Available online at www.sciencedirect.com
SCIENCE DIRECT®
Insect Biochemistry and Molecular Biology 35 (2005) 23–32
Insect Biochemistry and Molecular Biology
www.elsevier.com/locate/ibmb

Cloning and characterization of chymotrypsin- and trypsin-like cDNAs from the gut of the Hessian fly [*Mayetiola destructor* (say)] ☆

Yu Cheng Zhu[a], Xiang Liu[b], Ashoka A. Maddur[b], Brenda Oppert[b,c], Ming-Shun Chen[b,d,*]

[a] *USDA-ARS-JWDSRC, PO Box 346/141 Exp Stn Rd, Stoneville, MS 38776, USA*
[b] *Department of Entomology, USDA-ARS, Kansas State University, 123 Waters Hall, Manhattan, KS 66506, USA*
[c] *USDA-ARS, Biological Research Unit, 1515 College Avenue, Manhattan, KS 66502, USA*
[d] *USDA-ARS, Plant Science and Entomology Research Unit, 4008 Throckmorton Hall, Kansas State University, Manhattan, KS 66506, USA*

Received 16 August 2004; received in revised form 26 September 2004; accepted 29 September 2004

Abstract

Fifteen unique cDNA clones encoding trypsin- or chymotrypsin-like proteins were cloned and characterized from a gut cDNA library derived from Hessian fly [*Mayetiola destructor* (Say)] larvae. Based on sequence similarities, the cDNAs were sorted into five gene groups, which were named MDP1 to MDP5. Two of the gene groups, MDP1 and MDP2, encoded chymotrypsin-like proteins; the other three encoded putative trypsins. All deduced proteins have conserved His^{87}, Asp^{136}, and Ser^{241} residues for the catalytic triad and three pairs of cysteine residues for disulfide bridge configurations. The substrate specificity determination residue at position 235 was also conserved in the putative trypsins and chymotrypsins. In addition, all the deduced protein precursors had a typical secretion signal peptide and activation peptide. Northern blot analysis revealed that all these gene groups were exclusively expressed in the larval stage. The expression profiles for each gene group differed significantly in different ages of the larva, as well as in different tissues. Protease activity analysis of gut extract, using specific inhibitors, demonstrated that serine proteases were the major digestive enzymes in the gut of *M. destructor* larvae. Serine protease inhibitors inhibited as much as 90% proteolytic activities of gut extract, whereas inhibitors specific to other proteases, including cysteine proteases, aspartic proteases, and metallo-proteases, inhibited only 10–24% of gut protease activity.

Keywords: Hessian fly; *Mayetiola destructor*; Wheat; Serine protease; Gut proteinases

1. Introduction

The Hessian fly [*Mayetiola destructor* (Say)] is one of the most destructive pests of wheat (*Triticum aestivum* L.) (Hatchett et al., 1987; Buntin, 1999). The most effective measure for controlling this insect pest is through the release of resistant wheat cultivars (Ratcliffe and Hatchett, 1997). As a consequence, most of the research on *M. destructor* is related to host–plant resistance. This includes the identification and introgression of host–plant resistance genes (Ratcliffe et al., 2003; Martin-Sanchez et al., 2003; Williams et al., 2003), molecular mapping of *Avr* genes (Rider et al., 2002; Behura et al., 2004), study of the distribution of different biotypes (Ratcliffe et al., 1994, 2000), and the characterization of induced wheat genes after infestation (Williams et al., 2002; Jang et al., 2003). All of the resistance genes so far identified confer resistance through antibiosis (Ratcliffe and Hatchett, 1997, Harris et al., 2003). First instar–larvae feeding on resistant plants die within 4 days without developing into the second instar (Hatchett and Gallun, 1970; El Bouhssini

Abbreviations: kDa, kilo-Dalton; EST, Expressed Sequence Tag; PCR, Polymerase Chain Reaction; ORF, Open Reading Frame

☆ Mention of commercial or proprietary product does not constitute endorsement by the USDA.

*Corresponding author. Department of Entomology, USDA-ARS, Kansas State University, 123 Waters Hall, Manhattan, KS 66506, USA. Tel.: +7855324719; fax: +7855326232.
E-mail address: mchen@oznet.ksu.edu (M.-S. Chen).

0965-1748/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ibmb.2004.09.006

et al., 1998). Molecular mechanisms for the antibiosis are not yet known. Experimental results demonstrated that there is no significant difference in the amount of food ingested from host plants between virulent and avirulent larvae in the initial feeding (Gallun and Langston, 1963). This observation raised the possibility that the antibiosis might be caused by toxic molecules produced in resistant plants. One of the likely targets for such toxicity in the insect would be the gut, where food digestion and detoxification take place (Terra and Ferreira, 1994; Herrero et al., 2001). It is known that plants synthesize various toxic molecules upon infestation, such as inhibitors to herbivore's digestive enzymes (Karban and Baldwin, 1997; Moura and Ryan, 2001). Protease inhibitors from host plants have been found to have a detrimental effect on insect development and are being used as targets for bioengineering to generate resistant plants (Murdock et al., 1988; Burgess et al., 1994; Huang et al., 1997). Protease activity has been detected in the gut extract from *M. destructor* larvae (Shukle et al., 1985), but little is known about the molecular and biochemical aspects of these enzymes. To isolate protease genes that are expressed in the gut of *M. destructor* larvae, we systematically analyzed the genes expressed in the gut of the first instar–larvae following a transcriptomic approach. Here we report the cloning and characterization of 15 trypsin- and chymotrypsin-like cDNAs identified from this analysis.

2. Materials and methods

2.1. Insects

Hessian fly larvae used in this research were derived from a laboratory colony. The insects were originally collected from Ellis County, Kansas (Gagne and Hatchett, 1989). Since then, the insects have been maintained on susceptible wheat seedlings ('Newton' or 'Karl 92') in the greenhouse. The majority (95%) of the insects were biotype GP although biotypes A, B and others were also found in low frequencies (Harris and Rose, 1989).

