"MOLECULAR CHARACTERIZATION OF PROTEASE INHIBITORS FROM THE HESSIAN FLY, [*MAYETIOLA DESTRUCTOR* (SAY)]"

by

MADDUR APPAJAIAH ASHOKA

B. Sc. (Forestry), University of Agricultural Sciences, Bangalore, India, 1994
M. Sc. (Agricultural Biochemistry), University of Agricultural Sciences, Bangalore, India, 1998

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Entomology
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2005

ABSTRACT

Analysis of transcriptomes from salivary glands and midgut of the Hessian fly [*Mayetiola destructor* (Say)] identified a diverse set of cDNAs that were categorized into five groups, group I – V, based on their phylogenetic relationship. All five of these groups may encode putative protease inhibitors based on structural similarity with known proteins. The sequences of these putative proteins among different groups are highly diversified. However, sequence identity and structural analysis of the proteins revealed that all of them contained high cysteine residues that were completely conserved at their respective positions among these otherwise diversified proteins. Analysis of bacterial artificial chromosome (BAC) DNA for two groups, group I (11A6) and group II (14A4), indicated that group I might be a single copy gene or genes with low copy number whereas group II exists as multiple copies clustered within the Hessian fly genome. To test the inhibitory activity and specificity of these putative proteins, recombinant proteins were generated. Enzymatic analysis of the recombinant proteins against commercial and insect gut proteases demonstrated that recombinant proteins indeed are strong inhibitors of proteases with different specificities. Northern analysis of the representative members of five groups revealed that the group I-IV genes were expressed exclusively in the larval stage with variations among groups at different larval stages. The group V (11C4) genes were expressed in the late larval and pupal stage. Tissue specific gene expression analysis revealed that group I-IV genes were predominantly expressed in malpighian tubules whereas the group V genes were abundantly expressed in the salivary glands. Localization experiments with the antibody for representative members from group II (14A4) demonstrated that the protein was predominantly localized in the malpighian

tubules and in low amounts in the midgut, suggesting that malpighian tubules are the primary tissue of 14A4 inhibitor synthesis. The overall results indicated that the Hessian fly contains a complex network of genes that code for protease inhibitors which regulate protease activities through different developmental stages of the insect.

"MOLECULAR CHARACTERIZATION OF PROTEASE INHIBITORS FROM THE HESSIAN FLY, [*MAYETIOLA DESTRUCTOR* (SAY)]"

by

MADDUR APPAJAIAH ASHOKA

B. Sc. (Forestry), University of Agricultural Sciences, Bangalore, India, 1994
M. Sc. (Agricultural Biochemistry), University of Agricultural Sciences, Bangalore, India, 1998

A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Entomology
College of Agriculture

KANSAS STATE UNIVERSITY

Manhattan, Kansas

2005

Approved by: Approved by:

Co-Major Professor   Co-Major Professor
Gerald E. Wilde   Ming-Shun Chen
Dept. of Entomology   USDA-ARS/KSU

ABSTRACT

Analysis of transcriptomes from salivary glands and midgut of the Hessian fly [*Mayetiola destructor* (Say)] identified a diverse set of cDNAs that were categorized into five groups, group I – V, based on their phylogenetic relationship. All five of these groups may encode putative protease inhibitors based on structural similarity with known proteins. The sequences of these putative proteins among different groups are highly diversified. However, sequence identity and structural analysis of the proteins revealed that all of them contained high cysteine residues that were completely conserved at their respective positions among these otherwise diversified proteins. Analysis of bacterial artificial chromosome (BAC) DNA for two groups, group I (11A6) and group II (14A4), indicated that group I might be a single copy gene or genes with low copy number whereas group II exists as multiple copies clustered within the Hessian fly genome. To test the inhibitory activity and specificity of these putative proteins, recombinant proteins were generated. Enzymatic analysis of the recombinant proteins against commercial and insect gut proteases demonstrated that recombinant proteins indeed are strong inhibitors of proteases with different specificities. Northern analysis of the representative members of five groups revealed that the group I-IV genes were expressed exclusively in the larval stage with variations among groups at different larval stages. The group V (11C4) genes were expressed in the late larval and pupal stage. Tissue specific gene expression analysis revealed that group I-IV genes were predominantly expressed in malpighian tubules whereas the group V genes were abundantly expressed in the salivary glands. Localization experiments with the antibody for representative members from group II (14A4) demonstrated that the protein was predominantly localized in the malpighian

tubules and in low amounts in the midgut, suggesting that malpighian tubules are the primary tissue of 14A4 inhibitor synthesis. The overall results indicated that the Hessian fly contains a complex network of genes that code for protease inhibitors which regulate protease activities through different developmental stages of the insect.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to expresses my deep gratitude to my major advisor Dr. Ming-Shun Chen for his constant supervision, guidance, encouragement, great patience in my research, for teaching me both critically and philosophically and for his great support in all aspects.

I am very grateful to my co-major advisor, Dr. Gerald E. Wilde and to my committee members, Dr. Subbaratnam Muthukrishnan, Dr. Jian-Min Zhou for their valuable suggestions and advice on my research. I would like to express special thanks to Dr. Subbaratnam Muthukrishnan for his help on my dissertation. I am also thankful to Dr. John E. Shirley for his willingness to serve as my committee chairperson.

I am very grateful to Dr. Brenda Oppert for allowing me to work in her lab, for her inputs on this dissertation and for her overall support.

I would like to thank Dr. Yoonseong Park and Dr. Kun Yan Zhu for allowing me to use their lab facility. I would like to thank Dr. Jianfa Bai and Dr. Philine Wangemann for allowing me to use their gene expression and COBRE confocal facility, respectively. I would like to thank Dr. Mariam Burton for technical help on confocal microscopy.

I am very thankful to Dr. Sonny Ramaswamy for his advice, help and support during my degree program. I am thankful to the department faculty for teaching scientific knowledge and all the staff for their help.

I wish to thank all my colleagues in the lab, Dr. Xuming Liu, Xiang Liu, Sheila Prabhakar, Raman Bansal, Chitvan Khajuria for their kind co-operation, understanding and assistance. I would like to thank all my friends, Venki, Nandi, Navdeep, Prasanna,

1. Review of Literature

1.1 Proteolysis

Proteolysis is the breakdown of proteins into amino acids through the action of various proteases. Proteolysis is involved in various cellular and biological processes such as DNA replication, cell-cycle progression, cell proliferation, and migration, morphogenesis and tissue remodeling, neuronal outgrowth, haemostasis, wound healing, immunity, angiogenesis and apoptosis in all living organisms (Sternlicht and Werb, 2001). There are two forms of proteolysis: limited proteolysis and non-limited proteolysis (Thumm, 1993). Limited proteolysis leads to the activation of immature proteins, inactive proteins, or zymogens by cleavage of one or a limited number of peptide bonds. Non-limited proteolysis is a process where the active biological molecules are rendered inactive through hydrolysis of many peptide bonds in the molecule. The amino acid constituents generated through this hydrolysis are recycled back to protein synthesis. Alternatively these amino acids can be used for metabolism to produce energy. Limited and non-limited proteolysis are the two major tools for a given cell or organism to activate, inactivate, or dispose proteins at a given time to control its biological processes.

1.1.1 Limited Proteolysis. Limited proteolysis occurs in proteins that undergo protein secretion, hormone processing and enzyme activation (Fuller et al., 1988). Limited proteolysis often happens through the action of proteases which act sequentially or at different levels of a cascade in a given biological phenomenon. For example, limited proteolysis can lead to the activation of one enzyme, which in turn activates a subsequent target in a cascade. This type of proteolysis usually leads to the sequential activation or

1

termination of a biological process. Some of the classical examples of limited proteolysis are the activation of the mating pheromone α-factor by proteolytic processing enzymes in yeast, cascades controlling dorsal–ventral fate in drosophila, hemolymph clotting in arthropods, complement and blood clotting in vertebrates (Bussey, 1988; Fueller et. al., 1988; Krem and Cera, 2002). In yeast, the α-factor is synthesized as precursor with a molecular mass of 28 kDa. This precursor contains four tridecapeptides. The maturation of the α-factor occurs during passage through secretory pathway via limited proteolysis catalyzed by a serine proteinase, yscF. The processed form is necessary for the formation of diploid yeast cell from two haploid cells.

In animals, many cascades are controlled by upstream and downstream serine proteases. For example the dorso-ventral fate in *Drosophila*, hemolymph clotting in horse shoe crab, and complement and blood clotting in vertebrates consist of a series of upstream and downstream serine proteases. These serine proteases undergo zymogenic activation sequentially leading to terminal cleavage of substrate by the downstream proteases (Fig. 1). The terminal protease in *Drosophila* leads to activation of a toll pathway by activating the toll ligand whereas in horseshoe crab and in vertebrates the terminal serine protease cleaves the substrate leading to formation of blood clots.

1.1.2 Unlimited Proteolysis. Unlimited proteolysis inactivates active proteins upon receiving the intra or extra cellular signals. Under various physical and biological conditions, the normal half life of proteins vary from a few minutes to over 100 hours (Thumm, 1993). The turnover rate of proteins is determined by various physical and biological environments. The degradation of proteins is achieved by the action of specific exopeptidases or endopeptidases.

1.2 Regulation of proteolysis

Since it plays a vital role in various cellular processes, proteolysis must be strictly regulated. Deficiency in proteolytic regulation could lead to serious pathological conditions such as cancer, arthritis, neurodegenerative and cardiovascular diseases (Sternlicht and Werb, 2001). Proteolysis can be regulated by different mechanisms such as:

1) Specificity of proteases to their substrates
2) Enzyme activation and inactivation
3) Tagging
4) Compartmentalization
5) Inhibition

1.2.1 Specificity of proteases. Proteases involved in regulatory functions work in a concerted manner to process a given substrate. A great portion of proteases do not attack substrates at random, but display a high degree of specificity in identification and digestion of their substrates. For example, trypsins and chymotrypsins are serine proteases with different specificities. The specificity of proteases provides one way to avoid unwanted proteolysis.

1.2.2 Protease activation and inactivation. Another way of regulating proteolysis is through protease activation or inactivation. Many proteases such as chymotrypsin are synthesized as inactive zymogenic forms that need to be converted into active form by proteolytic excision (Khan and James, 1998; Lazure, 2000). For example, the inactive chymotrypsinogen is converted to active chymotrypsin by removal of a

peptide by trypsin. In addition to cleavage activation, some zymogenic proteases can be activated by other ways including binding with other proteins (Friedrich, 2003) or with salts (Huang et al., 2001b) or under certain temperature condition (Spiess et al., 1999). For example, *Staphylococcus aureus,* an human pathogen, secretes staphylocoagulase which activates prothrombin by binding to the proenzyme. The activity of human kallikrein 3 is regulated by salt concentration *in vivo*. The heat shock protein DegP (HtrA) has both molecular chaperone and proteolytic activities. At low temperatures it functions as a chaperone and at elevated temperatures it functions as a protease. Activation of inactive proteases is an important way to regulate proteolysis that happens only in certain conditions or specific developmental stages.

1.2.3 Tagging. Tagging is one way to regulate proteolysis in a unlimited proteolytic process. Tagging is achieved by a molecule called ubiquitin, a 76 amino acid protein in all eukaryotes. In a tagging process, a protein that is targeted for degradation is covalently linked between a lysine residue of the targeted protein and ubiquitin moiety of the ubiquitin molecule. This covalent linkage is catalyzed by ubiquitin activating enzymes. Once it is tagged, the protein will be degraded by proteosomes. Proteosome is a subcellular structure that consists of proteases and other constitutes. For example, the yeast 20S proteosome exhibits three distinct proteolytic activities, a chymotrypsin-like, trypsin-like and peptidyl glutamyl peptide hydrolyzing activity (Orlowski, 1990; Heinemeyer et al., 1991; Tanka et al., 1992). The mammalian cell contains a more complex structure called the 26 S proteosome, which contains a 26 S protease with Mr 1300-1500 kDa as well as other trypsin-like, chymotrypsin-like, and glutamyl proteases.

1.2.4 Compartmentalization. Compartmentalization of proteases is another way to prevent unwanted proteolysis in a proteolytic process. The two best known compartments for proteolysis are vacuole in yeast and lysosome in mammals, both of which contain various hydrolytic proteases. The hydrolytic enzymes in each compartment belong to aspartic, cysteine, or serine protease families. Targeted proteins are transported by various means to these compartments for degradation or for activation. For example, the *Saccharomyces cerevisiae* α-factor receptor (STE3) is transported to vacuole though endocytic pathway for degradation (Davis et. al., 1993). Besides vacuole and lysosomes, there are other organelles such as endoplasmic reticulum, golgi bodies, peroxisomes that also serve as compartments for proteolytic regulations.

In insects, compartmentalization of digestive enzymes has been observed. The midgut lumen is divided into endo- and ectoperitrophic space by the peritrophic membrane. Peritrophic membrane is made of protein-carbohydrate matrix. Proteases involved in initial digestion are found in the endoperotrophic space. The proteases involved in intermediate and final stages of digestion are found in ectoperitrophic space. In *Rhynchosciara ameircana* larvae trypsin and α–amylase are present in midgut cells, ecto- and endoperitophic spaces. Aminopeptidases and trehalase are present in midgut cells and ectoperitrophic space. The dipeptidases are present only in the midgut cells (Terra and Ferreira, 1994).

1.2.5 Inhibition. Another important way to regulate proteolysis is the inhibition of protease activity through the network of protease inhibitors. Protease inhibitors are present both inter- and intra-cellularly. The best known examples of protease inhibitors present intercellularly are inhibitors in blood. Vertebrate blood contains various serine

proteases that are involved in blood clotting. The activated enzymes are regulated by the action of various inhibitors that are circulating in the blood. In humans, deficiency of inhibitors circulating in the blood leads to various clinical disorders. For example deficiency of α-1 antitrypsin, a serine protease inhibitor of neutrophil elastase, leads to emphysema. Protease inhibitors found intracellularly are located in the cytoplasm or in the nucleus (Grigoryev et al., 1999) and regulate various proteases intracellularly. The lysosomes proteases are regulated by protease inhibitors if they are released into the cytoplasm (Turk et al., 2002). The caspases involved in apoptotic pathways are known to be regulated by various inhibitors of apoptosis (IAPs) after their release from lysosome.

The regulation of proteolysis by protease inhibitors has been well documented in various insects. For example, melanization is one of the defensive mechanisms adopted by insects against various pathogens (Soderhall and Cerenius, 1998). Melanization process involves activation of proteases leading to conversion of propeholoxidase (PPO) to phenoloxidase. The activated enzyme catalyzes formation of quinones which polymerize to form melanin. The quinones and reactive oxygen species produced during melanization are toxic to insect cells and microorganisms (Soderhall and Cerenius, 1998). Under normal conditions to avoid unwanted production of these reactive molecules, insects have evolved various serine protease inhibitors to control the activity of proteolytic enzymes (Sugumaran et al., 1985; Aspan et al., 1990; Boigegrain et al., 1992; Jiang and Kanost, 1997).

In *Leucophaea maderae,* protease activity is found in the posterior half of the midgut, whereas the anterior midgut contains protease inhibitor. The biological significance of the protease inhibitor in the anterior portion of the midgut is not known

(Engelmann et al., 1980). Since, trypsin inhibitors (YIs), subtilisn inhibitors 1 and 2 (SI1, and SI2), inhibit endogenous proteases from posterior midgut, it is believed that the endogenous inhibitors TIs, SI1, and SI2 act as regulators of proteases that are involved in food digestion and possibly protect gut tissues from degradation (Elpidina et al., 2001).

1.3 Proteases

Proteases are the enzymes that catalyze the irreversible breakdown of peptide bonds in proteolysis (Table 1). Based on functional groups present at the active site and on the catalytic mechanism, proteases are classified into four groups: serine, cysteine, aspartic, and metalloproteases (Hartely, 1960). According to International Union of Biochemistry and Molecular Biology nomenclature (Int. Union Bioch., 1992), proteinacious enzymes belong to one of the six recognized groups. The six groups are 1)oxidoreductases, 2)transferases, 3)hydrolases, 4)lyases, 5)isomerases, and 6)ligases. The six groups are further classified into different subclasses based on the type of bonds they cleave. Proteases belong to group 3 (hydrolases) and subgroup 4 (peptidases, act on peptide bonds) (Int. Union Bioch., 1992). The subclasses of peptidases fall into two sets, exopeptidases and endopeptidases. Exopeptidases cleave scissile bond proximal to amino or carboxy terminus of substrate whereas endopeptidases cleave internal scissile bonds of substrate. Based on the amino acid sequences, proteases are classified into different families (Rawlings and Barrett, 1993). A family is a group of peptidases that are related evolutionarily and have similar primary structure. Each family of peptidases are assigned with a code letter denoting the type of catalysis, i.e., S, C, A, M, T or U for serine, cysteine, aspartic, metallo, threonine- or unknown type, respectively. These families are

further subdivided into "clans" to accommodate sets of peptidases that have a common ancestor (Rawlings and Barrett, 1993). Presently there are more than 30 families of serine-, cysteine-, and metallo-proteases and a few families of aspartic-, threonine- and mixed- proteases (Barrett, 1999; Merops database). The families for known proteases are summarized in Table 2A-F.

1.3.1 Serine Proteases. Serine proteases are one of the most important protease super-families. Serine proteases are found in all organisms studied so far and carry out various physiological functions, such as digestion, defense, and protein activation (Yousef et al., 2003). Serine proteases are characterized by the presence of a serine residue at the active site. Based on primary structural homology, serine proteases are grouped into more than 20 families and each of these families is further subdivided into several clans. Members in each clan have a common ancestor (Rawlings and Barrett, 1993) (Table 2A). The primary structures of four clans including chymotrypsin (SA), subtilisin (SB), carboxypeptidase C (SC), and *Escherichia* D-Ala-D-Ala peptidase A (SE) are totally unrelated, suggesting at least four separate evolutionary origins for serine proteases. The clans SA, SB, and SC consists of a common catalytic triad of three amino acids, serine (nucleophile), aspartate (electrophile), and histidine (base) and have a common reaction mechanism. The three catalytic triads of the three clans are oriented in a similar manner, but the protein folds are quite different, indicating that these proteases have undergone convergent evolution over a period of time. The catalytic mechanisms of clans SE and SF (repressor LexA) are different from those of clans SA, SB, and SE, since they lack the classical Ser-His-Asp triad. Another interesting feature of serine proteases is

the conservation of glycine residues in the vicinity of the catalytic serine residue to form the motif Gly-Xaa-Ser-Yaa-Gly (Brenner, 1988).

1.3.2 Insect Serine Proteases. In many insects, serine proteases are the major digestive and regulatory enzymes (Applebaum, 1985; Terra and Ferreira, 1994). Among serine proteases, trypsin and chymotrypsin have important role in dietary protein digestion (Terra and Ferreira, 1994). Protein digestion occurs in insect midgut epithelium. Initially the proteins are cleaved into oligomers by serine proteases in the endoperotrophic space of ventriculus. These oligomers are further hydrolyzed by action of exopeptidases in the ectoperitrophic space and on microvillar membranes. The di- and tripeptides generated are cleaved by dipeptidases that are present in the midgut cells and glycocalyx. Trypsins and chymotrypsins are observed in midgut from various orders of insect which include Coleoptera, Diptera, Dictyoptera, Heteroptera, Hymenoptera, Lepidoptera, Orthoptera, Thysanura, (Gooding and Rolseth, 1976; Jany et al., 1978; Baker, 1981; Briegel, 1985; Sakal et al., 1988; Ferreira and Terra, 1989; Baumann, 1990; Christeller et al., 1990; Graf and Graf et al., 1991; Johnston et al., 1991; Lemos and Terra, 1992; Zinker and Polzer, 1992; Schumaker et al., 1993; Ferreira et al., 1994; Zhu et al., 2005; Volpicella et al., 2003; Zhu et al., 2003a). In contrast to vertebrate trypsins, insect trypsins are not activated by calcium ions (Levinski et al., 1977; Jany et al., 1978; Johnston et al., 1991; Lemos and Terra, 1992). Insect trypsins are unstable in acidic pH (Miller et al., 1974; Ward et al., 1975; Jany et al., 1978; Sakal et al., 1989) with the exception of some enzymes from *Musca domestica* and *Ostrinia nubilalis* (Lemos and Terra, 1992; Bernardi et al., 1996). Serine proteases with elastase specificity are less

studied in insects. Presence of elastase like enzymes is documented by Christeller et al., (1989), Christeller et al., (1992), Valaitis (1995).

1.4 Protease inhibitors

Protein protease inhibitors are polypeptides that inhibit the action of proteases (Laskowski and Kato, 1980, Travis and Salvesen, 1983). The existence of protease inhibitors in nature was first reported by Fermi and Pernossi (Birk, 2003). Protease inhibitors are widely distributed in various tissues of animals, plants and microorganisms (Laskowski and Kato, 1980). In plants protease inhibitors are found to be induced in response to insect attack (Ryan, 1990). Individual inhibitors inhibit proteases of their mechanistic class identified by Hartley (1960) except α2-macroglobulins which inhibit proteases of all classes. Accordingly, protease inhibitors are divided into 4 major mechanistic classes: inhibitors of serine proteases, inhibitors of cysteine proteases, inhibitors of aspartic proteases and inhibitors of metalloproteases. The protease inhibitors of these mechanistic classes are further classified into different families.

1.4.1 **Classification of Protease Inhibitors.** Originally protease inhibitors are grouped into 9 families based on amino acid sequence homology, topological relationships between disulfide bonds and the location of reactive sites (Laskowski and Kato, 1980). With the discovery of new protease inhibitors, the protease inhibitor families were expanded. Following the Laskowski and Kato (1980) homology criteria, Reeck et al., (1997) made a complete list and grouped inhibitors into 22 families (Table 3). Individual families are identified by a prominent member of that family without

considering functional and biological source. According to this classification, inhibitor families include members that may inhibit proteases of different mechanistic class, so called "cross class" inhibitors. The specificity of inhibitors largely relies on the exposed reactive-site loop (Laskowski and Kato, 1980; Laskowski, et al. 1987). The sequence of reactive site loop is highly variable whereas the scaffolding structure that supports the loop is quite conserved. The combination of a variable loop and a conserved scaffolding structure appears to be the basis for inhibitory selectivities, which is supported by the fact that differences in reactive-site loop sequence of turkey ovomucoid lead to differences in selectivities of the inhibitors (Bigler et al., 1993).

Protease inhibitors exist in different tissues and stages of organisms from bacterium to humans. Among the four mechanistic classes of protease inhibitors, namely serine, cysteine, aspartic and metallo protease inhibitors, serine protease inhibitors are most widely studied. According to the above classification, 15 out of 22 families are serine protease inhibitors. The 15 families of serine protease inhibitors are:

1.4.2 Bovine pancreatic trypsin inhibitor (Kunitz) family. The pancreatic trypsin inhibitor identified from bovine is a typical example of this family. This is the first inhibitor whose sequence and three dimensional structure of free enzyme and enzyme-inhibitor complex was determined. The inhibitors of this family contain four to eight disulfide bonds and two inhibitory domains. Inhibitors with two inhibitory domains are called double headed inhibitors. The inhibitors of this family have been found from humans, bovine, pig, birds, snails, and from snake venom.

1.4.3 Kazal serine protease inhibitor family. Secretory pancreatic trypsin inhibitor (Kazal type), an initial member of this group, is found in both vertebrates and invertebrates. The inhibitors belonging to this group have one to many reactive site domains. The number of disulfide bonds vary from 3 (bovine pancreatic tryspin inhibit) to 20 (avian egg white ovoinhibitor) (Laskowski and Kato, 1980).

1.4.4 Soybean trypsin inhibitor (Kunitz) families. Members in the Soybean trypsin inhibitor (Kunitz) families contain a single inhibitory domain that inhibits single protease molecule. This type of protease inhibitors have been called single-headed. The inhibitor in this family contains two disulfide bonds. In most soybean trypsin inhibitor kunitz family the P_1 is usually arginine.

1.4.5 Bowman-Birk inhibitors (BBI). Bowman-Birk inhibitors are isolated from all leguminous plants. BBIs usually have two inhibitory domains that can inhibit two protease molecules. They are ~70 amino acids in length and contain 7 disulfide bonds. BBIs are potent inhibitor of insect digestive trypsins and chymotrypsins (Levinsky et al., 1977; Sakal et al., 1989). Suzuki, et al., (1987) described structure of BB type protease inhibitor from peanuts. The inhibitor is tetramer and each monomer consists of two distinct domains. The inhibitor consists of two protruding sites that independently bind to two protease molecules. Under reduced conditions, the inhibitors lose 3-D structure but retain some inhibitory activity (Godbole et al., 1994).

1.4.6 Potato inhibitor I family. This family is also referred to as chymotrypsin inhibitor I because they inhibit chymotrypsin very strongly. The inhibitors of this group are noncovalent tetramers with four different subunits. The molecular mass of each

subunit is 10 kDa. Each subunit consists of one intra chain disulfide bond (Melville and Ryan, 1972). The tetramer binds four chymotrypsin molecules thus each subunit possess one chymotrypsin binding site.

1.4.7 Potato inhibitor II family. The molecular weight of the inhibitor is 20 kDa. The inhibitor is a dimer and each subunit has two binding sites for chymotrypsin. The inhibitor II preparations from different potato varieties are immunologically related (Ryan and Santarius, 1976). There is no consensus observed in the reactive site sequences of this family of inhibitors.

1.4.8 Squash inhibitor family. The inhibitors of this family are small molecular weight proteins ~3000 Da (Wieczorek et al., 1985). They consists of 27-33 amino acids residues and are cross-linked by three intra disulfide bonds. The reactive site peptide bond is located between amino acid residue 5 and 6. The inhibitors of this group are highly stable and rigid proteins. Due to rigid nature of the proteins, the structure is distorted and contains triple stranded antiparallel β sheets (Otlewski and Krowarsch, 1996).

1.4.9 Barley trypsin inhibitor family. The prototype of this group is barley trypsin inhibitor which is of 13 kDa protein and contains five disulfide bonds (Odani et al., 1983). The anionic and cationic inhibitors isolated from buckwheat seeds are highly pH and thermostable (Belozersky et al., 2000). The inhibitors of this group inhibit trypsin and to certain extent they inhibit chymotrypsin too.

1.4.10 *Ascaris* trypsin inhibitor family. The most striking characteristic of this family is the universal presence of ten cysteine residues that form five disulphide bonds.

Two disulfide bonds are located on either side of the reactive-site loop. The inhibitors contain two β sheets and each β sheet is composed of two antiparallel strands (Grasberger et al., 1994; Duggan et al., 1999). They lack true hydrophobic core. The reactive-site residues are hypervariable, and this hypervariability is due to the selective pressure brought by their host cognate enzymes (Hawley and Peanasky, 1992; Lu et al., 1998). Well characterized insect inhibitors that belong to this group are chymotrypsin inhibitors (AMCI) 1 to 3 isolated from larval hemolymph of *Apis mellifera.*

1.4.11 Locust inhibitor family. Members of this group are ~35-residues long and consist of three antiparallel β-strands stabilized by three disulfide bonds. The protease binding loop is located near the C-terminus and is flanked by two disulfides bonds. Well-characterized inhibitors of this family are from the migratory locust, *Locusta migratoria* and the desert locust, *Schistocerca gregaria*. The inhibitors are found in many tissues example the brain, fat body, ovaries of the insects (Simonet et al., 2005).

1.4.12 Ecotin family. Ecotin is the primary member of this family of inhibitors. Ecotin is a *Escherichia coli* periplasmic protein of 142 amino acids (McGrath et al., 1991). It consists of two β sheets that are sandwiched. It is a potent inhibitor of serine proteases that have wide substrate specificities. Each individual molecule of ecotin interact C-terminally to form a dimer. The dimer binds two protease molecules. Ecotin monomer consists of one disulfide bond near to the reactive site loop that provides stability to the molecule.

1.4.13 Serpin family. Serpins are found in both prokaryotes and eukaryotes (Potempa et al., 1994; Silverman et al., 2001). Their secondary structure consists of three

β- sheets, nine α-helices, and a reactive site loop (RSL) which confers specificity to protease recognition (Huber and Carrell, 1989; Crowther et al., 1992). The RSL is composed of ~17 amino acids. The RSL is a flexible region and is capable of distorting target protease upon entering into the enzymes active site.

1.4.14 Streptomyces subtilisin inhibitor family Streptomyces subtilisin inhibitors (SSI) are found widely in *Streptomyces* species. They exhibit inhibitory activity against both subtilisin BPN' and trypsin. The SSI is a dimeric molecule of identical subunits and is a 23 kDa protein (Takahashi and Fukada, 1985). The two subunits interact with two protease molecules.

1.4.15 Hirudin family. Hirudin is a potent thrombin inhibitor belongs to this family. Hirudin is a 65 amino acids protein with Mr of ~7000 Da. Hirudin consists of an N-terminal globular domain and an extended C-terminal domain. The N-terminal residues 1-3 form parallel β-strand with 214-217 residues of thrombin. The C-terminal domain makes numerous electrostatic interactions with the exosite of thrombin. The last five residues at C-terminal are in a helical loop. They form many hydrophobic contacts with thrombin (Rydel et al., 1990).

1.4.15 α-2-Macroglobulin family. Human α-2-macroglobulin, the major plasma protein belongs to this family. It contains four polypeptide subunits. Each subunit is 180 kDa. Each subunit consists of a bait region of 25-30 amino acid residues and a thioester bond, both located in the middle of the subunit. These two sites play important role in inhibiting the proteases (Sottrup-Jensenn and Birkedal-Hansen, 1989).

1.5 Diversity in proteases and protease inhibitors

Due to their roles in various biological processes, there is an increasing interest in the identification and functional characterization of proteases that are present in various organisms. At least 5% of the human genes codes for proteolytic process components including its proteases, inhibitors and cofactors (Richard, 2005). The Merops, a protease database, lists 461 proteases and homologues in human, 363 in *C. elegans*, 513 in *Drosophila*, 383 in *Mus musculus*, and 227 in *Rattus norvegicus* (Table 3) (López-Otín and Overall, 2002). Up to now, only a tiny portion of these protease are studied. The protease map of murine is similar to the human map, but protease families, kallikreins, cathepsins or matrix metalloproteases (MMPs) have evolved differently (Balbin et al., 2001; Yousef and Diamandis, 2001; Deussing et al., 2002) (Table 4). The *Drosophila melanogaster* gene content is considerably lower than vertebrate organisms, but a similar number of protease genes to that of humans are identified in the fly genome. Studies are needed to clarify the genetic, functional and evolutionary differences between the protease repertoires of these organisms. In addition to the universal job of proteolytic cleavage that is common to all organisms, there are many proteases that carry specific functions in different organisms. Studies in this field will uncover diverse new proteolytic substrates and new endogenous inhibitors that have the ability to regulate protease activity in physiological and diseased conditions.

Like proteases, protease inhibitors also occur widely in nature and are known to regulate various functions ranging from cell activity to cell death (Dickinson et al., 1995; Armstrong et al., 1996). Of the total genes identified in the genome, the inhibitors account for about 0.75% and 1.17% in *Drosophila* and humans respectively (Merops

database, July 25th 2005). In *Anopheles gambiae*, the inhibitors accounts for about 0.62%. The Merops database contains about 3690 inhibitor proteins that are grouped into different families. Of the four known class of protease inhibitors, serine protease inhibitors are known to participate in highly diversified actions in regulating proteolysis.

1.6 Insect protease inhibitors

As in other organisms, protease inhibitors are believed to occur abundantly in insects (Eguchi, 1993). Of the four mechanistic classes of inhibitors, serine protease inhibitors are more extensively studied in insects and will be reviewed in more detail later. For the other three mechanistic classes of inhibitors, only a few have been identified from a number of insects. These include several cysteine protease inhibitors from *Drosophila, Sarcopgaha, Spodoptera, Trichoplusia,* and *Bombax,* which are found to participate in insect defense, to block apoptosis, and to involve in pupal diapause (Suzuki and Natori, 1985; Yamamoto, 1999; Meier et al., 2000; Huang et al., 2001a; Goto and Denlinger, 2002; Liao et al., 2002; Liu and Chejanovsky, 2003). In addition, a few metalloprotease inhibitors have been identified from *Galleria melonella* and *Drosophila* (Wedde et al., 1998; Pohar et al., 1999). The immune response related inhibitor, IMPI from *Galleria melonella* is the first metalloprotease identified from insects. To date, no aspartic protease inhibitor has been identified from insects. A list of inhibitors identified from different insect species is presented in the Table 5.

1.6.1 Insect serine protease inhibitors. Serine protease inhibitors identified from different insects are known to perform diverse functions. The insect inhibitors identified

17

or partially characterized from the insect hemolymph belong to either low molecular weight proteins (Ramesh et al., 1988; Boigegrain, et al., 2000; Frobius et al., 2000; Gáspári et al., 2002; Shrivastava and Gosh, 2003) and or the serpin family (Kanost, 1990; Cherqui et al., 2001). The hemolymph inhibitors regulate the activity of serine proteases involved in prophenoloxidase cascade and also play a role in insect defense against pathogens and injury. Examples for low molecular weight inhibitors in insect hemolymph are the inhibitors identified from the pupa of housefly, *Musca domestica* L., which are heat-stable low molecular weight peptides and are potent phenoloxidase inhibitors (Tsukamoto et al., 1992). HiTI, a 7 kDa protein from the fly, *Haematobia irritans irritans,* contains a single Kunitz type domain and is known to play a role in antimicrobial defense (Azzolini et al., 2004). For serpin inhibitors, *Manduca sexta* hemolymph contains different isoforms of serpins. These serpins regulate the prophenoloxidase activation pathway in the hemolymph (Kanost et al., 1989; Kanost, 1990; Zhu et al., 2003b; Wang and Jiang, 2004; Tong and Kanost, 2005). The *Locusta migratoria* inhibitors (LMCI I and II) and the *Mythimna unipuncta* trypsin (MTI) and α–chymotrypsin (MCI) inhibitors also inhibit prophenoloxidase activation cascade (Boigregrain et al., 1992; Cherqui et al., 2001).

In addition to hemolymph inhibitors, serine protease inhibitors in other tissues with different functions have also been identified. The Kunitz-type cocoon shell-associated trypsin inhibitor (CSTI), is present in the silk of *Bombax mori* prevents inappropriate degradation of silk proteins during their secretion (Kurioka et al., 1999). Serine protease inhibitors in the midgut regulate midgut proteases of many insects (Gatehouse and Gatehouse, 1998; Elpidina, 2001; Kellenberger et al., 2003). Inhibitors

isolated from *Rhodnius prolixus* and *Dipetalogaster maximus* prevent blood clotting once injected into the hosts (Friedrich et. al., 1993; Lange et al., 1999; Francischetti et al., 2000; Campos et al., 2002; Schlott et al., 2002) and thus helps insects in feeding blood. The anti-hemostatic activity of thrombostatin plays an important role in maintaining the ectoparasitic lifestyle of horn flies (Zhang et al., 2002). Protease inhibitors are also components are venom. A protease inhibitor isolated from bee venom is highly stable at low pH and does not contain any SH groups (Shkenderov, 1973). The purified inhibitor inhibits trypsin and possibly play protective role of bee venom components from proteases of stung man or animal.

Some insect inhibitors are known to contain high cysteine residues. For example, the *Apis mellifera* cathepsin/chymotrypsin inhibitors 1-3 (AMCI 1-3) isolated from larval hemolymph belong to *Ascaris* family (Bania et al., 1999). The AMCI-1 inhibits both chymotrypsin and Cathepsin-G. All three isoforms contain 10 cysteine residues and lack tryptophan, tyrosine, and histidine. The major isoform AMCI-1 is a single-chain protein whereas AMCI-2 and -3 are two-chain proteins. The reactive site of inhibitors is Met30-Gln31. All the three inhibitors are 44% identical to *Ascaris* trypsin isoinhibitor (ATI) and belong to *Ascaris* family. NMR studies shows that AMCI-1, consists of two perpendicular sheets, several turns and a loop that host a protease binding site (Cierpicki et al., 2000). The *Drosophila* seminal fluid protein, Acp62F, is a trypsin protease inhibitor (Lung et al., 2002). It contains 10 cysteine residues that are highly conserved with *Ascaris* members. It is proposed that it could play a role in reducing the female life span. The *Anopheles gambiae* genome contains a few uncharacterized protease inhibitors that belong to *Ascaris* family. The *Galleria melonella* metalloprotease inhibitor, IMPI,

belongs to *Ascaris* family and contains 10 cysteine residues (Wedde et al., 1998; Clermont et al., 2004). The peptides SGCI, SGTI, PMP-D2 and PMP-C, isolated from locusts contain 6 cysteine residues that three intramolecular disulfide bonds. The peptide PMP-C contains a fucose moiety (Nakakura et al., 1992; Gaspari et al., 2002).

Given the abundance of insect species and only a small portion of protease inhibitors have been identified from a few insect species. There are lots of opportunities to identify new protease inhibitors which have great potential in understanding basic biology as well as practical applications.

1.7 Action of serine proteases and inhibition mechanism

The amino acid sequence analysis and X-ray crystallographic studies of many serine proteases have shown that the catalytic site lies in the cleft on the surface of enzyme molecule. The substrate chain lies along the active site cleft. There are subsites that adapted to bind amino acid side chains of a substrate on either side of active site cleft (Fig. 2). Individual subsites of an enzyme (….S_3, S_2, S_1, S_1', S_2', S_3'….) and the complementary part of a substrate (…..P_3, P_2, P_1, P_1', P_2', P_3'….) are named according to Berger and Schechter, (1970). In the center of the cleft is the catalytic triad consisting of His 57, Asp 102 and Ser 195 (chymotrypsinogen numbering). During catalysis, the initial step is to form an acyl enzyme intermediate between the substrate and serine residue at the active site of the enzyme. This covalent complex proceeds via negatively charged transition state to form a tetrahedral intermediate. During deacylation, the acyl-enzyme intermediate is hydrolyzed by a water molecule to release cleaved peptide and restores the Ser-hydroxyl of the enzyme (Fastrez and Fersht, 1973).

If a protein can interact with the active site of a protease in a manner similar to a substrate and block (or slow down) the activity of this protease, then this protein is called a protease inhibitor. According to Laskowski and Kato, (1980), the inhibition mechanism by which inhibitor suppress protease activity by acting as substrates of their target enzymes is referred to standard mechanism. Inhibitors that follow standard mechanism have a similar reactive site loop configuration. The loop consists of a scissile bond between two amino acid residues P_1 and P_1' on the substrate (Fig. 2). P_1 residue at the N-terminal end can either be lysine/arginine in trypsin inhibitors. In chymotrypsin inhibitors the P_1 residue could be either Tyr/Trp/Phe/Leu/Met. In elastase the P_1 residue is alanine/serine. P_1' residue is located adjacent to P_1 by a scissile bond. The inhibitors that obey standard mechanism encompass at least one disulfide bond at the reactive site peptide bond (Fig. 3). This ensures close proximity and no dissociation of two peptide chains when virgin inhibitor converted to cleaved inhibitor.

