295**: 156-163

Schmucker D (2007) Molecular diversity of Dscam: recognition of molecular identity in neuronal wiring. *Nat Rev Neurosci* **8**: 915-920

Schnitger AK, Kafatos FC, & Osta MA (2007) The melanization reaction is not required for survival of *Anopheles gambiae* mosquitoes after bacterial infections. *J Biol Chem* **282**: 21884-21888

Shi L, Li B, & Paskewitz SM (2006) Cloning and characterization of a putative inhibitor of melanization from *Anopheles gambiae*. *Insect Mol Biol* **15**: 313-320

Shi L & Paskewitz S (2006) Proteomics and insect immunity. *ISJ* **3**: 4-17

Shin SW, Park DS, Kim SC, & Park HY (2000) Two carbohydrate recognition domains of *Hyphantria cunea* lectin bind to bacterial lipopolysaccharides through O-specific chain. *FEBS Lett* **467**: 70-74

Silverman GA, Bird PI, Carrell RW, Church FC, Coughlin PB, Gettins PG, Irving JA, Lomas DA, Luke CJ, Moyer RW, Pemberton PA, Remold-O'Donnell E, Salvesen GS, Travis J, & Whisstock JC (2001) The serpins are an expanding superfamily of structurally similar but functionally diverse proteins. Evolution, mechanism of inhibition, novel functions, and a revised nomenclature. *J Biol Chem* **276**: 33293-33296

Song KH, Jung SJ, Seo YR, Kang SW, & Han SS (2006) Identification of up-regulated proteins in the hemolymph of immunized *Bombyx mori* larvae. *Comp Biochem Physiol Part D* **1**: 260-266

Sowell RA, Hersberger KE, Kaufman TC, & Clemmer DE (2007) Examining the proteome of *Drosophila* across organism lifespan. *J Proteome Res* **6**: 3637-3647

Stokstad E (2007) Entomology. The case of the empty hives. *Science* **316**: 970-972

Strand MR (2008) Insect hemocytes and their role in immunity. In *Insect Immunity,* Beckage NE (ed) pp 25-47. Elsevier

Stuart LM, Boulais J, Charriere GM, Hennessy EJ, Brunet S, Jutras I, Goyette G, Rondeau C, Letarte S, Huang H, Ye P, Morales F, Kocks C, Bader JS, Desjardins M, &

Ezekowitz RA (2007) A systems biology analysis of the *Drosophila* phagosome. *Nature* **445**: 95-101

Stuart LM & Ezekowitz RA (2005) Phagocytosis: elegant complexity. *Immunity* **22**: 539-550

Su XD, Gastinel LN, Vaughn DE, Faye I, Poon P, & Bjorkman PJ (1998) Crystal structure of hemolin: a horseshoe shape with implications for homophilic adhesion. *Science* **281**: 991-995

Sugumaran M (2002) Comparative biochemistry of eumelanogenesis and the protective roles of phenoloxidase and melanin in insects. *Pigment Cell Res* **15**: 2-9

Sun SC, Lindstrom I, Boman HG, Faye I, & Schmidt O (1990) Hemolin: an insect-immune protein belonging to the immunoglobulin superfamily. *Science* **250**: 1729-1732

Takehana A, Katsuyama T, Yano T, Oshima Y, Takada H, Aigaki T, & 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 U S A* **99**: 13705-13710

Takehana A, Yano T, Mita S, Kotani A, Oshima Y, & Kurata S (2004) Peptidoglycan recognition protein (PGRP)-LE and PGRP-LC act synergistically in *Drosophila* immunity. *EMBO J* **23**: 4690-4700

Terenius O (2008) Hemolin-A lepidopteran anti-viral defense factor? *Dev Comp Immunol* **32**: 311-316

Thomas JA (2005) Monitoring change in the abundance and distribution of insects using butterflies and other indicator groups. *Philos Trans R Soc Lond B Biol Sci* **360**: 339-357

Tong Y, Jiang H, & Kanost MR (2005) Identification of plasma proteases inhibited by *Manduca sexta* serpin-4 and serpin-5 and their association with components of the prophenol oxidase activation pathway. *J Biol Chem* **280**: 14932-14942

Tong Y & Kanost MR (2005) *Manduca sexta* serpin-4 and serpin-5 inhibit the prophenol oxidase activation pathway: cDNA cloning, protein expression, and characterization. *J Biol Chem* **280**: 14923-14931

Tsukamoto T, Ichimaru Y, Kanegae N, Watanabe K, Yamaura I, Katsura Y, & Funatsu M (1992) Identification and isolation of endogenous insect phenoloxidase inhibitors. *Biochem Biophys Res Commun* **184**: 86-92

Vestal ML & Campbell JM (2005) Tandem time-of-flight mass spectrometry. *Methods Enzymol* **402**: 79-108

Vierstraete E, Cerstiaens A, Baggerman G, Van den Bergh G, De Loof A, & Schoofs L (2003) Proteomics in *Drosophila melanogaster*: first 2D database of larval hemolymph proteins. *Biochem Biophys Res Commun* **304**: 831-838

Vierstraete E, Verleyen P, Baggerman G, D'Hertog W, Van den Bergh G, Arckens L, De Loof A, & Schoofs L (2004a) A proteomic approach for the analysis of instantly released wound and immune proteins in *Drosophila melanogaster* hemolymph. *Proc Natl Acad Sci U S A* **101**: 470-475

Vierstraete E, Verleyen P, Sas F, Van den Bergh G, De Loof A, Arckens L, & Schoofs L (2004b) The instantly released *Drosophila* immune proteome is infection-specific. *Biochem Biophys Res Commun* **317**: 1052-1060

Wang L & Ligoxygakis P (2006) Pathogen recognition and signalling in the *Drosophila* innate immune response. *Immunobiology* **211**: 251-261

Wang L, Weber AN, Atilano ML, Filipe SR, Gay NJ, & Ligoxygakis P (2006) Sensing of Gram-positive bacteria in *Drosophila*: GNBP1 is needed to process and present peptidoglycan to PGRP-SA. *EMBO J* **25**: 5005-5014

Wang Y, Cheng T, Rayaprolu S, Zou Z, Xia Q, Xiang Z, & Jiang H (2007) Proteolytic activation of pro-spatzle is required for the induced transcription of antimicrobial peptide genes in lepidopteran insects. *Dev Comp Immunol* **31**: 1002-1012

Wang Y & Jiang H (2007) Reconstitution of a branch of the *Manduca sexta* prophenoloxidase activation cascade in vitro: snake-like hemolymph proteinase 21 (HP21) cleaved by HP14 activates prophenoloxidase-activating proteinase-2 precursor. *Insect Biochem Mol Biol* **37**: 1015-1025

Wang Y & Jiang H (2006) Interaction of beta-1,3-glucan with its recognition protein activates hemolymph proteinase 14, an initiation enzyme of the prophenoloxidase activation system in *Manduca sexta*. *J Biol Chem* **281**: 9271-9278

Wang Y & Jiang H (2004) Purification and characterization of *Manduca sexta* serpin-6: a serine proteinase inhibitor that selectively inhibits prophenoloxidase-activating proteinase-3. *Insect Biochem Mol Biol* **34**: 387-395

Wang Y, Zhang P, Fujii H, Banno Y, Yamamoto K, & Aso Y (2004) Proteomic studies of lipopolysaccharide-induced polypeptides in the silkworm, *Bombyx mori*. *Biosci Biotechnol Biochem* **68**: 1821-1823

Watson FL, Puttmann-Holgado R, Thomas F, Lamar DL, Hughes M, Kondo M, Rebel VI, & Schmucker D (2005) Extensive diversity of Ig-superfamily proteins in the immune system of insects. *Science* **309**: 1874-1878

Webster J & Oxley D (2005) Peptide mass fingerprinting: protein identification using MALDI-TOF mass spectrometry. *Methods Mol Biol* **310**: 227-240

Whisstock JC & Bottomley SP (2006) Molecular gymnastics: serpin structure, folding and misfolding. *Curr Opin Struct Biol* **16**: 761-768

Wiegand C, Levin D, Gillespie J, Willott E, Kanost M, & Trenczek T (2000) Monoclonal antibody MS13 identifies a plasmatocyte membrane protein and inhibits encapsulation and spreading reactions of *Manduca sexta* hemocytes. *Arch Insect Biochem Physiol* **45**: 95-108

Willott E, Wang XY, & Wells MA (1989) cDNA and gene sequence of *Manduca sexta* arylphorin, an aromatic amino acid-rich larval serum protein. Homology to arthropod hemocyanins. *J Biol Chem* **264**: 19052-19059

Wyatt GR (1961) Biochemistry of Insect Hemolymph. *Annu Rev Entomol* **6**: 75-102

Wyatt GR & Pan ML (1978) Insect plasma proteins. *Annu Rev Biochem* **47**: 779-817

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 XQ, Gan H, & Kanost MR (1999) Immulectin, an inducible C-type lectin from an insect, *Manduca sexta*, stimulates activation of plasma prophenol oxidase. *Insect Biochem Mol Biol* **29**: 585-597

Yu XQ & Kanost MR (2004) Immulectin-2, a pattern recognition receptor that stimulates hemocyte encapsulation and melanization in the tobacco hornworm, *Manduca sexta*. *Dev Comp Immunol* **28**: 891-900

Yu XQ & Kanost MR (2003) *Manduca sexta* lipopolysaccharide-specific immulectin-2 protects larvae from bacterial infection. *Dev Comp Immunol* **27**: 189-196

Yu XQ & Kanost MR (2002) 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 XQ & Kanost MR (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 XQ & Kanost MR (1999) Developmental expression of *Manduca sexta* hemolin. *Arch Insect Biochem Physiol* **42**: 198-212

Yu XQ, Ling E, Tracy ME, & Zhu Y (2006) Immulectin-4 from the tobacco hornworm *Manduca sexta* binds to lipopolysaccharide and lipoteichoic acid. *Insect Mol Biol* **15**: 119-128

Yu XQ & Ma Y (2006) Calcium is not required for immulectin-2 binding, but protects the protein from proteinase digestion. *Insect Biochem Mol Biol* **36**: 505-516

Yu XQ, Tracy ME, Ling E, Scholz FR, & Trenczek T (2005) A novel C-type immulectin-3 from *Manduca sexta* is translocated from hemolymph into the cytoplasm of hemocytes. *Insect Biochem Mol Biol* **35**: 285-295

Yu XQ, Zhu YF, Ma C, Fabrick JA, & Kanost MR (2002) Pattern recognition proteins in *Manduca sexta* plasma. *Insect Biochem Mol Biol* **32**: 1287-1293

Zhang P, Aso Y, Yamamoto K, Banno Y, Wang Y, Tsuchida K, Kawaguchi Y, & Fujii H (2006) Proteome analysis of silk gland proteins from the silkworm, *Bombyx mori*. *Proteomics* **6**: 2586-2599

Zhao P, Li J, Wang Y, & Jiang H (2007a) Broad-spectrum antimicrobial activity of the reactive compounds generated *in vitro* by *Manduca sexta* phenoloxidase. *Insect Biochem Mol Biol* **37**: 952-959

Zhao P, Xia Q, Li J, Fujii H, Banno Y, & Xiang Z (2007b) Purification, characterization and cloning of a chymotrypsin inhibitor (CI-9) from the hemolymph of the silkworm, *Bombyx mori*. *Protein J* **26**: 349-357

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

Zou Z, Evans JD, Lu Z, Zhao P, Williams M, Sumathipala N, Hetru C, Hultmark D, & Jiang H (2007) Comparative genomic analysis of the *Tribolium* immune system. *Genome Biol* **8**: R177

Zou Z & Jiang H (2005) *Manduca sexta* serpin-6 regulates immune serine proteinases PAP-3 and HP8. cDNA cloning, protein expression, inhibition kinetics, and function elucidation. *J Biol Chem* **280**: 14341-14348

Zou Z, Najar F, Wang Y, Roe B, & Jiang H (in press) Pyrosequence analysis of expressed sequence tags for *Manduca sexta* hemolymph proteins involved in immune responses. *Insect Biochem Mol Biol*

CHAPTER 2 - High molecular weight complexes from gel filtration of *Manduca sexta* plasma

Introduction

Insects have many systems that work together to limit the spread of bacteria and other pathogens within their bodies. *Manduca sexta*, like other insects, has pattern recognition proteins that can bind to the foreign surfaces of bacteria or other pathogens (Yu et al, 2002). Some hemocytes, such as granular cells in *Pseudoplusia includens,* can also recognize certain foreign surfaces in the absence of other plasma proteins (Lavine & Strand, 2001). Recognition is followed by immune responses, which can include cellular responses like phagocytosis, encapsulation, and nodulation, as well as humoral responses like phenoloxidase activation and melanin deposition, production of reactive oxygen species, and the synthesis of antimicrobial peptides (Gillespie et al, 1997, Kanost et al, 2004, Lavine & Strand, 2002, Nappi & Christensen, 2005, Stanley & Miller, 2008, Strand, 2008).

The insect also needs to close and heal any wounds that may have accompanied infection (Theopold et al, 2004). Insects have open circulatory systems and therefore need strategies to both contain pathogens and to kill the pathogens as selectively as possible to minimize damage to the insect's own tissues. To contain a microbial infection, insects must halt the spread of the pathogens by sequestering or killing the pathogens. It seems likely that clotting plays a role in surrounding pathogens and providing a scaffold or framework for an immune response. In *Periplaneta americana* clotting localizes immune effectors, reduces spread of pathogens through the hemocoel, and is greater when the wound is accompanied with non-self molecules (e.g., LPS) (Haine et al, 2007).

We are interested in the protein complexes that form during immune response and how protein-protein interactions target the immune responses specifically toward the pathogen. Proteins already present in the hemolymph or quickly released by hemocytes

are available in the first hour of the immune response, before any induced synthesis of antimicrobial peptides or other proteins. Thus, we are interested in proteins of the early phase of immune response. These proteins include pattern recognition proteins and presumably other proteins involved in triggering the effecter responses mentioned above.

A high molecular weight complex of proteins was identified in *M. sexta* plasma collected under "non-sterile" conditions and separated by gel filtration chromatography (Beck et al, 1996). This fraction included phenoloxidase and could represent a protein complex involved in localizing phenoloxidase and thereby the melanization response. We repeated this work and further characterized similarities and differences in the "high molecular weight" fractions (HMW fractions) from gel filtration of immune-challenged and naïve plasma.

Materials and Methods

Insects

We originally obtained eggs for the *M. sexta* colony from Carolina Biological Supply. The larvae were reared on artificial diet (Dunn & Drake, 1983).

Preparation of hemolymph samples

For each treatment at least 22 day 1-3 fifth instar larvae were chilled on ice for approximately ten minutes. One of the second or third prolegs of each larva was snipped with scissors and hemolymph was collected into a sterile 15 or 50 ml centrifuge tube (Fisher). Hemolymph was then centrifuged in a swinging bucket rotor at 1750xg for 30 min to remove hemocytes. The supernatant (plasma) was mixed with diethylthiocarbamate (DETC) at 2.5 mg/ml to inhibit phenoloxidase activity. No elicitor was added to control plasma. Activated plasma contained either 0.06 mg/ml peptidoglycan (PGN) from *Micrococcus luteus* (Fluka) or 0.05 mg/ml lipopolysaccharide (LPS) from *E. coli* (Sigma). The control or activated plasma was incubated with gentle shaking at room temperature for two hours and then centrifuged at 10,000xg for 15-20 min prior to loading 16 ml of plasma on a gel filtration column.

Gel filtration chromatography

Gel filtration chromatography was performed using a column (100 x 2.5 cm) packed at 2.45 ml/min with 500 ml of Sepharose CL-4B (Amersham) in 50 mM sodium phosphate buffer, pH 6.8. The column was equilibrated at 2 ml/min with at least 500 ml of 50 mM sodium phosphate pH 6.8 with 2.5 mM reduced glutathione prior to loading a plasma sample. The column was eluted with 50 mM sodium phosphate pH 6.8 and 5 ml fractions were collected for at least 110 fractions. Two molecular weight standards, β-amylase from sweet potato (Sigma) and carbonic anhydrase from bovine erythrocyte (Sigma), were run on the gel filtration column.

Protein concentration of column fractions

The absorbance at 280 nm of the gel filtration fractions was measured in a Hitachi U-2000 spectrophotometer.

Protein concentration in the column fractions was analyzed using a Micro BCA Protein Assay Reagent Kit (Pierce), with bovine serum albumin (Pierce) as a standard. The assays were performed in a plate reader according to instructions from the manufacturer.

Phenoloxidase activity assay

A dopamine oxidation assay was used to test fractions for phenoloxidase and prophenoloxidase. 5 μl of each tested fraction were mixed with 200 μl of 2 mM dopamine in 50 mM sodium phosphate buffer, pH 6.5. The absorbance at 470 nm was monitored every 20 s for five min in a plate reader. The maximum slope, in milliOD/min, was recorded. One unit of PO activity was defined as ΔA_{470}/min = 0.001. To assay for the presence of prophenoloxidase, the same reaction was run with the addition of 2 μl of 10% (w/v) cetylpyridinium chloride (CPC). This detergent is known to activate prophenoloxidase into phenoloxidase (Hall et al, 1995).

SDS-PAGE

Protein samples were mixed with 2x SDS sample buffer (126 mM Tris-HCl ph 6.8, 20% glycerol, 4% SDS w/v, 2% β-mercaptoethanol v/v, 0.02% bromophenol blue w/v) and heated at 95°C for 5 min. SDS-PAGE was performed using 7% acrylamide

Tris-acetate gels in XT-tricine buffer (BioRad) or 4-12% acrylamide Bis-Tris NuPAGE gels in MOPS buffer (Invitrogen). A protein standard, either unstained Mark12 or pre-stained SeeBlue (Invitrogen), was run on each gel.

Gel staining

For Coomassie blue staining, gels were washed three times, at least five min each, in water and then incubated in BioSafe Coomassie G-250 stain (BioRad) for at least 1 h or overnight. Destaining was done with at least three changes of water.

For silver staining, gels were incubated in 50 ml of fixing solution (40% methanol, 0.05% formaldehyde) for at least 20 min on a shaker. Gels were then washed twice for five min each in distilled deionized water (ddH_2O). Fifty ml of 0.02% sodium thiosulfate was added and after 1 min the gels were washed twice in ddH_2O for 20 s. Next the gel was incubated in 0.1% silver nitrate for 10 min. After three 20 s ddH_2O washes, developing solution (3% sodium carbonate and 0.05% formaldehyde) was added and gels developed until the bands were apparent. Development was stopped by adding 5 ml of 2.3 M citric acid. Gels were incubated in 20% ethanol and 5% glycerol for at least 30 min before drying between wet cellophane sheets overnight.

Immunoblotting

Gels for immunoblotting were incubated briefly in western transfer buffer (20% methanol, 48 mM Tris, 40 mM glycine, and 1.3 mM SDS). Proteins were transferred from the gels to 0.45 μm nitrocellulose membranes with a Trans-Blot Semi-Dry Transfer Cell (BioRad) at 10-12 V for one h. The membranes were blocked with 3% dry milk in TTBS (0.05% Tween 20 in Tris-buffered saline, 25 mM Tris-HCl, pH 7.4, 137 mM NaCl, 2.7 mM KCl) for 30 min, after which a 1:2000 dilution of rabbit antiserum in 3% dry milk in TTBS was added and incubated for an additional two h. The different rabbit antisera used were each made to one of the following *M. sexta* proteins: serine proteinase-homolog-1 (SPH-1), serine proteinase-homolog-2 (SPH-2), β-glucan recognition protein-2 (βGRP2), lipophorin 1 and 2, hemocytin, and both prophenoloxidase-1 and -2 (PPO1&2). The membranes were washed three times with TTBS and then incubated for 1 h with a 1:3000 dilution of secondary antibody (goat-anti-rabbit IgG conjugated with alkaline phosphatase, BioRad) in TTBS with 1% dry milk.

After two washes with TTBS and one wash with TBS, the membrane was developed using an Alkaline Phosphatase Conjugate Substrate Kit (BioRad).

Protein digestion

The Nevada Proteomics Center at the University of Nevada at Reno analyzed selected protein bands after digestion with trypsin. Digestion followed a previously described protocol with some modifications (Rosenfeld et al, 1992). Samples were washed twice with 25 mM ammonium bicarbonate (pH 7.8) and 100% acetonitrile, reduced and alkylated using 10 mM dithiothreitol and 100 mM iodoacetamide, and incubated with trypsin.

Mass spectrometry

For MALDI TOF/TOF analysis, samples were spotted onto a MALDI target with a ZipTipμ-C18 (Millipore Corp., MA) and eluted with 70% acetonitrile, 0.2% formic acid and overlaid with 0.5 μl 5 mg/ml MALDI matrix (α-Cyano-4 hydroxycinnamic acid, 10 mM ammonium phosphate). Mass spectrometric data were collected using an Applied Biosystems 4700 Proteomics Analyzer MALDI TOF/TOF mass spectrometer with their 4000 Series Explorer software v. 3.6. The peptide masses were acquired in reflectron positive mode (1-keV accelerating voltage) from a mass range of 700-4,000 Da and averaging 1250 laser shots per mass spectrum. Each trypsin sample was internally calibrated on trypsin's autolysis peaks 842.51 and 2211.10 to within 20 ppm. Any trypsin sample failing to internally calibrate were analyzed under default plate calibration conditions of 150 ppm. Raw spectrum filtering/peak detection settings were S/N threshold of 3, and cluster area S/N optimization enabled at S/N threshold 10, baseline subtraction enabled at peak width 50. The twelve most intense ions from the MS analysis were subjected to MS/MS. For MS/MS analysis the mass range was 70 to precursor ion with a precursor window resolution of -1 to +4 Da, and average of 2500 laser shots for each spectrum, CID on, and metastable suppressor on. Raw spectrum filtering/peak detection settings were S/N threshold 5, cluster area S/N optimization at S/N threshold 6, and baseline subtraction enabled at peak width 50.

MALDI data analysis

MALDI data analysis peak lists were created using ABI's 4000 Series Explorer software v. 3.6 Peaks to MASCOT feature. MS peak filtering settings included mass range 700-4000 Da, minimum S/N filter 10, and peak density filter of 50 peaks per 200 Da with a maximum number of peaks set to 200. MSMS peak filtering included mass range of 60 Da to 20 Da below each precursor mass, minimum S/N filter 10, peak density filter of 50 peaks per 200 Da, and cluster area filter with maximum number of peaks 200. The filtered data were searched by Mascot v 2.1.03 (Matrix Science) using NCBI nr database (NCBI 20070725). Searches were performed without restriction to protein species, M_r, or pI and with variable oxidation of methionine residues and carbamidomethylation of cysteines. Maximum missed cleavage was set to 1 and limited to trypsin cleavage sites. Precursor mass tolerance and fragment mass tolerance were set to 20 ppm and ± 0.2 Da if internal calibration with trypsin peaks occurred; otherwise, the settings were 150 ppm and ± 0.2 Da. Specific bands were also searched with the full length *M. sexta* hemocytin protein sequence (Scholz, 2002).

The MALDI-TOF data from trypsin digestion was analyzed by using Aldente version 18/09/2007 (www.expasy.org) and searching Swiss-Prot and TrEMBL databases (releases of 13-Nov-2007). These searches included insect proteins other than *Drosophila melanogaster* ("other Insecta" selection) with mass between 0 and 300,000 Da, allowing 1 missed cleavage, variable methoinine oxidation and fixed carbamidomethylation of cysteines. Top protein hits after peptide mass fingerprinting were validated and all of the unmatched peaks searched again. If another significant hit occurred it was recorded and, in a second round of validation, the remaining peaks were searched again.

Results

We hypothesized that plasma protein complexes target PO activation to sites of wounding or infection. A previous report observed phenoloxidase in high molecular weight fractions identified by gel filtration chromatography of *M. sexta* plasma (Beck et al, 1996). To further investigate this phenomenon we performed gel filtration chromatography of plasma proteins using a Sepharose-CL4B column. A peak at

fractions 30-38 corresponded to proteins eluting in the void volume of the column, and we refer to these as "high molecular weight (HMW) fractions." A large peak from fractions 60-80 plus a smaller shoulder from fractions 80-90 represent the majority of the plasma proteins. A 200 kDa standard, β-amylase, had its peak elution in fraction 73, and the 29 kDa standard, carbonic anhydrase, peaked in fraction 86 (data not shown). The very large peak, which starts at fraction 60, corresponds with the elution of hexameric storage proteins that accumulate during the last larval feeding stages and that are around 450 kDa in their hexameric form (Ryan et al, 1985).

Increased prophenoloxidase detected in high molecular weight complexes from activated plasma

Gel filtration fractions were collected in buffer containing reduced glutathione to inhibit phenoloxidase (PO) activity during the separation. HMW fractions were tested for PO activity and, similarly, for proPO which can be activated to PO by the addition of the detergent cetylpyridinium chloride (CPC) (Hall et al, 1995). No PO activity was detected under these conditions in the absence of CPC. PO activity stimulated by CPC treatment (proPO activity) was detected in the protein fractions in the void volume as well as in a peak around fraction 80 (Fig. 2-1). ProPO specific activity in the void volume was at least fifteen times higher in plasma treated with either peptidoglycan (PGN) or lipopolysaccharide (LPS) than in control plasma. In comparison, the uncomplexed proPO activity detected around fraction 80 was at similar levels in control plasma and plasma activated with either PGN or LPS. Thus, a portion of the pool of proPO was shifted to HMW fractions after activation of plasma with PGN or LPS.

HMW fraction proteins identified by peptide mass fingerprinting

Proteins in the HMW fractions from either control plasma or plasma activated by LPS or PGN were separated by SDS-PAGE and stained with silver (Fig. 2-2) or Coomassie blue (Fig. 2-3). The overall band pattern of these three samples appeared to be quite similar. However, the silver-stained gel shows that HMW fractions from PGN-activated plasma had a strong band around 36 kDa that was fainter in LPS-treated plasma and absent in the control plasma.

To identify the protein constituents of the HMW fractions, Coomassie stained bands indicated in Fig. 2-3 were excised and analyzed by MALDI-TOF/TOF and peptide mass fingerprinting (PMF). Initially, bands over 100 kDa (L1-7) were selected from only the LPS-treated sample because this sample had more intense Coomassie staining, but the same pattern of bands was observed in the other two samples. Bands L1-7 were not identifiable by PMF regardless of whether the results were searched against NCBInr by Mascot (Table 1) or Uniprot databases by Aldente (data not shown). In a different SDS-PAGE separation of HMW fractions from PGN-treated plasma, the high molecular weight bands (P11-16) were submitted for Mascot PMF analysis (Fig 3B). Band P16 had a significant match to *B. mori* hemocytin (Kotani et al, 1995). We then searched all of the high molecular weight fraction bands against the *M. sexta* hemocytin sequence (Scholz, 2002), which has not yet been entered in the NCBI database. P16 had a much higher Mascot score for the *M. sexta* hemocytin sequence than the *B. mori* hemocytin sequence. All of the other bands with mass greater than 100 kDa, L1-7 and P11-15, also had tryptic peptides with significant match to the *M. sexta* hemocytin sequence (Table 2-1).

In addition to the bands containing hemocytin, other HMW fraction bands were analyzed by peptide mass fingerprinting using the Aldente program. The top hit returned from Aldente was validated and the remaining, unmatched tryptic fragments were searched again against the Uniprot (Swiss-Prot and TrEMBL) databases (Table 2-2). Bands at approximately 80 kDa from LPS- and PGN-treated plasma contained either proPO subunit 2 (L8), proPO subunit 1 (L9, P18) or both proteins (P17, P19). Arylphorin β was also detected in the LPS-treated plasma (L9, after validation of proPO). The control plasma also contained bands at ~80 kDa, but these bands contained arylphorins α (C21) and β (C22) but not proPO. Detection of proPO in the samples activated by PGN or LPS but not in the control is consistent with the proPO enzyme activity results shown in Fig. 2-1. Full length SPH-2 (43 kDa) was detected in LPS-activated and control plasma samples (L10, C23). The PGN-treated sample contained SPH-2 at 36 kDa, the expected size of cleaved SPH-2 (Yu et al, 2003).

HMW fraction proteins identified by immunoblot

Additional information about the proteins present in the HMW fractions was obtained by immunblot analysis. Serine proteinase-homolog-1 (SPH-1) was not detected in HMW fractions from control plasma but bands of appropriate size for full length and cleaved forms of SPH-1 were present in PGN- and LPS- treated plasma (Fig. 2-4A). Full-length serine proteinase-homolog-2 (SPH-2) was detected in control and LPS-treated plasma, while cleaved SPH-2 was detected in both PGN- and LPS- treated plasma. SPH-1 and SPH-2 together are co-factors for phenoloxidase activation when in their cleaved form (Lu & Jiang, 2008). Antiserum to beta-glucan recognition proteins 1 and 2 (βGRP-1 or -2) failed to react with the proteins in the void volumes of any of the types of plasma (data not shown and Fig. 2-4A).

A band above 191 kDa in all three plasma treatments reacted with apolipophorin-I&II antiserum (Fig. 2-4B). A band of this size is appropriate for apolipophorin-I, while the band detected above 64 kDa is the expected size for apolipophorin-II (75 kDa). The hemocytin antiserum recognized multiple bands above 200 kDa in all 3 treatments. The proPO antiserum detected proPO in LPS- and PGN-treated HMW fractions but not in the naïve HMW fractions (Fig. 2-4B). These results are consistent with the phenoloxidase activity assay, which showed much higher levels of prophenoloxidase in the void volume of plasma treated with microbial polysaccharides than in naïve plasma, and with the MS results which detected proPO in LPS- and PGN-activated plasma but not the control.

Discussion

We detected by both MS and immunoblot analyses multiple bands corresponding to hemocytin in control and immune activated plasma. The first insect hemocytin cDNA was discovered in *Bombyx mori*; it encodes a protein with 3,133 amino acids (Kotani et al, 1995). A homologous protein in *D. melanogaster*, hemolectin, was reported in 2001 (Goto et al, 2001), and the *M. sexta* hemocytin sequence was published in a PhD dissertation in 2002 (Scholz, 2002). Hemocytin is homologous to human von Willebrand factor (hvWF), a 2,813 amino acid adhesive protein important in blood clotting that is produced as a preproprotein and is cleaved to an active protein of 2,050 residues (Ruggeri, 2003). HvWF is capable of forming huge multimers, ~20,000 kDa, which are

confined to sites of tissue damage and are able to promote adhesion even under high shear stress conditions, like in arterioles (Ruggeri, 2003, Xie et al, 2001). Like hvWF, insect hemocytins contain multiple D, D', C, and B domains. However, insect hemocytins lack the A domains, which bind collagens and also function in platelet adhesion, and insect hemocytins also lack an RGD sequence, which is known for binding to integrins (Goto et al, 2001, Kotani et al, 1995, Ruggeri, 2003). While multimeric forms of hvWF, held together by disulfide bonds, are important in thrombus formation in vertebrates, only monomeric and dimeric forms have been detected in *D. melanogaster* (Goto et al, 2001).

Hemolectin was detected in media from hemocyte derived cells from *D. melanogaster* (Goto et al, 2001) and is expressed in *D. melanogaster* plasmatocytes and crystal cells (Goto et al, 2003). *D. melanogaster* plasmatocytes are phagocytes similar to lepidopteran granular hemocytes while *D. melanogaster* crystal cells are similar to lepidopteran oenocytoids: they contain proPO and lyse easily (Ribeiro & Brehelin, 2006). Hemolectin mutant larvae show more hemolymph bleeding after wounding (Goto et al, 2003). RNAi knockdown of hemolectin also led to a lack of fiber formation and much less bead aggregation than normally seen in assays of *D. melanogaster* hemolymph; the presence of proPO wasn't required for this observed bead aggregation (Scherfer et al, 2004). However, an *in vivo* study found crystal cells are essential for scab formation and effective healing in *D. melanogaster* larvae after puncture wounding (Galko & Krasnow, 2004).

The 2,230-2,321 region of *B. mori* hemocytin was proposed to be a carbohydrate recognition domain, but domain prediction tools don't identify a C-type lectin in the homologous region in *D. melanogaster*, leading to the recent suggestion that the discoidin domains (also known as coagulation factor V/VIII type C domains) are responsible for the cell agglutinating properties of the hemocytin homologs (Kotani et al, 1995, Lesch et al, 2007). Both *D. melanogaster* and *Apis mellifera* hemocytin homologs have a longer amino terminal region than the *B. mori* and *M. sexta* sequences, raising the possibility that the N-terminal region was missed during sequencing in *B. mori* and *M. sexta* (Lesch et al, 2007). The first residue of the reported *M. sexta* hemocytin sequence, Pro, is highly

unusual for the beginning of a protein sequence and thus consistent with the idea of a missing region of amino terminal sequence (Scholz, 2002).

Multiple bands corresponding to hemocytin were detected in the HMW fractions from all three treatments of plasma, ranging from ~136 kDa to a mass too large to run into the gel (>345 kDa). At least 6 or 7 bands were consistently observed and identified by MS, with the highest molecular weight bands (>200 kDa) staining the most strongly with the hemocytin antibody. Because these gels were run under reducing conditions we did not expect to observe bands corresponding to dimers. Furthermore, the monomeric mass of hemocytin (at least 352 kDa) should place it at the top of the gel. Sequence coverage from peptide mass fingerprinting is throughout the hemocytin sequence for most of the bands. Exceptions are band P16, which has the majority of its coverage in the first 750 residues, with single additional peptides at 1256 and 2910, and band P13 which starts coverage around 1137 and ends at 2640. The bands of hemocytin smaller than 345 kDa may contain multiple forms of cleaved hemocytin. Cleavage of hvWF is a mechanism for regulating multimer size, and human plasma contains hvWF subunit fragments (Ruggeri, 2003). To our knowledge, there has been no study of hemocytin cleavage in insects or its possible significance.

While hemocytin was present in HMW fractions from both control and activated plasma samples, there were two major differences in the HMW fractions from activated plasma compared with the control. HMW fractions from activated plasma had a much higher level of proPO and contained cleaved SPHs (cofactors for proPO activation). *M. sexta* SPHs are cleaved in response to immune challenge by an uncharacterized serine protease (Lu & Jiang, 2008, Yu et al, 2003). SPH-1 and SPH-2 have been previously identified as forming high molecular weight complexes during their purification from plasma (Wang & Jiang, 2004).

The presence of proPO in the high molecular weight complexes from PG- or LPS-activated plasma is correlated with the presence of cleaved SPHs. This is consistent with the finding that both proPO and PO can bind to the catalytic-like domain of SPH-1 and -2 *in vitro* (Yu et al, 2003). Our gel filtration results suggest that, in *M. sexta* plasma, cleaved SPHs associate more strongly with proPO than do full-length SPHs. After immune-challenge triggers SPH cleavage, the cleaved SPHs can then complex with

proPO in plasma, which explains the increased presence of proPO in HMW complexes after microbial elicitor stimulation. It is possible that SPHs, whether cleaved or uncleaved, associate with hemocytin and that cleaved SPHs serve as a linker between proPO and hemocytin. Hemocytin aggregates may form in plasma in response to sheer stress (*in vivo* due to bleeding from a wound or *in vitro* due to shaking on the shaker) regardless of the presence of activating microbial patterns such as LPS or PGN. These speculations will need to be tested in future experiments.

Previous gel filtration of *M. sexta* plasma collected under non-sterile conditions resulted in high molecular weight fractions that contained active PO at levels two or three times lower than the proPO present in these fractions (Beck et al, 1996). We have previously seen PO activity in the absence of CPC in high molecular weight fractions, at levels about ten times lower than proPO activity (E. Ragan, unpublished data), although no PO activity was detected in the HMW fractions examined in this study. One factor in this study that could impact PO activity was the use of reduced glutathione in the buffer in the column.

ProPO and lipophorin have been identified as components of *G. mellonella* clots by expression library screening with anti-clot antiserum (Li et al, 2002). ProPO and lipophorin were also detected on beads used to pull out clotting factors from *A. gambiae* larval hemolymph and, in *A. gambiae*, PO was required for clot formation and appeared to crosslink lipophorin (Agianian et al, 2007). Lipophorins have been implicated in coagulation for over twenty years and contain a single von Willebrand factor D-domain (Barwig, 1985, Gellissen, 1983, Theopold et al, 2004).

Hemocytin, lipophorin, and proPO, are present in the high molecular weight fractions from *M. sexta* plasma. Thus, most of the proteins in the HMW fractions share a possible role in clotting. It seems reasonable that the HMW fractions we detect in immune challenged *M. sexta* plasma are related to complexes required for locating active proPO where it is needed, at the surface of clots either at wound sites or surrounding pathogens.

References

Agianian B, Lesch C, Loseva O, & Dushay MS (2007) Preliminary characterization of hemolymph coagulation in *Anopheles gambiae* larvae. *Dev Comp Immunol* **31**: 879-888

Barwig B (1985) Isolation and characterization of plasma coagulen (PC) of the cockroach *Leucophaea madera* (Blatteria). *J. Comp. Physiol. B* **155**: 135-43

Beck G, Cardinale S, Wang L, Reiner M, & Sugumaran M (1996) Characterization of a defense complex consisting of interleukin 1 and phenol oxidase from the hemolymph of the tobacco hornworm, *Manduca sexta*. *J Biol Chem* **271**: 11035-11038

Dunn PE & Drake D (1983) Fate of bacteria injected into naive and immunized larvae of the tobacco hornworm, *Manduca sexta*. *J Invertebr Pathol* **41**: 77-85

Galko MJ & Krasnow MA (2004) Cellular and genetic analysis of wound healing in *Drosophila* larvae. *PLoS Biol* **2**: E239

Gellissen G (1983) Lipophorin as the plasma coagulen in *Locusta migratoria*. *Naturvissenschaften* **70**: 45-46

Gillespie JP, Kanost MR, & Trenczek T (1997) Biological mediators of insect immunity. *Annu Rev Entomol* **42**: 611-643

Goto A, Kadowaki T, & Kitagawa Y (2003) *Drosophila* hemolectin gene is expressed in embryonic and larval hemocytes and its knock down causes bleeding defects. *Dev Biol* **264**: 582-591

Goto A, Kumagai T, Kumagai C, Hirose J, Narita H, Mori H, Kadowaki T, Beck K, & Kitagawa Y (2001) A *Drosophila* haemocyte-specific protein, hemolectin, similar to human von Willebrand factor. *Biochem J* **359**: 99-108

Haine ER, Rolff J, & Siva-Jothy MT (2007) Functional consequences of blood clotting in insects. *Dev Comp Immunol* **31**: 456-464

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

Kanost MR, Jiang H, & Yu XQ (2004) Innate immune responses of a lepidopteran insect, *Manduca sexta*. *Immunol Rev* **198**: 97-105

Kotani E, Yamakawa M, Iwamoto S, Tashiro M, Mori H, Sumida M, Matsubara F, Taniai K, Kadono-Okuda K, & Kato Y (1995) Cloning and expression of the gene of hemocytin, an insect humoral lectin which is homologous with the mammalian von Willebrand factor. *Biochim Biophys Acta* **1260**: 245-258

Lavine MD & Strand MR (2002) Insect hemocytes and their role in immunity. *Insect Biochem Mol Biol* **32**: 1295-1309

Lavine MD & Strand MR (2001) Surface characteristics of foreign targets that elicit an encapsulation response by the moth *Pseudoplusia includens*. *J Insect Physiol* **47**: 965-974

Lesch C, Goto A, Lindgren M, Bidla G, Dushay MS, & Theopold U (2007) A role for hemolectin in coagulation and immunity in *Drosophila melanogaster*. *Dev Comp Immunol*

Li D, Scherfer C, Korayem AM, Zhao Z, Schmidt O, & Theopold U (2002) Insect hemolymph clotting: evidence for interaction between the coagulation system and the prophenoloxidase activating cascade. *Insect Biochem Mol Biol* **32**: 919-928

Lu Z & Jiang H (2008) Expression of *Manduca sexta* serine proteinase homolog precursors in insect cells and their proteolytic activation. *Insect Biochem Mol Biol* **38**: 89-98

Nappi AJ & Christensen BM (2005) Melanogenesis and associated cytotoxic reactions: applications to insect innate immunity. *Insect Biochem Mol Biol* **35**: 443-459

Ribeiro C & Brehelin M (2006) Insect haemocytes: what type of cell is that? *J Insect Physiol* **52**: 417-429

Rosenfeld J, Capdevielle J, Guillemot JC, & Ferrara P (1992) In-gel digestion of proteins for internal sequence analysis after one- or two-dimensional gel electrophoresis. *Anal Biochem* **203**: 173-179

Ruggeri ZM (2003) Von Willebrand factor. *Curr Opin Hematol* **10**: 142-149

Ryan RO, Anderson DR, Grimes WJ, & Law JH (1985) Arylphorin from *Manduca sexta*: carbohydrate structure and immunological studies. *Arch Biochem Biophys* **243**: 115-124

Scherfer C, Karlsson C, Loseva O, Bidla G, Goto A, Havemann J, Dushay MS, & Theopold U (2004) Isolation and characterization of hemolymph clotting factors in *Drosophila melanogaster* by a pullout method. *Curr Biol* **14**: 625-629

Scholz F (2002) Aktivierung von Hamozyten des Tabakschwarmers *Manduca sexta* nach bakteriellen Infektionen.

Stanley DW & Miller JS (2008) Eicosanoid actions in insect immunology. In *Insect Immunity,* Beckage NE (ed) pp 49-68. Elsevier

Strand MR (2008) Insect hemocytes and their role in immunity. In *Insect Immunity,* Beckage NE (ed) pp 25-47. Elsevier

Theopold U, Schmidt O, Soderhall K, & Dushay MS (2004) Coagulation in arthropods: defence, wound closure and healing. *Trends Immunol* **25**: 289-294

Wang Y & 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

Xie L, Chesterman CN, & Hogg PJ (2001) Control of von Willebrand factor multimer size by thrombospondin-1. *J Exp Med* **193**: 1341-1349

Yu XQ, Jiang H, Wang Y, & Kanost MR (2003) Nonproteolytic serine proteinase homologs are involved in prophenoloxidase activation in the tobacco hornworm, *Manduca sexta*. *Insect Biochem Mol Biol* **33**: 197-208

Yu XQ, Zhu YF, Ma C, Fabrick JA, & Kanost MR (2002) Pattern recognition proteins in *Manduca sexta* plasma. *Insect Biochem Mol Biol* **32**: 1287-1293

Tables

Table 2-1. Peptide mass fingerprinting analysis of bands 1-7, from LPS-treated plasma, and bands 11-16 from PGN-treated plasma.

Band	Obs.MW (kDa)	Mascot results, NCBI	Mascot Protein Score NCBI	Mascot Results, Hemocytin sequence	Mascot Protein Score Hemocytin	Mascot Protein C.I.%
L1	222	n/d	0	Hemocytin	87	100
L2	218	n/d	0	Hemocytin	150	100
L3	211	n/d	0	Hemocytin	159	100
L4	201	n/d	0	Hemocytin	96	100
L5	191	n/d	0	Hemocytin	25	99.712
L6	188	n/d	0	Hemocytin	172	100
L7	152	n/d	0	Hemocytin	15	97.049
P11	>345	n/d	0	Hemocytin	212	100
P12	253	n/d	0	Hemocytin	63	100
P13	220	n/d	0	Hemocytin	123	100
P14	200	n/d	0	Hemocytin	76	100
P15	180	n/d	0	Hemocytin	115	100
P16	136	Hemocytin (*B. mori*)	66	Hemocytin	270	100

A Mascot search was performed against both the NCBI database and the *M. sexta* hemocytin sequence, which is not present in the NCBI database. C.I. is the confidence interval.

Table 2-2. Aldente PMF search of bands L8-10, P17-20, and N21-23.

Band	Score*	Hits/Total	Coverage %	Accession #	Description	Observed mass	Calculated Mass
L8	160.32	31/203	52	Q25519	Phenoloxidase subunit 2.	83	80
L9	242.68	47/256	67	O44249	Phenoloxidase subunit 1.	81	79
L9 val.	69.91	28/209	49	P14297	Arylphorin subunit beta.	81	82
L9 val. #2	41.45	22/181	39	Q25490	Apolipophorin-2.	81	75
L10	89.64	23/182	59	Q8MUL9	Serine proteinase-like protein 2. (SPH-2)	43	43
P17	180.41	38/245	59	Q25519	Phenoloxidase subunit 2.	83	80
P17 val.	132.16	31/207	59	O44249	Phenoloxidase subunit 1.	83	79
P18	174.30	37/226	56	O44249	Phenoloxidase subunit 1.	80	79
P19	153.22	32/219	52	Q25519	Phenoloxidase subunit 2.	73	80
P19 val.	100.61	26/187	49	O44249	Phenoloxidase subunit 1.	73	79
P20	18.50	10/125	34	Q8MUL9	Serine proteinase-like protein 2. (SPH-2)	35	43
C21	166.09	43/245	66	P14296	Arylphorin subunit alpha.	83	82
C22	263.07	45/219	76	P14297	Arylphorin subunit beta.	81	82
C23	51.11	21/210	53	Q8MUL9	Serine proteinase-like protein 2. (SPH-2)	43	43

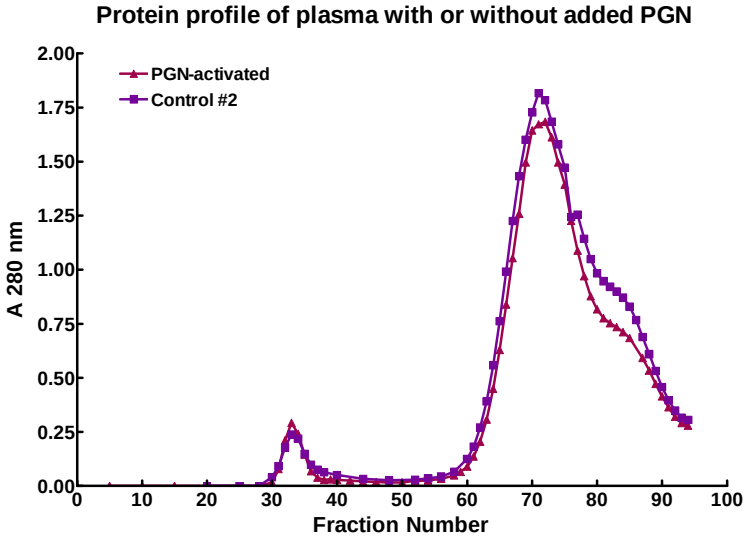
*Scores over 16.32 are statistically significant

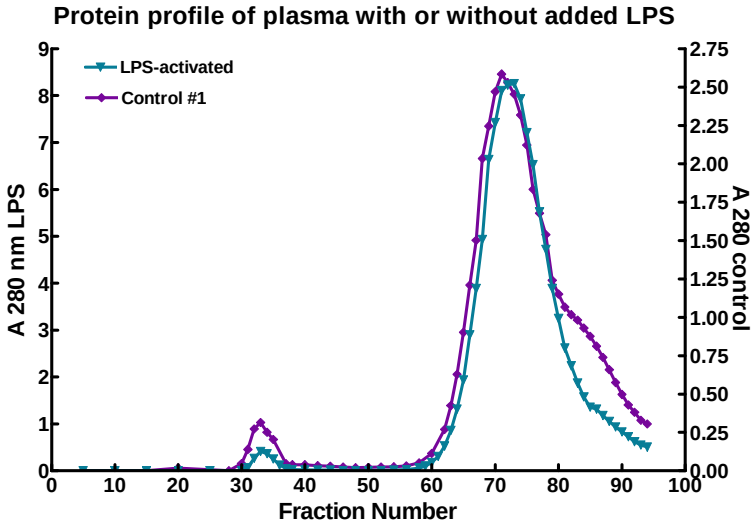
"Val." indicates protein was detected after validation of the protein listed above it.