2.2. cDNA library construction and sequencing

Two hundred guts were obtained by dissecting 3-day-old larvae (first instar) under a dissecting microscope. The guts were immediately transferred into TRI reagent^{TM} (Molecular Research Center, Inc. Cincinnati, OH). Total RNA was extracted from the tissue of the gut using TRI reagent according the procedure provided by the manufacturer. A cDNA library was constructed from the RNA sample using a 'SMART^{TM,} cDNA library construction kit' from Clontech (Palo Alto, CA) according to the protocol provided by the manufacturer with one modification. Instead of using the original phage vector, PCR fragments were cloned directly into a plasmid using a TOPO TA cloning kit (Invitrogen, Carlsbad, CA). Recombinant plasmid DNA was isolated with a Qiagen BioRobot-3000 and sequenced using an ABI 3700 DNA analyzer.

2.3. RNA isolation and Northern blot analysis

Total RNA was extracted from whole insects using TRI reagent^{TM} (see above). For Northern blot, equal amounts (5 μg) of total RNA were separated on a 1.2% agarose gel containing formaldehyde and blotted onto GeneScreen membrane (Perkin Elmer Life Science Inc., Boston, MA). The membrane was baked at 80 °C for 2 h to fix RNAs onto the membrane. The membranes were then hybridized separately to individual cDNA probes labeled with ^{32}P-dCTP using the random labeling kit from Stratagene (La Jolla, CA). Hybridization was carried out overnight at 42 °C in a plastic bag containing 15 ml hybridization solution (10% dextran sulfate/1% SDS/1 M NaCl, pH 8.0) as described by Chen et al. (2004). After hybridization, the membranes were washed twice with 2 × SSC at room temperature for 30 min, twice with 2 × 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. The membranes were then exposed to Kodak SR-5 X-ray film overnight.

2.4. Sequence analysis

Open-reading-frame (ORF) and sequence-similarity analysis were performed through the website (http://www.ncbi.nlm.nih.gov/) of the National Center for Biotechnology Information (Bethesda, MD). Molecular weight calculations and *p*I prediction of mature proteins were carried out with the 'Compute *p*I/Mw tool' (http://us.expasy.org/tools/pi_tool.html). The ExPASy Proteomics tools on the website (http://www.expasy.ch/tools/) of the Swiss Institute of Bioinformatics were used to process data of deduced protein sequences. ClustalW (Thompson et al., 1994; gap weight = 8, gap length weight = 2) was used in pairwise sequence comparison. The BioEdit (Ver. 5.09; Hall, 1999) program was used to conduct multiple-sequence alignment and phylogenetic analysis.

2.5. Analysis of tissue-specific expression through RT-PCR

RT-PCR was carried out to determine if specific proteases were expressed in the gut or salivary glands. Total RNA was separately prepared from dissected guts or salivary glands. The RNA samples were treated with RNase-free DNase-I (Promega, Madison, WI) to remove potential DNA contamination. The RNA was

then reverse transcribed into cDNA using an oligo-dT primer with SuperScript II reverse transcriptase (Invitrogen). PCR amplification was carried out for 45 cycles as follows: 30 seconds (s) at 94 °C; 30 s at 55 °C; 60 s at 72 °C. DNA fragments from the PCR reactions were separated on 1% agarose gels, stained with (0.5 μg/mL) ethidium bromide. DNA bands were photographed with a Kodak 290 digital camera, and band intensity was determined using Kodak 1D image analysis software (Version 3.54).

For primer design, we divided the protease cDNAs into five gene groups, according to their sequence similarity (see the Results section). Primer pairs specific to each gene group were synthesized (see primer sequences in Fig. 3). The specificity of the primer pairs to individual groups was tested experimentally using individual plasmid DNA samples as templates. Primers MDP1F + MDP1R could flank the corresponding regions of all five chymotrypsin (A-F) cDNAs in the MDP1 gene group to produce 657-bp fragments. Primers MDP2F + MDP2R were specific to the chymotrypsin cDNAs in the MDP2 gene group to produce 601-bp fragments. Primers MDP3F + MDP3R were specific to the trypsin cDNAs in the MDP3 gene group to produce 677-bp fragments. Primers MDP4F + MDP4R were specific to trypsin cDNAs in the MDP4 gene group to produce 669-bp fragments. Primers MDP5F + MDP5R were specific to the trypsin cDNAs in the MDP5A gene group to produce 486-bp fragments. As a control, a primer pair (MDrDF + MDrDR) was synthesized according to a cDNA that encodes a L27A-like ribosomal protein in the 60S subunit. The RT-PCR products (420 bp) from this ribosomal protein mRNA were used as an internal control to quantify the amounts of the protease RNAs.

2.6. Protease activity assay

Midguts of first instar *Mayetiola destructor* larvae were dissected under a light microscope in cold deionized water. The midgut was exposed by pulling the head with a fine needle, and the salivary gland and Malphigian tubules were removed. Midguts were collected in deionized water, mixed through pipetting, and pooled (20 guts per 20 μl water) in Eppendorf tubes on ice. The extract was centrifuged at 15,000*g* for 5 min, and the supernatant was aliquoted to 10 μl and frozen at −20 °C.

Accordingly, inhibition studies 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). Inhibitors included antipain, aprotinin, 4-amido-phenylmethylsulfonyl fluoride (APMSF), bestatin, chymostatin, ethyelenediamine tetraacetic acid (EDTA), ethylene glycol-bis (beta-aminoethyl-ether)-*N*,*N*,*N*′,*N*′-tetraacetate (EGTA), transexposysuccinyl-L-leucylamido-(4-guanido)-butane (E-64), soybean Kunitz trypsin inhibitor (SBTI), soybean Bowman-Birk trypsin inhibitor (SBBTI), and *N*-tosyl-L-phenyalanine chloromethyl ketone (TPCK). These inhibitors were obtained either from Pierce Chemical Company (Rockford, IL), Roche Applied Science (Indianapolis, IN), or Sigma Chemical Company (St. Louis, MO). Stock solutions were prepared in solvents per company specifications. The starting concentrations of the inhibitors are provided in the legend of Fig. 4. Protease inhibitors were added to universal buffer (Oppert et al., 1997) of pH 7.9, which is the optimal pH for proteolytic activity in mid-gut extract from Hessian fly larvae (data not shown). Inhibitors were preincubated with the gut extract (~0.8 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 casein, Molecular Probes, Eugene, OR) was added to each well to initiate the reaction. Inhibition was measured after 4 h incubation at 37 °C using a fluorescent microplate reader (Fluoroskan Ascent FL, Labsystems, Thermo Electron Corp., Milford, MA), with an excitation of 584 nm and emission of 620 nm. The percentage inhibition was calculated from incubations containing the inhibitor divided by the control incubations without the inhibitor, and multiplied by 100. The amount of the inhibitor resulting in 50% inhibition (IC_{50}) was calculated by linear regression from linear data points.