Laskowski and Kato (1980) gave the overall mechanism of enzyme-inhibitor interaction (Fig. 4). In their model, the enzyme and inhibitor interact to form a stable complex (C) with relatively small conformational change. This stable complex dissociates very slowly to give free enzyme and virgin (I) or cleaved inhibitor (I*). In I* the scissile bond at P_1-P_1' is cleaved. The enzyme and inhibitor interact predominantly in a canonical fashion. In complexes with their cognate enzymes, the reactive site loop of the inhibitor associates with catalytic residues in a similar way as that of bound substrates (Fig. 5). In trypsin like proteases, the segment amino terminal of the scissile bond of the inhibitor fits in to the enzyme as an antiparallel β-strand. This is facilitated through main chain hydrogen bond formation between P_3 and P_1. The flanking carboxy terminal

residue at P_2' position interacts to form one more hydrogen bond. Under this condition, the inhibitor reactive site is in close proximity to catalytic residues of the protease. The P_1 carbonyl carbon forms contact with Ser195 O^γ through van der Waals interaction. The carbonyl group of P_1 projects in to the oxyanion hole to form two hydrogen bonds with Gly 193N and Ser195N (Bode and Huber, 1992). The scissile peptide bond remains intact. The exposed P_1 residue of the inhibitor gets buried in the S_1 specificity pocket of the enzyme which is hydrophobic. The interaction of the P_1 side chain with S_1 specificity pocket are energetically most important and this determines the specificity of a given inhibitor for a particular protease. Inhibitors have ability to tolerate replacement of P_1 residue and still retain inhibitory activity either with retention of original specificity (eg. Arg $\longrightarrow$ Lys or Leu $\longrightarrow$ Met) or with change in specificity from tryptic to chymotryptic after Arg $\longleftrightarrow$ Phe replacement. The residue at P_2 position in all active inhibitors is neutral. In most inhibitor families, the P_1' position can tolerate broad range of residues with the exception that serine proteases rarely hydrolyze bonds with P_1' Pro. Inhibitors with P_1' Pro are not active (Qasim et al., 1995).

In addition to the primary interactions, the exposed side chains of residues around reactive site (P_9 to P_4') make hydrophobic interaction with protease subsites to form inhibitor-protease complex. The complex formation is accompanied by slight conformational changes which freezes the binding loop of the inhibitor. This change is essential to form tight, rigid and energetically favorable inhibitor-protease complex. The stable complex acts as energy sink which dissociates very slowly to give free enzyme and cleaved inhibitor.

Exceptions to the standard mechanism are α_2-macroglobulin, ecotin and serpins. α_2-macroglobulin inhibits the serine proteases by engulfing the proteinase after cleavage of α_2-macroglobulin molecules by their target enzymes (Sottrup-Jensen and Birkedal-Hansen, 1989). Ecotin, a non-canonical inhibitor from *E. coli* inhibits proteases that have wide specificities. Each ecotin homodimer inhibits two enzyme molecules. It has a less typical substrate like primary site and an antibody like secondary site, which allows inhibitor to bind tightly to proteases and to block the substrate binding site (McGrath et al., 1995). The primary site interactions that are less optimal are compensated by the additional secondary site interactions. Serpins to certain extent are canonical inhibitors. Serpins employ suicide substrate inhibition mechanism. The actions of the protease, both in initial recognition and in subsequent steps, are of a serine protease acting on a substrate peptide bond to cleave it. However, in serpin, the two chain form is not inhibitory because of the conformational changes. Initially, the proteinase recognizes the reactive bond on the reactive center loop (RCL) of serpin as a potential substrate to form Michaelis complex (Ye et al., 2001; Dobo' and Gettins, 2004). The protease sits on top of serpin RCL and the side chain of P_1 residue resides in the protease specificity pocket. The scissile bond in the RCL is cleaved and the catalytic serine residue in the protease is acetylated. The RCL inserts into the β-sheet A of serpin molecule. The protease undergoes translocation from top of serpin to bottom of serpin molecule to form stable serpin-protease complex (Dobo' and Gettins, 2004) (Fig. 6).

1.8 Structural characteristics of serine protease inhibitor families

Protease inhibitors are rigid and stable molecules. Most of the inhibitors are competitive inhibitors and form tight complex with proteases. Inhibitors of different mechanistic class contain β-sheet and α-helices. The number of β-sheet and α-helices vary among different mechanistic class. Inhibitors with single polypeptide chain (monomer), two polypeptide chains (dimer) and many polypeptide chins (multimer) exits in nature. For, example, potato inhibitor II is a dimer and potato inhibitor I and α2-macroglobulin are tetramers. The inhibitor consists of a reactive site domain that forms a loop which fits into the enzyme specificity pocket. The length of reactive site domain varies depending on the size of the inhibitor molecule. It may vary from a few residues such as in *Cucurbita maxima* trypsin inhibitor to forty five residues such as in serpins (Jiang et al., 1994). The number of reactive site domains varies in each inhibitor molecule. An inhibitor of single polypeptide chain may contain one to several reactive site domains. Based on the number of reactive site domains in the inhibitor molecule, they may be referred to as single headed if they contain one reactive site domain, double headed if they contain two reactive site domains and so on. The single headed inhibitor inhibits one enzyme molecule, like wise double headed can inhibit two molecules of the same or different enzymes.

The gene size and coding regions of protease inhibitors are generally small except in some cases of inhibitors that belong to serpin family, macroglobulin family, cysteine and metallo protease inhibitor families. The genes contain no to a few introns and many inhibitors are multigene families (Boulter, 1993; Ryan, 1990). The double and multi-

headedness of protease inhibitors have arisen by gene duplication and gene elongation and in some cases by gene fusion (Laskowski and Kato, 1980).

Of the protease inhibitors, the structures of canonical inhibitors of serine proteases have been studied extensively. Canonical inhibitors are proteins that range in size from 29 to 190 amino acids. Majority of inhibitors are purely β-sheet or mixed α/β proteins. They can also be α-helical or irregular proteins rich in disulfide bonds. The most striking feature of canonical inhibitor families is that the reactive site loops are very similar even though the P_3-P_3' segment is different among different families. Figure 7A-C shows inhibitors belonging to different families have an exposed binding loop which is unique to canonical inhibitors (Bode and Huber, 1992). The loop is highly flexible in uncomplexed state and becomes rigidified upon complex formation. The structures outside the binding loop posses different folding motifs (β-sheets and α-helices). The loop which is primary binding segment between the framework anchoring points has a flat shape that fits into the active-site cleft of cognate proteases. The inhibitors have compact shape and contain a hydrophobic core which in some cases formed by cross connecting disulfide bonds. The stability of uncleaved inhibitor domain towards unfolding is high. The disulfide bonds stabilize the inhibitor and the extent of stabilization depends on the number of thiol groups that form disulfide bond under native state.

1.9 Wheat production and Hessian fly

Wheat (*Triticum aestivum* L. em. Thell) is the national staple food staple of 43 countries and provides approximately 20% of the total calories for the people of the

world. The total acreage of wheat grown annually in U.S varies between 50 and 80 million acres. One of the major challenges to wheat production is yield loss and quality reduction caused by insects and diseases.

Among various pests of wheat, Hessian fly is one of the major insect pests. The Hessian fly was introduced into the United States by Hessian soldiers in 1776 (Gagne and Hatchett 1989) and it occurs in all major wheat growing regions of the United States (Ratcliffe and Hatchett, 1997). The annual loss from the Hessian fly infestation to the U. S. farmers is up to 5-10% of the wheat crop per year (Hatchett et al., 1987). In addition to wheat, the Hessian fly also damages barley, rye and *Triticale*. The insect damages the plant by sucking the plant juice. The ingested juice enters the insect gut where the juice constituents are degraded by action of various proteases. Protease activity in the gut was detected by Shukle et al., (1985). Recently Zhu et al., (2005) also showed the presence of serine protease activity in the gut of Hessian fly. Fifteen trypsin and chymotrypsin cDNAs have cloned and characterized. The initial protease activity and protease gene identification in Hessian fly indicate that there is a complex of protease network like that in other organisms. Understanding the regulation of this complex network of proteases and the regulation of proteolysis by protease inhibitors in the Hessian fly will not only contribute to the basic understanding of the proteolysis process, but may also have practical application such as finding new ways to manage this insect pest. Identification and characterization of proteases and protease inhibitors is the first step to understand the proteolytic system and its regulation. To date no protease inhibitors have been identified from the Hessian fly. My research focuses on identification and molecular

26

characterization of cDNAs encoding protease inhibitors from the Hessian fly. My results

will provide a basis to reveal the proteolytic system and its regulation in the future.

Figure 1: Serine protease cascade controlling (a) dorso-ventral polarity in *Drosophila* (b) hemolymph clotting in horseshoe crab (c) complement activation and (d) blood clotting in vertebrates. Square colored boxes represent zymogenic serine proteases. Colored oval shapes denote active serine proteases. Upstream proteases activate downstream proteases and substrates below them. Dashed boxes and ovals represent non-serine protease substrates and cleavage products. Adopted from Krem and Cera (2002).
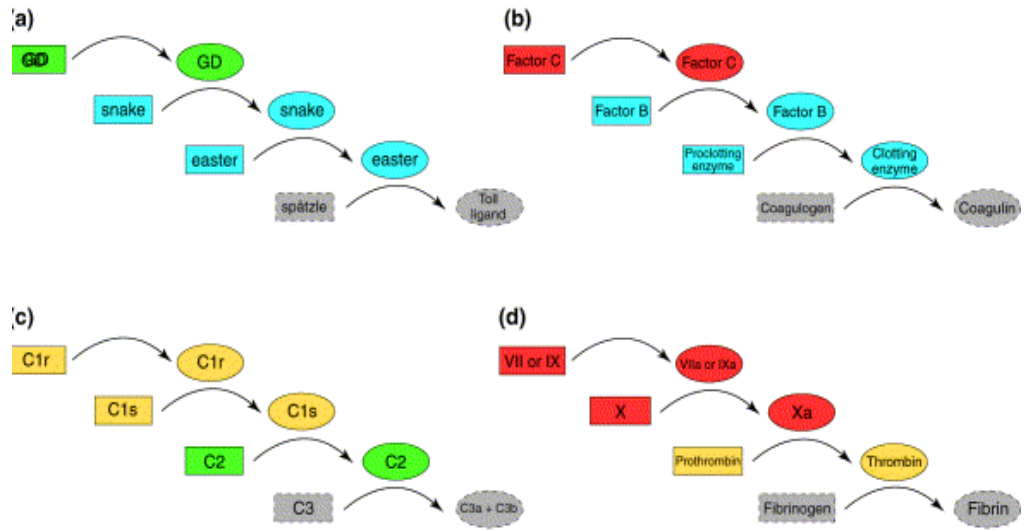
Figure 2: Schematic representation of specificity subsites of proteases and the corresponding complementary sites of substrates. The subsites of proteases, S_1, S_2, S_3… are away from the active site towards N-terminus. S_1', S_2' and S_3' are towards C-terminus. Arrow indicates the cleavage site. Substrate sites corresponding to subsites of enzymes P_1, P_2, P_3…. are towards N-terminus. P_1', P_2', P_3'……. are towards C-terminus.
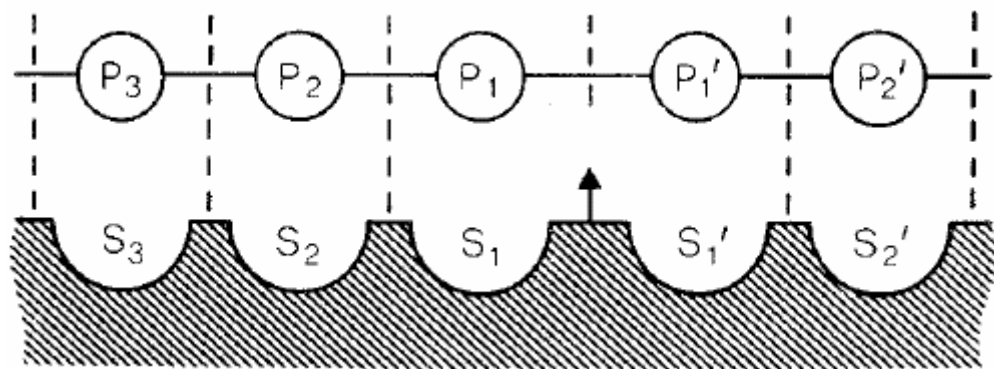
Figure 3: Schematic representation of inhibitor reactive site. I - represents the virgin inhibitor, I* - represents the cleaved inhibitor. Arrow in the virgin inhibitor represent the peptide bond cleavage site between P_1-P_1' residues. The reactive center site in the virgin inhibitor is encompassed by disulfide bond. The two peptide chains of cleaved (I*) inhibitor is secured by disulfide bond.

H_3N^+….. Cys P_5 P_4 P_3 P_2 P_1 ———↓——— $P^{'}_1$ $P^{'}_2$ $P^{'}_3$ $P^{'}_4$ $P^{'}_5$Cys ……COO$^-$

Virgin Inhibitor - I

S ———————————————————————— S

↕

H_3N^+….. Cys P_5 P_4 P_3 P_2 P_1-COO$^-$ $^+$HN-$P^{'}_1$ $P^{'}_2$ $P^{'}_3$ $P^{'}_4$ $P^{'}_5$ Cys ……COO$^-$

Cleaved Inhibitor – I*

S ———————————————————————— S
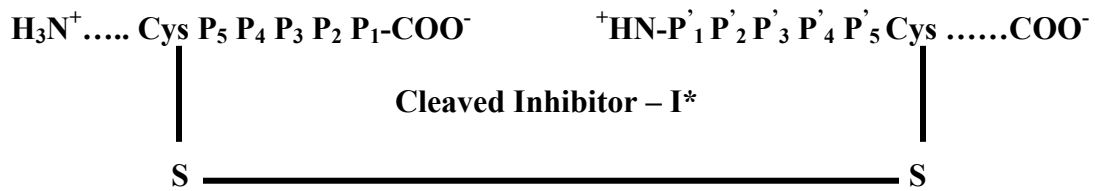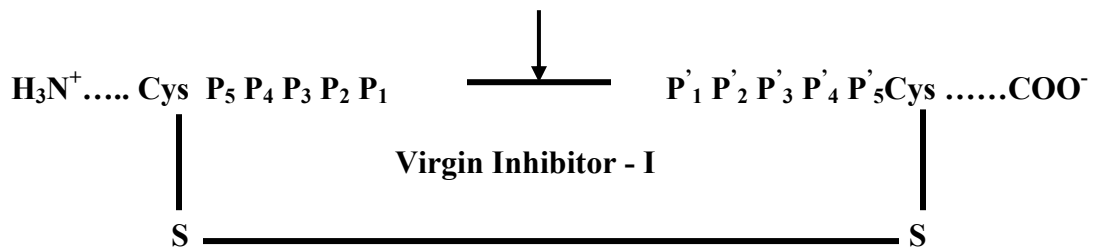
Figure 4: Enzyme-inhibitor interaction. Figure adopted from Laskowski and Kato (1980). E-enzyme, I and I^* - virgin and modified inhibitors, L and L^* - loose, noncovalent complexes of E with I and I^* respectively, C-stable enzyme inhibitor complex, X-long lived intermediate in the $E+I^*$ reaction

$$E + I \Longleftrightarrow L \Longleftrightarrow C \Longleftrightarrow X \Longleftrightarrow L^* \Longleftrightarrow E + I^*$$
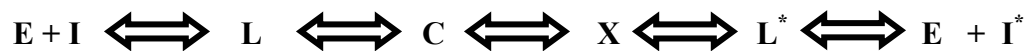
Figure 5: Complex formation between trypsin and ovomucoid inhibitor. Adopted from Bode and Huber (1992). The reactive site loop spans between scaffold-anchored pillars. The inhibitor is connected to its molecular core through intra disulfide bond and other spacer elements. The inhibitor is bound similarly as that of substrate (shown in bold line) through hydrogen bonds forming an antiparallel β-structure between P_1-P_3 of inhibitor and 216 to 218 of protease. The catalytic Ser195 O^{γ} is in contact with P_1 carbonyl carbon of inhibitor through van der Waals forces.
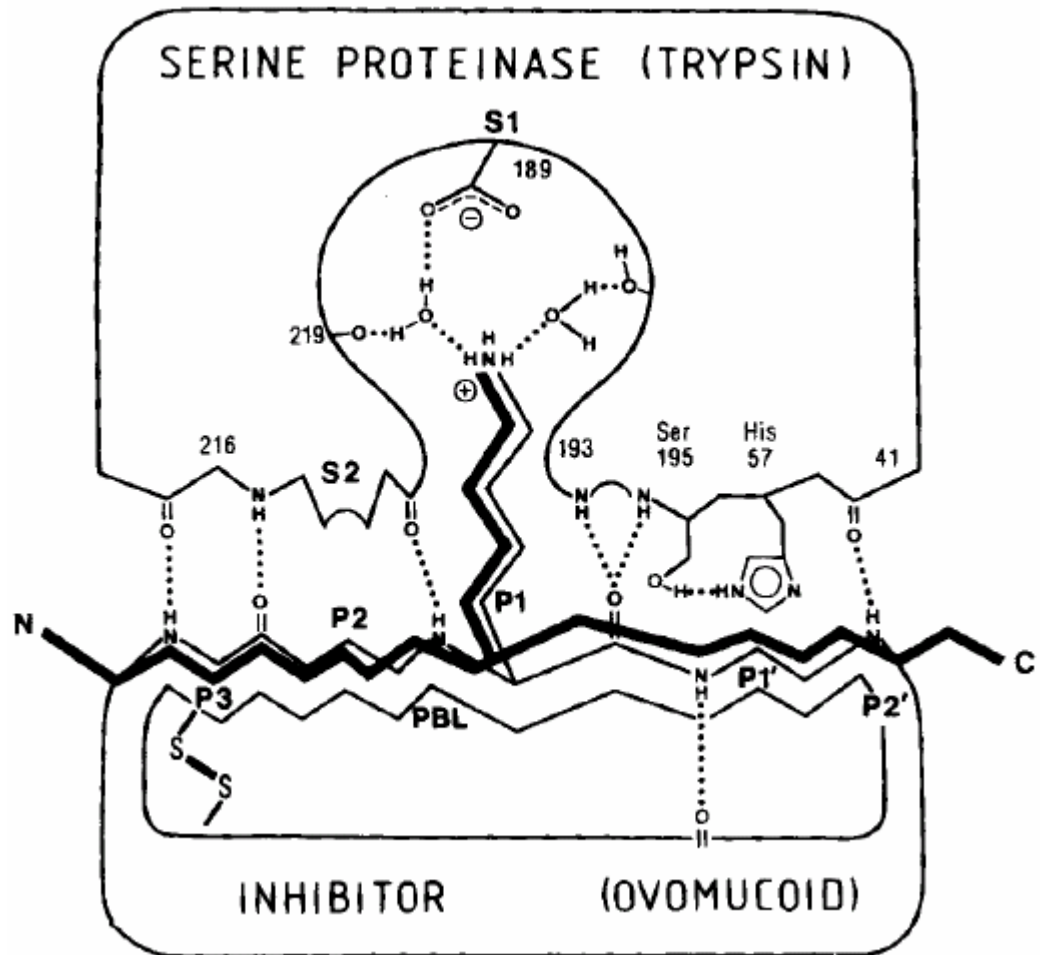
Fig. 6: The branched pathway mechanism of serpins. Figure adopted from Dobo' and Gettins, (2004). E–Protease, I–serpin, EI-noncovalent Michaelis-like complex, EI'-covalent acyl-enzyme intermediate, E-I*-covalent complex (E-I*), I*-cleaved serpin (I*). k_1, k_{-1}, k_2, k_3, k_4, and k_5 are rate constants. Reactive center loop of serpin - solid black color.
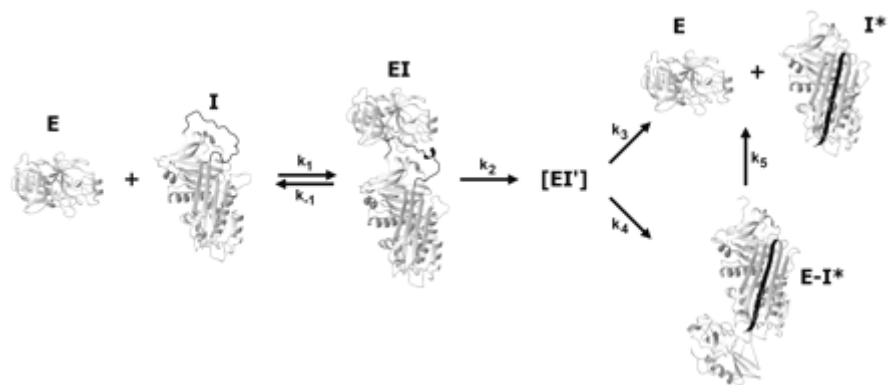
Figure 7: Ribbon model of canonical inhibitors showing the similar reactive site loop. Adopted from Bode and Huber (1992). A- Ribbon model of pancreatic basic trypsin inhibitor. Semisynthetic arginine15 is shown on the binding loop. Inhibitor consists of three stranded β–sheet and a C-terminal α-helix that form the core.
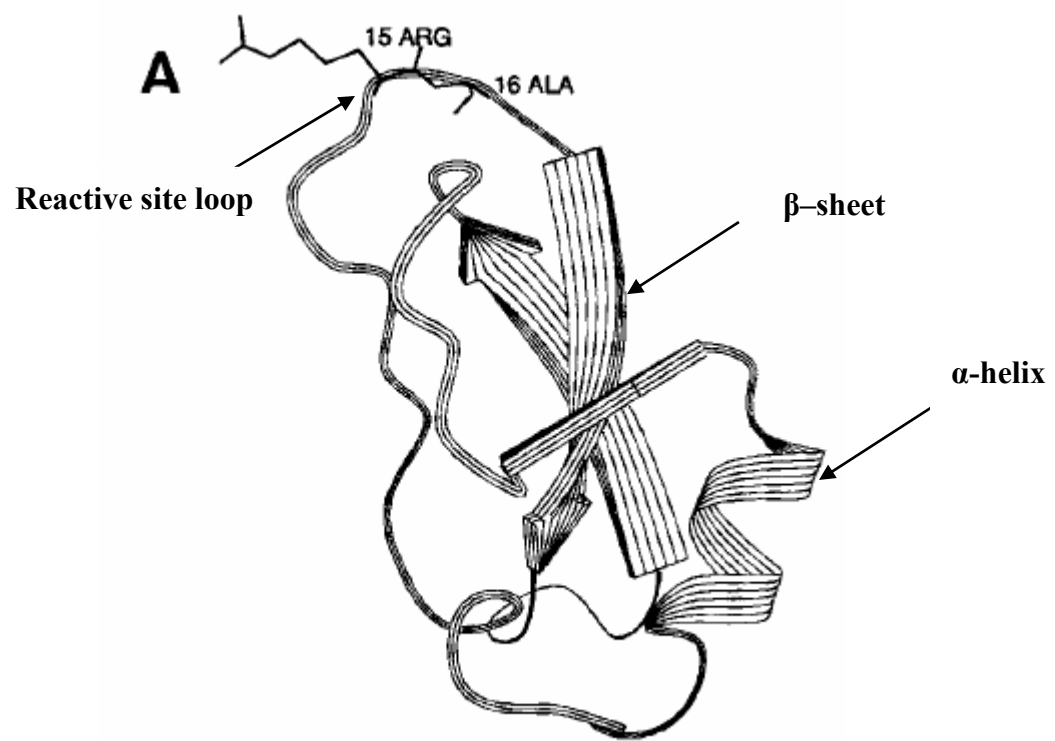
A

15 ARG

16 ALA

Reactive site loop

β–sheet

α-helix

Figure 7: Ribbon model of canonical inhibitors showing the similar reactive site loop. Adopted from Bode and Huber (1992). B-Ribbon model of eglin c. Semisynthetic leucine 45 (P_1) and arginine 46(P_1') is shown on the binding loop. The inhibitor consists of four stranded mixed antiparallel (β1, β4, β3) and parallel (β3, β2) β-sheet forms the core.
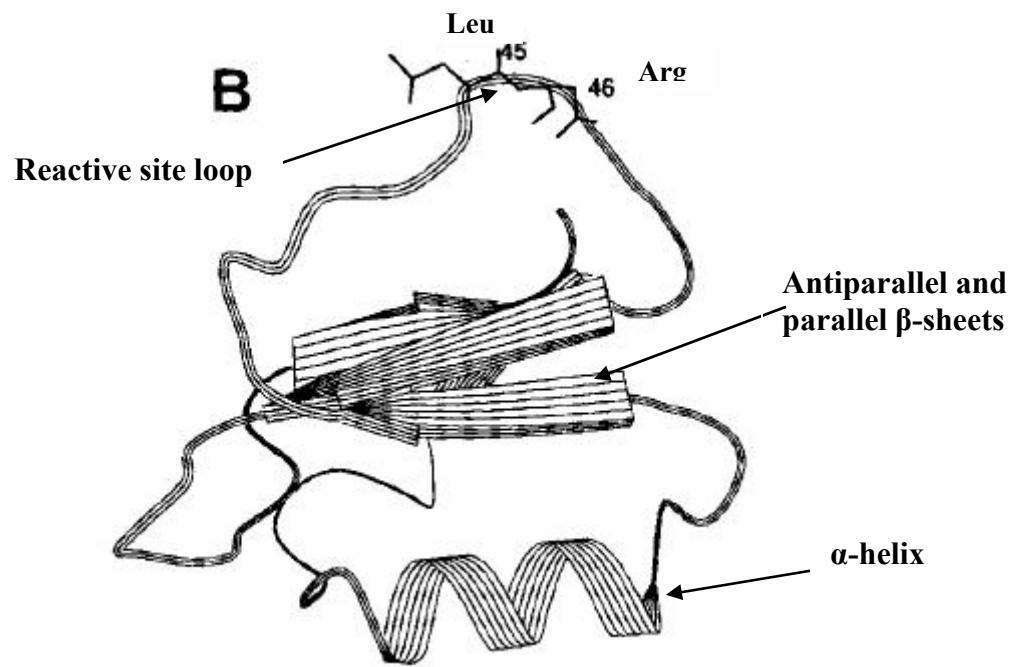
B

Leu

45

46

Arg

Reactive site loop

Antiparallel and parallel β-sheets

α-helix

Figure 7: Ribbon model of canonical inhibitors showing the similar reactive site loop. Adopted from Bode and Huber (1992). C-Ribbon model of squash seed inhibitor-1(CMTI-1). The semisynthetic arginine 5 (P_1) and isoleucine 6 (P_1') is shown on the loop. It consists of two stranded β-ladder.

C ARG 5 ILE 6

Reactive site loop

β-ladder
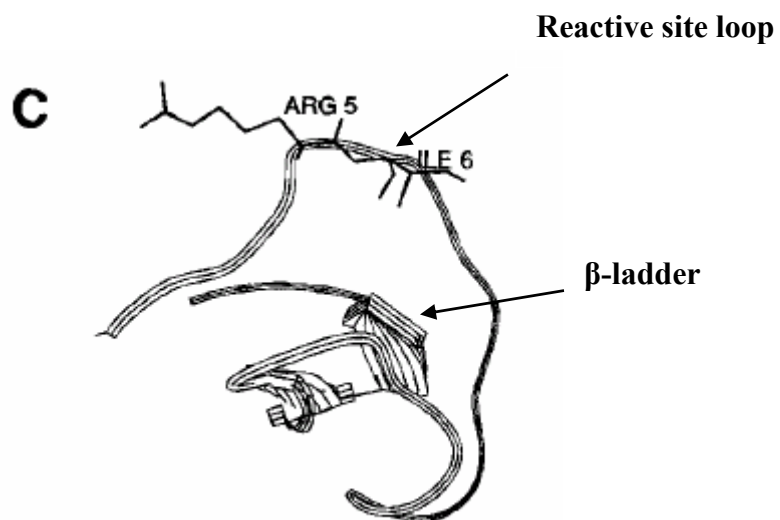
Table 1: Specificity of Proteases. Adopted from Rao et al., (1998).

Enzyme	Peptide bond cleaved
Trypsin	Lysin (or Arg) ▼-----
Chymotrypsin, subtilisin	Try (or Trp, Phe, Met) ↓------
Staphylococcus V8 protease	Asp (or Glu) ▼------
Papain	Phe (or Val, Leu)-Xaa ▼--
Thermolysin	---▼ Leu (or Phe) ↓---
Pepsin	-Phe (or Try, Leu) -Try (or Phe, Try)

▼ Arrow indicates the site action of protease. Xaa any amino acid residue.

Table 2A: Serine protease clans and families.

Clan	Family	Example
SA	S1	Chymotrypsin A (*Bos taurus*)
	S2	Stretogrisin A
	S3	Togavirin (Sindbis virus)
	S6	IgA1-Specific serine endopeptidase (*Neisseria gonorrhoeae*)
	S7	Flavivirin (yellow fever virus)
	S29	Hepatitis C virus NS3 polyprotein peptidase
	S30	Potyvuris P1 proteinase (plum pox virus)
	S31	Pestivirus polyprotein peptidase p80 (bovine viral diarrhea virus 1)
	S32	Equine arteritis virus serine endopeptidase (Equine arteritis virus)
	S35	Apple stem growing virus protease
SB	S8	Subtilisn (*Bacillus licheniformis*)
SC	S9	Prolyl oligopeptidase (*Sus scrofa*)
	S10	Carboxypeptidase Y (*Saccharomyces cerevisiae*)
	S15	X-Pro dipeptidyl-peptidase (*Lactococcus lactis*)
	S28	Lysosomal Pro-X carboxypeptidase (*Homo sapiens*)
	S33	Prolyl aminopeptidase (*Neisseria gonorrhoeae*)
	S37	PS-10 peptidase (*Streptomyces lividans*) (*Streptomyces lividans*)
SE	S11	D-Ala-D-Ala carboxypeptidase A (*G. stearothermophilus*)
	S12	D-Ala-D-Ala carboxypeptidase B (*Streptomyces sp.*)
	S12	D-Ala-D-Ala peptidase C (*Escherichia coli*)
SF	S24	Repressor LexA (*Escherichia coli*)
	S26	Signal peptidase I (*Escherichia coli*)
	S27	Eukaryotic signal peptidase
	S41	C-terminal processing peptidase-1 (*Escherichia coli*)
SH	S21	Assemblin (human herpesvirus 5)
	S23	Escherichia protease I
	S25	Multicatalytic endopeptidase ciomplex
SJ	S16	Lon-A peptidase (*Escherichia coli*)
	S50	Birnavirus Vp4 peptidase (infectious pancreatic necrosis virus)
SK	S14	Clp endopeptidase (subunit clp) (*Escherichia coli*)
	S41	C-terminal processing peptidase-1 (*Escherichia coli*)
	S49	Signal peptide peptidase A (*Escherichia coli*)
SP	S59	Nucleoporin 145 (*Homo sapiens*)
SR	S60	Lactoferrin (*Homo sapiens*)
S-	S46	Dipeptidyl-peptidase 7 (*Porphyromonas gingivalis*)
	S48	HetR endopeptidase (*Anabaena variabilis*)
	S54	Rhomboid-1 (*Drosophila melanogaster*)
	S62	Influenza A PA endopeptidase (influenza A virus)
	S63	EGF-like module containing mucin-like hormone receptor-like 2 (*Homo sapiens*)

Table 2B: Cysteine protease clans and families.

Clan	Family	Example
CA	C1	Papain (*Carica papaya*)
	C2	Calpain-2 (*Homo sapiens*)
	C6	Potato virus Y-type helper component peptidase
	C7	Chestnut blight fungus virus p29 peptidase
	C8	Chestnut blight fungus virus p48 peptidase
	C9	Sindbis virus-type nsP2 peptidase
	C10	Streptopain (*Streptococcus pyogenes*)
	C12	Ubiquitinyl hydrolase-L1 (*Homo sapiens*)
	C16	Murine hepatitis coronavirus papain-like endopeptidase 1
	C19	Ubiquitin-specific peptidase 14 (*Homo sapiens*)
	C21	Tymovirus endopeptidase (turnip yellow mosaic virus)
	C23	Carlavirus endopeptidase (apple stem pitting virus)
	C27	Rubella virus endopeptidase
	C28	Foot-and-mouth disease virus L-peptidase
	C31	Porcine respiratory and reproductive syndrome arterivirus-type Cysteine peptidase alpha (lactate-dehydrogenase-elevating virus)
	C32	Equine arteritis virus-type cysteine peptidase (porcine reproductive and respiratory syndrome virus)
	C33	Equine arterivirus Nsp2-type cysteine peptidase
	C36	Beet necrotic yellow vein furovirus-type papain-like endopeptidase
	C39	Bacteriocin-processing peptidase (*Pediococcus acidilactici*)
	C42	Beet yellows virus-type papain-like peptidase (beet yellows virus)
	C47	Staphopain A (*Staphylococcus aureus*)
	C51	D-alanyl-glycyl endopeptidase (*Staphylococcus aureus*)
	C54	ATG4 peptidase (*Saccharomyces cerevisiae*)
	C58	YopT peptidase (*Yersinia pestis*)
	C64	Cezanne deubiquitinating peptidase (*Homo sapiens*)
	C65	Otubain-1 (*Homo sapiens*)
	C66	IdeS peptidase (*Streptococcus pyogenes*)
	C67	CylD protein (*Homo sapiens*)
	C71	*Pseudomurein endoisopeptidase* Pei (Methanobacterium phage psiM2)
	C72	HopPtoN g.p. (*Pseudomonas syringae*)
CD	C11	Clostripain (*Clostridium histolyticum*)
	C13	Legumain (*Canavalia ensiformis*)
	C14	Caspase-1 (*Rattus norvegicus*)
	C25	Gingipain R (*Porphyromonas gingivalis*)
	C50	Separase (*Saccharomyces cerevisiae*)
CE	C5	Adenain (human adenovirus type 2)
	C48	Ulp1 endopeptidase (*Saccharomyces cerevisiae*)
	C55	YopJ endopeptidase (*Yersinia pseudotuberculosis*)
	C57	Vaccinia virus I7L processing peptidase (vaccinia virus)
	C63	African swine fever virus processing peptidase
CF	C15	Pyroglutamyl-peptidase I (*Bacillus amyloliquefaciens*)
CH	C46	Hedgehog protein (*Drosophila melanogaster*)
CL	C60	Sortase A (*Staphylococcus aureus*)
C-	C18	Hepatitis C virus endopeptidase 2 (hepatitis C virus)
	C40	Dipeptidyl-peptidase VI (*Bacillus sphaericus*)
	C53	Pestivirus Npro endopeptidase (classical swine fever virus)
	C70	AvrRpt2 g.p. (*Pseudomonas syringae*)

Table 2C: Metalloprotease clans and families.

Clan	Family	Example
MA	M1	Aminopeptidase N (*Homo sapiens*)
	M2	Angiotensin-converting enzyme peptidase unit 1 (*Homo sapiens*)
	M3	Thimet oligopeptidase (*Rattus norvegicus*)
	M4	Thermolysin (*Bacillus thermoproteolyticus*)
	M5	Mycolysin (*Streptomyces cacaoi*)
	M6	Immune inhibitor A (*Bacillus thuringiensis*)
	M7	Snapalysin (*Streptomyces lividans*)
	M8	Leishmanolysin (*Leishmania major*)
	M9	Microbial collagenase (*Vibrio alginolyticus*)
	M10	Collagenase 1 (*Homo sapiens*)
	M11	Gametolysin (*Chlamydomonas reinhardtii*)
	M12	Astacin (*Astacus astacus*)
	M13	Neprilysin (*Homo sapiens*)
	M26	IgA1-specific metalloendopeptidase (*Streptococcus sanguinis*)
	M27	Tentoxilysin (*Clostridium tetani*)
	M30	Hyicolysin (*Staphylococcus hyicus*)
	M32	Carboxypeptidase Taq (*Thermus aquaticus*)
	M34	Anthrax lethal factor (*Bacillus anthracis*)
	M35	Deuterolysin (*Aspergillus flavus*)
	M36	Fungalysin (*Aspergillus fumigatus*)
	M41	FtsH endopeptidase (*Escherichia coli*)
	M43	Cytophagalysin (*Cytophaga sp.*)
	M57	PrtB g.p. (*Myxococcus xanthus*)
	M60	Enhancin (Lymantria dispar nucleopolyhedrovirus)
	M61	Glycyl aminopeptidase (*Sphingomonas capsulata*)
	M64	IgA peptidase (*Clostridium ramosum*)
	M66	StcE peptidase (*Escherichia coli*)
	M72	Peptidyl-Asp metalloendopeptidase (*Pseudomonas aeruginosa*)
MC	M14	Carboxypeptidase A1 (*Homo sapiens*)
MD	M15	Zinc D-Ala-D-Ala carboxypeptidase (*Streptomyces albus*)
	M74	Murein endopeptidase (*Escherichia coli*)
ME	M16	Pitrilysin (*Escherichia coli*)
	M44	Pox virus metalloendopeptidase (vaccinia virus)
MF	M17	Leucyl aminopeptidase (*Bos taurus*)
MG	M24	Methionyl aminopeptidase 1 (*Escherichia coli*)
MH	M18	Aminopeptidase I (*Saccharomyces cerevisiae*)
	M20	Glutamate carboxypeptidase (*Pseudomonas sp.*)
	M28	Aminopeptidase S (*Streptomyces griseus*)
	M42	Glutamyl aminopeptidase (*Lactococcus lactis*)
MJ	M19	Membrane dipeptidase (*Homo sapiens*)
	M38	Beta-aspartyl dipeptidase (*Escherichia coli*)
MK	M22	O-sialoglycoprotein endopeptidase (*P. haemolytica*)
ML	M52	HybD endopeptidase (*Escherichia coli*)
	M63	Gpr peptidase (*Bacillus megaterium*)
MM	M50	S2P peptidase (*Homo sapiens*)
MN	M55	D-aminopeptidase DppA (*Bacillus subtilis*)

Table 2C: (continued)

Clan	Family	Example
MO	M23	Beta-lytic metalloendopeptidase (*Achromobacter lyticus*)
MP	M67	Poh1 peptidase (*Saccharomyces cerevisiae*)
M-	M29	Aminopeptidase T (*Thermus aquaticus*)
	M48	Ste24 endopeptidase (*Saccharomyces cerevisiae*)
	M49	Dipeptidyl-peptidase III (*Rattus norvegicus*)
	M56	BlaR1 peptidase (*Staphylococcus aureus*)
	M73	Camelysin (*Bacillus cereus*)
	M75	Imelysin (*Pseudomonas aeruginosa*)

Table 2D: Aspartic protease clans and families

Clan	Family	Example
AA	A1	Pepsin A (*Homo sapiens*)
	A2	HIV-1 retropepsin
	A3	Cauliflower mosaic virus-type endopeptidase
	A9	Spumapepsin (human spumaretrovirus)
	A11	Copia transposon (*Drosophila melanogaster*)
AB	A6	Nodavirus endopeptidase (flock house virus)
	A21	Tetravirus endopeptidase (Nudaurelia capensis omega virus)
AC	A8	Signal peptidase II (*Escherichia coli*)
AD	A22	Presenilin 1 (*Homo sapiens*)
	A24	type 4 prepilin peptidase 1 (*Pseudomonas aeruginosa*)
AF	A26	Omptin (*Escherichia coli*)
A-	A5	Thermopsin (*Sulfolobus acidocaldarius*)

Table 2E: Threonine and unknown protease clans and families.

Clan	Family	Example
T-	T5	Ornithine acetyltransferase precursor (*S. cerevisiae*)
U-	U4	Sporulation factor SpoIIGA (*Bacillus subtilis*)
	U9	Prohead peptidase (Enterobacteria phage T4)
	U32	Collagenase (*Porphyromonas gingivalis*)
	U35	Prohead peptidase (bacteriophage HK97)
	U40	Protein P5 murein endopeptidase (bacteriophage phi-6)
	U48	Prenyl peptidase 2 (*Saccharomyces cerevisiae*)
	U49	Lit peptidase (*Escherichia coli*)
	U57	YabG protein (*Bacillus subtilis*)
	U61	Muramoyl-tetrapeptide carboxypeptidase (*Escherichia coli*)
	U62	Microcin-processing peptidase 1 (*Escherichia coli*)

Table 2F: Clans of mixed proteases.