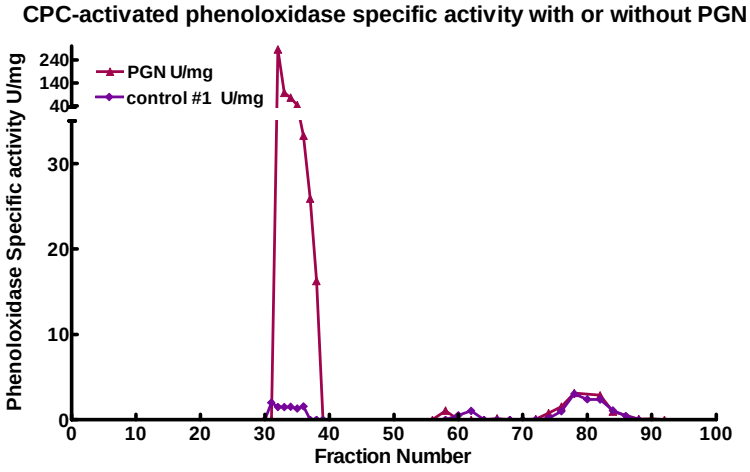
Figures

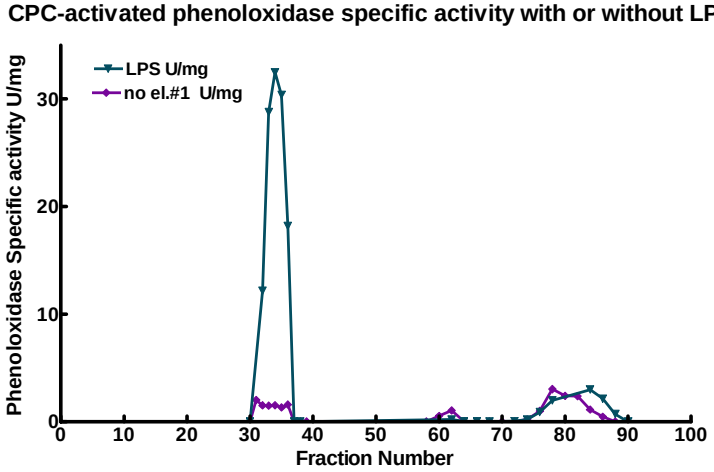
Figure 2-1. Isolation of very high molecular weight plasma protein fractions by gel filtration chromatography.

Total protein, measured by absorbance at 280 nm, of Sepharose CL-4B fractions from both PGN-activated and control plasma (A) or LPS-activated and control plasma (B). Phenoloxidase specific activity of fractions from PGN-activated and control plasma (C) or LPS-activated and control plasma (D).


Protein profile of plasma with or without added PGN
PGN-activated
Control #2
A 280 nm
Fraction Number


Protein profile of plasma with or without added LPS
LPS-activated
Control #1
A 280 nm LPS
A 280 control
Fraction Number


CPC-activated phenoloxidase specific activity with or without PGN
PGN U/mg
control #1 U/mg
Phenoloxidase Specific activity U/mg
Fraction Number


CPC-activated phenoloxidase specific activity with or without LPS
LPS U/mg
no el.#1 U/mg
Phenoloxidase Specific activity U/mg
Fraction Number

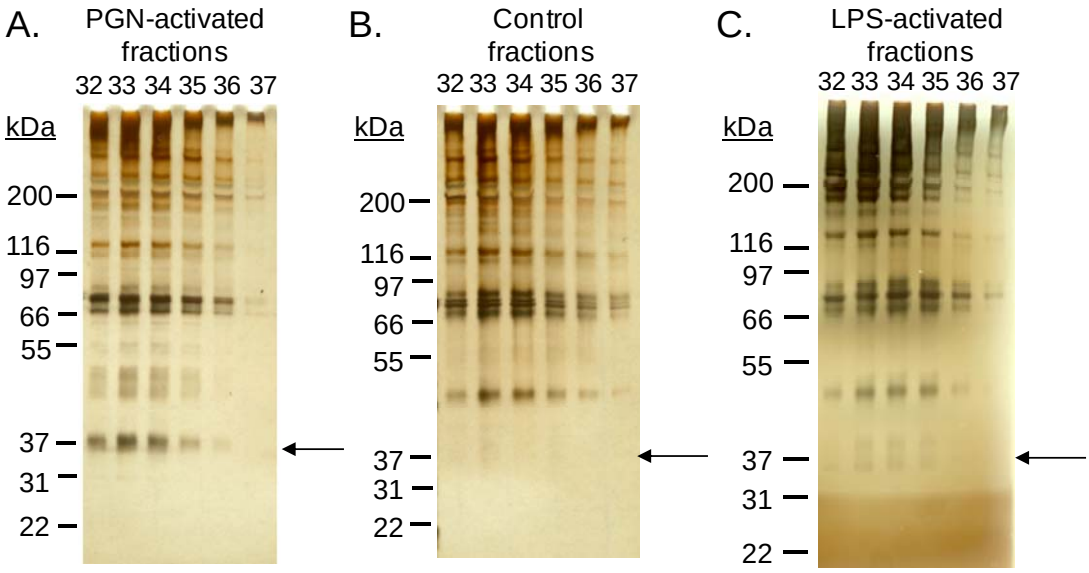


Figure 2-2. Analysis of high molecular weight fractions by SDS-PAGE.
Fractions 32-37 from gel filtration separations of plasma (Fig. 2-1) were analyzed by SDS-PAGE followed by silverstaining. A) PGN-activated plasma. B) Control plasma. C) LPS-activated plasma. Gels were silver stained. The arrows indicate a prominent band in PGN-activated fractions that is absent in control plasma fractions and very faint in LPS-activated fractions.

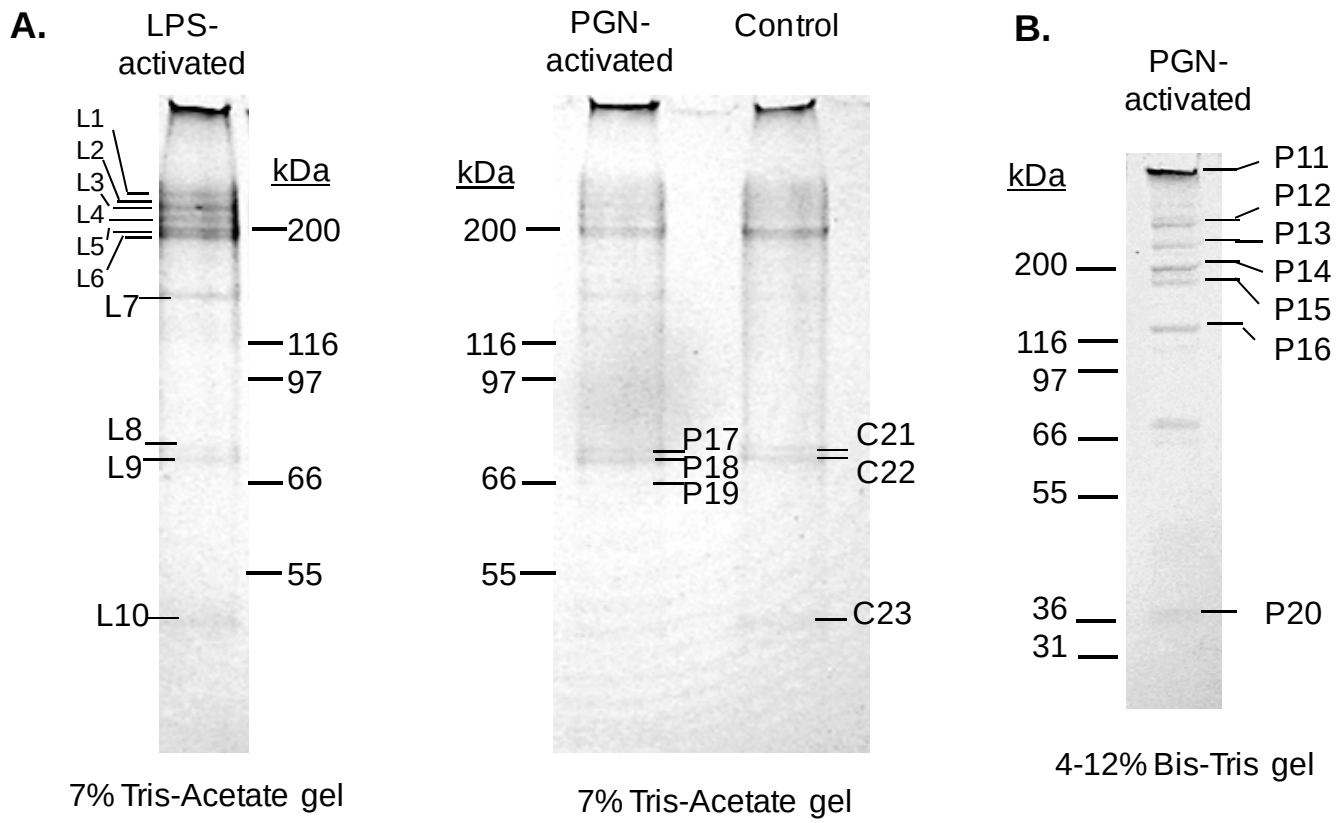


Figure 2-3. Bands from high molecular weight fractions chosen for MALDI-TOF analysis. Coomassie stained bands cut out for MS analysis are labeled. A) HMW fraction from either LPS- or PGN-activated plasma or control plasma on a 7% Tris-Acetate gel. B) HMW fraction from PGN-activated plasma on a 4-12 % Bis-Tris gel.

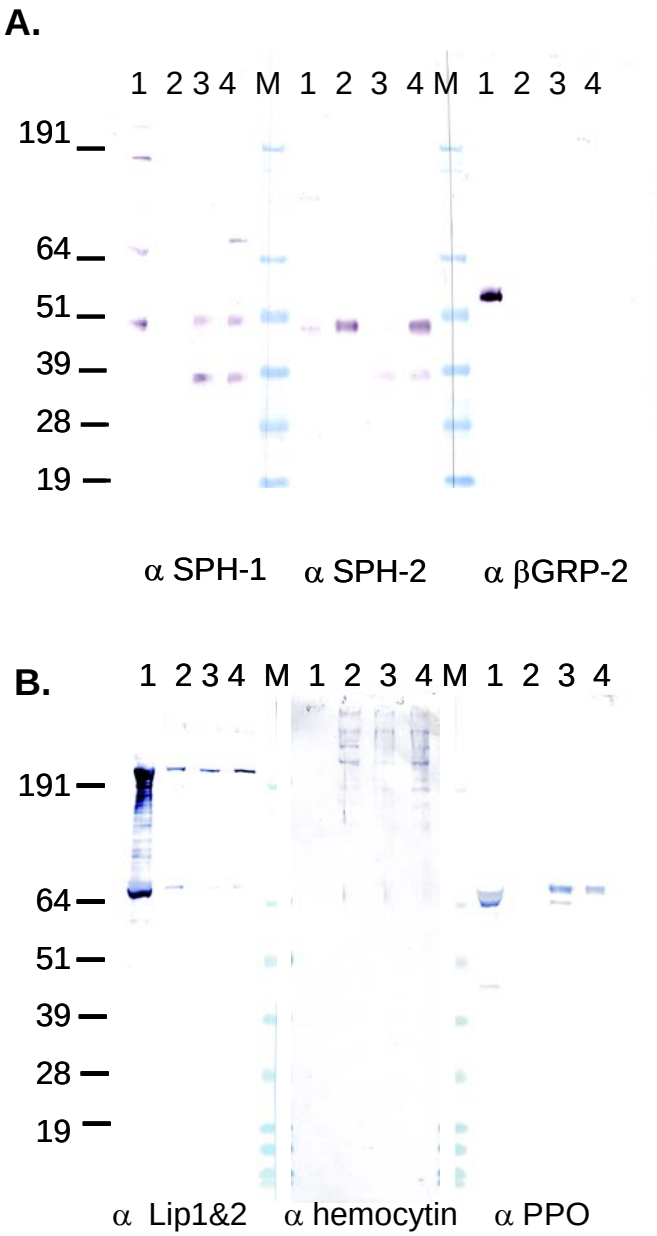


Figure 2-4. Immunoblots of high molecular weight fractions.

Control plasma prior to gel filtration (1, positive control), or HMW fractions from gel filtration of plasma from control (2), PGN-activated (3), or LPS-activated (4) larvae. M: See blue pre-stained standard. A) Antiserum to serine protease homolog-1 (SPH-1), serine protease homolog-2 (SPH-2) and beta-glucan recognition protein-2 (βGRP-2). B) Antiserum to lipophorin 1 and 2, hemocytin, and prophenoloxidase.

CHAPTER 3 - Identification of *Manduca sexta* plasma proteins that bind to bacteria or β-1,3-glucan

Introduction

Prior to mounting an immune response, an organism must recognize the invading pathogen as dangerous non-self (Janeway & Medzhitov, 2002, Matzinger, 2002, Sansonetti, 2006). Insect pattern recognition proteins have been discovered that can bind bacterial and fungal surface molecules such as β-1,3-glucans, mannan, peptidoglycan, lipopolysaccharide (LPS), and lipoteichoic acid (LTA) (Kanost et al, 2004, Yu et al, 2002). Two β-1,3-glucan recognition proteins (βGRP-1 and βGRP2) have been studied in *M. sexta*. They bind and aggregate yeast, Gram-negative bacteria, and Gram-positive bacteria and lead to an increase in prophenoloxidase (PPO) activation in the presence of laminarin (Jiang et al, 2004, Ma & Kanost, 2000). βGRPs are related to Gram-negative bacteria binding proteins (GNBPs) which can bind to LPS (Yu et al, 2002). Peptidoglycan recognition proteins (PGRPs) have been studied extensively in other insects, including *D. melanogaster* (Lemaitre & Hoffmann, 2007). The sequence for one *M. sexta* PGRP is known (Yu et al, 2002). Another pattern recognition protein, hemolin, contains four immunoglobulin (Ig)-like domains (Su et al, 1998). *M. sexta* hemolin binds to hemocytes (Ladendorff & Kanost, 1991) and to bacterial surface components, including LPS and LTA (Yu & Kanost, 2002). Four immulectins (IML-1-4) characterized in *M. sexta* are calcium-dependent (C-type) lectins, each with two carbohydrate-recognition domains (Yu et al, 1999, Yu & Kanost, 2000, Yu et al, 2005, Yu et al, 2006). IML-1 and -2 cause aggregation of bacteria (Yu et al, 1999, Yu & Kanost, 2000) and IML-2, -3, and -4 enhance encapsulation (Yu & Kanost, 2004, Yu et al, 2005, Yu et al, 2006). IML-2, -3, and -4 have all been shown to bind to several microbial polysaccharides, including LPS, LTA, and laminarin (β-1, 3-glucan), while IML-2 also binds mannose (Yu et al, 2005, Yu & Ma, 2006, Yu et al, 2006). Expression of many pattern recognition proteins increases upon immune challenge, including βGRP-

2 (Jiang et al, 2004), PGRP (Yu et al, 2002), hemolin (Yu & Kanost, 1999), and the four IMLs (Yu et al, 2002, Yu et al, 2005, Yu et al, 2006).

The connection between soluble pattern recognition proteins in plasma, which can bind to bacterial or fungal surfaces, and immune effector responses, requires signal transduction. The Toll pathway in *Drosophila melanogaster* involves soluble pattern recognition proteins that bind fungal or Gram-positive bacterial patterns and activate a serine protease cascade, which in turn activates spätzle, a cytokine. Spätzle binds to the Toll membrane receptor and triggers antimicrobial peptide production in response to fungi or Gram-positive bacteria (Lemaitre & Hoffmann, 2007). In *Manduca sexta,* an extracellular pathway of serine proteases is involved in activation of phenoloxidase and melanin production (Gorman et al, 2007). The first serine proteinase in the cascade, HP14, can combine with βGRP2 and bind to β-1,3-glucans (Wang & Jiang, 2006). HP14 has also been found to interact with peptidoglycan (Ji et al, 2004). After binding to microbial polysaccharides, HP14 autoactivates and starts the proteolytic cascade that culminates in prophenoloxidase activation.

We are interested the proteins involved in the steps from pattern recognition to an immune response. Investigators who have purified pattern recognition proteins by affinity to microbial polysaccharides have also pulled out other, often unidentified, proteins as well. For example, gram negative binding protein (GNBP) was purified from *Bombyx mori* by extracting hemolymph plasma proteins that bound to *Enterobacter clocae*. A prominent band at 50 kDa was GNBP and a band at 15 kDa was identified as lysozyme using Edman degradation. However, unidentified bands detected by SDS-PAGE were also eluted from the bacteria (Lee et al, 1996). Similarly, when *M. sexta* βGRP1 was purified from day 3 fifth instar plasma by binding to curdlan, SDS-PAGE revealed a band at 53 kDa was shown to be βGRP1, and an amino-terminal fragment of βGRP1 was detected in a band at 17 kDa. Additional bands were also visible on SDS-PAGE gels and not detected by the βGRP1 antibody (Ma & Kanost, 2000).

It is established that different bacteria, which have different surfaces, will bind different pattern recognition proteins. We also anticipate that other proteins involved in immune response, in addition to bona fide pattern recognition proteins, will bind, likely through interactions with the bound pattern recognition proteins. Thus, we hypothesized

that proteins beyond pattern recognition proteins will assemble at the surface of pathogens and may be represented in the unidentified bands seen on gels during pattern recognition protein purification. To investigate complexes of proteins that form during immune response, we used a proteomic approach to more thoroughly identify *M. sexta* hemolymph proteins that bind at the surface of bacteria.

Materials and Methods

Insects

We originally obtained eggs for the *Manduca sexta* colony from Carolina Biological Supply. The larvae are reared on artificial diet (Dunn & Drake, 1983).

Hemolymph collection for curdlan and mannan binding

Bar stage pharate pupae were chilled on ice and pinned behind the head capsule and posterior to the horn. For each experiment, five insects were cut up the dorsal midline, pinned open, and hemolymph was removed with a pipette. Approximately 0.25 ml bar stage hemolymph was collected per insect and placed immediately in 1 ml chilled anticoagulant saline (AC saline, 4 mM sodium chloride, 40 mM potassium chloride, 8 mM EDTA, 9.5 mM citric acid, 27 mM sodium citrate, 5% sucrose, 0.1 % polyvinylpyrollidone, and 1.7 mM PIPES) with 5 mg/ml of diethylthiocarbamate (DETC). The bar stage hemolymph was centrifuged at 10,000×*g* for 15 min. The supernatant was used for binding to curdlan or mannan-agarose beads.

Immune-activated larvae for the curdlan binding experiment were prepared as follows: day-1 fifth instar larvae were injected with 50 µl of 1 mg/ml *Micrococcus luteus* (Sigma). Naïve larvae were untreated. For immune-activated larvae for the *in vitro* bacteria-binding assays, day 1 fifth instar larvae were injected with 100 µl of 1 mg/ml *M. lutea*s (Sigma) while naïve larvae were left untreated. For both experiments, 24 h later larvae were bled from a clipped proleg into 15 ml or 50 ml centrifuge tubes containing DETC (amount estimated to give a final concentration of ~2.5 mg DETC/ml) on ice. At least five larvae were used for each treatment. For the curdlan binding experiment the tubes also contained 1 ml of AC saline for each 2 ml of hemolymph collected and the entire sample was centrifuged to pellet hemocytes; only the supernatant (plasma) was

used for the subsequent curdlan binding. Hemolymph samples for *in vitro* bacteria-binding were split in half and either left on ice (hemolymph) or centrifuged at 10,000×*g*, 20 min at 4°C to pellet hemocytes, and the resulting supernatant (plasma) or reserved hemolymph was immediately used for *in vitro* bacteria-binding.

Curdlan and mannan-agarose binding

Two and a half ml of bar stage plasma diluted 1:4 in AC saline or 5.5 ml of naïve or *M. luteus*-induced fifth instar plasma diluted 2:1 in AC saline was mixed with 30 mg of curdlan (from *Alcaligenes faecalis*, Sigma C7821) suspended in 350 µl of phosphate buffered saline (PBS, 0.01 M phosphate buffer, pH 7.4, 2.7 mM potassium phosphate, and 0.137 M sodium chloride). For mannan-agarose binding, 1.25 ml bar stage plasma in 6 ml of AC saline was mixed with 400 µl mannan agarose suspension (Sigma). The samples were mixed end over end for 20 min, then pelleted by centrifuging at 1,000×*g* for 5 min and washed three times with at least 3 ml of PBS. The curdlan pellets were then mixed with 500 µl of 2x SDS sample buffer. Mannan-agarose samples were mixed with 350 µl of 2x SDS sample buffer. The samples were heated at 95°C for 10 minutes. Curdlan samples were centrifuged at13,000×*g* for 3 minutes and then stored at -80°C. For mannan-agarose samples, the SDS-sample buffer after heating was stored at -20°C. Proteins were analyzed SDS-PAGE, as described in Chapter 2.

Formalin-treated E. coli

A single colony of *Escherichia coli* strain XL1Blue was grown for 22 h in 50 ml Luria broth at 37°C. The bacteria were pelleted by centrifuging at 3,000×*g* for 15 min and washed once with 0.85% NaCl (saline). Bacteria were fixed in 0.37% (w/v) formaldehyde in 0.85% saline at room temperature for two days, mixing end-over-end, and then washed three times in sterile saline and resuspended in 20 ml of sterile saline.

Acetic acid-treated bacteria

Acetic acid treatment of bacteria and *in vitro* bacterial binding assays were inspired by previously published methods (Lee et al, 1996, Sun et al, 1990). *Bacillus thuringiensis* strain HD133 (from Dr. Brenda Oppert, USDA, Manhattan, KS) was streaked on a nutrient broth plate and grown overnight at 30°C. *Serratia marcescens*

ATCC 8145 and *Staphylococcus aureus* (unnamed strain, from Dr. Jim Urban, KSU) were streaked on nutrient broth plates and grown overnight at 37°C. Single colonies were picked and grown over night in 50 ml nutrient broth at 30°C or 37°C, as appropriate. After 19 h, at an OD between 1 and 2, the culture was centrifuged and the pellet was re-suspended in 5 ml of sterile saline and 5 ml of 20% acetic acid. After 5 min at room temperature, 30 ml of 1 M Tris-HCl pH 8.2 was added, the bacteria were centrifuged, and the pellets weighed. Bacteria were resuspended based on weight, 1 ml of saline per 0.1 g of pellet.

Binding of plasma proteins to bacteria or curdlan

One ml of plasma or hemolymph was mixed end-over-end for 30 min with either 1.8 x10^9 formalin-treated *E. coli* in 150 µl or 2 mg of curdlan. After mixing, the bacteria or curdlan was pelleted by centrifuging at 13,000×*g* for 2 min. The plasma or hemolymph was removed, and the pellets were vortexed and washed three times with 1 ml of 0.1 M Tris HCl pH 8.2. Elution of proteins bound to the bacteria or curdlan was performed by adding 100 µl of 0.5 M NaCl, 0.1 M ammonium acetate pH 4, resuspending the bacteria or curdlan, and centrifuging again at 13,000×*g* for 2 min. The supernatant (elution fraction) was removed and neutralized with 12.5 µl 1M Tris-HCl pH 8. Then 17 µl of 6x SDS-sample buffer were added and the samples were heated to 95°C for 5 min prior to SDS-PAGE.

A second *in vitro* bacterial binding assay was performed with plasma or hemolymph collected from naïve or immune-activated day 2 or 3 fifth instar larvae. Aliquots of 1.6 ml of hemolymph or plasma were incubated with 0.32 ml of acetic acid-treated bacteria: *B. thuringiensis*, *S. aureus*, or *S. marcescens*. The samples were mixed end-over-end for 10 min and then centrifuged at 13,000×*g* for two min to pellet the bacteria. After vortexing and washing with 1.5 ml 0.1 M Tris-HCl pH 8.2 four times, the proteins bound to the bacteria were eluted by vortexing with 200 µl of 0.1 M ammonium acetate pH 4, 0.5 M NaCl and centrifuging again at 13,000×*g* for 2 min. The supernatant (elution fraction) was removed and mixed with 25 µl of 1M Tris-HCl pH 8.2 and 37 µl 6x SDS sample buffer and heated to 95°C for 5 min.

Proteins from the above *in vitro* bacterial binding assays were analyzed on 4-12% acrylamide Bis-Tris gels and stained with Coomassie Blue, silver, or immunoblotted. Bands visible on Coomassie blue-stained gels from immune-activated hemolymph samples were cut out and analyzed by MALDI-TOF/TOF.

Protein digestion and mass spectrometry

The Nevada Proteomics Center at the University of Nevada at Reno analyzed selected protein bands after digestion with trypsin, as described in Chapter 2.

Additionally, the filtered MALDI data were searched using Mascot against *M. sexta* ESTs, in FASTA format, and specific bands were also searched with the full length *M. sexta* hemocyte aggregation inhibitor protein sequences (Scholz, 2002) and the sequences of *M. sexta* HP1-24 catalytic domains (after cleavage at the predicted activation sites).

The MALDI-TOF data from trypsin digestion was analyzed by using Aldente as described in Chapter 2, with the exception that the mass range searched was between 0 and 100,000 Da. Top protein hits after peptide mass fingerprinting were validated, and all of the unmatched peaks were searched again. If another significant hit occurred it was recorded and, in a second round of validation, the remaining peaks were searched again.

Results

Bar-stage plasma proteins that bind curdlan

M. sexta prepupae develop a black bar on the dorsal side of their thorax starting about 12 hours before ecdysis to the pupal stage. Plasma collected from bar-stage *M. sexta* is a rich source of βGRP-2, a pattern recognition protein previously shown to bind to curdlan and to be involved in PO activation (Jiang et al, 2004). Other proteins from *M. sexta* bar-stage plasma also bind to curdlan but have not yet been identified. We hypothesized that additional plasma proteins binding at microbial surfaces could form protein complexes related to immune response. We mixed bar-stage *M. sexta* plasma with curdlan, removed the supernatant and washed the curdlan, and then removed any bound proteins by heating the curdlan in SDS reducing sample buffer. Proteins in the sample were separated by SDS-PAGE (Fig. 3-1).

Identification of curdlan-bound proteins by peptide mass fingerprinting

The twelve bands indicated in Fig. 3-1 were cut out of the gel and analyzed by trypsin digestion followed by MALDI-TOF/TOF analysis. Peptide mass fingerprinting (PMF) analysis, performed using Aldente, led to identification of at least one protein in each of the bands (Table 3-1). After validation of the top hit, additional searches led to a second protein identification in bands 5, 6, and 12, two additional identifications for bands 4 and 11, and three extra identifications in band 10. As expected, βGRP-2 was present in the most prominent band, 7, as well as bands 6, 8, 9, 10, and 11. Bands 10 and 11 additionally contained serine protease homologs 4 and 2 and band 10 also contained immulectin-2. The stained material at the bottom of the gel, band 12, also contained peptides from βGRP-2 and pattern recognition serine protease (HP14) and is presumably due to degraded forms of those proteins.

Peptides from the two highest molecular weight bands analyzed, 1 and 2, both matched apolipophorin-I. The highly abundant larval storage proteins arylphorin α and arylphorin β were identified in bands 3 and 4, respectively. Band 4 also contained apolipophorin-II and serpin-3, which was the size expected for a serpin-3-protease complex. Band 5 contained proPO-1 and proPO-2, while band 6 contained proPO-2 in addition to βGRP-2. It is likely that additional proteins were present in these bands in amounts below the MALDI-TOF detection limit and proteins in parts of the gel that weren't analyzed by MALDI-TOF remain unidentified as well.

Identification of curdlan and mannan-agarose bound proteins by immunoblotting

Using SDS-PAGE we compared bar-stage plasma proteins that bound to curdlan or mannan-agarose, and curdlan-binding plasma proteins from both naïve and bacteria-induced fifth instar larvae (Fig. 3-2). The protein profile of the fraction eluted from mannan is notably different than the elution from the β-1,3-glucan curdlan. Plasma from fifth-instar larvae has a similar pattern of proteins that bound to curdlan as bar stage plasma, although several bands were much stronger in the plasma from *M. luteus*-injected larvae as compared to plasma from the naïve larvae. We used immunoblots to obtain

additional information about bar-stage proteins bound to curdlan or mannan-agarose as well as to identify fifth-instar plasma proteins bound to curdlan.

Four proteins known to be involved in proPO activation, βGRP, prophenoloxidase activating proteinase-3 (PAP3), hemolymph proteinase 21 (HP21), and pattern recognition serine proteinase (HP14), were all detected in the curdlan bound fraction (Fig. 3-3A). Significantly, both PAP3 and HP14 were clearly in their cleaved and activated forms. PAP3 antibody also labeled a higher molecular weight band the same size as a band recognized by the serpin-3 antiserum, suggesting this ~74 kDa band is covalent complex with PAP-3 and serpin-3 (Zhu et al, 2003b). HP21 antibody recognized a band corresponding to full-length HP21 as well as lower molecular weight bands. HP21 can be active while still in its full length form (Gorman et al, 2007) and also may be active in cleaved forms (Wang & Jiang, 2007). ProPO, Iml2, and SPH-2 were also detected in the curdlan-bound fraction from bar-stage plasma (Fig. 3-3A). Immunoblot results are consistent with the detection of these proteins by MALDI-TOF.

We compared the proteins binding curdlan from fifth instar plasma with those from prepupal stage. The βGRP band was weaker in the fifth-instar samples than the bar-stage sample, which is consistent with the expression pattern of βGRP2, which normally begins to be expressed during the wandering stage and is highly expressed in the bar stage (Jiang et al, 2004). Although βGRP2 synthesis is also induced after immune challenge in larvae (Jiang et al, 2004), immunoblots showed only slightly stronger βGRP bands in the induced fifth instar plasma as compared to the naïve larvae (Fig. 3-3B). Immunoblots do not distinguish between the two βGRP proteins because both βGRP1 and βGRP2 antibodies cross-react. Larger differences were apparent between naïve and induced plasma samples with HP21, PAP3, and HP14 antibodies. HP21 antibody detected a band at ~32 kDa in the naïve sample, possibly a degradation product. In the induced sample HP21 recognized three bands, at ~60, 55, and 40 kDa, of which the one at 55 kDa is the size of full-length HP21. The PAP3 band pattern suggests that cleaved, active PAP3 is in the curdlan bound fraction from bacteria-injected insects while only a faint band, the size of full-length proPAP3, is visible in the curdlan bound fraction from naïve insects. HP14 was only detected in the induced sample, and its size indicates it is the uncleaved form. In addition, proPO, immulectin-2, cleaved SPH-1 and SPH-2, and

lysozyme were detected in the induced curdlan-bound sample, but not in the naïve control (Fig. 3-3B).

The βGRP2 also bound to mannan, but this band was much fainter than in the sample of curdlan-bound proteins (Fig. 3-3C). In contrast, there was a much more intense band corresponding to immulectin-2 that eluted from mannan-agarose, indicating a preference for binding to mannan rather than glucan for this pattern recognition protein. Small amounts of full-length proHP21, proPAP3, proPO and cleaved SPH-1 were also detected in the mannan bound sample.

Protein binding to bacterial surfaces

We hypothesized that different plasma proteins may bind, either directly or indirectly, to different types of microbial surfaces in immune responses. To test this hypothesis, we analyzed hemolymph proteins which bound *in vitro* to four different bacterial species, the gram-negative *E. coli* and *S. marcesens*, and the gram-positive *B. thuringiensis* and *S. aureus,* and compared these to hemolymph proteins that bound curdlan. Hemolymph from fifth-instar *M. sexta* collected 24 hours after injection of 100 μg of *M. luteus* was mixed with bacteria or curdlan, which were then washed three times with a low salt buffer at pH 7. Bound proteins were then eluted with a high ionic strength ammonium acetate buffer, pH 4.

Proteins eluted from the bacteria or curdlan were separated by SDS-PAGE. A total of 52 protein bands visible by Coomassie Blue staining (Fig. 3-4) were cut out, ranging from 14 bands for hemolymph incubated with *B. thuringiensis* to four bands for *S. aureus* incubated hemolymph (Fig. 3-4). Bands were digested with trypsin, and the resulting peptides analyzed by MALDI-TOF/TOF.

We searched the MALDI-TOF peaks against the Uniprot databases (Swiss-Prot and TrEmbl) using Aldente and against the NCBInr database using Mascot. Table 3-2 summarizes the peptide mass fingerprinting results for the 52 bands analyzed by Aldente and Mascot. Each band is labeled with two letters that indicate the bacterial species or curdlan incubated with the hemolymph, followed by the band number, as listed in Fig. 3-4. Bacterial proteins found in the samples are listed in Table 3-3. Eight bands did not have a protein identified by either Mascot or Aldente protein database searches: Bt4,

Cn2, Cn10, Sm4, Sm5, Sm10, Sm12 and Sm13. Of those eight bands, three had matches after searching the *M. sexta* ESTs (Table 3-4). Five peptides from Bt4 matched an EST similar to protein disulfide-isomerase-like protein ERp57 in *B. mori*; Cn10 matched ESTs for six different proteins including actin depolymerization factor, apolipoprotein III, thiol peroxiredoxin, actin, transgelin, and immune-induced protein; Sm10 matched three ESTs for histone H2A.

The Aldente program allowed us to validate a protein hit and search the remaining peaks for additional proteins. It was common to identify multiple proteins in a 1D SDS-PAGE band from hemolymph protein samples. Eight bands had at least one protein identified by Aldente that was not detected by Mascot (Table 3-5), while 10 proteins were detected by Mascot but not by Aldente (Table 3-6), although four of those have a Mascot score below the 95% confidence interval. An advantage of Mascot is that it considered both MALDI-TOF and TOF/TOF data in making its assignments and searched the entire NCBI database; with Aldente we restricted our searches to "Other Insecta" (insect proteins other than *D. melanogaster*) and separately ran searches against the "Bacteria" protein database.

Immune-related proteins bound to bacteria or curdlan

At least one form of proPO was found in all six samples. Both proPO proteins were present in ~80 kDa bands Bt2, Sa1, and Sm1, the same bands that also contained arylphorin α and β. proPO-2 and both isoforms of arylphorin were detected in Cn1, also at ~80 kDa. Immunoblot data for Bt, Sm, and Sa showing detection of proPO at the expected size support these results (Fig. 3-5B). Bands Bt3, Cn3, and Ec1 had apparent masses of ~63 kDa and all contained PO-1. The MS results are consistent with this PO1 being cleaved at the activation site Arg51, as no peptides from residues 1-51 were detected in these 3 samples.

Cationic peptide (CP8, ABO60878) was identified, by protein database search, in the 9 kDa bands from three samples, Bt13, Cn 12, and Sa4. The Mascot scores were significant in all three cases. These scores were strengthened by peptides that were supported by MS/MS: 4 peptides for Bt13 and Cn12 and 2 peptides Sa4. The Aldente program also identified CP8 in Bt13, but CP8 in Cn 12 and Sa4 was below the level of

significance with Aldente. The stronger identification by Mascot in this case shows the advantage of including MS/MS information, especially for a short sequence which naturally has fewer peptides available to match. EST search results confirmed CP8 in those three bands and also detected CP8 in Bt14.

Lysozyme was identified at about 13 kDa in four samples, Bt12, Cn11, Ec7, and Sa3, and its presence in these samples was confirmed by immunobloting (Fig. 3-5C). In the immunoblot for Sm there was a weak band detected by lysozyme antibody at the same size, suggesting that lysozyme was present but at levels too low to be detected by MS.

M. sexta peptidoglycan recognition protein was detected at about 17 kDa in both Bt 11 and Sm8, but not in samples incubated with *E. coli* or *S. aureus*, which was confirmed by immunoblot (Fig. 3-5D).

Serine protease homologs (SPHs) were detected in Cn7, Bt7, and Bt 8 by Mascot search of NCBInr (Table 3-2, immune-related proteins). Only SPH-2 in Bt8 was clearly significant (Mascot scores over 80 in this database search are in the 95% confidence interval (CI%)). The protein database CI% for SPH-1 in Cn7 was 77%, SPH-2 in Cn7 was 73%, and the CI for SPH-2 in Bt7 was 86% (see Table 3-6). Bt8 match ESTs similar to multi-binding protein from *B. mori* (BAF03496) and SPH-2 from *M. sexta* (Table 3-4). Cn7 also had EST matches to SPH-3 (3 peptides) and SPH-2 (1 peptide). On immunblot analysis (Fig. 3-5E) the Cn7 sample has two bands detected by SPH-1 antibody (SPH-2 antibody was not used with this sample) while SPH-2 antisera recognized bands at about 36 kDa in Bt, Sm, and Sa samples. MS analysis of bar stage plasma proteins bound to curdlan and eluted with SDS-sample buffer also detected SPH-2 and -4 (Table 3-1).

HP14 was recognized in Bt9 by Mascot search of NCBInr (Table 3-2, under pattern recognition serine protease) and in both Sm7 and Bt9 by search of protease catalytic domains (Table 3-2, bottom section). Bt9 and Sm7 had apparent masses of ~28 kDa, which is consistant with the expected size for the catalytic domain of HP14 (29 kDa). Immunoblot results with HP14 antibody showed a strong band at about 42 kDa in Bt and Sm samples, which is expected for the regulatory domain, while a faint band at 28 kDa is detected in the Sm but not Bt sample. The HP14 antiserum does recognize the pro domain more strongly than the catalytic domain (M. Gorman, unpublished observation),

which could explain the faint 28 kDa band in the Sm sample and its absence in the Bt sample.

Leureptin, a pattern recognition protein with 12 leucine-rich repeats and that binds to LPS (Zhu, 2001), was only observed by mass spectrometry in the proteins eluted from *E. coli* (Ec3 and Ec4) but was detected at least faintly at the expected size of 45 kDa in immunoblots for all the samples (Fig 3-5G).

Hemolin was only detected by MS in the curdlan sample. Its presence was verified by a strong signal from immunoblotting (Fig. 3-5H).

M. sexta hemocyte aggregation inhibitor protein (HAIP) (Kanost et al, 1994) was initially detected by MS, despite the vast majority of its sequence not being in the databases, due to its high similarity with *Mamestra brassicae* imaginal disc growth factor-like protein (Table 3-2, bottom). This led us to search our MS results against the HAIP sequence deduced from its cDNA sequence (Scholz, 2002). HAIP was identified in Bt5, Cn4, Cn5, Ec2, and Sa2. HAIP antiserum detected HAIP at ~53 kDa in all five samples (Fig. 3-5J).

Other insect proteins

Annexin IX was detected by both Aldente and Mascot in bands Cn8 (28 kDa) and Sm6 (32 kDa) and by Aldente in Bt8 (32 kDa) and Ec5 (33 kDa). While there are three isoforms of annexin IX in *M. sexta*, they differ only in their C-terminal 16 residues (Lesch, 2005). All of the peptides detected by PMF were in the common region of annexin IX, so no isoform-specific designation could be made. Immunoblots with a mouse monoclonal antibody to Annexin IX (kindly provided by Dr. Tina Trenczek) detected annexin IX in all three of the samples examined Bt, Sa, and Sm, although the band in the *S. marcescans* sample was fainter than the others (Fig. 3-5K). Annexin IX lacks a signal peptide and has an expected mass of 36 kDa; it was detected on the immunoblots at ~32 kDa.

Histones were detected in 3 lower molecular weight bands from the *S. marcesans* elution. Aldente matched peptides from Sm9 to histone H2B and histone H3 sequences from other insects and peptides from Sm11 were matched to a histone H4. The EST search found Sm9 matching an EST for *M. sexta* histone H3 and Sm10 matched ESTs for histone H2A.

Three different redox proteins were observed. A peroxiredoxin similar to *D. melanogaster* Jafrac1 was detected in the 20 kDa band Bt10 by Mascot search of both NCBInr and the *M. sexta* ESTs (38047571 and contig 6551). Peroxiredoxin was also detected in Cn9 (19 kDa) and Cn10 (17 kDa) by EST search only (contig 6551). By Aldente protein database and Mascot EST searches, Sm 3 (60 kDa) matched a protein disulfide-isomerase-like protein ERp57 from *B. mori* (Q587N3, contig 6705). A protein disulfide-isomerase-like protein ERp57 was also detected by EST search in band Bt4, at ~60 kDa (contig 6705). ERp57 belongs to the protein disulfide isomerase family and contains 2 thioredoxin domains. Finally, a thioredoxin protein was detected in band Cn12 (9 kDa) by Mascot search of NCBInr and ESTs (6560635, contig 6039).

A few other *M. sexta* proteins detected include transferrin (Cn1), serpin-1 (Cn6) and insecticyanin (Cn9).

Storage proteins and abundant cytoplasmic proteins

Some highly abundant *M. sexta* plasma proteins that were eluted from bacteria do not have an obvious role in immune response. These storage proteins, arylphorin α and β and methionine-rich storage protein 3, are present at very high concentration in plasma (Willott et al, 1989). They associate with the microorganisms at relatively low amount when compared with their concentration in plasma and therefore their presence is likely due to non-specific binding.

Also present were intracellular proteins, including elongation factors 1-α (Bt5, Ml2) and 2 (Bt1), ubiquitin and ribosomal protein s27a (Cn13), actin (Ec4, Bt6, Cn6), the actin-binding proteins transgelin (Bt10) and actin depolymerization factor (Cn10). These proteins may also be artifacts. It is interesting to note that several proteomic studies in *D. melanogaster* have also detected actin and one proposed an immune role for it (Vierstraete et al, 2004b). However, hemocyte rupture during immune response or hemolymph collection may trigger actin release, making the presence of actin and other intracellular proteins a non-specific side effect (Engstrom et al, 2004).

Bacterial proteins

Bacterial proteins were observed in several of the *E. coli* incubated bands, two of the *B. thuringiensis* bands, and two of the *S. marcesans* bands. Some of these bacterial

proteins could represent degradation of the bacteria, either due to the physical process of resuspending and washing the bacterial pellets or because of *M. sexta* proteases or other proteins (like lysozyme and cecropins) which can lyse bacteria.

Discussion

Challenges in using MALDI-TOF and peptide mass fingerprinting to identify M. sexta *proteins*

I used MALDI-TOF mass spectrometry and peptide mass fingerprinting to identify *M. sexta* hemolymph proteins associated with bacteria and therefore potentially involved in immune response. Because *M. sexta* lacks a sequenced genome, there were certain challenges in identifying proteins. Fortunately, sequences of many proteins are in the database; there are 616 entries for *M. sexta* proteins in the NCBI database. Unfortunately, some *M. sexta* proteins that have been sequenced and studied, such as hemocytin and HAIP, are not in the databases. Many other proteins have yet to be sequenced or studied in *M. sexta* and are obviously absent as well. For very well-conserved proteins, such as histones, transgelin, or elongation factor 2, peptide hits from *B. mori* or other insect species allowed identification using the UniProt or NCBI databases. Sequence identity generally has to be about 70% or higher for peptide mass fingerprinting to pick up hits across species (Biron et al, 2006). Cross-species identification is also how I initially identified HAIP, based on its similarity to *M. brassicae* imaginal disc growth factor-like protein. Similarly, in Chapter 2, one hemocytin band was identified by peptide matches to *B. mori* hemocytin. For HAIP and hemocytin, where we had access to the protein sequences (Scholz, 2002), we were able to input the *M. sexta* sequences into Mascot for peptide mass fingerprinting and confirmed the identification of these proteins. Finally, we also used Mascot to search all the known *M. sexta* ESTs, the ones currently available on NCBI from the antennal library and from a subtracted fat body cDNA library (Zhu et al, 2003a), as well as over 7,000 *M. sexta* ESTs recently compiled by Dr. Haobo Jiang. This enabled us to identify additional proteins and to obtain *M. sexta* sequences for some proteins that were initially identified by a hit to another insect species.

Bar-stage plasma bound to curdlan

We expect that different pathogens, with different surface patterns, will bind different pattern recognition proteins and activate different immune responses. We saw that bar-stage plasma proteins associated with mannan-agarose differed from those associated with the β-1,3-glucan curdlan. The predominant protein bound by curdlan was βGRP2, whereas mannan bound much less βGRP and instead exhibited a very strong affinity for immulectin-2. This is consistent with previous results in which βGRP was isolated by affinity to curdlan (Ma & Kanost, 2000) and immulectin-2 was isolated by its affinity to mannan (Yu & Kanost, 2000).

We are interested in how proteins interact with pathogens to trigger immune responses. We performed *in vitro* binding of *M. sexta* bar stage plasma with curdlan to see what proteins (in addition to βGRP2) would associate strongly enough with curdlan to withstand the washing. The majority of the proteins identified as associated with curdlan are known to be involved in immune responses. Lysozyme, involved in bacterial lysis and killing, was detected by immunoblot in the curdlan bound sample from bar-stage plasma. Pattern recognition proteins bound to curdlan and identified by MALDI-TOF analysis include βGRP2 and immulectin-2. Serine protease homologs (SPH) 2 and 4 were also detected; immulectin-2 and proPO have been found to bind to activated SPH-2 (Yu et al, 2003), and activated SPHs-1 and -2 are required for activation of proPO by PAPs (Lu & Jiang, 2008). SPH-4 is a similar protein that hasn't been as extensively characterized or studied as SPHs -1 and -2. The sizes of SPH-2 and -4 detected by SDS-PAGE (both at 36 and 31 kDa) suggest they are in their cleaved form.

We detected proteases responsible for proPO activation on the surface of curdlan by immunoblot. HP14 and PAP3 were clearly in cleaved, active forms while HP21, which may be activated even in the absence of cleavage (Gorman et al, 2007, Wang & Jiang, 2007), was present in both its full length form and some lower molecular weight species. Both of the known *M. sexta* phenoloxidases, subunits 1 and 2, bound to curdlan and were activated, based on size (64-67 kDa instead of the expected 79-80 kDa). An inhibitor of proPO activation, serpin-3, was detected in the curdlan bound fraction by MALDI-TOF. Serpin-3 is an ortholog of *D. melanogaster* serpin 27A and complexes with PAP3 to inhibit PPO activation (De Gregorio et al, 2002, Zhu et al, 2003b). The

size of the detected serpin-3, ~74 kDa, suggests that it was in a complex with a protease. This size is consistent with a complex of serpin-3 with PAP3 (Zhu et al, 2003b). A band of this size was also detected by both PAP3 and serpin-3 antibodies in immunoblot analysis of this sample. Additionally, four peptide fragments from the PAP3 catalytic domain were detected by MS in band 4, where serpin-3 was also detected, although the Aldente score was not above the significance threshold. The lack of statistical significance of the PAP3 score is not surprising given the number of proteins in band 4 as well as the presence of only the PAP3 catalytic domain in the band but the complete sequence in the database. A lower sequence coverage leads to a lower protein score in Aldente.

Identification of active proteases that are involved in proPO activation in the curdlan bound fraction suggests that they may assemble together at the surface of a pathogen as a proPO activation complex. It also makes sense that an inhibitor of proPO activation such as serpin-3 would also be found in that location, as a regulator of the PO activity.

Other proteins detected bound with curdlan by MALDI-TOF and PMF included the highly abundant storage proteins, arylphorins α and β (Willott et al, 1989), and apolipophorins 1 and 2. Lipophorin has been associated with clot formation in other insects, as discussed in Chapter 2. These proteins, particularly the arylphorins, are at extremely high concentrations in the plasma and may result from a low-affinity, non-specific association with curdlan.

Binding of hemolymph proteins to bacteria

Proteins associating with bacterial surfaces are candidates for participation in early stage effector responses. To screen for such proteins, we assayed hemolymph proteins that bound *in vitro* to different types of bacteria and to curdlan. This experiment differed from the bar-stage curdlan and mannan-agarose binding experiments mentioned above in that we used hemolymph (plasma with hemocytes) from fifth instar *M. sexta* and eluted bound proteins with an ammonium acetate buffer, pH 4. SDS-sample buffer, which was used for elution with curdlan and mannan agarose, would have resulted in a higher background of bacterial proteins. It is possible that some strongly binding

proteins, such as βGRPs, might not have been eluted from the bacteria with the less stringent elution conditions and therefore were not detected.

Leureptin is a pattern recognition protein that bound to all four bacterial species and to curdlan (Table 3-7). Leureptin contains 12 leucine rich repeats that can bind LPS and associate with hemocyte membranes (Zhu, 2001, Zhu et al, 2003a). It has been proposed to have a role in LPS clearance and phagocytosis; the presence of leureptin bound to *E. coli* and other bacterial surfaces in this study supports such a role in the early stages of immune response.