3. Results

In order to isolate genes that encode potential digestive proteases, we sequenced 1014 random clones from a gut cDNA library derived from first instar–larvae. GenBank searching with the cDNA sequences revealed 55 clones (or 5.4%) encoding various proteases. Among these 55 clones, 27 coded for chymotrypsin- or trypsin-like proteins, 19 coded for carboxypeptidases, six for signal peptidases, two for cycteine proteases (Cathepsin L), and one for lysosomal aspartic protease.

3.1. Chymotrypsin- and trypsin-like cDNAs

Of the 27 clones (all except one were full length) that encoded chymotrypsin- or trypsin-like proteins, 15 cDNAs were unique clones that encode different proteins. The GenBank-accession numbers for the unique clones, lengths of the cDNAs and predicted proteins, estimated molecular masses, *p*Is, putative secretion signal peptides, and activation peptides are listed in Table 1.

Sequence comparison indicated that these 15 cDNA clones could be sorted into five groups, which were

Table 1
Serine proteinase cDNAs and putative proteins from the gut of *M. destructor*

Clone name	GenBank number	cDNA (bp)	Encoded residues	Molecular mass(kDa)	*p*I	Signal peptide	Activation peptide	Putative proteinase
MDP1A	AY596471	1046	275	26.71	5.44	19	8	Chymotrypsin
MDP1B	AY596472	1049	275	26.81	5.44	19	8	Chymotrypsin
MDP1C	AY665658	953	275	26.78	5.44	19	8	Chymotrypsin
MDP1D	AY665659	1045	275	26.78	5.55	19	8	Chymotrypsin
MDP1E	AY665660	1068	275	26.81	5.44	19	8	Chymotrypsin
MDP1F	AY665661	1048	275	26.78	5.44	19	8	Chymotrypsin
MDP2A	AY596478	1052	269	26.24	8.66	21	10	Chymotrypsin
MDP2B	AY596479	1052	269	26.25	8.66	19	12	Chymotrypsin
MDP3A	AY596476	1116	268	26.11	8.41	17	14	Trypsin
MDP3B	AY596475	1065	268	26.12	7.72	17	14	Trypsin
MDP3C	AY665662	1121	268	26.09	7.72	17	14	Trypsin
MDP4A	AY596477	1077	273	25.76	5.79	23	13	Trypsin
MDP4B	AY596473	1075	273	25.71	5.79	23	13	Trypsin
MDP4C	AY596474	1078	273	25.76	5.79	23	13	Trypsin
MDP5A	AY669864	1147	263	25.67	5.75	17	13	Trypsin

Table 2
Pairwise comparison of cDNA and [putative protein] sequence identity (showing as % in upper-right corner) and phylogenetic distances (showing in lower-left corner) among the 15 serine proteinase clones

Clones	MDP1A	MDP1B	MDP1C	MDP1D	MDP1E	MDP1F	MDP2A	MDP2B	MDP3A	MDP3B	MDP3C	MDP4A	MDP4B	MDP4D	MDP5A
MDP1A		99[99]	90[99]	99[99]	97[99]	99[99]	48[25]	48[25]	51[28]	51[28]	51[28]	50[32]	50[31]	50[31]	49[25]
MDP1B	0.01		90[99]	99[99]	98[99]	99[99]	48[25]	48[25]	50[28]	51[28]	50[28]	50[31]	50[31]	50[31]	49[25]
MDP1C	0.02	0.01		90[99]	88[99]	90[99]	47[25]	47[25]	48[27]	49[28]	47[28]	50[31]	49[31]	50[31]	47[24]
MDP1D	0.02	0.01	0.01		97[99]	99[99]	48[25]	48[25]	50[28]	51[28]	50[28]	51[31]	51[31]	51[31]	49[25]
MDP1E	0.02	0.01	0.01	0.01		97[99]	48[25]	48[25]	50[28]	50[28]	50[28]	50[31]	50[31]	50[31]	49[25]
MDP1F	0.02	0.01	0.01	0.01	0.01		48[25]	48[25]	51[28]	51[28]	50[28]	51[31]	51[31]	51[31]	49[25]
MDP2A	2.65	2.66	2.67	2.66	2.68	2.66		98[99]	50[36]	51[37]	51[37]	52[40]	52[39]	52[40]	53[37]
MDP2B	2.67	2.67	2.68	2.67	2.69	2.67	0.01		51[36]	51[36]	51[36]	51[40]	52[39]	52[40]	53[38]
MDP3A	2.37	2.38	2.39	2.38	2.38	2.38	1.94	1.97		95[99]	99[99]	58[48]	58[48]	58[48]	62[60]
MDP3B	2.36	2.36	2.37	2.36	2.37	2.36	1.94	1.97	0.01		95[99]	58[48]	58[48]	58[48]	61[60]
MDP3C	2.36	2.36	2.37	2.36	2.37	2.36	1.94	1.97	0.02	0.01		57[48]	57[48]	57[48]	62[60]
MDP4A	2.17	2.17	2.18	2.17	2.18	2.17	1.77	1.79	1.40	1.38	1.38		99[99]	99[99]	57[49]
MDP4B	2.19	2.19	2.20	2.19	2.20	2.19	1.79	1.81	1.41	1.40	1.40	0.01		99[99]	57[49]
MDP4C	2.17	2.19	2.20	2.19	2.20	2.19	1.77	1.79	1.41	1.40	1.40	0.01	0.01		57[49]
MDP5A	2.50	2.51	2.53	2.50	2. 52	2.51	1.88	1.89	0.94	0.91	0.93	1.33	1.35	1.35	

named MDP1, MDP2, MDP3, MDP4, and MDP5, respectively. The MDP1 and MDP2 groups encode chymotrypsin-like proteins whereas the other three encode trypsin-like proteins. Members from the same group share more than 95% sequence identity at both nucleotide and amino acid levels (Table 2), except chymotrypsin MDP1-C, which has only 88–90% sequence identity compared with the other member in the same group at the nucleotide level. The lower sequence identity for chymotrypsin MDP1-C was caused by a short 3′ UTR showing 80 bp insertion/deletion (indel) (data not shown). Excluding this indel, sequence identity is also more than 95% between MDP1-C and other members in this group. Members from different groups share little sequence similarity at the nucleotide level. No sequence alignments between members from different groups could be found using BLASTN (data not shown). Pairwise comparison of the nucleotide sequences between members from different groups with ClustalW (Gap open = 10; Gap extension = 0.1) again revealed only low- (ranging from 47 to 62%) sequence identity. At the amino acid level, there were conservations around regions that are common to serine proteases between members from different groups (Table 2 and see Section 3.2).