Clan	Family	Example
PA	C3	Poliovirus-type picornain 3C (human poliovirus 1)
	C4	Nuclear-inclusion-a endopeptidase (plum pox virus)
	C24	Rabbit hemorrhagic disease virus 3C-like endopeptidase
	C30	Porcine transmissible gastroenteritis virus-type main endopeptidase (transmissible gastroenteritis virus)
	C37	Calicivirin (Southampton virus)
	C62	Gill-associated virus 3C-like peptidase
	S55	SpoIVB peptidase (*Bacillus subtilis*)
	S64	Ssy5 endopeptidase (*Saccharomyces cerevisiae*)
PB	C44	Amidophosphoribosyltransferase precursor (*Homo sapiens*)
	C45	Acyl-coenzyme A:6-aminopenicillanic acid acyl-transferase precursor (*Penicillium chrysogenum*)
	C59	Penicillin V acylase (*Bacillus sphaericus*)
	C69	Dipeptidase A (*Lactobacillus helveticus*)
	S45	Penicillin G acylase precursor (*Escherichia coli*)
	S58	Aminopeptidase DmpA (*Ochrobactrum anthropi*)
	T1	Archaean proteasome, beta component (*Thermoplasma acidophilum*)
	T2	Glycosylasparaginase precursor (*Homo sapiens*)
	T3	Gamma-glutamyltransferase 1 (*Escherichia coli*)
PC	C26	Gamma-glutamyl hydrolase (*Rattus norvegicus*)
	C56	PfpI endopeptidase (*Pyrococcus furiosus*)
	S51	Dipeptidase E (*Escherichia coli*)

Table 3: Protein families of protease inhibitors. Adopted from Reeck et al., (1997).

Sl. No.	Name	Approximate domain size and number of domains
1.	Bovine pancreatic trypsin inhibitor (Kunitz)	60 residues
2.	Kazal serine protease inhibitor family	55 residues; up to 7 domains
3.	Soybean trypsin inhibitor (Kunitz) family	180 residues
4.	Bowman-Birk inhibitor family	35 residues; 2 domains
5.	Potato inhibitor I family	70 residues
6.	Potato inhibitor II family	50 residues; 1 or 2 domains
7.	Squash inhibitor family	30 residues
8.	Barley trypsin inhibitor family	120 residues
9.	Thaumatin family	200 residues
10.	*Ascaris* trypsin inhibitor family	60 residues
11.	Locust inhibitor family	35 residues
12.	Ecotin family	140 residues; 2 subunits
13.	Serpin family	400 residues
14.	*Streptomyces* subtilisin inhibitor	110 residues
15.	Hirudin family	65 residues
16.	Cystatin family	110 residues; usually single domain but 3 inhibitory domains in kininogens and 8 in potato mutlistatin
17.	Calpastatin family	140 residues; 4 domains
18.	Potato carboxypeptidase inhibitor family	40 residues
19.	*Ascaris* carboxypeptidase inhibitor family	40 residues
20.	Collagenase inhibitor family	200 residues
21.	*Ascaris* pepsin inhibitor family	150 residues
22.	α-2-macroglobulin family	1500 residues; 2 or 4 subunits

Table 4: Proteases in humans and model species. Adopted from López-Otín and Overall, (2002).

Species	Total	Catalytic class of protease				
		Aspartic	Cysteine	Metallo	Serine	Threonine
Homo sapiens	461	18	121	159	140	23
Caenorhabditis elegans	353	26	93	151	62	21
Drosophila melanogaster	513	38	59	157	225	34
Mus musculus	383	11	93	120	137	22
Rattus norvegicus	227	10	41	77	80	19

Table 5: Inhibitors studied/identified from different insect species.

Inhibitor class	Insect species
Serine protease inhibitors	*Acanthocephala femorata, Acheta domesticus, Aedes aegypti, Antheraea mylitta, Antheraea perni, Apis mellifera, Bombyx mori, Dactylopius coccus, Dipetalogaster maximus, Drosophila melanogaster, Drosophila virilis, Haematobia irritans, Hemileuca oliviae, Leucophaea maderae, Locusta migratoria, Manduca sexta, Mythimna unipuncta, Nauphoeta cinerea, Philosamia cynthia ricini, Rhodnius prolixus, Schistocerca gregaria, Sarcophaga bullata, Tenebrionidae spp., Traitoma infestin, Trichoplusia ni*
Cysteine protease inhibitors	*Bombax mori, Drosophila melanogaster, Sarcophaga crassipalpis, Spodoptera littoralis, Trichoplusia ni*
Metallo protease inhibitors	*Drosophila melanogaster, Galleria melonella*
Aspartic protease inhibitors	-

2. Research objectives

Under normal physiological condition proteolytic processes in an organism are precisely regulated in various ways. One way to control proteolysis is through protease inhibitors. Precise levels of protease inhibitors are to regulate proteases. Because of their importance in regulating proteolysis, protease inhibitors are of importance in various areas such as medicine and agriculture. In medicine protease inhibitors are used as drugs against the HIV proteases. In agriculture, many plant protease inhibitors are identified from various plant species that inhibit insect digestive proteases. These molecules serve as attractive candidates in developing transgenic crops to control insect pests. The specific objectives of my research were:

(1) To identify cDNAs that encode protease-inhibitor-like proteins.

(2) To determine whether the cDNA encoded proteins possess inhibitor activity to specific proteases.

(3) To determine the expression profile of different genes at different developmental stages and in different tissues.

(4) To determine the tissue localization of protease inhibitor like proteins.

(5) To determine how the genes are organized in the Hessian fly genome.

Successful completion of the above objectives will provide important information on composition, specificity, function and genomic organization and expression regulation of protease inhibitor in the Hessian fly. These results will provide a foundation for future studies to understand the proteolysis and its regulation in this insect.

3. Materials and methods

3.1 Hessian flies

Hessian flies originally collected from Ellis county (Gagane and Hatchett, 1989) were maintained on susceptible wheat seedlings ('Newton' or 'Karl92') in an environmental chamber at 20°C and 12:12 (L:D) photoperiod. The biotype colony that was maintained consisted of the Great Plains.

3.2 cDNA library construction

The cDNA library was constructed as described previously using a 'SMART™' cDNA library construction kit from Clontech (Palo Alto, CA, USA) (Zhu et al., 2004). Briefly, three hundred midguts or two hundred pairs of salivary glands were obtained by dissecting 3-day-old larvae and immediately placed them into TRI reagent™ (Molecular Research Center, Inc.). Total RNA was then isolated from the respective tissues. The cDNA libraries were constructed following the protocol provided by Clontech (Palo Alto, CA, USA) with one modification. The PCR fragments were cloned directly into a plasmid using a TOPO TA cloning kit (Invitrogen, Carlsbad, CA, USA). Individual cDNA clones were selected and arrayed into 96 well plates. Plasmid DNA was isolated from the plates using a Qiagen BioRobot 3000 and sequenced in an ABI 3700 DNA analyzer.

3.3 Sequence analysis

Open-reading-frame (ORF) and sequence-similarity analysis were performed using the National Center for Biotechnology Information (Bethesda, MD) website (http://www.ncbi.nlm.nih.gov). The pI prediction of mature proteins was performed using 'Compute pI/Mw tool' (http://us.expasy.org/tools/pi_tool.html). The amino acid sequences of proteins were deduced from the cDNA through ExPASY proteomics tools available on the website of the Swiss Institute of Bioinformatics (http://www.expasy.ch/tools/). The BCM search launcher program (http://searchlauncher.bcm.tmc.edu/) was used to conduct ClustalW multiple-sequence alignment. The shading of conserved residues amino acid was boxed using the program available at (http://www.ch.embnet.org/software/BOX_form.html).

3.4 Phylogenetic analysis

The cDNA encoded protein sequences were used to construct the phylogenetic tree with the software package GeneBee, available at (http://www.genebee.msu.su/services/phtree_reduced.html). The boot strap values were obtained using the software package Phylipv4.0. The boot strap values obtained from the Phylip program were placed at respective branch points of the phylogenetic tree obtained from GeneBee.

3.5 BAC library construction

A BAC DNA library was constructed in our lab in collaboration with Amplicon, (Pullman, WA), a commercial research company. The inserts were ligated into the Hind

III site of pECBA1 vector with average size around 150 kb. The BAC clones were spotted on nylon membranes. Each membrane contained 384 grids and each grid was spotted with 8 BAC clones in duplicates. Totally each filter was spotted with 6144 BAC clones.

3.6 Probe preparation

To generate cDNA probes, cDNA clones were inoculated into 5ml of terrific broth containing 50μg/ml of ampicillin and were grown overnight at 37°C. Plasmid DNA was isolated from the culture according to Maniatis et al. (1982). The insert from individual clones were amplified through PCR amplification. PCR reactions were carried out in buffer containing 50ng of plasmid DNA, 0.25 mM dNTP's, 2.5 mM MgCl2, 10 pm of each forward (5PCR 5'-AAGCAGTGGTATCAACGCAGAGT-3') and reverse (CDCIII/3PCR 5'-ATTCTAGAGGCCGAGGCGGCCGA-3'), 5 units of Taq Polymerase, 2μl of 10x PCR buffer) with 40 cycles of 94°C for 1 min, 55°C for 2 min, 72°C for 2 min. The PCR products were separated on 1.2% agarose gel. DNA bands with appropriate size were excised from the agarose gel and purified using Geneclean Turbo PCR purification kit from Q Biogene Inc., (Carlsbad, CA).

Purified DNA fragments were radio-labeled using Prime-It® II Random primer labeling kit from Stratagene, (La Jolla, CA). Specifically 50ng of cDNA, 5μl of random 9-mer primers (27OD units/ml), in 35μl of water were boiled for 5 min. The boiled solution was put on ice for 2 min. To each tube 10μl of 5x dCTP buffer (contained 0.1mM of each dCTP, dATP, dTTP, dGTP), 5μl of ^{32}P (1mC) and 2μl of Klenow enzyme

(2 units/μl) was added and kept at 37°C for 1 hour. Unincorporated radioactive isotope was removed from the solution by passing through NucTrap® probe purification columns from Stratagene, (La Jolla, CA). The purified probes were used for hybridization.

3.7 BAC library screening

BAC membranes were washed in 500 ml of solution containing 50 mM Tris, pH 8.0, 1 M NaCl, 1 mM EDTA, and 0.1% SDS for one hour at room temperature for three times. The membranes were then prehybridized in a polythene bag containing 15ml of prehybridization buffer (10% dextran sulfate/1% SDS/ 1M NaCl, pH8.0) at 42°C for one hour. Hybridization was carried out overnight at 42°C in the same plastic bag by adding the labeled probe. The membranes were then washed twice with 2x SSC at room temperature for 30 min, twice with 2x SSC plus 1% SDS at 65 °C for 30 min, and twice with 0.1× SSC plus 1% SDS at room temperature for 30 min. After washing, the membranes were exposed to Kodak SR-5 X-ray film for 1 hour to overnight depending upon the signal. The X- ray films were developed by transferring the film first to GBX developer and then to fixer, obtained from Kodak Company, Rochenster, NY, for 1 and 4 min, respectively. Positive colonies were identified according to the positions of the arrays.

3.8 BAC DNA isolation and Southern blot

BAC DNA from positive clones was isolated as described by Maniatis et al., (1982) with a modification. Briefly, individual BAC clones were grown in 5 ml Terrific Broth (TB) medium containing 4x (15µg/ml) of chloromphenicol for 2 hours at 37°C. To increase the yield of BAC DNA, another 15µg/ml of chloramphenicol was added to the culture after two hours incubation. The culture was allowed to grow overnight and BAC DNA was then isolated.

For Southern blot analysis, BAC DNA was completely digested with EcoR I, BamH I, and Hind III (~5-10µg of BAC DNA, 1x buffer, 30 units of enzyme and 0.25mg of BSA) at 37°C overnight. DNA fragments were separated on 1% agarose gel and were transferred onto the GeneScreen Plus® hybridization membrane obtained from PerkinElmer Life Sciences, Inc. (Boston, MA). The membrane was baked at 80°C for 2 hours for cross linking. The prehybridization, hybridization, washing, exposing and developing conditions were same as mentioned in BAC library screening. The membranes were hybridized with 11A6 and 14A4 cDNA probes, respectively.

3.9 RNA extraction and Northern analyses

Total RNA was extracted from the different stages of Hessian fly larvae (0 day, 2 day, 4 day, 6 day, 12 day larvae, pupa and adult) using TRI reagent^{TM} following the protocol from Molecular Research Center, Inc., (Cincinnati, OH) with few changes. 100mg of larval tissue were homogenized in 1ml of TRI reagent using a tissue homogenizer. The homogenate was stored for 5 minutes at room temperature for dissociation of nucleoprotein complexes. To the homogenate, 02.ml of chloroform was

added, mixed and kept at room temperature for 10 minutes. The homogenate was centrifuged at 10,000xg for 10 minutes. The supernatant (0.5 ml) was transferred to a new eppendorf tube and the RNA was precipitated using an equal amount of isopropanol. The samples were stored at room temperature for 10 minutes and centrifuged at 10,000xg for 10 min at 4°C. The RNA pellet was washed with 1ml 75% DEPC treated ethanol and centrifuged at 10,000xg for 5min at 4°C. The 75% ethanol was removed and the pellet was air dried. The RNA pellet was dissolved in an appropriate amount of DEPC treated water. 5μg of total RNA was separated on 1.2% agarose gel containing formaldehyde and MOPS buffer. The RNA separated on the gel was blotted on the GeneScreen Plus® membrane. The RNA was cross linked onto the membrane by baking the membrane at 80°C for two hours. The prehybridization, hybridization, washing, exposing and developing conditions were same as mentioned in Southern analysis.

3.10 Real-Time PCR

The first instar Hessian fly larvae were dissected in 1x PBS saline. The salivary glands, midgut, malpighian tubules (Fig. 1), fatbody, and carcass were collected and were immediately put into TRI reagent™ (Molecular Research Center, Inc.) separately. Total RNA was isolated from all the tissues as well as from whole larvae. The cDNA was obtained from the total RNA through reverse transcription. For reverse transcription, 1 μg of total RNA, 0.1 nM (0.1 nm/μl) poly (T) was heat denatured at 72°C for 2 minutes in 15 μl of DEPC water. Then the solution was cooled on ice for 1 minute. After cooling, 3μl of first strand buffer, 1μl of 10mM dNTP and 1μl of superscript reverse transcriptase (from BD Biosciences), was added to the solution. After cDNA synthesis, the RNA in the

reaction mixture was removed by adding 1μl of 10units/μl DNase free RNase (Promega Corporation) and keeping the tubes at 37°C for 1 hr.

The cDNA mixture from all the tissues was diluted to get 1.25ng/μl. The diluted samples were used for normalization of RPS30 transcripts using a pair of primers (Table 1) for the RPS30 gene. The normalized samples were used for transcript quantification. For the real-time PCR analysis, 5μl of normalized transcript quantification was used as template. The PCR products were labeled with SYBR green dye, (Biorad, Hercules), during amplification. The real-time amplification protocol to amplify 11A6, Lg2A3, and 11C4 consisted of 95°C for 5 minutes, then 45 cycles of 95°C for 20 sec, 62°C for 45 sec, followed by 95°C for 1 min and 55°C for 10 sec for 80 cycles. The annealing and extension for 14A4 and LG2F7 were 63°C and 64.5^{o}C, respectively.

To make standard curves of different cDNAs, the recombinant pCR$^®$ 2.1-TOPO plasmid containing different cDNAs were isolated using Qiagen isolation kit (Qiagen Inc., Valencia, CA). The concentration was measured using Nanodrop. The plasmids were diluted and real-time PCR reaction was conducted with the DNA concentration ranging from 1 ng to 1 fg to derive the standard curve with the above PCR conditions. The copy number of plasmids ranged from 211303557 (1ng) to 211.303557 (1fg) for 11A6, 206333957.3 (1ng) to 206.3339573(1fg) for 14A4, 206848505.8 (1ng) to 206.8485058 (1fg) for Lg2A3, 204392643.2 (1ng) to 204.3926432 (1fg) for Lg2F7 and 207837986.1 (1ng) to 207.8379 for 11C4.

Ct value is the cycle number at which the fluorescence crosses the threshold line (default set for different genes) which is available during the reaction. The Ct values for each gene with different PCR DNA concentrations were used to establish a standard

curve against the copy number in log scale. The regression equation obtained from the curve for different cDNAs by plotting a regression line was used to determine the log copy number of different cDNAs in different tissues. The log copy number was converted to exponential numbers and plotted on the graph.

The real time PCR primers were designed using the software, Beacon designer version 2.0 from Biosoft International, Palo Alto, CA. The primer pairs specific to each gene (Table. primer sequences) were synthesized by IDT Inc., Caraville, IA and Invitrogen, Carlsbad, CA. The specificity of the primer pairs to each gene was tested experimentally using genomic DNA and cDNA from the reverse transcription. As a control, a primer pair of the ribosomal protein S30 was synthesized according to its cDNA sequence. The real-time PCR products from this ribosomal protein mRNA were used as an internal control for normalization of transcripts in different tissues.

3.11 Preparation of expression constructs

The vector used to construct the expression system was pCR® T7/NT-TOPO, from Invitrogen, Carlsbad, CA (Fig. 2). This vector contains the T7 promoter priming site, a region coding for polyhistidine (6xHis) and Xpress epitope, a enterokinase recognition site, cloning site, a T7 reverse priming site, and a T7 transcription termination region. It also contains f1 and pUC origin and ampicillin selectable marker genes. The length of 6x polyhistidine, Xpress epitope tag and enterokinase region together account for 35 amino acid residues.

The 11A6 and 14A4 open reading frame cDNA inserts were prepared through PCR (100ng DNA, 2.5mM of MgCl2, 0.2mM dNTP's, 10 pM of each forward and

reverse primer and 8 units of Taq polymerase) from the cDNA library using forward(D3) (5'-ATTGCGGCACCGTTGGCGGAA-3' and 5'-AATGAAGACAACAAACCAAA-3'), and reverse(U3) (5'-AATGAAACTATCGCCACATAA-3' and 5'-TGTTAAAGATTTGGATACAG-3') primers specific for the 11A6 and 14A4 cDNAs respectively. The prepared insert were ligated into the pCR® T7/NT-TOPO (Invitrogen, Carlsbad, CA, USA) expression vector containing the T7 promoter priming site, a region coding for polyhistidine (6xHis) and ampicillin and zeocin selectable marker genes. The ligation reaction consisted of 1μg of insert, 1μl of TOPO cloning mixture (Invitrogen, Carlsbad, CA, USA), 1μl salt solution (1.2M NaCl, 0.06M MgCl$_2$) and the reaction volume was made up to 6μl. The reaction was allowed for 20 minutes at room temperature.

Two microliter of ligation mixture was transformed on to TOPO-10 chemically competent cells supplied by Invitrogen, Carlsbad, CA, USA. Initially, the competent cells were thawed on ice for 5 minutes. The ligation mixture was mixed with competent cells and allowed to stand on ice for 30 minutes. Afterwards, the competent cells were heat shocked at 42°C for 30 seconds and immediately kept on ice for 2 minutes. The transformed was plated onto the ampicillin (50μg/ml) plates. The plates were incubated at 37°C overnight.

Positive clones from 11A6 and 14A4 were inoculated into 5 ml LB medium containing 1x ampicillin (50μg/ml), grown overnight, and plasmids were isolated as per the protocol described by Maniatis et al., (1982). The orientations of the ligated cDNAs inserts were determined through PCR using T7 primer (5'-TAATACGACTCACTATAGGG-3') and U3 upstream primer specific for the 11A6 and

14A4 cDNAs (100ng DNA, 2.5mM of $MgCl_2$, 0.2mM dNTP's, 10 pM of each forward and reverse primer and 8 units of Taq polymerase). The PCR products were separated on the gel and observed for the correct product size. The clones with appropriate product size and with right orientation were selected to transform into expression cells.

3.12 Protein expression

The expression constructs containing the 11A6 and 14A4 cDNA inserts were transformed into Rosetta Blue BL21-DE3 competent cells, Novagen®. A single colony of each 11A6 and 14A4 cDNA expression clones was inoculated into 5 ml LB medium containing 1x ampicillin (50μg/ml) and 1x chloromphenicol (34μg/ml) and allowed to grow over night at 37^oC. To a fresh 5 ml of LB, 50 μl of the overnight bacterial culture was inoculated and grown at 37^oC for 3 hours. The 5ml overnight grown culture was divided equally into two tubes. For one tube 10mM IPTG was added, and the tubes containing the culture with and without IPTG were grown for 3 hrs at 37^oC. The culture was centrifuged at 5000 xg for 10 minutes at 4^oC. The pellet was vortexed with 135 μl of TE buffer (10mM Tris, 1mM EDTA). Fifteen μl of 10x SDS sample buffer (10% glycerol, 2% SDS, 0.1% bromophenol blue, 0.5M Tris pH 6.8) was added, mixed and the samples were boiled for 10 minutes. The samples were stored on ice for 2 minutes and centrifuged at 8,000xg. The soluble fraction (supernatant) was separated on a 13% SDS-PAGE.

For large scale protein expression, 50 ml of overnight culture of each 11A6 and 14A4 was inoculated into 500 ml of LB medium separately and allowed to grow for 3 hrs

at 37°C. The culture was grown for 4 hrs at 37°C by adding 10 mM IPTG to induce protein.

3.13 Protein purification

Native protein of 11A6 and 14A4 were purified from the IPTG induced culture using the ProBond™ Purification System (Invitrogen, Carlsbad, CA). Bacterial culture (250 ml) pellet of cDNA clones 11A6 and 14A4 were resuspended separately in 25ml of native purification buffer, pH 8.0 (25mM Tris, 0.25M NaCl) with 0.1% triton X-100. Lysozyme (8mg/50ml of culture pellet) was added to the suspension and stored on ice for 1 hour. The solution was sonicated on ice for 5 minutes with 10 second bursts. The lysate was centrifuged at 8000xg for 15 min to pellet the debris. The supernatant was collected and used for purification of protein.

The nickel resin was packed into the column by pipetting 2 ml of resin into a 10ml purification column and allowed to settle by gravity. The supernatant was removed from the column, and the resin was washed by resuspending in 8 ml of sterile water by inverting the column. The resin was allowed to settle down by gravity. Again the supernatant was removed from the column by pipetting. After two washes with water, the resin was resuspended into 1x native binding buffer, pH 8.0 (25mM NaPO4, 0.25M NaCl) and allowed to settle. The supernatant was removed, and this step was repeated one more time. The column was used immediately to purify the His-tagged protein.

The native binding buffer was removed from the column, and 8ml of prepared protein lysate was applied onto the column. His-tagged fusion proteins were allowed to

66

interact and bind to the nickel resin for 1 hour at room temperature by gently inverting the column. The column material was allowed to settle by gravity, and the supernatant was removed. The column was washed twice with 1x native wash buffer, pH 8.0, and eluted with 6ml of native elution buffer. One ml fractions were collected and separated on the 13% SDS-PAGE. Fractions containing the protein were pooled. The pooled fractions were filtered in a macrosep® 3K centifugal devices supplied by PALL Life Sciences (Ann Arbor, Michigan) to concentrate and remove salts.

3.14 N-terminal His-tag cleavage

To remove the N-terminal His-tag from the fusion protein, in a reaction volume of 150μl, 4.2μg of 11A6 and 14A4 recombinant fusion proteins were taken separately. Fifteen μl of enterokinase buffer and 15 units of enterokinasemax^{TM}, Invitrogen (Carsbald, CA), were added and incubated at 37°C overnight. The mixture was separated on 13% SDS-PAGE. The reaction was scaled up and the mixture was passed through a nickel resin column to obtain mature protein.

3.15 Protein determination

Protein concentration was determined by the bicinchoninic acid method (Smith et al., 1985).

3.16 Inhibition assay

3.15.1 Assay with commercial enzymes

Five grams of first instar *Mayetiola destructor* larvae were collected and ground in 250μl of extraction buffer (50 mM Tris-HCl, pH 8.0, 0.1% TrtionX-100). The extract was centrifuged three times at 10,000xg, 15 minutes each at 4°C to remove the debris. The protease extract was aliquoted into eppendorf tubes and frozen at -20°C. Commercial enzymes trypsin, chymotrypsin and papain were purchased from Worthington Biochemical Corporation, Lakewood, NJ. The enzymes trypsin and chymotrypsin were dissolved in 0.001N HCl at 1mg/ml and 10mg/ml, respectively. Papain was dissolved in cysteine HCl buffer, (50 mM cysteine-HCl pH 6.2, 10 mM EDTA and 60 mM mercaptoethanol).

Inhibition assay were conducted by measuring the hydrolysis of casein conjugated to a fluorescent probe, BODIPY-TR-X (Molecular Probes, Eugene, OR) according to a method previously described (Oppert et al., 1997). The starting concentration of recombinant 11A6 and 14A4 inhibitors were 1.3μM and 2.56μM, respectively. The assays were carried out in 100μl. In the assay for trypsin, chymotrypsin and Hessian fly larval extract, a universal buffer, pH 7.8 was used. For papain, yellow mealworm and red flour beetle gut extracts, a universal buffer, pH 6.2 was used. Inhibitors were preincubated with trypsin, chymotrypsin, papain, and with gut extracts of Hessian fly, yellow mealworm and red flour beetle (∽0.5 gut equivalents) for 10 min at 37°C before the addition of substrate. At time 0, 10 μl of a stock solution of 10 μg/ml of fluorescent-labeled casein, BODIPY-TR-X, was added to each well to initiate the reaction. The plates

were incubated at 37°C for 4 hr and inhibitory activity was measured using a fluorescence microplate reader (Fluoroskan Ascent FL, Labsystems, Thermo Electron Corp., Milford, MA), with an excitation coefficient of 584 nm and emission of 620 nm. The percentage inhibition was calculated by dividing the value obtained from incubations with inhibitor by control incubations without inhibitor, and multiplied by 100. IC_{50} values were the amount of the inhibitor concentration resulting in 50% inhibition and were calculated using linear regression. IC_{50} of 11A6 and 14A4 towards trypsin and chymotrypsin were calculated. The residual enzyme activity was plotted against the inhibitor to enzyme molar ratios.

3.17 Protein sample preparation for Western analysis

The first instar Hessian fly larvae were dissected in 1x PBS, pH 7.4, (137mM NaCl, 2.7mM KCl, 8mM Na_2HPO_4, 2mM KH_2PO_4). Salivary gland pairs (125) and malpighian tubules (125) were collected separately in 50 μl extraction buffer (50mM Tris-HCl buffer, pH 8.0). The tissue was homogenized through pipetting action. The extract was centrifuged at 8000 xg and the supernatant was collected.

To prepare midgut and midgut extract, intact 100 midguts were collected from the dissected insects, placed on a slide and punctured with a needle. The released midgut contents, called midgut extract, were collected and subjected to electrophoresis. The midgut without gut contents were collected in 200 μl of extraction buffer and the total proteins were extracted as mentioned above.

Hemolymph was prepared from 6 day old insects. The insect cuticle was pricked with a needle and the hemolymph was collected in 1:1 anticoagulant solution (40mM

sodium citrate buffer, pH4.5, 150mM NaCl, 15mM EDTA and 2mM phenylmethylsulfonyl fluoride. The contents were centrifuged at 8000xg and supernatant was collected.

3.18 Antibody preparation

Antibody against the 14A4 peptide, (KDYYDHSSQNECKLPATFAK), derived from 14A4 was produced commercially by Bethyl Laboratories Inc., (Montgomery, TX). The 14A4 peptide was conjugated to KLH carrier and injected into rabbits. The immunosorbents were prepared by conjugating the peptide with agarose. Antibodies were affinity purified by passing hyperimmune serum from rabbits against immunosorbents.

3.19 Western blot

About 2 μg of protein from salivary gland, malpighian tubule, midgut and midgut extract along with purified recombinant protein 14A4 were separated on 13% SDS-PAGE (Laemmli, 1970). The protein samples were then blotted on to the Hybond™-P filter membrane (Amersham Pharmacia Biotech). The membrane was blocked with 5% milk in 1x PBS containing 0.1% TritonX-100 (PBST) for 1 hour at room temperature. The membrane was washed with 1x PBST 3 times for 15 minutes each at room temperature. Then the membrane was incubated with 2.5% milk containing 14A4 primary antibody (1: 1000) for 1 hour at room temperature. The membrane was washed with 1x PBST with several changes of buffer up to 4 hours. The blot was then incubated in 2.5% milk containing an anti rabbit secondary antibody (1:1000) for 1 hour at room temperature. The membrane was washed with 1x PBST up to 4 hrs with several changes

of buffer. The membrane was placed on filter paper and activated by placing it in a substrate solution containing equal volume of ECL solution A and B, (Amersham Pharmacia Biotech Inc.) for 5 minutes. The excess solution was removed and the membrane was covered with saran wrap. The filter was exposed to Kodak SR-5 X-ray film for 10 min and the film was developed as described previously.

3.20 Immunohistochemistry

Salivary glands, midgut, and malpighian tubules were collected by dissecting first instar Hessian fly larvae in 1x PBS. The tissues were immediately fixed in 4% paraformaldehyde overnight at 4°C. The tissues were washed six times, 15 minutes each with 1x PBS containing 1% TritonX-100 (1x PBST). Then the tissues were incubated overnight at 4°C in 1:1000 primary antibody diluted in 1x PBST. The tissues were washed eight times, 15 minutes each with 1x PBST. Then the tissues were blocked overnight at 4°C with 5% goat serum. The tissues were then washed with 1x PBST six times, 15 minutes each. The tissues were then incubated with 1:500 secondary antibody in 1x PBST overnight at 4°C and the samples were covered with aluminum foil to avoid light. The samples were removed and washed eight times in 1x PBST. Tissues were nuclei stained with DAPI (4'-6-Diamidino-2-phenylindole), 1:1000 in 1x PBST and washed four times. Tissues were mounted on a glass slide using mounting medium and the tissues were excited at 550nm under a confocal microscope (Leica Microsystems).

Figure 1: Alimentary canal of Hessian fly larva showing salivary gland (SG), foregut (FG), midgut (MG), hindgut (HG) and malpighian tubule (MT). Figure adopted from Grover et al., (1988).
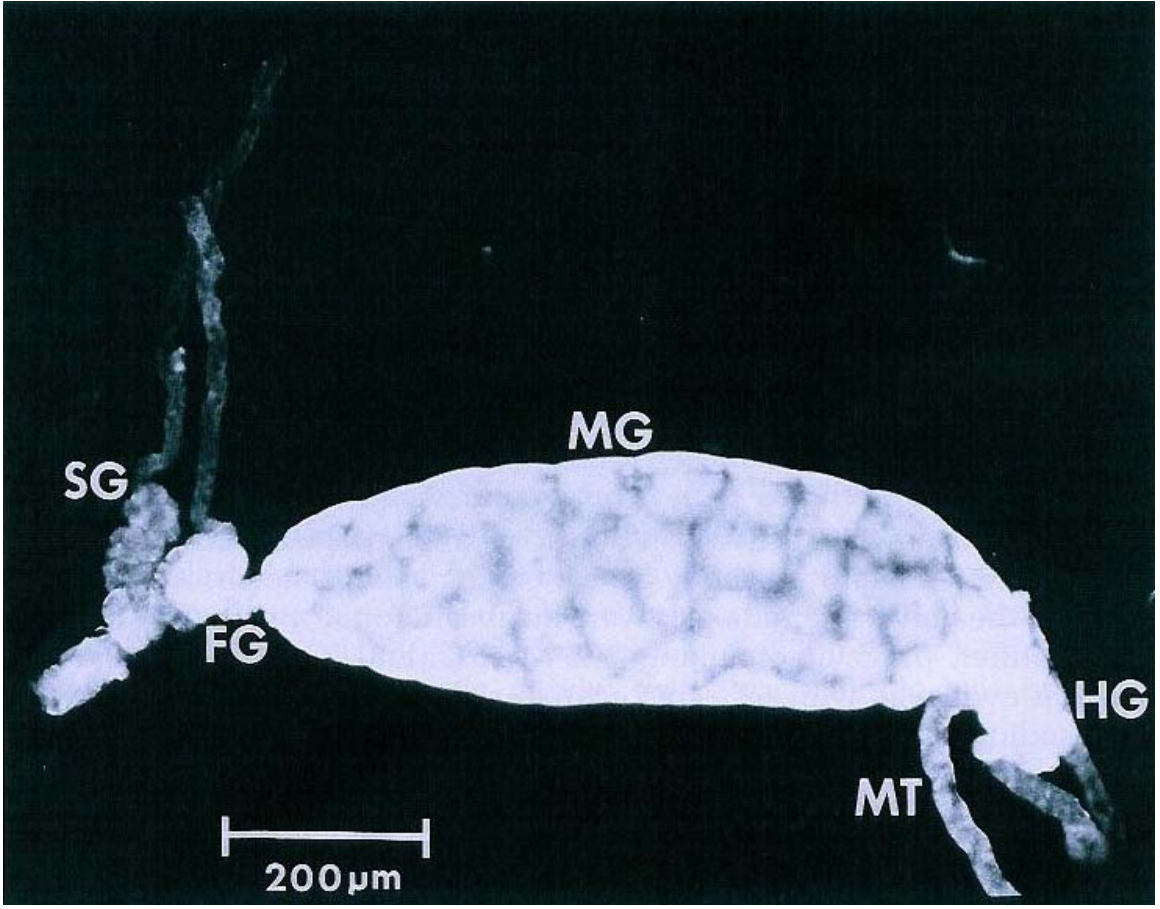
Figure 2: Expression construct used to clone the 11A6 and 14A4 cDNA inserts. The construct contains pUC origin and f1 ori sites, T7 polymerase priming site, RBS-ribosomal binding site, ATG-start codon, 6xHis-region coding for 6 histidine residues, region coding for Xpress epitope, EK-region coding for enterokinase cleavage site, multiple cloning site and T7 termination region and ampicillian marker gene.
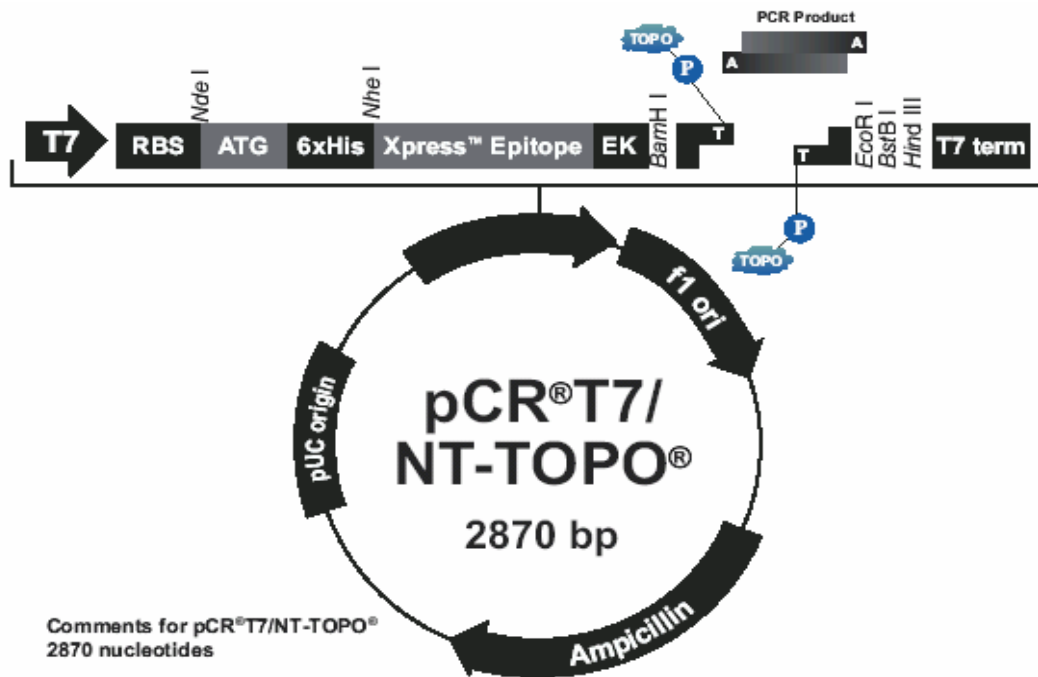
http://www.invitrogen.com/content/sfs/manuals/pcrt7topo_man.pdf

Table 1: Primer sequence of different protease inhibitor genes used in real-time PCR.

Clone Name	Primer Type	Sequence	Product Size
11A6	F	5'-GCTGCGTTGACAGTATTTTGG-3'	146
	R	5'-TCGCTACCGTCACAGTATTTG-3'	
14A4	F	5'-TACGAGGACTATTGCTGGACTC-3'	100
	R	5'-CAGGGCGACTCCAATTATAACC-3'	
Lg2A3	F	5'-TCTGCTCGTCTTCGTTACATC-3'	125
	R	5'-TTGCACAAGTGTCTCTAGCG-3'	
Lg2F7	F	5'-GACTATTGCTGGACTCTGAACG-3'	110
	R	5'-ACCGAGCAGAGCGATTCC-3'	
11C4	F	5'-CACCGAAAGACTGTGGACC-3'	110
	R	5'-GATTTGACGTTGATCTGGACAC-3'	
RPS30	F	5'-ACAAAATTGCCACTTTGGAAGG-3'	105
	R	5'-TCAAAGCTGCTCAATTCGTTAG-3'	

4. Results

A Hessian fly database containing more than 8000 ESTs had been established previously. Two thousand ESTs were obtained by random sequencing a cDNA library made from dissected guts (Zhu et. al., 2005). The other six thousand ESTs were derived from cDNA libraries made from dissected salivary glands (Chen et. al., 2003). Analysis of the EST database using GAP3 and other analytic tools revealed that there are several groups of genes that encode proteins with structural similarities to protease inhibitors. The sequences of the putative proteins among different groups are highly diversified. However, the number and positions of cysteine residues in the proteins are highly conserved. The purpose of this research was to determine the function of these diversified proteins and the genomic organization and expression regulation of the corresponding genes.

4.1 cDNAs encoding diversified proteins with conserved cysteine residues

Full length cDNA that encode proteins with conserved cysteine residues according to EST prediction were sequenced and characterized. Phylogenetic analysis of 29 unique proteins encoded by the cDNAs sorted the proteins into three lineages and five groups (Fig. 1). Lineage I and III each represent an independent group. Lineage II was further separated into three sublineages, each of which represents a different group. Sequence analysis revealed that members from the same group share greater than 85% sequence identity with E-value $< 10^{-50}$ while members from different groups share less than 35% identity with E-value >100.

4.1.1 Group I. The complete nucleotide and predicted amino acid sequences of a representative cDNA clone (11A6) from Group I are shown in the Fig. 2A. The cDNA contains 418 nucelotides and encodes a protein of 101 amino acids. The first 23 amino acids were predicted to consist of secretion signal peptides. This cDNA encodes a mature protein with Mr 9 kDa and pI 5.63.

Sequence alignment of the putative proteins deduced from ten cDNAs in this group is shown in Fig. 2B. The ten cDNAs encode very similar proteins with only a few amino acid substitutions. All the members contained 10 cysteine residues except 11A6 where only 9 cysteine residues were present. The cysteine residue at position 46 of 11A6 was substituted by an arginine residue. The cysteine residues were located at positions Cys^{36}, Cys^{46}, Cys^{50}, Cys^{55}, Cys^{64}, Cys^{68}, Cys^{74}, Cys^{76}, Cys^{88}, and Cys^{94}. The position of cysteine residues are completely conserved among different members of this group. The P_1 reactive site is at the 52nd position. The putative P_1 site residue for group 1 proteins is aspartic acid. Nucleotide sequence alignment of the ten cDNAs is shown in Fig. 2C. All the cDNAs contain a 5' inframe stop codon, start codon, stop codon and poly (A) signal at 3' end. The nucleotide identity between members varied from 95%-97% (data not shown).