A previous study showed that *H. cecropia* hemolin bound to *E. coli* (Sun et al, 1990). *M. sexta* hemolin binds to lipoteichoic acid (LTA), LPS, and zymosan, a fungal cell wall preparation (Yu & Kanost, 2002). Hemolin has also been shown to associate with the surface of hemocytes and inhibit hemocyte aggregation (Ladendorff & Kanost, 1991). I detected hemolin bound to curdlan from *Alcaligenes faecalis* in my *in vitro* binding assay.

An unexpected protein associated with *B. thuringiensis*, curdlan, and *S. aureus* is *M. sexta* CP8, which was first identified because it binds to IML-3 (Xiao Qiang Yu, personal communication). CP8 is 73% identical to *Antheraea mylitta* fungal protease inhibitor-1 (AmFPI-1), which was purified from larval hemolymph and found to inhibit a range of different serine proteases (Shrivastava & Ghosh, 2003). CP8 and AmFPI-1 are also both similar to *Galleria mellonella* inducible serine protease inhibitor-1 (ISPI-1) (Frobius et al, 2000). A recent proteomic analysis of *M. sexta* hemolymph detected a similar peptide by Edman degradation of a band at ~10 kDa from a SDS-PAGE gel. This band had 58% identity with *G. mellonella* ISPI-1 (Furusawa et al, 2008). The peptide they sequenced matches the CP8 sequence exactly, with the exception that the cysteines in CP8 were not identified by Edman degradation.

A protein that bound to curdlan and all bacterial species tested was HAIP (Table 3-7), named for its ability to prevent hemocyte aggregation during hemocyte cell culture (Kanost et al, 1994). A related protein in *D. melanogaster*, Ds47, was secreted from S2 cells (Kirkpatrick, 1995). More recently, both HAIP and Ds47 have been identified as members of a chitinase gene family "glycosyl hydrolase family 18" and also the imaginal disc growth factor-like (Idgf) family, named for related glycosyl hydrolase family 18

proteins in *D. melanogaster* that can stimulate imaginal disc proliferation (Kawamura et al, 1999, Scholz, 2002). The proteins with the highest similarity to HAIP by blast search (e value 0.0) are Idgf-like proteins from other lepidopteran insects: *Pieris rapae* (AAT36640), *Mamestra brassicae* (ABC79625), and *Bombyx mori* (BAF73623). The *M. brassicae* Idgf stimulated cell proliferation in cell lines derived from *M. brassicae* fat body and hemocytes without the need for insulin (Zhang et al, 2006).

Other proteins homologous to HAIP are *Anopheles gambiae* BR1 and BR2 (Blast result from searching with HAIP of 2e-127 & 4e-133, respectively). These proteins were identified in two ~32 kDa spots that increased in intensity in 2D gels after immune challenge with *E. coli* or *M. luteus*. It was determined that these spots belonged to two ~50 kDa proteins cleaved after exposure to bacteria. Like *M. sexta* HAIP (Kanost et al, 1994), these two proteins are constitutively expressed and the increased intensity of the 33 kDa spots was due to proteolytic processing after immune stimulation (Shi & Paskewitz, 2004). We do not have any evidence that HAIP is cleaved like its relatives in *A. gambiae*.

Large scale proteomic studies have identified *D. melanogaster* Idgfs in larval plasma or hemolymph. Ds47, Idgf-3, Idgf-4, and Idgf-5 were identified in *D. melanogaster* plasma (Karlsson et al, 2004). Idgf-2 was strongly upregulated in hemolymph after 25 minutes in response to *M. luteus* and *S. cerevisiae* (Vierstraete et al, 2004b) but was found to be suppressed after 72 hours in response to *Beauveria bassiana* while levels of IDGF-3 were unchanged (Levy et al, 2004). The study by Furusawa and coworkers of *M. sexta* hemolymph identified peptides that matched IDGF-like proteins in both 1D and 2D gel analysis (2008). All of these peptides can be matched to the *M. sexta* HAIP sequence except one, which has minor amino acid changes that result in a molecular weight change of just less than 1.0 Da, which is within the peptide mass tolerance they used for their analysis.

HAIP could have a role in pattern recognition. It bound to bacterial and curdlan samples, and it also has an effect on hemocytes. *M. sexta* HAIP has also been shown to bind to heparin and chitin (Scholz, 2002). Some mammalian glyosyl hydrolase family 18 members are expressed in phagocytic cells like macrophages and dendridic cells (Funkhouser & Aronson, 2007). Members of the glycosyl hydrolase family 18 found in

humans have been suggested to be involved in innate immunity. One example is chitotriosidase, which has an e value of 1e-42 after a blast search with the HAIP sequence, and is secreted by activated human macrophages (Boot et al, 1995). LPS and other pro-inflammatory stimuli increase chitotriosidase synthesis in macrophages (Di Rosa et al, 2005). HAIP may also be involved in preventing excessive hemocyte aggregation during immune response (Kanost et al, 2004) or response to tissue damage.

Annexin IX was also detected, either by MS or immunoblotting, in all five samples. *M. sexta* annexin IX was first discovered by screening cDNA-expression libraries from induced hemocytes and fat body with monoclonal antibodies against *M. sexta* hemocytes (Lesch, 2005). *M. sexta*-annexin IX is 33-34 kDa and has three isoforms that differ in the last 16 amino acid residues, which is very similar to the *Bombyx mori* annexin IX (Xia et al, 2001). Annexin IX was detected in hemocytes and fat body but not in plasma of *M. sexta* larvae, which is reasonable given the lack of a signal peptide in the sequence (Lesch, 2005).

Insect annexins (numbered from IX to XIII) are classified in the "B" family and lack any one-to-one orthologs with vertebrate annexins (Moss & Morgan, 2004). *B. mori* annexin IX is present at high levels in portions of the anterior silk gland that persist through metamorphosis (Kaneko et al, 2006), and *B. mori* annexin XIII is thought to be involved in wing morphogenesis (Matsunaga & Fujiwara, 2002). *D. melanogaster* annexin X binds membranes containing phosphatidylserine (Johnston et al, 1990) while *D. melanogaster* annexin XI and *A. gambiae* annexin 10B bind *Plasmodium berghei* ookinetes (Kotsyfakis et al, 2005a). Annexin antibody localizes to the extracellular cell membrane in an *A. gambiae* hemocyte-like cell line (4a-3A) (Kotsyfakis et al, 2005b). Annexin IX mRNA was upregulated 2.5-fold one and a half and three hours after septic injury in *D. melanogaster* and listed under "coagulation" due to "homology to annexin V, a vertebrate anticoagulant factor" (De Gregorio et al, 2001). Annexin IX protein was increased in the cytoplasmic fraction of *D. melanogaster* hemocyte-like mbn-2 cells after 6 h of LPS treatment (Loseva & Engström, 2004)

M. sexta PGRP-1 was only detected in samples eluted from *B. thuringiensis* and *S. marcescens* but not from *S. aureus*. Peptidoglycan, an essential component of both Gram-positive and Gram-negative bacteria cell walls, is a polymer of a repeating

disaccharide unit, β-1,4 linked N-acetylglucosamine and N-acetylmuramic acid (Kurata et al, 2006). These polymer chains are cross-linked by short peptides in which the third residue is either a lysine, as in most Gram positive bacteria, or a diaminopimelic acid (DAP) in Gram-negative bacteria and *Bacillus* species. These two types of peptidoglycan, Lys-type peptidoglycan and DAP-type peptidoglycan, can be recognized by different PGRPs, which are present in invertebrates and vertebrates (Dziarski & Gupta, 2006).

A blast search of the known *M. sexta* PGRP-1 shows it is more similar (E value = 1e-37) to the peptidoglycan-binding domain of PGRP-LC than to any of the other *D. melanogaster* PGRPs. PGRP-LC binds to DAP-type peptidoglycan but does not bind to Lys-type peptidoglycan, making it a selective recognizer of Gram-negative and *Bacillus* type bacteria (Lemaitre & Hoffmann, 2007). Some PGRPs from other insect species are even more similar to *M. sexta* PGRP-1. *Antheraea mylitta* PGRP-like protein has an E value of 2e-82 and PGRP-A from *Samia cynthia ricini* has an E value of 6e-81 when compared with the *M. sexta* PGRP-1 sequence. PGRP-A protein from *Samia cynthia ricini*, a wild silk worm, has recently been studied. Binding assays with radiolabeled peptidoglycan indicate that *S. cynthia* PGRP-A bound to DAP-type cross-linked peptidoglycan (from *Bacillus* cell wall) and to uncross-linked Lys-type peptidoglycan (from *M. luteus*) but not to cross-linked Lys-type peptidoglycan (Onoe et al, 2007). Thus, my observation that *M. sexta* PGRP binds to only *B. thuringiensis* and *S. marcescans* and not *S. aureus* is consistent with its similarities to the PGRP-LC from *D. melanogaster* and PGRP-A from *S. cynthia.* Binding to DAP-type peptidoglycan allows selective recognition of and response to Gram-negative and *Bacillus* type bacteria. There are at least two PGRP proteins in *B. mori* (AAL32058 and BAA77210) which have 38% identity to each other. Additional PGRP proteins are expected to be present in *M. sexta* and may bind different types of peptidoglycan.

M. sexta histones were eluted from *S. marcesens* (Tables 3-2 and 3-5). Histones H2A, H2B, H3 and H4 from hemocytes of the Pacific white shrimp, *Litopenaeus vannamei*, were recently found to possess antimicrobial activity (Patat et al, 2004). Histone 2A and a mix of H2B/H4 inhibited growth of *M. luteus* completely at 4.5 μM and 3 μM, respectively. A synthetic peptide from the N-terminal sequence of histone 2A

(H2A 2-39) inhibited bacterial growth by 50% at 0.5-1 μM. This paper was the first suggestion of histones as antimicrobial proteins in invertebrates, although the antimicrobial properties of vertebrate histones were first noted in 1958 (Hirsch, 1958) and now appear to be well accepted (Parseghian & Luhrs, 2006).

One notable location of histones, where antimicrobial properties are important, is in neutrophil extracellular traps (NETs). First described by Brinkmann and coworkers in 2004, human neutrophils activated with LPS or other stimulators (interleukin-8, phorbol myristate acetate) flattened and formed very narrow fibers. DNA intercalating dyes stained the NETs, and NETs reacted with antibodies against histones H1, H2a, H2B, H3 and H4. DNase treatment (but not protease treatment) caused degradation of the NETs. These NETs were also found to lead to bacterial killing, presumably due to a high local concentration of antibacterial granular contents. Specifically, histone 2A was shown *in vitro* to kill bacteria such as *S. aureus* at concentrations as low at 140 nM (2 μg/ml) in 30 minutes (Brinkmann et al, 2004).

NET formation has subsequently been found in many different vertebrates, including humans, mice, rabbits, horses, cows and fish (Brinkmann & Zychlinsky, 2007). The mechanism of NET formation is a type of programmed cell death, different than apoptosis, and seems to be induced by reactive oxygen species. NET formation involves activation of neutrophils and subsequent reactive oxygen species production. Next, the nuclear membrane disintegrates and then nuclear material mixes with granule contents. Finally, the NETs are released from the cell and the cell dies. This process takes 45-240 minutes after activation (Fuchs et al, 2007). NETs can bind Gram-positive and Gram-negative bacteria and fungi, killing them even in the absence of phagocytosis. Most bacteria are susceptible to being killed by histone 2A so it is expected that the presence of histones in the NETs contributes to microbial killing, although the relative contribution of histones compared with other constituents, such as antimicrobial peptides, is not known (Brinkmann & Zychlinsky, 2007).

While NETs have not been detected in invertebrates, the histones bound to *S. marcesens* might be biologically relevant because of potential antibacterial properties. Their correlation with NET formation in vertebrates is also intriguing. Certainly NET

types of webs to contain and kill invading bacteria could be particularly valuable in the open circulatory systems of insects.

EST search of band Cn10 bound to curdlan identified six proteins, including apolipophorin III and immune-induced protein-1. Immune induced protein-1 is similar to noduler from *Antheraea mylitta*, which binds to microorganisms and hemocytes and is important for the noduation response (Gandhe et al, 2006). There are many reports suggesting a role for apolipophorin III in immune response (reviewed in (Schmidt et al, 2008)). This is further complicated by the fact that apolipophorin III and immune induced protein-1 are the same size and hard to separate from each other (Kanost et al, 1988). The other proteins detected in this band were actin depolymerization factor, actin, transgelin, and thiol peroxiredoxin. This is a large number of proteins to detect in one band by MS analysis. *A. mylitta* noduler binds to a variety of bacteria and yeast as well as to curdlan and *A. mylitta* hemocytes during *in vitro* binding assays (Gandhe et al, 2006). Knock down of noduler by RNAi (confirmed by immunblotting) led to a higher load of both *S. aureus* (10x) and *E. coli* (4-5x) in hemolymph after 30 min and a 3-5 fold decrease in nodules formed in response to *S. aureus* or *E. coli* but no significant change in phenoloxidase activity or phagocytosis activity. Noduler is concentrated in nodules, as shown by immunoblotting, and proposed to be a "facilitator of nodulation" (Gandhe et al, 2006).

Three different redox proteins were observed in six different bands eluted from curdlan, *S. marcesens*, or *B. thuringiensis*: a peroxiredoxin at 20 kDa in bands Bt10, Cn9, and Cn10, protein disulfide- isomerase-like protein ERp57 at 60 kDa in Sm3 and Bt4, and a thioredoxin protein in the 9 kDa band Cn12. These proteins have been found in other proteomic studies and may be involved in protecting the host insect from the reactive oxygen species generated during immune response.

Peroxiredoxins reduce hydrogen peroxide, peroxynitrite, and organic hydroperoxides and are reactivated after reduction by thioredoxin or another electron donor (Wood et al, 2003). Proteomic studies found *D. melanogaster* peroxiredoxin (Jafrac1) was present in higher amounts in hemolymph 25 min after *M. luteus* or *S. cerevisiae* challenge and 4 h after LPS challenge (Vierstraete et al, 2004a, Vierstraete et al, 2004b).

Protein disulfide-isomerases are known for helping proper protein folding in the endoplasmic reticulum. Additionally, in mammals, they have been found associated with the surface of cells, to be secreted from activated platelets, and to have a role in platelet aggregation (Manickam et al, 2008, Jordan et al, 2005). Also, one isoform of protein disulfide isomerase was upregulated in the nuclear fraction of *D. melanogaster* hemocyte-like mbn-2 cells after 30 min of LPS treatment (Loseva & Engstrom, 2004).

Thioredoxin can reduce hydrogen peroxide and reduce peroxiredoxins (information from NCBI conserved domain database, cd02947). Human thioredoxin-1 can be secreted from cells, including activated lymphocytes (Rubartelli et al, 1992), and has been found to be chemotactic, important for recruiting cells to sites of infection (Bertini et al, 1999). Recently human thioredoxin-1 was shown to specifically bind to tumor necrosis factor receptor superfamily member 8 (CD30) on the surface of lymphocytes (Schwertassek et al, 2007). *B. mori* thioredoxin is up-regulated in fatbody in response to injection of microorganisms or extreme temperatures (Kim et al, 2007).

Based on these other systems, it seems reasonable that protein disulfide-isomerase, peroxiredoxin, and thioredoxin may be released from hemocytes and play a role in the immune response. Redox molecules represent an exciting new area of signal transduction (Jordan and Gibbins, 2006) and provide a way early immune responses such as phenoloxidase activation that generate reactive oxygen species could trigger other downstream cellular responses. *B. mori* paralytic peptide, a cytokine involved in hemocyte response, is activated by serine proteases after reactive oxygen species are generated (Ishii et al, 2008).

Binding of hemolymph proteins to microbial surfaces involves the early stage (30 min or less) of the immune response. Over twenty five years ago Dunn and Drake observed significant decreases in circulating bacteria one h after injection into *M. sexta.* This decrease was not dependent on previous immune challenge of the insects (Dunn & Drake, 1983). Bacteria were able to grow in cell-free hemolymph, even in the presence of melanization which occurred one or two hrs after addition of bacteria. Also, total hemocyte counts were observed to increase starting at 30 min and peak at 1 h after bacteria injection (Horohov & Dunn, 1982). Finally, light microscopy investigations showed that hemocytes clump together minutes after injection and that larger clumps

were seen at 30 min. The initial decrease in circulating bacteria appeared to be due to nodulation, with phagocytosis occurring slightly later, from 2-8 h (Horohov & Dunn, 1983). Thus, very early stage immune responses are known to involve interactions between hemocytes and bacteria.

The early stage of nodulation response, which is also called microaggregation of hemocytes, occurs quickly, within the first 30 minutes, and is mediated by eicosanoids (Miller & Stanley, 2001). Eicosanoids have also been recently shown to be involved in hemocyte migration towards an *E. coli* chemotactic peptide (Merchant et al, 2008) and eicosanoids stimulate release of proPO from oenocytoids (Shrestha & Kim, 2008).

Our study suggests possible other factors that could be involved in early stage immune response signal transduction. They include HAIP, which could affect both hemocyte aggregation and proliferation, and thioredoxin and peroxidase which have the potential to mediate signaling in response to reactive oxygen species. I believe early stage immune response is a complex interplay between factors in the plasma and hemocytes, challenging us to bring together two aspects of insect immunity humoral and cellular responses, which are often thought of separately.

Acknowledgements

I would like to thank Dr. Kristin Michel for translating portions of Dr. Lesch's and Dr. Scholz's PhD dissertations from German and Dr. Tina Trenczek for the annexin IX antibody.

References

Bertini R, Howard OM, Dong HF, Oppenheim JJ, Bizzarri C, Sergi R, Caselli G, Pagliei S, Romines B, Wilshire JA, Mengozzi M, Nakamura H, Yodoi J, Pekkari K, Gurunath R, Holmgren A, Herzenberg LA, Herzenberg LA, & Ghezzi P (1999) Thioredoxin, a redox enzyme released in infection and inflammation, is a unique chemoattractant for neutrophils, monocytes, and T cells. *J Exp Med* **189**: 1783-1789

Biron DG, Brun C, Lefevre T, Lebarbenchon C, Loxdale HD, Chevenet F, Brizard JP, & Thomas F (2006) The pitfalls of proteomics experiments without the correct use of bioinformatics tools. *Proteomics* **6**: 5577-5596

Boot RG, Renkema GH, Strijland A, van Zonneveld AJ, & Aerts JM (1995) Cloning of a cDNA encoding chitotriosidase, a human chitinase produced by macrophages. *J Biol Chem* **270**: 26252-26256

Brinkmann V, Reichard U, Goosmann C, Fauler B, Uhlemann Y, Weiss DS, Weinrauch Y, & Zychlinsky A (2004) Neutrophil extracellular traps kill bacteria. *Science* **303**: 1532-1535

Brinkmann V & Zychlinsky A (2007) Beneficial suicide: why neutrophils die to make NETs. *Nat Rev Microbiol* **5**: 577-582

De Gregorio E, Han SJ, Lee WJ, Baek MJ, Osaki T, Kawabata S, Lee BL, Iwanaga S, Lemaitre B, & Brey PT (2002) An immune-responsive Serpin regulates the melanization cascade in *Drosophila*. *Dev Cell* **3**: 581-592

De Gregorio E, Spellman PT, Rubin GM, & Lemaitre B (2001) Genome-wide analysis of the *Drosophila* immune response by using oligonucleotide microarrays. *Proc Natl Acad Sci U S A* **98**: 12590-12595

Di Rosa M, Musumeci M, Scuto A, Musumeci S, & Malaguarnera L (2005) Effect of interferon-gamma, interleukin-10, lipopolysaccharide and tumor necrosis factor-alpha on chitotriosidase synthesis in human macrophages. *Clin Chem Lab Med* **43**: 499-502

Dunn PE & Drake D (1983) Fate of bacteria injected into naive and immunized larvae of the tobacco hornworm, *Manduca sexta*. *J Invertebr Pathol* **41**: 77-85

Dziarski R & Gupta D (2006) The peptidoglycan recognition proteins (PGRPs). *Genome Biol* **7**: 232

Engstrom Y, Loseva O, & Theopold U (2004) Proteomics of the *Drosophila* immune response. *Trends Biotechnol* **22**: 600-605

Frobius AC, Kanost MR, Gotz P, & Vilcinskas A (2000) Isolation and characterization of novel inducible serine protease inhibitors from larval hemolymph of the greater wax moth *Galleria mellonella*. *Eur J Biochem* **267**: 2046-2053

Fuchs TA, Abed U, Goosmann C, Hurwitz R, Schulze I, Wahn V, Weinrauch Y, Brinkmann V, & Zychlinsky A (2007) Novel cell death program leads to neutrophil extracellular traps. *J Cell Biol* **176**: 231-241

Funkhouser JD & Aronson NN,Jr (2007) Chitinase family GH18: evolutionary insights from the genomic history of a diverse protein family. *BMC Evol Biol* **7**: 96

Furusawa T, Rakwal R, Nam HW, Hirano M, Shibato J, Kim YS, Ogawa Y, Yoshida Y, Kramer KJ, Kouzuma Y, Agrawal GK, & Yonekura M (2008) Systematic investigation of the hemolymph proteome of *Manduca sexta* at the fifth instar larvae stage using one- and two-dimensional proteomics platforms. *J. Proteome Res.* **7**: 938-959

Gandhe AS, Arunkumar KP, John SH, & Nagaraju J (2006) Analysis of bacteria-challenged wild silkmoth, *Antheraea mylitta* (lepidoptera) transcriptome reveals potential immune genes. *BMC Genomics* **7**: 184

Gorman MJ, Wang Y, Jiang H, & Kanost MR (2007) *Manduca sexta* hemolymph proteinase 21 activates prophenoloxidase-activating proteinase 3 in an insect innate immune response proteinase cascade. *J Biol Chem* **282**: 11742-11749

Hirsch JG (1958) Bactericidal action of histone. *J Exp Med* **108**: 925-944

Horohov DW & Dunn PE (1983) Phagocytosis and nodule formation by hemocytes of *Manduca sexta* larvae following injection of *Pseudomonas aeruginosa*. *J Invertebr Pathol* **41**: 203-213

Horohov DW & Dunn PE (1982) Changes in the circulating hemocyte population of *Manduca sexta* larvae following injection of bacteria. *J Invertebr Pathol* **40**: 327-339

Ishii K, Hamamoto H, Kamimura M, & Sekimizu K (2008) Activation of the silkworm cytokine by bacterial and fungal cell wall components via a reactive oxygen species-triggered mechanism. *J Biol Chem* **283**: 2185-2191

Janeway CA,Jr & Medzhitov R (2002) Innate immune recognition. *Annu Rev Immunol* **20**: 197-216

Ji C, Wang Y, Guo X, Hartson S, & 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, Ma C, Lu ZQ, & Kanost MR (2004) Beta-1,3-glucan recognition protein-2 (betaGRP-2)from *Manduca sexta*; an acute-phase protein that binds beta-1,3-glucan and lipoteichoic acid to aggregate fungi and bacteria and stimulate prophenoloxidase activation. *Insect Biochem Mol Biol* **34**: 89-100

Johnston PA, Perin MS, Reynolds GA, Wasserman SA, & Sudhof TC (1990) Two novel annexins from *Drosophila melanogaster*. Cloning, characterization, and differential expression in development. *J Biol Chem* **265**: 11382-11388

Jordan PA & Gibbins JM (2006) Extracellular disulfide exchange and the regulation of cellular function. *Antioxidants redox signaling* **8**: 312

Jordan PA, Stevens JM, Hubbard GP, Barrett NE, Sage T, Authi KS, & Gibbins JM (2005) A role for the thiol isomerase protein ERP5 in platelet function. *Blood* **105**: 1500-1507.

Kaneko Y, Takaki K, Iwami M, & Sakurai S (2006) Developmental profile of annexin IX and its possible role in programmed cell death of the *Bombyx mori* anterior silk gland. *Zoolog Sci* **23**: 533-542

Kanost MR, Bradfield JY, Cook KE, Locke J, Wells MA, & Wyatt GR (1988) Gene structure, complementary DNA sequence, and developmental regulation of a low molecular weight hemolymph protein from *Locusta migratoria*. *Arch Insect Bioch Phys* **8**: 203-218.

Kanost MR, Jiang H, & Yu XQ (2004) Innate immune responses of a lepidopteran insect, *Manduca sexta*. *Immunol Rev* **198**: 97-105

Kanost MR, Zepp MK, Ladendorff NE, & Andersson LA (1994) Isolation and characterization of a hemocyte aggregation inhibitor from hemolymph of *Manduca sexta* larvae. *Arch Insect Biochem Physiol* **27**: 123-136

Karlsson C, Korayem AM, Scherfer C, Loseva O, Dushay MS, & Theopold U (2004) Proteomic analysis of the *Drosophila* larval hemolymph clot. *J Biol Chem* **279**: 52033-52041

Kawamura K, Shibata T, Saget O, Peel D, & Bryant PJ (1999) A new family of growth factors produced by the fat body and active on *Drosophila* imaginal disc cells. *Development* **126**: 211-219

Kim YJ, Lee KS, Kim BY, Choo YM, Sohn HD, & Jin BR (2007) Thioredoxin from the silkworm, *Bombyx mori*: cDNA sequence, expression, and functional characterization. *Comp Biochem Physiol B Biochem Mol Biol* **147**: 574-581

Kirkpatrick (1995) An abundantly secreted glycoprotein from *Drosophila melanogaster* is related to mammalian secretory proteins produced in rheumatoid tissues and by activated macrophages. *Gene* **153**: 147

Kotsyfakis M, Ehret-Sabatier L, Siden-Kiamos I, Mendoza J, Sinden RE, & Louis C (2005a) *Plasmodium berghei* ookinetes bind to *Anopheles gambiae* and *Drosophila melanogaster* annexins. *Mol Microbiol* **57**: 171-179

Kotsyfakis M, Vontas J, Siden-Kiamos I, & Louis C (2005b) The annexin gene family in the malaria mosquito *Anopheles gambiae*. *Insect Mol Biol* **14**: 555-562

Kurata S, Ariki S, & Kawabata S (2006) Recognition of pathogens and activation of immune responses in *Drosophila* and horseshoe crab innate immunity. *Immunobiology* **211**: 237-249

Ladendorff NE & Kanost MR (1991) Bacteria-induced protein P4 (hemolin) from *Manduca sexta*: a member of the immunoglobulin superfamily which can inhibit hemocyte aggregation. *Arch Insect Biochem Physiol* **18**: 285-300

Lee WJ, Lee JD, Kravchenko VV, Ulevitch RJ, & Brey PT (1996) Purification and molecular cloning of an inducible gram-negative bacteria-binding protein from the silkworm, *Bombyx mori*. *Proc Natl Acad Sci U S A* **93**: 7888-7893

Lemaitre B & Hoffmann J (2007) The host defense of *Drosophila melanogaster*. *Annu Rev Immunol* **25**: 697-743

Lesch C (2005) Untersuchugen zu moglichen Bindungs- und Akitvierungsfaktoren der zellularen Immunantwort bei *Manduca sexta* L. PhD Thesis. Justus-Liebig-Universität Giessen

Levy F, Bulet P, & Ehret-Sabatier L (2004) Proteomic analysis of the systemic immune response of *Drosophila*. *Mol Cell Proteomics* **3**: 156-166

Loseva O & Engström Y (2004) Analysis of signal-dependent changes in the proteome of *Drosophila* blood cells during an immune response. *Mol Cell Proteomics* **3**: 796-808

Lu Z & Jiang H (2008) Expression of *Manduca sexta* serine proteinase homolog precursors in insect cells and their proteolytic activation. *Insect Biochem Mol Biol* **38**: 89-98

Ma C & Kanost MR (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

Manickam N, Sun X, Li M, Gazitt Y, & Essex DW (2008) Protein disulphide isomerase in platelet function. *Br J Haematol* **140**: 223-229

Matsunaga TM & Fujiwara H (2002) Identification and characterization of genes abnormally expressed in wing-deficient mutant (flugellos) of the silkworm, *Bombyx mori*. *Insect Biochem Mol Biol* **32**: 691-699

Matzinger P (2002) The danger model: a renewed sense of self. *Science* **296**: 301-305

Merchant D, Ertl RL, Rennard SI, Stanley DW, & Miller JS (2008) Eicosanoids mediate insect hemocyte migration. *J Insect Physiol* **54**: 215-221

Miller JS & Stanley DW (2001) Eicosanoids mediate microaggregation reactions to bacterial challenge in isolated insect hemocyte preparations. *J Insect Physiol* **47**: 1409-1417

Moss SE & Morgan RO (2004) The annexins. *Genome Biol* **5**: 219

Onoe H, Matsumoto A, Hashimoto K, Yamano Y, & Morishima I (2007) Peptidoglycan recognition protein (PGRP) from eri-silkworm, *Samia cynthia ricini*; protein purification and induction of the gene expression. *Comp Biochem Physiol B Biochem Mol Biol* **147**: 512-519

Parseghian MH & Luhrs KA (2006) Beyond the walls of the nucleus: the role of histones in cellular signaling and innate immunity. *Biochem Cell Biol* **84**: 589-604

Patat SA, Carnegie RB, Kingsbury C, Gross PS, Chapman R, & Schey KL (2004) Antimicrobial activity of histones from hemocytes of the Pacific white shrimp. *Eur J Biochem* **271**: 4825-4833

Rubartelli A, Bajetto A, Allavena G, Wollman E, & Sitia R (1992) Secretion of thioredoxin by normal and neoplastic cells through a leaderless secretory pathway. *J Biol Chem* **267**: 24161-24164

Sansonetti PJ (2006) The innate signaling of dangers and the dangers of innate signaling. *Nat Immunol* **7**: 1237-1242

Schmidt O, Theopold U, Beckage NE (2008) Insect and Vertebrate Immunity: Key similarities versus differences. In *Insect Immunology,* Beckage NE (ed) pp 1-23. Elsevier

Scholz F (2002) Aktivierung von Hamozyten des Tabakschwarmers *Manduca sexta* nach bakteriellen Infektionen. PhD Thesis. Justus-Liebig-Universität Giessen

Schwertassek U, Balmer Y, Gutscher M, Weingarten L, Preuss M, Engelhard J, Winkler M, & Dick TP (2007) Selective redox regulation of cytokine receptor signaling by extracellular thioredoxin-1. *EMBO J* **26**: 3086-3097

Shi L & Paskewitz SM (2004) Identification and molecular characterization of two immune-responsive chitinase-like proteins from *Anopheles gambiae*. *Insect Mol Biol* **13**: 387-398

Shrestha S & Kim Y (2008) Eicosanoids mediate prophenoloxidase release from oenocytoids in the beet armyworm *Spodoptera exigua*. *Insect Biochem Mol Biol* **38**: 99-112

Shrivastava B & Ghosh AK (2003) Protein purification, cDNA cloning and characterization of a protease inhibitor from the Indian tasar silkworm, *Antheraea mylitta*. *Insect Biochem Mol Biol* **33**: 1025-1033

Strand MR (2008) Insect hemocytes and their role in immunity. In *Insect Immunity,* Beckage NE (ed) pp 25-47. Elsevier

Su XD, Gastinel LN, Vaughn DE, Faye I, Poon P, & Bjorkman PJ (1998) Crystal structure of hemolin: a horseshoe shape with implications for homophilic adhesion. *Science* **281**: 991-995

Sun SC, Lindstrom I, Boman HG, Faye I, & Schmidt O (1990) Hemolin: an insect-immune protein belonging to the immunoglobulin superfamily. *Science* **250**: 1729-1732

Vierstraete E, Verleyen P, Baggerman G, D'Hertog W, Van den Bergh G, Arckens L, De Loof A, & Schoofs L (2004a) A proteomic approach for the analysis of instantly released wound and immune proteins in *Drosophila melanogaster* hemolymph. *Proc Natl Acad Sci U S A* **101**: 470-475

Vierstraete E, Verleyen P, Sas F, Van den Bergh G, De Loof A, Arckens L, & Schoofs L (2004b) The instantly released *Drosophila* immune proteome is infection-specific. *Biochem Biophys Res Commun* **317**: 1052-1060

Wang Y & Jiang H (2007) Reconstitution of a branch of the *Manduca sexta* prophenoloxidase activation cascade in vitro: snake-like hemolymph proteinase 21 (HP21) cleaved by HP14 activates prophenoloxidase-activating proteinase-2 precursor. *Insect Biochem Mol Biol* **37**: 1015-1025

Wang Y & Jiang H (2006) Interaction of beta-1,3-glucan with its recognition protein activates hemolymph proteinase 14, an initiation enzyme of the prophenoloxidase activation system in *Manduca sexta*. *J Biol Chem* **281**: 9271-9278

Willott E, Wang XY, & Wells MA (1989) cDNA and gene sequence of *Manduca sexta* arylphorin, an aromatic amino acid-rich larval serum protein. Homology to arthropod hemocyanins. *J Biol Chem* **264**: 19052-19059

Wood ZA, Schroder E, Robin Harris J, & Poole LB (2003) Structure, mechanism and regulation of peroxiredoxins. *Trends Biochem Sci* **28**: 32-40

Xia QY, Fujii H, Kusakabe T, & Banno Y (2001) Identification of three annexin IX isoforms generated by alternative splicing of the carboxyl-terminal exon in silkworm, *Bombyx mori*. *Insect Biochem Mol Biol* **32**: 9-14

Yu XQ, Gan H, & Kanost MR (1999) Immulectin, an inducible C-type lectin from an insect, *Manduca sexta*, stimulates activation of plasma prophenol oxidase. *Insect Biochem Mol Biol* **29**: 585-597

Yu XQ, Jiang H, Wang Y, & Kanost MR (2003) Nonproteolytic serine proteinase homologs are involved in prophenoloxidase activation in the tobacco hornworm, *Manduca sexta*. *Insect Biochem Mol Biol* **33**: 197-208

Yu XQ & Kanost MR (2004) Immulectin-2, a pattern recognition receptor that stimulates hemocyte encapsulation and melanization in the tobacco hornworm, *Manduca sexta*. *Dev Comp Immunol* **28**: 891-900

Yu XQ & Kanost MR (2002) 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 XQ & Kanost MR (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 XQ & Kanost MR (1999) Developmental expression of *Manduca sexta* hemolin. *Arch Insect Biochem Physiol* **42**: 198-212

Yu XQ, Ling E, Tracy ME, & Zhu Y (2006) Immulectin-4 from the tobacco hornworm *Manduca sexta* binds to lipopolysaccharide and lipoteichoic acid. *Insect Mol Biol* **15**: 119-128

Yu XQ & Ma Y (2006) Calcium is not required for immulectin-2 binding, but protects the protein from proteinase digestion. *Insect Biochem Mol Biol* **36**: 505-516

Yu XQ, Tracy ME, Ling E, Scholz FR, & Trenczek T (2005) A novel C-type immulectin-3 from *Manduca sexta* is translocated from hemolymph into the cytoplasm of hemocytes. *Insect Biochem Mol Biol* **35**: 285-295

Yu XQ, Zhu YF, Ma C, Fabrick JA, & Kanost MR (2002) Pattern recognition proteins in *Manduca sexta* plasma. *Insect Biochem Mol Biol* **32**: 1287-1293

Zhang J, Iwai S, Tsugehara T, & Takeda M (2006) MbIDGF, a novel member of the imaginal disc growth factor family in *Mamestra brassicae*, stimulates cell proliferation in two lepidopteran cell lines without insulin. *Insect Biochem Mol Biol* **36**: 536-546

Zhu Y (2001) Identification of immune-related genes from the tobacco hornworm, *Manduca sexta*, and characterization of two immune-inducible proteins, Serpin-3 and Leureptin.

Zhu Y, Johnson TJ, Myers AA, & Kanost MR (2003a) Identification by subtractive suppression hybridization of bacteria-induced genes expressed in *Manduca sexta* fat body. *Insect Biochem Mol Biol* **33**: 541-559

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

Tables

Table 3-1. Aldente MALDI-TOF identification of bar-stage plasma proteins in curdlan-bound fraction.

Sample name	Score*	Hits/Total	Coverage %	Accession #	Description	Observed mass (kDa)	Calculated mass (kDa)
1	1577.48	113/197	51	Q25490	Apolipophorin-I.	208	289
2	560.86	86/204	37	Q25490	Apolipophorin-I.	190	289
3	813.16	55/199	70	P14296	Arylphorin subunit alpha.	79	82
4	487.49	47/203	67	P14297	Arylphorin subunit beta.	74	82
4, 1 val.	145.35	23/156	37	Q25490	Apolipophorin-II.	74	75
4, 2 val.	40.53	13/133	29	Q867T1	Serpin 3	74	51
5	1038.86	64/197	71	O44249	Phenoloxidase subunit 1. (proPO-1)	67	79
5, 1 val.	614.11	42/133	47	Q25519	Phenoloxidase subunit 2. (proPO-2)	67	80
6	114.81	30/214	42	Q25519	Phenoloxidase subunit 2. (proPO-2)	64	80
6, 1 val.	43.99	18/184	36	Q8ISB6	Beta-1,3-glucan-binding protein 2.	64	52
7	137.06	24/464	47	Q8ISB6	Beta-1,3-glucan-binding protein 2.	51	52
8	242.14	25/151	50	Q8ISB6	Beta-1,3-glucan-binding protein 2.	47	52
9	116.89	20/173	39	Q8ISB6	Beta-1,3-glucan-binding protein 2.	43	52
10	88.60	22/186	39	Q8ISB6	Beta-1,3-glucan-binding protein 2.	36	52
10, 1 val.	61.86	17/164	41	Q56SF4	Serine proteinase-like protein 4. (SPH-4)	36	43
10, 2 val.	58.10	13/147	44	Q8MUL9	Serine proteinase-like protein 2. (SPH-2)	36	43
10, 3 val.	45.93	12/134	40	Q9NBV9	Immulectin-2. (IML-2)	36	37
11	86.96	18/178	41	Q8ISB6	Beta-1,3-glucan-binding protein 2.	31	52
11, 1 val.	33.18	12/160	32	Q8MUL9	Serine proteinase-like protein 2. (SPH-2)	31	43
11, 2 val.	30.66	12/148	27	Q56SF4	Serine proteinase-like protein 4. (SPH-4)	31	43
12	28.06	14/172	21	Q69BL0	pattern recognition serine proteinase (HP14)	2	72
12, 1 val.	100.51	21/158	39	Q8ISB6	Beta-1,3-glucan-binding protein 2.	2	52
*Scores over 16.36 are significant							

Table 3-2. *M. sexta* proteins eluted from *in vitro* bacterial binding assays and identified by MS.

Protein description					Aldente search results				Mascot search		
Immune related proteins	**Band**	**Mw Theo.**	**Mw obs.**	**Signal peptide**	**Aldente Score**	**Hits/Peaks**	**Cov (%)**	**Uniprot Accession**	**Mascot score***	**NCBI Accession**	**Species[a]**
Cationic peptide CP8.	Bt13	11	9	yes	29.51	9/224	62	A4LA63	301	gi|134103857	
Cationic peptide CP8.	Cn12	11	9	yes	7.88	9/176	58	A4LA63	335	gi|134103857	
Cationic peptide CP8.	Sa4	11	9	yes	3.03	6/176	58	A4LA63	201	gi|134103857	
Hemolin.	Cn5	44	48	yes	140.17	24/275	71	P31398	506	gi|511297	
Leureptin.	Ec3	47	52	yes	131.5	32/159	65	Q86RS5	762	gi|27733411	
Leureptin.	Ec4	47	45	yes	64.91	24/214	53	Q86RS5	161	gi|27733411	
Lysozyme.	Bt12	14	13	yes	70.34	16/190	79	Q9TWY1	n/d	n/d	
Lysozyme.	Cn11	14	13	yes	25.23	15/145	75	Q9TWY1	222	gi|233964	
Lysozyme.	Ec7	14	13	yes	27.42	16/140	75	Q9TWY1	136	gi|233964	
Lysozyme.	Sa3	14	13	yes	31.94	15/170	73	Q9TWY1	241	gi|233964	
Pattern recognition serine proteinase precursor (HP14)	Bt9	29[#]	28	yes	n/d	n/d	n/d	n/d	71*	gi|39655053	
Peptidoglycan recognition protein 1A.	Bt11	22	17	yes	130.96	15/210	66	Q86D95	331	gi|27733423	
Peptidoglycan recognition protein 1A.	Sm8	22	16.5	yes	56.01	11/250	48	Q86D95	542	gi|27733423	
Peptidoglycan recognition protein 1B.	Bt11	22	17	yes	131.64	15/210	66	Q86RS6	n/d	n/d	
Peptidoglycan recognition protein 1B.	Sm8	22	16.5	yes	56.3	11/250	48	Q86RS6	543	gi|27733409	
Phenoloxidase subunit 1.	Bt2	79	82	no	132.44	20/135	42	O44249	n/d	n/d	
Phenoloxidase subunit 1.	Bt3	79	61	no	41.72	29/303	40	O44249	97	gi|74763772	
Phenoloxidase subunit 1.	Cn3	79	63	no	168.6	37/290	56	O44249	226	gi|74763772	
Phenoloxidase subunit 1.	Ec1	79	65	no	56.08	33/310	49	O44249	n/d	n/d	
Phenoloxidase subunit 1.	Sa1	79	82	no	128.12	24/146	48	O44249	n/d	n/d	
Phenoloxidase subunit 1.	Sm1	79	87	no	277.86	47/278	75	O44249	162	gi|74763772	
Phenoloxidase subunit 2.	Bt2	80	82	no	139.77	22/157	36	Q25519	155	gi|75038472	
Phenoloxidase subunit 2.	Cn1	80	82	no	24.7	13/110	23	Q25519	n/d	n/d	
Phenoloxidase subunit 2.	Sa1	80	82	no	168.98	28/174	42	Q25519	n/d	n/d	
Phenoloxidase subunit 2.	Sm1	80	87	no	203.95	28/199	43	Q25519	351	gi|75038472	
Serine protease homolog (SPH-1)	Cn7	31[#]	34	yes	n/d	n/d	n/d	n/d	73*	gi|27733421	

Serine protease homolog 2 (SPH-2)	Bt7	36[#]	39	yes	n/d	n/d	n/d	n/d	75*	gi|21630233	
Serine protease homolog 2 (SPH-2)	Bt8	36[#]	32	yes	n/d	n/d	n/d	n/d	107	gi|21630233	
Serine protease homolog 2 (SPH-2)	Cn7	36[#]	34	yes	n/d	n/d	n/d	n/d	73*	gi|21630233	
Storage Proteins	**Band**	**Mw Theo.**	**Mw obs.**	**Signal peptide**	**Aldente Score**	**Hits/Peaks**	**Cov (%)**	**Uniprot Accession**	**Mascot score***	**NCBI Accession**	**Species[a]**
Arylphorin subunit alpha.	Bt2	82	82	yes	142.34	25/182	41	P14296	171	gi|114240	
Arylphorin subunit alpha.	Cn1	82	82	yes	514.93	39/163	65	P14296	n/d	n/d	
Arylphorin subunit alpha.	Sa1	82	82	yes	296.65	35/209	57	P14296	250	gi|114240	
Arylphorin subunit alpha.	Sm1	82	87	yes	32.15	18/171	26	P14296	n/d	n/d	
Arylphorin subunit beta.	Bt2	82	82	yes	234.38	33/215	53	P14297	186	gi|1168527	
Arylphorin subunit beta.	Cn1	82	82	yes	322.99	40/203	64	P14297	n/d	n/d	
Arylphorin subunit beta.	Cn7	82	34	yes	29.82	32/310	37	P14297	n/d	n/d	
Arylphorin subunit beta.	Sa1	82	82	yes	300.97	43/252	63	P14297	171	gi|159491	
Arylphorin subunit beta.	Sm1	82	87	yes	179.66	32/231	51	P14297	162	gi|159491	
Methionine-rich storage protein 3.	Ec2	87	55	yes	25.96	19/229	24	Q25518	n/d	n/d	
Structural proteins	**Band**	**Mw Theo.**	**Mw obs.**	**Signal peptide**	**Aldente Score**	**Hits/Peaks**	**Cov (%)**	**Uniprot Accession**	**Mascot score***	**NCBI Accession**	**Species[a]**
Actin.	Ec4	42	45	no	22.59	13/190	52	Q4JQ54	138	gi|91088367	
Actin.	Bt6	42	43	no	81.98	17/272	60	Q4PKE5	368	gi|59894747	
Actin.	Cn6	42	42	no	82.96	21/226	62	P04829	86	gi|116222085	
Transgelin. (BOMMO)	Bt10	21	20	no	19.12	9/260	57	Q1HQ70	131	gi|114051357	*Bombyx mori*
Other insect proteins	**Band**	**Mw Theo.**	**Mw obs.**	**Signal peptide**	**Aldente Score**	**Hits/Peaks**	**Cov (%)**	**Uniprot Accession**	**Mascot score***	**NCBI Accession**	**Species[a]**
Annexin IX	Bt8	36	32	no	17.97	14/228	45	Q5F321	n/d	n/d	
Annexin IX	Cn8	36	28	no	45.62	17/204	48	Q5F321	210	gi|58864720	
Annexin IX	Ec5	36	33	no	27.63	17/192	48	Q5F321	n/d	n/d	
Annexin IX	Sm6	36	32	no	161.97	25/234	67	Q5F321	360	gi|58864720	
Elongation factor 1-alpha.	Bt5	45	52	no	44.7	16/198	47	Q9BP07	127	gi|12836925	
Elongation factor 2. (BOMMO)	Bt1	95	99	no	174.24	34/211	42	Q1HPK6	260	gi|28627569	*Bombyx mori*
Histone H2B.	Sm9	14	12	no	24.51	8/203	50	Q17EF1			
Histone H3.3.	Sm9	15	15	no	17.59	14/217	45	P84250			
Histone H4.	Sm11	11	11	no	24.99	12/167	62	P84048			
Insecticyanin B form.	Cn9	21	19	yes	20.56	11/250	58	Q00630	374	gi|124527	
Insecticyanin A form.	Cn9	21	19	yes	n/d	n/d	n/d	gi|124527	252	gi|124151	
Peroxiredoxin	Bt10	22	20	no	n/d	n/d	n/d	n/d	115	gi|38047571	*Drosophila yakuba*

Protein disulfide-isomerase like protein (ERp57, BOMMO)	Sm3	55	61	Yes	19.27	18/324	42	Q587N3			*Bombyx mori*
Serpin-1.	Cn6	44	42	Yes	65.67	19/245	49	Q25498	n/d	n/d	
Thioredoxin-like protein	Cn12	12	12	No	n/d	n/d	n/d	n/d	147	gi|6560635	
Transferrin.	Cn1	73	82	Yes	36.05	14/124	28	P22297	n/d	n/d	
HAIP-like proteins	**Band**	**Mw Theo.**	**Mw obs.**	**Signal peptide**	**Aldente Score**	**Hits/Peaks**	**Cov (%)**	**Uniprot Accession**	**Mascot score***	**NCBI Accession**	**Species[a]**
Imaginal disc growth factor-like protein...	Bt5	48	53	Yes	19.35	11/182	33	Q0QJJ5	n/d	gi|85726208	*Mamestra brassicae*
Imaginal disc growth factor-like protein...	Cn4	48	52	Yes	29.28	13/230	30	Q0QJJ5	214	gi|85726208	*Mamestra brassicae*
Imaginal disc growth factor-like protein...	Sa2	48	52	Yes	16.51	10/194	20	Q0QJJ5	167	gi|85726208	*Mamestra brassicae*
HAIP and Protease sequence search	**Band**	**Mw Theo.**	**Mw obs.**	**Signal peptide**	**Aldente Score**	**Hits/Peaks**	**Cov (%)**	**Uniprot Accession**	**Mascot score****	**NCBI Accession**	**Species[a]**
Hemocyte aggregation inhibitory peptide (HAIP)	Bt5	48	53	Yes	n/d	n/d	n/d	n/d	179	n/a	
HAIP	Cn4	48	52	Yes	n/d	n/d	n/d	n/d	577	n/a	
HAIP	Cn5	48	48	Yes	n/d	n/d	n/d	n/d	36	n/a	
HAIP	Ec2	48	55	Yes	n/d	n/d	n/d	n/d	36	n/a	
HP14	Sm7	28[#]	29	Yes	n/d	n/d	n/d	n/d	57	gi|39655053	
HP14	Bt 9	28[#]	29	Yes	n/d	n/d	n/d	n/d	78	gi|39655053	

[a] Species is *Manduca sexta* unless otherwise noted
*Mascot scores over 80 are in the 95% confidence interval
**Mascot scores over 35 from HAIP and protease sequence search are in the 99% confidence interval
[#] size of catalytic (or catalytic-like) domain only

Table 3-3. Bacterial proteins eluted from *in vitro* bacterial binding assays.