3.2. Structure of the deduced proteins

Sequence comparison of the fifteen deduced protease precursors is shown in Fig. 1. Forty-six residues are conserved among all sequences. Most of the conserved residues are located in the vicinity of the histidine and

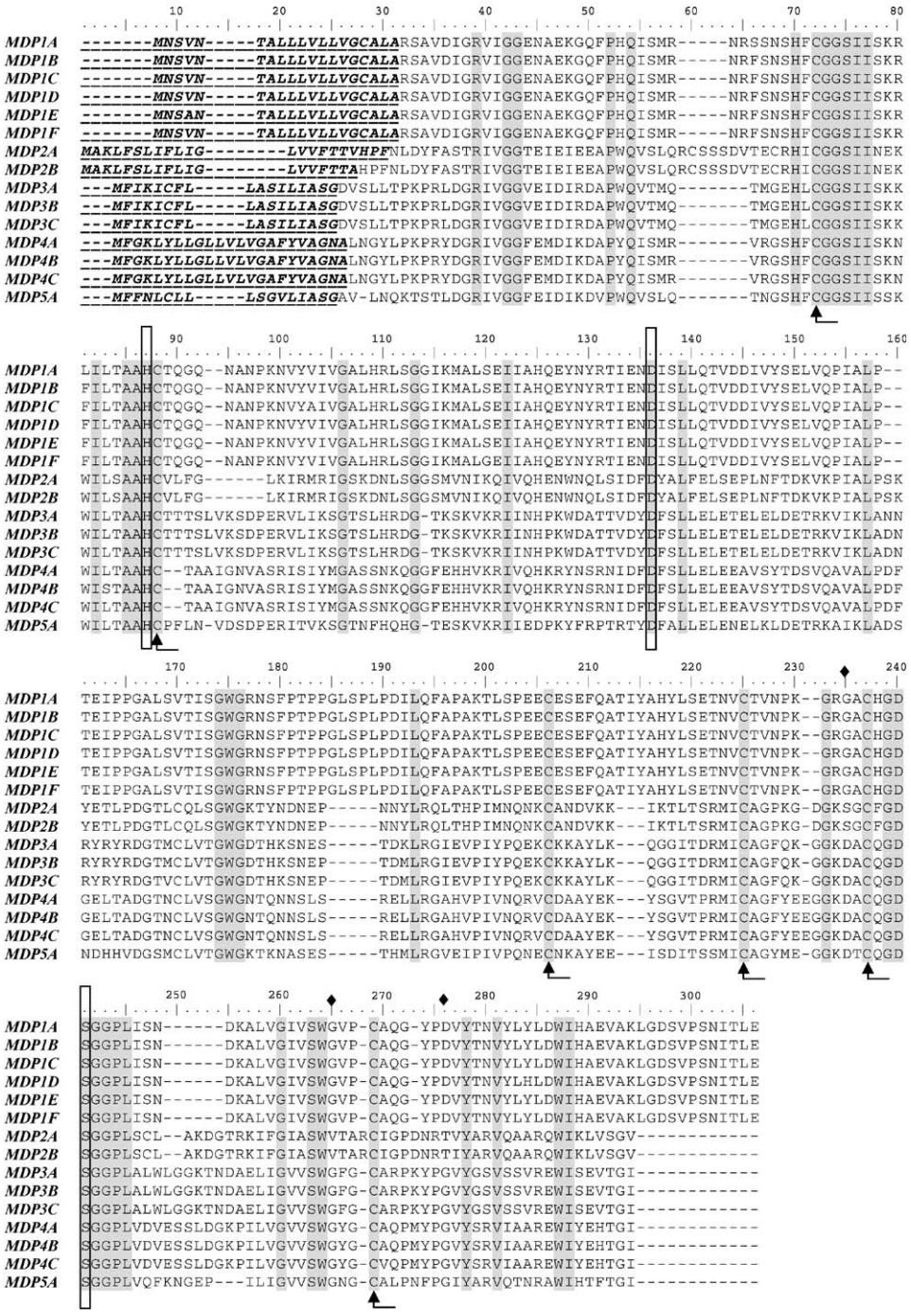


Fig. 1. Multiple alignment of the predicted amino acid sequences of the 15 proteinase-like precursors from *M. destructor*. Putative secretion signal peptides were bold and underlined. Functionally important residues H, D, and S (active sites) were boxed. Cysteines corresponding to the sites of the predicted disulfide bridges were marked with arrow lines at the bottom. Trypsin specificity determinant residues were indicated with (♦) on the top of sequences. Identical residues among all fifteen sequences are highlighted with gray background. Hyphens represent sequence alignment gaps.

serine active site, including a highly conserved N-terminus (IVGG or VIGG preceded by an arginine or lysine), which marks the N-termini of the active enzyme (Jany and Haug, 1983; Wang et al., 1993). The three critical residues, His^{87}, Asp^{136}, and Ser^{241} (numbers are assigned according to the multiple alignments) are all conserved to form the catalytic triad that is characteristic of serine proteases (Kraut, 1977; Wang et al., 1993; Peterson et al., 1994). Other important residues, including D^{235}, G^{265}, and G^{276} (♦), which define the substrate binding pocket, also are conserved in all trypsin-like proteins encoded by the MDP3 to MDP5

gene groups. The D^{235} residue determines specificity in both invertebrate and vertebrate trypsins by stabilizing the Lys or Arg residue at the substrate cleavage site through ionic interactions (Hedstrom et al., 1992; Wang et al., 1993). The D^{235} residue was replaced by either a Gly or Ser residue in all the chymotrypsin-like proteins encoded by the MDP1 or MDP2 gene groups.

Among all the deduced proteins, there are six cysteine residues that are conserved at positions 72, 88, 206, 225, 237, and 269 (marked by arrows). These cysteine residues are predicted to occur in disulfide bridge configurations in trypsins and chymotrypsins (Wang et al., 1993).