4.1.2 Group II. The nucleotide and predicted amino acid sequences of a representative cDNA clone (14A4) are shown in Fig. 3A. The cDNA contains 500 nucleotides and encodes a mature protein with Mr 9.57 kDa and pI 7.67.

There are nine members in this group. The amino acid alignment for these members is shown in Fig. 3B. The predicted signal peptide is 21 amino acids except S17B2, which was predicted to have 27 amino acids for the secretion signal peptide.

Most of the members are very similar with only a few amino acid substitutions, except S17B2 and Sg6G2, which have substantial amino acid substitutions, an insertion/deletion, and different endings. The number and position of cysteine residues are completely conserved in all members. The cysteine residues are located at positions Cys^{36}, Cys^{37}, Cys^{50}, Cys^{55}, Cys^{67}, Cys^{79}, Cys^{81}, Cys^{93}, and Cys^{99}. P_1 reactive site residue is lysine at the 52nd position except in Sg6G2, in which P_1 residue is an isoleucine. Figure 3C shows the nucleotide sequence alignment of nine cDNAs of this group. More nucleotide variations are observed in cDNA clones S17B2 and Sg6G2. The cDNAs contain a 5' inframe stop codon, start codon, stop codon, and poly (A) signal at 3' end. The nucleotide identity between members ranged between 83%-97% (data not shown).

4.1.3 Group III. The nucleotide and deduced amino acid sequences of a representative member (LG2A3) in group III is shown in Fig. 4A. The cDNA contains 423 nucleotides and encodes a protein with 96 amino acids. The predicted signal peptide is 21 amino acids. Therefore, the cDNA can encodes a mature protein of 75 amino acids with predicted M_r 8.47 kDa and pI 4.80.

There are seven members in this group. Alignment of the amino acid sequences of the members is shown in Fig. 4B. The predicted signal peptide contains 21 residues, except for clone Lg2G5, which has a predicted signal peptide of 24 residues. The increased length in the signal peptide of Lg2G5 is due to an insertion/deletion in this protein. Members in this group are very similar with only a few amino acid substitutions and a small insertion/deletion in one member. All the members contain 8 cysteine residues, except Pg10A8, which contains an arginine instead of cysteine at position 73. The position of the cysteine residues are at Cys^{33}, Cys^{46}, Cys^{51}, Cys^{63}, Cys^{73}, Cys^{75},

Cys^{87}, and Cys^{93}. The predicted P$_1$ site residue is lysine at the 48th position in the sequence alignment. The nucleotide sequence alignment is shown in Fig. 4C. The Sg6C8 contains an addition segment in the cDNA. The identity between members varied from 94%-99% (data not shown).

4.1.4 Group IV. The nucleotide and predicted amino acid sequences of a representative clone (LG2F7) in this group is shown in Fig. 5A. The cDNA contains 535 nucleotides and encodes a protein with 103 amino acids. The first 21 amino acid residues are predicted to constitute a secretion signal peptide. Thus, this cDNA encodes a mature protein of 82 amino acids with a predicted M$_r$ 9.53 kDa and pI 4.64. There is another member that belongs to this group. The two members share 90% amino acid sequence identity (Fig. 5B). Both proteins contain 10 cysteines in the mature protein. In addition, there is an extra cysteine residue in LG2F7 in the signal peptide region. The cysteine residues are located at positions Cys^2, Cys^{23}, Cys^{32}, Cys^{33}, Cys^{46}, Cys^{51}, Cys^{63}, Cys^{75}, Cys^{77}, Cys^{89}, and Cys^{95}. Arginine is the predicted P$_1$ site residue located at the 53rd position of the protein. Figure 5C shows the nucleotide alignment between LG2F7 and Pg7F5. Both cDNAs contain addition and deletion of nucleotides at the 5' end. The nucleotide identity between the two members is 92% (data not shown).

4.1.5 Group V. There is only one member in group 5. The cDNA and deduced amino acid sequences are shown in Fig. 6. The cDNA contains 447 nucleotides and encodes a protein of 118 amino acids. The first 26 residues are predicted to constitute a secretion signal peptide. Therefore, this cDNA encodes a mature protein of 92 amino acids with a predicted M$_r$ 10.25 kDa and pI 6.78. The protein contains 9 cysteine residues

which are located at Cys^{32}, Cys^{40}, Cys^{54}, Cys^{59}, Cys^{66}, Cys^{82}, Cys^{84}, Cys^{96}, and Cys^{104}. The reactive site at P_1 site tyrosine is located at the 56th position.

4.2 Structural similarity between groups and with known protease inhibitors

The common feature in the protein structure among different groups is that all members contain similar numbers (7 to 10) of cysteine residues in the mature proteins. Pair-wise comparison between individual groups demonstrated that the positions for many of the cysteine residues are highly conserved (Fig. 7A and data not shown). Particularly, the positions for the last four residues are completely conserved among all groups (Fig. 7B). Despite the conservation in the number and position of residues, the overall sequence identity among groups is very low, with only 20 to 30% identity.

Interestingly, similar conservation patterns also exist between all these groups and several known protease inhibitors from the Ascaris trypsin inhibitor family (Fig. 7B). This observation indicates that the five groups from the Hessian fly may also be protease inhibitors. Indeed, database searching revealed that the number and position of cysteine residues are 100% conserved between some members from the five groups and known protease inhibitors or protease inhibitor-like proteins in the database (Fig. 7C).

4.3 Inhibitory activity of recombinant proteins towards serine proteases

Structural conservation between the identified proteins and known protease inhibitors suggested that the putative proteins might be protease inhibitors as well. To

determine if this is the case, two cDNAs from two different groups were selected to generate recombinant proteins for activity assays.

4.3.1 Generation of Recombinant Proteins. Individual expression constructs were made using the pCRT7NTTOPO® expression system. The constructs were introduced into *E. coli* strains, and high yield expression of recombinant protein was achieved under induced conditions. Figure 8 shows the total proteins from *E. coli* that contain 11A6 and 14A4 constructs under induced or uninduced conditions along with a positive control (8A3-1C). The protein profile is similar in uninduced and induced samples, except for the presence of one additional protein band in the induced cells. This extra band in each of the induced lane represents the recombinant protein 11A6 or 14A4. The molecular masses of the induced proteins are of the expected size.

4.3.2 Purification of Recombinant Proteins. To purify the recombinant proteins 11A6 and 14A4, bacteria containing the respective constructs were grown in a large volume of LB medium and were induced by adding IPTG. The cells were lysed, sonicated and centrifuged.

The protein extracts were used for purification using the ProBond™ purification system (Invitrogen, Carlsbad, CA) following the protocol provided by the manufacturer, with a few modifications as mentioned in the methods. Protein extracts containing fusion proteins were passed through the nickel column to purify 11A6 or 14A4 recombinant proteins separately. A single step purified the recombinant proteins to homogeneity. Mature proteins were obtained by removing the N-terminal fusion part using enterokinase enzyme. The recombinant proteins with and without the N-terminal His-tag are shown in

Fig. 9. The sizes of the purified recombinant proteins were as expected. Large amounts of mature proteins were obtained by scaling up.

4.3.3 Inhibitory Activity of Recombinant Protein 11A6. To determine if the recombinant proteins are indeed protease inhibitors, inhibition assays with the recombinant proteins were conducted against various proteases. As shown in Fig. 10, recombinant protein 11A6 inhibited the trypsin activity as much as 62% at a molar ratio 1.54:1. The IC_{50} (amount of inhibitor required to inhibit the 50% of enzyme activity) for trypsin is 0.52 μM (Table 1). Surprisingly, 11A6 also inhibited papain, (a cysteine protease) by ~30%. The 11A6 inhibitor was ineffective against chymotrypsin. The 11A6 protein inhibited the Hessian fly protease activity in gut extract by 35%. The protein also inhibited gut proteases of red flour beetle, *Tribolium castaneum*, and yellow meal worm, *Tenebrio molitor*, by 82% and 60%, respectively.

The impact of 11A6 protein on trypsin activity was examined. As shown in the Fig. 11 a gradual decrease in trypsin enzymatic activity was observed with the increase in inhibitor concentration within the molar ratio from 0:1 to 3.1:1. The trypsin activity decreased to 37.89% at a molar ratio of 1.5:1. After that further increase in inhibitor concentration had almost no impact on trypsin enzymatic activity.

The inhibition of papain by 11A6 protein was much less than that of trypsin. As shown in Fig. 12, the activity of papain activity was decreased to ~70% at a 1.5:1 molar ratio.

4.3.4 Inhibitory Activity of Recombinant Protein 14A4. As shown in Fig. 13, recombinant protein 14A4 inhibited trypsin activity as much as ~68% and chymotrypsin activity by 54% at a molar ration of 3.2:1. The IC_{50} values for trypsin and chymotrypsin are 0.9 and 2.32 μM, respectively (Table 1). 14A4 had no significant impact on the activity of papain. Recombinant protein 14A4 also inhibited the protease activity in gut extracts from Hessian fly, red flour beetle and yellowmeal worm. The 14A4 protein inhibited Hessian fly protease activity as much as 62%, red flour beetle 82%, and yellow meal worm by 86%.

Figure 14 shows the effect of recombinant 14A4 protein on activity of trypsin. The molar ratio of inhibitor to trypsin was increased gradually from 0:1 to 4.1:1. With an increase in ratio of inhibitor to enzyme, trypsin activity was decreased. Trypsin activity was 32.54% at a molar ratio of 3.1:1 (Table 1). No further increase in inhibition was observed beyond this molar ratio.

The effect of the 14A4 protein on the activity of chymotrypsin was studied (Fig. 15). The molar ratio of 14A4 to chymotrypsin increased gradually from 0:1 to 4.3:1. The chymotrypsin activity decreased gradually with an increase in inhibitor, and the activity was 45.75% at a molar ratio of 3.2:1 (Table 1). Little further increase in inhibition was observed beyond this molar ratio.

4.4 Developmental regulation of gene expression

Northern blot analysis was conducted to study the differential gene expression profile in different developmental stages of the insect. The RNA was extracted from different larval stages, pupa and adult of Hessian fly and blots were prepared as mentioned in methods. The blots were hybridized with their respective P^{32} labeled cDNA probes. 18S RNA was used as control to show equal loading. As shown in Fig. 16, all the protease inhibitor genes were exclusively expressed in the larval stages of the insect except group I and group V, which are also expressed in the pupal stage. The expression patterns varied widely among different inhibitor groups.

The group I genes are abundantly expressed from 4-12 days, weakly expressed in pupa. The highest level of RNA was detected in 6-day old larvae when the larva transists from first to second instar. No RNA could be detected in 0- to 2- day old larvae.

The expression of the genes in group II was detected in 0-day old larvae, but the expression level was significantly elevated in 2-day and 4-day old larvae. No RNA could be detected in 12 day old larvae, pupae and adults.

The expression profiles of group III and group IV genes were similar. Very low level expression of genes was detected in 2-day old larvae. The expression level gradually increased as the larvae became older and reached a maximum in 4-day old larvae. After that the expression level decreased and became undetectable in pupae and adults.

The expression of group V genes was observed only in later instar larvae and pupae. The RNA was found abundantly in 12-day larvae and pupae. No RNA was detected in 0-, 2-, and 4-day old larvae as well as in adult.

4.5 Tissue specific expression

To quantitatively determine transcript levels in different tissues of the larva, real-time PCR was conducted with primers specific to each of the five groups. The plasmids containing their respective cDNAs were used to generate Ct values. The Ct (it is the point at which the fluorescence crosses the threshold line) values of the amplification curves were used to plot against the log copy number to generate a standard curve and linear regression equation for each gene. The regression equation was used to determine the copy number of genes in different tissues.

Figure 17B, 18B and 20B shows the expression profile of group I, group II and group IV genes in different tissue types. The expression profiles for all three gene groups were similar in different tissues. All the gene groups were expressed in malpighian tubules with the highest copy number. Apart from malpighian tubules, they were also expressed in midgut and salivary glands but at a low level. The copy number of the respective genes in midgut and salivary glands were similar. A very low level of expression was observed in fatbody and carcass tissues.

The tissue specific expression profile of group III genes is shown in Fig. 19B. The genes were expressed in malpighian tubules at a high level, followed by midgut and salivary glands. The expression was again very low in fatbody and carcass.

A different expression profile was observed for group V genes. The gene group was predominantly expressed in salivary glands followed by fatbody and malpighian tubules (Fig. 21B). A low level of expression was detected in the midgut and carcass.

4.6 Immunolocalization of 14A4

Due to limitation of resources, only one antibody was generated. The antibody was produced against 20 amino acid peptides synthesized according to the 14A4 protein. The antibody was affinity purified against the same peptide. Insect tissues including malpighian tubules, salivary glands, and midgut were dissected from the first instar larvae and used for histochemical studies.

As shown in Fig. 22, the 14A4 protein was predominantly present in the malpighian tubule. Within this tissue, the protein was predominantly localized at the tip and in the middle segment of the tubule. Within these two regions, the protein was present in a few cells that are located at the tip whereas in the middle segment the protein was present in a greater number of cells. Also the protein was abundantly present in these cells. A low level of protein was detected in the midgut cell (Fig. 23). The figure shows the enlarged midgut cell. In the midgut the protein was distributed through out the cytoplasm as indicated by the presence of small specks of granules. Inside the cytoplasm, the protein was concentrated in some areas. No protein could be detected in the nucleus. Very low levels of protein were detected in the salivary gland (Fig. 24). The protein was detected in the cells located at the bottom of the basal region and also in the tubular cells.

Western analysis was conducted to confirm the localization of 14A4 protein in different tissues and to check the specificity of 14A4 antibody. Immunoblot data shows that the 14A4 was present in the hemolymph, midgut, and midgut extract in large amounts whereas in salivary glands lesser amounts of the inhibitor were found (Fig. 25). The presence of protein in low amounts in salivary glands was in agreement with the

localization data. The specificity of 14A4 antibody was confirmed with the presence of a bright 14A4 recombinant protein band.

4.7 Organization of genes

4.7.1 Identification of positive BAC clones. The variations in deduced amino acid sequence of cDNAs of a given group suggested that they might be members of a family of genes. To test this hypothesis, an Hessian fly BAC DNA library was analyzed through hybridization using P^{32} labeled 11A6 and 14A4 cDNAs as probes. Probe 11A6 identified six BAC clones which were named 2J21, 2M12, 7H20, 8L4, 8L18, and 8D19 (Figure 26). Similarly, probe 14A4 identified five positive BAC clones which were named 9O13, 10O2, 11O4, 14C5, and 15M14 (Fig. 27).

4.7.2 Southern Blot analysis of positive BAC clones. Multiple BAC clones positive to 11A6 and 14A4 probes were subjected to Southern analysis. For the 11A6 probe, a single band was observed in all three BACs that were cut with two different restriction enzymes (Fig. 28). The sizes of the DNA fragments that were positive to probe 11A6 were ~1 and ~2.5 kb for EcoRI and HindIII, respectively.

For the 14A4 probe, a different hybridization pattern was observed (Fig. 29). In this case, multiple bands with size ranging from 0.5 to 3 kb were detected in each BAC cut with different restriction enzymes. When BAC clones were digested with EcoRI, 3 to 4 positive bands were detected. The hybridization pattern was similar for each BAC. However, when the BAC clones were digested with HindIII, the hybridization bands were larger in size and pattern varied for different BACs.

Figure 1: Phylogenetic analysis of proteins encoded by the cDNAs. Boot strap values are indicated at the branching sites. Phylogenetic tree shows three lineages. One of the lineages (Lineage II) contains 3 sublineages. The three sublineages in lineage II together with lineage I and lineage III represent five different linkage groups, Group I - V.
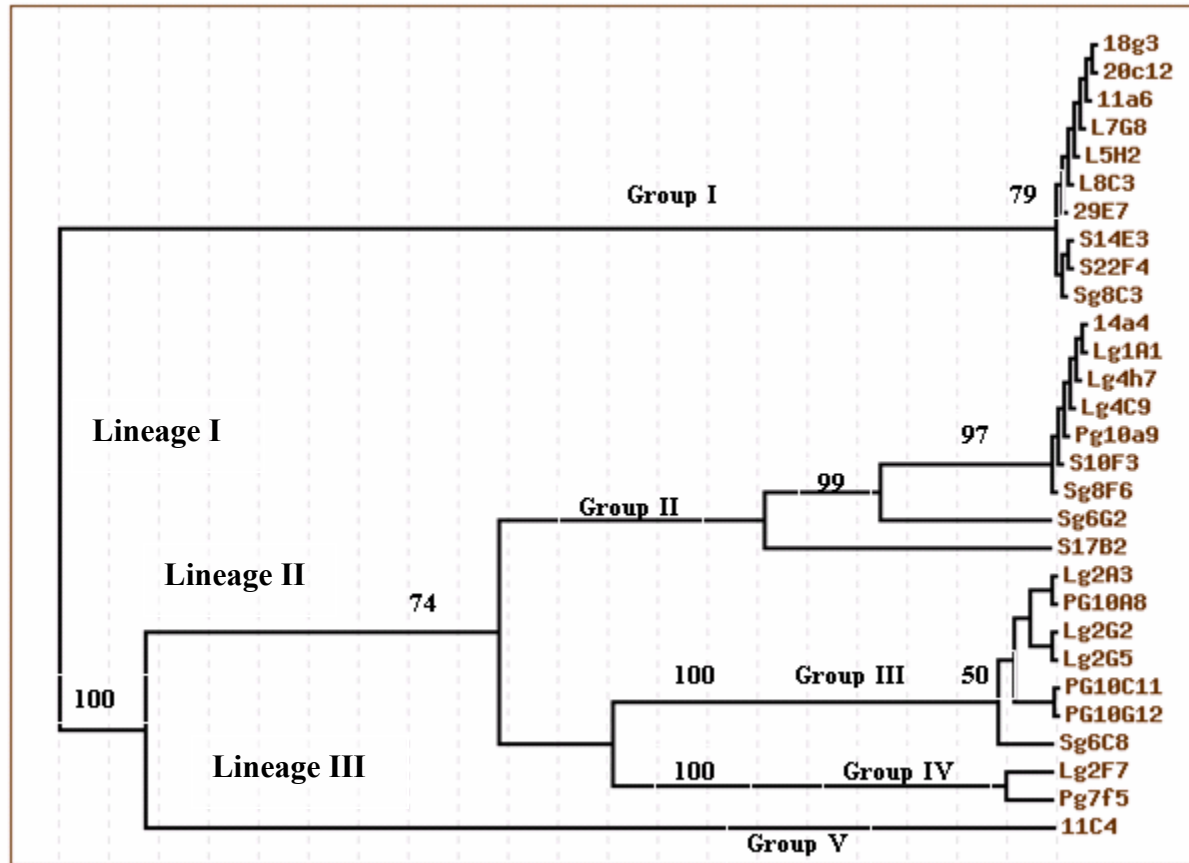
Figure 2A: Nucleotide and amino acid sequence of a representative cDNA from Group I. The amino acid is shown under the nucleotide sequence. Signal peptide is denoted by underline. * indicates the stop codon. Cysteine residues are in bold letters.

11A6

```
-52              ccggggaccgtctatagttttttcgaatagaaaatagaagcttctgtacgat
                                                 *

1    atggattcgaaaatgttcttattgatggtcgctgcgttgacagtattttggattacatca
     M  D  S  K  M  F  L  L  M  V  A  A  L  T  V  F  W  I  T  S

61   ggtattgcggcaccgttggcggaagatgagccagtcgatagaacatgtacaagactaaat
     G  I  A  A  P  L  A  E  D  E  P  V  D  R  T  C  T  R  L  N

121  gagttttacgcagaacgctcaagtcgttgtgatgacaaatactgtgacggtagcgaaata
     E  F  Y  A  E  R  S  S  R  C  D  D  K  Y  C  D  G  S  E  I

181  tatatgaagtgcactcaagattgtagtcatggatacagatgccgttgtataacaggaaca
     Y  M  K  C  T  Q  D  C  S  H  G  Y  R  C  R  C  I  T  G  T

241  aagagagattcccacggcaattgcatcgaagagaataaatgtaatgctacaaaattgtgg
     K  R  D  S  H  G  N  C  I  E  E  N  K  C  N  A  T  K  L  W

301  ctttgagtgtgttagaactatatttcggctggttatgtggcgatagtttcatttggtaga
     L  *

361  agttta
```

Figure 2B: Sequence alignment of putative proteins encoded by Group I cDNAs. Sequence variations among group members are bold and underlined. N-terminal signal peptide is bold and underlined. P_1 represent the putative amino acid residue at the reactive site center. Grey boxes represent the conserved cysteine residues.

```
                 1                                                    P₁       65
        11a6     MDSKMFLLMVAALTVFWITSGIAAPLAEDEPVDRTCTRLNEFYAERSSRCDDKYCDGSEI
        18g3     MDSKMFLLMVAALTVFWITSGIAAPLAEDEPVDRTCRRLNEFYVECSSRCDDKYCDGSEI
        20c12    MDSKMFLLMVAALTVFWITSGIAAPLAEDEPVDRTCRRLNEFYVECSSRCDDKYCDGSEI
        L7G8     MDSKMFLLMVAALTVFWITSGIAAPLAEDEPVDRTCRRLNEFYVECSSRCDDKYCDGSEI
        L5H2     MDSKMFSLMLAALTVFWITSGIAAPLAEDEPVDRTCRRLNEFYVECSSRCDDKYCDGSEI
        L8C3     MDSKMFLLMVAALTVFWITSGIAAPLADDEPVERTCRRLNEFYVECSSRCDDKYCDGSEI
        29E7     MDSKMFLLMVAALTVFWITSGIAAPLAEDEPVDRTCRRLNEFYVECSSRCDDKYCDGSEI
        S14E3    MDSKMFLLMVAALTVFWITSGIAAPLEEDEPVDRTCRRLNEFYVECSSRCDDKYCDGSEI
        S22F4    MDSKMFLLMVAALTVFWVTSGIAAPLEEDEPVDRTCRRLNEFYVECSGRCDDKYCDGSEI
        Sg8C3    MDSKMFLLMVAALTVFWITSGIAAPLEEDEPVDRTCRRLNEFYVECSNRCDDKYCDGSEI


                 66                                       106
        11a6     YMKCTQDCSHGYRCRCITGTKRDSHGNCIEENKCNATKLWL
        18g3     YMKCTQDCSHGYRCRCITGTKRDSHGNCIEENKCNATKLWL
        20c12    YMKCTQDCSHGYRCRCIPGTKRDSHGNCIEENKCNATKLWL
        L7G8     YMKCTQDCSHGYRCRCITGTKRDSHGNCIEENKCNATKLWL
        L5H2     YMKCTQDCSHGYRCRCITGTKRDSHGNCIEENKCNATKLWL
        L8C3     YMKCTQDCSHGYRCRCITGTKRDSHGNCIEENKCNATKLWL
        29E7     YMKCTQDCSHGYRCRCITGTKRDSHGNCIEENKCNATKLWL
        S14E3    YMKCTQDCSHGYRCRCKTGTKRDSRGNCIEENKCNTTKLWL
        S22F4    YMKCTQDCSHGYRCRCKTGTKRDSRGNCIEENKCNATKLWL
        Sg8C3    YMKCTQDCSHGYRCRCKTGTKRDSRGNCIEENKCNTTKLWL
```

Figure 2C: Nucleotide sequence alignment of Group I cDNAs. Sequence variations among group members are bold and colored. The 5' inframe stop codon, start codon, stop codon and poly (A) tail signal are bold and underlined.

```
        1                                                             60
11A6    attacggccggggaccgtctatagtttttttcgaatagaaaatagaagcttctgtacgata
18g3    ---------------gtctatagtttttttcgaatagaaaatagaagcttctgtacgata
20C12   ---------------gtctatagtttttttcgaatagaaaatagaagcttctgtacgata
L7G8    -------------cagtctatagtttttttcgaatagaaaatagaagcttctgtacgata
L5H2    -------------cagtctatagtttttttcgaatagaaaatagaagcttctgtacgata
L8C3    -------------cagtctatagtttttttcgaatagaaaatagaagcttctgtacgata
29E7    -------------cagtctatagtttttttcgaatagaaaatagaagcttctgtacgata
S14E3   -------------cagtctatagtttttttcgaatagaaaatagaagcttctgtacgata
S22F4   --------------agtctatagtttttttcgaatagaaaatagaagcttctgtacgata
Sg8C3   -------------cagtctatagtttttttcgaatagaaaatagaagcttctgtacgata

        61                                                            120
11A6    tggattcgaaaatgttcttattgatggtcgctgcgttgacagtatttttggattacatcag
18g3    tggattcgaaaatgttttttattgatggtcgctgcgttgactgtatttttggattacatcag
20C12   tggattcgaaaatgttttttattgatggtcgctgcgttgacagtatttttggattacatcag
L7G8    tggattcgaaaatgttttttactgatggtcgctgcgttgacagtatttttggattacatcag
L5H2    tggattcgaaaatgttttttcattgatgcttgctgcgttgacagtatttttggattacatcag
L8C3    tggattcgaaaatgttttttattgatggtcgctgcgttgacagtatttttggattacatcag
29E7    tggattcgaaaatgttttttattgatggtcgctgcgttgacagtatttttggattacatcag
S14E3   tggattcgaaaatgttttttattgatggttgctgcgttgacagtatttttggattacatcag
S22F4   tggattcgaaaatgttttttattgatggttgctgcgttgacagtatttttgggttacatcag
Sg8C3   tggattcgaaaatgttttttattgatggttgctgcgttgacagtatttttggattacatcag

        121                                                           180
11A6    gtattgcggcaccgttggcggaagatgagccagtcgatagaacatgtacaagactaaatg
18g3    gtattgcggcaccgttggcggaagatgagccagtcgatagaacatgtagaagactaaatg
20C12   gtattgcggcaccgttggcggaagatgagccagtcgatagaacatgtagaagactaaatg
L7G8    gtattgcggcaccgttggcggaagatgagccagtcgatagaacatgtagaagactaaatg
L5H2    gtattgcggcaccgttggcggaagatgagccagtcgatagaacatgtagaagactaaatg
L8C3    gtattgcggcaccgttggcggatgatgagccagtcgaaagaacatgtagaagactaaatg
29E7    gtattgcggcaccgttggcggaagatgagccagtcgatagaacatgtagaagactaaatg
S14E3   gtattgccgcaccgttggaggaagatgagccagtcgatagaacatgtagaagactaaatg
S22F4   gtattgccgcaccgttggaggaagatgagccagtcgatagaacatgtagaagactaaatg
Sg8C3   gtattgccgcaccgttggaggaagatgagccagtcgatagaacatgtagaagactaaatg

        181                                                           240
11A6    agttttacgcagaacgctcaagtcgttgtgatgacaaatactgtgacggtagcgaaatat
18g3    agttttacgtagaatgctcaagtcgttgtgatgacaaatactgtgacggtagcgaaatat
20C12   agttttacgtagaatgctcaagtcgttgtgatgacaaatactgtgacggtagcgaaatat
L7G8    agttttacgtagaatgctcaagtcgttgtgatgacaaatactgtgacggtagcgaaatat
L5H2    agttttacgtagaatgctcaagtcgttgtgatgacaaatactgtgacggtagcgaaatat
L8C3    agttttacgtagaatgctcaagtcgttgtgatgacaaatactgtgacggtagcgaaatat
29E7    agttttacgtagaatgctcaagtcgttgtgatgacaaatactgtgacggtagcgaaatat
S14E3   agttttacgtagaatgctcaagtcgttgtgatgacaaatactgtgacggtagcgaaatat
S22F4   agttttacgtagaatgctcaggtcgttgtgatgacaaatactgtgacggtagcgaaatat
Sg8C3   agttttacgtagaatgctcaaatcgttgtgatgacaaatactgtgacggtagcgaaatat
```

96

```
        241                                                              300
11A6    atatgaagtgcactcaagattgtagtcatggatacagatgccgttgtataacaggaacaa
18g3    atatgaagtgcacacaagattgtagtcatggatacagatgccgttgtataacaggaacaa
20C12   atatgaagtgcacacaagattgtagtcatggatacagatgccgttgtataccaggaacaa
L7G8    atatgaagtgcacacaagattgtagtcatggatacagatgccgttgtataacaggaacaa
L5H2    atatgaagtgcacacaagattgtagtcatggatacagatgccgttgtataacaggaacaa
L8C3    atatgaagtgcacacaagattgtagtcatggatacagatgccgttgtataacaggtacaa
29E7    atatgaagtgcacacaagattgtagtcatggatacagatgccgttgtataacaggaacaa
S14E3   atatgaagtgcacacaagattgtagtcatggatacagatgccgttgtaaacaggtacaa
S22F4   atatgaagtgcacacaagattgtagtcatggatacagatgccgttgtaaacaggtacaa
Sg8C3   atatgaagtgcacacaagattgtagtcatggatacagatgccgttgtaaacaggtacaa


        301                                                              360
11A6    agagagattcccacggcaattgcatcgaagagaataaatgtaatgctacaaaa-ttgtgg
18g3    agagagattcccacggcaattgcatcgaagagaataaatgtaatgctacxaaaattgtgg
20C12   agagagattcccacggcaattgcatcgaagagaataaatgtaatgctacaaaa-ttgtgg
L7G8    agagagattcccacggcaattgcatcgaagagaataaatgtaatgctacaaaa-ttgtgg
L5H2    agagagattcccacggcaattgcatcgaagagaataaatgtaatgctacaaaa-ttgtgg
L8C3    agagagattcccacggcaattgcatcgaagagaataaatgtaatgctacaaaa-ttgtgg
29E7    agagagattcccacggcaattgcatcgaagagaataaatgtaatgctacaaaa-ttgtgg
S14E3   agagagattcccgcggcaattgcatcgaagagaataaatgtaatactacaaaa-ttgtgg
S22F4   agagagattcccgcggcaattgcatcgaagagaataaatgtaatgctacaaaa-ttgtgg
Sg8C3   agagagattcccgcggcaattgcatcgaagagaataaatgtaatactacaaaa-ttgtgg


        361                                                              420
11A6    ctt**tga**gtgtgttagaactatatttcgg**c**tggttatgtggcgatagtttcatttggtag-
18g3    ctttgagtgtgttagaactatatttcggttggttatgtggcgatagtttcatttggtaa-
20C12   ctttgagtgtgttagaactatatttcggttggttatgtggcgatagtttcatt**c**ggtaa-
L7G8    ctttgagtgtgttagaactatatttcggttggttatgtggcgatagtttcatttggtaa-
L5H2    ctttgagtgtgttagaactatatttcggttggttatgtggcga-ggt**gg**ca---gg**a**aa**t**
L8C3    ctttgagtgtgttagaactatatttcggttggttatgtggcgatagtttcatttggtaa-
29E7    ctttgagtgtgttagaactatatttcggttggttatgtggcgatagtttcatttggtaa-
S14E3   ctttgagtgtgtt**g**gaactatatttcggttggttatgtggcgat**g**gtttcatttggt**g**a-
S22F4   ctttgagtgtgttagaactatatttcggttggttatgtggcgat**g**gtttcatttggt**g**a-
Sg8C3   ctttgagtgtgtt**g**gaactatatttcggttggttatgtggcgat**g**gtttcatttggt**g**a-


        421                                                              480
11A6    aa-gttt**a**--------------------------------------------------
18g3    aa-gttt----**aaataaa**gtttattttgg--------------gg----aatatgtc--
20C12   aa-gttt----aaataaagtttattttgg--------------gg----aatatgtc--
L7G8    aa-gttt----aaataaagtttattttgg--------------gg----aatatg**ct**--
L5H2    aa**tgt**a**tgaag**aaataa**t**gt**a**tatttttt**tttgaaataaaatat**ggt**ctg**aatata**tgt**a
L8C3    aa**a**gttt----aaataaagtttattttgg--------------gg**g**---aatat**accc**a
29E7    aa-gttt----aaataaagtttattttgg--------------gg----aatatgtc**ta**
S14E3   aa-gttt----aaataaagtttattttgg--------------gg----aatatg**gc**aa
S22F4   aa-gttt----aaataaagtttattttgg--------------gg----aatatgtc**c**a
Sg8C3   aa-gttt----aaataaagtttattttgg--------------gg----aatatgtc**a**a
```

Figure 3A: Nucleotide and amino acid sequence of a representative cDNA from Group II. The amino acid is shown under the nucleotide sequence. Signal peptide is denoted by underline. * indicates the stop codon. Cysteine residues are in bold letters.

14A4

-69   atatctacg

-60 aggactattgctggactc**tga**acgtgtttaatattcgaagaaaagaaaaataatcctaaa
                      *

1   **atg**tctttgaaattggttataattggagtcgccctgctcggcttcgttgcatcttcattg
    M S L K L V I I G V A L L G F V A S S L

61 ggtaatgaagacaacaaaccaaacttatgttgtcctcaaaacgaaaaactttatgactat
    G N E D N K P N L **C** **C** P Q N E K L Y D Y

121 cattatccgtgcaaaaaaaacatgtgcaaagactattacgaccattccagtcaaaacgaa
    H Y P **C** K K N M **C** K D Y Y D H S S Q N E

181 tgcaaacttccggccacttttgctaaaaacaaatattgcgattgtaaggatggttactta
    C K L P A T F A K N K Y **C** D **C** K D G Y L

241 cgcactaatgcaggtccttgtgttaagaaagaagattgcaatttcttgataaaaaatgag
    R T N A G P **C** V K K E D **C** N F L I K N E

301 catcat**tga**tttgaaatgaaacgcaactgataaggaattggcaatgataaaaaaactgta
    H H *****

361 tccaaatctttaacaattt**aaaaataaa**gcaaaataaaaattccattttcctggtgaagg

421 gaacaacttgc

Figure 3B: Sequence alignment of putative proteins encoded by Group II cDNAs. Sequence variations among group members are bold and underlined. N-terminal signal peptide is bold and underlined. P_1 represent the putative amino acid residue at the reactive site center. Grey boxes represent the conserved cysteine residues.

```
                  1                                                  P₁   55
14a4      MSLKLVIIGVALLGFVASSL------GNEDNKPNLCCPQNEKLYDYHYPCKKNMC
Lg1A1     MSLKLVIIGVALLGFVASSL------GNEDNKPNLCCPQNEKLYDYHYPCKKNMC
Lg4h7     MSLKLVIIGVALLGFVASSL------GYEDNKPNLCCPQNEKLYDYHYPCKKNMC
Lg4C9     MSLKLVIIGVALLGFVASSL------GYEDSKPNLCCPQNEKLYDYHYPCKKNMC
Pg10a9    MSLKLVIIGVALLGFVASSL------GYEDNKPNLCCPQNEKLYDYHYPCKKNMC
S10F3     MSLKLVIIGVALLGFVAFSL------GNEDTKPNLCCPQNEKLYDYHYPCKKNMC
Sg8F6     MCLKLVIIGVALLGFVASSL------GYEDNKPNLCCPQNEKLYDYHYPCKKNMC
S17B2     MCLKLVIIGVALLGFVASTFIPLPPSGNEDTKPNLCCPQNEKLYDYNYPCKKNMC
Sg6G2     MSLKLVIIGVALLGFVASSL------GNEDNKPNFCCPQNEKLYDYNYPCKITMC


          56                                                          108
14a4      KDYYDHSSQNECKLPATFAKNKYCDCKDGYLRTNAGPCVKKEDCNFLIKNEHH
Lg1A1     KDYYDHSSQNECKLPATFAKNKYCDCEDGYLRTNAGPCVKKEDCNFLIKNEHH
Lg4h7     KDLYDHSSQNECKLPATFAKNKYCDCMDGYLRTNAGPCVKKEDCNFLIKNEHH
Lg4C9     KDLYDHSSQNECKLPATFAKNKYCDCMDGYLRTNAGPCVKKEDCNFLIKNEHH
Pg10a9    KDLYDHSSQNECKLPATFAKNKYCDCMDGYLRTNAGPCVEKEDCNFLIKNEHH
S10F3     KDFYDHSSQTECKLPATFAKNKYCDCMDGYLRTNAGPCVKQEDCNFLIKNKHH
Sg8F6     KDYYDHSSQNECKLPATFAKNKYCDCMDGFLRTNAGPCVKKEDCNFLIKNEHH
S17B2     RDFFDYSSKKECNLPATYAKNKYCDCMEGYWRLDDGPCVKSENCDILLQNQNTK
Sg6G2     KDLYDHSKKKECNLPATFAKNKYCECMEGYYHSVKGPCVREEDCNFLI
```

101

Figure 3C: Nucleotide sequence alignment of Group II cDNAs. Sequence variations among group members are bold and colored. The 5' inframe stop codon, start codon, stop codon and poly (A) tail signal are bold and underlined.

```
          1                                                           60
14A4      -atatctacgaggactattgctggactctgaacgtgtttaatattcgaagaaaagaaaaa
Lg1A1     --tatcaacgaggactattgctggactctgaacgtgtttaatattcgaagaaaagaaaaa
Lg4H7     --tatcaacgaggactattgctggactctgaacgtgtttaatattcgaagaaaagaaaaa
Lg4C9     -atatcaacgaggactattgctggactctgaacgtgtttaatattcgaagaaaagaaaaa
Pg10a9    --tatcaacgaggactattgctggactctgaacgtgtttaatattcgaagaaaagaagaa
S10F3     --taacagcaagaactattgctggactctgaacgtgtttaatattcgcagaaaagaaaaa
Sg8F6     --tatcaacgaggactattgctggactctgaacgtgtttaatattcgaagaaaagaaaaa
S17B2     gatatcaacgaggactattgctggactctgaacgtgtttaatattcgaagaaaagaaaaa
Sg6G2     gatatcaacgaggactattgctggactctgaacgtgtttaatattcgaagaaaagaaaaa

          61                                                          120
14A4      -taatcctaaaatgtctttgaaattggttataattggagtcgccctgctcggcttcgttg
Lg1A1     -taatcctaaaatgtctttgaaattggttataattggagtcgccctgctcggcttcgttg
Lg4H7     ataatcctaaaatgtctttgaaattggttataattggagtcgccctgctcggcttcgttg
Lg4C9     -taatcctaaaatgtctttgaaattggttataattggagtcgccctgctcggcttcgttg
Pg10a9    -taatcctaaaatgtctttgaaattggttataattggagtcgccctgctcggcttcgttg
S10F3     -taatcgaaaaatgtctttgaaattggttataattgggggtcgctctgctcggcttcgttg
Sg8F6     -taatccaaaaatgtgtttgaaattggttataattggagtcgccctgctcggcttcgttg
S17B2     -taatcctaaaatgtgtttgaaattggttataattggagtcgctttgctcggcttcgttg
Sg6G2     -taatccaaaaatgtctttgaaattggttataattggagtcgctctgcttggcttcgttg

          121                                                         180
14A4      catcttcatt-----------------gggtaatgaagacaacaaaccaaacttatgtt
Lg1A1     catcttcatt-----------------gggtaatgaagacaacaaaccaaacttatgtt
Lg4H7     catcttcatt-----------------gggttatgaagacaacaaaccaaacttatgtt
Lg4C9     catcttcatt-----------------gggttatgaagacagcaaaccaaacttatgtt
Pg10a9    catcttcatt-----------------gggttatgaagacaacaaaccaaacttatgtt
S10F3     cattttcatt-----------------gggtaatgaagacaccaaaccaaacttatgtt
Sg8F6     catcttcatt-----------------gggttatgaagacaacaaaccaaacttatgtt
S17B2     catctacattcattccgctacctccatcgggtaatgaagacaccaaaccaaacttatgtt
Sg6G2     catcttcatt-----------------gggtaatgaagacaacaaaccaaacttttgtt

          181                                                         240
14A4      gtcctcaaaacgaaaaactttatgactatcattatccgtgcaaaaaaaacatgtgcaaag
Lg1A1     gtcctcaaaacgaaaaactttatgactatcattatccgtgcaaaaaaaacatgtgcaaag
Lg4H7     gtcctcaaaacgaaaaactttatgactatcattatccgtgcaaaaaaaacatgtgcaaag
Lg4C9     gtcctcaaaacgaaaaactttatgactatcattatccgtgcaaaaaaaacatgtgcaaag
Pg10a9    gtcctcaaaacgaaaaactttatgactatcattatccgtgcaaaaaaaacatgtgcaaag
S10F3     gtcctcaaaacgaaaaactttatgattatcattatccgtgcaaaaaaaacatgtgcaaag
Sg8F6     gtcctcaaaacgaaaaactttatgactatcattatccgtgcaaaaaaaacatgtgcaaag
S17B2     gtcctcaaaacgaaaaactttatgattataattatccgtgcaaaaaaaacatgtgtagag
Sg6G2     gtcctcaaaacgaaaaactttatgattataattatccgtgcaaaataaccatgtgtaaag

          241                                                         300
14A4      actattacgaccattccagtcaaaacgaatgcaaacttccggccacttttgctaaaaaca
Lg1A1     actattacgaccattccagtcaaaacgaatgcaaacttccggccacttttgctaaaaaca
Lg4H7     acttatacgaccattccagtcaaaacgaatgcaaacttccggccacttttgctaaaaaca
Lg4C9     acttatacgaccattccagtcaaaacgaatgcaaacttccggccacttttgctaaaaaca
Pg10a9    acttatacgaccattccagtcaaaacgaatgcaaacttccggccacttttgctaaaaaca
S10F3     acttttacgaccattccagtcaaaccgaatgcaaacttccggccacttttgctaaaaaca
Sg8F6     actattacgaccattccagtcaaaacgaatgcaaacttccggccacttttgctaaaaaca
S17B2     actttttcgactattccagtaagaaggaatgcaatcttccggccacttatgctaaaaaca
Sg6G2     acttatacgaccattccaaaaagaaggaatgcaatcttccggccacttttgctaaaaaca
```