Bacterial proteins				Aldente search results					Mascot search results		
Description	**Band**	**Calc. mass**	**Obs. Mass**	**Aldente Score**	**Hits/ Peaks**	**Cov (%)**	**Uniprot Accession**	**Species**	**Mascot score**	**NCBI Accession**	**Species**
30S ribosomal protein S15	Bt13	10.6	9	n/d	n/d	n/d	n/d	n/d	129	gi|30021897	*Bacillus cereus* ATCC 14579
5-keto-4-deoxyuronate isomerase	Ec6	31	29	126.56	21/170	58	Q46938	ECOLI	507	gi|16130747	*Escherichia coli* K12
60 kDa chaperonin.	Sm2	57	68	186.93	43/264	74	O66206	SERMA	779	gi|6225138	SERMA
DNA-binding protein HU	Bt14	9.6	8	n/d	n/d	n/d	n/d	n/d	146	gi|30261605	*Bacillus anthracis* str. Ames
DNA-binding protein HU-alpha	Ec8	9.5	9	n/d	n/d	n/d	n/d	n/d	110	gi|15804591	*Escherichia coli* O157:H7 EDL933
D-ribose-binding periplasmic protein.	Ec6	28	29	74.34	13/149	64	P02925	ECOLI	n/d	n/d	n/d
Malate dehydrogenase.	Ec5	32	33	239.73	19/175	80	A1AGC9	ECOK1	754	gi|26249817	*Escherichia coli* CFT073
Outer membrane protein A.	Sm7	36	28	124.83	25/219	77	P04845	SERMA	617	gi|129141	SERMA
Periplasmic oligopeptide-binding protein.	Ec1	58	65	60.07	29/277	60	P23843	ECOLI	465	gi|26247572	*Escherichia coli* CFT073

Table 3-4. EST database identification of proteins eluted from *in vitro* bacterial binding assays.

The Mascot program was used to search MALDI-TOF/TOF data against compiled *M. sexta* ESTs from NCBI and a pyrosequencing EST project (Haibo Jiang). Identification of the ESTs is based on BLAST searches against the NCBI database

Sample	Name	Contig or Accession #	Protein Score	Total Ion Score	Total Ion CI%	Protein Score CI%	Peptide count
Bt1	Elongation factor 2	gi|10763916|	176	139	100	100	6
Bt1	Elongation factor 2	Contig6411	171	139	100	100	5
Bt2	Arylphorin	Contig7105	157	129	100	100	6
Bt2	Arylphorin	Contig7225	154	129	100	100	7
Bt2	Arylphorin	Contig6916	142	129	100	100	3
Bt2	Arylphorin alpha	Contig7216	115	103	100	100	5
Bt2	Arylphorin alpha	Contig7217	97	88	100	99.999	3
Bt2	PPO1&2	Contig7215	89	73	100	99.992	6
Bt2	PPO1&2	Contig7213	81	72	100	99.948	4
Bt2	Arylphorin	Contig7226	74	47	99.998	99.733	8
Bt2	Arylphorin alpha	Contig6236	67	56	100	98.835	2
Bt2	Arylphorin	Contig7202	65	56	100	98.022	3
Bt3	PPO1&2	Contig7159	87	46	99.997	99.988	11
Bt4	ERp57 (PDI-like)	Contig6705	162	129	100	100	5
Bt5	Elongation factor 1 alpha	Contig7207	104	58	100	100	10
Bt5	HAIP	Contig7147	98	131	100	100	3
Bt5	Similar to p50 βGRP (*B. mori*)	Contig6004	97	88	100	99.999	1
Bt5	EF1alpha	gi|3719506|	85	58	100	99.98	5
Bt5	EF1alpha	Contig7069	66	58	100	98.534	2
Bt5	EF1alpha	gi|3658559|	63	58	100	96.641	1
Bt6	Actin	Contig7165	197	198	100	100	5
Bt6	Actin	Contig7141	184	141	100	100	7
Bt6	Actin	Contig6831	129	122	100	100	2
Bt6	Actin	Contig7124	125	195	100	100	2
Bt6	Actin	Contig7161	110	192	100	100	2
Bt6	Actin	gi|26055765|	109	94	100	100	3
Bt6	Actin	Contig4438	104	98	100	100	1
Bt6	Actin	Contig7094	102	76	100	100	4
Bt6	Actin	gi|3859637|	100	76	100	100	4
Bt6	Actin	gi|3719571|	99	91	100	100	2
Bt7	SPH-2	Contig5656	62	57	100	96.563	1
Bt8	Similar to multi-binding protein (*B. mori*)	gi|14860775|	193	167	100	100	4
Bt8	Similar to multi-binding protein (*B. mori*)	Contig4547	123	167	100	100	2
Bt8	SPH-2	Contig5656	85	80	100	99.982	1
Bt9	No EST hit						
Bt10	Peroxiredoxin	Contig6551	257	221	100	100	6
Bt10	Similar to transgelin (*B. mori*)	Contig5374	119	90	100	100	4

Sample	Name	Contig or Accession #	Protein Score	Total Ion Score	Total Ion CI%	Protein Score CI%	Peptide count
Bt11	PGRP	gi|14860840|	333	258	100	100	9
Bt11	PGRP	Contig6444	306	258	100	100	7
Bt12	No EST hit						
Bt13	CP8	Contig6888	299	247	100	100	6
Bt13	CP8	Contig6921	292	247	100	100	6
Bt13	CP8	Contig6795	154	209	100	100	4
Bt13	CP8	Contig7084	130	118	100	100	2
Bt13	Similar to RPS30 (*B. mori*)	Contig6809	61	45	99.997	95.469	3
Bt14	CP8	Contig6888	62	42	99.993	96.401	3
Bt14	CP8	Contig6921	59	42	99.993	92.48	3
Sample	**Name**	**Contig or Accession #**	**Protein Score**	**Total Ion Score**	**Total Ion CI%**	**Protein Score CI%**	**Pep. count**
cn1	No EST hit						
cn2	No EST hit						
cn3	PPO1&2	Contig7159	220	135	100	100	16
cn3	PPO1&2	Contig7195	85	135	100	99.982	7
cn3	PPO1&2	Contig4466	85	57	100	99.981	4
cn4	HAIP	gi|10764042|	258	200	100	100	8
cn4	HAIP	Contig7147	240	351	100	100	5
cn4	HAIP	Contig4477	69	64	100	99.175	1
cn5	Hemolin	gi|18957964|	382	309	100	100	9
cn5	Hemolin	Contig6907	365	309	100	100	8
cn5	Hemolin	Contig6879	237	230	100	100	6
cn5	Hemolin	gi|8276525|	209	172	100	100	5
cn5	Hemolin	gi|14863329|	206	172	100	100	4
cn5	Hemolin	gi|3738723|	199	172	100	100	4
cn5	Hemolin	gi|10764014|	121	94	100	100	5
cn5	Hemolin	Contig6822	110	94	100	100	3
cn6	Actin	Contig7165	65	57	100	98.067	2
cn6	Actin	Contig4438	64	57	100	97.452	1
cn6	Actin	Contig6831	63	57	100	96.865	1
cn6	Actin	Contig7141	58	57	100	89.13	1
cn6	Serpin-1	gi|8276517|	56	47	99.997	85.67	2
cn6	Serpin-1	Contig6815	55	47	99.997	78.805	2
cn6	Serpin-1	Contig5962	52	47	99.997	64.825	1
cn7	Similar to mannose-binding protein associated serine protease	gi|14860852|	77	60	100	99.872	3
cn7	SPH-3	Contig6436	75	60	100	99.807	3
cn7	SPH-3	Contig7083	69	60	100	99.282	3
cn7	SPH-2	Contig5656	60	55	100	94.425	1
cn8	Annexin	Contig6948	132	144	100	100	7
cn8	Iml-4	Contig7106	55	28	99.805	79.287	6
cn9	Insecticyanin	Contig6998	241	210	100	100	5
cn9	Peroxiredoxin	Contig6551	207	180	100	100	5
cn9	Insecticyanin	Contig6219	117	91	100	100	4
cn10	Actin depol. factor	Contig6714	139	108	100	100	4

Sample	Name	Contig or Accession #	Protein Score	Total Ion Score	Total Ion CI%	Protein Score CI%	Pep. count
cn10	Actin depol. factor	gi|12007749|	138	108	100	100	5
cn10	Apolipophorin III	Contig7045	73	55	100	99.679	3
cn10	Apolipophorin III	Contig7155	69	55	100	99.23	3
cn10	Peroxiredoxin	Contig6551	64	56	100	97.51	2
cn10	Actin	Contig7049	63	55	100	96.936	2
cn10	Similar to transgelin (*B. mori*)	Contig5374	59	30	99.9	92.651	4
cn10	Immune induced protein-1	gi|14863334|	54	27	99.786	72.695	4
cn11	Lysozyme	Contig7173	197	134	100	100	9
cn11	Lysozyme	Contig6863	191	134	100	100	9
cn11	Lysozyme	Contig6986	183	134	100	100	8
cn11	Lysozyme	gi|14860785|	110	83	100	100	4
cn11	Lysozyme	Contig5577	104	72	100	100	4
cn12	CP8	Contig6888	333	292	100	100	5
cn12	CP8	Contig6921	327	292	100	100	5
cn12	CP8	Contig7084	165	153	100	100	2
cn12	CP8	Contig6795	155	246	100	100	3
cn12	Thioredoxin-like	Contig6039	140	102	100	100	5
cn12	Thioredoxin-like	gi|3658502|	139	102	100	100	5
cn12	Thioredoxin-like	Contig7011	135	102	100	100	5
cn12	Thioredoxin-like	gi|3658503|	124	102	100	100	4
cn13	Ribosomal protein S27A	Contig6778	109	78	100	100	4
cn13	Ubiquitin	Contig4258	108	78	100	100	4
cn13	Ribosomal protein S27A	gi|10763932|	108	78	100	100	5
cn13	Ubiquitin & ribosomal protein s27a	Contig7061	107	78	100	100	5
cn13	Ubiquitin	gi|14860751|	105	78	100	100	5
cn13	Ubiquitin	Contig6085	103	78	100	100	3
cn13	Ubiquitin & ribosomal protein s27a	Contig3192	103	78	100	100	3
cn13	Ubiquitin	Contig6329	99	78	100	100	3
cn13	Ubiquitin	gi|12007767|	94	78	100	99.997	3

Sample	Name	Contig or Accession #	Protein Score	Total Ion Score	Total Ion CI%	Protein Score CI%	Pep. count
ec1	No EST hit						
ec2	Similar to p50 βGRP (*B. mori*)	Contig6004	59	50	99.999	92.304	1
ec3	Leureptin	Contig6569	341	356	100	100	5
ec3	Leureptin	Contig6618	103	69	100	100	6
ec3	Leureptin	gi|14863342|	103	69	100	100	5
ec4	Actin	Contig7141	107	92	100	100	4
ec4	Leureptin	Contig6569	102	87	100	100	4
ec4	Actin	Contig6831	84	74	100	99.976	2
ec4	Actin	Contig7165	84	92	100	99.975	3
ec4	Actin	Contig7161	83	74	100	99.969	2
ec4	Actin	Contig4438	80	74	100	99.939	1
ec4	Actin	Contig7124	80	92	100	99.937	2
ec5	No EST hit						
ec6	No EST hit						
ec7	Lysozyme	Contig6863	119	62	100	100	9

ec7	Lysozyme	Contig7173	114	62	100	100	8
ec7	Lysozyme	Contig6986	111	62	100	100	8
ec7	Lysozyme	Contig5577	94	62	100	99.998	4
ec8	MBF2 (*Samia cynthia*)	Contig4132	75	43	99.994	99.811	4

Sample	**Name**	**Contig or Accession #**	**Protein Score**	**Total Ion Score**	**Total Ion CI%**	**Protein Score CI%**	**Pep. count**
sa1	Arylphorin alpha	Contig7216	184	171	100	100	6
sa1	Arylphorin alpha	Contig7217	165	153	100	100	4
sa1	Arylphorin alpha	Contig6236	144	126	100	100	3
sa1	Arylphorin	Contig7180	133	126	100	100	2
sa1	Arylphorin	Contig7202	132	126	100	100	2
sa1	Arylphorin	Contig7226	86	44	99.996	99.986	11
sa1	Arylphorin alpha	Contig6891	72	44	99.996	99.656	4
sa1	Arylphorin	Contig6916	71	41	99.99	99.48	5
sa1	Arylphorin	Contig7225	65	41	99.99	97.929	7
sa1	Arylphorin	Contig7105	62	41	99.99	95.672	5
sa2	HAIP	Contig7147	222	287	100	100	4
sa2	HAIP	gi|10764042|	194	136	100	100	8
sa3	Lysozyme	Contig7173	212	138	100	100	10
sa3	Lysozyme	Contig6863	206	138	100	100	10
sa3	Lysozyme	Contig6986	188	138	100	100	8
sa3	Lysozyme	gi|14860785|	115	88	100	100	4
sa3	Lysozyme	Contig5577	108	76	100	100	4
sa4	CP8	Contig6888	199	158	100	100	5
sa4	CP8	Contig6921	193	158	100	100	5
sa4	MBF2 (*Samia cynthia*)	Contig6722	179	146	100	100	4
sa4	CP8	Contig6795	105	138	100	100	3
sa4	CP8	Contig7084	82	70	100	99.96	2

Sample	**Name**	**Contig or Accession #**	**Protein Score**	**Total Ion Score**	**Total Ion CI%**	**Protein Score CI%**	**Pep. count**
sm1	PPO1&2	Contig7215	219	182	100	100	10
sm1	PPO1&2	Contig7159	138	63	100	100	15
sm1	PPO1&2	Contig6916	125	103	100	100	4
sm1	Arylphorin	Contig7105	116	103	100	100	4
sm1	Arylphorin	Contig7225	116	103	100	100	5
sm1	PPO1&2	Contig7195	82	63	100	99.965	7
sm1	PPO1&2	Contig6970	82	62	100	99.961	4
sm1	PPO1&2	Contig7213	74	65	100	99.773	4
sm2	No EST hit						
sm3	ERp57 (PDI-like)	Contig6705	180	126	100	100	7
sm3	ERp57 (PDI-like)	Contig6611	104	87	100	100	3
sm3	ERp57 (PDI-like)	Contig6428	89	63	100	99.992	5
sm4	No EST hit						
sm5	No EST hit						
sm6	Annexin	Contig6948	237	261	100	100	11
sm7	No EST hit						
sm8	PGRP	gi|14860840|	378	289	100	100	10
sm8	PGRP	Contig6598	222	203	100	100	3
sm8	PGRP	Contig6444	207	221	100	100	6

Sample	Name	Contig or Accession #	Protein Score	Total Ion Score	Total Ion CI%	Protein Score CI%	Pep. count
sm9	Histone H3	Contig4822	73	74	100	99.686	6
sm10	Histone H2A	gi|10763973|	112	99	100	100	3
sm10	Histone H2A	gi|10764117|	106	99	100	100	2
sm10	Histone H2A	Contig4479	62	56	100	95.961	1
sm11	No blast hit	Contig5842	217	346	100	100	4
sm12	MBF2 (*Samia cynthia*)	Contig6472	213	201	100	100	2
sm12	MBF2 (*Samia cynthia*)	Contig6722	167	145	100	100	3
sm12	No blast hit	Contig208	63	57	100	96.792	1

Table 3-5. Proteins from *in vitro* bacterial binding detected by Aldente but not Mascot protein database search.

Band	Function	Description	Aldente score
Bt2	immune	Phenoloxidase subunit 1.	132.44
Cn1	immune	Phenoloxidase subunit 2.	24.7
Cn1	storage	Arylphorin subunit alpha.	514.93
Cn1	storage	Arylphorin subunit beta.	322.99
Cn1	other	Transferrin.	36.05
Cn6	other	Serpin-1.	65.67
Ec1	immune	Phenoloxidase subunit 1.	56.08
Ec2	storage	Methionine-rich storage protein 3.	25.96
Ec6	bacteria	D-ribose-binding periplasmic protein.	74.34
Ml1	immune	Phenoloxidase subunit 1.	100.71
Sa1	immune	Phenoloxidase subunit 2.	168.98
Sa1	immune	Phenoloxidase subunit 1.	128.12

Aldente scores greater than 16 are significant.

Table 3-6. Proteins from *in vitro* bacterial binding detected by Mascot but not Aldente protein database search.

Band	Function	Description	Mascot score	Mascot C.I.%	Peptides matched
Bt7	immune	serine proteinase-like protein 2 (SPH-2)	75	85.7	5
Bt8	immune	serine proteinase-like protein 2 (SPH-2)	107	100.0	6
Bt9	immune	pattern recognition serine proteinase precursor (HP14)	71	61.4	2
Bt10	other	thiol peroxiredoxin [Bombyx mori]	82	96.9	4
Cn7	immune	serine protease-like protein (SPH-1)	73	76.8	3
Cn7	immune	serine proteinase-like protein 2 (SPH-2)	73	72.7	5
Cn9	other	Insecticyanin A form (INS-a)	252	100.0	6
Cn12	immune	cationic peptide CP8 precursor	335	100.0	5
Cn12	other	thioredoxin-like protein	147	100.0	5
Sa4	immune	cationic peptide CP8 precursor	201	100.0	5

Table 3-7. Summary of *M. sexta* proteins found bound to bacteria or curdlan.

The first column under each type of bacteria (or curdlan) summarizes the bound proteins that were detected by Mascot search of NCBInr (MSn) or detected by Mascot search of *M. sexta* HAIP and active protease sequences (MSp) (data from Table 3-2). The middle column identifies proteins in each sample detected by Mascot search or the *M. sexta* ESTs results in Table 3-5 from (MSe). The last column summarizes proteins detected by immunoblotting (IB) (results shown in Fig. 3-5). The summary of EST and immunoblot results does not distinguish between the arylphorins α and β or between proPO-1 and -2.

Protein	***S. aureus***	***B. thuringiensis***	***S. marcesens***	***E. coli***	**curdlan**
Actin.		MSn MSe		MSn MSe	MSn MSe
Actin depol. Factor					MSe
Annexin-IX	IB	MSn IB	MSn MSe IB	MSn	MSn MSe
Apolipophorin III					MSe
Arylphorin subunit alpha.	MSn MSe	MSn MSe	MSn MSe		MSn
Arylphorin subunit beta.	MSn	MSn	MSn		MSn
βGRP p50-like (*B. mori*)		MSe		MSe	
Cationic peptide CP8.	MSn MSe	MSn MSe			MSn MSe
Elongation factor 1-alpha	MSn	MSn MSe			
Elongation factor 2. (*B. mori*)		MSn MSe			
HAIP	MSp MSe IB	MSp MSe IB	IB	MSp IB	MSp MSe IB
Hemolin.				n/d	MSn MSe IB
Histones			MSn MSe		
HP14	faint	MSp IB	MSp IB		
Immulectin-2				IB	IB
Immulectin-4					MSe
Immune-induced protein-1					MSe
Insecticyanin					MSn MSe
Leureptin.	IB	IB	IB	MSn MSe IB	IB
Lysozyme	MSn MSe IB	MSn IB	MSn IB	MSn MSe IB	MSn MSe IB
Mannose-binding serine protease-like					MSe
MBF2? (*Samia cynthia*)	MSe		MSe	MSe	
Methionine-rich storage protein 3.				MSn	
Multi-binding protein (*B. mori*)-like		MSe			

Protein	*S. aureus*	*B. thuringiensis*	*S. marcesens*	*E. coli*	Curdlan
Peroxiredoxin		MSn MSe			MSe
PGRP	n/d	MSn MSe IB	MSn MSe IB		
ProPO-1	MSn IB	MSn MSe IB	MSn MSe IB	MSn	MSn MSe
ProPO-2	MSn	MSn	MSn		MSn
Protein disulfide-isomerase like		MSe	MSn MSe		
RPS30 (*B. mori*)-like		MSe			
Serpin-1.					MSn MSe
SPH-1				n/d	MSn IB
SPH-2	IB	MSn MSe IB	IB		MSn
SPH-3					MSe
Thioredoxin-like protein					MSn MSe
Transferrin.					MSn
Transgelin. (*B. mori*)		MSn MSe			MSe
Ubiquitin & ribosomal protein s27a					MSe

Figures

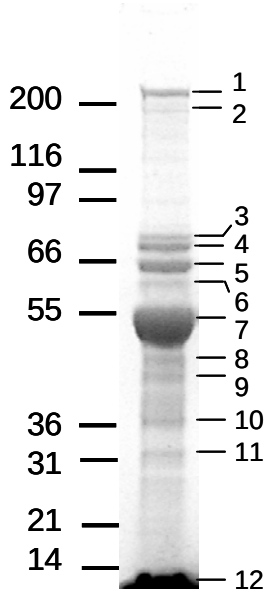


Figure 3-1. Plasma proteins from bar-stage prepupae which bound to curdlan.

SDS-PAGE (4-12 % bis-Tris gel) and Coomassie staining of plasma proteins eluted from curdlan. Numbers correspond to bands cut out for MALDI-TOF/TOF analysis (Table 3-1).

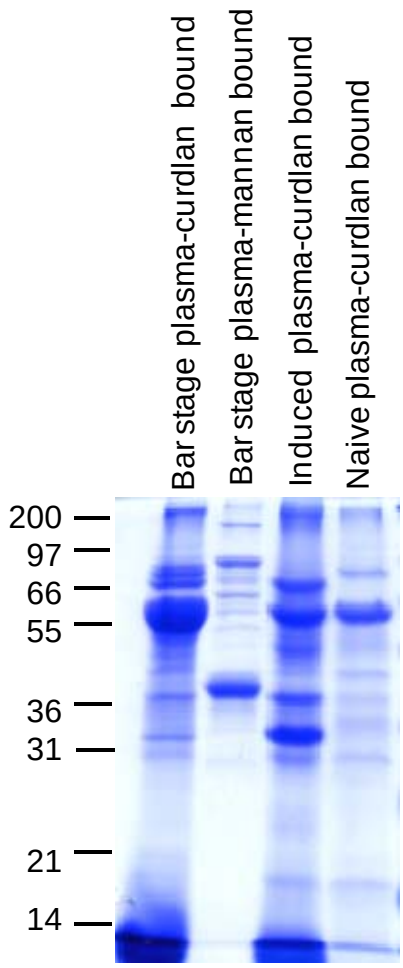
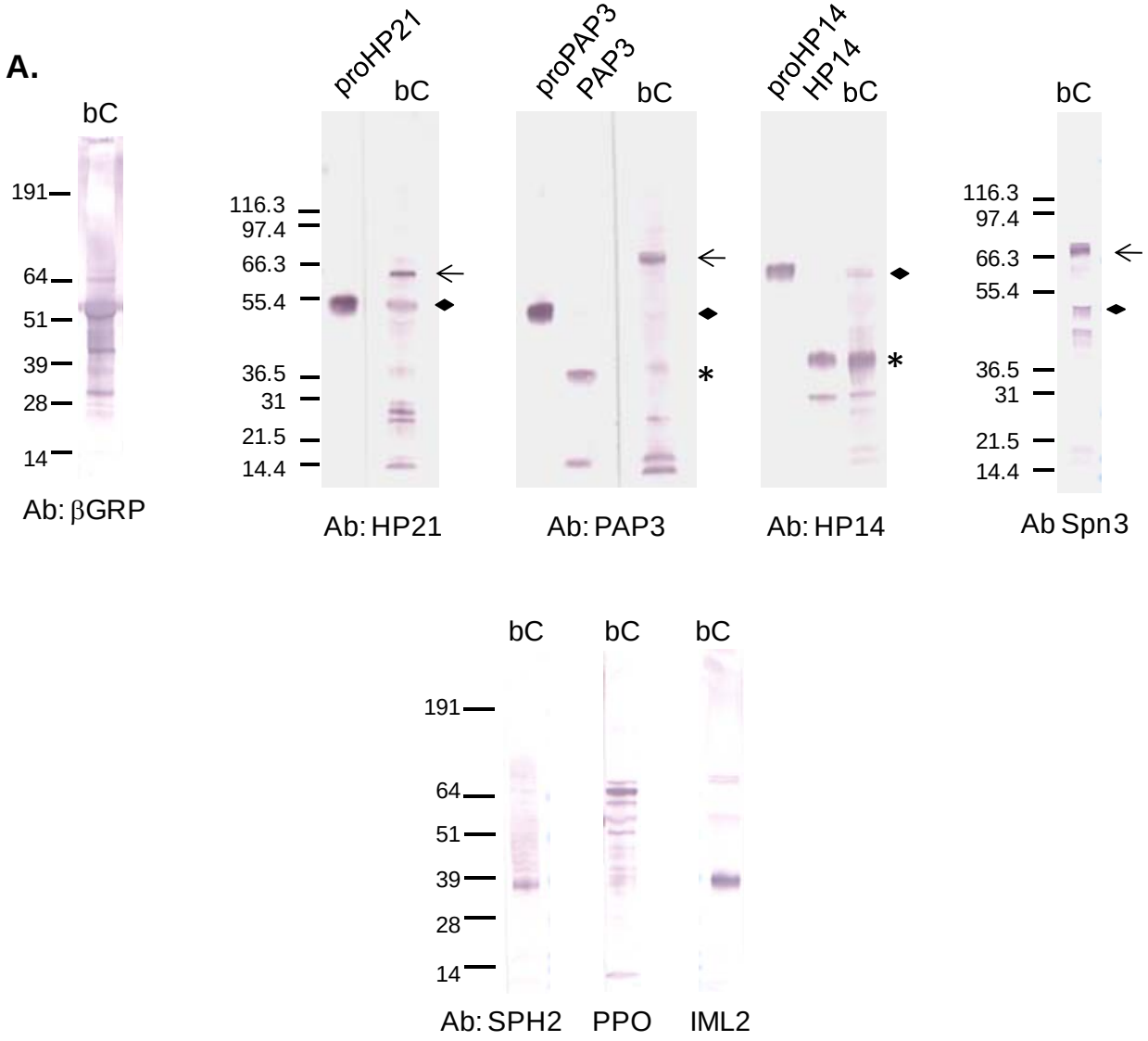
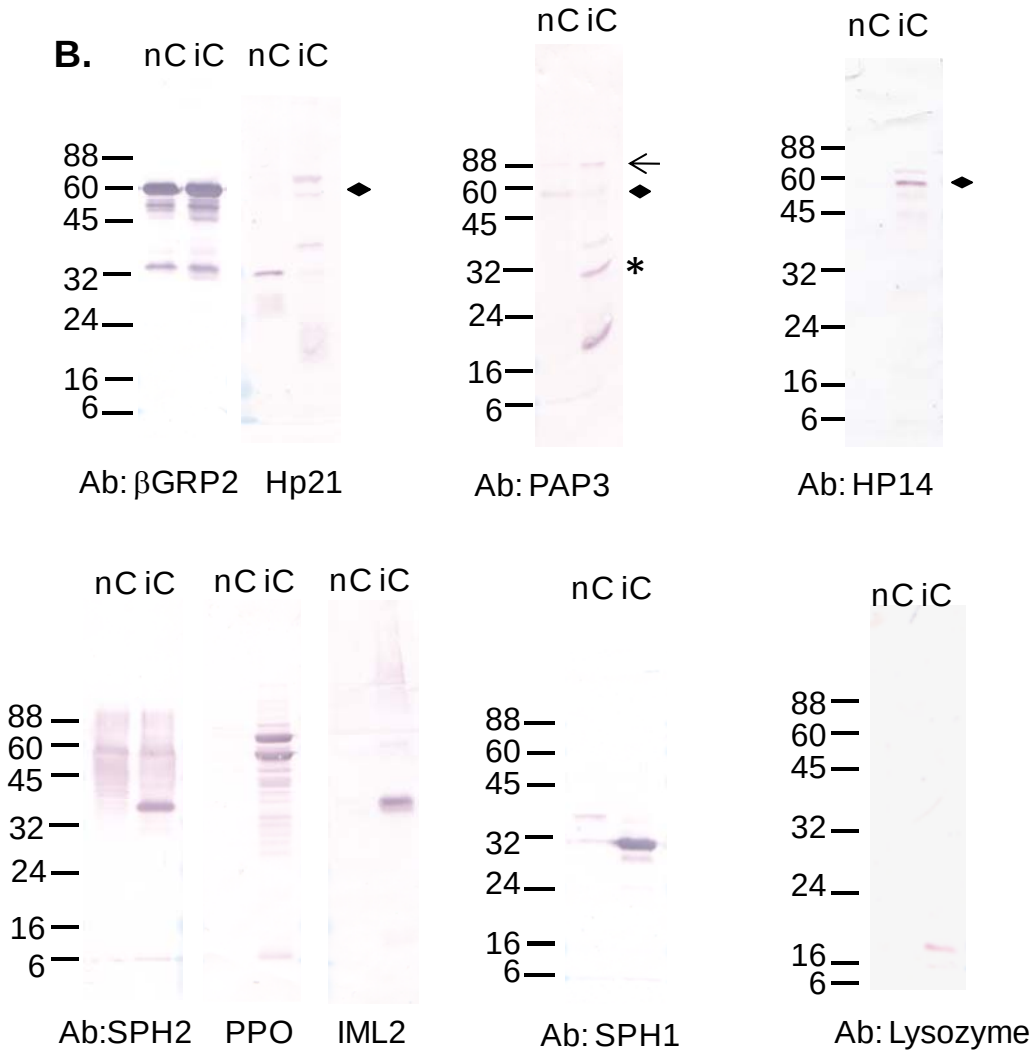


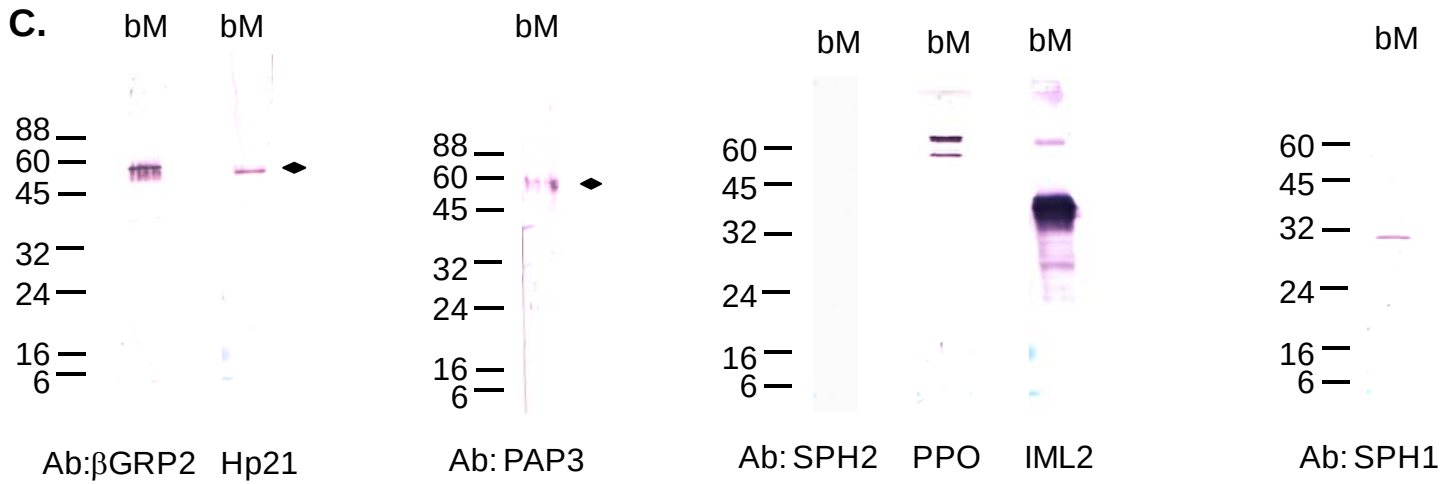
Figure 3-2. Analysis of plasma proteins bound to curdlan or mannan.
Separation by SDS-PAGE on a 10% gel.

Figure 3-3. Immunoblot analysis of *M. sexta* plasma proteins eluted from curdlan or mannan-agarose.

A) Plasma proteins from bar-stage prepupae, bound to curdlan (bC), after SDS-PAGE on a 4-12% acrylamide Bis-Tris gel. Some lanes are purified proteases either in their zymogen form (proHP21, proPAP3, proHP14) or activated form (PAP-3, HP14). B) Plasma proteins from naïve (nC) or induced (iC) 5th instar insects after binding curdlan. C) Bar-stage plasma proteins bound to mannan-agarose (bM). Blots B and C after SDS-PAGE on a 10% acrylamide gel. Full-length proteases or serpin are marked with a diamond. Putative serpin-protease complexes are marked with an arrow. Predicted protease catalytic domains are indicated by an asterisk.


A.
bC
191
64
51
39
28
14
Ab: βGRP
proHP21
bC
116.3
97.4
66.3
55.4
36.5
31
21.5
14.4
Ab: HP21
proPAP3
PAP3
bC
Ab: PAP3
proHP14
HP14
bC
*
Ab: HP14
bC
116.3
97.4
66.3
55.4
36.5
31
21.5
14.4
Ab Spn3
bC
bC
bC
191
64
51
39
28
14
Ab: SPH2
PPO
IML2


B.
nC iC
nC iC
88
60
45
32
24
16
6
Ab: βGRP2
Hp21
nC iC
Ab: PAP3
nC iC
Ab: HP14
nC iC
nC iC
nC iC
Ab:SPH2
PPO
IML2
nC iC
Ab: SPH1
nC iC
Ab: Lysozyme


C.
bM
bM
88
60
45
32
24
16
6
Ab:βGRP2
Hp21
bM
88
60
45
32
24
16
6
Ab: PAP3
bM
bM
bM
60
45
32
24
16
6
Ab: SPH2
PPO
IML2
bM
60
45
32
24
16
6
Ab: SPH1

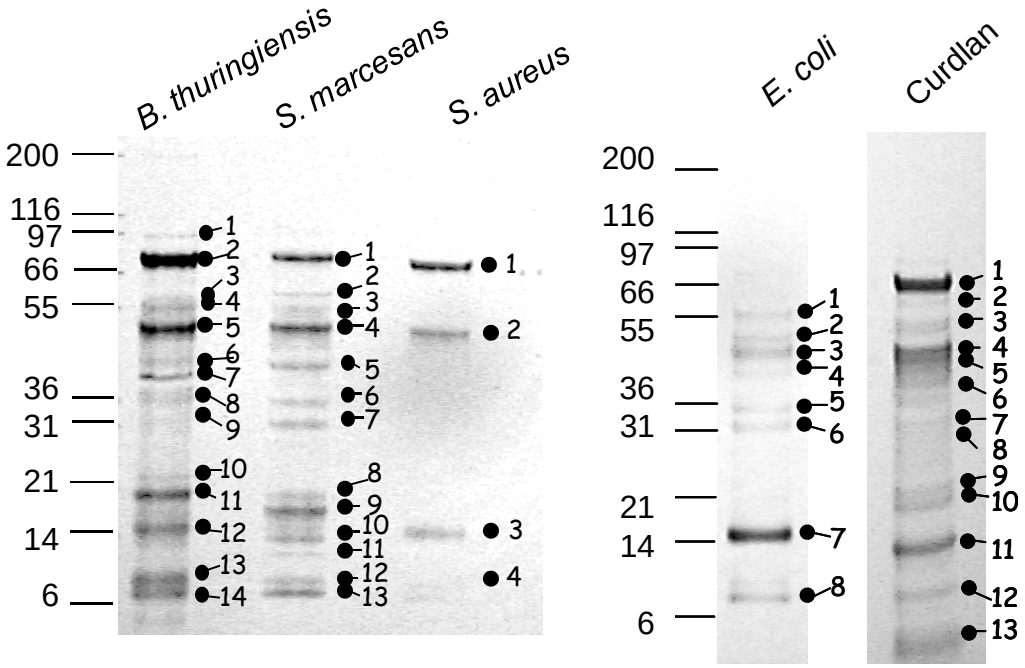
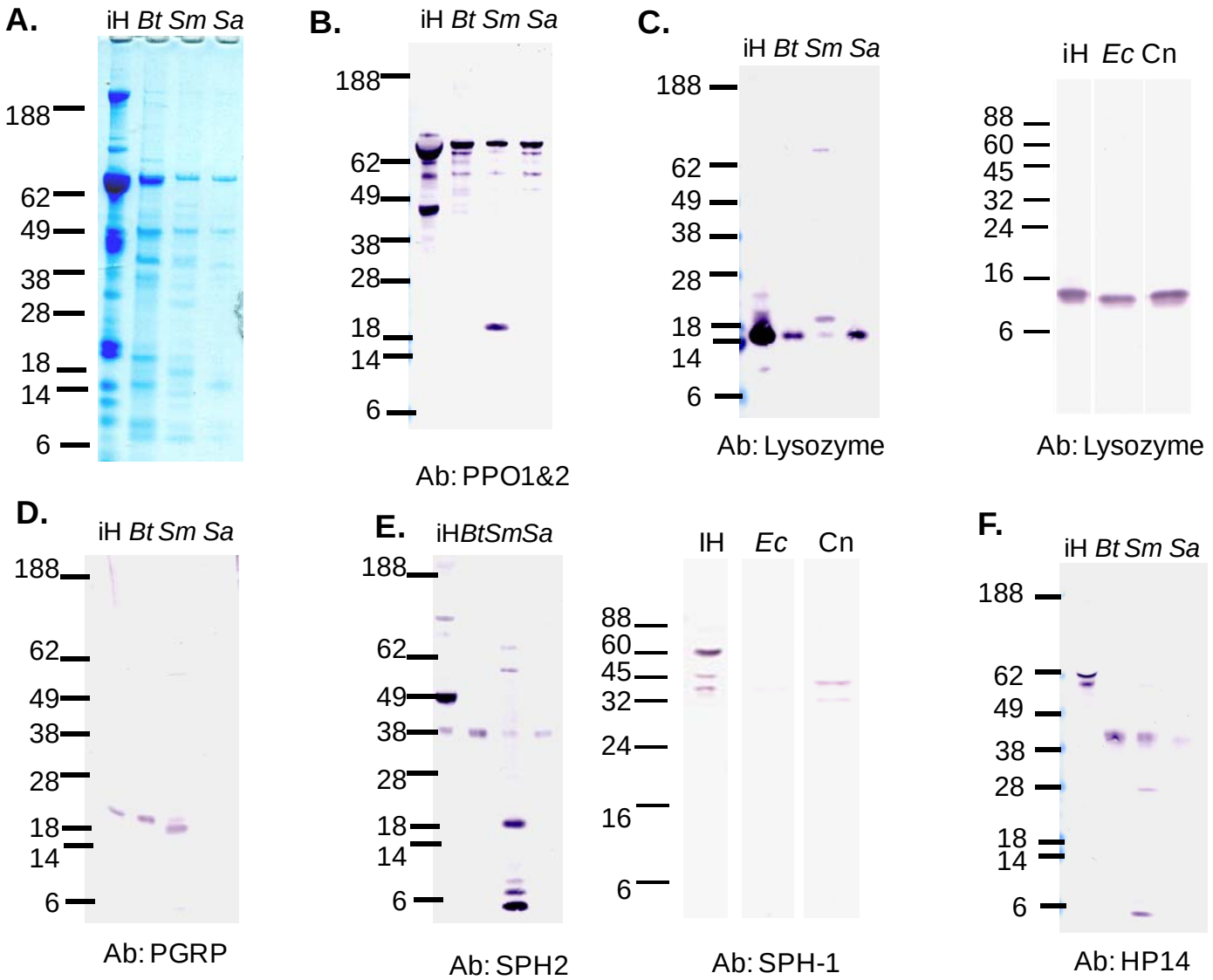


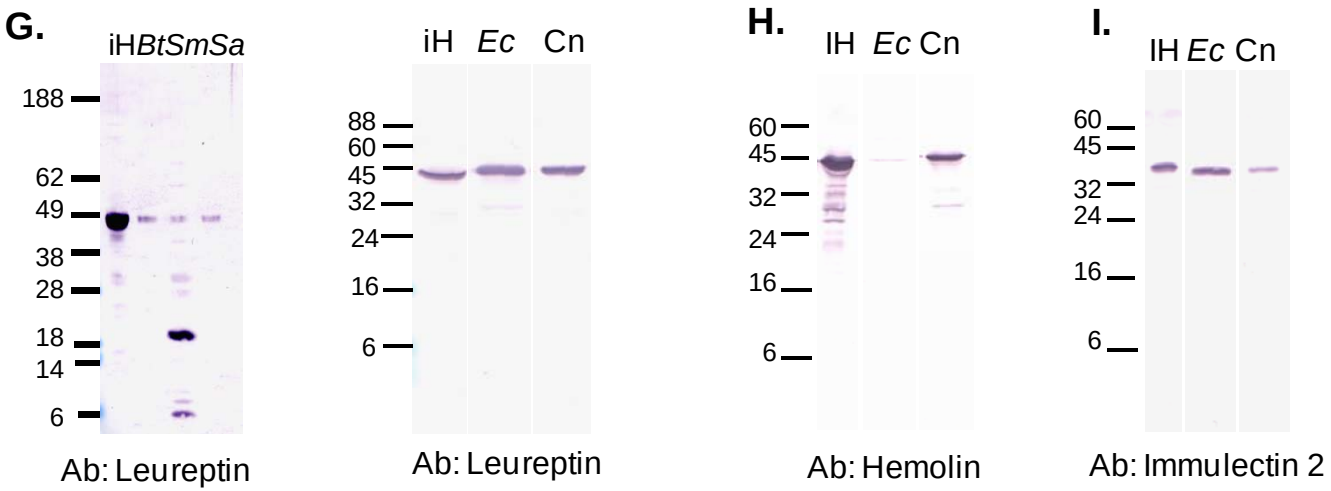
Figure 3-4. Hemolymph proteins recovered from *in vitro* bacterial binding assays.

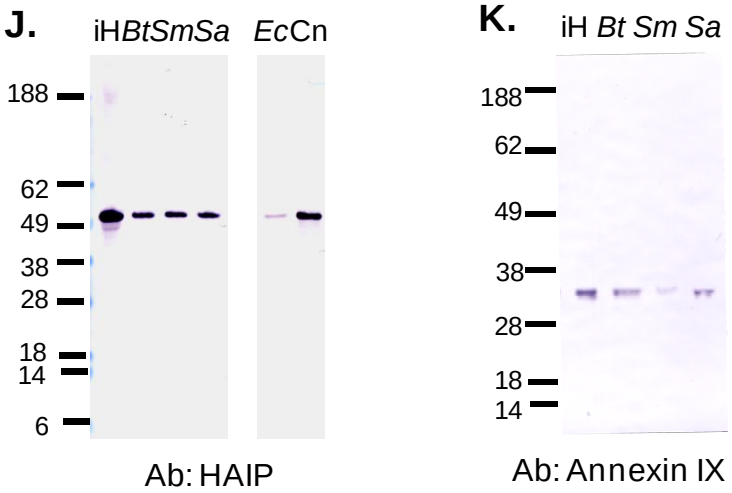
M. sexta bacteria-induced hemolymph was incubated with *B. thuringiensis*, *S. marcescans*, *S. aureus*, or with *E. coli* or curdlan from *A. faecalis*. After washing, bound proteins were eluted with 0.1 M ammonium acetate, 0.5 M NaCl at pH 4, and then separated by SDS-PAGE, followed by staining with Coomassie blue. Labeled bands were cut out and further analyzed by trypsin digestion and MALDI-TOF/TOF.

Figure 3-5. Coomassie stained gel and immunoblots from *in vitro* bacterial binding elution fractions.

A. Coomassie stained gel with induced hemolymph, prior to *in vitro* bacterial binding (positive control) and elution fractions from *B. thuringiensis*, *S. marcesans*, and *S. aureus*. Immunoblots: antiserum to prophenoloxidase (B), lysozyme (C), PGRP (D), SPH-2 (E left) and SPH-1 (E right), HP 14 (F), leureptin (G), hemolin (H), immulectin 2 (I), HAIP (J) and annexin IX (K).


A.
iH Bt Sm Sa
188
62
49
38
28
18
14
6
B.
iH Bt Sm Sa
188
62
49
38
28
18
14
6
Ab: PPO1&2
C.
iH Bt Sm Sa
188
62
49
38
28
18
14
6
Ab: Lysozyme
iH Ec Cn
88
60
45
32
24
16
6
Ab: Lysozyme
D.
iH Bt Sm Sa
188
62
49
38
28
18
14
6
Ab: PGRP
E.
iHBtSmSa
188
62
49
38
28
18
14
6
Ab: SPH2
IH Ec Cn
88
60
45
32
24
16
6
Ab: SPH-1
F.
iH Bt Sm Sa
188
62
49
38
28
18
14
6
Ab: HP14


G.
iH*BtSmSa*
188
62
49
38
28
18
14
6
Ab: Leureptin
iH *Ec* Cn
88
60
45
32
24
16
6
Ab: Leureptin
H.
IH *Ec* Cn
60
45
32
24
16
6
Ab: Hemolin
I.
IH *Ec* Cn
60
45
32
24
16
6
Ab: Immulectin 2


J.
iH*BtSmSa* *Ec*Cn
188
62
49
38
28
18
14
6
Ab: HAIP
K.
iH *Bt Sm Sa*
188
62
49
38
28
18
14
Ab: Annexin IX

CHAPTER 4 - Analysis of *Manduca sexta* serpin-1 isoforms

Introduction

Serpins are a superfamily of protease inhibitors found in all multicellular eukaryotes, some prokaryotes, and poxviruses (Law et al, 2006). Most serpins inhibit serine proteases although some serpins inhibit cysteine proteases and other serpins are noninhibitory. When serpins inhibit proteases the serpin structure changes dramatically. The serpin initially folds into a metastable state that includes 3 beta sheets, A, B, and C, and 17-25 residues which extend above the rest of the protein and form the reactive center loop. This reactive center loop is a bait for proteases. During protease attack the protease catalytic Ser forms an ester bond with the acyl group of the serpin P1 residue (Dementiev et al, 2006). After this cleavage between the P1 and P1' residues, the amino-terminal portion of the reactive center loop rapidly rearranges and inserts as a beta strand in the A beta sheet (Whisstock & Bottomley, 2006). This massive rearrangement results in the protease moving 70 angstroms away from where it started. The active site of the protease is distorted in this acyl-enzyme intermediate and prevented from reacting with water so the protease remains attached to the cleaved serpin (Dementiev et al, 2006). This form of the serpin, with the reactive site loop inserted in the A beta sheet, is more energetically favorable for the serpin than the native, metastable state in which it started (Whisstock & Bottomley, 2006).

Serine proteases are involved in blood cascades in mammals and insects, and extracellular serpins are candidates for modulating the activity of serine proteases in the hemolymph (blood). Serpins have been shown to inhibit serine proteases involved in melanization in *Manduca sexta* (Tong & Kanost, 2005, Wang & Jiang, 2004, Zhu et al, 2003), *Drosophila melanogaster* (De Gregorio et al, 2002), and *Anopheles gambiae* (Michel et al, 2006). *D. melanogaster* serpins also regulate serine proteases involved in Toll signaling during immune response and dorsal ventral embryonic patterning (Green et al, 2000, Hashimoto et al, 2003, Ligoxygakis et al, 2003).

Serpin-1 is present at high levels in *M. sexta* plasma, approximately 0.4 mg/ml (Kanost, 1990, Kanost et al, 1995). *M. sexta* serpin-1 has 12 different copies of exon 9

that undergo mutually exclusive alternative splicing to produce 12 different protein isoforms (Fig. 4-1A). These isoforms all differ in their carboxyl-terminal 39-46 residues, which include the P1 residue in the reactive center loop (Figure 4-1B, C, D) (Jiang et al, 1994, Jiang et al, 1996, Li et al, 1999). The 12 serpin-1 isoforms have different inhibitory selectivities for proteases, which has been tested by expressing all 12 isoforms and mixing them with various commercially available human or bacterial proteases (Jiang & Kanost, 1997). We can classify the 12 serpin-1 isoforms as likely to inhibit enzymes that are trypsin-like, chymotrypsin-like, or elastase-like based on their inhibitory activities and known or predicted P1 residue (Table 4-1). Of the over thirty known serine proteases in *M. sexta*, only two have been found to form serpin-protease complexes with serpin-1 isoforms. Serpin-1J was found to inhibit activation of the phenoloxidase pathway and forms a complex with PAP3 *in vitro* (Jiang & Kanost, 1997, Jiang et al, 2003) while serpin-1I can complex with HP14 *in vitro* (Wang & Jiang, 2006).