3.3. Differential expression in different developmental stages

To examine the expression profiles of different proteases in different developmental stages, Northern blot analysis was performed with RNA samples derived from larvae (at different ages), pupae, and adults. The blots were examined separately for each gene group with probes specific for each. As shown in Fig. 2, all of the proteases were exclusively expressed in the larval stage. No RNA could be detected in pupae or adults. Within the larval stage, the RNA levels corresponding to each gene group differed significantly in larvae of different ages. The MDP1 gene group (chymotrypsin MDP1-A to MDP1-F) was expressed in all larval stages with the maximum amount expressed in 4-day-old larvae. In comparison, the MDP2 gene group (chymotrypsin MDP2-A and MDP2-B) was mainly expressed in 4- and 6-day-old larvae. The MDP3 gene group (trypsin MDP3-A to MDP3-C) was expressed at a low level in 2-days old larvae, significantly increased in 4- and 6-days old larvae, and then disappeared in 12-day-old larvae. The expression for the MDP4 gene group (trypsin MDP4-A to MDP4-C) was unique. It was highly expressed in 0- (freshly hatched) and 12-day-old larvae, but the expression level was significantly less in 2- and 4-day-old larvae. The MDP5 gene group (trypsin MDP5-A) was abundantly expressed in freshly hatched larvae, but no RNA could be detected in larvae after 4 days.

3.4. Analysis of tissue-specific expression through RT-PCR

RT-PCR was carried out to examine tissue-specific expression of the five gene groups in the gut and salivary glands. All PCR reactions generated DNA fragments with expected sizes. As shown in Fig. 3, the majority of the protease groups were expressed in both the gut and salivary glands. The only group that was specifically expressed in the gut was the trypsin MDP5 group.

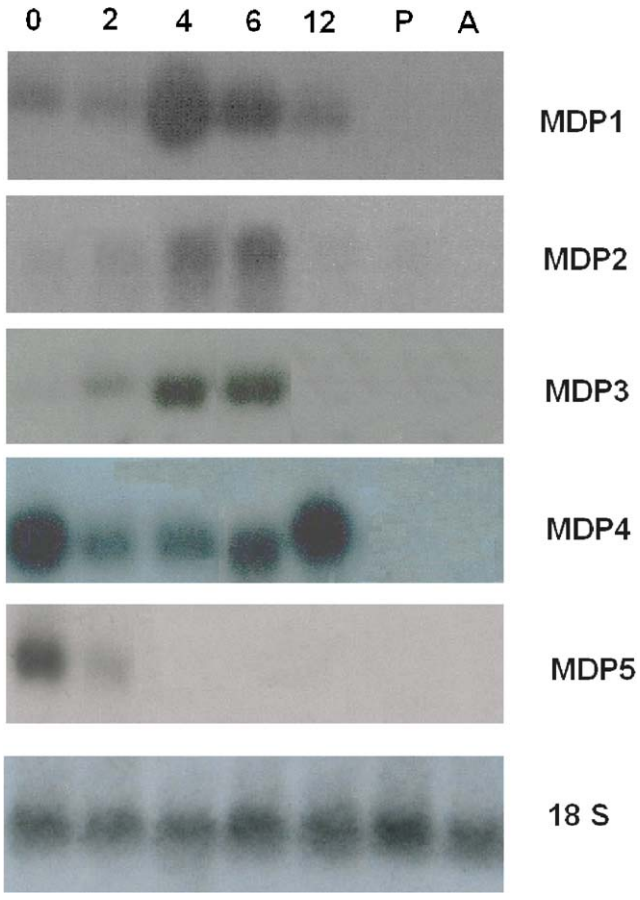
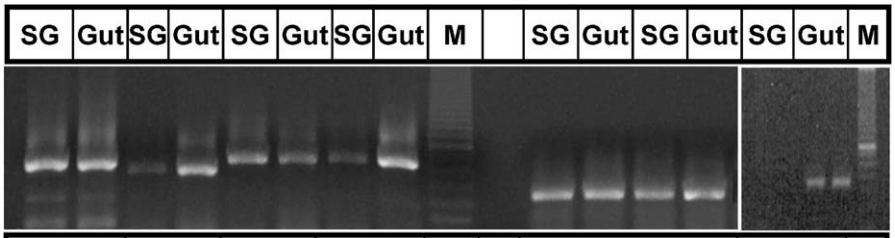


Fig. 2. Northern blot analysis. RNA samples were extracted from 0-day (freshly hatched, nonfeeding), 2-day, 4-day, 6-day and 12-day-old larvae, pupae (P), and adults (A), respectively, as indicated in the figure. Probes specific to MDP1, MDP2, MDP3, MDP4, and MDP5 gene groups were used for blot analysis. Hybridization and washing were performed as described in the Materials and Methods section. The bottom panel of the figure shows the hybridization image of a probe derived from an 18-S rRNA cDNA clone as a control for equal loading.

3.5. Protease activity in gut extract

The isolation of trypsin- and chymotrypsin-like cDNAs from a gut library, and subsequent detection of transcripts in gut by RT-PCR, indicated that serine proteinases were major digestive enzymes in *M. destructor* larvae. To further determine the composition of gut proteases, specific protease inhibitors were used to categorize proteolytic activities contributed by different digestive enzymes. From the data in Table 3 and Fig. 4, serine proteases were the major digestive enzymes because serine protease inhibitors inhibited as much as 90% of the proteolytic activities of gut extract. Inhibitors specific to cysteine proteases, aspartic proteases, or metallo-proteases were less effective, causing only 10–24% inhibition. One exception was the inhibitor bestatin, which caused nearly 90% inhibition. Bestatin



Primer Sequences			
MDP1F	ACATCAAATTTCAATGCGAAATCGA	MDP1R	GCATGAATCCAATCGAGGTACA
MDP2F	GACTCGAATTGTCGGTGGAACA	MDP2R	TCCATCTTTGGCAAGACAACTT
MDP3F	TAGCGAGCATTTTGATTGCATCA	MDP3R	GATCAACTCGGCATCATTTGTT
MDP4F	ATTGGTGTTGGTTGGTGCCTTC	MDP4R	TCCAGAGAACTTTCCACATCG
MDP5F	ATGGACGTATGGACATTTGTTTGAAG	MDP5R	CAAGTTCAAGTAATGCAAAGTCATACG

Fig. 3. Analysis of tissue-specific expression through RT-PCR analysis. Primer names are given at bottom of corresponding lanes. $M = 100$ bp molecular markers.