103

```
        301                                                           360
14A4    aatattgcgattgtaaggatggttacttacgcactaatgcaggtccttgtgttaagaaag
Lg1A1   aatattgcgattgtgaggatggttacttacgcactaatgcaggtccttgtgttaagaaag
Lg4H7   aatattgcgattgtatggatggctacttacgcactaatgcaggtccttgtgttaagaaag
Lg4C9   aatattgcgattgtatggatggctacttacgcactaatgcaggtccttgtgttaagaaag
Pg10a9  aatattgcgattgtatggatggctacttacgcactaatgcaggtccttgtgttgagaaag
S10F3   aatattgcgattgtatggatggttacttacgcactaatgcaggtccttgtgttaaacaag
Sg8F6   aatattgcgattgtatggatggtttcttacgcactaatgcaggtccttgtgttaagaaag
S17B2   aatattgtgattgtatggaaggttactggcgccttgatgatggtccttgtgttaagagcg
Sg6G2   aatattgcgaatgtatggaaggttactaccacagtgttaaaggtccttgtgttagagaag

        361                                                           420
14A4    aagattgcaatttcttgataaaaaatgagcat-c------attgatttgaaatgaaacgc
Lg1A1   aagattgcaatttcttgataaaaaatgagcat-c------attgatttgaaatgaaacgc
Lg4H7   aagattgcaatttcttgataaaaaatgagcat-c------attgatttgaaatgaaacgc
Lg4C9   aagattgcaatttcttgataaaaaatgagcat-c------attgatttgaaatgaaacgc
Pg10a9  aagattgcaatttcttgataaaaaatgagcat-c------attgatttgaaatgaaacgc
S10F3   aagattgcaatttcttgataaaaaataagcat-c------attgatttaaaatgaaacgc
Sg8F6   aagattgcaatttcttgataaaaaatgagcat-c------attgatttgaaatgaaacgc
S17B2   aaaattgcgatatcttgttacaaaatcagaatacgaaatgattgatttgaaatgaaacgc
Sg6G2   aagattgcaatttcttgat--------------------tt-----aaaatgaaacgc

        421                                                           480
14A4    aactgataaggaattggcaatgataaaaaaa---ctgtatccaaatctttaacaatttaa
Lg1A1   Aactgataaggaattggcaatgataaaaaaa---ctgtatccaaatctttaacaatttaa
Lg4H7   aactgataaggaattggcaacgataaaaaa----ctgtattcaaattttttaacaatttaa
Lg4C9   aactgataaggaattggcaatgataaaaaa----ctgtattcaaattttttaacaatttaa
Pg10a9  aactgatgaggaattggcaatgataaaaaa----ctgtattcaaattttttaacaatttaa
S10F3   aacggataaagaattggcaatgataaaaaaa---ttgtattcaaatctttaacaatttaa
Sg8F6   aactgataaagaattggcaatgataaaaaa----ctgtattcaaattttttgacaatttaa
S17B2   aacggataaggaatttgcaatgatgaaaaa----ctgtattcaaatctttaacaatttaa
Sg6G2   aacggataaagaattggcaatgataaaaaaaaaactgtattcaaatctttaacaatttaa

        481                                                       531
14A4    aaataaagcaaaataaaaattccattttcctggtgaagggaacaacttgca
Lg1A1   aaataaagcaaaataaaaattccattttcctggtgaagggaacaacttgca
Lg4H7   aa-taaagcaaaataaaaattccattttcctggtgaagggaacaacttgaa
Lg4C9   aa-taaagcaaaataaaaattccattttcctggtgaagggaacaacttaaa
Pg10a9  aa-taaagcaaaaaaaaaa--------------------------aa
S10F3   aaataaagcaaaataaaaattacattttcctggtgaagggaataacttata
Sg8F6   aa-taaagcaaaataaaaattccattttcctggtgaagggaacaacttgaa
S17B2   aa-taaagcaaaataaaaattccatttttctggtgaatggaacaacttaag
Sg6G2   aa-aaaagcaaaataaaaattccatttttctggtgaagggaaaaa----aa
```

104

Figure 4A: Nucleotide and amino acid sequence of a representative cDNA from Group III. The amino acid is shown under the nucleotide sequence. Signal peptide is denoted by underline. * indicates the stop codon. Cysteine residues are in bold letters.

Lg2A3

```
-68                                                                  tatcaacg

-60 aggactattgctggactctgaacgtatttaatattcgaagcaaagaaaaataatcgaaaa
                      *
1   atgtctttgaaattggttataattggaatcgctctgctcggtttcgttgcatcttcattg
    M   S   L   K   L   V   I   I   G   I   A   L   L   G   F   V   A   S   S   L

61  ggtgaagacgcccaaccaaattcagcttgccctccaaacgaaaccctttatgattttgtc
    G   E   D   A   Q   P   N   S   A   C   P   P   N   E   T   L   Y   D   F   V

121 gatccgtgcaggaaagacacttgtgaaaaggcaatcaaaaacgtcggattttacgaatgc
    D   P   C   R   K   D   T   C   E   K   A   I   K   N   V   G   F   Y   E   C

181 ggtggctacaaaactgaacatccagtttgcaattgtgtcgacaacttttatcgcaataaa
    G   G   Y   K   T   E   H   P   V   C   N   C   V   D   N   F   Y   R   N   K

241 gagggtcactgtgtttcgatagatcaatgcagatcagaaataaatggataaaggcaccgg
    E   G   H   C   V   S   I   D   Q   C   R   S   E   I   N   G   *

301 aataaaggcaacggataaggaatttgcaatgataaaaaaactgtattcaaatctttaaca

361 atttaaaataaagcaaaataaaaattccattttttctggtgaagggaacaacttcc
```

Figure 4B: Sequence alignment of putative proteins encoded by Group III cDNAs. Sequence variations among group members are bold and underlined. N-terminal signal peptide is bold and underlined. P_1 represent the putative amino acid residue at the reactive site center. Grey boxes represent the conserved cysteine residues.

```
         1                                                    P₁          56
Lg2A3    MSLKLVIIGIALLGFVASSL---GEDAQPNSACPPNETLYDFVDPCRKDTCEKAIK
PG10A8   MSLKLVIIGIALLGFVASSL---GEDAQPNSACPPNETLYDFVDPCRKDTCEKAIK
PG10C11  MSLKLVIIGIALLGFVASSL---GEDAQPNSACPPNETLYDFVDPCRKDTCEKAMK
PG10G12  MSLKLVIIGIALLGFVASSL---GEDAQPNSACPPNETLYDFVDPCRKDTCEKAMK
Lg2G2    MSLKLVIIGIALLGFVASAL---GEDTQPNSACPPNETLYDFVDPCRKDTCEKAIK
Lg2G5    MSLKLVIIGIALLGFVASALDVEGEDTQPNSACPPNETLYDFVDPCRKDTCEKAIK
Sg6C8    MSLKLVIIGVALLGFVAFSL---GDGTQPNSACPPNETLYDFVDPCRKDTCEKAIK

         1                                              99
Lg2A3    NVGFYECGGYKTEHPVCNCVDNFYRNKEGHCVSIDQCRSEING
PG10A8   NVGFYECGGYKTEHPVRNCVDNFYRNKEGHCVSIDQCRSEING
PG10C11  DTGFYECGGHKTEHPVCNCVENTYRNKEGRCVSLDQCKSEING
PG10G12  DTGFYECGGHKTEHPVCNCVENTYRNKEGRCVSLDQCKSEING
Lg2G2    NVGFYECGGHKTEHPVCNCVDNMYRNKAGHCVSIDQCRSEING
Lg2G5    NVGFYECGGHKTEHPVCNCVDNMYRNKAGHCVSIDQCRSEING
Sg6C8    NIGFYECGGYKVEHPVCNCVDNMYRNKENRCVSIDQCRSEING
```

108

Figure 4C: Nucleotide sequence alignment of Group III cDNAs. Sequence variations among group members are bold and colored. The 5' inframe stop codon, start codon, stop codon and poly (A) tail signal are bold and underlined.

```
          1                                                            60
Lg2A3     ---tatcaacgaggactattgctggactctgaacgtatttaatattcgaagcaaagaaaa
PG10A8    ---tatcaacgaggactattgctggactctgaacgtatttaatattcgaagcaaagaaaa
PG10C11   ---tatcaacgaggactattgctggactctgaacgtatttaatattcgaagcaaagaaaa
PG10G12   ---tatcaacgaggactattgctggactctgaacgtatttaatattcgaagcaaagaaaa
Lg2G2     ---tatcaacgaggacaattgctggactctgaacgtatttaatattcgaagcaaagaaaa
Lg2G5     -----tcaacgaggacaattgctggactctgaacgtatttaatattcgaagcaaagaaaa
Sg6C8     tcatatcaacgaggactattgctggactctgaacgaatattaaattcgaagcaaagaaaa

          61                                                           120
Lg2A3     ataatcgaaaaatgtctttgaaattggttataattggaatcgctctgctcggtttcgttg
PG10A8    ataatcgaaaaatgtctttgaaattggttataattggaatcgctctgctcggtttcgttg
PG10C11   ataatcgaaaaatgtctttgaaattggttattattggaatcgctctgctcggcttcgttg
PG10G12   ataatcgaaaaatgtctttgaaattggttattattggaatcgctctgctcggcttcgttg
Lg2G2     ataatcgaaaaatgtctttgaaattggttataattggaatcgctctgctcggcttcgttg
Lg2G5     ataatcgaaaaatgtctttgaaattggttataattggaatcgctctgctcggcttcgttg
Sg6C8     ataatcgaaaaatgtctttgaaattggttataattggagtcgctctgctcggcttcgttg

          121                                                          180
Lg2A3     catcttcattgg---------gtgaagacgcccaaccaaattcagcttgccctccaaacg
PG10A8    catcttcattgg---------gtgaagacgcccaaccaaattcagcttgccctccaaacg
PG10C11   catcttcattgg---------gtgaagacgcccaaccaaattcagcttgccctccaaacg
PG10G12   catcttcattgg---------gtgaagacgcccaaccaaattcagcttgccctccaaacg
Lg2G2     catctgcattgg---------gtgaagacacccaaccaaattcagcttgccctccaaacg
Lg2G5     catctgcattggatgtggaaggtgaagacacccaaccaaattcagcttgccctccaaacg
Sg6C8     cattttcattgg---------gtgatggcacccaaccaaattcagcttgccctccaaacg

          181                                                          240
Lg2A3     aaaccctttatgattttgtcgatccgtgcaggaaagacacttgtgaaaaggcaatcaaaa
PG10A8    aaaccctttatgattttgtcgatccgtgcaggaaagacacttgtgaaaaggcaatcaaaa
PG10C11   aaaccctttatgattttgtcgatccgtgcaggaaagacacttgtgaaaaggcaatgaaag
PG10G12   aaaccctttatgattttgtcgatccgtgcaggaaagacacttgtgaaaaggcaatgaaag
Lg2G2     aaaccctttatgattttgtcgatccgtgcagaaaagacacttgtgaaaaggcaatcaaaa
Lg2G5     aaaccctttatgattttgtcgatccgtgcagaaaagacacttgtgaaaaggcaatcaaaa
Sg6C8     aaaccctttatgatttcgtcgatccgtgcaggaaagacacttgtgaaaaggcaatcaaaa

          241                                                          300
Lg2A3     acgtcggattttacgaatgcggtggctacaaaactgaacatccagtttgcaattgtgtcg
PG10A8    acgtcggattttacgaatgcggtggctacaaaactgaacatccagttcgcaattgtgtcg
PG10C11   acaccggattttacgaatgcggggggccacaaaactgaacatccagtttgcaattgtgtcg
PG10G12   acaccggattttacgaatgcggggggccacaaaactgaacatccagtttgcaattgtgtcg
Lg2G2     acgtcggattttacgaatgcggggggccacaaaactgaacatccagtttgcaattgtgtcg
Lg2G5     acgtcggattttacgaatgcggggggccacaaaactgaacatccagtttgcaattgtgtcg
Sg6C8     acatcggattttacgaatgcggggggctacaaagttgaacatccagtttgcaattgtgtcg

          301                                                          360
Lg2A3     acaacttttatcgcaataaagagggtcactgtgtttcgatagatcaatgcagatcagaaa
PG10A8    acaacttttatcgcaataaagagggtcactgtgtttcgatagatcaatgcagatcagaaa
PG10C11   aaaacacgtatcgcaataaagaaggtcgctgtgtttcgttggatcaatgcaaatcagaaa
PG10G12   aaaacacgtatcgcaataaagaaggtcgctgtgtttcgttggatcaatgcaaatcagaaa
Lg2G2     acaatatgtatcgcaataaagcgggtcactgtgtttcgatagatcaatgccgatcagaaa
Lg2G5     acaatatgtatcgcaataaagcgggtcactgtgtttcgatagatcaatgccgatcagaaa
Sg6C8     acaatatgtatcgcaataaagagaatcgctgtgtttcgatagatcaatgcagatcagaaa
```

```
         361                                                                      420
Lg2A3    taaatgga**taa**aggca**c**cggaataaaggcaacggataaggaatt**t**gcaatgataaaaaaa
PG10A8   taaatggataaaggca**c**cggaataaaggcaacggataaggaatt**t**gcaatgataaaaaaa
PG10C11  taaatggataaaggcaacggaataaaggcaacggataaggaattggcaatgataaaaaa-
PG10G12  taaatggataaaggcaacggaataaaggcaacggataaggaattggcaatgataaaaaa-
Lg2G2    taaatggataaaggcaacggaataaaggcaacggataaggaattggcaatgataaaaaaa
Lg2G5    taaatggataaaggcaacggaataaaggcaacggataaggaattggcaatgataaaaaaa
Sg6C8    taaatggataaaggcaa**t**ggaataaaggcaacggataaggaattggcaatgataaaaaaa


         421                                                                      480
Lg2A3    -ctgtattcaaatctttaacaattt**aaaataaa**gcaaaataaaaattccatttttctggt
PG10A8   -ctgtattcaaatctttaacaatttaaaataaagcaaaataaaaattccatttttctggt
PG10C11  -ctgtattcaaatctttaa**a**aatttaaaataaagcaaaataaaaattccatttttctggt
PG10G12  -ctgtattcaaatctttaa**a**aatttaaaataaagcaaaataaaaattccatttttctggt
Lg2G2    actgtattcaaatctttaacaatttaaaataaagcaaaataaaaattccatttttctggt
Lg2G5    actgtattcaaatctttaacaatttaaaataaagcaaaataaaaattccatttttctggt
Sg6C8    actgtattcaaatctttt**t**a--atttaaaataaagcaaaataaaaattccatttttctggt


         481                  504
Lg2A3    gaagggaacaactt**c**caaaaaaaa
PG10A8   gaagggaacaacttgcaaaaaaaa
PG10C11  gaagggaa--------aaaaaaaa
PG10G12  gaagggaacaacttgcaaaaaaaa
Lg2G2    gaagggaacaacttgcaaaaaa**g**a
Lg2G5    gaagggaacaacttgca**t**aaaaaa
Sg6C8    gaagggaacaacttg-a**t**aaaaaa
```

Figure 5A: Nucleotide and amino acid sequence of a representative cDNA from Group IV. The amino acid is shown under the nucleotide sequence. Signal peptide is denoted by underline. * indicates the stop codon. Cysteine residues are in bold letters.

Lg2F7

```
-112          ccagtgtgatggatatctgcagaattcgcccttaagcagtggtatcaacgc
                     *

-60  agagtggccattaccgcgagaacgtgtttaatattcgaagaaaagaaaaataatccaaaa

1    atgtgtttgaaattgcttataattggaattgctctgctcgtcttcgttacatcttcattg
     M  C  L  K  L  L  I  I  G  I  A  L  L  V  F  V  T  S  S  L

61   ggtgaatgtaatcaagacatccaatcaaagaactgttgtcctgaaaacgaaacactttat
     G  E  C  N  Q  D  I  Q  S  K  N  C  C  P  E  N  E  T  L  Y

121  gatcatgattatccgtgcgctagagacacttgtgcaaaggcattcgacgaccccgcattt
     D  H  D  Y  P  C  A  R  D  T  C  A  K  A  F  D  D  P  A  F

181  tacgaatgcaaatttgtggcacaaaaatctaaatttcgatattgcaattgtaacgacaat
     Y  E  C  K  F  V  A  Q  K  S  K  F  R  Y  C  N  C  N  D  N

241  atgtttcgcaataaagatggttactgtgtttcgattgaacaatgcttattagaaataaat
     M  F  R  N  K  D  G  Y  C  V  S  I  E  Q  C  L  L  E  I  N

301  ggttcaatatgaaaggcaacagaaaaggaatttccaatgatagaaaagtgtattcaaatc
     G  S  I  *

361  tttaacaatttaaaataaaggaaaataaaaaaaattctttgttttttggtgaaggcaacaa

421  ctt
```

Figure 5B: Sequence alignment of putative proteins encoded by Group IV cDNAs. Sequence variations among group members are bold and underlined. N-terminal signal peptide is bold and underlined. P_1 represent the putative amino acid residue at the reactive site center. Grey boxes represent the conserved cysteine residues.

```
          1                                                        P₁        56
Lg2F7  MCLKLLIIGIALLVFVTSSLGECNQDIQSKNCCPENETLYDHDYPCARDTCAKAFD
Pg7f5  MSLKLVIIGIALHVFVTSSLGECNQDFQSKNCCPQNEILYDHDYPCARDTCAKAFD


          57                                                  103
Lg2F7  DPAFYECKFVAQKSKFRYCNCNDNMFRNKDGYCVSIEQCLLEINGSI
Pg7f5  DPAFYECKFVAQKSKYQYCNCNDTMFRNKAGQCVSIEECLLEINGSI
```

Figure 5C: Nucleotide sequence alignment of Group IV cDNAs. Sequence variations among group members are bold and colored. The 5' inframe stop codon, start codon, stop codon and poly (A) signal are bold and underlined.

```
        1                                                           60
Lg2F7   ccagtgtgatggatatctgcagaattcgcccttaagcagtggtatcaacgcagagtggcc
Pg7f5   -------------------------------------------------catattagca


        61                                                          120
Lg2F7   a---ttaccgcgagaacgtgtttaatattcgaagaaaagaaaaataatccaaaaatgtgt
Pg7f5   agaattattgccggactctg---aacgt--gtagaaaagaaaaataattcaaaaatgtct


        121                                                         180
Lg2F7   ttgaaattgcttataattggaattgctctgctcgtcttcgttacatcttcattgggtgaa
Pg7f5   ttgaaattggttataattggaattgctctgcatgtcttcgttacatcttcattgggtgaa


        181                                                         240
Lg2F7   tgtaatcaagacatccaatcaaagaactgttgtcctgaaaacgaaacactttatgatcat
Pg7f5   tgtaaccaagacttccaatcaaagaactgttgtcctcaaaacgaaatactttatgatcat


        241                                                         300
Lg2F7   gattatccgtgcgctagagacacttgtgcaaaggcattcgacgaccccgcattttacgaa
Pg7f5   gattatccgtgcgctagagacacttgtgcaaaggcattcgacgaccccgcattttacgaa


        301                                                         360
Lg2F7   tgcaaatttgtggcacaaaaatctaaatttcgatattgcaattgtaacgacaatatgttt
Pg7f5   tgcaaatttgtggccaaaaatctaaatatcaatattgcaattgtaacgacactatgttt


        361                                                         420
Lg2F7   cgcaataaagatggttactgtgtgtttcgattgaacaatgcttattagaaataaatggttca
Pg7f5   cgcaataaagctggtcagtgtgtgtttcgatagaagaatgcttattagaaataaatggttca


        421                                                         480
Lg2F7   atatgaaaggcaacagaaaaggaatttccaatgatagaaaagtgtattcaaatctttaac
Pg7f5   atatgaaaggtaacggatacggaatttgtaatgatagaaaagtgtattcaaatcttttac


        481                                                         540
Lg2F7   aatttaaaataaaggaaaataaaaaaaattctttgtttttgg-tgaaggcaacaacttaa
Pg7f5   aatttaaaataaaggaaaataaaaaaaattctttgtttttgggtgaaggcaacaacttac
```

Figure 6: Nucleotide and amino acid sequence of the only one cDNA from Group V. The amino acid is shown under the nucleotide sequence. Signal peptide is denoted by underline. * indicates the stop codon. Cysteine residues are in bold letters.

11C4

```
-28                                                    catatcgtttgttaacttcaacaagaat

1    atgaattcgttaattggtttgattaaaattgttttttttattggaattgttgttatttct
      M  N  S  L  I  G  L  I  K  I  V  F  F  I  G  I  V  V  I  S

61   gtacatccagctttggctgaaaacaactcagaatgtgtttcatctacaccgaaagactgt
      V  H  P  A  L  A  E  N  N  S  E  C  V  S  S  T  P  K  D  C

121  ggaccaaataaggtattgaatccgcgcttcttacatcgatgctttttatccacaatgcatc
      G  P  N  K  V  L  N  P  R  F  L  H  R  C  F  Y  P  Q  C  I

181  acttcgactgaatcgtgtccagatcaacgtcaaatcaaatttaaattgtcgccaagacct
      T  S  T  E  S  C  P  D  Q  R  Q  I  K  F  K  L  S  P  R  P

241  gaatgcgcatgcagtgatggatttgcttataaatcaattgatacgtgtgcaccgattaat
      E  C  A  C  S  D  G  F  A  Y  K  S  I  D  T  C  A  P  I  N

301  tcacctgaatgtggacaattgggttacagttggccaccacttccttacacatcatagctt
      S  P  E  C  G  Q  L  G  Y  S  W  P  P  L  P  Y  T  S  *

361  atttgaattacattttttgcgaatatctccgattataatcgtcaaaattttccgttattga

421  aataaataacccaaaatcc
```

Figure 7A: Alignment of two representative proteins from two different groups. The conserved cysteine residues between the two proteins are highlighted in grey boxes. Alignments between other groups revealed similar conservation patterns (data not shown).

Identities = 24/100 (24%)

```
11A6 : 1   MDSKMFLLMVAALTVFWITSGIAAPLAEDEPVDRTCTRLNE-----FYAERSSRCDDKYC 60
           M  K+ ++ VA L          +A+ L  ++    C   NE      Y + + C D Y
14A4 :     MSLKLVIIGVALLGF------VASSLGNEDNKPNLCCPQNEKLYDYHYPCKKNMCKDYYD


11A6 : 61  DGSEIYMKCTQDCSHGYRCRCITGTKRDSHGNCIEENKCN 100
            S+   K     +   C C  G  R + G C+++  CN
14A4 :     HSSQNECKLPATFAKNKYCDCKDGYLRTNAGPCVKKEDCN
```

121

Figure 7B: Multiple alignments of deduced amino acid sequences of *M. destructor* with *Ascaris* family members, *Ascaris* trypsin inhibitor (ATI) and Bombina skin trypsin inhibitor (BSTI). The P_1 reactive site residues are boxed. Towards C-terminal end of P_1 the residues were named as P_1', P_2', P_3'….. and towards N-terminal end of P_1 the residues were named P_2, P_3, P_4…. . Identical residues of all inhibitors are highlighted with black background and in clear boxes. Gray shades represent the loosely conserved amino acid residues. Conserved cysteine residues are shown in asterisk (*****) below the sequence.

```
                    10          20          30          40      P₁    50          63
11A6    ------------------APLAEDEPVDRTCTRLNEFYAERSSRCDDKYCDGS---EIYMKCT
14A4    --------------------NEDNKPN--LCCPQNEKLYDYHYPCKKNMCKDYYDHSSQNECK
LG2A3   --------------------EDAQPN---SACPPNETLYDFVDPCRKDTCEKAIKNVGFYECG
LG2F7   -------------------------CNQDIQSKNCCPENETLYDHDYPCARDTCAKAFDDPAFYECK
11C4    -----------------NNSECVSSTP-KDCGPNKVLNPRFLHRCFYPQCITS-----TESCP
ATI     EAEKCTKPNEQWTKCGGCEGTCAQKI------------V----PCTR------------ECK
BSTI    -NFVCPPGQTFQTCASSCPKTCETRN-------------KLVLCDK-------------KCN
                                                *         *           *
```

```
                    70          80          90          100         114
11A6    ------QDCSHGYRCRCIT--GTKRDSHGNCIEENKCNATKLWL-------
14A4    LPA----TFAKNKYCDCKD--GYLRTNAGPCVKKEDCNFLIKNEHH-----
LG2A3   --G----YKTEHPVCNCVD--NFYRNKEGHCVSIDQCRSEING--------
LG2F7   FVA----QKSKFRYCNCND--NMFRNKDGYCVSIEQCLLEINGSI------
11C4    DQRQIKFKLSPRPECACSDGFAYKSIDTCAPINSPECGQLGYSWPPLPYTS
ATI     --P---------PRCECIASAGFVRDAQGNCIKFEDCPK------------
BSTI    ------------QRCGCISGTVLKSKDSSECVHPSKC--------------
                 *  *            *        *
```

123

Figure 7C: Alignment of deduced *M. destructor* 11A6 protein with a trypsin-inhibitor-like protein from *Caenorhabditis elegans*. The conserved cysteine residues between the two proteins are highlighted in grey boxes.

Identities = 24/69 (34%)

```
11A6 :30   EPVDRTCTRLNEFYAERSSRCDDKYCDGSEIYMKCTQDCSHGYRCRCITGTKRDSHGNCI 89
           +P++R C R NE +     + C D C+    +  C Q C+ G  C C  G RD+ G C+
TI   :493  KPINRRC-RSNEKFEPCKTVCSDTKCNEEPRF--CPQVCTGG--CVCQEGFFRDNSGKCV 548



11A6 :90   EENKCNATK 98
            +N C+A K
TI   :549  TQNDCDAQK 557
```

Figure 8: Total proteins on SDS-PAGE of *E. coli* (BL21-DE3) containing expression constructs 11A6, 14A4, and 8A3. The construct 8A3 expresses an unrelated protein at high level in this system as demonstrated before and serve as a positive control in this experiment. U - uninduced, I -induced, M - protein standards. Left hand side of gel picture shows the size of protein standards. Boxed area shows the induced recombinant protein.
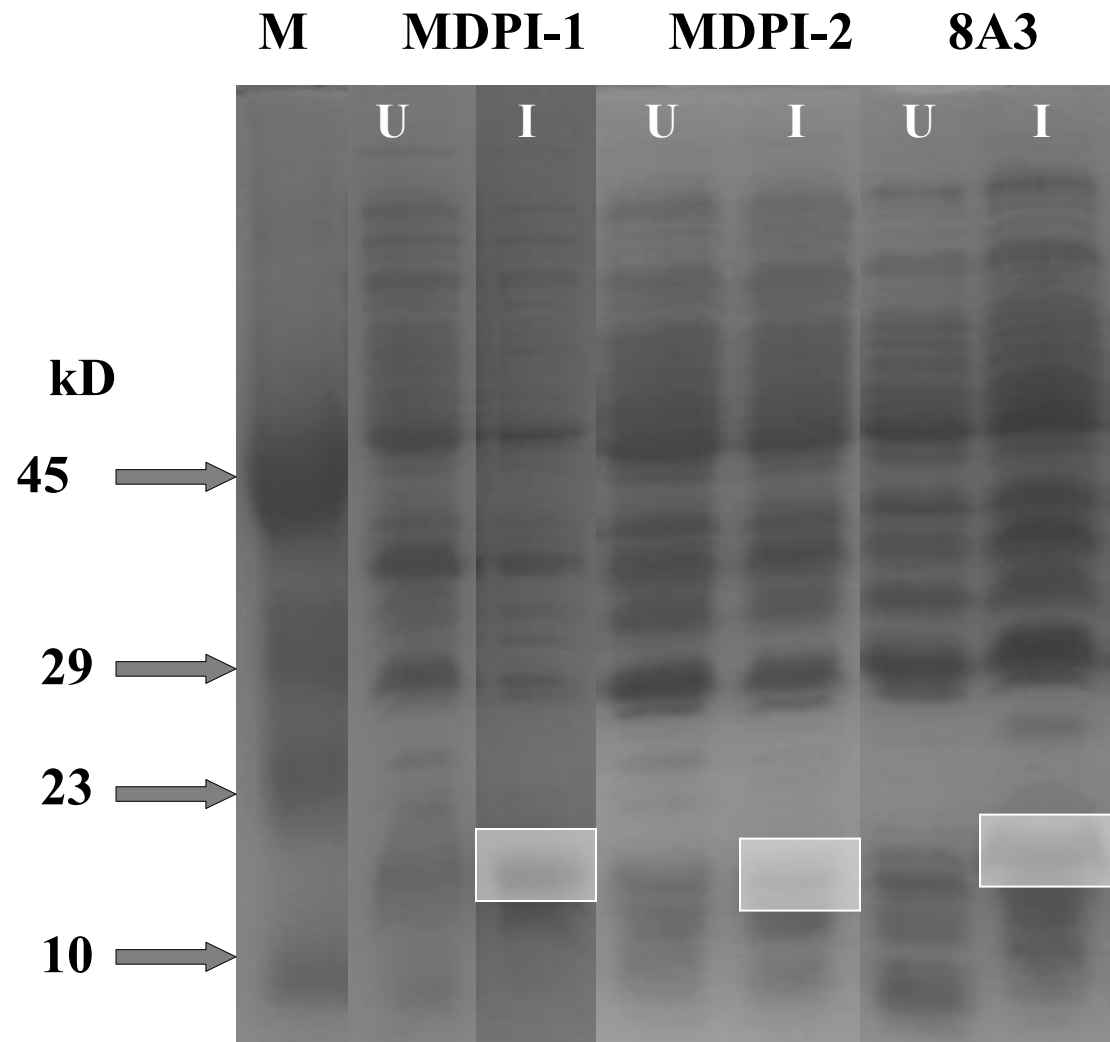
Figure 9: Purified recombinant proteins 11A6 and 14A4 protein with (+) and without (-) N-terminal His-tag. The N-terminal His-tag was removed by using the enterokinase enzyme. Left side of figure shows the sizes of the protein standards.
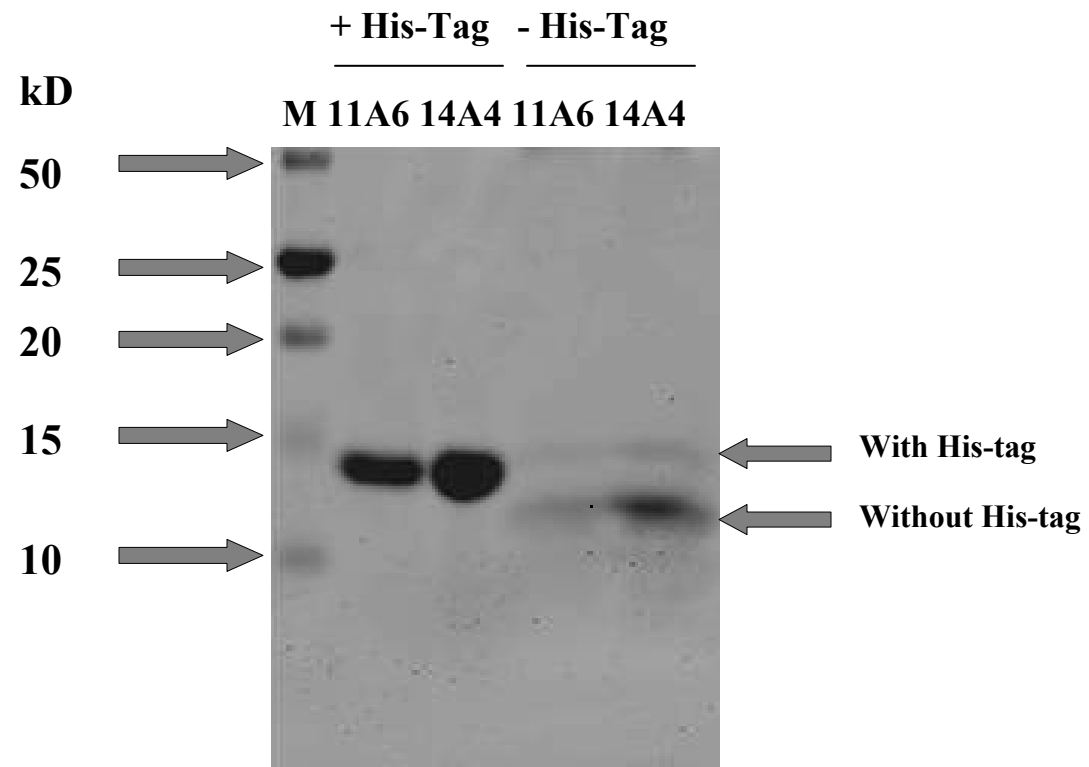
Figure 10: Inhibitory activity of recombinant protein 11A6 toward various proteases and insect gut extracts. YMW-Yellowmeal worm, RFB-Red flour beetle. Recombinant inhibitor was incubated with different proteases at 37°C for 10 minutes before adding substrate. The protease activity was measured by monitoring the hydrolysis of casein conjugated with BODIPY-TR-X.
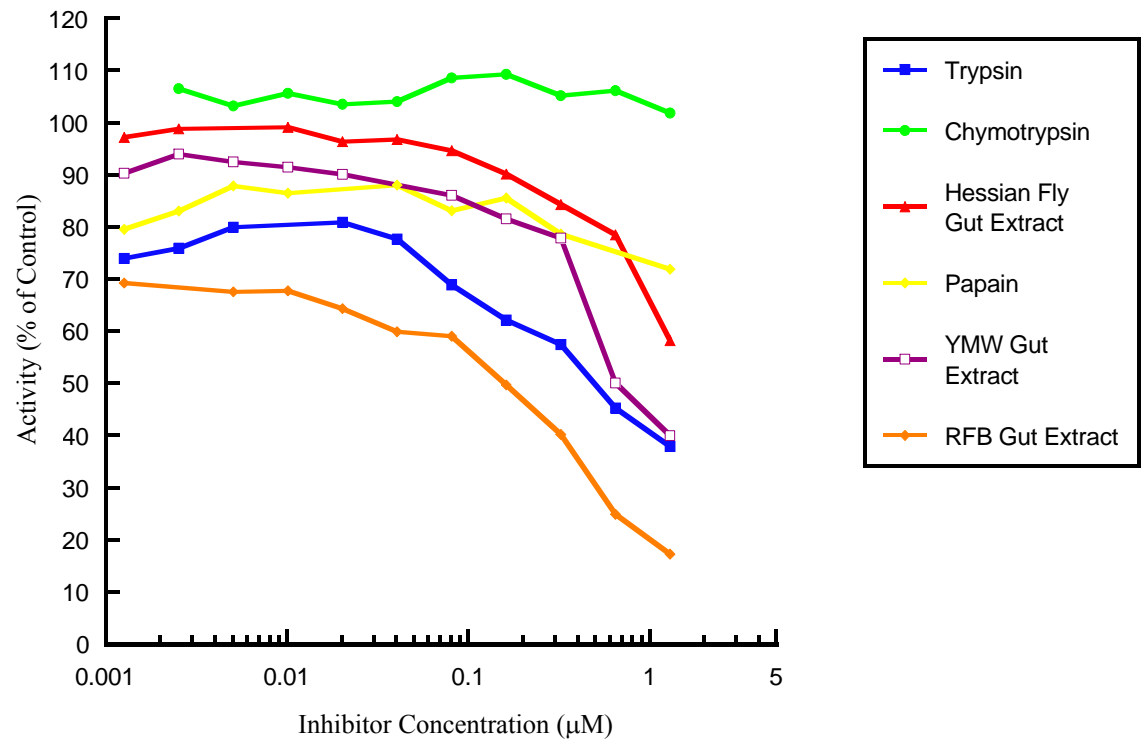
Figure 11: Stoichiometry of inhibition of trypsin activity. Recombinant 11A6 protein was incubated with trypsin at 37°C for 10 minutes at molar ratios ranging from 0:1 to 3.1:1. The trypsin activity was measured by monitoring the hydrolysis of casein conjugated to BODIPY-TR-X. X-axis shows the molar ratio between inhibitor and trypsin and Y-axis shows trypsin activity in percentage.
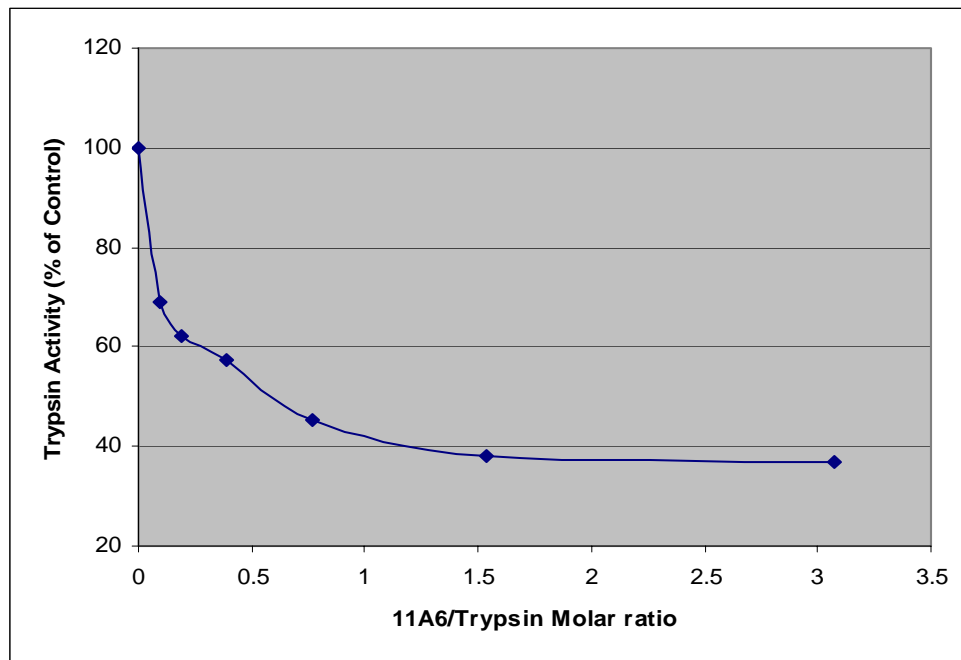
Figure 12: Stoichiometry of inhibition of papain activity. Recombinant 11A6 protein was incubated with trypsin at 37°C for 10 minutes at molar ratios ranging from 0:1 to 1.5:1. The papain activity was measured by monitoring the hydrolysis of casein conjugated to BODIPY-TR-X. X-axis shows the molar ratio between inhibitor and trypsin and Y-axis shows trypsin activity in percentage.
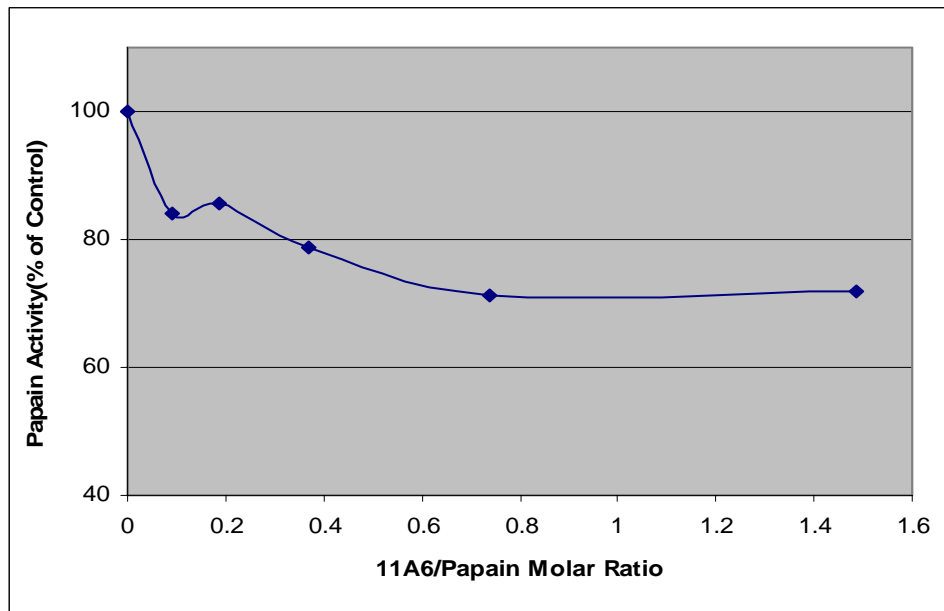
135