Bombyx mori antitrypsin and antichymotrypsin II were the first example of serpins that share a common sequence for most of the protein but differ in their C-termini, including the P1 residue (Sasaki, 1991). A serpin-1 homolog in another lepidopteran, *Mamestra configurata*, is 60% identical to *M. sexta* serpin-1 in their common regions. *M. configurata* serpin-1a has a reactive center loop that is 73% identical to serpin-1Z while McSerpin-1b and c share all but two residues with each other (they are perhaps different alleles) but their reactive center loops don't share sequence similarity with any of the *M. sexta* serpin-1 isoforms (Chamankhah et al, 2003). Quite recently *M. configurata* has been found to have a total of 9 serpin-1 variants (Hegedus et al. in press).

We know the mutually exclusive alternative splicing of serpin-1 creates 12 different transcripts from just one gene, but we do not yet understand how this process is regulated. Often alternative splicing is regulated depending on developmental stage or cell type and, in some cases, is influenced by external stimuli such as receptor stimulation and cellular stress (Stamm et al, 2005, Tarn, 2007). Interestingly, most vertebrate exons have three or fewer possible alternative exons, suggesting that many mechanisms of splicing regulation that have been studied in vertebrates may not directly apply to a situation with many mutually exclusive alternative exons, such as serpin-1 (Crayton et al,

2006, Graveley et al, 2004, Graveley, 2005). A notable example of mutually exclusive alternative splicing is *D. melanogaster* Dscam which has four alternatively spliced sites and the potential of generating 38, 016 different transcripts. Splicing in Dscam exon nine is differentially regulated during development and across different tissues (Neves et al, 2004) while the 48 exon 6 variants seem to be randomly selected (Olson et al, 2007).

Northern blot results showed serpin-1 expression in fat body and, less strongly, in hemocytes, but the mRNA was not detected in midgut or integument (Kanost et al, 1995). Analysis of 17 cDNA clones from hemocytes and 21 from fat body provided evidence of the expression of serpin-1 A-G, J, K, and Z (Jiang et al, 1994). The hemocyte clones were well distributed over the different isoforms but 19 of the 21 fat body clones were serpin-1F. This suggested a fat body specific preference for isoform F. These results led us to wonder if all 12 serpin-1 isoforms are expressed in fat body and hemocytes and about the relative abundance of the serpin-1 mRNA in the tissues and the protein isoforms in the hemolymph. We used immunoaffinity chromatography to purify serpin-1 from plasma, which was then separated by 2D-PAGE. Serpin-1 isoforms were identified by MALDI-TOF/TOF analysis of proteolytic fragments. Additionally, we have identified some naturally occurring serpin-1 protease complexes that provide insight into some of the endogenous *M. sexta* proteases that serpin-1 can inhibit, bringing us closer to our ultimate goal of understanding the function of serpins and proteases in hemolymph of *M. sexta* and other insects.

Materials and Methods

Insects and rearing

We originally obtained eggs for the *M. sexta* colony from Carolina Biological Supply. The insects are reared on artificial diet (Dunn & Drake, 1983).

Primer design and PCR

Primers were designed using the primer 3 program (Invitrogen). RNA was purified from midguts of naïve fifth-instar larvae as described previously (Dittmer et al, 2004). A RNeasy Midi Kit (Qiagen) was used to extract RNA from hemocytes or fat body of fifth instar larvae from both naïve insects and insects 24 hours after injection of

100 µl of 1 mg/ml *Micrococcus luteus*. RNA was treated with Turbo DNA-free (Ambion) to remove any contaminating genomic DNA. cDNA was synthesized in 20 µl reactions with the Superscript III kit using an oligo-(dT) primer (Invitrogen). For both fat body samples, 5.36 µg RNA was used while 1.18 µg RNA was used for naïve hemocytes and 2.06 µg RNA for induced hemocytes. PCR was performed using 0.5 µl of midgut, naïve fat body or induced fat body cDNA, 1 µl of naïve hemocyte cDNA, or 0.6 µl of induced hemocyte cDNA with 0.5 µl forward primer, 0.5 µl reverse primer, and 22.5 µl Platinum PCR Supermix (Invitrogen) in a total volume of 25 µl. PCR reactions were run for 30 or 35 cycles (30 s at 94°C, 30 s at 50°C and 25 s at 72°C). The PCR products were analyzed by electrophoresis on a 1% agarose gel and stained with ethidium bromide. The PCR products from the 35 cycle naïve hemocyte reactions were excised from the gel, purified (Qiagen kit), and sequenced at the DNA sequencing facility (IA State) using the sequence specific primers listed in Table 4-1.

Collection of hemolymph

Fifth instar larvae were chilled on ice for approximately 10 min. One of the second or third prolegs of each larva was snipped with scissors, and hemolymph was collected into a sterile tube containing 2.5 mg diethylthiocarbamate (DETC) per larva. Larvae were held in a U-shape and gently massaged with fingers until hemolymph stopped dripping. Hemolymph was then centrifuged at 10,000×*g* for 15 min to remove hemocytes. Samples were either used immediately or frozen in aliquots at -80°C.

For one experiment, 4 ml of plasma from naïve day 1 fifth instar larvae was incubated with either 50 µl of 40 µg/ml *M. luteus* (Sigma M3770) or 100 µl of 10 mg/ml LPS (Sigma L2630) for 40 min. The LPS-treated plasma was subsequently used for immunoaffinity of serpin-1 followed by 1D-PAGE. All other immunoaffinity isolation purifications of serpin-1 used untreated naïve plasma.

In vitro *culture of fat body*

Fat body was dissected from three fifth instar larvae. Fat body from each individual was placed in a well of a 16 well culture plate and washed twice with 1 ml of SF 900 serum free media (Invitrogen). Fat body was incubated for 18 h in 1.3 ml of SF

900 media supplemented with liquid penicillin G and streptomycin sulfate (Invitrogen) at final volumes of 100 U/ml and 100 μg/ml, respectively.

Preparation of serpin-1 antibody column

Recombinant his-tagged serpin-1I (Jiang & Kanost, 1997) was expressed in *E. coli,* purified by nickel-NTA chromatography and SDS-PAGE, and sent to Cocalico (Reamstown PA) for injection into a rabbit. The serum from the final bleed from this rabbit was labeled KSU 178. For the serpin-1-antibody column, 2.7 ml of protein-A bead slurry (Sigma P9424) was prewashed two times with 12 ml phosphate buffered saline (PBS, 0.01 M phosphate buffer, pH 7.4, 2.7 mM potassium phosphate, and 0.137 M sodium chloride) and centrifuged at 100×*g* for 5 min. The protein-A beads (2 ml) and serpin-1 antiserum (2 ml) were mixed end-over-end for 2 h at room temperature and washed two times with PBS. The serpin-1 antibody beads were next washed twice with 12 ml of 0.2 M sodium borate, pH 9 and then resuspended in 12 ml of that same buffer. Eighty mg of the cross-linker dimethylpimelimidate dihydrochloride (Sigma D8388) was added and the beads were mixed for 1 h. The beads were centrifuged for 5 min, 100×*g*, then washed once with 0.2 M ethanolamine pH 8, and mixed overnight in 12 ml of 0.2 M ethanolamine pH 8. The beads were washed twice with PBS, and then 11.5 ml of elution buffer (100 mM glycine, 10% ethylene glycol, pH 3) was mixed with the serpin-1 antibody beads as a pre-elution step to remove any uncoupled antibody. The beads were washed three times with PBS. Serpin-1 antibody beads were stored in PBS containing 0.05% sodium azide at 4°C.

Immunoaffinity purification of serpin-1 isoforms

Serpin-1 antibody beads were mixed in a 15 ml tube with 4 ml of day 1 and 2 fifth instar plasma in 8 ml PBS for 2 h. The beads were centrifuged at 100×*g* and the plasma was removed with a transfer pipette. Beads were washed three times with 12 ml PBS and then transferred to a 10 ml disposable chromatography column (BioRad). The column was filled with PBS for a fourth wash and then eluted with ten 1 ml applications of elution buffer (100 mM glycine, 10% ethylene glycol, pH 3). Elution fractions were collected into 100 μl aliquots of 1 M sodium phosphate pH 8 for neutralization. After 10 ml of elution buffer, 3 ml of PBS was added for elutions 11-13. The column was washed

with at least 10 ml of PBS before being equilibrated with PBS containing 0.05% sodium azide for storage.

SDS-PAGE, silver staining and immunoblotting

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed using 4-12% bis-Tris NuPAGE gels (Invitrogen). Protein samples were mixed with 2x SDS sample buffer containing β-mercaptoethanol and heated at 95°C for 5 min. Gels were run in MOPS or MES buffer (Invitrogen) at 200 V until the dye front reached the bottom of the gel. A protein standard, either the unstained Mark12 or the pre-stained SeeBlue (Invitrogen), was run alongside the samples.

For silver staining, gels were incubated in 50 ml of fixing solution (40% methanol, 0.05% formaldehyde) for at least 20 min on a shaker. Gels were then washed twice for 5 min in water. Fifty ml of 0.02% sodium thiosulfate was added for 1 min and then the gels were washed twice in water for 20 s. Gels were incubated in 0.1% silver nitrate for 10 min. After three 20 s water washes, developing solution (3% sodium carbonate and 0.05% formaldehyde) was added and gels were allowed to develop until the desired band intensity appeared. Development was stopped by adding about 5 ml of 2.3 M citric acid. Gels were incubated in gel drying solution (20% ethanol, 5% glycerol) for at least 30 min before drying between wet cellophane sheets overnight.

Gels for immunoblotting were incubated briefly in western transfer buffer (20% methanol, 48 mM Tris, 40 mM glycine, and 1.3 mM SDS). Proteins from the gels were transferred to 0.45 μ nitrocellulose membranes with a Trans-Blot Semi-Dry Transfer Cell (BioRad) at 10-12 V for 1 h. Blots with Mark12 unstained standards were first stained with 0.25% Ponceau S (Sigma) in 0.75% acetic acid for 5 min and destained briefly in water. Positions of markers were marked in pencil followed by further destaining in water. Positions of spots 4 and 6 were also marked on the two blots for use with the serpin-1B and serpin-1F antiserum. The membranes were blocked with 3% dry milk in TTBS (0.05% Tween 20 in tris buffered saline, 25 mM Tris-HCl, pH 7.4, 137 mM NaCl, 2.7 mM KCl) for 30 min. After blocking, primary antiserum was added and incubated with the blot for an additional 2 h. For most primary antiserum a 1:2000 dilution was used, including antiserum to serpin-1 (KSU 178), HP1, HP6, HP8, PAP3 and PPO1&2.

The antiserum to HP6 and HP8 was generously provided by Dr. Haobo Jiang, Oklahoma State University. Rabbit antisera against serpin-1 isoform-specific peptides were also used (Yang Wang and Mike Kanost, unpublished data). The antiserum to serpin-1B was used at a 1:500 dilution and the antiserum to serpin-1F was used at 1:3,000 dilution. After incubation with primary antibody, membranes were washed three times with TTBS and then incubated for 1 h with a 1:3000 dilution of the secondary antibody (goat-anti-rabbit IgG conjugated with alkaline phosphatase, BioRad) in TTBS with 1% dry milk. After two washes with TTBS and one wash with TBS, the membrane was developed using an Alkaline Phosphatase Conjugate Substrate Kit (BioRad).

Protein concentration assay

Protein concentrations were determined using the Coomassie Plus Protein Assay Reagent (Pierce) with a bovine serum albumin standard as directed by the manufacturer.

Concentration of protein samples

Affinity purified serpin-1 elution fractions with protein concentrations of 50 μg/ml protein or less were concentrated using a Centriplus YM-3 centrifugal filter device (Amicon) at 3,000×*g*, 4°C, until the volume was reduced from 9 ml to 2 ml and the concentration of the sample was ~110 μg/ml. These 2 ml were further concentrated using two Microcon YM3 centrifugal filter units (Amicon) at 13,000×*g* at 4°C. Four elution fractions that had higher initial protein concentrations (60-130 μg/ml) were also individually concentrated in Microcon YM3 centrifugal filter units. The retentates were pooled and used for 2D SDS-PAGE.

Sample preparation for 2D SDS-PAGE

Whole *M. sexta* plasma (16 μl, from three day 3 fifth instar larvae), SF 900 media from cultured fat body (55 μl from well 1 and 45 μl of well 2, pooled together), or concentrated affinity purified serpin-1 (90 μl) was prepared for isoelectric focusing using the ReadyPrep 2-D Cleanup Kit (BioRad) according to the manufacturer's instructions. Briefly, serpin-1, media, or plasma was mixed with proteomic grade water (BioRad) to make 100 μl total volume in a 1.5 ml tube and vortexed together with 300 μl of

precipitating agent 1. After 15 min on ice, 300 µl of precipitating agent 2 was added and the sample was vortexed. The tube was then centrifuged at 14,000×*g*, 4°C, for 5 min. The supernatant was removed, and the pellet was washed with 40 µl of wash reagent 1, then centrifuged again at 14,000×*g*, 4°C, for 5 min. The liquid was removed and 25 µl of proteomic grade water (BioRad) was added. After vortexing, 1 ml of wash reagent 2 (stored at -20°C) and 5 µl of wash 2 additive were added to the pellet, and the tube was vortex mixed for 1 min. The tube was placed in the -20°C freezer for 30 min with the exception of vortex mixing for 30 s at two 10 min intervals. The mixture was centrifuged at 13,000×*g*, 4°C, for 5 min and the supernatant was removed. The pellet was air dried for about 3 min before addition of 190 µl of a strongly chaotropic buffer (7 M urea, 2 M thiourea, 4% CHAPS, 50 mM DTT, 0.2% Bio-Lyte 4.7/5.9 ampholytes (BioRad), 0.002% bromophenol blue). The tube was vortex mixed for 30 s, incubated at room temperature 5 min, vortexed again 1 min, and centrifuged at 14,000×*g* for 5 min. The sample was used immediately for rehydrating an IEF strip.

Isoelectric Focusing (IEF)

Eleven cm IEF strips with a 4.7-5.9 pH range (BioRad) were rehydrated overnight in a disposable equilibration/rehydration tray (BioRad). The protein sample (190 µl) in strongly chaotropic buffer was pipetted along the length of the tray. The plastic backing was removed from the IEF strip, which was placed gel surface down in the sample. After 0.5 h, 1 ml of mineral oil was layered on top of the strip and incubated for 16 h at room temperature.

The strip was removed from the rehydration tray, and the oil was drained off. The strip was rested plastic side down on a paper towel to further remove any remaining mineral oil and unabsorbed sample. Filter paper pieces moistened with water were placed over the wires on each end of a well in an 11 cm IEF tray (BioRad). The strip was placed gel side down in the tray with the ends covering the filter paper and the strip end with a + placed to the left. IEF was conducted using the following parameters: Step 1, 250 V with a 15 min rapid ramp; Step 2, 8,000 V for 1 h slow ramp; Step 3 30,000 V-h rapid ramp; Step 4 500 V hold for 24 h. The run was complete after step 3 and stopped during step 4.

Second dimension SDS-PAGE

After completion of IEF, the strip was removed to a disposable equilibration tray. The strip was washed first with 2% dithiothreitol in equilibration buffer (6 M urea, 2% SDS, 0.05 M Tris/HCl pH 8.8, 20% glycerol) for 12 min and then drained and washed in equilibration buffer with 2.5% iodoacetamide for 12 min.

The strip was then rinsed in MOPS SDS running buffer and placed in the IPG well of an 4-12% bis-Tris Criterion gel (BioRad). Melted agarose overlay [0.5% agarose in 1% MOPS buffer (Invitrogen) with 0.002% bromophenol blue] was added on top of the strip. The gel box was filled with MOPS buffer (Invitrogen), 5 μl of Mark12 protein standard (Invitrogen) was loaded, and the gel was electrophoresed at 200 V until the dye reached the bottom of the gel.

Immunoblotting was performed as explained previously. Gels for Coomassie blue staining were washed three times, at least 5 min each, in water and then incubated in BioSafe Coomassie G-250 stain (BioRad) for at least 1 h or overnight. Destaining was performed with at least three changes of water.

Protein digestion

The 2D spots from whole plasma were digested with trypsin and analyzed by MALDI-TOF/TOF at the Biochemical Research Services Laboratory (BRSL, University of Kansas). The remaining bands and spots were analyzed by mass spectrometry at the Nevada Proteomics Center (University of Nevada at Reno) and detailed information about those methods is described as follows. Selected protein spots were digested in bicarbonate buffer with trypsin or a double digest of lysine-C and aspartic-N (LysC/AspN). Other spots were digested by glutamic-C (GluC) in either bicarbonate or phosphate buffer. Trypsin digestion followed the protocol previously described with some modifications (Rosenfeld et al, 1992). Samples were washed twice with 25 mM ammonium bicarbonate (pH 7.8) and 100% acetonitrile, reduced and alkylated using 10 mM dithiothreitol and 100 mM iodoacetamide, and incubated with digestive enzyme. Samples for trypsin digestion were mixed with 75 ng sequencing grade modified porcine trypsin (Promega) in 25 mM ammonium bicarbonate for 6 h at 37^{o}C. For the Lys C/AspN double digest, samples were incubated with 150 ng sequencing grade modified

LysC (Princeton Separations Inc) in 100 mM ammonium bicarbonate for 6 h at 30°C followed by 40 ng endoproteinase Asp-N (Sigma-Aldrich) in 100mM ammonium bicarbonate for 12 h at 37°C. GluC digestions were performed in two different buffers. For GluC bicarbonate, 150 ng of sequencing grade modified GluC (Princeton Separations Inc) was incubated with samples in 25 mM ammonium bicarbonate for 6 h at 30°C. For GluC phosphate, samples were incubated with 150 ng sequencing grade modified Glutamic-C (Princeton Separations Inc) in 50 mM sodium phosphate pH 7.8 for 4 h at 30°C.

Mass spectrometry

MALDI TOF/TOF analysis was performed as described in Chapter 2. Each trypsin sample was internally calibrated on trypsin's autolysis peaks 842.51 and 2211.10 to within 20 ppm. Any trypsin sample failing to internally calibrate and samples from other digestions were analyzed under default plate calibration conditions of 150 ppm. For the samples digested with trypsin, certain known trypsin and background masses were excluded from MS/MS analysis: 842.51, 870.54, 1045.56, 1126.56, 1420.72, 1531.84, 1940.94, 2003.07, 2211.10, 2225.12, 2239.14, 2283.18, 2299.18, 2678.38, 2807.31, 2914.51, 3094.62, 3337.76, and 3353.75. Similarly, selected masses were excluded from the LysC/AspN MS/MS analysis: 1192.69, 1716.87, 1743.83, 801.45, 1601.76, 1974.00, 2187.07, 2818.36, 1072.53, 1120.65, 1064.56, and 971.56.

MALDI data analysis

The raw data generated from the ABI 4700 were extracted, and a peak list was created by GPS Explorer software v 3.6 (Applied Biosystems). Analyses were performed as combination MS + MS/MS. MS peak filtering included mass range 700-4000 Da, minimum S/N filter 10. A peak density filter of 50 peaks per 200 Da with a maximum number of peaks set to 65. MSMS peak filtering included mass range of 60 Da to 20 Da below each precursor mass. Minimum S/N filter 10, peak density filter of 50 peaks per 200 Da, cluster area filter used with maximum number of peaks 65. The filtered data were searched by Mascot v 1.9.05 (Matrix Science) using NCBI nr database (NCBI 20070908), containing 5,454,477 sequences. Searches were performed without restriction to protein species, M_r, or pI and with variable oxidation of methionine residues and

carbamidomethylation of cysteines. All of the digests had the maximum missed cleavages set to 1. For trypsin digests, precursor mass tolerance and fragment mass tolerance were set to 150 ppm (or 20 ppm, if internally calibrated) and ± 0.2 Da. Trypsin cleavages were limited to C-terminal K or R (not KP or KR). For the LysC/AspN double digest, cleavages were limited to C-terminal K and N-terminal D cleavage sites. Precursor mass tolerance and fragment mass tolerance were set to 150 ppm and ± 0.4 Da, respectively. For Glu-C digests, precursor mass tolerance and fragment mass tolerance were set to 150 ppm and ± 0.2 Da, respectively. Cleavages were limited to glutamic-C (C-terminal DE) cleavage sites in the phosphate buffer and glutamic-C (C-terminal E) cleavage sites in the bicarbonate buffer.

MS peak list masses and intensities were also used to search the Swiss-Prot and TrEMBL databases using Aldente (http://ca.expasy.org/tools/aldente/) in the fall of 2007. The default settings were used with the following exceptions: predefined taxon was set to "other insecta," MW max was 100,000 Da, and for samples digested with trypsin and calibrated to 20 ppm, the spectrometer shift max was set at 0.05 Da and the slope max to 50 ppm. Default spectrometer settings were used for other enzyme digestions and the trypsin digestions done at the University of Kansas. Individual MS/MS peaks from MALDI-TOF/TOF were also analyzed using Mascot MS/MS Ion search (http://www.matrixscience.com) using peptide tolerance of 20 ppm, MS/MS mass tolerance of 0.2 Da, peptide charge 1+, carbamidomethylation of cysteine as a fixed modification and methionine oxidation as a variable modification. For the data from the University of Kansas, the default peptide tolerance (1.2 Da) and MS/MS tolerance (0.6 Da) were used.

MALDI data analysis peak lists were also created using ABI's 4000 Series Explorer software v. 3.6 Peaks to MASCOT feature. MS peak filtering settings included mass range 700-4000 Da, minimum S/N filter 0, and peak density filter of 65 peaks per 200 Da with a maximum number of peaks set to 200. MSMS peak filtering included mass range of 60 Da to 20 Da below each precursor mass, minimum S/N filter 0, peak density filter of 65 peaks per 200 Da, and cluster area filter with maximum number of peaks 200. The filtered data were searched by Mascot v 2.1.03 (Matrix Science) using NCBI nr database (NCBI 20070908), containing 5,454,477 sequences. Searches were

performed without restriction to protein species, M_r, or pI and with variable oxidation of methionine residues and carbamidomethylation of cysteines. Maximum missed cleavage was set to 1 and limited to appropriate digestive enzyme and buffer cleavage sites. Precursor mass tolerance and fragment mass tolerance were set to 150 ppm (or 20 ppm) and ± 0.2 Da. These files were then analyzed using Proteome Software's Scaffold software.

Results

Tissue expression of specific isoforms by reverse transcriptase-PCR

In *M. sexta* serpin-1, mutually exclusive alternative splicing of exon 9 leads to production of twelve different mRNAs in hemocytes (Jiang et al, 1996). We wanted to know more about the relative expression of individual serpin-1 isoforms in different tissues and after an immune challenge. I designed isoform-specific forward primers to the 12 alternatively spliced versions of exon 9 and a common reverse primer in exon 10, which encodes the 3' untranslated region (Table 4-2). PCR with cDNA from hemocytes of naïve larvae and from larvae injected with bacteria (induced) showed bands of the expected size for all 12 isoforms (Fig. 4-2 A and B). The products from naïve hemocytes were sequenced using the primers indicated in Table 4-1 and found to be the expected isoforms (results not shown), demonstrating that the exon 9 primers specifically amplified the intended cDNA isoforms. Transcripts for all 12 serpin-1 isoforms were also detected in naïve and induced fat body, and midgut samples (Fig. 4-2 C-E). The isoform expression levels within the tissues are similar, suggesting that regulation of splice site selection is similar in fat body, hemocytes, and midgut.

One transcript, serpin-1A, was more abundant in both hemocytes and fat body 24 h after immune challenge than in controls. This difference can be seen most clearly in Fig. 4-2 F, where naïve and bacteria-induced samples corresponding to isoforms A, B, and Z are run next to each other. Isoform A was more intense in the induced hemocyte and fat body lanes, while the intensities of B and Z are quite similar despite immune challenge.

Immunoaffinity purification of serpin-1 and serpin-1 protease complexes from plasma

Serine protease cascades are part of the insect innate immune responses (proPO and spätzle activation) and are triggered in response to immune challenge. Serine proteases are likely involved in other physiological processes as well and are inhibited by serpins. We are interested in identifying endogenous proteases that complex with serpin-1 and identifying individual serpin-1 isoforms in plasma. To obtain a more concentrated and pure pool of serpin-1, we immunoaffinity purified serpin-1 from the plasma. Immunoaffinity purification of *M. sexta* serpins followed by MALDI-TOF has been used in previous studies to detect covalent serpin-protease complexes (Tong et al, 2005, Zou & Jiang, 2005). The polyclonal antibodies used in this study recognize all 12 isoforms because the majority of the serpin-1 sequence is identical in all of the isoforms.

When we compared plasma activated with *M. luteus* or LPS with control plasma by immunoblot, a putative serpin-1 protease complex band appeared more abundant in the *M. luteus* and LPS activated samples (Fig. 4-3A). To identify the serpin isoforms and proteases in these complexes, we isolated them by imunoaffinity chromatography. Four ml of LPS activated plasma were mixed with the serpin-1 antibody resin. Elution fractions were analyzed by 1D-SDS-PAGE and detected by immunoblotting with serpin-1 antibody, Coomassie blue staining, or silver staining (Fig. 4-3 B, C, and D). Bands at ~47 kDa (diamond) are full-length serpin-1 isoforms. Also observed were shorter, cleaved forms of serpin-1 (circle) and higher molecular weight bands suspected to be SDS stable serpin-1-protease complexes. The bands labeled 1, 2, and 3 were cut from the Coomassie stained gel for trypsin digestion and MALDI-TOF/TOF mass spectrometry analysis.

Peptide mass fingerprinting (PMF) of MALDI-TOF data was carried out using Aldente software, and serpin-1 isoform specific peaks subjected to MS/MS were analyzed by Mascot (Table 4-3). PMF identified the two prophenoloxidase proteins (proPO), especially proPO1, in band 1 and serpin-1 in bands 2 and 3. ProPO1 peptides in band 1 were validated and the remaining peptides in band 1 were searched against the database a second time. Band 1 was found to also contain proPO2 and, after a second round of validation, serpin-1 had a score that was just above the threshold. The faint

presence of serpin-1 may explain the weak antibody staining with this top band that is mostly proPO. Band 1's size, ~82 kDa, is consistent with the expected size for full length proPO1 (80 kDa) and proPO2 (78 kDa).

Band 2, with an estimated mass of 77 kDa, contained multiple peptides expected for serpin-1 as well as isoform-specific peaks for both serpin-1J and serpin-1A (Fig. 4-4 A). The presence of serpin-1J was further confirmed by Mascot analysis of MALDI-TOF/TOF data. Serpin-1J was the top hit for the 2422 Da peptide MS/MS spectra (Fig. 4-5 C) while the MS/MS spectra of the 2341 Da peptide was insufficient for confirmation of identity as serpin-1A (e-value of 19, spectra not shown). Additional searches after serpin-1 validation did not identify a protease that might be complexed with serpin-1J or A.

Peptide mass fingerprinting of band 3 identified serpin-1K and, after validation of serpin-1K, multiple peaks for *M. sexta* chymotrypsin (Peterson et al, 1995) were detected. The three most intense chymotrypsin peaks are marked on the spectrum (Fig. 4-4 B). An isoform-specific peptide for serpin-1K was further confirmed by MS/MS (Fig. 4-4 D). The detected chymotrypsin peptides do not include the activation peptide, as expected for an active protease complexed with a serpin (Fig. 4-5 A). The estimated mass of band three is 67 kDa, very similar to the calculated mass of a complex of serpin-1K (38.1 kDa without signal peptide and after protease cleavage at P1') with activated chymotrypsin (24.8 kDa).

Serpin-1 isoforms A, D, E, J, and K have an isoform-specific tryptic peptide prior to or ending at the P1' residue (Fig. 4-1). The remaining serpin-1 isoforms lack an isoform specific tryptic peptide prior to the P1' residue and therefore we cannot identify those isoforms from a serpin-1 protease complex because the region of serpin-1 from P1' to the carboxyl termini is lost when the P1-P1' bond is cleaved by the target protease. We are thus unable to rule out the presence of additional serpin-1 isoforms in either band 2 or 3.

2D gel analysis of serpin-1 isoforms in whole plasma

To analyze relative serpin-1 isoform abundance at the protein level and identify the serpin-1 isoforms present in plasma, we separated serpin-1 isoforms by 2D-PAGE.

The isoelectric point and mass of the serpin-1 isoform proteins were calculated excluding the signal peptide (cleaved after Ala16 (Kanost et al, 1989)) and ignoring the possible glycosylation at Asn85. All of the isoforms were plotted together on a graph showing expected migration on an isoelectric focusing strip with pH range of 4.7-5.9 followed by SDS-PAGE (Fig. 4-6). All 12 isoforms have very similar molecular weights, and very little isoform separation is possible on the basis of size. However, the sequence differences in the variable carboxyl-terminal region do result in differences in charge that should allow separation of most forms by isoelectric focusing.

The isoform predicted to be the most acidic is 1J, at an expected pI of 4.88. Several pairs of isoforms are predicted to migrate very close together. Isoforms C and H are predicted right next to each other, F and B are slightly more basic and overlap at pI of ~4.95, and I and G overlap at about 5.05. Next we predict the presence of isoform Z (5.17) and isoform D (5.36). Isoforms K and A are also predicted to be near to one another, at pI 5.44 and 5.46, while serpin-1 E is the most basic at an expected pI of 5.70. Based on the predicted serpin-1 isoform migration, we anticipated we would achieve a partial separation of serpin-1 isoforms by using very narrow range isoelectric focusing strips with the pH range 4.7-5.9.

Initial characterization of serpin-1 isoforms in naïve larval plasma was performed using 2D-PAGE. Eleven spots of the expected size of serpin-1 were detected by Coomassie staining, excised from the gel, and analyzed by trypsin digestion (Fig. 4-7). All eleven spots were found to contain serpin-1. Six of the 12 isoforms, A, E, F, J, K, and Z, were identified on the basis of peptide mass fingerprinting, with 2 or more isoform-specific peaks identified in a spot (Table 4-4). Serpin-1I was identified in spots SpER4 and SpER5 by 1 PMF peak that was confirmed by MS/MS analysis (Table 4-5). MS/MS analysis also confirmed the presence of isoform Z in spot SpER7, A in spot SpER9, and E in spots SpER10 and SpER11. The PMF and MS/MS results are summarized in Table 4-6. Seven of the 12 isoforms were identified, including at least one isoform from each spot.

2D-PAGE of serpin-1 secreted by fat body

Previous studies suggested that the fat body might not express the full range of serpin-1 isoforms (Jiang et al, 1994, Molnar et al, 2001). The serpin-1 protein profile from plasma was compared with that from proteins secreted by fat body cultured *in vitro*. Coomassie staining of the 1D SDS-PAGE separation of the media shows the presence of a large number of proteins (Fig. 4-8 A). Immunoblotting confirmed the presence of serpin-1 in these fat body secretions (Fig. 4-8 B). Media from fat body incubations 1 and 2 was pooled prior to analysis by 2D-PAGE. Many proteins are seen in the 2D Coomassie stained gel while serpin-1 proteins were specifically detected by immunoblotting (Fig. 4-9). At least eight spots corresponding to serpin-1 isoforms indicates that fat body secretes multiple serpin-1 isoforms, in a pattern similar to that observed in 2D gels of whole plasma.

Selection of proteases for cleavage prior to MALDI-TOF/TOF analysis

Peptides with masses between 700 and 3500 Da are reasonably expected to show up in our MS analysis. Several of the unidentified isoforms from whole plasma, B, C, G, and H, have only one isoform-specific tryptic peptide of appropriate size for MALDI-TOF/TOF analysis, while the other isoforms have at least three. Therefore, we explored the uses of other proteases for peptide mass fingerprinting. Table 4-7 shows the expected number of serpin-1 isoform-specific peptides with a mass between 700 and 3500 Da from four different protease digestions. The combination of AspN and LysC gives cleavage N-terminal to Asp and C-terminal to Lys. GluC (also known as *Staphylococcus aureus* V8 protease) has been reported to cleave predominately after Glu residues in bicarbonate buffer but after either Glu or Asp when used in phosphate buffer (Houmard & Drapeau, 1972). The isoform-specific peptide predictions assume GluC protease cleavage after Glu residues in bicarbonate buffer and cleavage after both Asp and Glu in phosphate buffer. Some isoforms which are poorly covered by trypsin digestion, such as B and H, are expected to give more definitive results with LysC/AspN or GluC (either buffer). Isoforms C and G remain more challenging to identify because few isoform-specific peptides are generated by any of the protease treatments. Serpin-1C has only one isoform-specific peptide in the 700-3500 Da mass range in three out of the four protease

treatments but cleavage after Asp and Glu would yield 3 suitably sized peptides. Serpin-1G has just one isoform-specific peptide from two of the treatments and two expected from either GluC digestion. Figure 4-10 lists the specific peptide masses expected from trypsin, LysC/AspN, and GluC digestions and highlights the relevant protease cleavage sites in the protein sequences.

2D-PAGE of immunoaffinity purified serpin-1 isoforms

Serpin-1 isoforms were further characterized by 2D-PAGE using immunoaffinity purified serpin-1 from naïve plasma. Immunoaffinity purification removed other plasma proteins and facilitated higher serpin-1 concentrations. These higher concentrations and the absence of other proteins enabled visualization of higher molecular weight spots (55-66 kDa) that are putative serpin-1 protease complexes (Fig. 4-11 A and B). Spots corresponding to serpin-1 isoforms or putative serpin-1 protease complexes were cut out from three gels and analyzed by either trypsin digestion (Fig. 4-11C), digestion by LysC/AspN (Fig. 4-11D), or digestion by GluC in either bicarbonate buffer or phosphate buffer (Fig. 4-11E). These spot numbers are consistent with spot location across the three gels used for the different protease treatments but are different than the numbers used for the whole plasma gel (Fig. 4-7).

Table 4-8 lists the top Mascot results after searching the NCBInr database using data from MALDI-TOF/TOF analysis of protease digested spots. An isoform was considered to be identified for the purpose of this table (see column 4, "Isoform") if a spot had at least 2 isoform-specific masses identified by PMF or one PMF mass that was further confirmed by MS/MS analysis. Spots which clearly contain serpin-1, but in which the specific isoform(s) could not be identified, have a question mark listed under the isoform column. Spots where no *M. sexta* protein was indentified are excluded from the table. The last two columns list the number of isoform specific peptides identified by PMF and, if applicable, also identified by subsequent MS/MS analysis (MS/MS results listed in table 4-10). Ideally we would like to have two MS/MS spectra confirmed for each identified isoform (see Table 4-11 for MS/MS spectra summary) but in some cases, like protease-complexes, at most only 1 isoform specific peptide is available.

Spots 2-14 and 41 are expected to be serpin-1 alone while higher molecular weight spots 16-21 and 30-35 are putative serpin-1 protease complexes. Spots 1 and 15, at the extreme ends of the IEF strip, were found by trypsin digestion to have unresolved serpin-1 isoforms and were not picked for subsequent analysis by the other proteases. Two of the fainter serpin-1 spots, 3 and 5, did not have a specific isoform detected by any of the protease digestions. Spot 5, with its lower molecular weight, is probably cleaved serpin-1. Trypsin digestion followed by MALDI-TOF/TOF and Mascot analysis was adequate for identification of serpin-1 isoforms in spot 2 (J), spots 7, 8, and 9 (I), spots 10 and 41 (Z), spot 11 (K), spot 12 (A), and spots 13 and 14 (E). Multiple isoforms were detected in a few spots. PMF (but not MS/MS) identified two E peaks in spot 12 although spot 12 also contained serpin-1A (supported by 4 PMF peaks, 3 verified by MS/MS). Spot 12 neighbors spot 13, which was shown to contain serpin-1E. Also, in addition to the 3 Z peaks (1 confirmed by MS/MS) in spot 10, a serpin-1A peak corresponding to residues 337-355 was detected. The pI of serpin-1A cleaved between P1-P1' is 5.1, close to the pI of spot 10, which raises the possibility that the serpin-1A detected in spot 10 is cleaved serpin-1A. Spot 41, which is also isoform Z, is above spot 10.

Serpin-1 F and H have predicted tryptic peptides of very similar size, 1751.83 and 1751.88 Da, respectively (see Fig. 4-10). A peptide with observed mass 1752.029 Da was present in spot 4, which was inconclusive by peptide mass fingerprinting at the 150 ppm error tolerance appropriate for that spot. However, MS/MS analysis of this peptide led to its identification as serpin-1H (Fig. 4-12). Two additional peaks from LysC/AspN digestion were identified as serpin-1H peptides in spot 4. Thus we are confident of the presence of serpin-1H in spot 4.

PMF results from LysC/AspN digestion complemented the results from trypsin digestion by identifying two peptides each for serpin-1Z in spot 41 and serpin-1K in spot 11. One peptide from each of those spots was confirmed by MS/MS analysis (see Table 4-8). Multiple peptides were identified by PMF and MS/MS for serpin-1E in spots 13 and 14. LysC/AspN digestion also led to MS/MS identification of a serpin-1B peptide in spot 6. GluC cleavage in bicarbonate buffer helped confirm the presence of serpin-1J in

spot 2 and of serpin-1B in spot 6 by MS/MS. The same serpin-1B peptide was also detected by GluC cleavage in phosphate buffer.

The observed positions of all of the identified isoforms after isoelectric focusing was remarkably similar to the expected locations based on predicted pI (Fig. 4-6). Only three isoforms remained undetected: C, D, and G.

Identification of putative serpin-1 protease complexes

Detection of the serpin-1 isoform(s) present in protease complexes was more difficult than detecting full length serpin-1 isoforms because the target protease cleaves the serpin at the P1 residue, releasing a carboxyl-terminal sequence that contains much of the variable sequence (see Fig. 4-1C). As mentioned earlier, only isoforms A, D, E, J, and K, have an isoform-specific tryptic peptide after they are cleaved. A search of the raw MALDI-TOF peaks from trypsin treated samples for predicted peptides from Lys336 to the P1 residue for the remaining isoforms in putative complex bands (spots 16- 21, 31-34) failed to reveal any additional isoform specific information (data not shown).

Despite the challenges, analysis of higher molecular weight spots 16-21 and 30-35 that are putative serpin-1-protease complexes, yielded some interesting results. Spots 16-21 and 30-34 all clearly contain serpin-1. Mascot and Aldente PMF of spots 31 and 32 detected one peak for serpin-1E and, after validation by Aldente, spot 31 also contained six peptides for the catalytic domain of *M. sexta* hemolymph proteinase 8 (HP8) (Table 4-9). The protease sequences in the databases (NCBI and Uniprot) are of the complete proteases, which are usually zymogens. We would generally expect a protease to be cleaved and activated prior to binding a serpin and therefore only the catalytic domain would be present in a serpin-protease complex. Therefore we constructed a list of catalytic domain sequences from known *M. sexta* proteases (HP1-24 (Jiang et al, 2005)) and used Mascot to search those sequences against our MALDI-TOF/TOF data. Using this method, at least four HP8 catalytic domain peptides were found in spots 19, 20, 31, and 32 (Table 4-9) although only spots 31 and 32 had a significant protein score for HP8 because those spots also generated HP8 MS/MS data. Three MS/MS spectra for HP8 proteins were generated from tryptic peptides (2 in spot 31 and 1 in spot 32) while another HP8 peptide was detected by MS/MS in spot 19 by LysC/AspN digestion (Table

4-10). Immunoblotting after 2D-PAGE with antibody to HP8 supported these results, showing a signal for HP8 corresponding to spots 18-20, 31, and 32 (Fig. 4-13). Spot 19 contained serpin-1J based on PMF and MS/MS. Based on PMF and the one possible isoform-specific peptide after serpin cleavage, spots 19 and 20 both appear to contain serpin-1A while spots 31 and 32 contain serpin-1E, although those peptides were not confirmed by MS/MS. These results suggest that serpin-1 isoforms J, A, and E complex with, and therefore can inhibit, HP8.

Spot 21 clearly contained serpin-1 but a specific isoform was not identified. Validation of serpin-1 in spot 21 by Aldente allowed identification of *M. sexta* hemocyte protease-1 (HP1) in the same spot (data not shown). HP1 was also detected in spot 21 in the Mascot search against the activated *M. sexta* protease sequences by PMF (Table 4-9). Two HP21 peptides were confirmed by Mascot MS/MS analysis (Table 4-10, highlighted), one of which was also identified by the Scaffold program (Table 4-10, underline). This identification was also confirmed by immunoblotting with HP1 antibody (Fig. 4-14 A and B). The HP1 antibody also recognized two small, neighboring spots that were not detected on the Coomassie gel. PMF coverage from both Aldente and Mascot is shown in Fig. 4-14C. Interestingly, two peptides were detected by Aldente that weren't included in the sequence of the active protease searched by Mascot because they are upstream of the putative HP1 activation site. A cleaved serpin-1 isoform has a mass of ~38 kDa and HP1 cleaved at its predicted activation site, R154, has a mass of 26.6 kDa. The expected serpin-1-HP1 mass is ~64.6 kDa, quite close to the observed size of the serpin-1-HP1 complex on the gel (~66 kDa). However, a cleavage site slightly more N-terminal in the HP1 sequence, say at K136, could also give a similar result (in this case, a 66.5 kDa complex).

Additional immunoblotting of immunoaffinity purified serpin-1 isoforms after 2D-PAGE was performed with antibodies to hemolymph proteinase 6 (HP6) and prophenoloxidase activating proteinase-3 (PAP3). PAP3 was previously found in a complex with serpin-1J. No signal was detected on either of these blots (data not shown).

Isoform specific antibody testing using 2D-PAGE

2D-PAGE of serpin-1 isoforms also gave us an opportunity to test some isoform specific antisera previously made to peptides corresponding the C-terminal region of serpin-1B and serpin-1F. Figure 4-15 shows immunoblots using the serpin-1F antiserum (A) and serpin-1B antiserum (B). Both blots were initially stained with Ponceau S and the positions of spots 4 and 6, two very prominent spots, were marked. MS/MS results suggest the presence of serpin-1B in spot 6 (Table 4-11) and, since both serpin-1B and serpin-1F have the same predicted pI (4.95), we would expect them both to migrate to the same position on a 2D gel. After destaining and developing the immunoblot, the position of the signal corresponded to spot 6 in both cases. No other specific staining was observed, although the background for the serpin-1B antibody was relatively high due to the low dilution (1:500) required for the primary antibody incubation.

Discussion

Previous investigations of serpin-1 in *M. sexta* plasma or tissues have tended to look at all serpin-1 isoforms as an aggregate. However, the 12 serpin-1 isoforms are functionally different due to differences in their reactive center loops. The serpin-1 isoforms could be present in the plasma in different amounts due to regulation at the post-transcriptional, translational, or post-translational levels, although details about any of these possibilities, including alternative splicing of serpin-1 isoforms, are still unknown. 1D-PAGE is incapable of separating the very similarly sized serpin-1 isoforms. One previous study attempted to separate *M. sexta* serpin-1 isoforms from hemolymph, fat body, epidermis and cuticle by native PAGE (Molnar et al, 2001). The authors made polyclonal antibodies in mice by injecting ~49kDa hemolymph proteins cut out of an SDS-PAGE gel. These antibodies detected 7 discrete bands in the native gel from hemolymph, which the authors assigned to the 12 serpin-1 isoforms based on estimated pI. This identification of the serpin-1 isoforms was not confirmed by any other method and it is possible that the mouse antibodies recognized other plasma proteins that, because of size, were present in the 1D band used for antibody production. For example, serpins 3, 4, 5, 6, and 7 have also been subsequently identified in plasma (Tong &

Kanost, 2005, Wang & Jiang, 2004, Zhu et al, 2003) and these serpins might have been present in Molnar et al.'s 1D band used for antibody production.

Early work on serpin-1 included separation of *M. sexta* hemolymph by 2D-PAGE using a pH range of 3.5 to 10. An antibody made against *M. sexta* serpin-1 purified from plasma recognized approximately 10 spots (Kanost, 1990, Kanost et al, 1989). Serpin-1 isoforms in *Mamestra configurata* were analyzed by 2D-PAGE with pH ranges of 3-6 and 5-8 (Chamankhah et al, 2003). Spots were identified as serpin-1 using antibody staining. The recently published proteomic analysis of *M. sexta* hemolymph identified 11 spots on a 4-7 pH range strip as serine protease inhibitors (spots 18-28) and labeled four of these as serpin-1 (spots 20, 21, 26, and 27) (Furusawa et al, 2008). I further analyzed their data and determined all 11 serine protease inhibitor spots are serpin-1 and that isoform-specific peptides are present for four of the 12 isoforms. Spots 18, 19, and 22 did not contain any isoform specific peptides while spot 20 contained one peptide for serpin-1I, spot 22 contained two peptides specific for serpin-1Z, and spots 26 and 27 both contained two serpin-1E peptides. Spots 23, 24, and 25 contained 1, 4, and 2 peptides specific for serpin-1K. Spot 28, which has a lower observed molecular weight based on the SDS-PAGE gel, has one serpin-1K specific peptide. Based on the observed molecular weight of the spot and the location of the serpin-1K peptide, spot 28 is likely serpin-1K cleaved after its P1 residue.

We improved upon these earlier studies by using very narrow pH range isoelectric focusing strips, from pH 4.7 to 5.9, to maximize separation of serpin-1 isoforms by 2D-PAGE. We also used immunoaffinity purified serpin-1 proteins to eliminate other plasma proteins of similar size and to obtain higher concentrations of serpin-1. We used MALDI-TOF/TOF to identify serpin-1 isoforms present in the 2D spots.

We detected 8 of the 12 serpin-1 isoforms in 2D-PAGE spots by identifying at least 2 isoform-specific peptides by MS/MS analysis. A ninth isoform, serpin-1F, was identified by peptide mass fingerprinting and immunoblotting. We consistently found serpin-1E migrating as two spots, one at a pI of about 5.6 and the other about 5.7. Neither spot has PMF coverage of the last 14 residues by either trypsin or AspN/LysC digestion. Allelic variation involving charged amino acids in the isoform-specific region of serpin-1E is a possible explanation for the two distinct spots containing this isoform.

For example, if an additional D or E replaced an uncharged amino acid, the pI would change from 5.7 to 5.57. Another potential explanation is difference in N-linked glycosylation, which has been shown to change the pI but not apparent molecular weight of proteins (Packer et al, 1996).

Detection of individual serpin-1 isoforms by MALDI-TOF/TOF is challenging because only the C-terminal 39-46 amino acids are isoform-specific. Additionally, the end of the common portion of serpin-1 (residues 337-352) contains a long region with no trypsin cleavage sites and multiple glutamic acids, making it a difficult region to detect by protease digestion and MS. This impacts identification of individual isoforms, which begin their unique sequences at residue 353. For example, trypsin digestion of serpin-1 isoforms B, C, G, and H gives an initial isoform-specific peptide that is larger than 4,600 Da. Given that peptides in mass range of 750-3500 kDa are best detected by MALDI-TOF/TOF, these very large fragments are not useful and reduce the total number of peptides available for isoform specific detection. Myung and coworkers used four different digestive enzymes, trypsin, chymotrypsin, AspN, and LysC, to obtain 100% sequence coverage of their 20 kDa protein by MALDI-TOF/TOF (Myung et al, 2007). We similarly used multiple digestive enzymes to improve our detection of specific serpin-1 isoforms. Trypsin digestion was adequate for identification of five of the 12 isoforms but accurate identification of serpin-1H in spot 4 and serpin-1B peptide in spot 6 required information from the LysC/AspN treatment. Digestion by GluC in bicarbonate buffer enhanced results from the other digestions by providing a second MS/MS spectrum for serpin-1J (in spot 2) and serpin-1B (in spot 6). Three isoforms, C, D, and G, were not identified in any of the spots, although one mass from GluC phosphate digestion of spot 8p matched a predicted serpin-1G peptide (see Table 4-8). The lack of identification of C and G is not surprising because they had the fewest isoform-specific peptides generated by the protease digestions and are thus hard to identify by MALDI-TOF/TOF. Serpin-1C could be in spot3 and also with serpin-1H in spot4. G is expected to be quite near serpin-1I and may be in spots 7 and/or 8. The largest mystery is D, since it should have multiple tryptic and LysC/AspN peptides (see Fig. 4-10) however, none were detected. No prominent spot was seen between spot 41 (Z) and spot 11 (K), which is the region where we would expect to see isoform D.