Table 3
Effect of protease inhibitors on the hydrolysis of BODIPY conjugated casein by *M. destructor* midgut extract

Different class of inhibitors	Concentration	BODIPY-TR-X hydrolysis (% of control)	IC_{50}
Serine protease inhibitors			
Aprotinin	13.5 μM	19.72	2.38 μM
APMSF	0.5 mM	52.25	0.51 mM
Chymostatin	5.0 mM	16.97	82 μM
SBTI	0.03 mM	11.63	2.2 μM
SBBTI	1.5 mM	10.43	1.2 μM
TPCK	0.1 mM	18.41	12 μM
Cysteine protease inhibitors			
Antipain	0.175 mM	76.01	—
E-64	0.1 mM	89.0	—
Aspartic protease inhibitor			
Pepstatin	0.05 mM	80.93	—
Metallo protease inhibitors			
Bestatin	2.8 mM	10.54	0.83 mM
EDTA	0.05 mM	135.0	—
EGTA	5.0 mM	90.59	—

inhibits metalloproteases, aminopeptidase, and exopeptidases (Taylor et al., 1993).

4. Discussion

In this report, we have identified five groups of cDNA clones that encode chymotrypsin- or trypsin-like proteins. Members from the same group encode similar proteins. These group members could represent different alleles since the cDNA library was made from multiple insects. If that is true, these cDNAs could be converted into molecular markers for genetic mapping. Different cDNAs from the same group could also represent different genes that arose by gene duplication. Indeed, most of the genes encoding secreted proteins in the salivary glands or gut of the Hessian fly have multiple copies, which are clustered within a short chromosome region (Chen et al., 2004). Clustered organization of trypsin genes has been reported in *A. gambiae* (Muller et al., 1993). Different cDNAs in the same group from *M. destructor* could be recent duplicates in the early phase for diversification.

In contrast to highly conserved sequences within the same group, members from different groups encode very different proteins. Sequence identities between proteins

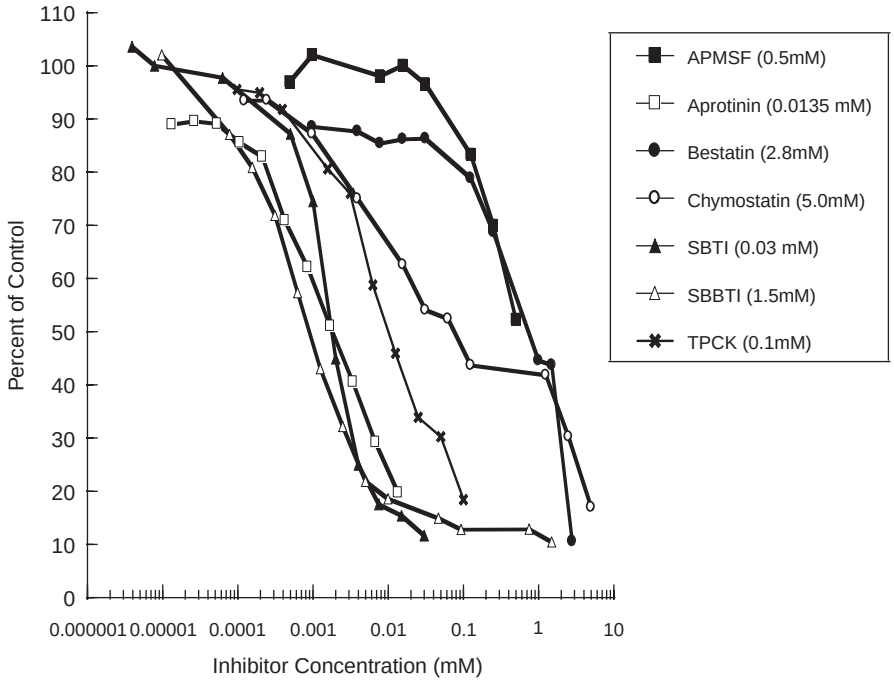


Fig. 4. Casein hydrolysis by *M. destructor* larval gut extract in the presence of various protease inhibitors.

from different groups ranged from 24% to 60% (Table 2). On the other hand, sequence identity ranged from 44% to 51% between these proteins and serine proteases from other insects (data not shown). The greater homology between proteases from different species (*M. destructor* and *A. gambiae*, for example) than between proteins from the same species indicates that these genes are orthologues that existed before the divergence of the two species.

The deduced amino acid sequences of the 15 proteins showed several structural features typical of serine proteases (Kraut, 1977), including: (i) the amino acid residues considered to determine trypsin or chymotrypsin substrate specificities; (ii) conserved histidine and serine catalytic sites; (iii) the catalytic triad; and (iv) six cysteine residues, all located at conserved positions. All these features suggested that these cDNAs encode active trypsins or chymotrypsins. All deduced proteins contained at their N-termini a typical secretion signal peptide consisting of 17–23 amino acids (Table 2, Fig. 1). Between the predicted signal peptides and the mature enzyme sequences, all deduced proteins contained a stretch of 8–14 amino acids characteristic of putative activation peptides. In addition to the three pairs of conserved cysteines that are believed to form disulfide bonds, additional cysteine residues were also found in some of the deduced proteins. These extra cysteine residues included one at position 170 in all deduced trypsins and one at position 247 in the chymotrypsin MDP2 group (Fig. 1). In addition, all proteinase precursors, except for the trypsin MDP4 group, had at least one cysteine on the putative signal or activation peptide. Extra cysteine residues have been found in vertebrate, as well as in some insect serine proteases (Colebatch et al., 2002; Jongsma et al., 1996; Zhu and Baker, 1999; Zhu et al., 2003). It has been speculated that free cysteine residues might be responsible for the vulnerability of the enzymes to some typical thio-reacting inhibitors (Jongsma et al., 1996; Gatehouse et al., 1997). The exact role and the pairing mode for disulfide bonding of these extra cysteines have yet to be established.