Figure 13: Inhibitory activity of recombinant protein 14A4 toward various proteases and insect gut extracts. YMW-Yellowmeal worm, RFB-Red flour beetle. Recombinant inhibitor was incubated with different proteases at 37°C for 10 minutes before adding substrate. The protease activity was measured by monitoring the hydrolysis of casein conjugated with BODIPY-TR-X.
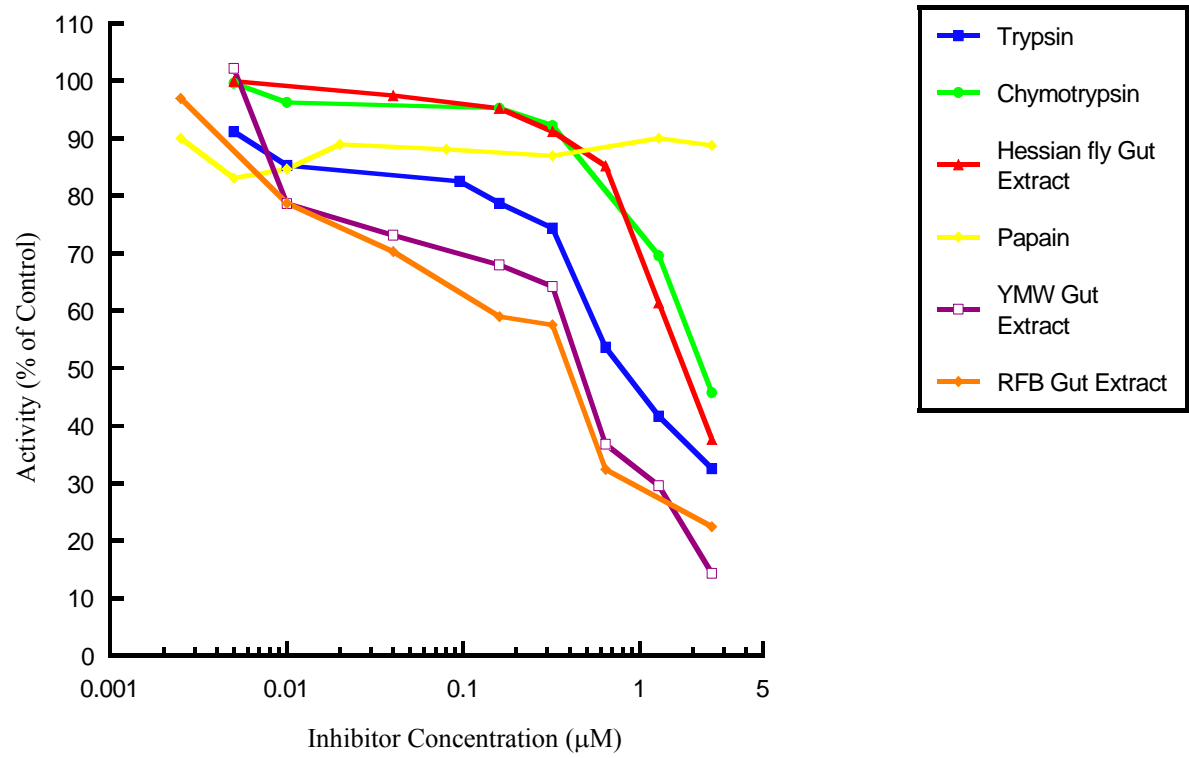
Figure 14: Stoichiometry of inhibition of trypsin activity by 14A4. Recombinant 14A4 protein was incubated with trypsin at 37°C for 10 minutes at molar ratios ranging from 0:1 to 4.1:1 before adding substrate. The trypsin activity was measured by monitoring the hydrolysis of casein conjugated to BODIPY-TR-X. X-axis shows the inhibitor and trypsin molar ration and Y-axis shows trypsin activity in percentage.

Figure 15: Stoichiometry of inhibition of chymotrypsin activity. Recombinant 14A4 protein was incubated with chymotrypsin at 37°C for 10 minutes at molar ratios ranging from 0:1 to 4.3:1 before adding substrate. The trypsin activity was measured by monitoring the hydrolysis of casein conjugated to BODIPY-TR-X. X-axis shows the inhibitor and trypsin molar ration and Y-axis shows trypsin activity in percentage.
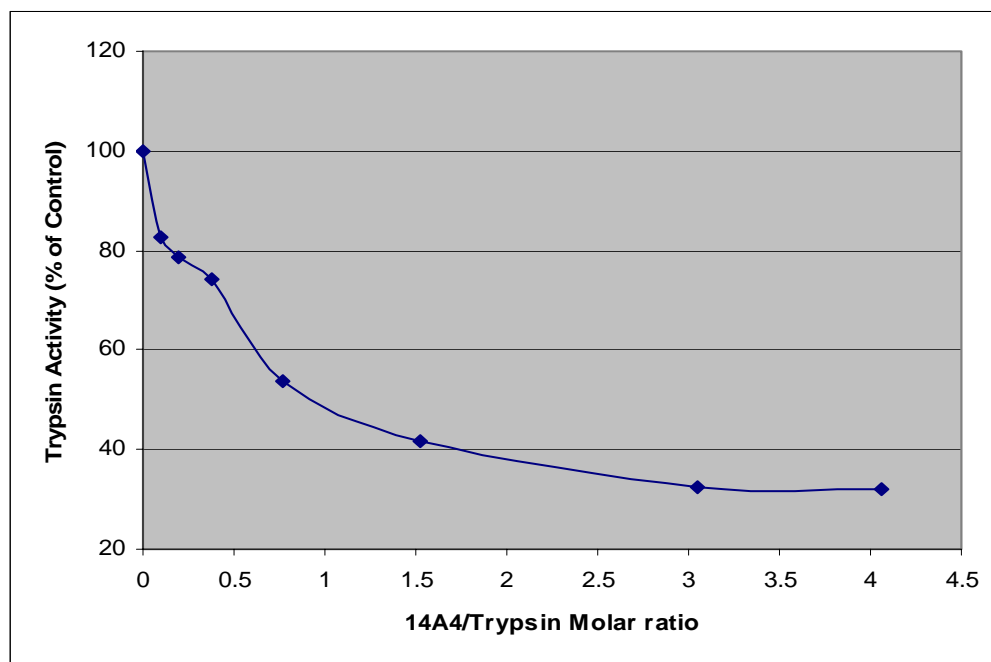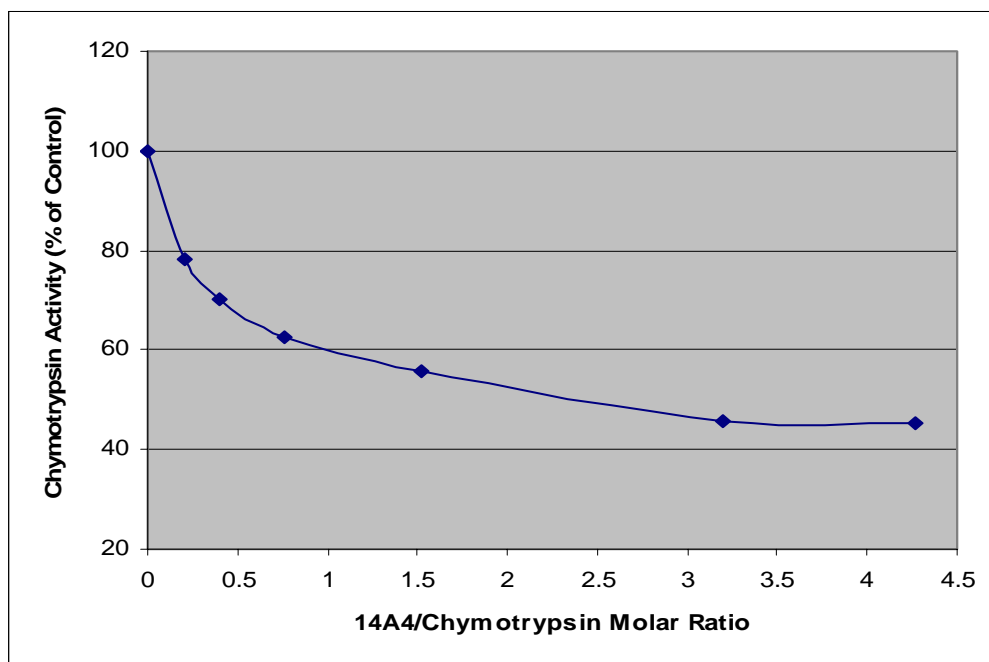
Figure 16: Northern blot analysis of RNA samples prepared from insects at different developmental stages. RNA samples were extracted from 0-day (freshly hatched, nonfeeding), 2-day, 4-day, 6-day and 12-day-old larvae, pupae (P), and adults (A), respectively, as indicated in the figure. Probes specific to group I (11A6), group II (14A4), group III (Lg2A3), group IV (Lg2F7), and group V (11C4) gene groups were used for blot analysis. Bottom panel of the figure shows the image of 18s rRNA stained with ethidium bromide to show equal loading of RNA.

Figure 17A: Regression plot of threshold cycle against the log starting copy number of 11A6 in the pCR2.1-TOPO plasmid. Correlation coefficient– 0.999 Slope: -3.749 Intercept: 45.265 , Y = -3.749X + 45.265

Figure 17B: Quantitative expression of 11A6 gene in different tissue types. MG-Midgut, MT-Malpighian tubule, SG-Salivary gland, FB-Fatbody. 11A6 is abundantly expressed in malpighian tubules and similar levels in midgut and salivary glands. Very low levels of expression in fatbody and carcass.
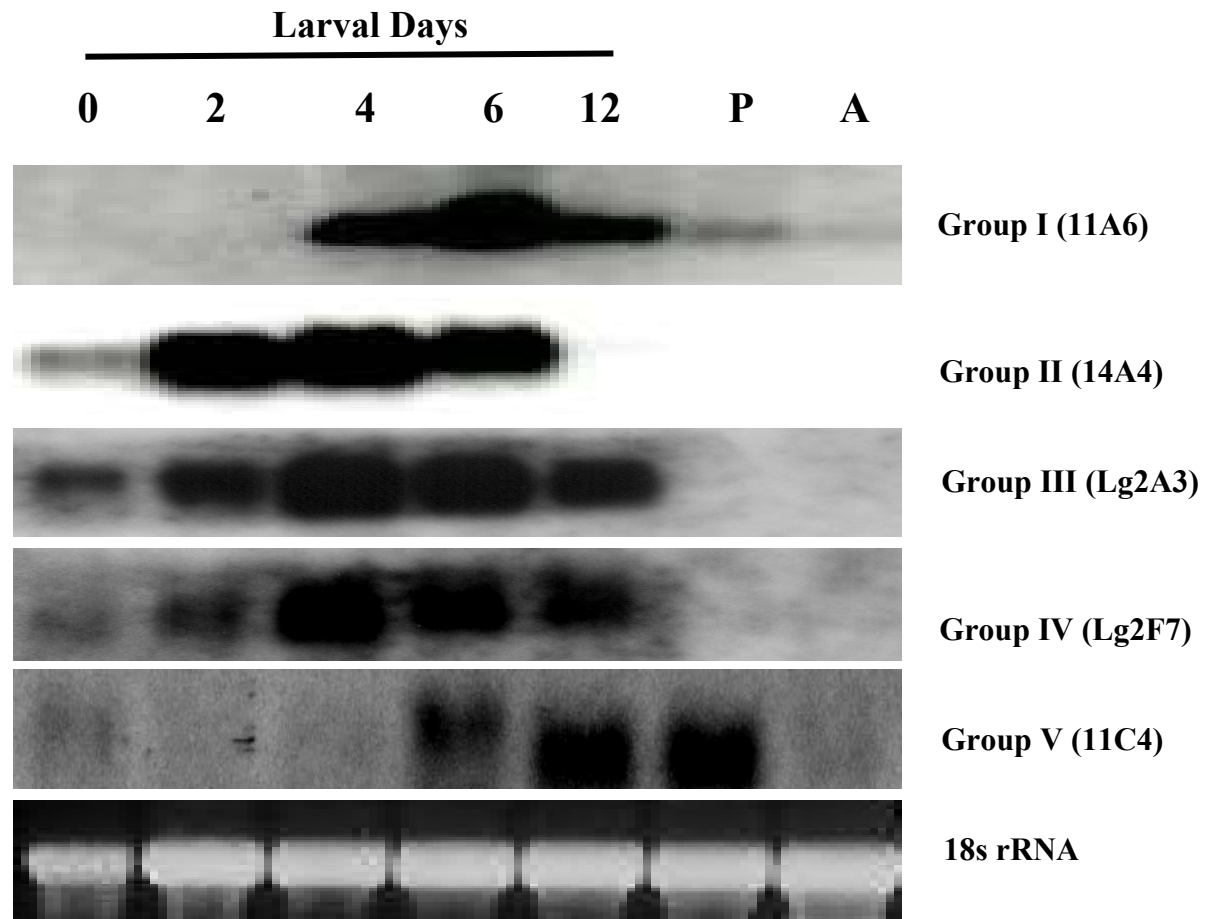
A

B

145

Figure 18A: Regression plot of threshold cycle against the log starting copy number of 14A4 in the pCR2.1-TOPO plasmid. Correlation coefficient– 0.999 Slope: -3.628 Intercept: 42.370, Y = -3.628X + 42.370

Figure 18B: Quantitative expression of 14A4 gene in different tissue types. MG-Midgut, MT-Malpighian tubule, SG-Salivary gland, FB-Fatbody. 14A4 is abundantly expressed in malpighian tubules, similar levels in midgut and salivary glands. Very low levels of expression in fatbody and carcass.
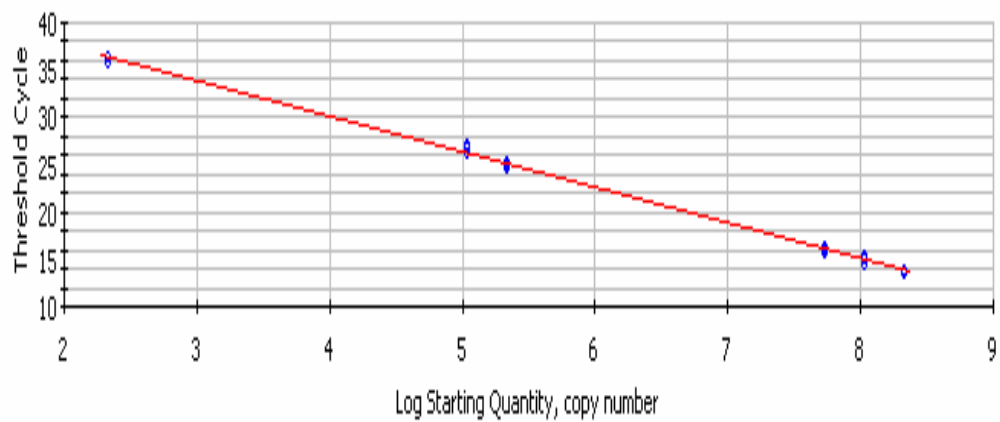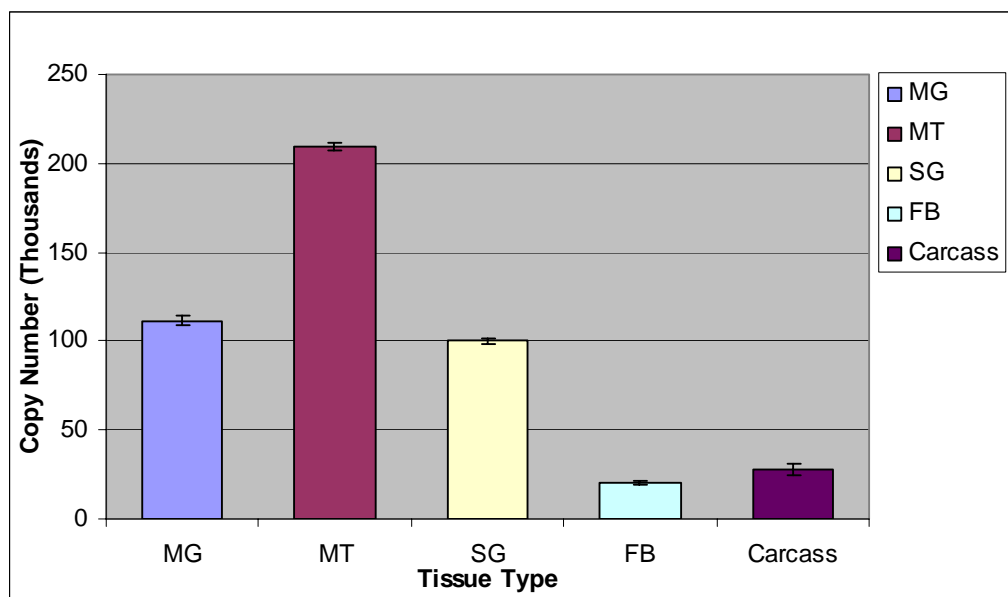
A

B

Figure 19A: Regression plot of threshold cycle against the log starting copy number of Lg2A3 in the pCR2.1-TOPO plasmid. Correlation coefficient– 0.999 Slope: -3.626 Intercept: 49.810, Y = -3.626X + 49.810

Figure 19B: Quantitative expression of Lg2A3 gene in different tissues. MG-Midgut, MT-Malpighian tubule, SG-Salivary gland, FB-Fatbody. Lg2A3 is abundantly expressed in malpighian tubules followed by midgut. Very low levels of expression in salivary glands, fatbody and carcass.
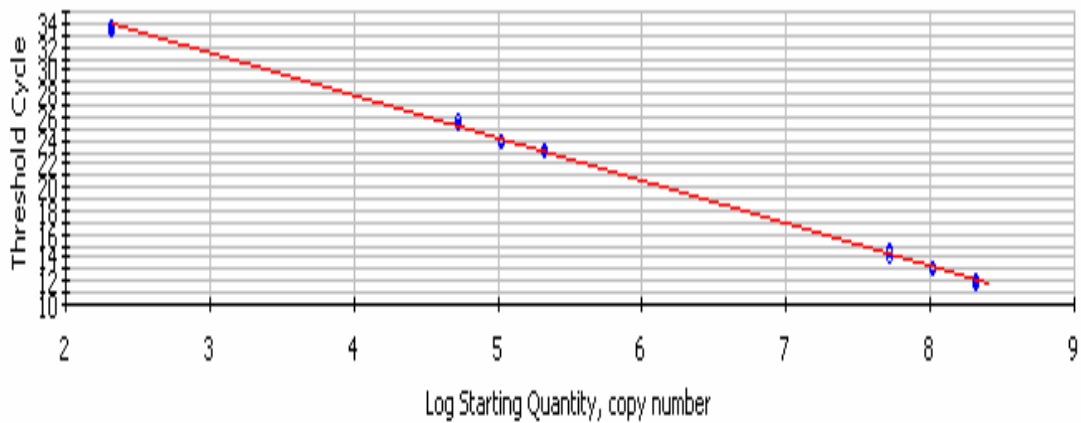
A

B
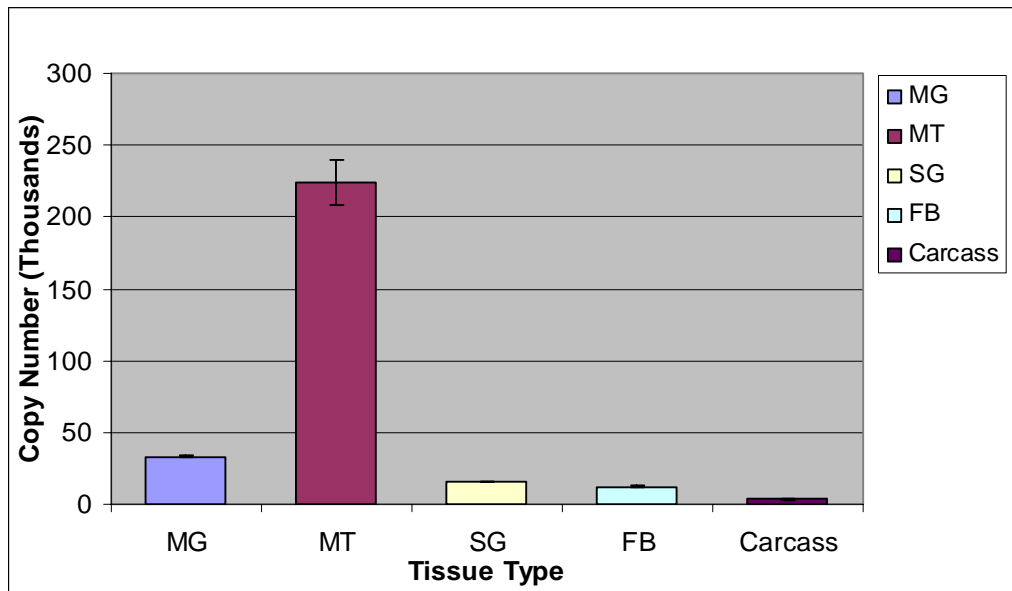
Figure 20A: Regression plot of threshold cycle against the log starting copy number of Lg2F7 in the pCR2.1-TOPO plasmid. Correlation coefficient– 0.999 Slope: -3.821 Intercept: 47.184, Y = -3.821X + 47.184

Figure 20B: Quantitative expression of Lg2F7 gene in different tissues. MG-Midgut, MT-Malpighian tubule, SG-Salivary gland, FB-Fatbody. Lg2F7 is abundantly expressed in malpighian tubules, similar levels of expression in midgut and salivary glands. Very low levels of expression in fatbody and carcass.
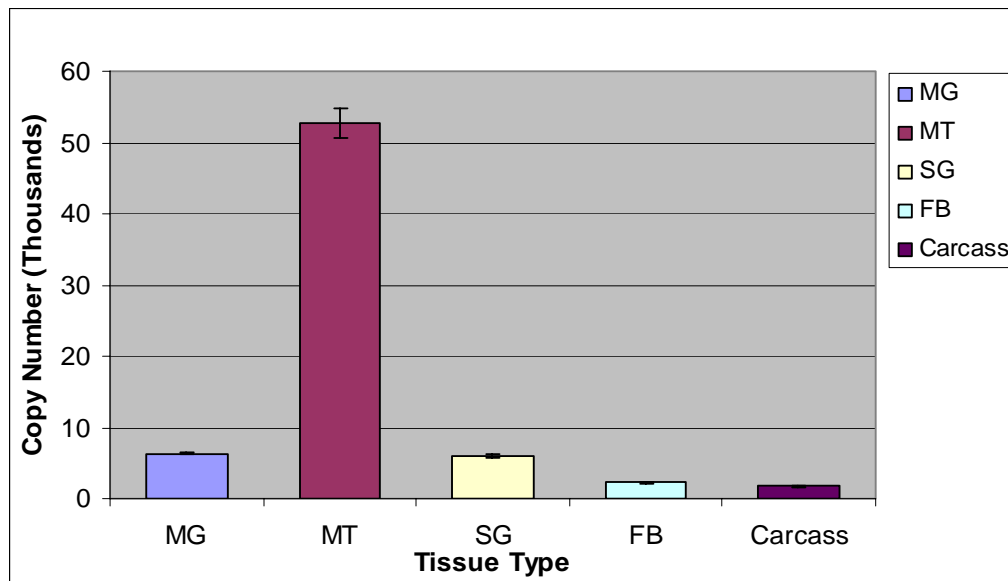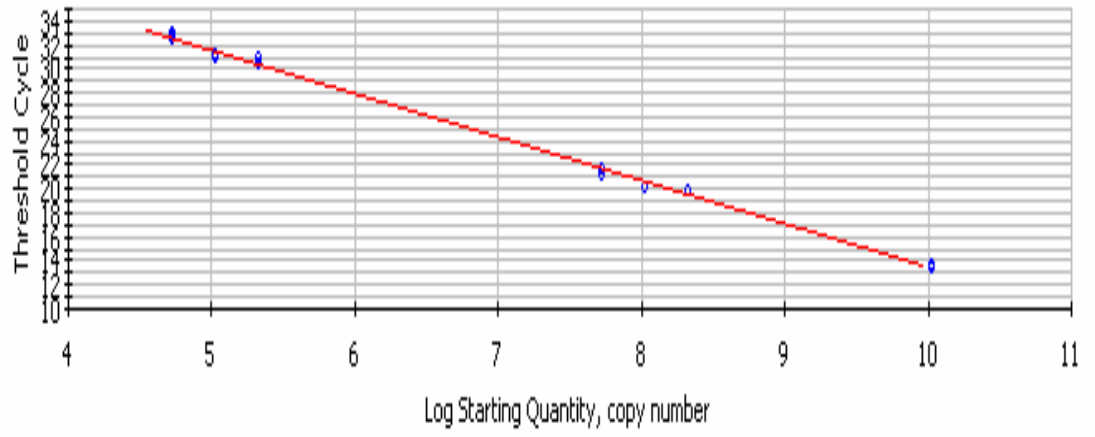
A

B

151
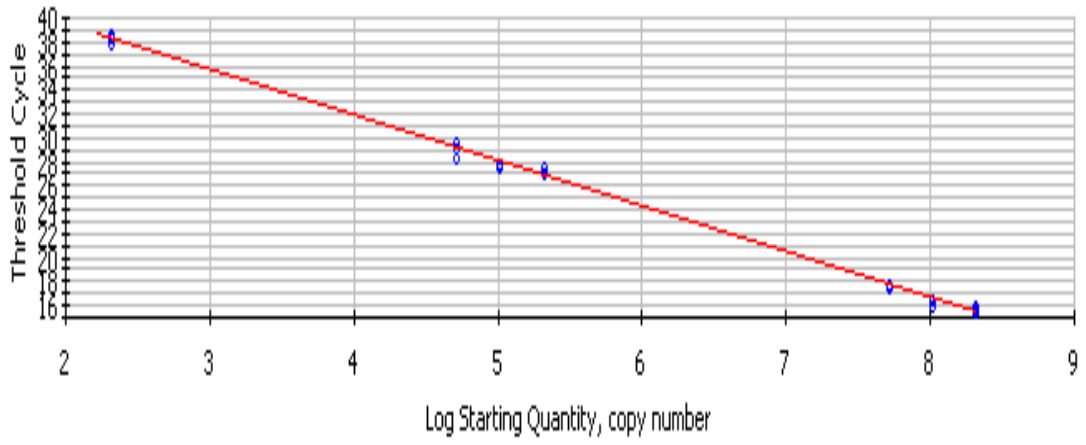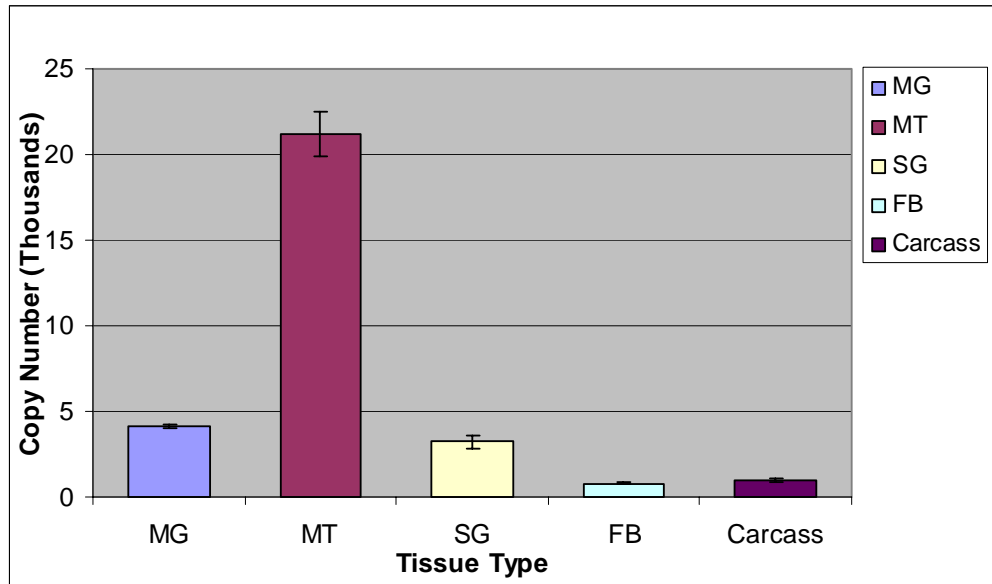
Figure 21A: Regression plots of threshold cycle against the log starting copy number of 11C4 in the pCR2.1-TOPO plasmid. Correlation coefficient– 0.998 Slope: -3.628 Intercept: 42.070, Y = -3.628X + 47.184

Figure 21B: Quantitative expression of 11C4 gene in different tissues. MG-Midgut, MT-Malpighian tubule, SG-Salivary gland, FB-Fatbody. 11C4 is abundantly expressed in salivary glands, similar levels of expression in fatbody and malpighian tubules. Low levels of expression in midgut and carcass.

A
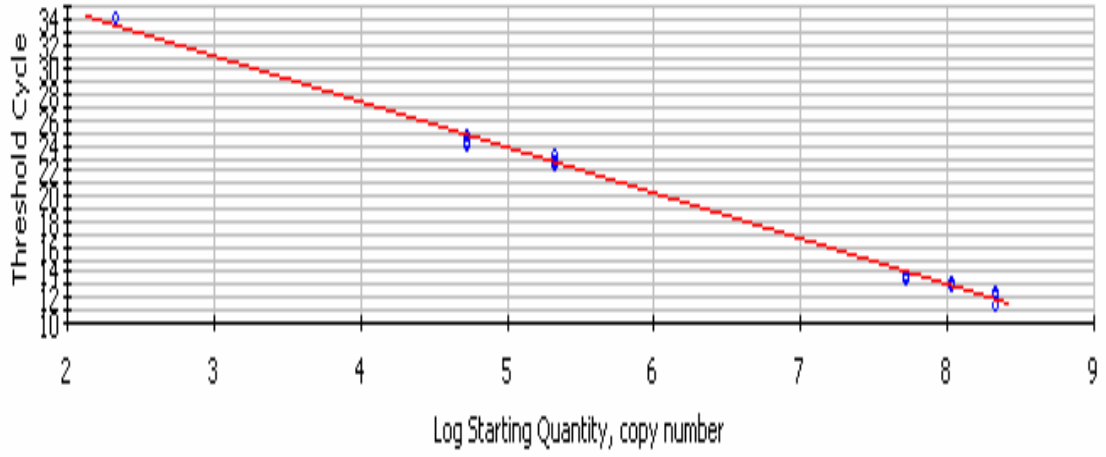
B
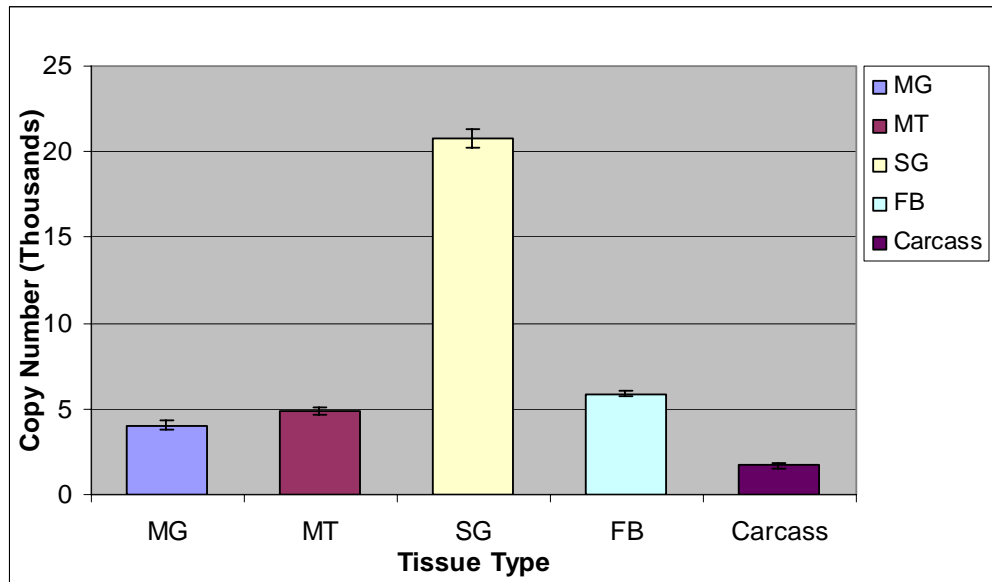
Figure 22: Immunohistochemical localization of 14A4 inhibitor in the malpighian tubule of *Mayetiola destructor*. Primary antibody of 1:1000 and anti rabbit secondary antibody, 1:500, conjugated with Cy-3 was used to localize the 14A4 protein. Arrows indicate the presence of 14A4. No localization of 14A4 protein in the control.
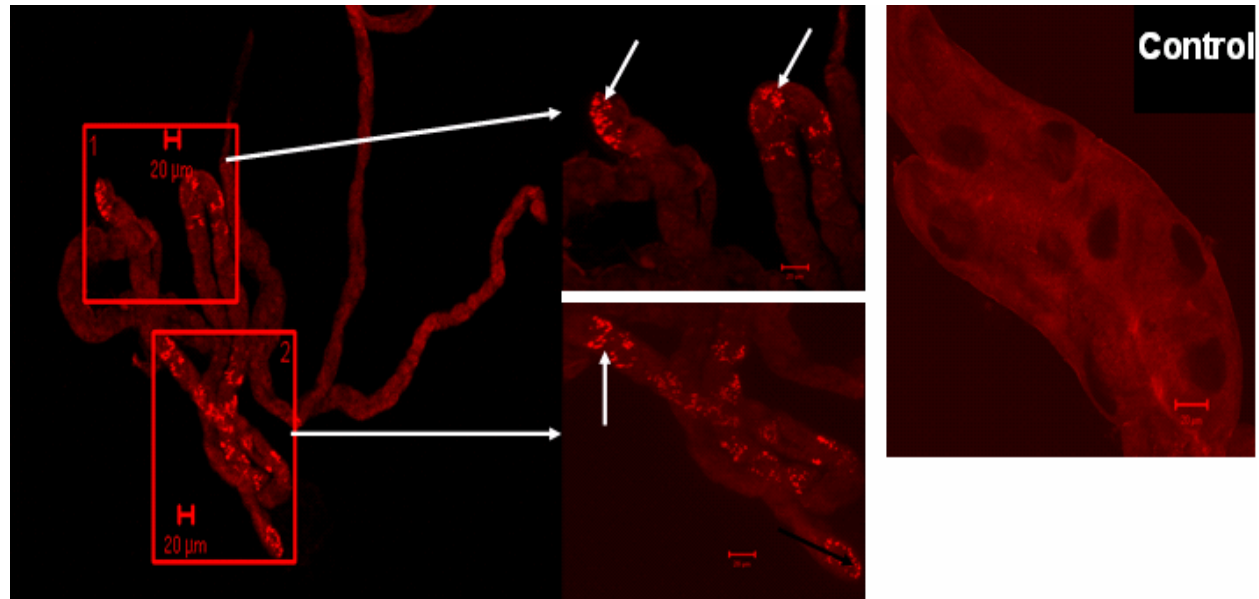
Figure 23: Immunohistochemical localization of 14A4 inhibitor in the midgut of *Mayetiola destructor*. Primary antibody of 1:1000 and anti rabbit secondary antibody, 1:500, conjugated with Cy-3 was used to localize the 14A4 protein. Arrows indicate the presence of 14A4. No localization of 14A4 protein in the control.
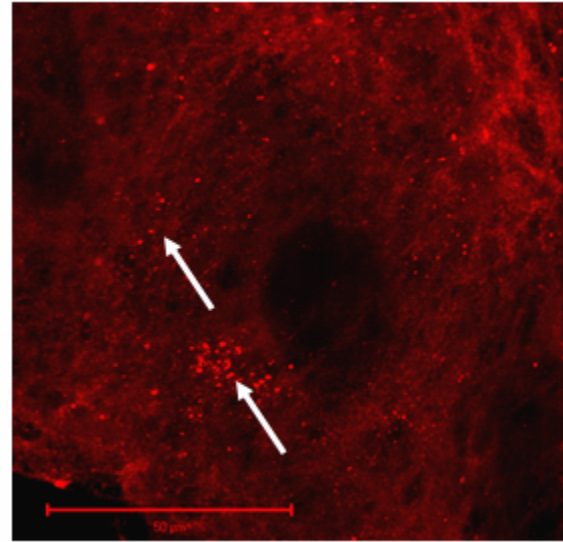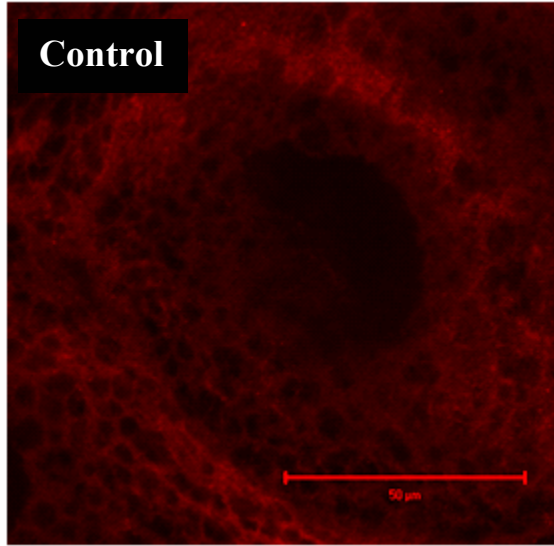
Figure 24: Immunohistochemical localization of 14A4 inhibitor in the salivary glands of *Mayetiola destructor*. Primary antibody of 1:1000 and anti rabbit secondary antibody, 1:500, conjugated with Cy-3 was used to localize the 14A4 protein. Arrows indicate the presence of 14A4. No localization of 14A4 protein in the control. Box A shows both basal and tubular region and box B shows the tubular region of salivary gland.
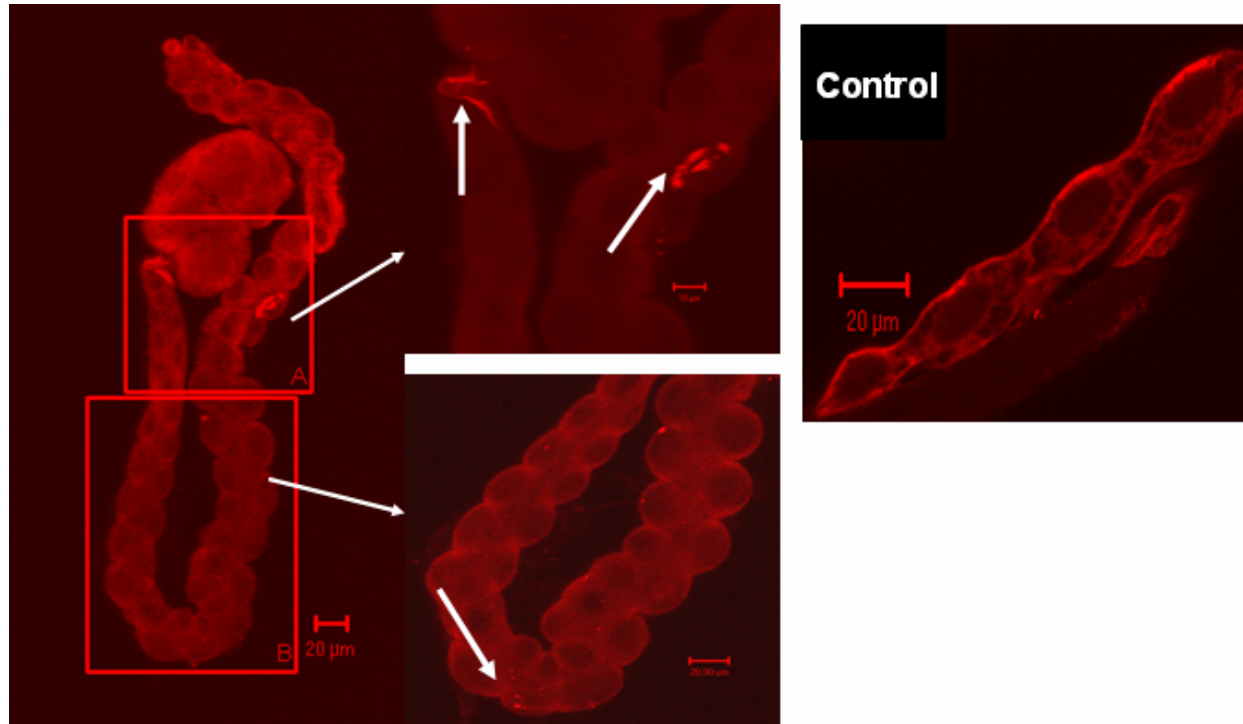
Figure 25: Western analysis to confirmation antibody specificity and tissue specific localization of 14A4 in different tissues of *Mayetiola destructor*. Primary antibody of 1:1000 and anti rabbit secondary antibody, 1:1000, was used to identify the presence of 14A4 protein. HL-Hemolymph, SG-Salivary gland, MGE-Midgut extract, MG-Midgut, RP-Recombinant 14A4.

