Alimentary tract proteinases of the Southern corn rootworm (Diabrotica undecimpunctata howardi) and the potential of potato Kunitz proteinase inhibitors for larval control.

Macgregor, James Mylne

Alimentary tract proteinases of the Southern corn rootworm (*Diabrotica undecimpunctata howardi*) and the potential of potato Kunitz proteinase inhibitors for larval control.

A thesis submitted by James Mylne Macgregor B.Sc. in accordance with the requirements of the University of Durham for the degree of Doctor of Philosophy.

Department of Biological Sciences,
University of Durham.
September 2001.

Abstract.

Proteolytic digestion by larval *Diabrotica undecimpunctata howardi* (Barber) (*D. undecimpunctata*) has been investigated with the aim of producing transgenic plants possessing enhanced resistance to this economically important crop pest. Biochemical characterisation *in vitro* by pH dependent hydrolysis and inhibition assays incorporating E-64, pepstatin A and soybean Kunitz trypsin inhibitor showed the majority of hydrolytic activity occurs at pH 5.5 and is performed by cysteine and aspartic endopeptidases.

Cysteine and aspartic proteinase encoding clones were isolated from a larval alimentary tract cDNA library. Four cathepsin L-like cysteine proteinases and two cathepsin D-like aspartic proteinase cDNA clones were identified by coding homology to known proteinase sequences. Analysis of primary and secondary sequence features revealed *D. undecimpunctata* aspartic proteinase 1 exhibits features associated with cathepsins E and is proposed to be a *D. undecimpunctata* cathepsin E-like aspartic proteinase.

Cathepsin D-like aspartic proteinase inhibitors of the potato Kunitz proteinase inhibitor (PKPI) family have been isolated by PCR and expressed employing the pET expression system (Novagen). *In vitro* assays demonstrated the inhibitory activity of PKPI-A and PKPI-B inhibitors against larval *D. undecimpunctata* alimentary tract proteolytic enzymes. To the authors knowledge this work represents the first reporting of the expression and purification of biologically active PKPI proteins. *In vitro* assays incorporating oryzacystatin I and PKPI proteins resulted in increased inhibition of proteolytic activity compared to single inhibitor and uninhibited control reactions.

Inhibition assays provide evidence for the potential of a dual proteinase inhibitor strategy to arrest protein hydrolysis by larval *D. undecimpunctata*, preventing essential amino acid absorption. Further research is necessary to characterise the properties of the digestive enzymes isolated in this work and the inhibitory spectrum of PKPI proteins. Transgenic crops expressing a combination of oryzacystatin and PKPI proteins would be predicted to show enhanced resistance to insect herbivores by virtue of digestive proteolysis inhibition.

Acknowledgements.

I would like to express my appreciation for the financial support provided by the BBSRC and Syngenta PLC for this C.A.S.E studentship. My gratitude is also due to the Department of Biological Sciences, Durham University for the use of their facilities.

My appreciation is extended to Dr John Gatehouse and Dr Jane Cayley for assistance and supervision during the pursuit of this project. My appreciation also goes to Gillian Davison for maintaining the *D. undecimpunctata* cultures and for being flexible with changing plans.

Many thanks go to the following people;

Ilona Bausch.

Samantha Connors.

Philip Pattison and Jaqueline Snowden

Terry Pratchett for his obnoxious but accurate little phrase.

Ana Gabriella Renetario Rojas.

Vincent Townshend and Lynne Conlon.

Also to Florence Reeves, Jack O'Grady and Colin Illet,

who are sadly no longer with us.

To my parents possibly the greatest thanks for their advice and support.

Abbreviations

a.a.	-	amino acid.
bp	-	base pair.
cDNA	-	copy deoxyribonucleic acid.
D. undecimpunctata	-	*Diabrotica undecimpunctata howardi* Barber.
DNA	-	deoxyribonucleic acid.
EPPO	-	European and Mediterranean Plant Protection Organisation.
g	-	gram.
hr	-	hour.
L	-	litre.
M	-	Molar.
min	-	minute.
MMLV-RT-		Moloney Murine Leukemia Virus reverse transcriptase.
mRNA	-	messenger ribonucleic acid.
MCS	-	multiple cloning site.
O.D.	-	optical density.
PAGE	-	polyacrylamide gel electrophoresis.
PCR	-	polymerase chain reaction.
pfu	-	plaque forming units.
RNA	-	ribonucleic acid.
rpm	-	revolutions per minute.
sec	-	second.
Taq	-	*Thermus aquaticus.*
U.V.	-	ultraviolet.
v/v	-	volume to volume ratio.
w/v	-	weight to volume ratio.

Amino acid and base codes.

A	-	adenine.
C	-	cytosine.
G	-	guanine.
T	-	thymine.

A	Ala	Alanine
C	Cys	Cysteine
B	Asx	Asparagine or aspartic acid
D	Asp	Aspartic
E	Glu	Glutamic acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
Y	Tyr	Tyrosine
Z	Glx	Glutamine or glutamic acid

Contents.

List of figures.

Appendices.