I was hoping for better detection of serpin-1C with GluC protease because it is commonly thought that GluC cleaves after both Glu and Asp when used in phosphate buffer, based on early reports and this is consistent with the enzyme cleavage rules on Aldente (for example, Glu C [phosphate] cleaves C-terminal side of D or E unless if P is C-term to D or E, or if E is C-term to D or E) (Drapeau et al, 1972, Houmard & Drapeau, 1972). However, another report suggests a 3,000 fold faster cleavage after Glu than Asp regardless of buffer (Sorensen et al, 1991). There are also subtle differences in enzyme specificity, with cleavages occurring at different rates depending on the P1' residue after the Asp or Glu (Sorensen et al, 1991). In this experiment, using GluC in either phosphate or bicarbonate buffer gave fewer PMF and MS/MS results than trypsin or LysC/AspN digestions, probably because GluC cleavage is less reliable than trypsin or AspN and LysC cleavages. Other digestive enzymes, such as chymotrypsin, could also be considered for increasing coverage in future experiments.

Use of multiple digestive enzymes (both LysC/AspN and GluC bicarbonate) helped identify a protease, HP8, in spot 19. Identification of an active protease, which has only its catalytic domain after SDS-PAGE, is challenging because the sequence in the database is the full length protease zymogen, while only the catalytic domain is present in a complex. Less coverage leads to a lower score and a higher likelihood of a false negative. Additionally, a protease complexed with a serpin is necessarily in a spot containing at least two different proteins which further reduces the confidence of the individual identifications by PMF. Given these challenges, increased information about protease complexed with serpins is another advantage of using multiple protease digestions of matched spots.

Identifying serpin-1 isoforms is a greater challenge when they are complexed with a protease. After cleavage at the P1 residue, at most only 1 isoform-specific tryptic peptide remains (in the case of A, D, E, J, K) and, for the other isoforms, there is no remaining isoform-specific tryptic peptide. MS/MS analysis of higher molecular weight spot 19 led to identification of an isoform-specific peptide for serpin-1J by trypsin digestion and to two different peptides matching HP8, one each from LysC/AspN and GluC bicarbonate digestions. Based on single isoform-specific peptides indicated by

PMF, we also suspect the presence of serpin-1J in spot 19, serpin-1A in spot 19 and 20, and serpin-1E in spots 31, 32, 33, and 34.

We took advantage of our antisera to *M. sexta* proteases and used immunoblotting to aid identification of serpin-protease complexes. Spots 18, 19, 20, 31, and 32 all reacted with antisera to HP8. Spots 19, 20, 31, and 32 contained at least four HP8 tryptic peptides by PMF search of *M. sexta* protease catalytic domains. MS/MS analysis after trypsin digestion further confirmed two HP8 peptides in spot 31 and one HP8 peptide in spot 32 while two HP8 peptides were detected by MS/MS in spot 19 (one by LysC/AspN and the other by GluC bicarbonate).

All together, MALDI-TOF/TOF and immunblot results suggest that serpin-protease complexes form between HP8 and serpin-1J, -1A, and -1E. Serpin-1 isoforms A, E, and J are the only three with known trypsin-like specificity. Serpin-1D has an Arg as its putative P1 residue but is followed by a Pro and did not inhibit any commercially available proteases during *in vitro* tests (Jiang & Kanost, 1997). HP8 is predicted to have trypsin-like specificity (Jiang et al, 2005) so it is not surprising that HP8 would be inhibited by serpin-1A, E, and J. *In vitro* mixing of active HP8 with serpin-1A, -1E, and -1J results in complex formation between HP8 and these three isoforms (Chunju An, unpublished data). HP8 has previously been found in a complex with serpin-6 (Zou & Jiang, 2005). HP8 is similar to *Holotrichia diomphalia* proPO-activating factor-III, which cleaves a serine-protease homolog named pro-PO activating factor-II (Kim et al, 2002, Zou & Jiang, 2005), and to *D. melanogaster* easter and spätzle processing enzyme, both of which cleave the cytokine spätzle during Toll activation (Jang et al, 2006, Jiang et al, 2005).

An additional serpin-protease complex was found in spot 21, with unidentified isoform(s) of serpin-1 and the serine protease HP1. This protease complex was also confirmed by immunoblotting with both HP1 and serpin-1 antibodies. HP1 has a predicted trypsin-like specificity and has been found in complexes with serpin-4 and serpin-5 (Tong et al, 2005). Key proteases in serine protease cascades are often regulated by multiple serpins (Zou & Jiang, 2005). Peptide mass fingerprinting coverage from Aldente included two peptides just upstream of the putative HP1 activation site. While the size of spot 21 suggests that a cleaved form of HP1 is complexed with serpin-1, it is

possible that HP1 cleavage occurred at a site slightly more N-terminal than predicted in the HP1 sequence. Many protease zymogens have an activation site followed by IIGG. The predicated cleavage site in HP1, AQGR*VFGS, is less conserved than many of the other activation sites in *M. sexta* hemolymph proteases (Jiang & Kanost, 2000, Jiang et al, 2005).

Antisera had been previously generated against peptides corresponding to the isoform-specific region of serpin-1B and serpin-1F (Yang Wang and Michael Kanost, unpublished data). Use of these antisera with 2D immunoblots helped support their isoform-specific detection of serpin-1 isoforms B and F and further strengthened the identification of serpin-1B and -1F in spot 6. No cross reactivity was seen with any of the other serpin-1 isoforms. While the co-migration of serpin-1B and –1F leaves open the possibility that the peptide antisera to those two isoforms cross-react, the isoform-specific amino acid sequences are more divergent between B and F than between certain other isoforms, such as F and H (Jiang et al, 1996). Additional peptide-based antisera could be made for identification of the other full-length serpin-1 isoforms on 2D gels but are not expected to be suitable for serpin-1-protease complex detection because such a large portion of the reactive center loop is cleaved upon complexing with the protease. Antisera to isoform-specific peptides have been generated for *M. configurata* serpin-1a and -1b/c. Antiserum to McSerpin-1a did not react with serpins separated by 2D-PAGE on 3-6 or 5-8 pH range strips, whereas three spots reacted with the McSerpin-1b/c antibodies (Chamankhah et al, 2003).

Jiang and coworkers found that recombinant serpin-1J could form a complex with active prophenoloxidase activating proteinase-3 (PAP3) purified from plasma and that PAP3 was inhibited to half its activity at a serpin:protease ratio of 6.7 (2003). Based on theoretical pIs, the PAP3 catalytic domain complexed with serpin-1J would run at about 5.2, in the range of the narrow IEF strips used in this study. However, IEF of the related PPAE protease in *B. mori* indicated a pI 2 pH units more basic than the predicted pI (Satoh et al, 1999) and unpublished observations with *M. sexta* PAP3 also suggested PAP3 is more basic than its predicted pI (Maureen Gorman and Michael Kanost). Both MS and immunoblots results failed to detect a complex formed between PAP3 and serpin-1J in this experiment, although since we used naïve plasma, PAP3 would be at low

concentration in our sample and unlikely to be activated. PAP3 is strongly upregulated in both fat body and hemocytes upon immune challenge (Jiang et al, 2003) and is completely inhibited by serpin-3 at a serpin:protease ratio of 2 and by serpin-6 at a serpin-protease ratio of 5 (Wang & Jiang, 2004, Zhu et al, 2003). Serpin-3, which is also upregulated upon immune challenge, is believed to be the physiologically most important inhibitor of PAP3 (Zhu et al, 2003).

While narrow range IEF strips give good separation by pI, a limitation is the inability to detect serpin-1 protease complexes that migrate outside of that strip range. One possible advantage to a 1D or larger range 2D gel is detection of serpin-protease complexes with different pIs. Antibody based immunoaffinity purification of *M. sexta* serpins and analysis by 1D-PAGE followed by MALDI-TOF has been used previously to identify serpin-protease complexes (Tong et al, 2005, Zou & Jiang, 2005). Using this method we were able to identify a serpin-1K complex with *M. sexta* chymotrypsin, which has an expected complex pI of 7.73 and is therefore too basic to be separated on the 4.7-5.9 range IEF strips used for 2D analysis. Finding serpin-1K complexed with chymotrypsin, a digestive protease made by the midgut (Peterson et al, 1995) supports the hypothesis that serpins in *M. sexta* plasma may contribute to regulating digestive enzymes that escape from the digestive system (Kanost, 1990). We also observed mRNA expression of all twelve serpin-1 isoforms in midgut by PCR. We do not yet know which cells in the midgut express serpin-1 or which direction, into hemolymph or gut lumen, serpin-1 is secreted.

Mechanisms and regulation of alternative splicing

There are multiple types of alternative splicing which lead to different transcripts, including internal exons that are variably included (cassette exons), alternative 5' or 3' splice sites, intron retention, and mutually exclusive alternative exons (Blencowe, 2006, Stamm et al, 2005). In serpin-1, the twelve versions of exon 9 are used mutually exclusively (Jiang et al, 1994). Tandem arrays of mutually exclusive exons are seen in other arthropods but vertebrate genes with mutually exclusive exons typically contain only two exons; vertebrate genes with more than three mutually exclusive exons are quite rare (Crayton et al, 2006, Graveley et al, 2004). Several mechanisms have been proposed

for mutually exclusive alternative exon usage in systems with just two available exons, including steric hindrance due to closely spaced exons or the binding of splicing factors, use of the minor U12 spliceosome for one of the splice sites, and the nonsense-mediated decay pathway when inclusion of both exons causes a frameshift and premature stop codon (Chang et al, 2007, Graveley, 2005, Smith, 2005). These mechanisms fail to explain how such splicing can occur in situations with larger numbers of alternatively spliced exons (Graveley, 2005, Smith, 2005). All types of alternative splicing rely on components of the spliceosome and trans-acting factors that are activators or repressors of splicing (House & Lynch, 2008, Park et al, 2004).

While vertebrates lack long arrays of tandem exons, they can still generate many isoforms through mechanisms other than mutually exclusive alternative splicing. Human neurexin3, which has the potential to form 1728 transcripts due to alternative splicing, does not have long tandem arrays of exons. Of its nine exon regions that can be alternatively spliced, three have one exon, five have two exons, and only one has three exons. Alternative splicing in neurexin3 involves alternative 5' and 3' splice sites and cassette exons and is more complicated than simply usage of mutually exclusive exons (Rowen et al, 2002).

However, in insects, extreme numbers of mRNAs can be formed entirely through mutually exclusive alternative splicing. One notable example of a gene with many mutually exclusive exons is Dscam, best studied in *D. melanogaster* where it has the potential to generate 38,016 different isoforms. Three exon clusters in DmDscam have many mutually exclusive exons: cluster 4 has 12, cluster 6 has 48, and cluster 9 has 33 (Olson et al, 2007). All three of these regions encode extracellular immunoglobulin repeats involved in neuronal wiring and immune response (Boehm, 2007, Schmucker, 2007). A fourth alternatively spliced location, cluster 17, has two variable exons that encode different transmembrane regions. It is significant that alternative exons from the same cluster have not been detected spliced together even though all of the exons have splice sites (Graveley, 2005). This is similar to *M. sexta* serpin-1, where only 1 cDNA with multiple copies of exon 9 (in this case, 9B spliced to 9C) has ever been found (Jiang et al, 1994).

Understanding the factors and mechanisms involved in the mutually exclusive splicing of exons in DmDscam cluster 6 is an area of active research. It appears that all 48 different variants of exon 6 are generated in a stochastic manner. A docking site is conserved in the intron after exon 5 and each of the 48 exon 6 variants have an upstream selector sequence complementary to the docking site. The docking site can only bind one selector sequence at a time, leading to the splicing of just one exon 6 onto exon 5 (Anastassiou et al, 2006, Graveley, 2005). Interestingly, this mechanism is not used with the mutually exclusive alternative splicing of exon clusters 4 or 9. Another important factor for splicing of Dscam exon 6 is a heterogeneous nuclear riboprotein hrp36, which is essential for preventing multiple exon 6 inclusion in S2 cells, but has no effect on exons from clusters 4 or 9 (Olson et al, 2007). Because of the complexity of genes with multiple alternative exons, as well as the complexity of the spliceosome and splicing factors, we are just beginning to understand some of the many mechanisms for mutually exclusive exon splicing and their regulation.

When different isoforms are expressed there is the possibility that regulation of splicing leads to preferential expression of one isoform over another under certain conditions such as developmental stage or tissue type (Stamm et al, 2005, Tarn, 2007). In the case of *M. sexta* serpin-1, an over-abundance of Spn-1F in a fat body cDNA library and the absence of a band on a native gel of fat body sample lead to the suggestion that not all of the serpin-1 isoforms are expressed in *M. sexta* fat body (Jiang et al, 1994, Molnar et al, 2001). In this study we made the first isoform-specific primers to serpin-1 and found, through semi-quantitative, isoform specific PCR, that all 12 isoforms are expressed at similar levels in hemocytes, fat body, and midgut. These results indicate that serpin-1F is not over-expressed relative to the other isoforms in fat body, a result that is also supported at the protein level by a similar pattern of serpin-1 expression in plasma and secreted by fat body cultured *in vitro*. Our RT-PCR results also showed a brighter band intensity in serpin-1A in fat body and hemocytes after immune challenge, which raises the possibility that serpin-1A is upregulated after immune challenge. This observation needs to be further explored, possibly by real-time PCR.

Other alternative splicing of serpin isoforms in other insects.

In addition to *M. sexta* serpin-1, several other insect serpins show alternative splicing that leads to multiple isoforms with different reactive center loops. The first discovered example of two serpins with identical sequence up until the reactive center loop are *B. mori* antichymotrypsin II and antitrypsin I, which are identical through residue 336. These two serpins share 46% similarity in their carboxyl termini (Sasaki, 1991).

The lepidopteran *M. configurata* has a serpin-1 homolog with multiple sequences at the C-terminal end. While the *M. configurata* serpin-1 was detected in a midgut cDNA library screen, Northern blotting of serpin-1 in *M. configurata* showed no expression in midgut, low levels of expression in foregut, and higher levels of expression in fat body and hemocytes (Chamankhah et al, 2003). Chamankhah and coworkers speculated that the presence of serpin-1 in the midgut cDNA library was due to contamination with fat body or other tissues. They also reported that McSerpin-1a was localized to the midgut lumen and that McSerpin-1b/c isoforms were associated with the sub-cuticular layer and basement membranes of tissues including muscle, midgut, and trachea. When a *M. sexta* midgut cDNA library was screened with serpin-1 antibody serpin-1 clones were also obtained (M. Kanost, unpublished data).

A serpin in the flea *Ctenocephalides felis* (order Siphonaptera) also has a similar gene structure, with a common region (6 exons) followed by cluster of at least 15 variable exons encoding the reactive center loop (Brandt et al, 2004). Northern blot results using a probe to the constant region showed expression in all life stages except low levels of expression in pupae and little if any in egg. Isoform specific probes to 14 and 15 showed slightly different patterns of expression, most interestingly that isoform 14 is strongly upregulated after blood feeding. *C. felis* serpin protein was detected by immunohistochemistry in unfed midgut and fat body cells (Brandt et al, 2004).

In *A. gambiae*, AgSpn10 contains 4 isoforms with different reactive center loops. All four of these isoforms lack a secretion signal peptide (Danielli et al, 2003). *D. melanogaster* serpin 4 (DmSpn4) encodes 8 variants with 4 different reactive center loops and three different cellular locations: DmSpn4A is retained in the ER, DmSpn4B-D are secreted, and DmSpn4E-H are intracellular (Bruning et al, 2007). The reactive center

loops are the same in isoforms A & E, B & F, C & G, and D & H. In *D. melanogaster* some proteases inhibited by the DmSpn4 isoforms have been identified. For example, inhibitory activity of Dm Spn4A against the proprotein convertase furin has been demonstrated by several groups (Osterwalder et al, 2004). Richer et al. (2004) showed that DmSpn4A both inhibits and forms serpin-protease complexes with human furin and *Drosophila* PC2 (*amontillado*); based on its ER retention signal they suggest DmSpn4A may regulate furin activity in the secretory pathway (Richer et al, 2004). Inhibition studies of all four DmSpn4 reactive center loops, using DmSpn4 isoforms E-H, showed that only DmSpn4E is able to inhibit furin (Oley et al, 2004). DmSpn4F, G, and H formed complexes with neutrophil elastase but also were significantly cleaved by the elastase. Both DmSpn4F and G formed complexes with chymotrypsin but again were also cleaved as substrates. Both DmSpn4F and G showed inhibition of cathepsins L and S and formed SDS stable complexes on non-reducing gels (Bruning et al, 2007).

Conclusions

Isoform-specific information about serpin-1 mRNA and protein levels indicates the potential upregulation of serpin-1A mRNA upon immune challenge. The 2D gel analysis of individual serpin-1 isoforms in the plasma suggests the levels of serpin-1D are low, but the strong presence of other isoforms, including two serpin-1E variants with different pIs. I also detected serpin-1 protease complexes with *M. sexta* proteases HP1, HP 8 and chymotrypsin, which suggests serpin-1 isoforms regulate these endogenous proteases. Serpin-1K complexed with chymotrypsin and serpins A, E, and J complexed with HP8.

The 2D-PAGE, MALDI-TOF/TOF, and immunblot analysis strategies used for identifying serpin-protease complexes can be useful in other systems where proteases and serpins form complexes, including mammalian blood. Alternative splicing is a common strategy for increasing protein diversity across higher eukaryotes and often the identification of different protein isoforms is cited as an advantage of proteomic studies. In practice, isoform identification can be challenging. Techniques for identifying individual protein isoforms by multiple protease digestions and MS analysis, as demonstrated in this study, can be helpful in proteomic studies of other systems.

Acknowledgements

This research was funded by NIH grant GM41247. I would like to thank Rebekah Woolsey from the Nevada Proteomics Center for sample digestion and mass spectrometry. The Nevada Proteomics Center is supported by NIH Grant Number P20 RR-016464 from the INBRE Program of the National Center for Research Resources. I also thank Dr. Rollie Clem for the use of the BioRad IEF unit.

References

Anastassiou D, Liu H, & Varadan V (2006) Variable window binding for mutually exclusive alternative splicing. *Genome Biol* **7**: R2

Blencowe BJ (2006) Alternative splicing: new insights from global analyses. *Cell* **126**: 37-47

Boehm T (2007) Two in one: dual function of an invertebrate antigen receptor. *Nat Immunol* **8**: 1031-1033

Brandt KS, Silver GM, Becher AM, Gaines PJ, Maddux JD, Jarvis EE, & Wisnewski N (2004) Isolation, characterization, and recombinant expression of multiple serpins from the cat flea, *Ctenocephalides felis*. *Arch Insect Biochem Physiol* **55**: 200-214

Bruning M, Lummer M, Bentele C, Smolenaars MM, Rodenburg KW, & Ragg H (2007) The Spn4 gene from *Drosophila melanogaster* is a multipurpose defence tool directed against proteases from three different peptidase families. *Biochem J* **401**: 325-331

Chamankhah M, Braun L, Visal-Shah S, O'Grady M, Baldwin D, Shi X, Hemmingsen SM, Alting-Mees M, & Hegedus DD (2003) *Mamestra configurata* serpin-1 homologues: cloning, localization and developmental regulation. *Insect Biochem Mol Biol* **33**: 355-369

Chang WC, Chen YC, Lee KM, & Tarn WY (2007) Alternative splicing and bioinformatic analysis of human U12-type introns. *Nucleic Acids Res* **35**: 1833-1841

Crayton ME,3rd, Powell BC, Vision TJ, & Giddings MC (2006) Tracking the evolution of alternatively spliced exons within the Dscam family. *BMC Evol Biol* **6**: 16

Danielli A, Kafatos FC, & Loukeris TG (2003) Cloning and characterization of four *Anopheles gambiae* serpin isoforms, differentially induced in the midgut by *Plasmodium berghei* invasion. *J Biol Chem* **278**: 4184-4193

De Gregorio E, Han SJ, Lee WJ, Baek MJ, Osaki T, Kawabata S, Lee BL, Iwanaga S, Lemaitre B, & Brey PT (2002) An immune-responsive Serpin regulates the melanization cascade in *Drosophila*. *Dev Cell* **3**: 581-592

Dementiev A, Dobo J, & Gettins PG (2006) Active site distortion is sufficient for proteinase inhibition by serpins: structure of the covalent complex of alpha1-proteinase inhibitor with porcine pancreatic elastase. *J Biol Chem* **281**: 3452-3457

Dittmer NT, Suderman RJ, Jiang H, Zhu YC, Gorman MJ, Kramer KJ, & Kanost MR (2004) Characterization of cDNAs encoding putative laccase-like multicopper oxidases and developmental expression in the tobacco hornworm, *Manduca sexta*, and the malaria mosquito, *Anopheles gambiae*. *Insect Biochem Mol Biol* **34**: 29-41

Drapeau GR, Boily Y, & Houmard J (1972) Purification and properties of an extracellular protease of *Staphylococcus aureus*. *J Biol Chem* **247**: 6720-6726

Dunn PE & Drake D (1983) Fate of bacteria injected into naive and immunized larvae of the tobacco hornworm, *Manduca sexta*. *J Invertebr Pathol* **41**: 77-85

Furusawa T, Rakwal R, Nam HW, Hirano M, Shibato J, Kim YS, Ogawa Y, Yoshida Y, Kramer KJ, Kouzuma Y, Agrawal GK, & Yonekura M (2008) Systematic investigation of the hemolymph proteome of *Manduca sexta* at the fifth instar larvae stage using one- and two-dimensional proteomics platforms. *J. Proteome Res.* **7**: 938-959

Graveley BR (2005) Mutually exclusive splicing of the insect Dscam pre-mRNA directed by competing intronic RNA secondary structures. *Cell* **123**: 65-73

Graveley BR, Kaur A, Gunning D, Zipursky SL, Rowen L, & Clemens JC (2004) The organization and evolution of the dipteran and hymenopteran Down syndrome cell adhesion molecule (Dscam) genes. *RNA* **10**: 1499-1506

Green C, Levashina E, McKimmie C, Dafforn T, Reichhart JM, & Gubb D (2000) The necrotic gene in *Drosophila* corresponds to one of a cluster of three serpin transcripts mapping at 43A1.2. *Genetics* **156**: 1117-1127

Hashimoto C, Kim DR, Weiss LA, Miller JW, & Morisato D (2003) Spatial regulation of developmental signaling by a serpin. *Developmental Cell,* **5**: 945-950

Houmard J & Drapeau GR (1972) Staphylococcal protease: a proteolytic enzyme specific for glutamoyl bonds. *Proc Natl Acad Sci U S A* **69**: 3506-3509

House AE & Lynch KW (2008) Regulation of alternative splicing: more than just the ABCs. *J Biol Chem* **283**: 1217-1221

Jang IH, Chosa N, Kim SH, Nam HJ, Lemaitre B, Ochiai M, Kambris Z, Brun S, Hashimoto C, Ashida M, Brey PT, & Lee WJ (2006) A Spatzle-processing enzyme required for toll signaling activation in *Drosophila* innate immunity. *Dev Cell* **10**: 45-55

Jiang H & Kanost MR (2000) The clip-domain family of serine proteinases in arthropods. *Insect Biochem Mol Biol* **30**: 95-105

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

Jiang H, Wang Y, Gu Y, Guo X, Zou Z, Scholz F, Trenczek TE, & Kanost MR (2005) Molecular identification of a bevy of serine proteinases in *Manduca sexta* hemolymph. *Insect Biochem Mol Biol* **35**: 931-943

Jiang H, Wang Y, Huang Y, Mulnix AB, Kadel J, Cole K, & Kanost MR (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, & Kanost MR (1994) Mutually exclusive exon use and reactive center diversity in insect serpins. *J Biol Chem* **269**: 55-58

Jiang H, Wang Y, Yu XQ, Zhu Y, & Kanost M (2003) Prophenoloxidase-activating proteinase-3 (PAP-3) from *Manduca sexta* hemolymph: a clip-domain serine proteinase regulated by serpin-1J and serine proteinase homologs. *Insect Biochem Mol Biol* **33**: 1049-1060

Kanost M (1990) Serine protease inhibitors from the serpin gene family in *Manduca sexta* and *Drosophila melanogaster*. In *Molecular Insect Science,* Hagedorn, H.H. et al. (ed) pp 139-146. Plenum Press: New York

Kanost MR, Prasad SV, Huang Y, & Willott E (1995) Regulation of serpin gene-1 in *Manduca sexta*. *Insect Biochem Mol Biol* **25**: 285-291

Kanost MR, Prasad SV, & Wells MA (1989) Primary structure of a member of the serpin superfamily of proteinase inhibitors from an insect, *Manduca sexta*. *J Biol Chem* **264**: 965-972

Kim MS, Baek MJ, Lee MH, Park JW, Lee SY, Soderhall K, & Lee BL (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

Law RH, Zhang Q, McGowan S, Buckle AM, Silverman GA, Wong W, Rosado CJ, Langendorf CG, Pike RN, Bird PI, & Whisstock JC (2006) An overview of the serpin superfamily. *Genome Biol* **7**: 216

Li J, Wang Z, Canagarajah B, Jiang H, Kanost M, & Goldsmith EJ (1999) The structure of active serpin 1K from *Manduca sexta*. *Structure* **7**: 103-109

Ligoxygakis P, Roth S, & Reichhart JM (2003) A serpin regulates dorsal-ventral axis formation in the *Drosophila* embryo. *Curr Biol* **13**: 2097-2102

Michel K, Suwanchaichinda C, Morlais I, Lambrechts L, Cohuet A, Awono-Ambene PH, Simard F, Fontenille D, Kanost MR, & Kafatos FC (2006) Increased melanizing activity

in *Anopheles gambiae* does not affect development of *Plasmodium falciparum*. *Proc Natl Acad Sci U S A* **103**: 16858-16863

Molnar K, Holderith Borhegyi N, Csikos G, & Sass M (2001) Distribution of serpins in the tissues of the tobacco hornworm (*Manduca sexta*) larvae. Existence of new serpins possibly encoded by a gene distinct from the serpin-1 gene. *J Insect Physiol* **47**: 675-687

Myung JK, Frischer T, Afjehi-Sadat L, Pollak A, & Lubec G (2007) Mass spectrometrical analysis of the processed metastasis-inducing anterior gradient protein 2 homolog reveals 100% sequence coverage. *Amino Acids*

Neves G, Zucker J, Daly M, & Chess A (2004) Stochastic yet biased expression of multiple Dscam splice variants by individual cells. *Nat Genet* **36**: 240-246

Oley M, Letzel MC, & Ragg H (2004) Inhibition of furin by serpin Spn4A from *Drosophila melanogaster*. *FEBS Lett* **577**: 165-169

Olson S, Blanchette M, Park J, Savva Y, Yeo GW, Yeakley JM, Rio DC, & Graveley BR (2007) A regulator of Dscam mutually exclusive splicing fidelity. *Nat Struct Mol Biol*

Osterwalder T, Kuhnen A, Leiserson WM, Kim YS, & Keshishian H (2004) *Drosophila* serpin 4 functions as a neuroserpin-like inhibitor of subtilisin-like proprotein convertases. *J Neurosci* **24**: 5482-5491

Packer NH, WIlkins MR, Golaz O, Lawson MA, Gooley AA, Hochstrasser DF, Redmond JW, & Williams KL (1996) Characterization of human plasma glycoproteins separated by two-dimensional gel electrophoresis. *Biotechnology (N Y)* **14**: 66-70

Park JW, Parisky K, Celotto AM, Reenan RA, & Graveley BR (2004) Identification of alternative splicing regulators by RNA interference in *Drosophila*. *Proc Natl Acad Sci U S A* **101**: 15974-15979

Peterson AM, Fernando GJ, & Wells MA (1995) Purification, characterization and cDNA sequence of an alkaline chymotrypsin from the midgut of *Manduca sexta*. *Insect Biochem Mol Biol* **25**: 765-774

Richer MJ, Keays CA, Waterhouse J, Minhas J, Hashimoto C, & Jean F (2004) The Spn4 gene of *Drosophila* encodes a potent furin-directed secretory pathway serpin. *Proc Natl Acad Sci U S A* **101**: 10560-10565

Rosenfeld J, Capdevielle J, Guillemot JC, & Ferrara P (1992) In-gel digestion of proteins for internal sequence analysis after one- or two-dimensional gel electrophoresis. *Anal Biochem* **203**: 173-179

Rowen L, Young J, Birditt B, Kaur A, Madan A, Philipps DL, Qin S, Minx P, Wilson RK, Hood L, & Graveley BR (2002) Analysis of the human neurexin genes: alternative splicing and the generation of protein diversity. *Genomics* **79**: 587-597

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

Satoh D, Horii A, Ochiai M, & Ashida M (1999) Prophenoloxidase-activating enzyme of the silkworm, *Bombyx mori*. *JBC* **274**: 7441-7453

Schmucker D (2007) Molecular diversity of Dscam: recognition of molecular identity in neuronal wiring. *Nat Rev Neurosci* **8**: 915-920

Smith CW (2005) Alternative splicing--when two's a crowd. *Cell* **123**: 1-3

Sorensen SB, Sorensen TL, & Breddam K (1991) Fragmentation of proteins by S. aureus strain V8 protease. Ammonium bicarbonate strongly inhibits the enzyme but does not improve the selectivity for glutamic acid. *FEBS Lett* **294**: 195-197

Stamm S, Ben-Ari S, Rafalska I, Tang Y, Zhang Z, Toiber D, Thanaraj TA, & Soreq H (2005) Function of alternative splicing. *Gene* **344**: 1-20

Tarn WY (2007) Cellular signals modulate alternative splicing. *J Biomed Sci* **14**: 517-522

Tong Y, Jiang H, & Kanost MR (2005) Identification of plasma proteases inhibited by *Manduca sexta* serpin-4 and serpin-5 and their association with components of the prophenol oxidase activation pathway. *J Biol Chem* **280**: 14932-14942

Tong Y & Kanost MR (2005) *Manduca sexta* serpin-4 and serpin-5 inhibit the prophenol oxidase activation pathway: cDNA cloning, protein expression, and characterization. *J Biol Chem* **280**: 14923-14931

Wang Y & Jiang H (2006) Interaction of beta-1,3-glucan with its recognition protein activates hemolymph proteinase 14, an initiation enzyme of the prophenoloxidase activation system in *Manduca sexta*. *J Biol Chem* **281**: 9271-9278

Wang Y & Jiang H (2004) Purification and characterization of *Manduca sexta* serpin-6: a serine proteinase inhibitor that selectively inhibits prophenoloxidase-activating proteinase-3. *Insect Biochem Mol Biol* **34**: 387-395

Whisstock JC & Bottomley SP (2006) Molecular gymnastics: serpin structure, folding and misfolding. *Curr Opin Struct Biol* **16**: 761-768

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

Zou Z & Jiang H (2005) *Manduca sexta* serpin-6 regulates immune serine proteinases PAP-3 and HP8. cDNA cloning, protein expression, inhibition kinetics, and function elucidation. *J Biol Chem* **280**: 14341-14348

Tables

Table 4-1. Known and predicted P1-P1' residues and protease specificity of serpin-1 isoforms.

Predicted P1-P1` residues are shown in italics and baced on Jiang et al. 1994. The specificity information and known P1-P1’ residues is from Jiang and Kanost 1997. Serpin-1D failed to cleave any of the commercially available proteases tested.

Serpin-1	P1-P1’	Specificity
A	R-Q	Trypsin
B	A-S	Elastase
C	*Y-S*	Chymotrypsin-like
D	*R-P*	Possibly trypsin-like
E	*K-K*	Trypsin-like
F	V-D	Elastase
G	*I-T*	Elastase-like
H	Y-V	Chymotrypsin
I	L-S	Chymotrypsin & Elastase
J	R-C	Trypsin-like
K	Y-S	Chymotrypsin
Z	Y-L	Chymotrypsin

Table 4-2. Primers used for RT-PCR to amplify transcripts from serpin-1 isoforms. Isoform specific forward primers were made to the 12 alternatively-spliced exon 9 variants and were used with a common reverse primer complementary to the 3' untranslated region, encoded by exon 10. The predicted size for the product of PCR from each exon-9 specific forward primer with the common reverse primer is listed.

Location	Primers	Primer sequences (5' to 3')	Product size (bp)
Exon 9A	Spn1A	CATAACACGACAAGCGAGAC	160
Exon 9B	Spn1B	TACCGGCGAGTTTGATACTA	160
Exon 9C	Spn1C	TTCTTCGTATGAACCTGTCG	161
Exon 9D	Spn1D	AAGTGTACGTCCACCCACTC	158
Exon 9E	Spn1E	AGTTATACCGCCCGTCCTA	155
Exon 9F	Spn1F	TTCATCGCTGTCGTAGACTC	192
Exon 9G	Spn1G	TTGAACCACCTGTTATCGAA	153
Exon 9H	Spn1H	CGTGGAATCCATAGACAATTT	178
Exon 9I	Spn1I	TCGGTATAGTTGCGCTATCA	179
Exon 9J	Spn1J	TGACAGACAGATGTTGTTCTGA	177
Exon 9K	Spn1K	TTCATTTCGTGCCAAAAGT	150
Exon 9Z	Spn1Z	TTGGCATAGCATATCTGTCG	167
Exon 10	Spn1Rev	ATTCCAACCGGACGTTATT	

Table 4-3. Mass spectrometry analysis of 1D bands 1, 2, and 3 using Aldente for PMF and Mascot for MS/MS analysis.

Bands as marked in Figure 2-3. Aldente protein scores over 16.31 are statistically significant. Experimental m/z mass values are only shown for serpin-1 isoform specific peptides. Isoform specific peptides supported by further fragmentation by MALDI-TOF/TOF and analysis by Mascot have an additional score and E value; peptides that failed to return a hit with an E value of less than 1 are marked "F" in place of the Mascot score. Sequence coverages are based on the entire sequence in the database.

Spot #	Name	Spn-1 Isoform specific peptides	Accsn. #	Score	Matched/ searched	Seq. Cov	Spn-1 specific peptide m/z	Modif.	Position	Mascot score	Mascot E-value
Band1	PPO1	n/a	O44249	409	45/191	69%					
"	PPO2	n/a	Q25519	22	15/146	24%					
"	Serpin 1	n/a	Q25500	16.6	12/131	39%					
Band 2	Serpin 1	J (2)	Q25499	151	20/160	63%	2422.192		337-359	41	5.70E-03
"	"	"					1870.935	1/1 CAM	382-397	F	
"	Serpin 1	A (1)	Q25498	129	19/160	60%	2341.15		337-358	F	
Band 3	Serpin 1	K (1)	Q25491	226	26/177	70%	1951.941		337-355	64	3.70E-05
"	"	"	"	"	"	"	3169.631		326-355	F	
"	Chymotrypsin	n/a	Q25503	29	12/151	38%					

Table 4-4. Aldente PMF search of MALDI-TOF data from putative serpin-1 spots from *M. sexta* whole plasma (spots as shown in Figure 4-7).

Spot #	Name	Isoform specific peptides	Accession #	Score	Matched/ searched masses	Cov %	Mw (kDa)	pI
SpER1	Serpin 1.	J (2)	Q25499	36.2	14/218	39	44	4.9
SpER2	Serpin 1.	F (2)	Q25497	71.26	24/368	64	44	5.1
SpER3	Serpin 1.	F (2)	Q25497	43.09	12/210	50	44	5.1
SpER4	Serpin 1.	I (1)	Q25493	92.77	18/239	52	44	5.2
"	Serpin 1.	F (1)	Q25497	46.04	18/239	53	44	5.1
"	Serpin 1.	Z (1)	Q25492	44.83	18/239	53	43	5.2
SpER5	Serpin 1.	I (1)	Q25493	36.59	21/399	45	44	5.2
	Serpin 1.	Z (1)	Q25492	31.81	23/399	54	43	5.2
	Serpin 1.	K (1)	Q25491	20.07	21/399	48	44	5.6
SpER6	Serpin 1.	Z (2)	Q25492	140.49	24/181	62	43	5.2
SpER7	Serpin 1.	Z (4)	Q25492	182.51	22/181	54	43	5.2
SpER8	Serpin 1.	K (3)	Q25491	52.55	18/262	54	44	5.6
"	Serpin 1.	A (2)	Q25498	49.66	17/262	55	44	5.6
SpER9	Serpin 1.	A (4)	Q25498	121.79	22/233	52	44	5.6
SpER10	Serpin 1.	E (3)	Q25500	127.93	25/312	58	44	5.9
	Serpin 1.	A (2)	Q25498	81.23	24/312	59	44	5.6
	Serpin 1.	K (1)	Q25491	75.17	23/312	58	44	5.6
SpER11	Serpin 1.	E (3)	Q25500	160.07	31/327	73	44	5.9

Table 4-5. Mascot search of MALDI-TOF/TOF data from serpin-1 peptides identified in Table 4-4.

Sample name	Isoform	Obsrv. Mass	Start	End	Sequence	CI %	E value
SpER1	J	2422.190	337	359	AFIEVNEEGAEAAAANAFILTDR	16	8.4E-01
SpER4	I	1543.850	370	381	FVVNKPFYFNIR	93.6	6.4E-02
SpER5	I	1544.000	370	381	FVVNKPFYFNIR	88	1.2E-01
SpER7	Z	2867.469	337	364	AFIEVNEEGAEAAAANAFGIAYLSAVIR	98.4	1.6E-02
SpER9	A	1627.940	366	378	YFVANKPFIFLLR	90.8	9.2E-02
SpER10	E	1973.990	337	355	AFIEVNEEGAEAAAANVIR	99.98	2.1E-04
SpER11	E	1973.990	337	355	AFIEVNEEGAEAAAANVIR	98	2.0E-02

Table 4-6. Summary of mass spectrometry results for identification of serpin-1 isoforms from spots cut from 2D-PAGE of whole *M. sexta* plasma.

	Serpin-1 isoforms detected	
Spot	MALDI-TOF (Aldente)	MALDI-TOF/TOF (Mascot)
SpER1	J (2)	-
SpER2	F (2)	-
SpER3	F (2)	-
SpER4	Serpin-1	I (1)
SpER5	Serpin-1	-
SpER6	Z (2)	-
SpER7	Z (4)	Z (1)
SpER8	K (3), A(2)	-
SpER9	A (4)	A (1)
SpER10	E (3), A (2)	E (1)
SpER11	E (3)	E (1)

Numbers in parenthesis indicate the number of isoform specific peptides detected by the MS methods.

Table 4-7. Summary of serpin-1 isoform-specific peptides expected from different digestive enzymes.

The number of serpin-1 isoform-specific peptides expected assuming 0 missed cleavages, peptide mass between 700 and 3500 Da, and peptides with possible modifications only listed once.

Enzyme	Serpin-1 isoform											
	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**Z**
Trypsin	3	1	1	4	4	3	1	1	3	3	4	3
GluC Bicarb.	0	3	1	2	0	2	2	3	2	3	2	0
GluC Phos.	2	3	3	3	3	2	2	4	2	3	2	3
LysC/AspN	3	3	1	3	4	4	1	4	2	2	4	3

Table 4-8. Peptide mass fingerprinting results using Mascot and NCBInr for immunoaffinity-purified serpin-1 separated by 2D-PAGE.

List of top Mascot PMF hits from the NCBInr database for the spots identified as *M. sexta* protein. Mascot protein scores are for the entire protein sequence, CI% is the confidence interval. Isoform designations are indicated if there are at least 2 isoform-specific peptides detected by PMF or one isoform-specific peptide detected by both PMF and MS/MS. MS/MS results listed in more detail in Table 4-10.

Trypsin spot	Protein Name	Accession No.	Isoform	Protein Score	Total Ion Score	Total Ion C.I.%	Protein Score C.I.%	Pep. Count	Isoform spec. pep. by PMF	Isoform spec. pep. by MS/MS
2t	serpin 1	gi|1378130	J	1,080	887	100	100	21	2 (J)	1
3t	serpin 1	gi|1378130	?	894	649	100	100	25	0	0
4t	serpin 1	gi|1378122	H	1,010	841	100	100	21	1 (H)	1
5t	serpin 1	gi|1378130	?	991	811	100	100	20	0	0
6t	serpin 1	gi|1378123	?	1,010	864	100	100	19	1 (F)	0
7t	serpin 1	gi|1378127	I	1,000	768	100	100	24	2 (I)	0
8t	serpin 1	gi|1378127	I	896	735	100	100	19	2 (I)	1
9t	serpin 1	gi|1378127	I	1,040	867	100	100	21	2 (I)	2
10t	serpin 1	gi|1378132	A	990	824	100	100	19	1 (A)	1
10t	serpin 1	gi|1378128	Z	878	683	100	100	21	3 (Z)	1
41t	serpin 1	gi|1378128	Z	995	812	100	100	20	3 (Z)	2
11t	serpin 1	gi|1378124	K	1,090	866	100	100	23	3 (K)	3
12t	serpin 1	gi|1378132	A	1,170	970	100	100	21	4 (A)	3
12t	serpin 1	gi|1378125	E	825	662	100	100	19	2 (E)	0
13t	serpin 1	gi|1378125	E	894	721	100	100	21	3 (E)	2
14t	serpin 1	gi|1378125	E	1,000	831	100	100	21	4 (E)	2
16t	serpin 1	gi|1378126	?	706	510	100	100	22	1 (G)	0
17t	serpin 1	gi|1378125	E	933	731	100	100	22	2 (E)	0
18t	serpin 1	gi|431347	?	486	354	100	100	17	0	0
19t	serpin 1	gi|1378130	J	585	453	100	100	17	1 (J)	1
19t	serpin 1	gi|1378132	?	453	321	100	100	17	1 (A)	0
20t	serpin 1	gi|1378132	?	676	517	100	100	19	1 (A)	0
21t	serpin 1	gi|134436	?	126	59	99.947	100	11	0	0
30t	serpin 1	gi|1378132	?	120	80	100	100	8	0	0
31t	serpin 1	gi|1378125	?	263	90	100	100	19	1 (E)	0

32t	serpin 1	gi|1378125	?	223	127	100	100	13	1 (E)	0
33t	serpin 1	gi|1378125	E	717	463	100	100	24	3 (E)	1
34t	serpin 1	gi|1378125	E	540	315	100	100	23	3 (E)	1
35t	Phenoloxidase subunit 1	gi|74763772	n/a	382	192	100	100	24	n/a	n/a
35t	Phenoloxidase subunit 2	gi|75038472	n/a	118	28	41.356	100	16	n/a	n/a

LysC/AspN Spot	Protein Name	Accession No.	Isoform	Protein Score	Total Ion Score	Total Ion C.I.%	Protein Score C.I.%	Pep. Count	Isoform spec. pep. By PMF	Isoform spec. pep. By MS/MS
2	serpin 1	gi|134436	?	118	50	98.933	100	9	0	0
3	serpin 1	gi|134436	?	100	20	0	99.939	11	0	0
4	serpin 1	gi|1378122	H	462	385	100	100	13	2 (H)	2
5	serpin 1	gi|134436	?	358	233	100	100	14	0	0
6	serpin 1	gi|134436	B	404	295	100	100	16	2 (B)	1
6	serpin 1	gi|1378123	F	380	272	100	100	16	2 (F)	0
7	serpin 1	gi|1378127	?	272	172	100	100	13	0	0
8	serpin 1	gi|1378123	F	403	287	100	100	17	2 (F)	0
9	serpin 1	gi|1378127	?	331	244	100	100	14	0	0
10	serpin 1	gi|431347	?	184	72	99.992	100	15	0	0
41	serpin 1	gi|1378128	Z	413	336	100	100	13	2 (Z)	1
11	serpin 1	gi|1378124	K	272	196	100	100	13	2 (K)	1
12	serpin 1	gi|1378125	E	278	183	100	100	15	2 (E)	0
13	serpin 1	gi|1378125	E	447	353	100	100	13	3 (E)	2
14	serpin 1	gi|1378125	E	509	400	100	100	16	4 (E)	3
19	hemolymph proteinase 8 [Manduca sexta]	gi|56418397		81	60	99.878	95.862	6	n/a	0

GluC Bicarbonate Spot	Protein Name	Accession No.	Isoform	Protein Score	Total Ion Score	Total Ion C.I.%	Protein Score C.I.%	Pep. Count	Isoform spec. pep. By PMF	Isoform spec. pep. By MS/MS
2b	serpin 1	gi|1378130	J	154	80	99.999	100	11	2 (J)	1
4b	serpin 1	gi|134436	?	119	29	0	100	13	0	0
5b	serpin 1	gi|134436	?	108	25	0	99.991	13	0	0
6b	serpin 1	gi|134436	B	132	45	95.9	100	13	3 (B)	1
8b	serpin 1	gi|134436	?	121	33	24.457	100	13	0	0
9b	serpin 1	gi|134436	?	115	33	24.787	99.998	13	0	0
10b	serpin 1	gi|17942678	?	117			99.999	16	0	0
14b	serpin 1	gi|134436	?	107			99.989	14	0	0

GluC Phosphate Spot	Protein Name	Accession No.	Isoform	Protein Score	Total Ion Score	Total Ion C.I.%	Protein Score C.I.%	Pep. Count	Isoform spec. pep. By PMF	Isoform spec. pep. By MS/MS
4p	serpin 1	gi|1378122	H	162	28	0	100	16	2 (H)	0
4p	serpin 1	gi|134436	B	134			100	16	2 (B)	0
6p	serpin 1	gi|134436	B	118	42	91.19	100	13	2 (B)	1
7p	serpin 1	gi|134436	?	80			95.025	10	1 (B)	0
8p	serpin 1	gi|1378126	?	132			100	14	1 (G)	0
9p	serpin 1	gi|1378127	I	225	142	100	100	14	2 (I)	0
11p	serpin 1	gi|1378124	?	118			100	12	0	0
12p	serpin 1	gi|134436	?	301	109	100	100	16	0	0
13p	serpin 1	gi|134436	?	240	129	100	100	16	0	0
14p	serpin 1	gi|1378129	?	175	52	99.399	100	14	0	0

Table 4-9. Mascot peptide mass fingerprinting search of *M. sexta* protease catalytic domains

Mascot hits with a protein score of 50% or greater from searching known *M. sexta* protease catalytic domain sequences. Corresponding MS/MS results are highlighted in Table 4-10.

Trypsin spot	Protein Name	Accession No.	Protein Score	Protein Score C.I.%	Total Ion Score	Total Ion C.I.%	Protein MW	Protein PI	Pep. Count	Pep. By MS/MS
19t	HP8	gi|56418397	19	64.927			28334.1	5.18	4	0
20t	HP8	gi|56418397	18	61.543			28334.1	5.18	4	0
21t	HP1	tr|O44330|	36	99.418	18	99.992	26622.9	5.86	4	2
31t	HP8	gi|56418397	94	100	62	100	28334.1	5.18	6	2
32t	HP8	gi|56418397	33	98.784	9	99.978	28334.1	5.18	5	1

Table 4-10. MS/MS results from immunoaffinity-purified serpin-1

The majority of the proteins were detected by Mascot by searching the NCBInr database. The three peptides detected in a Mascot search of *M. sexta* catalytic protease domains are highlighted in pink. Two additional MS/MS spectra were detected by a combination of Mascot and XTandem analysis using the scaffold program and are underlined.