Figure 26: Identification of bacterial artificial chromosomes (BAC) clones that contain 11A6 cDNA related genes. ^{32}P labeled 11A6 cDNA probe was hybridized with BAC DNA membrane. Arrows represent the positive BAC clones for 11A6.
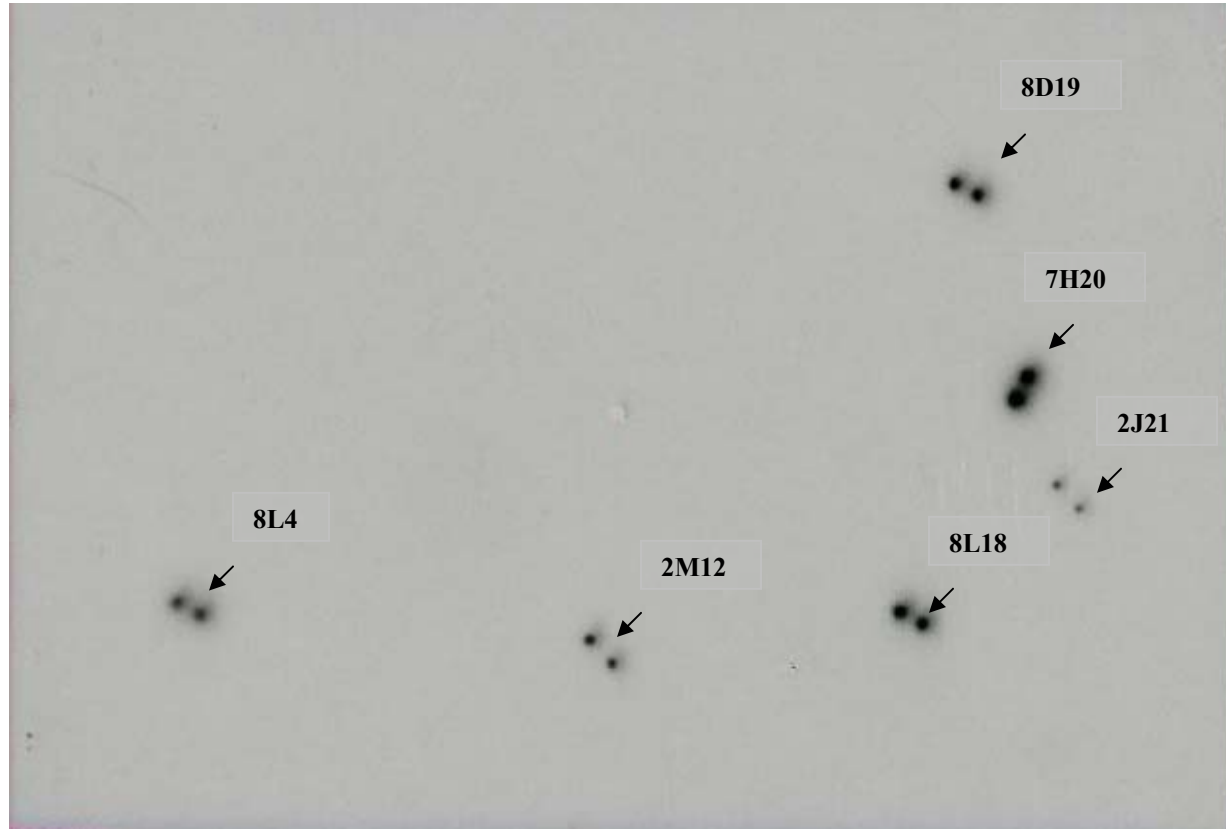
Figure 27: Identification of bacterial artificial chromosomes (BAC) clones that contain 14A4 cDNA related genes. ^{32}P labeled 14A4 cDNA probe was hybridized with BAC DNA membrane. Arrows represent the positive BAC clones for 14A4.
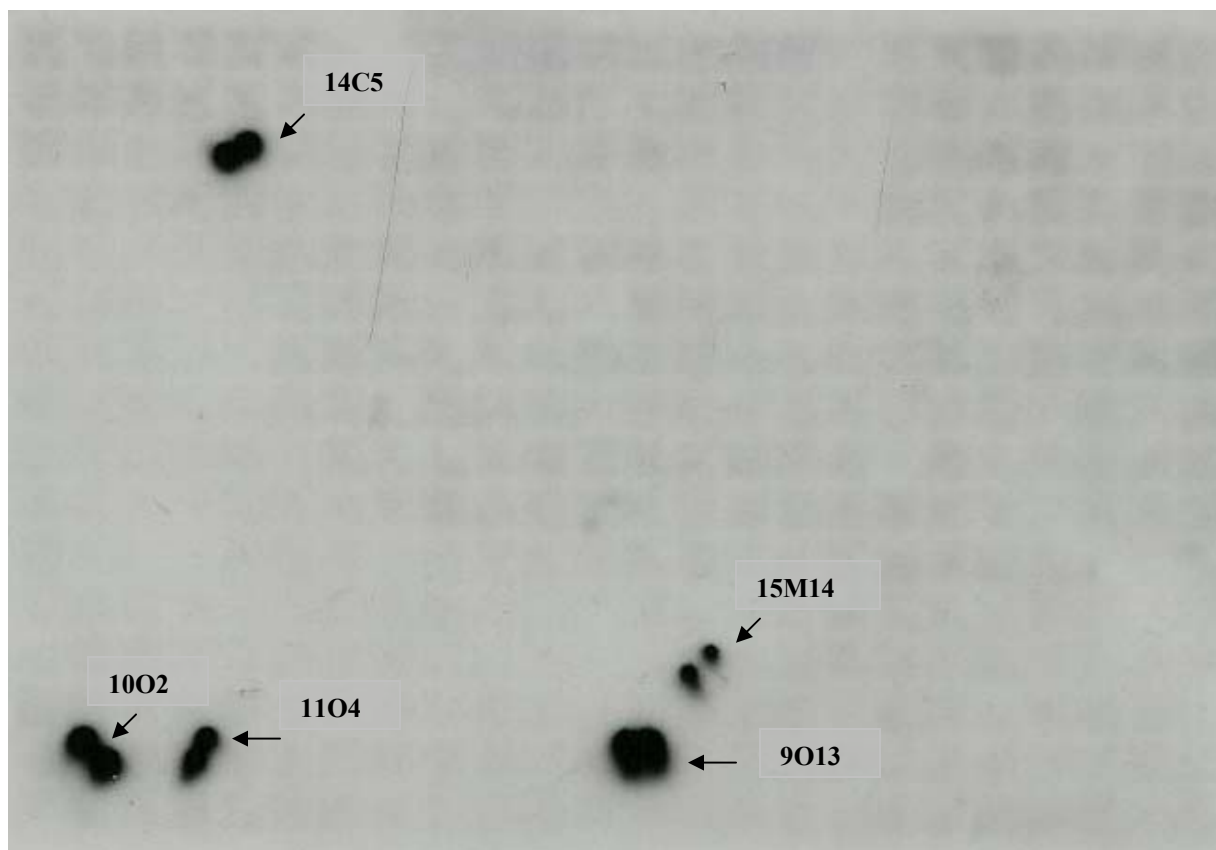
Figure 28: Southern blot analysis of BAC DNA using 11A6 cDNA probe. Five μg of BAC DNA was digested with different restriction enzymes overnight and the products were separated on 1% agarose gel. The agarose separated DNA was transferred on to nylon membrane and hybridized with ^{32}P labeled 11A6 cDNA probe. A- EcoRI digest, B-HindIII digest. Lane 1-8L4 BAC DNA, lane 2-8L18, lane 3-8D19. DNA markers are indicated at the left.
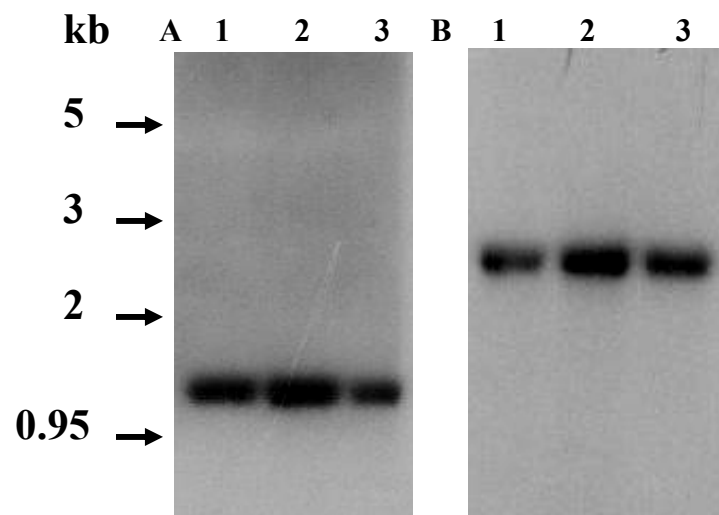
Figure 29: Southern blot analysis of BAC DNA using 14A4 cDNA probe. Five μg of BAC DNA was digested with different restriction enzymes overnight and the products were separated on 1% agarose gel. The agarose separated DNA was transferred on to nylon membrane and hybridized with ^{32}P labeled 14A4 cDNA probe. A-EcoRI digest, B-HindIII digest. Lane 1-10O2, lane 2-11O4, lane 3-14C5, lane 4-15M14 and lane 5-9013. DNA markers are indicated at the left.
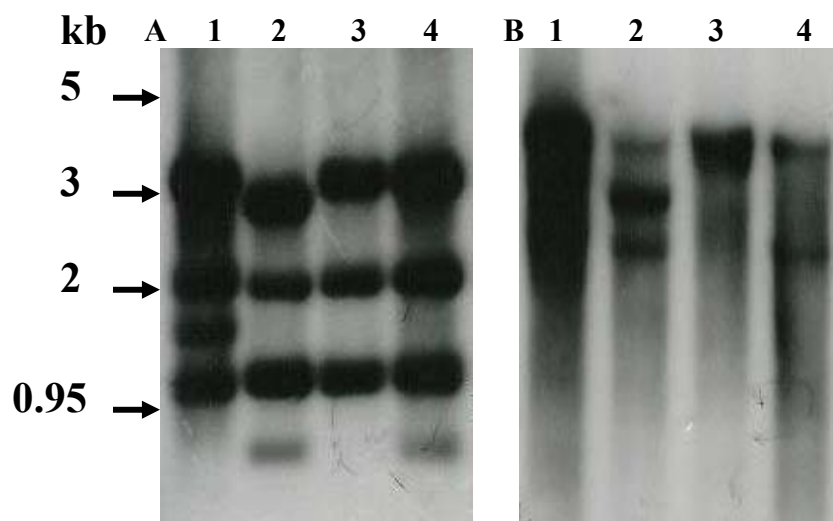
Table 1: Effect on recombinant *Mayetiola destructor* inhibitors on BODIPY-TR-X casein hydrolysis by serine proteases.

Inhibitor	Enzyme	Maximum Inhibition (%)	Molar Ratio	IC_{50} (µM)
11A6	Trypsin	62.11	1.5 : 1	0.52
	Chymotrypsin	-	-	-
14A4	Trypsin	67.46	3.1 : 1	0.90
	Chymotrypsin	54.25	3.2 : 1	2.32

5. Discussion

5.1 Structural similarity to *Ascaris* inhibitors

Fifteen families of serine protease inhibitors have been identified (Reeck et al., 1997). Each family is characterized by the number of cysteine residues and disulfide bonds (Laskowski and Kato, 1980). For example, the pancreatic trypsin inhibitor Kunitz family contains 4 to 8 disulfide bonds. The Bowman-Birk inhibitors contain 7 disulfide bonds. All of the protease inhibitors I have identified belong to the *Ascaris* family, which have five disulfide bonds (Bernard and Peanasky, 1993).

All the identified inhibitors contain cysteine residues that are highly conserved. The cysteine residues present in these proteins might be involved in forming disulfide bonds. Formation of these disulfide bonds could help in optimal folding and maintaining three dimensional structures for the inhibitory activity even with a high degree of diversification in amino acid sequence. Also they might be involved in holding the two chains of the inhibitor by encompassing the reactive site after cleavage by protease (Laskowski and Kato, 1980). This is particularly true in *Ascaris* inhibitors. They entirely rely on disulfide bonds for proper folding because they lack a definite hydrophobic region (Bernard and Peanasky, 1993; Cierpicki et al., 2000; Zang and Maizels, 2001), and at least one disulfide bond encompasses the reactive site (Bernard and Peanasky, 1993).

Although the identified inhibitors contain cysteine residues, the number of cysteine residues varies from 7 to 10. This variation in the number did not affect the protease inhibitory activity of recombinant inhibitors 11A6 and 14A4, suggesting that the inhibitors can sustain the variation in the number of cysteine residues without altering the protein scaffold required for their activity. The *Ascaris* members, including the insect

171

inhibitors *Apis mellifera* chymotrypsin/cathepsin-G inhibitor (AMCI-1) and *Drosophila* Acp62F, contain 10 cysteine residues (Bearnard and Peanasky, 1993; Bania et al., 1999; Lung et al., 2002). However, some members like von Willebrand factor contain 11 cysteines, mouse mucin-6 has 8 cysteines and *Anisakis simplex* coagulation inhibitor, Ax-SPI-3 has 9 cysteine residues (Lu et al., 1998; Katsumi et al., 2000; Escande et al., 2004). The von Willebrand factor forms 5 intra and 1 inter chain disulfide bonds (Katsumi et al., 2000). In isolated protease inhibitors, 7 cysteine residues are conserved with *Ascaris* members. Out of these 7 residues the C-terminal 4 cysteine residues are completely conserved, suggesting that these residues might be absolutely required for proper folding of the protein. In *Ascaris* family members not only are the 10 cysteine residues conserved but the patterns of the disulfide bonds are also completely conserved (Bearnard and Peanasky, 1993).

The isolated protease inhibitors contain a cysteine residue at P_2 which is highly conserved with the P_2 cysteine of *Ascaris* inhibitors (Bernard and Peanasky, 1993). In contrast to this, the cysteine residue present at P_2' position in Ascaris inhibitors is located at P_3' position in inhibitors identified from Hessian fly. This change in cysteine position is possible if the change does not alter the substrate binding to the enzyme (Bode and Huber, 1991; Stanssens et al., 1996). The P_2' located cysteine in Ascaris inhibitors forms a disulfide bond, covering the reactive site, with a cysteine located at N-terminal 22nd position (Bearnard and Peanasky, 1993). Hence, it is possible that the P_3' located cysteine might be involved in forming a disulfide bond encompassing the reactive site in these inhibitors.

5.2 Diversity in protease inhibitors

The identified cDNA clones were categorized into five groups based on their evolutionary relationship. The five groups of cDNA clones identified possibly code for protease inhibitors. Individual members within the same group encode similar proteins which could represent different alleles because the cDNA library was made from multiple Hessian fly insects. Alternatively, these individual cDNAs could represent different genes that arose by gene duplication. Protease inhibitor genes that occur in multiple copies are known to have arisen by gene duplication (Ohta, 1994; Prakash et al., 1996; Mukhopadhyay, 1999). Indeed, the cDNAs isolated from the Hessian fly salivary glands and midguts are usually present in multiple copies which are clustered within a short region of the chromosome (Chen et al., 2004). Clustered organization of protease inhibitor genes has been reported in plants and animals (Heibges et al., 2003; Puente and Lopez-Otin, 2004).

Individual members within the same group are highly conserved whereas members from different groups encode diversified proteins. The sequence identity between different members ranged from 25% to 58% (data not shown). The maximum overall identity of protease inhibitors with known members of *Ascaris* inhibitors is 34%. Out of this 34% identity, 13% is contributed by the complete conservation of cysteine residues whereas the percentage of cysteine content of *Ascaris* members varies from ~8% to ~18% (Babin et al., 1984; Katsumi et al., 2000). Without the cysteine residues the overall identity of protease inhibitors with *Ascaris* members is 21%, which is close to that among random sequence alignments. This suggests that the proteins might have undergone divergent evolution leading to diversification while maintaining functional

integrity. This is particularly the case in Ascaris inhibitors. The *Ancylostoma ceylanicum* anticoagulant protein, AceAp1, is 40-44% identical to two inhibitors, AcApc2 and AcAP5, of *A. caninum*, but all of them have the fundamental protein backbone of *Ascaris* inhibitor family (Harrison et al., 2002). It is believed that divergent evolution of protease inhibitors has taken place to counteract the rapidly evolving cognate proteases (Zang and Maizels, 2001). Even though they have undergone divergent evolution over a period of time, their inhibitory activity is maintained without loosing the cysteine residues at their respective positions (Harrison et al., 2002; Zang and Maizels, 2001).

5.3 Diversity in reactive center region

The reactive center region residues on either side of P_1, except the P_2 and P_3' cysteine, are highly variable. This variation in the reactive center region leads to variation in their specificity towards their target proteases (Laskowski and Kato, 1980; Stanssens et al., 1996). This is true because the recombinant proteases exhibit different specificities towards commercial proteases. Also the *Ascaris* family hookworm anticoagulant inhibitors, AcAPc2 and AcAP5 exhibit different specificity. Even though both contain a P_1 arginine, AcAP5 inhibits fXa, a trypsin like enzyme, whereas AcAPc2 does not. This is due to the variability of amino acid residues at the reactive center region (Stanssens et al., 1996). Similarly with the locust inhibitory peptides, PMP-D2 and HI, both contain an arginine specific for trypsin at P_1, but they inhibit chymotrypsin (Kellenberger et al., 2003). The variation in the reactive center region is the common phenomenon found in protease inhibitors (Laskowski and Kato, 1980; Hawley and Peanasky, 1992; Gettins,

174

2002) which is due to the occurrence of mutations in the reactive center region that have taken place to overcome the evolving target proteases (Nielsen et al., 1998). High variation in the amino acid residues at the reactive center region has been observed in functionally important domains of various serine protease inhibitors (Babin et al., 1984; Hill and Hastie, 1987; Laskowski Jr. et al., 1987; Jiang et al., 1994; Goodwin et al., 1996).

In the reactive center region the P_1 residue is also highly variable. The group I proteins contain a predicted P_1 residue aspartic acid which is different from basic and aromatic amino acids present in other members. The inhibitors with basic amino acids lysine and arginine inhibit trypsin (Laskowski and Kato, 1980). The P_1 aspartic acid residue is not specific to trypsin inhibition. It is not known how the P_1 aspartic acid reacts with active site residues of cognate enzyme. Variation in the P_1 reactive site is known to occur in inhibitors (Laskowski and Kato, 1980). For example, the pancreatic trypsin inhibitor family members contain one of the following residues- arginine, lysine, alanine, leucine, aspartic acid, methionine, serine, valine, glutamine, or tyrosine as a P_1 residue (Laskowski and Kato, 1980). The insect double headed Kazal inhibitor, rhodniin, contains aspartic acid as a P_1 residue in its second inhibitor domain, Rhod2, whereas the first domain, Rhod1, contains histidine as a P_1 residue (van de Locht et al., 1995).

The members in group II except Sg6G2, contain lysine residue at P_1 and the members in group III and IV contain lysine and arginine, respectively, at P_1. The members with lysine and arginine inhibit trypsin (Laskowski and Kato, 1980). As demonstrated by recombinant protein 14A4, the protein inhibited trypsin activity, suggesting that it is a trypsin inhibitor. In addition to inhibiting trypsin it also inhibited

chymotrypsin activity. Evaluation of the reactive center region residues revealed that the 14A4 protein contains a hydrophobic residue proline at P_4 position which is not found in group I and V proteins. Although the P_1 residue confers specificity, it is suggested that the residues between P_3 and P_{10} are also responsible in binding interactions of Kunitz inhibitors (Kraunsoe et al., 1996; Pritchard and Dufton, 1999; Milestone, 2000). Also, the Kunitz type serine protease inhibitor secreted by *A. ceylanicum* is known to inhibit chymotrypsin, trypsin, pancreatic and neutrophil elastase with its P_1 residue methionine because it contains two extra residues at P_3 and P_{12} (Milestone et al., 2000). It is possible that the presence of proline and also other residues in this region might be influencing the interaction of 14A4 inhibitor with different proteases.

In contrast to the presence of lysine at P_1 in all other members in group II, Sg6G2 contains an isoleucine, a hydrophobic amino acid at the P_1 site. The isoleucine at P_1 site is also present in a serpin, MEPI (Gettins, 2002), but it not known how the P_1 isoleucine interacts with the active site residues of the target enzymes. The chymotrypsin enzyme is known to selectively cleave peptide bonds on the carboxyl side of hydrophobic or aromatic amino acids (Lu et al., 1997).

Even though group I (11A6) proteins have an aspartic acid, which can follow canonical mechanism of inhibition, the molar ratio suggests otherwise. The stiochiometry ratio of inhibitor to trypsin enzyme is greater than 1:1. The anticoagulant inhibitor contains two domains, Rhod1 and Rhod2. Rhod1 contains histidine as P_1 whereas Rhod2 contains aspartic acid (van de Locht et al., 1995). The Rhod1 forms a reactive site loop which follows the canonical conformation whereas the Rhod2 reactive site loop does not follow this conformation even though the rhodniin molecule binds at a 1:1 ratio to the

thrombin enzyme. This is because the Rhod2 binding loop contains 3 negatively charged residues including P_1 aspartic acid and 1 positively charged residue in addition to the presence of the 1-amino acid insertion (Friedrich et al., 1993). The Rhod1 binds to the active site of the thrombin enzyme but Rhod2 loop binds to the exosite (van de Locht et al., 1995). It might be a similar case as that of Rhod2 in 11A6 inhibitor which contains 2 negatively charged residues at the P_1 and P_2 site and 1 positively charged residue at the P_1' site which can restrict canonical loop formation. Hence 11A6 inhibitor might inhibit the proteases noncanonically by binding noncovalently which is also suggested by its molar ratio of inhibition. The molar ratio of 14A4 inhibitor is also greater than 1:1 which implicates that this inhibitor also inhibits the proteases by binding noncovalently. It could be true because binding of inhibitors to proteases did not prevent the hydrolysis of casein conjugated fluorescence substrate completely. Bothrojaracin, isolated from the *Bothrops jararaca*, exhibits inhibitory effects by binding to the anion-binding site but does not substantially decrease the amidolytic activity of thrombin (Zingali et al., 1993). One explanation for not inhibiting the protease activity completely is that binding of inhibitors noncovalently to protease might sterically hinder the access of substrate to the active site of protease. Another possibility is that binding of inhibitors to protease might change the conformation of proteases resulting in lower activity.

5.4 Cross specificity towards different proteases

The recombinant protease inhibitors exhibited different specificities towards different proteases. Protease inhibitor 11A6 inhibited trypsin and also papain at lower levels, whereas 14A4 inhibited trypsin and chymotrypsin, suggesting that these

recombinant inhibitors posses different specificities. This difference could be due to the variation in the reactive center region amino acid residues that impart specificity. The hookworm anticoagulant inhibitors have different specificities towards their target proteases due to variation in their reactive center region (Stanssens et al., 1996). It is possible that the recombinant inhibitors might have domains that could inhibit more than one protease. The human LEKT1 has 15 domains and inhibits plasmin, trypsin, elastase, subtilisin A and cathepsin G (Mitsudo et al., 2003). Also, the 11A6 and 14A4 recombinant inhibitors might act more as substrates for chymotrypsin and papain proteases, respectively, because no inhibition was observed. It is known that serpin-3 acts as more substrate for chymotypsin whereas it inhibits propehnoloxidase, trypsin, elastase and plasmin completely (Zhu, 2001). This difference in protease specificities is likely caused by conformational effects of peptide bonds in proteases and inhibitor at the contact region (Narumi et al., 1993).

In addition to inhibiting the commercial proteases, the two recombinant proteins also inhibited Hessian fly gut protease activity with different levels. The Hessian fly midgut contains digestive proteases which are predominantly serine proteases, trypsins and chymotrypsins (Zhu et al., 2005). The gut extract proteases that are not specific to inhibition by recombinant proteins may use recombinant proteins as substrates at different capacities. Hence, this could lead to differences in levels of inhibition by two recombinant proteins.

Further, both recombinant inhibitors inhibited gut protease activity of yellow mealworm and red flour beetle. The inhibition of commercial papain by 11A6 recombinant inhibitor was less than the inhibition of gut protease activity. No significant inhibition of papain by 14A4 recombinant was observed. Both the insect protease extracts contain cysteine proteases (Thai and Houseman, 1990; Oppert et al., 2005). Yellowmeal worm, in addition to its predominant cysteine proteases activity, also contains serine protease activity (Thai and Houseman, 1990). The ability to inhibit cysteine proteases in addition to inhibiting serine proteases suggests that the recombinant proteins are cross-class inhibitors. This cross class inhibition observed by recombinant inhibitors could be due to the conformation of the inhibitors that might be well suited to inhibit cysteine proteases. Another possibility is that the inhibitors might contain the reactive P_1 site residue that might be recognized by the cysteine proteases which could lead to their inhibition after cleavage. The endopin 2 contains P_1 recognition site that are distinct for papain and elastase. Hence it inhibits both these enzymes (Hawng et al., 2002). Cross class specificity is also observed in a serpin, squamous cell carcinoma antigen 1 (SSCA 1) that inhibits lysosomal cysteine proteases, cathepsins K, L, and S (Schick et al., 1998) and CrmA that inhibits ICE/CED-3-like cysteine proteases (Ray et al., 1992). It is not known why the recombinant inhibitors inhibited gut cysteine proteases to a higher extent than the commercial enzyme papain, but it is possible that the recombinant inhibitors might be more specific to insect cysteine proteases than to commercial papain.

5.5 Differential expression pattern of genes

Except for group V (11C4) all cDNAs are predominantly expressed in the larval stage. The expression patterns are quite different. This suggests that cDNAs might have different functional targets and specificities towards their cognate enzymes that are involved in various biological processes. The cDNAs from group II, III and IV are expressed during the initial developmental stages of larva. Possibly these cDNAs have roles during early development stages of the larva. In mosquitoes, chymotrypsin inhibitors are known to occur at different developmental stages and play a role in the developmental process (Horler and Briegel, 1997).

In contrast to the above four groups of cDNAs, group V was abundantly expressed in late larval and pupal stage. It is possible that these cDNAs might be regulating proteases present at the pupal stage. Thus they might facilitate pupation and may have a role in insect developmental processes. A similar kind of observation is described in *Drosophila*. The *Drosophila* kil-1 gene is expressed during prepupal stage and regulates pupal proteolytic activity (Kress et al., 2004). Expression of kil-1 inhibitor in prepupa coincides with decreased protease activity in the molting fluid. Expression of serine protease inhibitors in larvae, prepupae and pupae is documented in *Aedes*, and *Drosophila* (Horler and Briegel, 1997; Kress et al., 2004).

5.6 Tissue specific expression

All the cDNA groups except group V are abundantly expressed in malpighian tubules. This abundant gene expression in malpighian tubules coincides with the localization of group II protein 14A4, suggesting that malpighian tubules are the major

site of synthesis. Malpighian tubules are organs that are involved in urine secretion and excretion of organic metabolites (Sözen et al., 1997). Hessian fly have three malpighian tubules. All three tubules merge at single point and are connected to the midgut at the posterior end. The immunolocalized protein in the malpighian tubules is present predominantly in the cells located at the tip and middle region of the tubule. *Drosophila* has of two pair (anterior and posterior) of malpighian tubules which consist of two tubules. Each tubule pair is divided into six regions: the initial segment present at the tip, transition segment, very long middle segment, lower tubule, upper ureter, and lower ureter (Sözen et al., 1997). These different segments are composed of different cell types and are genetically distinct. The first three regions are composed of principal and stellate cells where as the last three regions are composed of very tiny cells. These tiny cells secrete neurohormones into the hemolymph that regulate muscle contractility or ion transport (Sözen et al., 1997). Our speculation is that the cells located at the tip and middle regions of malpighian tubules are involved in synthesizing and secreting inhibitors into the hemolymph and to other parts of the insect body to regulate proteolytic processes. Indeed, the presence of 14A4 inhibitor was confirmed by the localization in the hemolymph (Fig. 25). Hence, these inhibitors may function as extracellular proteins.

Hessian fly malpighian tubules are freely floating in the hemolymph where various serine proteases are present. These protease inhibitors might play a role in protecting the malpighian tubules from these destructive proteases by regulating their activity inside malpighian cells or near the vicinity. Alternatively, they might be involved in regulating the hydrolytic proteases in their own cells but the presence of proteolytic enzymes in Hessian fly malpighian tubules has not been demonstrated. The expression of

an aminopeptidase gene and the presence of protease activities are documented in the malpighian tubules of *Trichoplusi ni* (Wang et al., 2005).

In addition to malpighian tubules, expression was observed in the midgut but to a lesser extent. The synthesized proteins in the midgut could be secreted into the gut extract and hemolymph in addition to their presence in the midgut tissue. Localization of group II protein was confirmed in the midgut cells, and further shown by Western analysis that these proteins are present in midgut extract and hemolymph. Serine proteases are part of the prophenoloxidase cascade system in the insect hemolymph (Ashida, 1990; Soderhall and Cerenius, 1998). The inhibitors secreted into the hemolymph could regulate the hemolymph proteolytic activity of proteases and thus may have a role in the insect defense system. Alternatively, in Hessian fly digestive processes occur in the midgut. The midgut contains primarily serine proteases (Zhu et al., 2005). These inhibitors might be involved in protecting midgut cells from destructive protease activity.

Apart from malpighian tubules and midgut, proteases are also expressed in salivary glands. It is possible that synthesized proteins are secreted outside the glands in other parts of the insect body to regulate proteases. No protease activity was detected when one pair of salivary gland extract was used for the activity assay. Also the salivary gland structure is modified in such a way that the tip of the basal region is connected to mouth parts and the filament region is free and extends into the hemocele. The protein synthesized in the salivary gland could be secreted into the hemolymph. Indeed not much of the group II (14A4) protein was detected in the salivary glands (Fig. 25). Also, most of the cDNAs isolated from Hessian fly salivary glands and midgut encode secreted proteins

(Liu et al., 2004, Chen et al., 2004). Together midgut and salivary glands may act as secondary sites of synthesis of these proteins.

In contrast to the above four groups of cDNAs, group V cDNAs are abundantly expressed in salivary gland followed by fatbody. Northern data shows group V cDNAs are expressed in late larval and pupa stages. The proteins synthesized in salivary glands probably are secreted outside the salivary glands to regulate the activity of proteases that occur during pupal stages (Kress et al., 2004). Hence these proteins might have a role in pupa formation. A similar kind of expression pattern of kil-1 gene has been noticed in *Drosophila*. The KIL-1 protein is synthesized in prepupal salivary glands and released into the molting fluid during pupation, which coincides with the loss of protease activity in pupa (Kress et al., 2004). Group V expression data is similar to that of KIL-1, suggesting that inhibitors are involved in developmental processes. The group V cDNAs, in addition to their predominant expression in salivary glands, are also expressed in fatbody. In Hessian fly, the fat body is present in the hemocele region. The inhibitors synthesized in the fat body might be released into insect hemolymph and may participate in regulating the proteases of prophenoloxidase cascades. In *Manduca*, the fatbody synthesized serpins play an important role in regulating the serine proteases of prophenoloxidase cascades (Zhu, 2001; Tong, 2005).

5.7 Gene organization

Southern analysis of positive BAC clones with group I (11A6) probe identified one band, suggesting this is single copy gene in the BAC. Whether this gene is a single copy in the whole genome remains to be determined. However, the screening of BAC library generated 6 BAC clones. The BAC library constructed contains 6144 clones. If the Hessian fly genome is ~160 Mb (personal communication with Dr. Jeff Stuart), then 1066 BAC clones represent 1 Hessian fly genome. Hence 6144 BAC clones approximately represent 6(5.8) Hessian fly genomes. Assuming that the BAC clones positive to 11A6 probe were normally represented, then our data is consistent with the notion that the group I gene is either a single copy gene or a low copy number gene in the genome since as expected 6 BAC clones were identified from the BAC screening that represents the ~6 times Hessian fly genome.

In contrast, multiple bands of group II (14A4) genes ranging from one to five were observed in the BACs. This suggests that these genes occur in multiple copies in the BAC and hence in the genome. This is true because the nucleotide identity between members range from 83% - 95% (data not shown). Also the fact that a few members in this group contain additional peptide segments not present in other proteins indicates that the group II members represent genes that occur in clusters in the BAC. The serpin genes in rats exist in three clusters (Puente and López-Otín et al., 2004). Multiple copies of serine protease inhibitor genes have been documented in many other organisms (Potempa et al., 1994; Lu et al., 1998; Yan et al., 2001; Heibges et al., 2003). Presence of multiple copies of inhibitor genes in the genome is an added advantage because each individual

gene can undergo evolution independently and so can counter the evolving target proteases in a better way (Zang and Maizels, 2001).

6. Conclusions

➢ The identified cDNAs code for diversified proteins with highly conserved cysteine residues suggesting the maintenance of functional integrity during evolution.

➢ The recombinant inhibitors 11A6 and 14A4 inhibited different proteases suggesting that the cDNA encoded proteins indeed are protease inhibitors with different specificities.

➢ The representative genes are predominantly expressed in first instar stage and their expression patterns are highly differential suggesting that inhibitors may have different specificities and functional roles.

➢ Group I-IV genes were predominantly expressed in the malpighian tubules whereas group V was found in the salivary glands. This implicates that the proteins are exclusively synthesized in the malpighian tubules and salivary glands respectively.

➢ The group II protein, 14A4, is abundantly localized in the malpighian tubules suggesting that the primary route of synthesis and secretion of group protein is malpighian tubules and might regulate different functions spatially.

➢ The analysis of BACs for group I and group II genes revealed single and multiple bands respectively implying that the group I is a single or low copy number gene whereas group II is a multi copy gene in the genome.

7. Reference

Applebaum, S. W. (1985) Biochemistry of digestion. In Comparative Physiology, Biochemistry and Pharmacology of Insects (Edited by Kerkut, G.A. and Gilbert, L.I.), pp. 279-311, Pergamon Press, Oxford.

Armstrong, R. C., Aja, T., Xiang, J., **Gaur**, S., **Krebs**, J. F., Hoang, K., Bai, X., **Korsmeyer, S. J.**, **Karanewsky**, D. S., **Fritz, L. C. and Tomaselli**, K. J. (1996) Fas-induced activation of the cell death-related protease CPP32 is inhibited by Bcl-2 and by ICE family protease inhibitors. *J. Biol. Chem.* **271**: 16850-16855.

Ashida, M. (1990) The prophenoloxidase cascade in insect immunity. *Res. Immunol*. 141: 908-910.

Aspan, A., Sturtevant, J., Smith, V. J. and Soderhall, K. (1990) Purification and characterization of a propehnoloxidase activating enzyme from crayfish blood cells. *Insect Biochem*. **20**: 709-718.

Azzolini, S, S., Santos, J. M. C., Souza, A. F., Torquato, R. J. S., Hirata, I. Y., Andreotti, R., and Tanaka, A. S. 2004. Purification, characterization, and cloning of a serine proteinase inhibitor from the ectoparasite *Haematobia irritans irritans* (Diptera: Muscidae). *Exp. Parasitol.* **106**: 103-109.

Babin, D. R., Peanasky, R. J. and Goos, S. M. (1984) The isoinhibitors of chymotrypsin/elastase from *Ascaris lumbricoides*: the primary structure. *Arch. Biochem. Biophy*. **232**:143-161.

Baker, J. E. (1981) Resolution and partial characterization of the digestive proteinases from larvae of the black carpet beetle. In Current Topics in Insect Endocrinology and Nutrition. Ed. by Bhaskaran, G., Friedman, S. and Rodriguez, J. G. Plenum Press, New York. pp 283-315.