Chapter 1. Introduction.

1.1. Summary; arable crops past, present and future.

The goal of agricultural practice over the centuries has been to develop methods of improving the quality and yield of cultivated crops. Selective breeding of related plant species has produced the elite varieties with improved traits such as increased crop quality, quantity and pathogen resistance grown by today's commercial cultivators. New technologies have revolutionised crop production, weed control by herbicides and pathogen control by pesticides allowing viable large-scale yearly monoculture farming practices to be adopted.

Yearly monoculture has provided an environment in which herbivorous pests of crop species are able to flourish due to the availability of host plants. Despite modern agricultural practices crop yield losses are estimated at $500 billion per annum worldwide, 14 % of total production being lost to insect pests (Oerke *et al.*, 1994). Agricultural losses in the developing world are a major concern considering there are over 800 million people whose diets are insufficient for their nutritional requirements (Pinstrup-Anderson, 2001).

Synthetic pesticides are major weapons in the control of crop pathogens such as nematodes, mites, fungi, bacteria and insect herbivores. Pesticide usage is associated with detrimental side effects primarily non-target toxicity by environmental contamination and human consumption health issues (Pimentel, 1992). Biotechnology offers a means to engineer future crop plants to produce strains resistant to

pathogens, reducing dependency on of inherently toxic synthetic pesticides.

1.2. Genus *Diabrotica*; agricultural pests of North America.

Diabrotica (*Insecta, Coleoptera, Chrysomelidae, Galerucinae, Luperini, Chevrolat*) species are the most damaging insect pests of the corn producing states of the North American continent, they are also present as economically important pest species of commercial crops in Canada, Mexico and Central America (Levine and Oloumi-Sadeghi, 1991). *Diabrotica virgifera virgifera* have also colonised Europe being first discovered in Serbia in 1992 and since rapidly spread being found in Hungary (1995), Croatia (1995), Romania (1996), Bosnia and Herzegovina (1997), Bulgaria (1998), Montenegro (1998), Italy (1998), Slovakia (2000), Switzerland (2000) and the Ukraine in 2001 (EPPO, 2002).

In total between 20 to 25 million acres of corn are treated annually in the United States to protect from larval corn rootworm feeding (Fuller *et al.*, 1997). Annual costs in North America from crop losses and pesticide usage due to *Diabrotica* species are estimated at $1 billion (Metcalf, 1986). Colonisation of Europe by *Diabrotica virgifera virgifera* and the possibility of subsequent colonisation by other *Diabrotica* species would be expected to result in similar economic cost to European commercial agricultural.

The genus includes *Diabrotica undecimpunctata howardi* Barber (*D. undecimpunctata*) the Southern corn rootworm also known as the spotted cucumber beetle, *Diabrotica barberi* Smith & Lawrence (*D. barberi*) the Northern corn rootworm,

Diabrotica virgifera virgifera LeConte (*D. virgifera*) the Western corn rootworm, *Diabrotica virgifera zeae* Kryson & Smith (*D. v. zeae*) the Mexican corn rootworm and *Diabrotica balteata* LeConte (*D. balteata*) known as the banded cucumber beetle.

The estimated total cost in the state of Georgia totalled $ 2,801,000 in 1995 for treating and yield loss of field corn primarily due to the billbug, lesser cornstalk borer, wireworms, western and southern corn rootworms, 175,000 acres needing treatment (McPherson *et al.*, 1996). The more polyphagous of the corn rootworms *D. undecimpunctata* is also an economic pest of cultivated melon (Gould, 1944), peanut, bean, pea, potato, beet, tomato, eggplant and cabbage crops (Stoner, 1996).

On the appearance of the primary host plants adult *Diabrotica* beetles devour the aerial regions such as cotyledons and stems, feeding appears as irregular holes in leaves and damaged growing tips. The larval stage is the most destructive in the life cycle yield being reduced by larval feeding on the root system and boring into germinating seeds. Wilting, loss of stand, stunting of host plant growth and 'goosenecking' all typify corn rootworm larval damage to commercial crops.

Diabrotica species in addition to feeding damage cause a secondary loss of commercial crop yield by infectious plant pathogen transmission. *Diabrotica* act as vectors for plant pathogens such as squash mosaic virus, maize chlorotic mottle virus, cowpea severe mosaic virus and harbour bacterium that cause bacterial wilt of cucurbits (Gould, 1969; Gergerich *et al.*, 1986). Greater control of *Diabrotica* species would therefore offer a dual benefit to commercial crop growers reducing *Diabrotica* direct feeding damage and pathogen transmission both causes of commercial crop yield loss.