Different insects use different types of proteases as major digestive enzymes (Terra and Ferreira, 1994). Digestive cysteine proteinases constitute the major gut proteolytic activity in many coleopterans (Murdock et al., 1987; Matsumoto et al., 1997). Proteinase activity assay using various inhibitors demonstrated that serine proteinases were the major digestive enzymes in the gut of *M. destructor* larvae. As shown in Table 3 and Fig. 4, cysteine proteinase inhibitors reduced the proteinase activity of gut extract only marginally (~10%), but serine proteinase inhibitors suppressed gut proteinase activity as much as 90%. This observation is consistent with our transcriptomic analysis, from which we have found only two identical cysteine proteinase clones compared with 27 chymotrypsin or trypsin clones. We are currently generating antibodies which will be used for assays to establish the connections between the detected serine protease activity in the gut and the specific protease genes we have identified.

Northern blot analysis revealed that all of the proteases were exclusively expressed in the larval stage (Fig. 2). Considering the fact that adults of this insect do not feed, the exclusive expression of these genes in the larval stage indicated a role in protein digestion of ingested food. Although all of the protease genes were expressed in the larval stage, the transcriptional control of these genes seemed to be quite different. The expression level of RNA corresponding to different groups differed greatly in larvae at different ages. In addition, the abundance of the transcripts detected by probes specific to individual gene groups also differed in different larval stages. Differential regulation of similar proteases was also observed in other insects (Muller et al., 1993), but the biological implication of differential expression remains to be resolved. RT-PCR analysis revealed that all of these gene groups except MDP5 were expressed in both the gut and salivary glands (Fig. 3). *M. destructor* larvae have sucking mouthparts and these proteinases may also be injected into host plant tissue for partial digestion before sucking the juice up (Haseman, 1930; Hatchett et al., 1990; Cohen, 1998). The function of diversified serine proteinase genes and differential expression of these genes in different developmental stages remains to be characterized. Further research will be also needed on the expression of different proteinase genes in response to feeding and alteration in host plants and on the development of techniques to manipulate protein digestion in *M. destructor* larvae.

Acknowledgements

This is Contribution No. 05-34-J from the Kansas Agricultural Experiment Station, Manhattan, Kansas. Hessian fly voucher specimens (No. 150) are located in the KSU Museum of Entomological and Prairie Arthropod Research, Kansas State University, Manhattan, Kansas. The authors want to thank Drs. Gerald R. Reeck and Srini Kambhampati for reviewing an earlier version of the manuscript.

References

Behura, S., Valicente, F.H., Rider Jr., S.D., Chen, M.S., Jackson, S., Stuart, J.J., 2004. Recombination suppression in the Hessian fly revealed by a genetic map and linkage to avirulence. Genetics 167, 343–355.

Buntin, G.D., 1999. Hessian fly (Diptera: Cecidomyiidae) injury and loss of winter wheat grain yield and quality. J. Econ. Entomol. 92 (5), 1190–1197.

Burgess, E.P.J., Main, C.A., Stevens, P.S., Christeller, J.T., Gatehouse, A.M.R., Laing, W.A., 1994. Effects of protease inhibitor concentration and combinations on the survival, growth and gut enzyme activities of the black field cricket, Teleogryllus commodus. J. Insect Physiol. 40 (9), 803–811.

Chen, M.S., Fellers, J.P., Stuart, J.J., Reese, J.C., Liu, X., 2004. A group of related cDNAs encoding secreted proteins from Hessian fly [*Mayetiola destructor* (Say)] salivary glands. Insect Mol. Biol. 13 (1), 101–108.

Cohen, A.C., 1998. Solid-to-liquid feeding: the inside(s) story of extra-oral digestion in predaceous Arthropoda. Am. Entomol. 44, 103–117.

Colebatch, G., Cooper, P., East, P., 2002. cDNA cloning of a salivary chymotrypsin-like protease and the identification of six additional cDNAs encoding putative digestive proteases from the green mirid, *Creontiades dilutus* (Hemiptera: Miridae). Insect Biochem. Mol. Biol. 32, 1065–1075.

El Bouhssini, M., Hatchett, J.H., Wilde, G.E., 1998. Survival of Hessian fly (Diptera: Cecidomyiidae) larvae on wheat cultivars carrying different genes for antibiosis. J. Agric. Entomol. 15 (3), 183–193.

Gagne, R.J., Hatchett, J.H., 1989. Instars of the Hessian fly (Diptera: Cecidomyiidae). Ann. Entomol. Soc. Am. 82 (1), 73–79.

Gallun, R.L., Langston, R., 1963. Feeding habits of Hessian fly larvae on P32-labeled resistance and susceptible wheat seedlings. J. Econ. Entomol. 56 (5), 702–706.

Gatehouse, L.N., Shannon, A.L., Burgess, E.P.J., Christeller, J.T., 1997. Characterization of major midgut proteinase cDNAs from *Helicoverpa armigera* larvae and changes in gene expression in response to four proteinase inhibitors in the diet. Insect Biochem. Mol. Biol. 27, 929–944.

Hall, T.A. (Ed.), 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, Nucl. Acids. Symp. Ser. 41, 95–98.

Harris, M.O., Rose, S., 1989. Temporal changes in egglaying behavior of the Hessian fly. Entomol. Exp. Appl. 53 (1), 17–29.

Harris, M.O., Stuart, J.J., Mohan, M., Nair, S., Lamb, R.J., Rohfritsch, O., 2003. Grasses and gall midges: plant defense and insect adaptation. Ann. Rev. Entomol. 48, 549–577.

Haseman, L., 1930. The Hessian fly larva and its method of taking food. J. Econ. Entomol. 23 (2), 316–319.

Hatchett, J.H., Gallun, R.L., 1970. Genetics of the ability of Hessian fly, *Mayetiola destructor,* to survive on wheats having different genes for resistance. Ann. Entomol. Soc. Am. 63 (5), 1400–1407.

Hatchett, J.H., Starks, K.J., Webster, J.A., 1987. Insect and mite pests of wheat. In: Wheat and Wheat improvement. Agronomy Monograph No. 13, pp. 625–675.

Hatchett, J.H., Kreitner, G.L., Elzinga, R.J., 1990. Larval mouthparts and feeding mechanism of the Hessian fly (Diptera: Cecidomyiidae). Ann. Entomol. Soc. Am. 83 (6), 1137–1147.

Hedstrom, L., Szilagyi, L., Rutter, W.J., 1992. Converting trypsin to chymotrypsin: the role of surface loops. Science 255, 1249–1253.