Balbin, M., **Fueyo, A**., **Knauper, V**., **Lopez, J. M**., **Alvarez, J**., **Sanchez, L. M**., **Quesada, V**., **Bordallo, J**., **Murphy, G**. **and Lopez-Otin, C**. (2001) Identification and enzymatic characterization of two diverging counterparts of human interstitial collagenase (MMP-1) expressed at sites of embryo implantation. *J. Biol. Chem.* **276**: 10253-10262.

Bania, J., Stachowiak, D. and Polanowski, A. (1999) Primary structure and properties of the cathepsin G/chymotrypsin inhibitor from the larval hemolymph of *Apis mellifera*. *Eur. J. Biochem.* **262**: 680-687.

Barrett, A. J. (1999) Peptidases: a view of classification and nomenclature. In Proteases New Perspectives Ed. by Turk, V. Birkhauser. pp 1-12.

Baumann, E. (1990) Isolation and partial characterization of a chymotrypsin-like endoprotease from cockroach intestinal system. *Insect Biochem.* **20**: 761-768.

Berger, A. and Schechter, I. (1970) Mapping the active site of papain with the aid of peptide substrates and inhibitors. *Phil. Trans. R. Soc. London* **B257**: 249-264.

Bernard, V. D. and Peanasky, R. J. (1993) The serine protease inhibitor family from *Ascaris suum*: chemical determination of the five disulfide bridges. *Arch. Biochem. Biophy.* **303**: 367-376.

Bernardi, R., Tedeschi, G., Ronchi, S. and Palmieri, S. (1996) Isolation and some molecular properties of a trypsin-like enzyme from larvae of European corn borer *Ostrinia nubilalis* Hübner (Lepidoptera: pyralidae). *Insect Biochem. Mol. Biol.* **26**: 883-889

Bigler, T. L., Lu, W., Park, S. J., Tashiro, M., Wieczork, M., Wynn, R. and Laskowski-Jr. M. (1993) Binding of amino acid side chains to preformed cavities: Interaction of serine proteinases with turkey ovomucoid third domains with coded and noncoded P(1) residues. *Protein Sci.* **2**: 786-799.

Birk, J. (2003) Plant protease inhibitors: Significance in nutrition, plant protection, cancer prevention, and genetic engineering. Published by Springer pp 111-120.

Bode, W. and Huber, R. (1991) Proteinase-protein inhibitor interaction. *Biomed. Biochim. Acta* **50**: 437-446.

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

Boigregrain, R-A., Mattras, H., Brehelin, M., Paroutaud, P., and Coletti-Previero, M-A. (1992) Insect immunity: Two proteinase inhibitors from hemolymph of *Locusta migrtoria. Biochem. Biophy. Res. Comm.* **189**: 790-793.

Boigegrain, R., Pugniere, M., Paroutaud, P., Castro, B., and Brehelin, M. (2000) Low molecular weight serine protease inhibitors from insects are proteins with highly conserved sequences. *Insect Biochem. Mol. Biol.* **30**: 145–152.

Bolter, C. J. and Jongsma, M. A. 1995. Colorado potato beetles (*Leptinotarsa decemlineata*) adapt to proteinase-inhibitors induced in potato leaves by methyl jasmonate. *J. Insect Physiol.* **41**: 1071

Brenner, S. (1988) The molecular evolution of genes and proteins: a tale of two serines. *Nature* **334:** 528-530.

Bussey, H. (1988) Proteases and the processing of precursors to secreted proteins in yeast. Yeast 4:17-26.

Campos, I. T. N., Amino, R., Sampaio, C. A. M., Auerswald, E. A., Friedrich, T., Lemaire, H. J., Schenkman, S. and Tanaka, A. S. (2002) Infestin, a thrombin inhibitor presents in *Triatoma infestans* midgut, a Chagas' disease vector: gene cloning, expression and characterization of the inhibitor. *Insect Biochem. Mol. Biol.* **32**(9): 991-997.

Chen, M. S., Fellers, J. P., Stuart, J. J., Reese, J. C. and Liu, X. (2004) A group of related cDNAs encoding secreted proteins from Hessian fly [*Mayetiola destructor* (Say)] salivary glands. *Insect Mol Biol.* **13**: 101-108.

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

Christeller, J. T. and Shaw, B. D. (1989) The interaction of a range of serine protease inhibitors with bovine trypsin and *Costelytra Zealandica* trypsin. *Insect Biochem.* **19**: 233-241.

Christeller, J. T., Laing, W. A., Shaw, B. D. and Burgess, E. P. J. (1990) Characterization and partial purification of the digestive proteases of the black field cricket, *Teleogryllus commodious* (Walker): elastase is the major component. *Insect Biochem.* **20**:157-164.

Christeller, J. T., Laing, W. A., Markwick, N. P. and Burgess, E. P. J. (1992) Midgut protease activities in 12 phytophagous lepidopteran larvae: Dietary and protease inhibitor interactions. *Insect Biochem. Mol. Biol.* **22**: 735-746.

Cierpicki, T., Bania, J. and Otlewski, J. (2000) NMR solution structure of *Apis mellifera* chymotrypsin/cathepsin G inhibitor-1 (AMCI-1): structural similarity with Ascaris protease inhibitors. *Protein Sci.* 9: 976-984.

Clermont, A., Wedde, M., Seitz, V., Podsiadlowski, L., Lenze, D., Hummel, M. and Vilcinskas, A. (2004) Cloning and expression of an inhibitor of microbial metalloproteinases from insects contributing to innate immunity**.** *Biochem. J.* **382**: 315-322.

Crowther, D. C., Evans, D. L. and Carrell, R. W. (1992) Serpins: implications of a mobile reactive centre. *Curr. Opin. Biotechnol.* **3**: 399-407.

Davis, N. G., Horecka, J. L. and Sprague, G. F. Jr. (1993) Cis- and trans-acting functions required for endocytosis of the yeast pheromone receptors. *J Cell Biol*.122: 53-65.

Deussing, J., **Kouadio, M**., **Rehman, S**., **Werber, I**., **Schwinde, A**. and **Peters, C**. (2002) Identification and characterization of a dense cluster of placenta-specific cysteine and peptidase genes and related genes on mouse chromosome 13. *Genomics* **79**: 225-240.

Dickinson, J. L., Bates, E. J., Ferrante, A. and Antalis, T. M. (1995) Plasminogen activator inhibitor 2 inhibits tumor necrosis factor alpha-induced apoptosis. Evidence for an alternate biological function. *J. Biol. Chem.* **270:** 27894-27904.

Dobó, J. and Gettins, P. G. W. (2004) α_1-Proteinase inhibitor forms initial non-covalent and final covalent complexes with elastase analogously to other serpin-proteinase pairs, suggesting a common mechanism of inhibition. *J. Biol. Chem.* **279**: 9264-9269.

Duggan, B. M., Dyson, H. J. and Wright, P. E. (1999) Inherent flexibility in a potent inhibitor of blood coagulation, recombinant nematode anticoagulant protein c2. *Eur. J. Biochem.* **265**: 539–548.

Eguchi, M. 1993. Protein protease inhibitors in insects and comparison with mammalian inhibitors. *Comp. Biochem. Physiol.* **105B**:449-456.

Elpidina, E. N., Vinokurov, K. S., Rudenskaya, Y. A., Dunaevsky, Y. E. and Zhuzhikov, D. P. (2001) Proteinase inhibitors in *Nauphoeta cinerea* midgut. *Arch. Insect Biochem. Physiol.* **48**: 217-222.

Engelmann, F., and Geraerts, W. P. M. (1980) The proteases and the protease inhibitor in the midgut of *Leucophaea maderae*. *J. of Insect Physiol.* **26**: 703-710.

Escande, F., Porchet, N., Bernigaud, A., Petitprez, D., Aubert, J. P. and Buisine, M. P. (2004) The mouse secreted gel-forming mucin gene cluster. *Biochim. Biophys. Acta* **1676**: 240-250.

Fastrez, J. and Fersht, A. R. (1973) Demonstration of the acyl-enzyme mechanism for the hydrolysis of peptides and anilides by chymotrypsin. *Biochemistry* **12**: 2025-2034.

Ferreira, C., and Terra, W. R. (1989) Spatial organization of digestion, secretory mechanisms and digestive enzyme properties in *Pheropsophus aequinoctialis* (Coleoptera: Carabidae). *Insect Biochem.* **19**: 383-391.

Ferreira, C., Capella, A. N., Sitnik, R. and Terra, W. R. (1994) Properties of the digestive enzymes and the permeability of the peritrophic membrane of *Spodoptera frugiperda* (Lepidoptera) larvae. *Comp. Biochem. Physiol*. **107A**: 631-640.

Francischetti, I. M., Ribeiro, J. M., Champagne, D. and Andersen, J. (2000). Purification, cloning, expression, and mechanism of action of a novel platelet aggregation inhibitor from the salivary gland of the blood-sucking bug, *Rhodnius prolixus. J. Biol. Chem.* 275(17): 12639-12650.

Friedrich, T., Kroger, B., Bialojan, S., Lemaire, H. G., Hoffken, H. W., Reuschenbach, P., Otte, M. and Dodt, J. (1993) A Kazal-type inhibitor with thrombin specificity from *Rhodnius prolixus. J. Biol. Chem.* **268**: 16216–16222.

Friedrich, R., Panizzi, P., Fuentes-Prior, P., Richter, K., Verhamme, I., Anderson, P. J., Kawabata, S., Huber, R., Bode, W. and Bock, P. E. (2003) Staphylocoagulase is a prototype for the mechanism of cofactor-induced zymogen activation. *Nature* 425: 535–539.

Frobius, A. C, Kanost, M. R., Gotz, P. and 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.

Fuller, R. S., Sterne, R. E., and Thorner, J. (1988) Enzymes required for yeast prohormone processing**.** *Annu. Rev. Physiol.* **50**: 345-362.

Gagne, R. J. and Hatchett, J. H. (1989) Instars of the Hessian fly (Diptera: Cecidomyiidae). *Ann. Entomol. Soc. Am.* **82**: 73-79.

Gáspári, Z., Patthy, A., Gráf, L. and Perczel, A. (2002) Comparative structure analysis of proteinase inhibitors from the desert locust, *Schistocerca gregaria. Eur. J. Biochem.* **269**: 527-537.

Gatehouse, A. M. R. and Gatehouse, J. A. (1998) Identifying proteins with insecticidal activity: use of encoding genes to produce insect-resistant transgenic crops. *Pestic. Sci.* **52**: 165-175.

Gettins, P. G. W. (2002) Serpin structure, mechanism, and function. *Chem. Rev.* **102**: 4751-4803.

Godbole, S. A., Krishna, T. G. and Bhatia, C. R. (1994) Purification and characterization of protease inhibitors from pigeon pea [*Cajanus cajan* (L)] seeds. *J. Sci. Food Agric.* **64**: 87-93.

Gooding, R. H. and Rolseth, B. M. (1976) Digestive process of hematophagous insects. XI. Partial purification and some properties of six proteolytic enzymes from the tsetse fly *Glossina morsitans morsitans* Westwood (Diptera: Glossinidae). *Can. J. Zool.* **54**: 1950-1959.

Goodwin, R. L., **Baumann, H**. **and Berger, F. G**. (1996) Patterns of divergence during evolution of alpha 1-proteinase inhibitors in mammals. *Mol. Biol. Evol.* **13**: 346-358

Goto, S. G. and Denlinger, D. L. (2002) Genes encoding two cystatins in the flesh fly *Sarcophaga crassipalpis* and their distinct expression patterns in relation to pupal diapause. *Gene* **292**: 121-127.

Graf, R. and Briegel, H. (1985) Isolation of trypsin isozymes from the mosquito *Aedes aegypti* (L.). *Insect Biochem.* **15**: 611-618.

Graf, R., Boehlen, P. and Briegel, H. (1991) Structural diversity of trypsin from different mosquito species feeding on vertebrate blood. *Experientia* **47**: 603-609.

Grasberger, B. L., Clore, G. M. and Gronenborn, A. M. (1994) High-resolution structure of *Ascaris* trypsin inhibitor in solution: direct evidence for a pH-induced conformational transition in the reactive site. *Structure* **2**: 669-678.

Grigoryev, S. A., Bednar, J., and Woodcock, C. L. (1999) MENT, a heterochromatin protein that mediates higher order chromatin folding, is a new serpin family member. *J. Biol. Chem.* **274**: 5626-5636.

Grover Jr., P. B., Ross, D. R. and Shukle, R. H. (1988) Identification of digestive carbohydrates in larvae of the Hessian fly, *Mayetiola destructor* (Say) (Diptera: Cecidomyiidae), *Arch. Insect Biochem. Physiol.* **8**: 59–72.

Harrison, L. M., Nerlinger, A., Bungiro, R. D., Cordova, J. L., Kuzmic, P., Cappello, M. (2002) Molecular characterization of *Ancylostoma* inhibitors of coagulation factor Xa. Hookworm anticoagulant activity in vitro predicts parasite blood feeding in vivo. *J. Biol. Chem.* **277**: 6223-6229.

Hartley, B. S. (1960) Proteolytic enzymes. *Annu. Rev. Biochem.* **29:** 45-72.

Hatchett, J. H., Starks, K. J. and Webster, J. A. (1987) Insect and mite pest of wheat. In Wheat and Wheat Improvement. Agronomy Monograph No. 13, pp. 625-675.

Hawley, J. H. and Peanasky, R. J. (1992) *Ascaris suum*: are trypsin inhibitors involved in species specificity of Ascarid nematodes? *Exp Parasitol.* **75**(1): 112-128.

Heibges, A., Glaczinski, H., Ballvora, A., Salamini, F. and Gebhardt, C. (2003) Structural diversity and organization of three gene families for Kunitz-type enzyme inhibitors from potato tubers (*Solanum tuberosum*L.). *Mol. Gen. Genomics* **269**: 526 – 534.

Heinemeyer, W., Kleinschmidt, J. A., Saidowsky, J., Escher, C. and Wolf, D. H. (1991) Proteinase yscE, the yeast proteasome/multicatalytic-multifunctional proteinase: mutants unravel its function in stress induced proteolysis and uncover its necessity for cell survival. *EMBO J.* 10: 555-562.

Hill, R. E. and Hastie, N. D. (1987) Accelerated evolution in the reactive center regions of serine protease inhibitors. *Nature* 326: 96-99.

Horler, E. and Briegel, H. (1997) Chymotrypsin inhibitors in mosquitoes: activity profile during development and after blood feeding. *Arch Insect Biochem. Physiol.* **36**: 315-333.

Huang, Q., Deveraux, Q. L., Maeda, S., Stennicke, H. R., Hammock, B. D. and Reed, J. C. (2001a) Cloning and characterization of an inhibitor of apoptosis protein (IAP) from *Bombyx mori*. *Biochim. Biophys. Acta*. **1499**: 191-198.

Huang, X., Knoell, C. T., Frey, G., Hazegh-Azam, M. and Tashjian, A. H. Jr. (2001b) Modulation of recombinant human prostate-specific antigen: activation by Hofmeister salts and inhibition by azapeptides. *Biochemistry* **40**: 11734–11741.

Huber, R. and Carrell, R.W. (1989) Implications of the three-dimensional structure of α 1-antitrypsin for structure and function of serpins. *Biochemistry* **28:** 8951-8966

Hwang, S. R., Steineckert, B., Toneff, T., Bundey, R., Logvinova, A. V., Goldsmith, P. and Hook, V. Y. (2002) The novel serpin endopin 2 demonstrates cross-class inhibition of papain and elastase: localization of endopin 2 to regulated secretory vesicles of neuroendocrine chromaffin cells**.** *Biochemistry* **41**: 10397-10405.

International Union of Biochemistry. (1992) Enzyme nomenclature. Academic Press Inc., Orlando, Fla.

Jany, K. D., Haung, H., Pfleiderer, G. and Ishay, J. (1978) Enzymatic and chemical properties of an endopeptidase from the larva of the hornet, *Vespa crabro*. *Biochemistry* **17**: 4675-4682.

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

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

Johnston, K. A., Lee, M. J., Gatehouse, J. A. and Anstee, J. H. 1991. The partial purification and characterization of serine protease activity in midgut of larval *Helicoverpa armigera*. *Insect Biochem*. **21**(4): 389-397.

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

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

Katsumi, A., Tuley, E. A., Bodó, I. and Sadler, J. E. (2000) Localization of disulfide bonds in the cysteine knot domain of human von willebrand factor. *J. Biol. Chem*. **275**: 25585-25594.

Kellenberger, C., Ferrat, G. Leone, P., Darbon, H. and Roussel, A. (2003) Selective inhibition of trypsins by insect peptides: Role of P6-P10 Loop. *Biochemistry* **42**: 13605–13612.

Khan, A. R. and James, M. N. (1998) Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes. *Protein Sci.* **7**:815–836.

Kraunsoe, J. A., Claridge, T. D. and Lowe, G. (1996) Inhibition of human leukocyte and porcine pancreatic elastase by homologues of bovine pancreatic trypsin inhibitor. *Biochemistry* **35**: 9090-9096.

Krem, M. M. and Cera, E. D. (2002) Evolution of enzyme cascades from embryonic development to blood coagulation. *Trends in Bioch. Sci*. **27**: 67-74.

Kress, H., Jarrin, A., Thüroff, E., Saunders, R., Weise, C., Busch, M. C. A., Knapp, E. W., Wedde, M. and Vilcinskas, A. (2004) A Kunitz type protease inhibitor related protein is synthesized in *Drosophila* prepupal salivary glands and released into the moulting fluid during pupation. *Insect Biochem. Mol. Biol.* **34**: 855-869.

Kurioka, A., Yamazaki, M. and Hirano, H. (1999) Primary structure and possible functions of a trypsin inhibitor of *Bombyx mori*. *Eur. J. Biochem.* **259**: 120–126.

Lange, U., Keilholz, W., Schaub, G. A., Landmann, H., Markwardt, F. and Nowak, G. (1999) Biochemical characterization of a thrombin inhibitor from the bloodsucking bug, *Dipetalogaster maximus*. *Haemostasis*. **29**: 204-211.

Laskowski, M. J. and Kato, I. (1980) Protein inhibitors of proteinases. *Annu. Rev. Biochem*. **49**: 593-626.

Laskowski, M. Jr, Kato, I., Ardelt, W., Cook, J., Denton, A., Empie, M. W., Kohr, W. J., Park, S. J., Parks, K. and Schatzley, B. L. (1987) Ovomucoid third domains from 100 avian species: isolation, sequences, and hypervariability of enzyme-inhibitor contact residues. *Biochemistry* **26**: 202-221.

Lazure, C. (2002) The peptidase zymogen proregions: nature's way of preventing undesired activation and proteolysis. *Curr. Pharm. Des.* 8: 511–531.

Lemos, F. J. A. and Terra, W. R. (1992) Soluble and membrane-bound forms of trypsin-like enzymes in *Musca domestica* larval midguts. *Insect Biochem. Mol. Biol.* **22**: 613-619.

Levinsky, H., Birk, Y. and Applebaum, S. W. (1977) Isolation and characterization of a new trypsin like enzyme from *Tenebrio molitor* L. larvae. *Int. J. Pept. Protein Res.* **10**: 252-264.

Liao, W. T., Yang, Y. and Wu, X. F. (2002) Expression and functional analysis of an inhibitor of apoptosis protein from *Trichoplusia ni. Biochem. Biophys. Res. Commun.* **293**: 675-679.

Liu, Q., Qi, Y. and Chejanovsky, N. (2003) Identification and classification of the *Spodoptera littoralis* nucleopolyhedrovirus inhibitor of apoptosis gene. *Virus Genes.* **26**:143-149.

Liu, X., Fellers, J. P., Wilde, G. E., Stuart, J. J. and Chen, M. S. (2004) Characterization of two genes expressed in the salivary glands of the Hessian fly, *Mayetiola destructor* (Say). *Insect Biochem. Mol. Biol.* **34**: 229-237.

López-Otín, C. and Christopher, M. (2002) Protease degradomics: A new challenge for proteomics. *Mol. Cell. Biol.* 3: 509-513.

Lu, W., Apostol, I., Qasim, M., Warne, N., Wynn, R., Zhang, W. L., Anderson, S., Chiang, Y. W., Ogin, E., Rothberg, I., Ryan, K. and Laskowski, M., Jr. (1997) Binding of amino acid side-chains to S1 cavities of serine proteinases. *J. Mol. Biol.* **266**: 441-461.

Lu, C. C., Nguyen, T., Morris, S., Hill, D. and Sakanari, J. A. (1998) *Anisakis simplex*: mutational bursts in the reactive site centers of serine protease inhibitors from an Ascarid nematode. *Exp. Parasitol.* **89**: 257-261.

Lung, O., Tram, U., Finnerty, C. M., Eipper-Mains, M. A., Kalb, J. M. and Wolfner, M. F. (2002) The *Drosophila melanogaster* seminal fluid protein Acp62F is a protease inhibitor that is toxic upon ectopic expression**.** *Genetics* **160**: 211-224.

McGrath, M. E., Erpel, T., Browner, M. F. and Fletterick, R. J. (1991) Expression of the protease inhibitor ecotin and its co-crystallization with trypsin. *J. Mol. Biol.* **222**:139-142.

McGrath, M. E., Gillmor, S. A. and Fletterrick, R. J. (1995) Ecotin: Lessons on survival in a protease-filled world. *Protein Sci.* **4**: 141-148.

Meier, P., Silke, J., Leevers, S. J. and Evan, G. I. (2000) The *Drosophila* caspase DRONC is regulated by DIAP1. *EMBO J.* **19**: 598-611.

Melville, J. C. and Ryan, C. A. (1972) Chymotrypsin inhibitor I from potatoes: Large scale preparation and characterization of its subunit components. *J. Biol. Chem.* **247**: 3445-3453.

Miller, J. W., Kramer, K. J. and Law, J. H. (1974) Isolation and partial characterization of the larval midgut trypsin from the tobacco hornworm, *Manduca sexta*, Johannson (Lepidoptera:Sphingidae). *Comp. Biochem. Physiol.* **48B**: 117-129.

Milstone, A. M., Harrison, L. M., Bungiro, R. D., Kuzmič, P. and Cappello, M. (2000) A Broad spectrum Kunitz type serine protease inhibitor secreted by the hookworm *Ancylostoma ceylanicum. J. Biol. Chem.* **275**: 29391-29399.

Mitsudo, K., Jayakumar, A., Henderson, Y., Frederick, M. J., Kang, Y., Wang, M., El-Naggar, A. K., and Clayman, G. L. (2003) Inhibition of serine proteinases plasmin, trypsin, subtilisin A, cathepsin G, and elastase by LEKTI: A kinetic analysis. *Biochemistry* **42:** 3874 – 3881.

Mukhopadhyay, D. (2000) The molecular evolutionary history of a winged bean a-chymotrypsin inhibitor and modeling of its mutations through structural analyses. *J. Mol. Evol.* **50**: 214 – 223.

Nakakura, N., Hietter, H., Van Dorsselaer, A. and Luu, B. (1992) Isolation and structural determination of three peptides from the insect *Locusta migratoria*. Identification of a deoxyhexose-linked peptide. *Eur. J. Biochem.* **204:** 147-153.

Narumi, H., Hishida, T., Sasaki, T., Feng, D.-F. and Doolittle, R. F. (1993) Molecular cloning of silkworm (*Bombax mori*) antichymotrypsin: a new member of the serpin superfamily of proteins from insects. *Eur. J. Biochem.* 214: 181-187.

Neurath, H. (1984) Evolution of proteolytic enzymes. *Science* **224**: 350-357.

Nielsen, E. W, Fure, H., Winge, P. and Mollnes, T. E. (1998) Identification of a C-->T mutation in the reactive-site coding region of the C1-inhibitor gene and its detection by an improved mutation-specific polymerase chain reaction method. *Scand. J. Immunol.* **47**: 273-276.

Odani, S., Koide, T. and Ono, T. (1983) The complete amino acid sequence of barley trypsin inhibitor. *J. Biol. Chem.* **258**: 7998-8003.

Ohta, T. 1994 Further examples of evolution by gene duplication revealed through DNA sequence comparisons. *Genetics* **138**:1331-1337.

Oppert, B., Kramer, K. J. and McGaughey, W. H. (1997) Rapid microplate assay of proteinase mixtures. *BioTechnology* **23**: 70–72.

Oppert, B., Morgan, T. D., Hartzer, K. and Kramer, K. J. (2005) Compensatory proteolytic responses to dietary proteinase inhibitors in the red flour beetle, *Tribolium castaneum* (Coleoptera: Tenebrionidae). *Comp. Biochem. Physiol.* Part C **140**: 53-58.

Orlowski, M. (1990) The multicatalytic proteinase complex, a major extra lysosomal proteolytic system. *Biochemistry* **29**: 10289-10297.

Otlewski, J. and Krowarsch, D. (1996) Squash inhibitor family of serine proteinases. *Acta Biochim. Pol.* **43**: 431-444.

Pohar, N., Godenschwege, T. A. and Buchner, E. (1999) Invertebrate tissue inhibitor of metalloproteinase: structure and nested gene organization within the synapsin locus is conserved from *Drosophila* to human. *Genomics* **57**: 293–296.

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

Prakash, B., Selvaraj, S., Murthy, M. R. N., Sreerama, Y. N., Rao, R. D. and Gowda, L. R. (1996) Analysis of the amino acid sequences of plant Bowman-Birk inhibitors. *J. Mol. Evol.* **42**: 560 – 569.

Pritchard, L. and Dufton, M. (1999) Evolutionary trace analysis of the Kunitz/BPTI family of proteins: functional divergence may have been based on conformational adjustment. *J. Mol. Biol.* **285**: 1589-1607.

Puente, X. S. and López-Otín, O. (2004) A genomic analysis of rat proteases and protease inhibitors. *Genome Res.* **14**: 609-622.

Qasim, M.A., Ranjbar, M. R., Wynn, R., Anderson, S. and Laskowski Jr. M. (1995) Ionizable P$_L$ residues in serine proteinase inhibitors undergo large pK shifts on complex formation. *J. Biol. Chem.* **270**: 27419-27422.

Ramesh, N., Sugumaran, M., and Mole, J. E. (1988) Purification and characterization of two trypsin inhibitors from the hemolymph of *Manduca sexta* larvae. *J. Biol. Chem.* **263**: 11523-11527.

Rao, M. B., Tanksale, A. M., Ghatge, M. S. and Deshpande, V. V. (1998) Molecular and biotechnological aspects of microbial proteases. *Microbiol. Mol. Biol. Rev.* **62**: 597-635.

Ratcliffe, R. H. and Hatchett, J. H. (1997) Biology and genetics of the Hessian fly and resistance in wheat: Edited by K. Bondari, In New Developments in Entomology pp 47-56. Research Signpost Scientific Information Guild, Trivandrum, India.

Rawlings, N. D. and Barrett, A. J. (1993) Evolutionary families of peptidases. *Biochem. J.* **290:** 205-218.

Ray, C. A., Black, R. A., Kronheim, S. R., Greenstreet, T. A., Sleath, P. R., Salvesen, G. S. and Pickup, D. J. (1992) Viral inhibition of inflammation: cowpox virus encodes an inhibitor of the interleukin-1 beta converting enzyme. *Cell* **69**: 597-604.

Reeck, G. R., Kramer, K. J., Baker, J. E., Kanost, M. R., Fabrick, J. A. and Behnke, C. A. (1997) Proteinase inhibitors and resistance of transgenic plants to insects. In: CAROZZI, N. and KOZIEL, M., eds. *Advances in insect control: the role of transgenic plants*. London, Taylor and Francis, 157-183.

Richard, I. (2005). The genetic and molecular bases of monogenic disorders affecting proteolytic systems. *J. Mol. Genet.* **42**: 529-539.

Ryan, C. A. and Santarius, K. (1976) Immunological similarities of proteinase inhibitors from potatoes. *Plant Physiol.* **58**: 683-685.

Ryan, C. A. (1990) Protease inhibitors in plants: Genes for improving defenses against insects and pathogens. *Annu. Rev. Phytopathol.* **28**: 425-429.

Rydel, T. J., Ravichandran, K. G., Tulinsky, A., Bode, W., Huber, R., Roitsch, C. and Fenton, J. W. 2nd. (1990) The structure of a complex of recombinant hirudin and human alpha-thrombin. *Science* **249**: 277-280.

Sakal, E., Applebaum, S. W. and Birk, Y. (1988) Purification and characterization of *Locusta migratoria* chymotrypsin. *Int. J. Pep. Prot. Res*. **32**: 590-598.

Sakal, E., Applebaum, S. W. and Birk, Y. (1989) Purification and characterization of digestive trypsins from the digestive tract of *Locusta migratoria. Int. J. Pept. Protein Res*. **34**: 498-505.

Schick, C., Pemberton, P. A., Shi, G. P., Kamachi, Y., Cataltepe, S., Bartuski, A. J., Gornstein, E. R., Bromme, D., Chapman, H. A. and Silverman, G. A. (1998) Cross-class inhibition of the cysteine proteinases cathepsins K, L, and S by the serpin squamous cell carcinoma antigen 1: A kinetic analysis. *Biochemistry* **37**: 5258-5266.

Schlott, B., Wohnert, J., Icke, C., Hartmann, M., Ramachandran, R., Guhrs, K. H., Glusa, E., Flemming, J., Gorlach, M., Grosse, F. and Ohlenschlager O. (2002) Interaction of Kazal-type inhibitor domains with serine proteinases: biochemical and structural studies. *J Mol. Biol.* **318:** 533-546.

Schumaker, T. T. S., Cristofoletti, P. T. and Terra, W. R. (1993) Properties and compartmentalization of digestive carbohrases and proteases in *Sacptotrigona bipunctata* (Apidae:Meliponiae) larvae. *Apidologie*. 24: 3-17.

Shkenderov, S. (1973) A protease inhibitor in bee venom-Identification, partial purification and some properties. *FEBS Lett*. **33**:343-347.

Shrivastava, B. and Ghosh, A. K. (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.

Shukle, R. H., Murdock, L. L. and Gallun, R. L. (1985) Identification and partial characterization of a major gut proteinase from larvae of the Hessian fly, *Mayetiola destructor* (Say) (Deptera: Cecidomyiidae), *Insect Biochem.* **15**: 93–101.

Silverman, G. A., Bird, P. I., Carrell, R. W., Church, F. C., Coughlin, P. B., Gettins, P. G., Irving, J. A., Lomas, D. A., Luke, C. J., Moyer, R. W., Pemberton, P. A., Remold-O'Donnell, E., Salvesen, G. S., Travis, J. and Whisstock, J. C. (2001) The serpins are an expanding superfamily of structurally similar but functionally diverse proteins. *J. Biol. Chem.* **276**: 33293-33296.

Simonet, G., Breugelmans, B., Proost, P., Claeys, I., Van Damme, J., De Loof, A. and Broeck, V. J. (2005) Characterization of two novel pacifastin-like peptide precursor isoforms in the desert locust (*Schistocerca gregaria*): cDNA cloning, functional analysis and real-time RT-PCR gene expression studies. *Biochem. J*. **388**: 281-289.

Soderhall, K. and Cerenius, L. (1998) Role of the propeholoxidase-activating system in invertebrate immunity. *Curr. Opin. Immunol*. 10: 23-28.

Sottrup-Jensenn, L. and Birkedal-Hansen, H. (1989) Human fibroblast collagenase-α-macroglobulin interactions. Localization of cleavage sites in the bait regions of five mammalian α-macroglobulins. *J. Biol. Chem*. **264**: 393-401.

Sözen, M. A., Armstrong, J. D., Yang, M., Kaiser, M. and Dow, J. A. T. (1997) Functional domains are specified to single-cell resolution in a *Drosophila* epithelium. *Proc. Natl. Acad. Sci. USA* **94**: 5207-5212.

Spiess, C., Beil, A., and Ehrmann, M. (1999) A temperature-dependent switch from chaperone to protease in a widely conserved heat shock protein. *Cell* 97: 339–347.

Stanssens, S., Bergum, P. W., Gansemans, Y., Jespers, J., Laroche, Y., Huang, S., Maki, S., Messens, J., Lauwereys, M., Cappello, M., Hotez, P. J., Lasters, I. and Vlasuk, G. P. (1996) Anticoagulant repertoire of the hookworm *Ancylostoma caninum. Proc. Natl. Acad. Sci. USA* **93**: 2149-2154.

Sternlicht, M. D. and Werb, Z. (2001) How matrix metalloproteinases regulate cell behavior. *Annu. Rev. Cell Dev. Biol.* **17**: 463-516.

Sugumaran, M., Saul, S. J. and Ramesh, N. (1985) Endogenous protease inhibitors prevent undesired activation of prophenolase in insect hemolymph. *Biochem. Biophys. Res. Commun.* **132**: 1124-1129.

Suzuki, T. and Natori, S. (1985) Purification and characterization of an inhibitor of the cysteine protease from the hemolymph of *Sarcophaga peregrina* larvae. *J. Biol. Chem*. **260**: 5115-5120.

Suzuki, A., Tsunogae, Y., Tanaka, I., Yamane, T., Ashida, T., Norioka, S., Hara, S. and Ikenaka, T. (1987) The structure of Bowman-Birk type protease inhibitor A-II from peanut (*Arachis hypogaea*) at 3.3A^{o} resolution. *J. Biochem*. (Tokyo) **101**: 267-274.

Takahashi, K. and Fukada, H. (1985) Calorimetric studies of the binding of *Streptomyces* subtilisin inhibitor to subtilisin of Bacillus subtilis strain N'. *Biochemistry* **24:** 297 – 300.

Tanaka, K., Tamura, T., Yoshimura, T. and Ichihara, A. (1992) Proteasomes: protein and gene structures. *New Biol*. **4**: 173-187.

Terra, W. R. and Ferreira, C. (1994) Insect digestive enzymes: properties, compartmentalization and function. *Comp. Biochem. Physiol*. **109B**: 1-62.

Thie, N. M. R. and Houseman, J. G. (1990) Cysteine and serine proteolytic activities in larval midgut of yellow mealworm, *Tenebrio molitor* L. (Coleoptera:Tenebrionidae). *Insect Biochem*. 20: 741–744.

Thumm, M. Biological control through proteolysis. Yeast as a model of the eukaryotic cell. In Innovations in Proteases and Inhibitors. Ed. By Frances X. Aviles. Published by Walter De Gruyter. 1993, pp63-80.

Tong, Y. (2004) Structure and function of serpins from hemolymph of the tobacco hornworm, *Manduca sexta*. PhD thesis submitted to Kansas State University.

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

Travis, J. and Salvesen, G. (1983) Control of coagulation and fibrinolysis by plasma proteinase inhibitors. *Behring. Inst. Mitt.* **73**: 56-65.

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

Turk, V., Turk, B., Guncar, G., Turk, D. and Kos, J. (2002) Lysosomal cathepsins: structure, role in antigen processing and presentation, and cancer. *Adv. Enzyme Regul.* **42**: 285-303.

Valaitis, A. P., (1995) Gypsy moth midgut proteinases: Purification and characterization of luminal trypsin, elastase and the brush border membrane leucine aminopeptidase. *Insect Biochem. Mol. Biol.* **25**: 139-149.

van de Locht, A., Lamba, D., Bauer, M., Huber, R., Friedrich, T., Kroger, B., Hoffken, W. and Bode, W. (1995) Two heads are better than one: crystal structure of the insect derived double domain Kazal inhibitor rhodniin in complex with thrombin. *EMBO J.* **14**: 5149-5157.

Vizioli, J., Catteruccia, F., Della Torre, A., Reckmann, I. and Müller, H-M. (2001) Blood digestion in the malaria mosquito *Anopheles gambiae* Molecular cloning and biochemical characterization of two inducible chymotrypsins. *Eur. J. Biochem.* **268**: 4027-4035.

Volpicella M., Ceci L. R., Cordewener J., America T., Gallerani R., Bode W., Jongsma M. A. and Beekwilder, J. (2003) Properties of purified gut trypsin from *Helicoverpa zea*, adapted to proteinase inhibitors. *Eur. J. Biochem.* **270**: 10-19.

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

Wang, P., Zhang, X. and Zhang, J. (2005) Molecular characterization of four midgut aminopeptidase N isozymes from the cabbage looper, *Trichoplusia ni*. *Insect Biochem. Mol. Biol.* **35**: 611-620.

Ward, C. W. (1975) Properties and specificity of the major anionic trypsin-like enzyme in the keratinolytic larvae of the webbing clothes moth. *Biochim. Biophys. Acta* **319**: 201-211.

Wedde, M., Weise, C., Kopacek, P., Franke, P., and Vilcinskas, A. (1998) Purification and characterization of an inducible metalloprotease inhibitor from the hemolymph of greater wax moth larvae, *Galleria mellonella*. *Eur. J. Biochem.* **255**: 535-543.

Wieczorek, M., Otlewski, J., Cook, J., Parks, K., Leluk, J., Wilimowska-Pelc A., Polanowski, A., Wilusz, T. and Laskowski, M. Jr. (1985) The squash family of serine protease inhibitors. Amino acid sequence and association equilibrium constants of inhibitors from squash, summer squash, zucchini, and cucumber seeds. *Biochem. Biophys. Res. Commun.* **126**: 646-652.

Yamamoto, Y., Watabe, S., Kageyama, T. Takahashi, S. Y. (1999) Purification and characterization of *Bombyx* cysteine proteinase specific inhibitors from the hemolymph of *Bombyx mori*. *Archiv. Insect Biochem. Physiol.* **42**:119-129.

Yan, J., Cheng, Q., Li, B-N., and Aksoy, S. (2001) Molecular characterization of two serine proteases expressed in gut tissue of the African trypanosome vector, *Glossina morsitans morsitans*. *Insect Mol. Biol.* **10**: 47-56.

Yousef, G. M. and Diamandis, E. P. (2001) The new human tissue kallikrein gene family: structure, function, and association to disease. *Endocrine Rev.* **22**: 184-204.

Yousef, G. M., Kopolovic, A. D., Elliott, M. B., and Diamandis, E. P. (2003) Genomic overview of serine proteases, *Biochem. Biophy. Res. Commun.* **305**: 28–36.

Zhang, X. and Maizels, R. M. (2001) Serine proteinase inhibitors from nematodes and the arms race between host and pathogen. *Trends Bioch. Sci.* **26**: 191-197.

Zhang, D., Cupp, M. S., and Cupp, E. W. (2002) Thrombostatin purification, molecular cloning and expression of novel anti-thrombin protein from horn fly saliva. *Insect Biochem. Mol. Biol.* **32:** 321–330.

Zhu, Y. (2001) Identification of immune-related genes from the tobacco hornworm, *Manduca sexta*, and characterization of two immune-inducible protein, serpin-3 and leureptin. PhD thesis submitted to Kansas State University.

Zhu, Y. C., Zeng, F. and Oppert, B. (2003a) Molecular cloning of trypsin-like cDNAs and comparison of proteinase activities in the salivary glands and gut of the tarnished plant bug *Lygus lineolaris* (Heteroptera: Miridae). *Insect Biochem. Mol. Biol.* **33**(9): 889-899.

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

Zhu Y. C., Liu, X., Maddur, A. A., Oppert, B. and Chen, M. S. (2005) Cloning and characterization of chymotrypsin- and trypsin-like cDNAs from the gut of the Hessian fly [*Mayetiola destructor* (Say)]. *Insect Biochem. Mol Biol*. **35**: 23-32.

Zingali, R. B., Jandrot-Perrus, M., Guillin, M.-C. and Bon, C. (1993) Bothrojaracin, a new thrombin inhibitor isolated from *Bothrops jararaca* venom: characterization and mechanism of thrombin inhibition. *Biochemistry* **32,** 10794-10802.

Zinkler, D. and Polzer, M. (1992) Identification and characterization of digestive proteinases from the firebrat *Thermobia domestica*. *Comp. Biochem. Physiol*. **103B:** 669-673.