MS/MS results from Mascot search of NCBI (regular), *M. sexta* activated protease sequences (pink), or Scaffold MS/MS by Mascot and Xtandem (underlined)

Trypsin Spot	Isoform	Isoform Spec. pep. By MS/MS	Calc. Mass	Obsrv. Mass	± da	± ppm	Start	End	Sequence	Ion score	C. I. %
2t	J	1	2422.1885	2422.175	-0.014	-6	337	359	AFIEVNEEGAEAAAANAFILTDR	174	100
4t	H	1	1751.8785	1752.0287	0.1502	86	382	396	ANDLYLFNGICVQPK	93	100
6t	F	1	1881.9646	1882.1062	0.1416	75	367	381	TIEFHADRPFFFNLK	24	0
7t	I	1	1543.8419	1543.8444	0.0025	2	370	381	FVVNKPFYFNIR	29	52.19
9t	I	2	1543.8419	1543.9646	0.1227	79	370	381	FVVNKPFYFNIR	61	99.922
"	"	"	1618.743	1618.8699	0.1269	78	382	395	SNGQHLFNGICFQP	92	100
10t	A	1	2341.1458	2341.136	-0.01	-4	337	358	AFIEVNEEGAEAAAANAFITR	206	100
10t	Z	1	1792.9169	1792.9185	0.0016	1	365	379	SPVFNADHPFVFFLR	65	99.989
41t	Z	2	1792.9169	1792.9272	0.0103	6	365	379	SPVFNADHPFVFFLR	80	100
"	"	"	2867.4573	2867.4468	-0.011	-4	337	364	AFIEVNEEGAEAAAANAFGIAYLSAVIR	108	100
11t	K	3	1385.7252	1385.7233	-0.002	-1	356	366	ITFYSFHFVPK	54	99.849
"	"	"	1468.8198	1468.822	0.0022	1	367	378	VEINKPFFFSLK	51	99.725
"	"	"	1951.9396	1951.9462	0.0066	3	337	355	AFIEVNEEGAEAAAANAFK	126	100
12t	A	3	1326.7205	1326.7213	0.0008	1	379	390	FNGLALFNGVFK	74	99.999
"	"	"	1627.9359	1627.9403	0.0044	3	366	378	YFVANKPFIFLLR	34	87.097
"	"	"	2341.1458	2341.1404	-0.005	-2	337	358	AFIEVNEEGAEAAAANAFITR	201	100
13t	E	2	1068.6927	1068.7885	0.0958	90	361	369	FRVIPPVLK	42	92.536
"	"	"	1973.9927	1974.1367	0.144	73	337	355	AFIEVNEEGAEAAAANVIR	169	100

14t	E	2	1566.8215	1566.949	0.1275	81	370	381	FHVDRPFFFNLK	54	99.506
"	"	"	1973.9927	1974.144	0.1513	77	337	355	AFIEVNEEGAEAAAANVIR	184	100
19t	J	1	2422.1885	2422.1807	-0.008	-3	337	359	AFIEVNEEGAEAAAANAFILTDR	132	100
21t	HP1	2	1027.5208	1027.5305	0.0097	9	56	63	LGEYDFKR	12	99.971
21t	HP1	"	1335.6804	1335.689	0.0086	6	46	55	RWEANELYVR	6	99.876
31t	HP8	2	1161.6085	1161.6031	-0.005	-5	39	48	YVLTAAHCVK	41	100
"	"	"	1588.8363	1588.8431	0.0068	4	184	198	LITDKQLCAGGVEGK	21	99.998
32t	HP8	1	1037.5374	1037.5372	-2E-04	0	176	183	SVYERVER	9	99.978
33t	E	1	1973.9927	1973.9961	0.0034	2	337	355	AFIEVNEEGAEAAAANVIR	74	99.999
34t	E	1	1973.9927	1973.9965	0.0038	2	337	355	AFIEVNEEGAEAAAANVIR	50	99.694

LysC/AspN Spot	Isoform	Isoform Spec. pep. By MS/MS	Calc. Mass	Obsrv. Mass	± da	± ppm	Start	End	Sequence	Ion score	C. I. %
4	H	2	1566.7985	1566.8564	0.0579	37	384	396	DLYLFNGICVQPK	101	100
"		"	1412.7321	1412.7864	0.0543	38	371	381	DVNRPFYFNLK	63	100
6	B	1	1214.6204	1214.6744	0.054	44	370	378	DRPFYFELK	50	98.839
41	Z	1	1305.6738	1305.7319	0.0581	44	371	380	DHPFVFFLRQ	64	99.948
11	K	1	1385.7252	1385.7603	0.0351	25	356	366	ITFYSFHFVPK	44	94.756
13	E	2	1068.6927	1068.7197	0.027	25	361	369	FRVIPPVLK	51	98.941
"	"	"	1183.6259	1183.6586	0.0327	28	373	381	DRPFFFNLK	49	98.374
14	E	3	1068.6927	1068.7067	0.014	13	361	369	FRVIPPVLK	50	98.659
"	"	"	1183.6259	1183.6438	0.0179	15	373	381	DRPFFFNLK	49	98.358
"	"	"	2300.2244	2300.2388	0.0144	6	337	358	AFIEVNEEGAEAAAANVI	49	98.471
19	HP8	1	1458.7198	1458.7936	0.0738	51	124	134	DEHPWMVLLRY	60	99.878

GluC bicarbonate spot	Isoform	Isoform Specific peptides	Calc. Mass	Obsrv. Mass	± da	± ppm	Start	End	Sequence	Ion score	C. I. %
2b	J	1	1526.8036	1526.8453	0.0417	27	385	397	HLLFSGICIQPEI	81	99.999
6b	B	1	1322.6528	1322.7026	0.0498	38	367	376	VHIDRPFYFE	45	95.9

19b	HP8	1	1333.6820	1334.6900	0.0194	15	260	271	VAGWGKTETRTE	44	95.000

GluC phos.	Isoform	Isoform Specific peptides	Calc. Mass	Obsrv. Mass	± da	± ppm	Start	End	Sequence	Ion score	C. I. %
6p	B	1	1322.6528	1322.6995	0.0467	35	367	376	VHIDRPFYFE	42	91.19

Table 4-11. Summary of MS/MS results from immunoaffinity-purified serpin-1.

Spot #	Trypsin	LysC/AspN	GluC bicarbonate	GluC phosphate
2	J (1)		J (1)	
4	H (1)	H (2)		
6		B (1)	B (1)	B (1)
9	I (2)			
10	Z (1)			
41	Z (2)	Z (1)		
11	K (3)	K (1)		
12	A (2)			
13	E (2)	E (2)		
14	E (3)	E (3)		
19	J (1)	HP8 (1)	HP8 (1)	
21	HP1(2)			
31	HP8 (2)			
32	HP8 (1)			
33	E (1)			
34	E (1)			

Serpin-1 isoform is indicated by letter. The number of peptides confirmed by MS/MS with a 90% confidence or greater is listed in the parentheses.

Figures

Figure 4-1. Gene map, amino acid sequence, and structure of serpin-1.

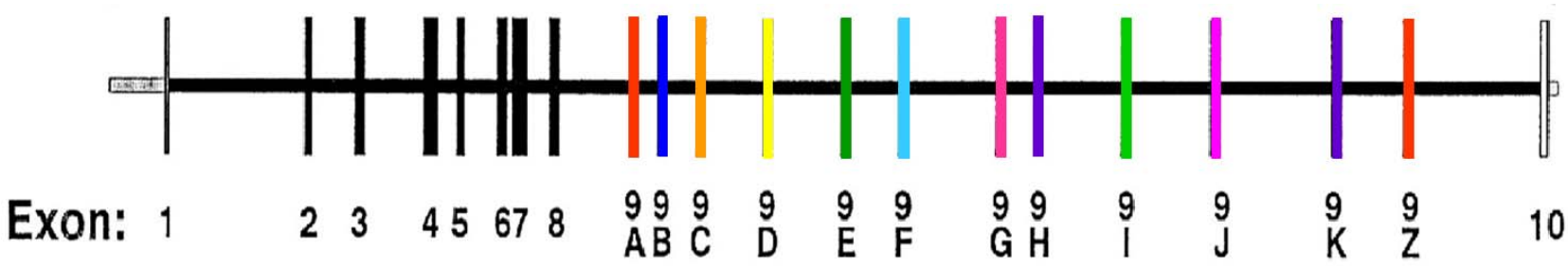
A) Serpin-1 gene map (Jiang et al. 1994). Common exons are shown as black (1-8) and white (10) bars. Exon 10 encodes the serpin-1 3' untranslated region and the 5' untranslated region is shown in gray. The alternatively spliced variants of exon 9 are indicated by colored bars (9A-Z). Exons are connected by introns (black lines).

B) Amino acid sequence of full length serpin-1A. The serpin-1A specific region is highlighted in cyan; the rest of the protein is the same in all 12 serpin-1 isoforms. The signal peptide is shown in red italics and the putative glycosylation site, Asn85, is highlighted in yellow. The reactive center loop from P15 to P5' is underlined and the protease cleavage site, between P1 and P1', is marked with a line.

C) An alignment of the carboxyl-termini of the 12 serpin-1 isoforms. The first shown residue is Lys 336 and an arrow marks the beginning of exon 9, at residue 353. Known P1 residues are underlined. The first isoform-specific tryptic peptide for each isoforms is shown in italics and dotted underline.

D) Protein structure of serpin-1K (PDB 1SEK). The alternatively spliced region is shown in green and the P1 residue, Tyr359, is shown in the "stick" format

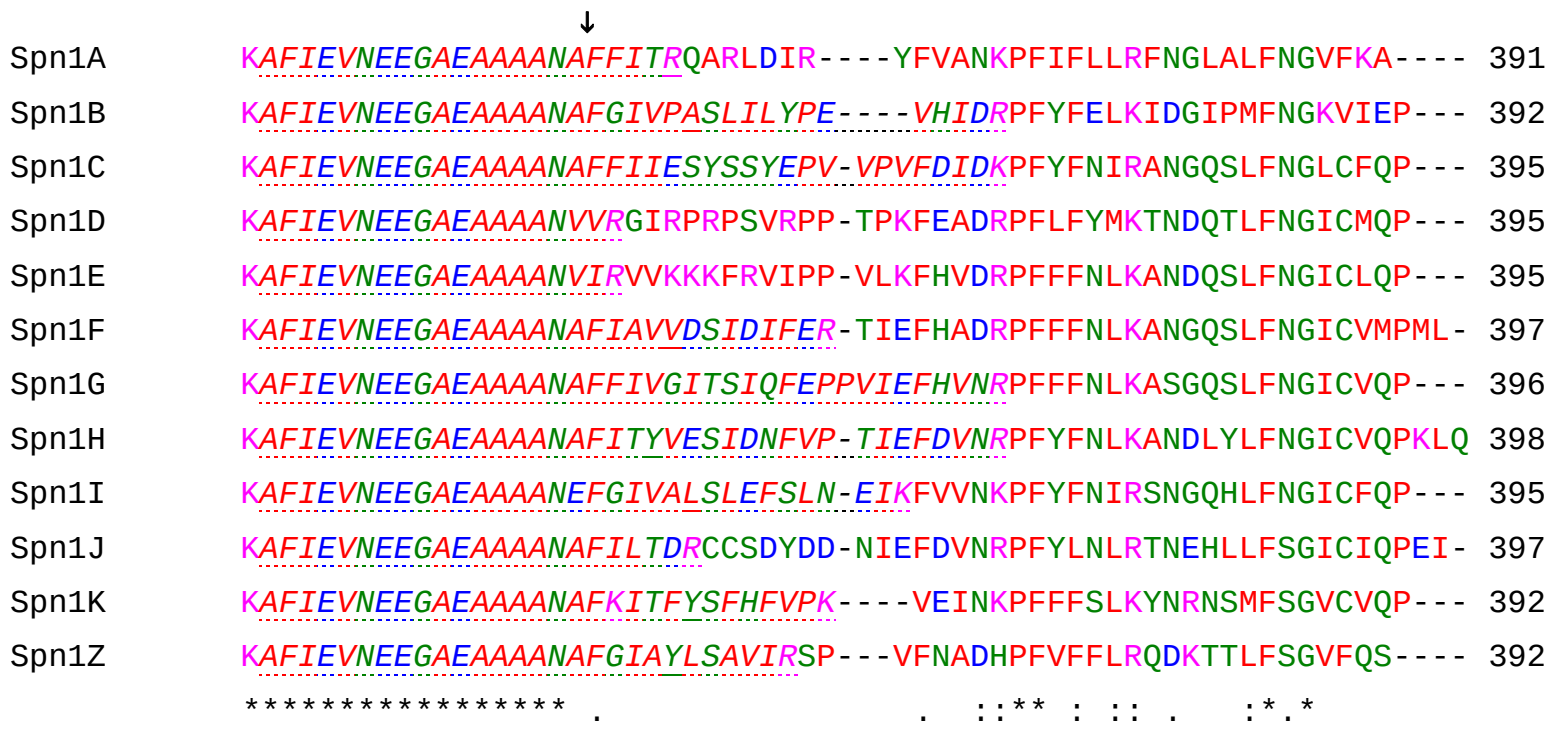
A.



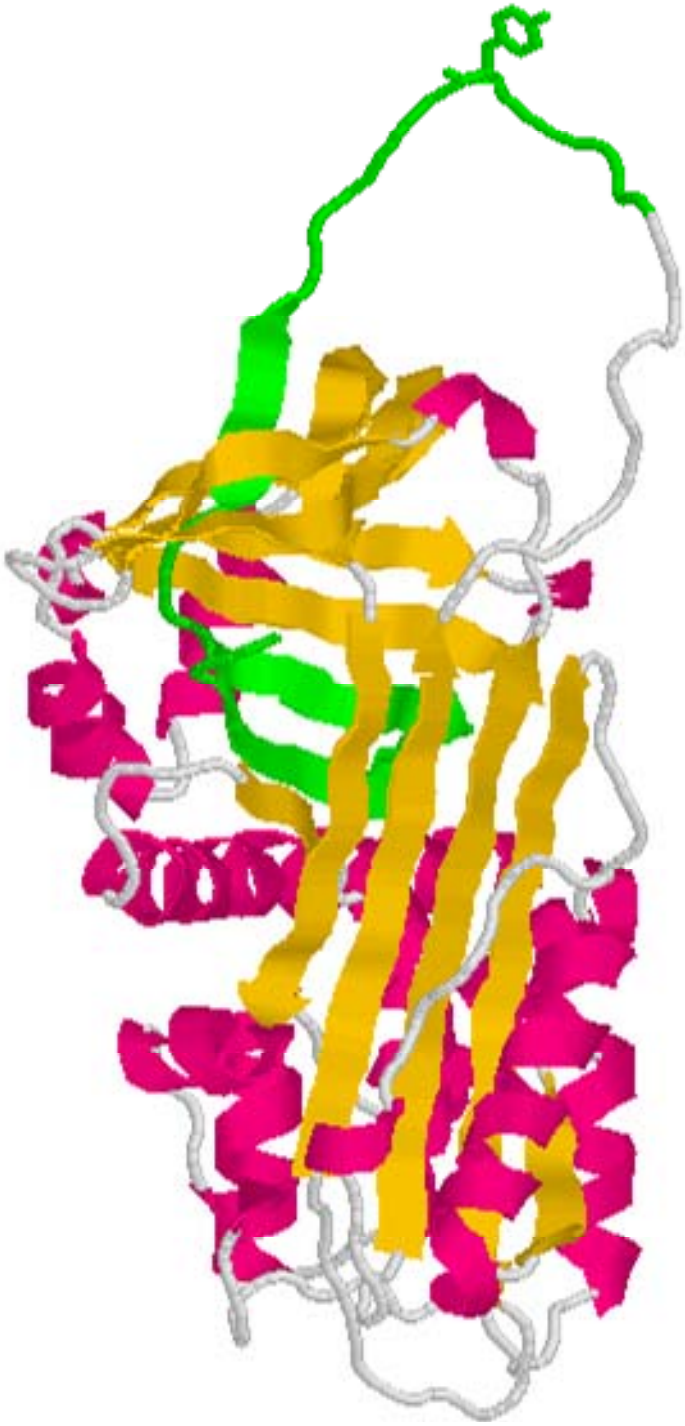
B.

```
  1  MKIIMCIFGL AALAMAGETD LQKILRESND QFTAQMFSEV VKANPGQNVV LSAFSVLPPL GQLALASVGE SHDELLRALA
 81  LPNDNVTKDV FADLNRGVRA VKGVDLKMAS KIYVAKGLEL NDDFAAVSRD VFGSEVQNVD FVKSVEAAGA INKWVEDQTN
161  NRIKNLVDPD ALDETTRSVL VNAIYFKGSW KDKFVKERTM DRDFHVSKDK TIKVPTMIGK KDVRYADVPE LDAKMIEMSY
241  EGDQASMIII LPNQVDGITA LEQKLKDPKA LSRAEERLYN TEVEIYLPKF KIETTTDLKE VLSNMNIKKL FTPGAARLEN
321  LLKTKESLYV DAAIQKAFIE VNEEGAEAAA ANAFFITR|QA RLDIRYFVAN KPFIFLLRFN GLALFNGVFK A
                                          P1 P1'
```

C.


↓
Spn1A KAFIEVNEEGAEAAAANAFFITRQARLDIR----YFVANKPFIFLLRFNGLALFNGVFKA---- 391
Spn1B KAFIEVNEEGAEAAAANAFGIVPASLILYPE----VHIDRPFYFELKIDGIPMFNGKVIEP--- 392
Spn1C KAFIEVNEEGAEAAAANAFFIIESYSSYEPV-VPVFDIDKPFYFNIRANGQSLFNGLCFQP--- 395
Spn1D KAFIEVNEEGAEAAAANVVRGIRPRPSVRPP-TPKFEADRPFLFYMKTNDQTLFNGICMQP--- 395
Spn1E KAFIEVNEEGAEAAAANVIRVVKKKFRVIPP-VLKFHVDRPFFFNLKANDQSLFNGICLQP--- 395
Spn1F KAFIEVNEEGAEAAAANAFIAVVDSIDIFER-TIEFHADRPFFFNLKANGQSLFNGICVMPML- 397
Spn1G KAFIEVNEEGAEAAAANAFFIVGITSIQFEPPVIEFHVNRPFFFNLKASGQSLFNGICVQP--- 396
Spn1H KAFIEVNEEGAEAAAANAFITYVESIDNFVP-TIEFDVNRPFYFNLKANDLYLFNGICVQPKLQ 398
Spn1I KAFIEVNEEGAEAAAANEFGIVALSLEFSLN-EIKFVVNKPFYFNIRSNGQHLFNGICFQP--- 395
Spn1J KAFIEVNEEGAEAAAANAFILTDRCCSDYDD-NIEFDVNRPFYLNLRTNEHLLFSGICIQPEI- 397
Spn1K KAFIEVNEEGAEAAAANAFKITFYSFHFVPK----VEINKPFFFSLKYNRNSMFSGVCVQP--- 392
Spn1Z KAFIEVNEEGAEAAAANAFGIAYLSAVIRSP---VFNADHPFVFFLRQDKTTLFSGVFQS---- 392
********** . . ::** : :: . :*.*

D.



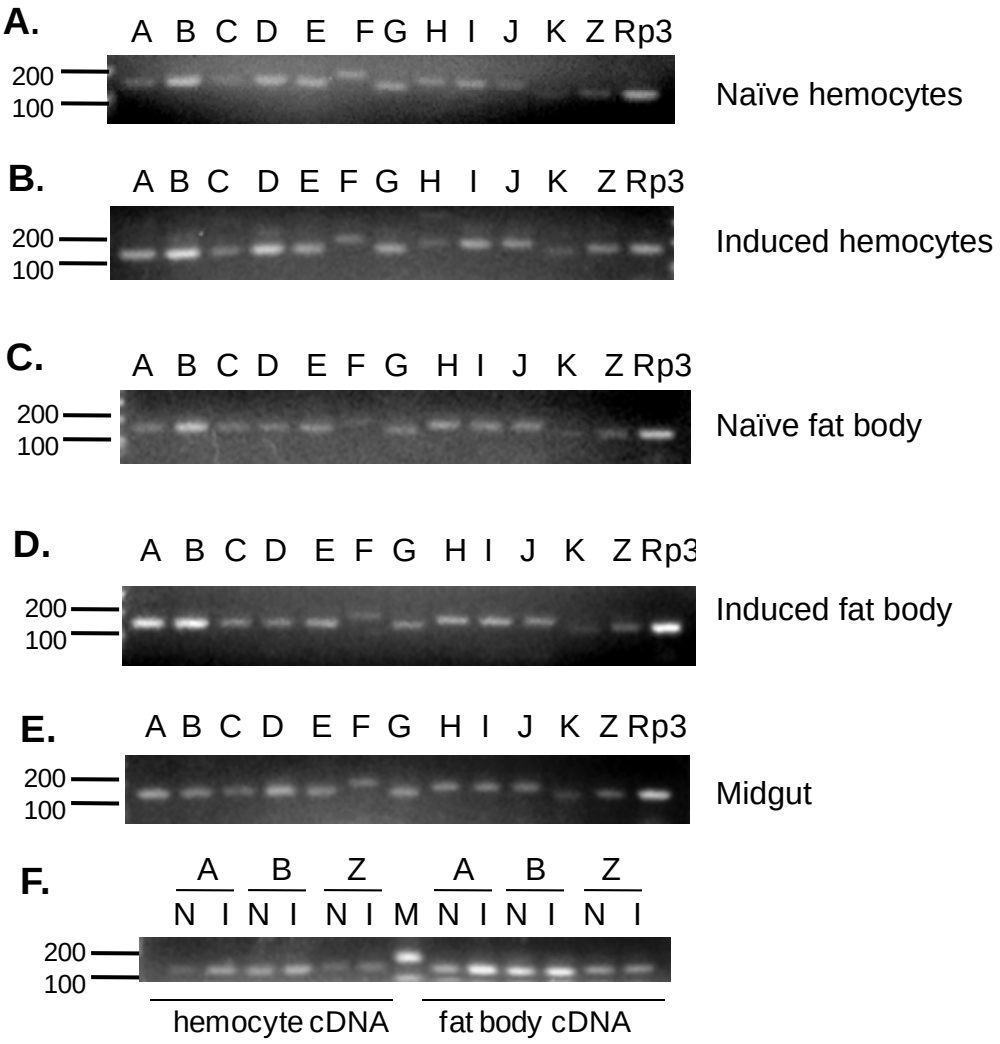


Figure 4-2. Semi-quantitative reverse-transcriptase PCR of serpin-1 isoforms. Each lane is marked with the exon-9 primer used for the PCR reaction. All reactions also used a common serpin-1 primer from exon 10. A) Naive hemocyte cDNA, 35 cycles. B) Induced hemocyte cDNA, 35 cycles. C) Naive fat body cDNA, 30 cycles. D) Induced fat body cDNA, 30 cycles. E) Midgut cDNA, 30 cycles. F) Regular or induced isoform-specific samples A, B, and Z, hemocyte samples on left, fat body on right. Rp3: ribosomal protein 3 primers, run for 18 cycles in A and B, 30 cycles in C, D, and E; M: marker. Serpin-1 primers used and expected band sizes are listed in Table 4-2.

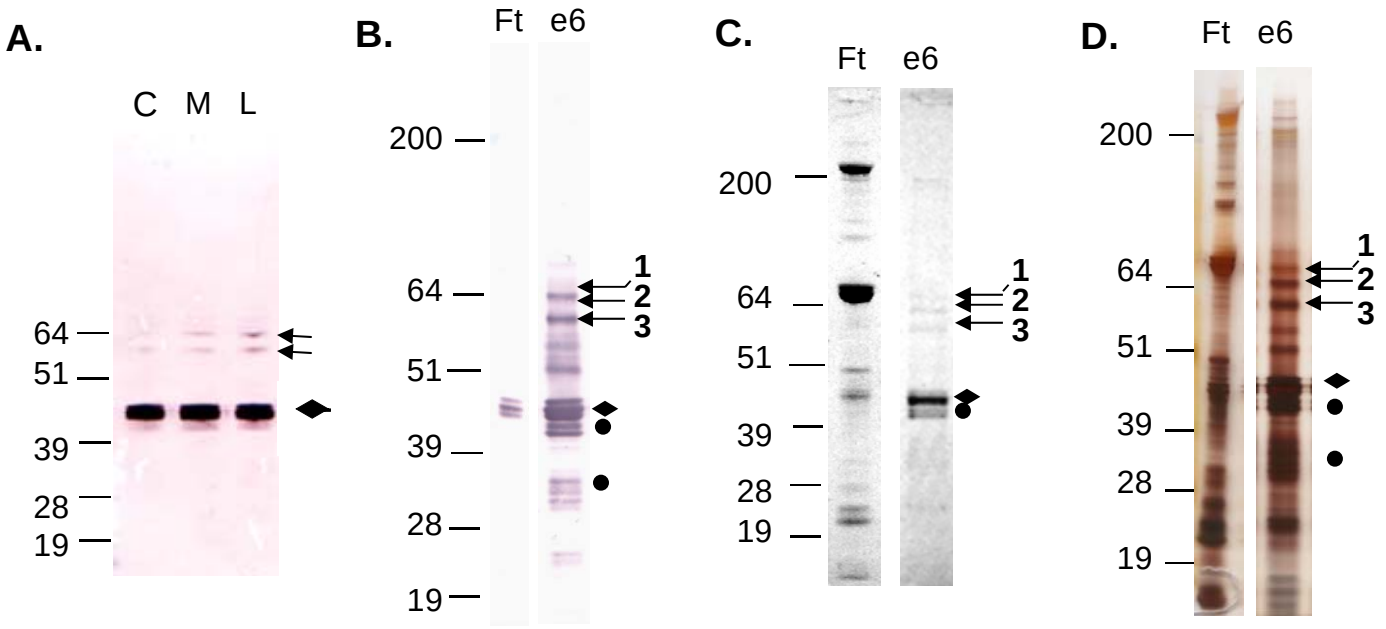
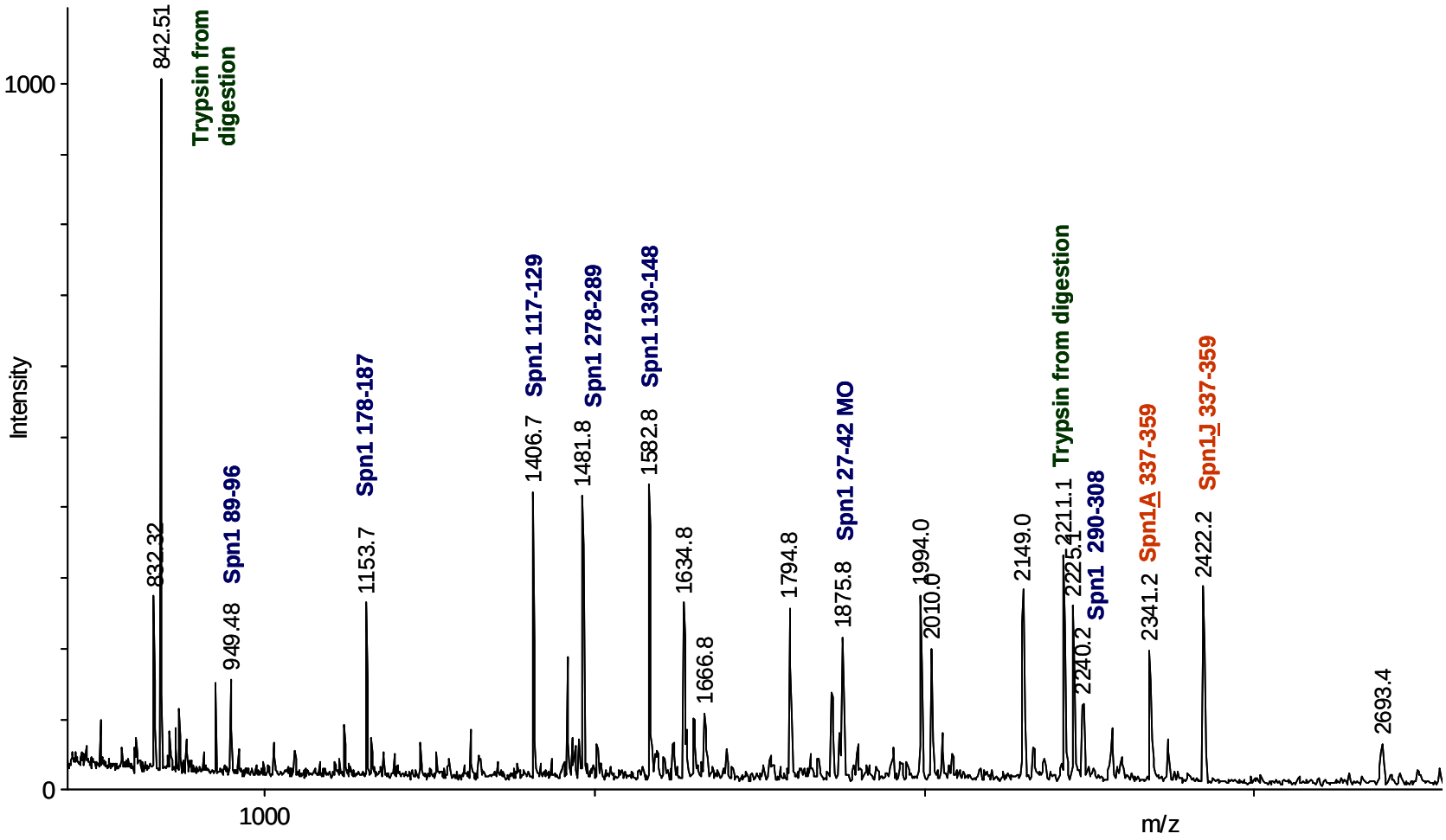


Figure 4-3. Immunoaffinity purification of serpin-1 and two putative serpin-1 protease complexes from *M. sexta* plasma.
A) Immunoblot using serpin-1 antisera to analyze *M. sexta* plasma incubated without elicitor (C, control) or activated in the presence of *M. luteus* (M) or LPS from *E. coli* (L). Arrows indicate putative serpin-1 protease complexes. B, C, and D) Immunoaffinity purification of serpin-1 from LPS incubated plasma. B) Immunoblot using serpin-1 antibody to analyze fractions elute from serpin-1 antibody column. Ft: Flow through; e6: elution fraction 6. Bands 1-3 further analyzed by MALDI-TOF/TOF are marked with arrows. Unmodified serpin-1 is indicated with a diamond. Cleaved serpin-1 is indicated by a circle. C) and D) Same fractions as in part B, detected by Coomassie blue (C) or silver staining (D).

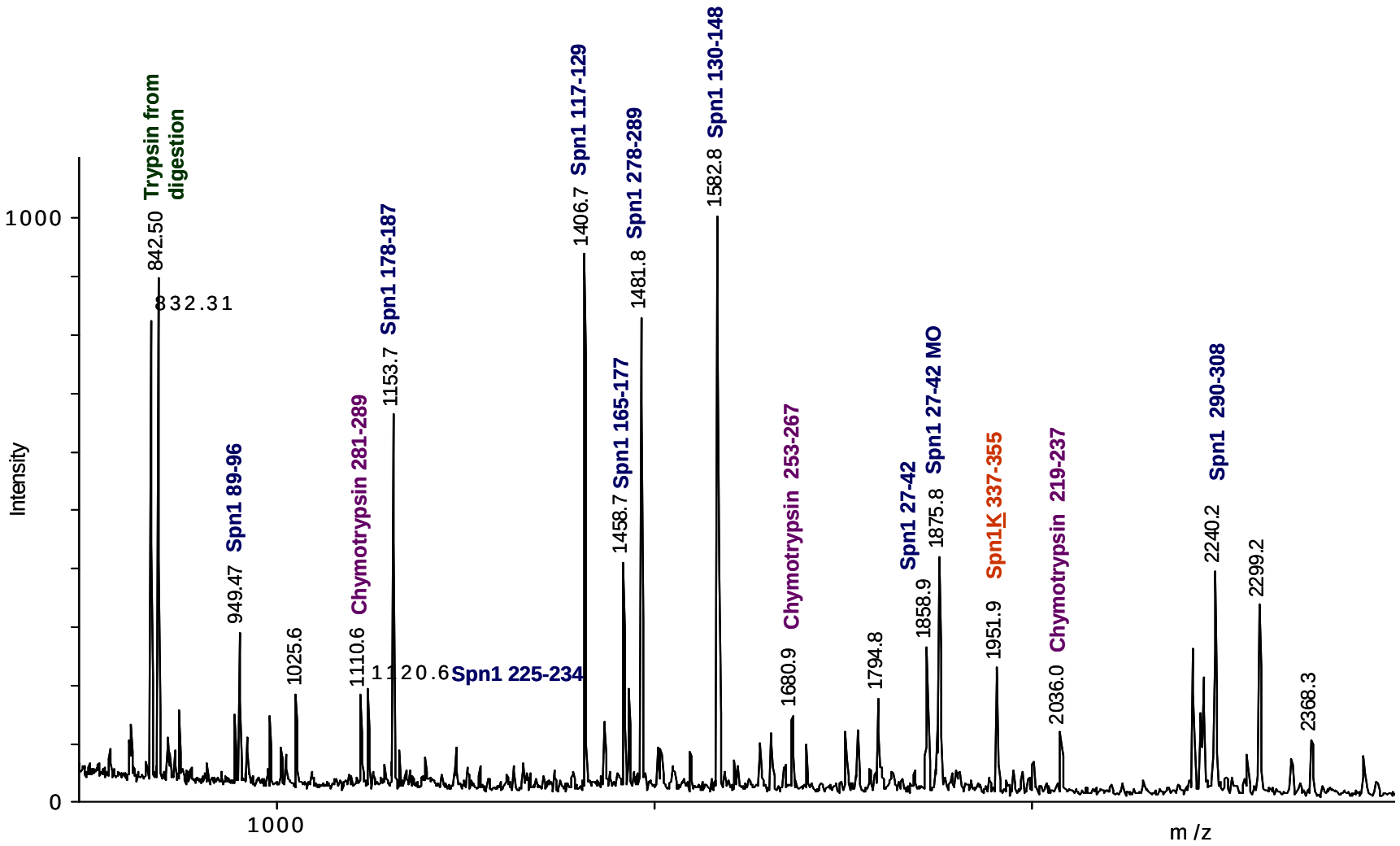
Figure 4-4. MALDI-TOF and TOF/TOF spectra from bands 2 and 3.

MALDI-TOF spectra from bands 2 (A) and 3 (B). Peaks corresponding to exons 1-8 of serpin-1 are labeled in blue; MO stands for oxidized methionine. Isoform-specific peaks are labeled in orange and chymotrypsin specific peaks in purple. MALDI-TOF/TOF spectra of serpin-1J peptide 2422 from band 2 (C) and serpin-1K specific peptide 1951 from band 3 (D).

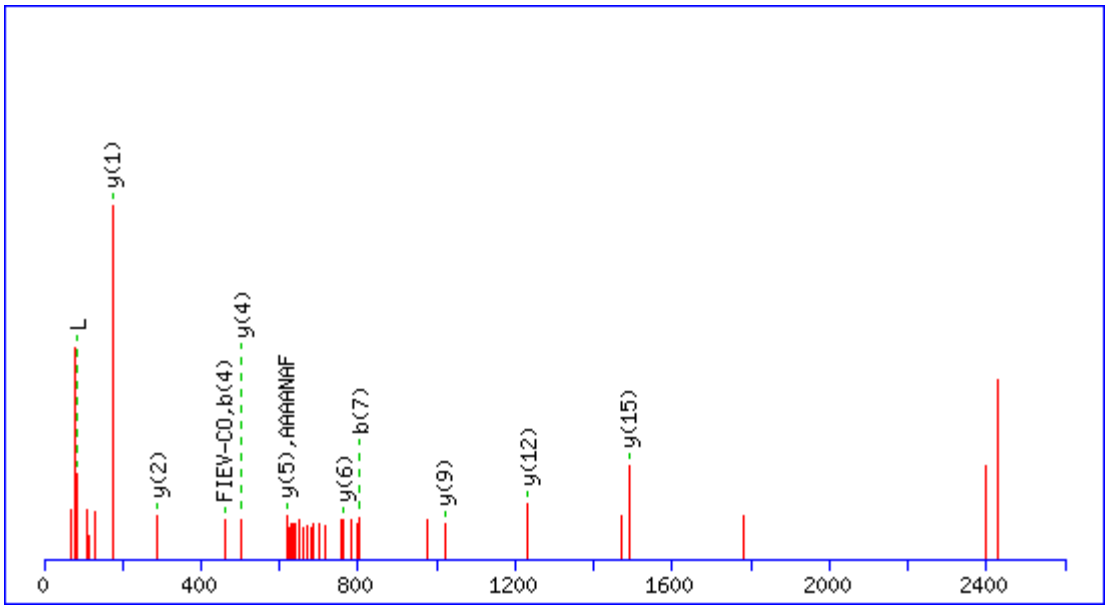
A.


Intensity
1000
0
1000
m/z
842.51
Trypsin from digestion
832.32
949.48
Spn1 89-96
1153.7
Spn1 178-187
1406.7
Spn1 117-129
1481.8
Spn1 278-289
1582.8
Spn1 130-148
1634.8
1666.8
1794.8
1875.8
Spn1 27-42 MO
1994.0
2010.0
2149.0
2211.1
Trypsin from digestion
2225.1
2240.2
Spn1 290-308
2341.2
Spn1A 337-359
2422.2
Spn1J 337-359
2693.4

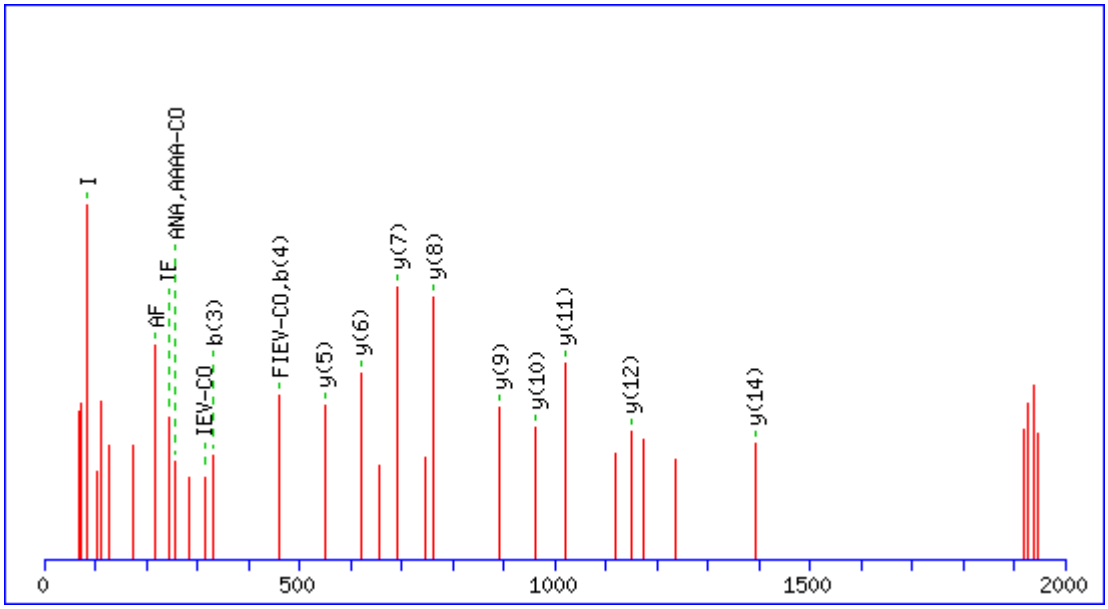
B.


Intensity
1000
0
1000
m/z
842.50
Trypsin from digestion
832.31
949.47
Spn1 89-96
1025.6
1110.6
Chymotrypsin 281-289
1153.7
Spn1 178-187
1120.6
Spn1 225-234
1406.7
Spn1 117-129
1458.7
Spn1 165-177
1481.8
Spn1 278-289
1582.8
Spn1 130-148
1680.9
Chymotrypsin 253-267
1794.8
1858.9
Spn1 27-42
1875.8
Spn1 27-42 MO
1951.9
Spn1K 337-355
2036.0
Chymotrypsin 219-237
2240.2
Spn1 290-308
2299.2
2368.3

C. AFIEVNEEGAEAAAANAFILTDR



D. AFIEVNEEGAEAAAANAFK



A.

```
1   ---------- -------rsf nfeehledit aygyltkygi praeeirrme eeeiaqsriv ggssssvgqf pyqaglvitl
81  prgtaacggs llsnrrVLTA AHCWWDGQNQ ASRFVVVLGS NRLFSGGVRl etrDIVMHGS WNPNLVRndi amirlpsnvg
161 fnnninvial psgsqlnlnf agerAIASGF GRtrdganid gslnhvtldv iannvcsrTF PLLIQSSNIC TSGANGRstc
241 hgdsggplaa trNNRPLLIG VTSFGHRDGC QRGHPAAFAR VTSYDAWIRR n
```

B.

```
1   mkIIMCIFGL AALAMAGETD LQKilrESND QFTAQMFSEV VKANPGQNVV LSAFSVLPPL GQLALASVGE SHDELLRala
81  lpndnvtkDV FADLNRgvra vkgvdlkmas kiyvakGLEL NDDFAAVSRD VFGSEVQNVD FVKsveaaga inkWVEDQTN
161 NRIKNLVDPD ALDETTRSVL VNAIYFKgsw kdkfvkerTM DRDFHVSKdk tikVPTMIGK kdvrYADVPE LDAKMIEMSY
241 EGDQASMIII LPNQVDGITA LEQKlkdpka lsraeerLYN TEVEIYLPKF KIETTTDLKE VLSNMNIKKL FTPGAARLEN
321 LLKTKESLYV DAAIQKAFIE VNEEGAEAAA ANAFKitfys fhfvpkvein kpfffslkyn rnsmfsgvcv q
```

Figure 4-5. Complex band 3 contains serpin-1K and chymotrypsin.

Band 3 peptide mass fingerprinting coverage of chymotrypsin (A) and Serpin-1K (B) from band 3. Tryptic peptides identified by mass spectrometry are shown in blue capital letters. The activation site of chymotrypsin is highlighted in yellow, the serpin-1K P1 residue is in cyan.

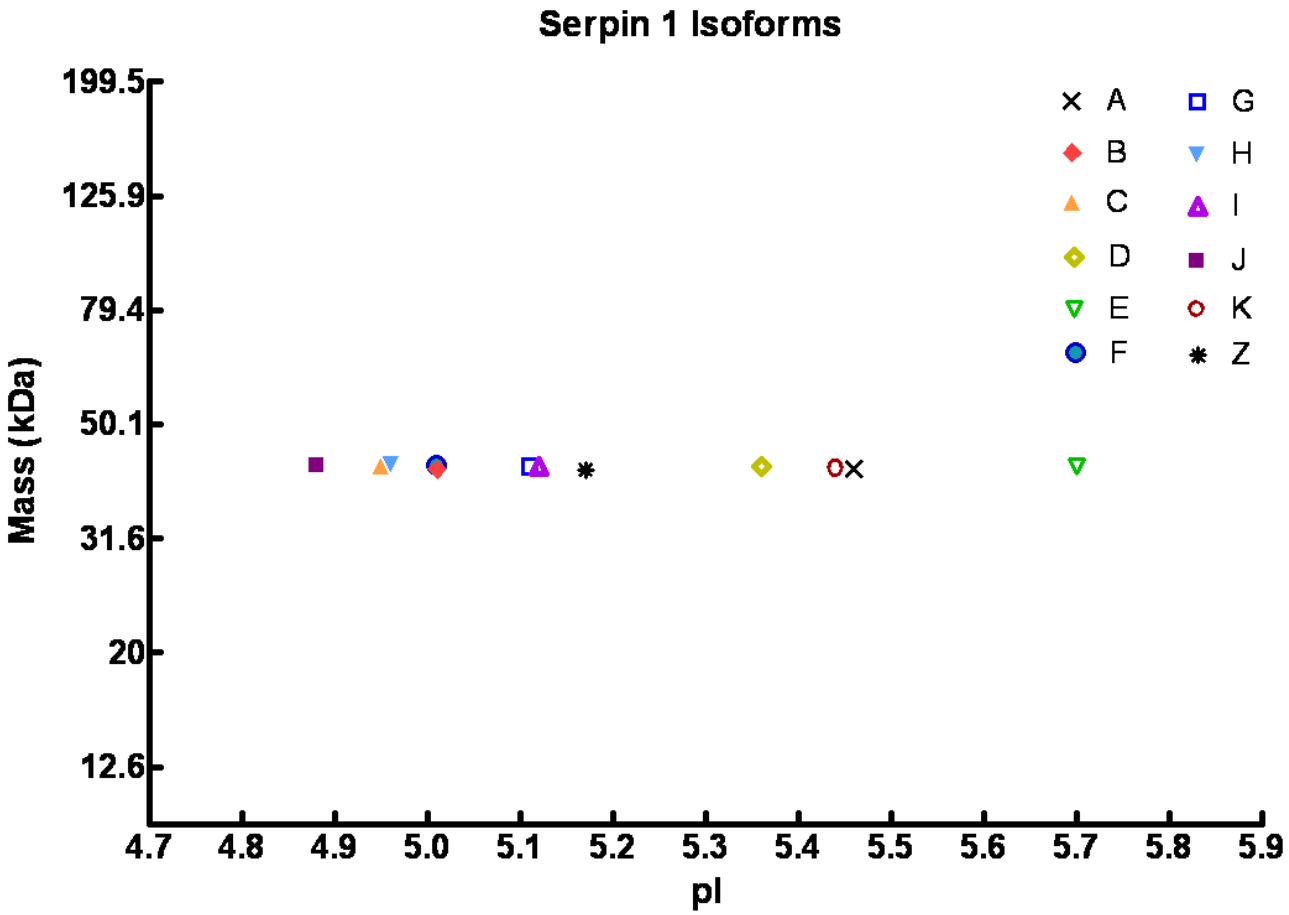


Figure 4-6. Expected serpin-1 isoform migration with separation by 2D-PAGE.

This model assumes isoelectric focusing on a 4.7 – 5.9 pH range strip followed by SDS-PAGE and is based on the calculated isolectric points and masses of each serpin-1 isoform.

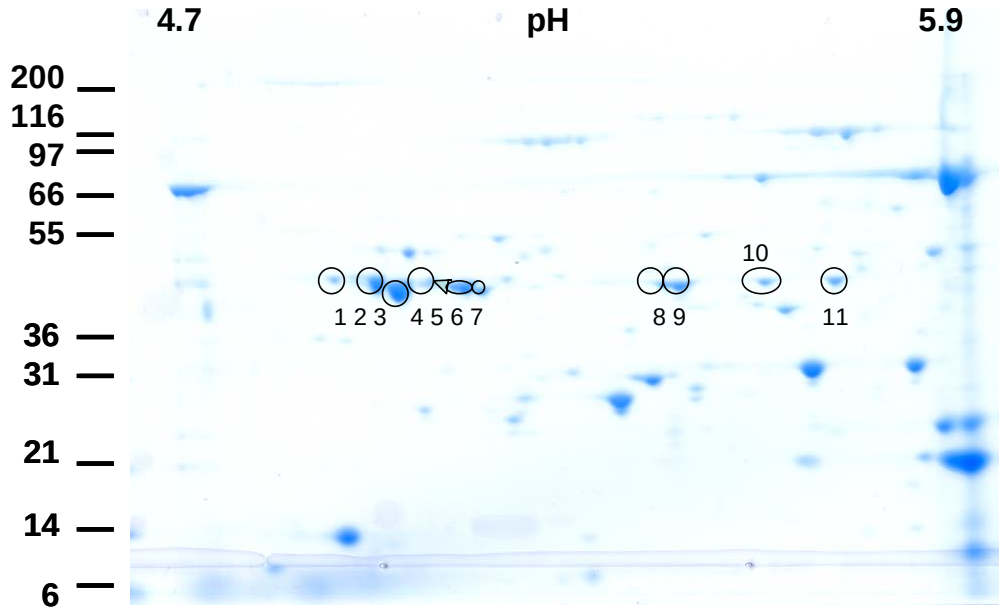


Figure 4-7. 2D gel electrophoretic separation of *M. sexta* larval plasma proteins.

Separation by IEF on a 4.7 -5.9 pH range strip was followed by SDS-PAGE on a 4-12% bis-tris gel. Spots SpER1-11 were cut out, digested with trypsin, and resulting peptides were analyzed by MALDI-TOF and MALDI TOF/TOF.