Herrero, S., Oppert, B., Ferre, J., 2001. Different mechanisms of resistance to Bacillus thuringiensis toxins in the indianmeal moth. Appl. Environ. Biol. 67 (3), 1085–1089.

Huang, Y., Nordeen, R.O., Di, M., Owens, L.D., McBeath, J.H., 1997. Expression of an engineered cecropin gene cassette in transgenic tobacco plants confers disease resistance to *Pseudomonas syringae pv. tabaci*. Phytopathology 87 (5), 494–499.

Jang, C.S., Kim, J.Y., Haam, J.W., Lee, M.S., Kim, D.S., Li, Y.W., Seo, Y.W., 2003. Expressed sequence tags from a wheat-rye translocation line (2BS/2RL) infested by larvae of Hessian fly [*Mayetiola destructor* (Say)]. Plant Cell Rep. 22 (2), 150–158.

Jany, K.D., Haug, H., 1983. Amino acid sequence of the chymotryptic protease II from the larvae of the hornet, *Vespa crabro*. FEBS Lett. 158, 98–102.

Jongsma, M.A., Peters, J., Stiekema, W.J., Bosch, D., 1996. Characterization and partial purification of gut proteinase of *Spodoptera exigua* Hübner (Lepidoptera: Noctuidae). Insect Biochem. Mol. Biol. 26, 185–193.

Karban, R., Baldwin, I.T., 1997. In: Karban, Baldwin (Eds.), Induced Response to Herbivory. University of Chicago Press, Chicago, IL, pp. 108–115.

Kraut, J., 1977. Serine proteases: structure and mechanism of catalysis. Ann. Rev. Biochem. 46, 331–358.

Martin-Sanchez, J.A., Gomez-Colmenarejo, M., Del Moral, J., Sin, E., Montes, M.J., Gonzalez-Belinchon, C., Lopez-Brana, I., Delibes, A., 2003. A new Hessian fly resistance gene (H30) transferred from the wild grass *Aegilops triuncialis* to hexaploid wheat. Theor. Appl. Genet. 106 (7), 1248–1255.

Matsumoto, I., Emori, Y., Abe, K., Arai, S., 1997. Characterization of a gene family encoding cysteine proteinases of *Sitophilus zeamais* (Maise weevil), and analysis of the protein distribution in various tissues including alimentary tract and germ cells. J. Biochem. 121, 464–476.

Moura, D.S., Ryan, C.A., 2001. Wound-inducible proteinase inhibitors in pepper. Differential regulation upon wounding, systemin, and methyl jasmonate. Plant Physiol. 126 (1), 289–298.

Muller, H.-M., Crampton, J.M., Torre, A.D., Sinden, R., Crisanti, A., 1993. Members of a trypsin gene family in *Anopheles gambiae* are induced in the gut by blood meal. EMBO J. 12 (7), 2891–2900.

Murdock, L.L., Brookhart, G., Dunn, P.E., Foard, D.E., Kelley, S., Kitch, L., Shade, R.E., Shukle, R.H., Wolfson, J.L., 1987. Cysteine digestive protenases in Coleoptera. Comp. Biochem. Physiol. 87B, 783–787.

Murdock, L.L., Shade, R.E., Pomeroy, M.A., 1988. Effects of E-64, a cysteine proteinase-inhibitor, on cowpea weevil growth, development, and fecundity. Environ. Entomol. 17, 467–469.

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

Peterson, A.M., Barillas-Mury, C.V., Wells, M.A., 1994. Sequence of three cDNAs encoding an alkaline midgut trypsin from *Manduca sexta*. Insect Biochem. Mol. Biol. 24, 463–471.

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

Ratcliffe, R.H., Safranski, G.G., Patterson, F.L., Ohm, H.W., Taylor, P.L., 1994. Biotype status of Hessian fly (Diptera: Cecidomyiidae) populations from the eastern United States and their response to 14 Hessian fly resistance genes. J. Econ. Entomol. 87 (4), 1113–1121.

Ratcliffe, R.H., Cambron, S.E., Flanders, K.L., Bosque-Perez, N.A., Clement, S.L., Ohm, H.W., 2000. Biotype composition of Hessian fly (Diptera: Cecidomyiidae) populations from the southeastern, Midwestern, and northwestern United States and virulence to resistance genes in wheat. J. Econ. Entomol. 93 (4), 1319–1328.

Ratcliffe, R.H., Ohm, H.W., Patterson, F.L., 2003. Breeding wheat for Hessian fly resistance. Plant Breed. Rev. 22, 221–296.

Rider Jr., S.D., Sun, W., Ratcliffe, R.H., Stuart, J.J., 2002. Chromosome landing near avirulence gene vH13 in the Hessian fly. Genome 45 (5), 812–822.

Shukle, R.H., Murdock, L.L., 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 (1), 93–101.

Taylor, A., Peltier, C.Z., Torre, F.J., Hakamian, N., 1993. Inhibition of bovine lens leucine aminopeptidase by bestatin: number of binding sites and slow binding of this inhibitor. Biochemistry 32 (3), 784–790.

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

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucl. Acid Res. 22, 4673–4680.

Wang, S., Magoulas, C., Hickey, D.A., 1993. Isolation and characterization of a full-length trypsin-encoding cDNA clone from the lepidopteran insect, *Choristoneura fumiferana*. Gene 136, 375–376.

Williams, C.E., Collier, C.C., Nemacheck, J.A., Liang, C., Cambron, S.E., 2002. A lectin-like wheat gene responds systemically to attempted feeding by avirulent first-instar Hessian fly larvae. J. Chem. Ecol. 28 (7), 1411–1428.

Williams, C.E., Collier, C.C., Sardesai, N., Ohm, H.W., Cambron, S.E., 2003. Phenotypic assessment and mapped markers for *H31*, a new wheat gene conferring resistance to Hessian fly (Diptera: Cecidomyiidae). Theor. Appl. Genet. 107 (8), 1516–1523.

Zhu, Y.C., Baker, J.E., 1999. Characterization of midgut trypsin-like enzyme and molecular cloning of three trypsinogen cDNAs from lesser grain borer, *Rhyzopertha dominica* (Coleoptera: Bostrichidae). Insect Biochem. Mol. Biol. 29, 1053–1063.

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