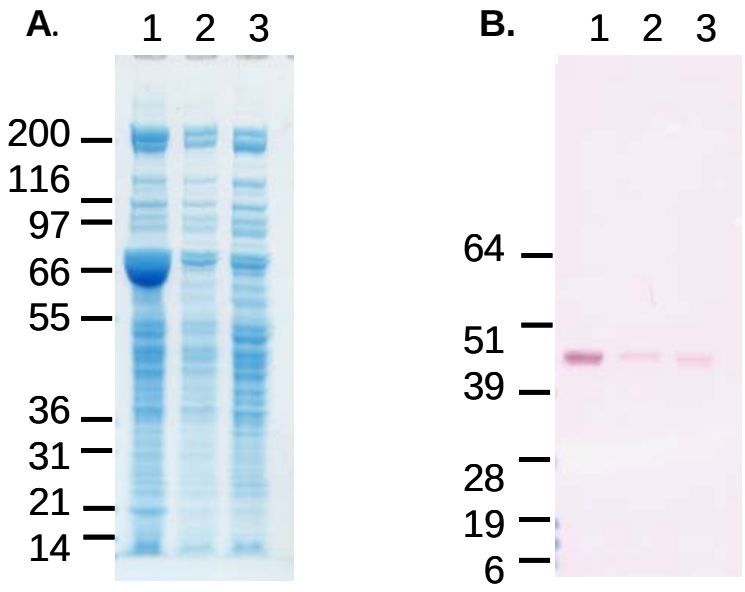


Figure 4-8. SDS-PAGE of proteins secreted by fat body cultured in vitro. Fat body samples from three day 3 5th instar larvae were individually cultured overnight in 1.2 ml SF-900 media with penicillin and streptomycin. The media was collected, separated by SDS-PAGE, and stained with Coomassie blue (A) or immunoblotted with detection by serpin-1 antibody (B). Lane numbers indicate media from culture of fat body from three different individuals.

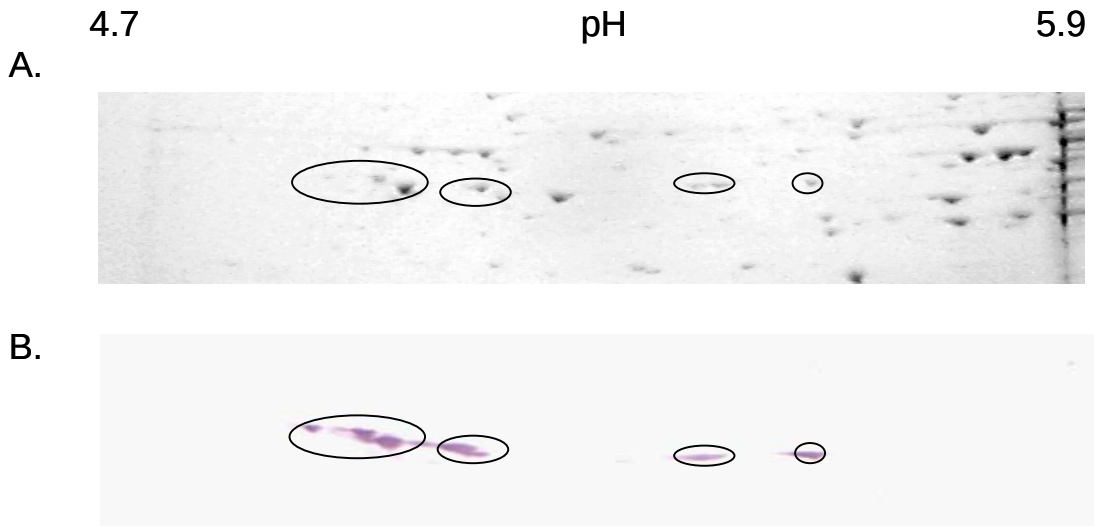


Figure 4-9. 2D gel analysis of proteins secreted by fat body in vitro.
Isoelectric focusing was carried out using strips with pH range of 4.7-5.9. The gel regions shown are from 36 to 55 kDa. A. Proteins visualized using Coomassie blue stain. B. Immunoblotting with detection by serpin-1 antisera. Multiple spots corresponding to serpin-1 isoforms are present, suggesting the secretion of multiple serpin-1 isoforms from fat body. Spots that labeled with serpin-1 antibody are circled on the stained gel.

Figure 4-10. Peptides from trypsin, LysC/AspN and Glu-C phosphate digestion of the isoform specific region of serpin-1.

A. and B. Trypsin digestion C. and D. LysC/AspN digestion E. and F. Glu-C phosphate digestion. Serpin-1 isoform specific peaks from protease digestion (A, C, and E) and locations of protease cleavage sites (B, D, and F).

A.

Isoform	Mono [M+]	Position	Mod.
A	2341.15	337-358	
A	1627.94	366-378	
A	1326.72	379-390	
B	4621.37	337-378	
B	1091.56	379-388	
B	1107.55	379-388	MSO
C	5089.48	337-381	
C	1552.72	382-395	Cys_CAM
D	1959.98	337-355	
D	1557.93	356-369	
D	1563.77	370-381	
D	1638.73	382-395	Cys_CAM
D	1654.73	382-395	MSO&CAM

Isoform	Mono [M+]	Position	Mod.
E	1973.99	337-355	
E	765.52	363-369	
E	1566.82	370-380	
E	1576.74	382-395	Cys_CAM
F	3181.57	337-366	
F	1881.96	367-381	
F	1751.83	382-397	Cys_CAM
G	5140.62	337-382	
G	1477.71	383-396	Cys_CAM
H	5085.48	337-381	
H	1751.88	382-396	Cys_CAM

Isoform	Mono [M+]	Position	Mod.
I	3495.75	337-369	
I	1543.84	370-381	
I	1618.74	382-395	Cys_CAM
J	2422.19	337-359	
J	2826.21	360-381	Cys_CAM
J	1870.94	382-397	Cys_CAM
K	1951.94	337-355	
K	1385.73	356-366	
K	1468.82	367-378	
K	1226.53	382-392	Cys_CAM
K	1242.53	382-392	CAM+MSO
Z	2867.46	337-364	
Z	1792.92	365-379	
Z	1086.55	383-393	

B.

```
Spn1B KAFIEVNEEGAEAAAANAFGIVPASLILYPEVHIDRPFYFELKIDGIPMFNGKVIEP--- 392
Spn1K KAFIEVNEEGAEAAAANAFKITFYSFHFVPKVEINKPFFFSLKYNRNSMFSGVCVQP--- 392
Spn1D KAFIEVNEEGAEAAAAN-VVRGIRPRPSVRPPTPKFEADRPFLFYMKTNDQTLFNGICMQP--- 395
Spn1E KAFIEVNEEGAEAAAAN-VIRVVKKKFRVRPPVLKFHVDRPFFFNLKANDQSLFNGICLQP--- 395
Spn1F KAFIEVNEEGAEAAAAN-AFIAVVDSIDIFERTIEFHADRPFFFNLKANGQSLFNGICVMPML- 397
Spn1H KAFIEVNEEGAEAAAAN--AFITYVESIDNFVPTIEFDVNRPFYFNLKANDLYLFNGICVQPKLQ 398
Spn1G KAFIEVNEEGAEAAAANAFFIVGITSIQFEPPVIEFHVNRPFFFNLKASGQSLFNGICVQP--- 396
Spn1A KAFIEVNEEGAEAAAAN-AFFITRQARLDIRYFVANKPFIFLLRFNGLALFNGVFKA---- 391
Spn1Z KAFIEVNEEGAEAAAAN-AFGIAYLSAVIRSPVFNADHPFVFFLRQDKTTLFSGVFQS---- 392
Spn1J KAFIEVNEEGAEAAAANAFILTDRCCSDYDDNIEFDVNRPFYLNLRTNEHLLFSGICIQPEI- 397
Spn1C KAFIEVNEEGAEAAAAN-AFFIIESYSSYEPVVPVFDIDKPFYFNIRANGQSLFNGLCFQP--- 395
Spn1I KAFIEVNEEGAEAAAANEFGIVALSLEFSLNEIKFVVNKPFYFNIRSNGQHLFNGICFQP--- 395
      ******************
```

C.

Isoform	Mono [M+]	Position	Mod.
A	2809.43	337-362	
A	1125.60	363-371	
A	2213.26	372-390	
B	3425.76	337-370	
B	1214.62	371-378	
B	978.47	380-388	
C	3680.77	337-363	
C	2490.20	376-395	Cys_CAM
D	3498.89	337-369	
D	1216.62	373-381	
D	1423.63	384-395	Cys_CAM
E	2300.22	337-358	
E	1068.69	361-369	
E	1183.63	373-381	
E	1391.66	384-395	Cys_CAM
F	2206.10	337-358	
F	1377.68	362-372	
F	1183.63	373-381	
F	1751.83	382-397	Cys_CAM
G	5140.62	337-382	
G	1477.71	383-396	Cys_CAM
H	2629.27	337-361	
H	1081.52	362-370	
H	1412.73	371-381	
H	1566.80	384-396	Cys_CAM
I	3495.75	337-369	
I	606.36	370-374	
I	2556.22	375-395	Cys_CAM
J	2151.06	337-357	
J	697.24	356-360	Cys_CAM
J	637.28	364-368	
J	3258.67	369-397	Cys_CAM
K	1951.94	337-355	
K	1385.73	356-366	
K	885.49	372-378	
K	1658.74	379-392	Cys_CAM
Z	3482.76	337-370	
Z	1305.67	371-380	
Z	1086.55	383-392	

D.

```
Spn1B KAFIEVNEEGAEAAAANAFGIVPASLILYPEVHIDRPFYFELKIDGIPMFNGKVIEP--- 392
Spn1K KAFIEVNEEGAEAAAANAFKITFYSFHFVPKVEINKPFFFSLKYNRNSMFSGVCVQP--- 392
Spn1D KAFIEVNEEGAEAAAAN-VVRGIRPRPSVRPPTPKFEADRPFLFYMKTNDQTLFNGICMQP--- 395
Spn1E KAFIEVNEEGAEAAAAN-VIRVVKKKFRVRPPVLKFHVDRPFFFNLKANDQSLFNGICLQP--- 395
Spn1F KAFIEVNEEGAEAAAAN-AFIAVVDSIDIFERTIEFHADRPFFFNLKANGQSLFNGICVMPML- 397
Spn1H KAFIEVNEEGAEAAAAN--AFITYVESIDNFVPTIEFDVNRPFYFNLKANDLYLFNGICVQPKLQ 398
Spn1G KAFIEVNEEGAEAAAANAFFIVGITSIQFEPPVIEFHVNRPFFFNLKASGQSLFNGICVQP--- 396
Spn1A KAFIEVNEEGAEAAAAN-AFFITRQARLDIRYFVANKPFIFLLRFNGLALFNGVFKA---- 391
Spn1Z KAFIEVNEEGAEAAAAN-AFGIAYLSAVIRSPVFNADHPFVFFLRQDKTTLFSGVFQS---- 392
Spn1J KAFIEVNEEGAEAAAANAFILTDRCCSDYDDNIEFDVNRPFYLNLRTNEHLLFSGICIQPEI- 397
Spn1C KAFIEVNEEGAEAAAAN-AFFIIESYSSYEPVVPVFDIDKPFYFNIRANGQSLFNGLCFQP--- 395
Spn1I KAFIEVNEEGAEAAAANEFGIVALSLEFSLNEIKFVVNKPFYFNIRSNGQHLFNGICFQP--- 395
      ****************
```

E.

Isoform	Mono [M+]	Position	Mod.
A	1735.92	348-363	
A	3275.86	364-391	
B	1888.02	348-366	
B	858.41	371-376	
B	1301.69	381-392	
C	1137.59	348-358	
C	1488.69	359-371	
C	2618.30	374-395	Cys_CAM
D	2586.47	348-371	
D	1431.71	374-384	
D	1308.61	385-395	Cys_CAM

Isoform	Mono [M+]	Position	Mod.
E	2915.78	348-373	
E	1368.71	374-384	
E	1276.63	385-395	Cys_CAM
F	1132.60	348-359	
F	2801.41	374-397	Cys_CAM
G	2405.28	348-370	
G	3024.53	371-396	Cys_CAM
H	1240.62	348-359	
H	819.42	363-369	
H	1597.81	372-384	
H	1692.91	385-398	Cys_CAM

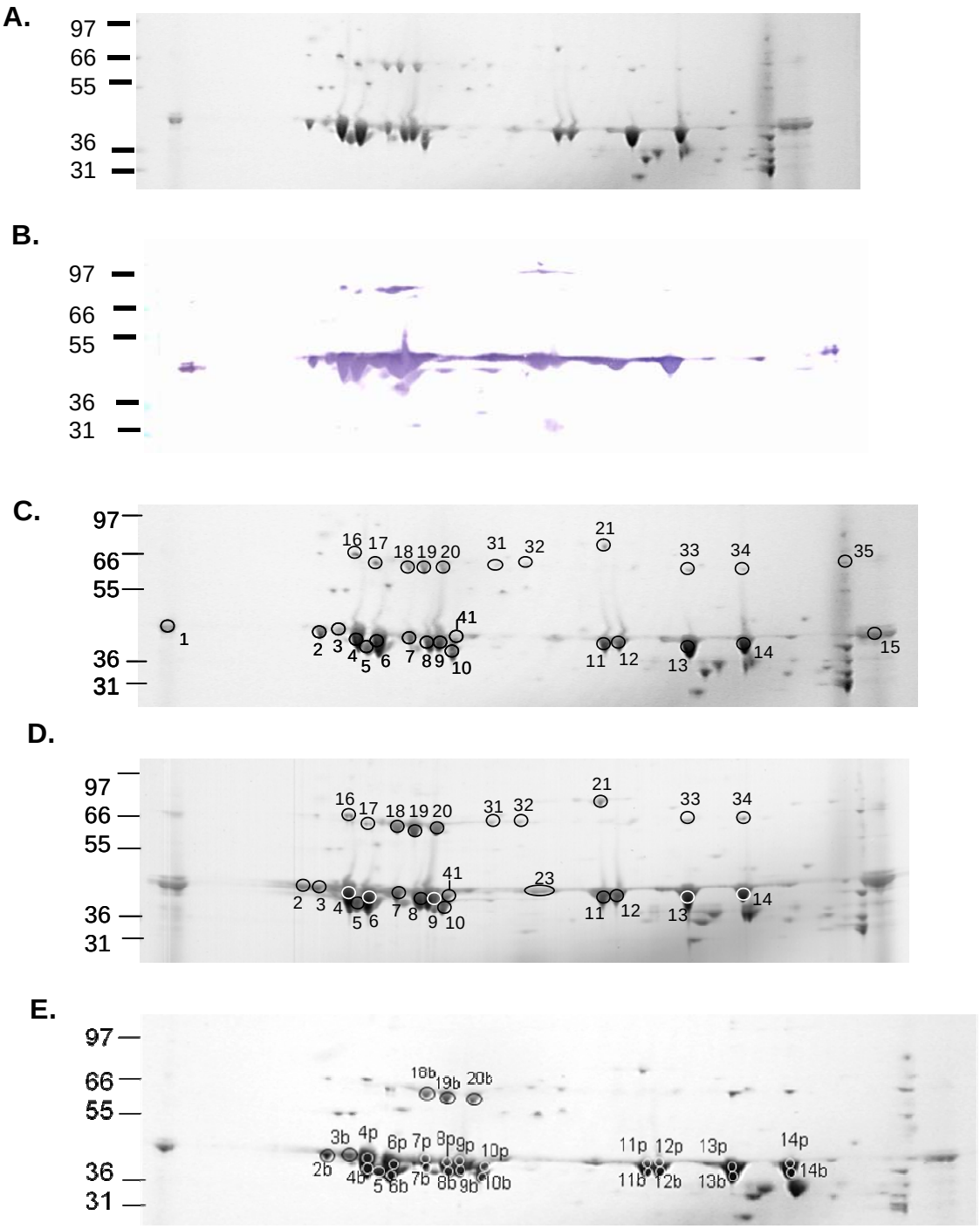
Isoform	Mono [M+]	Position	Mod.
I	948.54	354-362	
I	3384.74	368-395	Cys_CAM
J	1077.56	348-358	
J	1635.86	372-384	
J	1413.72	385-396	Cys_CAM
K	2358.23	348-368	
K	2880.43	369-392	Cys_CAM
Z	2409.26	348-371	
Z	1305.67	372-381	
Z	1214.64	382-392	

F.

```
Spn1B KAFIEVNEEGAEAAAANAFGIVPASLILYPEVHIDRPFYFELKIDGIPMFNGKVIEP--- 392
Spn1K KAFIEVNEEGAEAAAANAFKITFYSFHFVPKVEINKPFFFSLKYNRNSMFSGVCVQP--- 392
Spn1D KAFIEVNEEGAEAAAAN-VVRGIRPRPSVRPPTPKFEADRPFLFYMKTNDQTLFNGICMQP--- 395
Spn1E KAFIEVNEEGAEAAAAN-VIRVVKKKFRVIPPVLKFHVDRPFFFNLKANDQSLFNGICLQP--- 395
Spn1F KAFIEVNEEGAEAAAAN-AFIAVVDSIDIFERTIEFHADRPFFFNLKANGQSLFNGICVMPML- 397
Spn1H KAFIEVNEEGAEAAAAN--AFITYVESIDNFVPTIEFDVNRPFYFNLKANDLYLFNGICVQPKLQ 398
Spn1G KAFIEVNEEGAEAAAANAFFIVGITSIQFEPPVIEFHVNRPFFFNLKASGQSLFNGICVQP--- 396
Spn1A KAFIEVNEEGAEAAAAN-AFFITRQARLDIRYFVANKPFIFLLRFNGLALFNGVFKA---- 391
Spn1Z KAFIEVNEEGAEAAAAN-AFGIAYLSAVIRSPVFNADHPFVFFLRQDKTTLFSGVFQS---- 392
Spn1J KAFIEVNEEGAEAAAANAFILTDRCCSDYDDNIEFDVNRPFYLNLRTNEHLLFSGICIQPEI- 397
Spn1C KAFIEVNEEGAEAAAAN-AFFIIESYSSYEPVVPVFDIDKPFYFNIRANGQSLFNGLCFQP--- 395
Spn1I KAFIEVNEEGAEAAAANEFGIVALSLEFSLNEIKFVVNKPFYFNIRSNGQHLFNGICFQP--- 395
      ******************
```

Figure 4-11. Immunoaffinity-purified serpin-1 separated by isoelectric focusing and SDS-PAGE.

A. Coomassie stained gel. B. Immunoblot detected using serpin-1 antibody. C.-E. Spots from Coomassie stained gels picked for protease digestion followed by MALDI-TOF and MALDI-TOF/TOF mass spectrometry analysis. C. Trypsin digestion. D. LysC/AspN double digestion. E. Region of spot selected for GluC digestion in either bicarbonate buffer (spot number followed by b) or phosphate buffer (spot number followed by p).


A.
97
66
55
36
31
B.
97
66
55
36
31
C.
97
66
55
36
31
16 17 18 19 20 31 32 21 33 34 35
1 2 3 4 5 6 7 8 9 10 41 11 12 13 14 15
D.
97
66
55
36
31
16 17 18 19 20 31 32 21 33 34
2 3 4 5 6 7 8 9 10 41 23 11 12 13 14
E.
97
66
55
36
31
16b 19b 20b
2b 3b 4p 4b 5 6b 6p 7b 7p 8b 8p 9b 9p 10b 10p
11b 11p 12b 12p 13b 13p 14b 14p

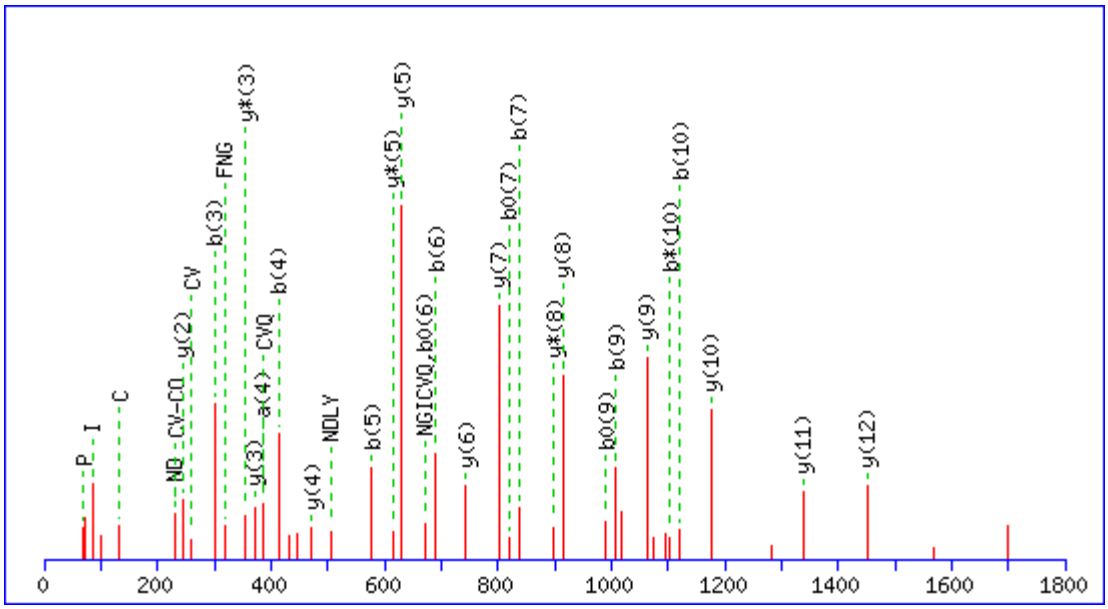


Figure 4-12. MS/MS spectra for trypsin spot 4, peptide 1752.

Identified as serpin-1H peptide ANDLYLFNGICVQPK.

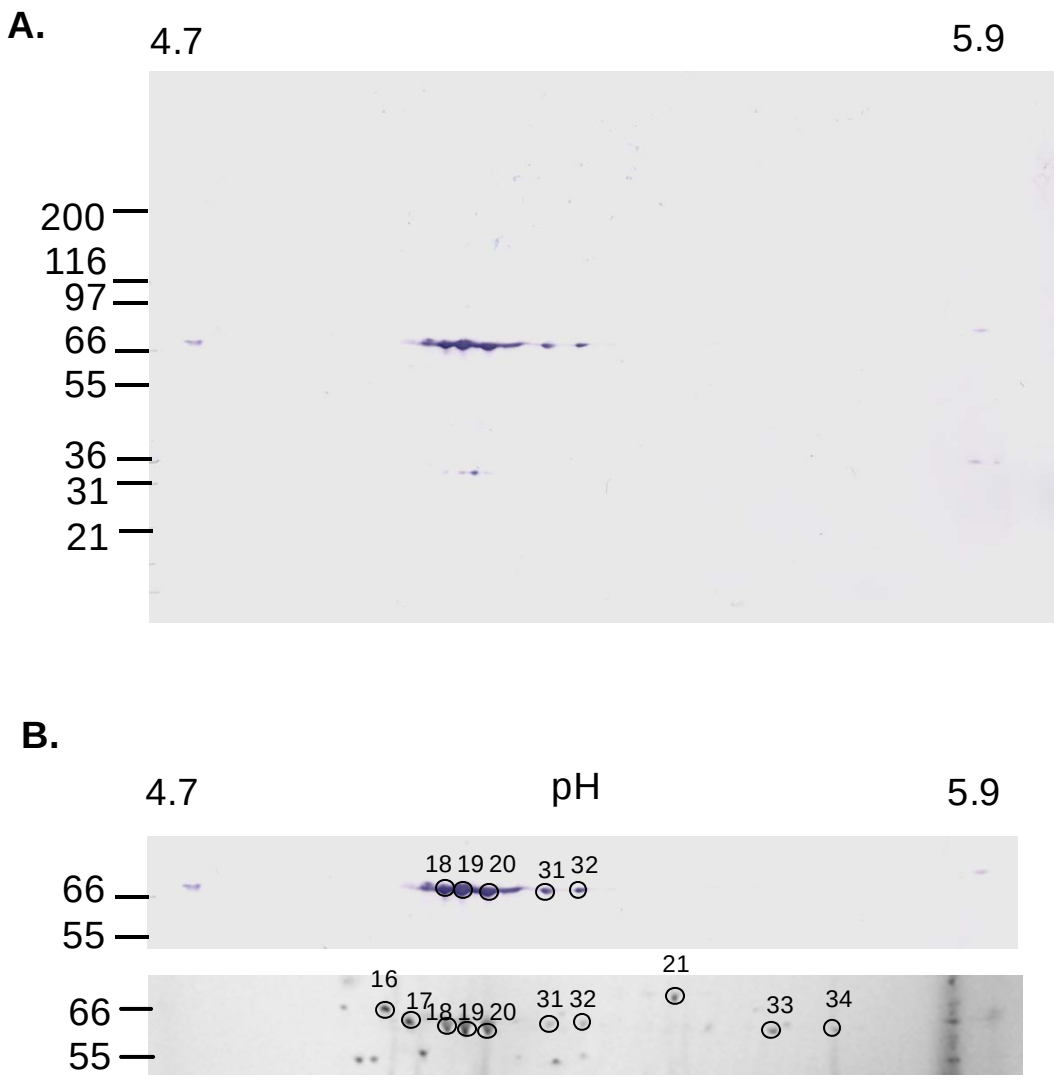


Figure 4-13. Identification of HP8 in serpin-1 complexes.

A. Immunoblot of 2D gel of proteins isolated by immunoaffinity to serpin-1, detected with HP8 antibody.

B. Comparison of HP8 immunoblot and Coomassie stained gel. Spots 18, 19, 20, 31, and 32 are detected by the HP8 antibody.

Figure 4-14. Identification of HP1 in serpin-1 complexes.

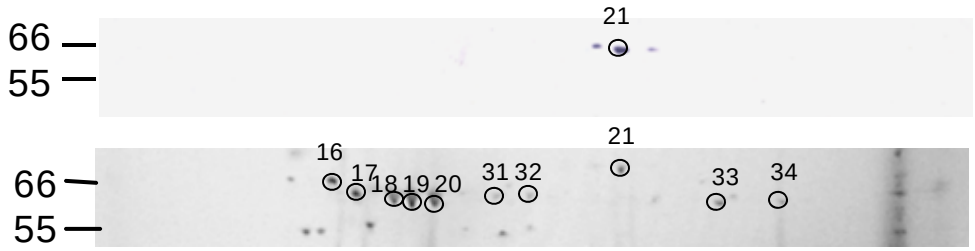
A. Immunoblot of 2D gel of proteins isolated by immunoaffinity to serpin-1, detected with HP1 antibody.

B. Comparison of HP1 immunoblot and Coomassie stained gel. Spot 21 is detected by the HP1 antibody.

C. HP1 sequence coverage of spot 21 after trypsin digestion, MALDI-TOF, and PMF using Aldente (capital blue letters) or PMF using Mascot search of predicted activated protease sequences (underlining). Two peptides identified by Aldente, GCGLSTR and GCGLSTRAQGR, are N-terminal to the putative HP1 activation site at Arg154 (highlighted in green) and therefore were not present in the activated protease sequences searched by Mascot.

A.

B.



C.

```
mifvfllllv svcsekgdys elqhpawedi ieeggmhigg grskrfiqln pnqgnvayqa  60
ctlpsgkqgh crhlrfciqe dfkqdfvkfm nyvcviqgqs mgvccpedtt vggpeglagd 120
lpatapqege detllkinqa qnrGCGLSTR AQGRVFGSRP ANPRewpwma sitpegfeqy 180
cggvlitdrh vltaahctrR WEANELYVRL GEYDFKRtnd trSYNFRVVE Kvqhvdfeis 240
nyhhdiailk ldkpaifnty vwpiclpppg lsienetvtv igwgtqwygg phshvlmevs 300
fpiwthqnci evhtnsifde sicagghegg rdacqgdsgg plmyqmpsgr wavvgivswg 360
vrCGEPNHPG IYTRvdkyig wimenar                                     387
```

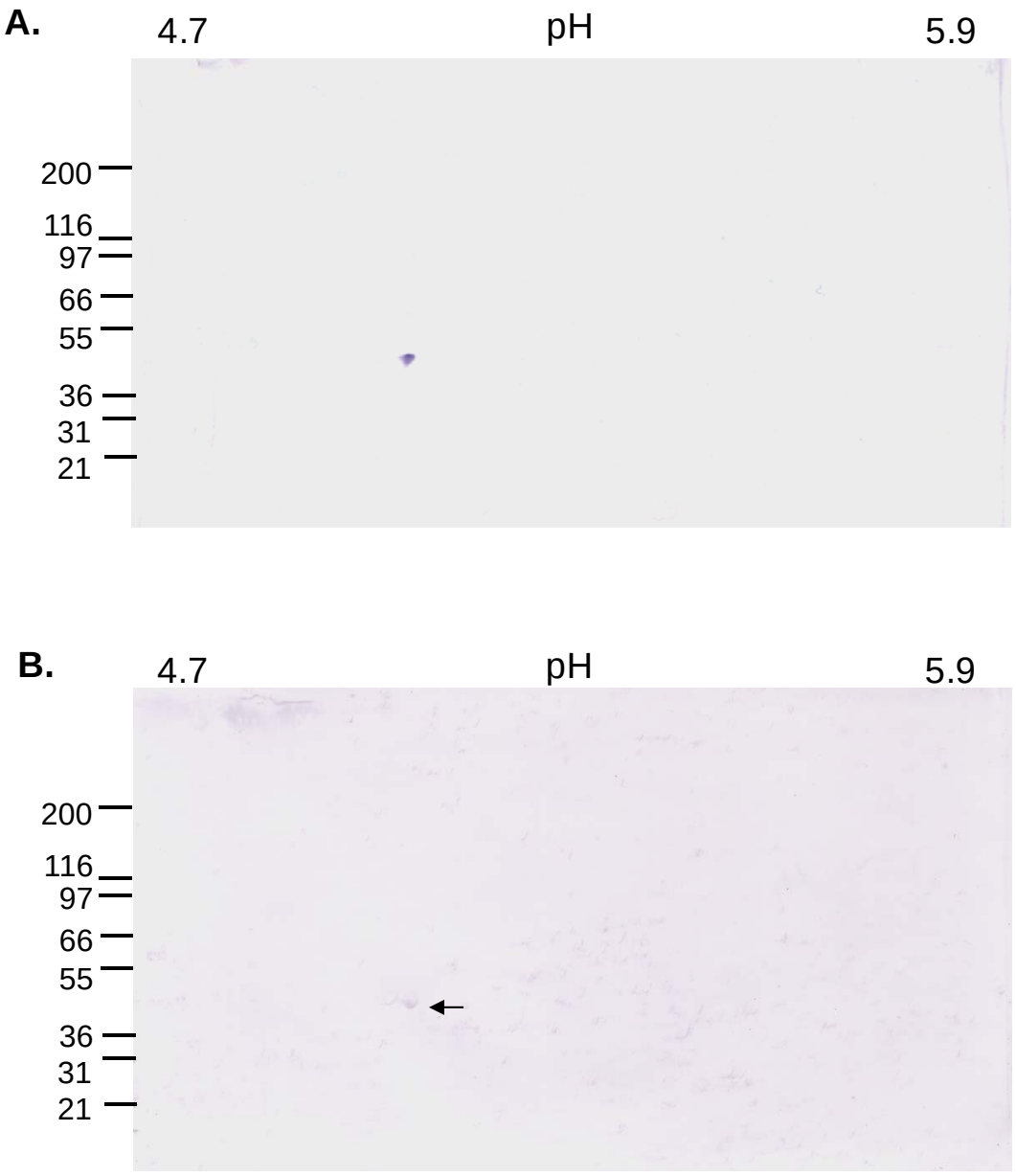


Figure 4-15. Immunoblot identification of serpin-1B and serpin-1F in spot 6. Immunoblot of 2D gel of proteins isolated by immunoaffinity to serpin-1, detected with antibody raised against Spn-1F peptide (A) or raised against Spn-1B peptide (B). The spot visible in B is fainter and marked with an arrow.

CHAPTER 5 - Conclusions and future perspectives

Three research projects were described in this dissertation. Chapter 2 discussed proteins present in high molecular weight complexes after gel filtration of *M. sexta* plasma. Chapter 3 covered *M. sexta* hemolymph proteins that bound to curdlan and different bacteria. Finally, in Chapter 4, individual serpin-1 isoforms were studied at the mRNA and protein levels. Here I present a short summary of each section and suggest ideas for future research in each of these areas.

Gel filtration

Gel filtration of *M. sexta* plasma after incubation with peptidoglycan, LPS, or no elicitor led to detection of a protein peak in the void volume of the gel filtration column that we referred to as the "high molecular weight" (HMW) fractions. SDS-PAGE of the HMW fractions, followed by MALDI-TOF/TOF and immunblotting, led to the identification of many proteins present in these fractions. Hemocytin was detected HMW fractions from in all three plasma samples. While we have a sequence for both *M. sexta* and *B. mori* hemocytins, they both may be incomplete at their N-termini (Lesch et al, 2007). Five-prime RACE could be used to get a complete cDNA sequence for *M. sexta* hemocytin, and we could then look for a propeptide cleavage site similar the ones found in *D. melanogaster* hemolectin and human Von Willebrand factor (Goto et al, 2001). It also would be interesting to study the disulfide bonding that may occur between different hemocytin proteins. While Von Willebrand factor is known to form large multimers, *D. melanogaster* hemocytin has been only shown to form monomers and dimers (Goto et al, 2001). Running hemocytin samples on low percentage poly-acrylamide or agarose gels under reducing and non-reducing conditions would allow us to look for dimer and higher multimer formation.

High molecular weight fractions from plasma samples activated by peptidoglycan or LPS contained cleaved serine protease homologs and significant levels of proPO. The control plasma contained full length serine protease homologs and much lower levels of proPO. This leads us to hypothesize that cleaved serine-protease homologs are important

for the co-presence of proPO. Further investigation could examine if SPH cleavage impacts its binding to proPO. Previous studies have found that proPO binds the protease-like domains of SPH-1 and SPH-2 (Yu et al, 2003). Partially cleaved SPHs migrated farther into native polyacrylamide gels than fully cleaved SPHs purified from hemolymph, suggesting that full cleavage of SPHs or interaction with plasma factors may be important for formation of high molecular weight SPH complexes (Lu and Jiang, 2008). We found full-length SPHs and hemocytin together in the high molecular weight complexes isolated from plasma. It is possible that interactions with hemocytin help the formation of SPH high molecular weight complexes.

Figure 5-1 shows a model of how proteins in the high molecular weight fraction may interact with each other. The middle of Fig. 5-1 shows plasma prior to any immune challenge or wounding stress. Some of these proteins, such as proPO and possibly hemocytin, may be stored in oenocytoids or may already be present in plasma. Changes after wounding and mixing in the absence of microbial elicitor are shown at the top of the model. Wounding and mixing of plasma components may induce a conformation change of in hemocytin, leading to a change from a globular protein to a more extended confirmation that might form dimers or higher multimers (Siedlecki et al, 1996). Also present in high molecular weight complexes after shaking are uncleaved proSPHs, which may self associate. ProSPHs might also interact with hemocytin directly, as shown in the model.

Wounding and mixing in the presence of an immune elicitor such as peptidoglycan led to the cleavage of serine protease homologs and the presence of proPO in the HMW fractions. A model of what these complexes might look like is shown at the bottom of Fig. 5-1. Cleavage of SPHs is likely performed by a currently unidentified serine protease, which is itself presumably activated in response to recognition of an immune challenge, similar to the serine protease cascade in during activation of proPAP2 and proPAP3. The significant presence of proPO only after incubation with elicitor leads us to speculate that proPO binds much more strongly with cleaved SPHs than the uncleaved form.

The importance of wounding and mixing of plasma for the formation of high molecular weight complexes and the possibility of a direct interaction between hemocytin

and proSPHs need to be further investigated. Future gel filtration experiments can enhance our understanding of the role that gentle shaking of hemolymph plays in the formation of these HMW complexes. Mixing plasma on the shaker may mimic on a larger scale what happens at a wound site where hemolymph is flowing out of the insect. This is the location where clotting needs to occur and changes in hemocytin protein structure may be important for such clot formation. We are also interested in looking at the ultra structure of high molecular weight complexes using transmission electron microscopy.

Curdlan and bacteria binding assays

The *M. sexta* plasma proteins we detected bound to curdlan support the idea of the formation of complexes, including βGRP, HP14, HP21, PAP3, SPHs, and PO at the surface of a pathogen. βGRP, SPHs, and PO were detected by MS and immunoblotting and are likely present at higher levels than the proteases HP21 and PAP3, which were only detected by immunoblot. Immulectin-2, SPH-1, and proPO bound to mannose-agarose. A model which integrates the results from curdlan and mannose binding on a hypothetical fungal surface is presented in Fig. 5-2.

M. sexta hemolymph proteins that bound to curdlan, *S. aureus*, *B. thuringiensis*, *S. marcesens*, and *E. coli* are annexin IX, HAIP, leureptin, lysozyme, and proPO. Annexin IX, HAIP and leureptin could have a role in recognizing foreign, dangerous surfaces while lysozyme and proPO have antimicrobial properties. *M. sexta* PGRP-1 did not bind the gram-positive *S. aureus*, which has Lys-type peptidoglycan, but did bind *B. thuringiensis* and *S. marcesens*, which have DAP-type peptidoglycan. Thus, we can expect PGRP-1 to be a sensor for Gram-negative and *Bacillus* species bacteria but not other Gram-positive bacteria.

Three samples, *S. aureus*, *B. thuringiensis*, and curdlan, bound CP8. Its role at the surface of bacteria requires additional study. The *S. marcesens* sample bound three different *M. sexta* histones, which have been found to have antimicrobial activity in shrimp and in vertebrates. Histones are also present in neutrophil extracellular traps (NETs), which are known to be formed in a wide range of vertebrates. Future research could investigate whether extrusion of DNA and histones from hemocytes is involved in

encapsulation or other cellular responses to bacteria in *M. sexta* that could have similarity with NET formation in vertebrates.

Serpin-1

All twelve serpin-1 isoforms are expressed in hemocytes, fat body, and midgut. *In situ* hybridization could be done to investigate which cells in the midgut express serpin-1. Serpin-1 could be secreted into the lumen of the midgut or into plasma. To learn know more about the direction of serpin-1 secretion from midgut cells we could culture tied-off midgut and probe the presence of serpin-1 by immunoblot using the culture media and the interior of the midgut as samples.

Our preliminary results also indicate that serpin-1A expression is increased 24 hours after bacterial injection. We can further investigate this finding with either real time PCR or Northern blotting with isoform specific primers.

The mechanism of mutually exclusive alternative splicing in M. sexta has not been studied. Future research could investigate regulation of serpin-1 alternative splicing. One of the first 38 serpin-1 cDNA clones had exon 9B and 9C together (Jiang et al, 1994). While no other cDNAs with multiple copies have been found, we haven't investigated how often this occurs.

Serpin-1A, -E, and –J were found in covalent complexes with HP8 and an unidentified serpin-1 isoform complexed with HP1. Future studies can compare the kinetics of these serpin-protease complexes with the kinetics for serpin-4 and -5 with HP1 and serpin-6 with HP8. We can also further investigate the role of serpin-1K in inhibiting chymotrypsin. How often do gut proteases escape into the hemolymph and how important is this function?

References

Goto A, Kumagai T, Kumagai C, Hirose J, Narita H, Mori H, Kadowaki T, Beck K, & Kitagawa Y (2001) A *Drosophila* haemocyte-specific protein, hemolectin, similar to human von Willebrand factor. *Biochem J* **359**: 99-108

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

Lesch C, Goto A, Lindgren M, Bidla G, Dushay MS, & Theopold U (2007) A role for hemolectin in coagulation and immunity in *Drosophila melanogaster*. *Dev Comp Immunol*

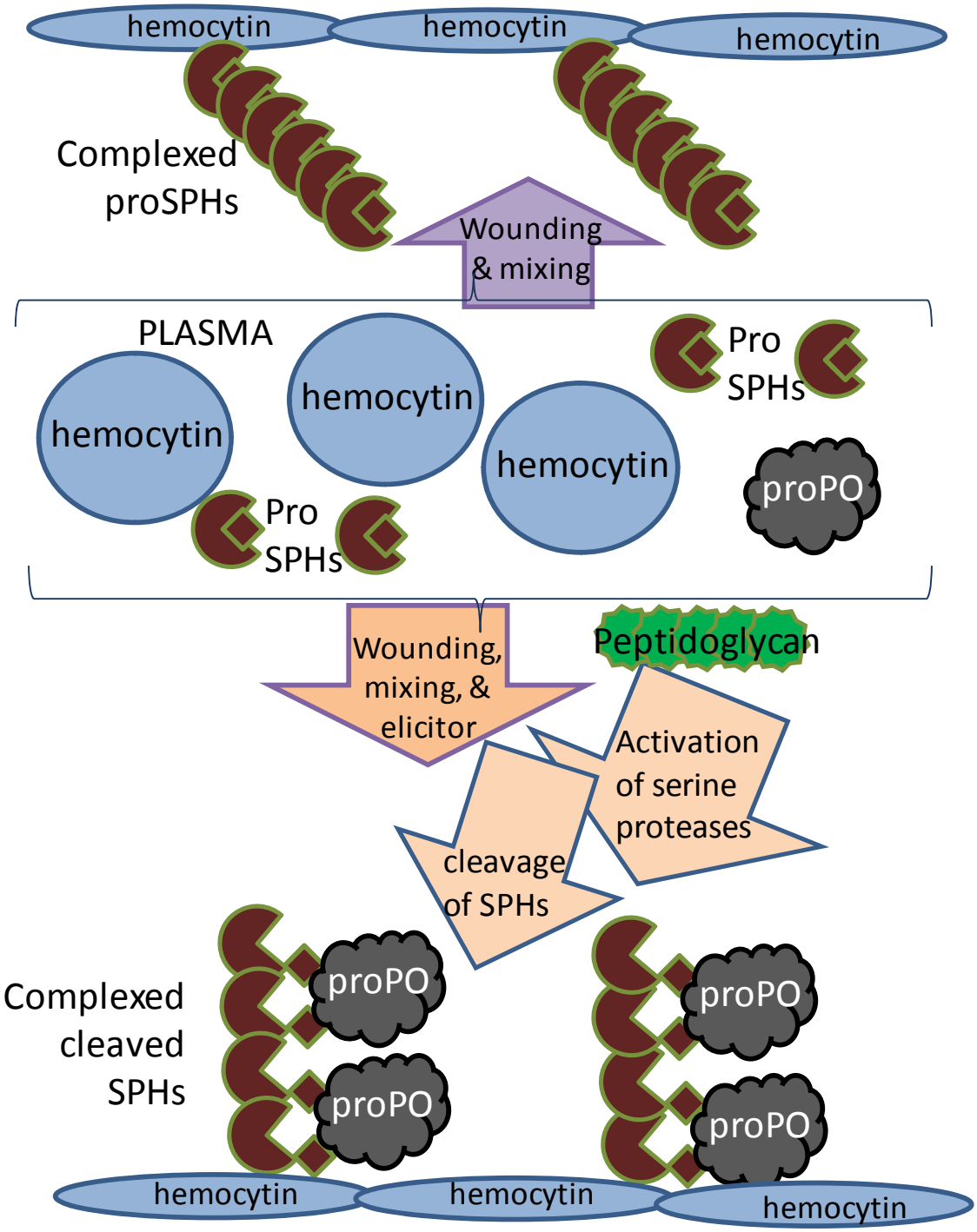
Lu Z & Jiang H (2008) Expression of *Manduca sexta* serine proteinase homolog precursors in insect cells and their proteolytic activation. *Insect Biochem Mol Biol* **38**: 89-98

Siedlecki CA, Lestini BJ, Kottke-Marchent K, Eppell SJ, Wilson DL, & Marchant RE (1996) Sheer-dependant changes in the three-dimensional structure of human von Willebrand factor. *Blood* **88**: 2939-2950.

Yu XQ, Jiang H, Wang Y, & Kanost MR (2003) Nonproteolytic serine proteinase homologs are involved in prophenoloxidase activation in the tobacco hornworm, *Manduca sexta*. *Insect Biochem Mol Biol* **33**: 197-208

Figures

Figure 5-1. A model of high molecular weight complexes formed in *M. sexta* plasma. Hemocytin, uncleaved serine protease homologs (proSPHs), and prophenoloxidase are present in naïve *M. sexta* plasma (middle of figure). After shaking, high molecular weight complexes containing hemocytin and uncleaved are formed (top of figure). If shaking occurs in the presence of a microbial polysaccharide, such as peptidoglycan, the proSPHs are cleaved to SPHs and proPO is also detected (bottom of figure).


hemocytin
hemocytin
hemocytin
Complexed
proSPHs
Wounding
& mixing
PLASMA
hemocytin
hemocytin
hemocytin
Pro
SPHs
Pro
SPHs
proPO
Wounding,
mixing, &
elicitor
Peptidoglycan
Activation
of serine
proteases
cleavage
of SPHs
Complexed
cleaved
SPHs
proPO
proPO
proPO
proPO
hemocytin
hemocytin
hemocytin

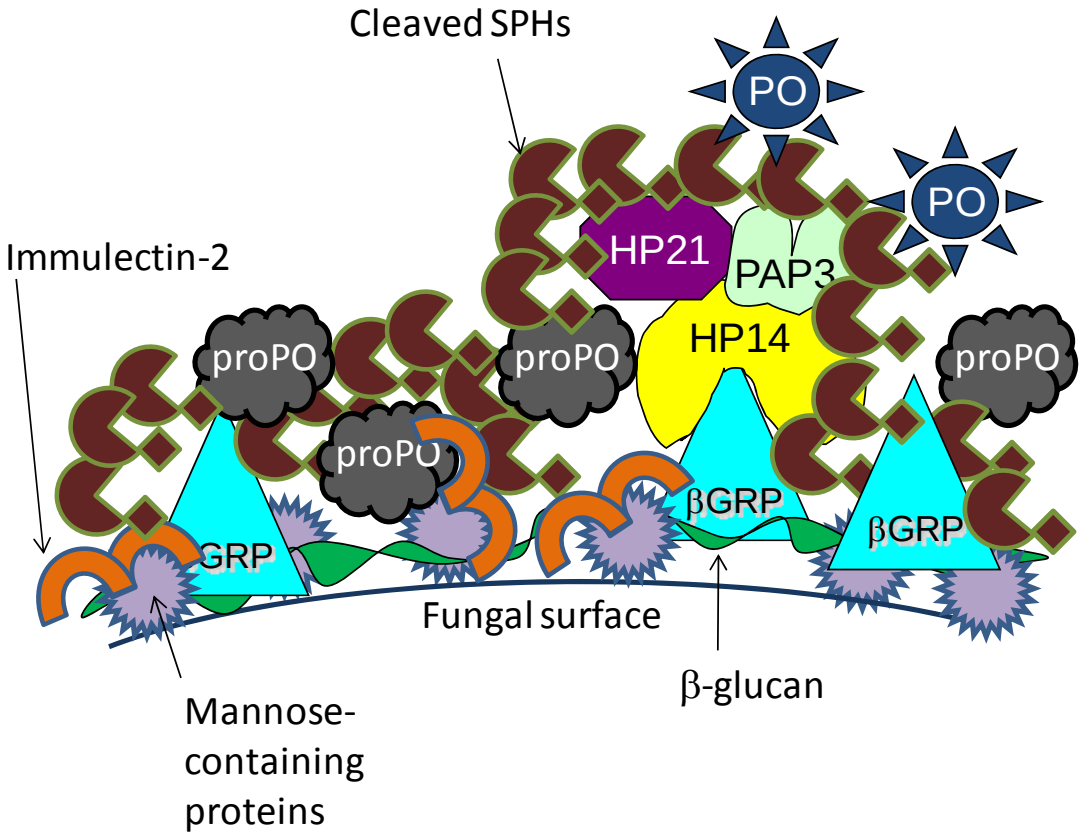


Figure 5-2. A model of immune-complex formation on a fungal surface containing β-1,3-glucan and mannose.

The fungal surface outside the plasma membrane contains β-glucans and mannose containing proteins. βGRP binds the β-glucans and immulectin-2 binds mannose. HP14 binds βGRP and activates proHP21, which activates proPAP3, leading to activation of proPO to PO in the co-presence of cleaved SPHs. Cleaved SPHs and proPO interact with each other and also bind other components of this complex, including immulectin-2 and possibly βGRP.