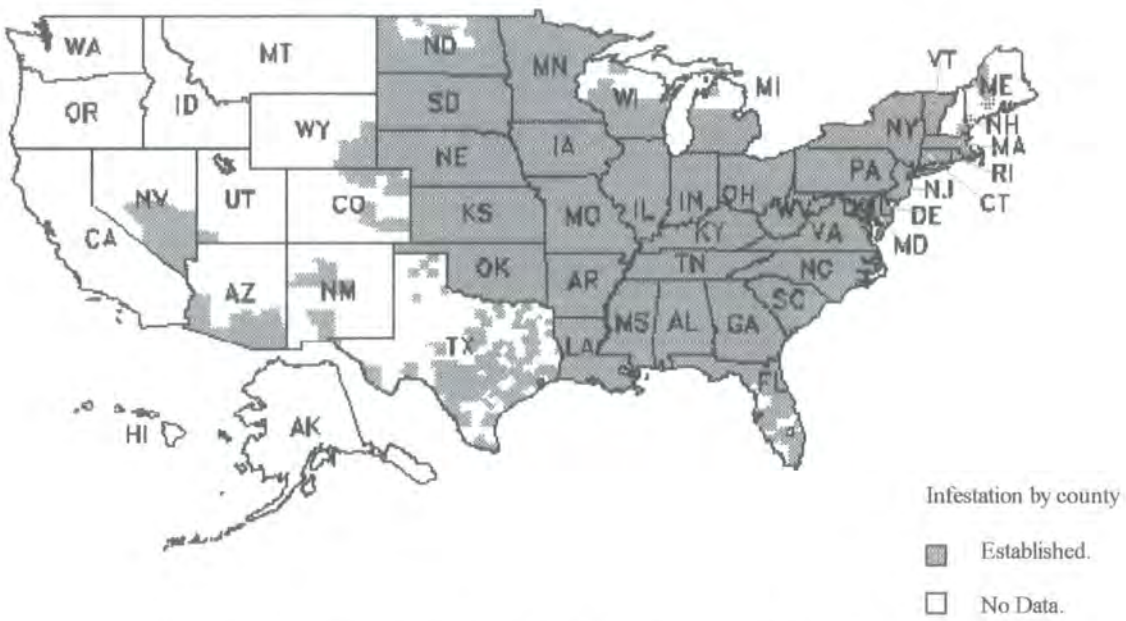
1.2.1. *D. undecimpunctata* life cycle and distribution.

Diabrotica species are native to the American continent having a combined range covering the majority of the United States also including large areas of Mexico and Central America. *Diabrotica* species are also present in central and eastern Canada (Bousquet, 1995, see online database). Phylogenetically *Diabrotica* are separated into two clades fucata; *D. undecimpunctata* and *D. balteata* and virgifera; *D. barberi, D. virgifera* and *D. v. zeae* (Szalanski *et al.*, 2000). Life cycle differences also occur between the former and latter corn rootworms. Adult *D. undecimpunctata* and *D. balteata* overwinter in foliage, whereas the eggs of the other corn rootworm species overwinter.

D. undecimpunctata adults migrate from their northern and eastern ranges to overwinter in the more temperate southern states of North America. Overwintering *D. undecimpunctata* adults emerge in late March and re-extend their range into Canada to the north (see Bousquet, 1995 online database), Mexico to the south and east of the Rocky Mountains (see figure 1.2.1a.). Adults feed on the pollen and petals of host plants before the growing season commences. *D. undecimpunctata* are lesser pests of the American corn belt due to migration from the southern states, the endogenous *D. virgifera virgifera* and *D. barberi* being the primary *Diabrotica* corn pests in this region (Levine and Oloumi-Sadeghi, 1991).

Overwintering females lay first generation eggs from late April to early June, up to five hundred eggs can be deposited by each female at the base of host plants. The eggs incubate for seven to ten days before hatching. Larvae feed on the roots and stems

Reported distribution of *D. undecimpunctata* across the Continental United States.
26/05/2000 Data retrieved from National Agricultural Pest information system (NAPIS)



The centre for environmental and regulatory information systems does not certify to the accuracy or completeness of this map.

Explanation of data source and legend keys (NAPIS).

An infestation status represents the currently reported pest infestation for a given county and is derived by examining the Observation-Date and Pest-Status of every record selected from the NAPIS DATABASE by the search criteria. The infestation status values may be interpreted as follows:

'Established' Of the selected records for the county, at least one contains Pest-Status established and it is not chronologically followed by a record containing Pest-Status of not known to be established, being eradicated, or has been eradicated.

Figure 1.2.1a. Distribution of *D. undecimpunctata* in the United States.

of the host plant for two to four weeks before pupation occurs. First generation adults emerge from late June to early July and repeat the life-cycle, second generation adults emerge from September to November. Two generations and sometimes a partial third are produced each year (Carter, 1996).

All *Diabrotica* larvae are similar in appearance being approximately ten millimetres in length, yellow-white in colour with three pairs of legs near the head and a pair of prolegs near the tip of the abdomen. Adult *Diabrotica* beetles are approximately eight to ten millimetres in length and can be distinguished by the markings and colouration of the wing covers. *D. undecimpunctata* possess a black head and legs, yellow wing covers have twelve black spots arranged in three rows, *D. virgifera* also possess a black head and legs and are distinguished by three black stripes on yellow wing covers. *D. barberi* are olive green this colouration extending to the head and wing covers the latter not possessing any markings, *D. balteata* are similar in colouration to *D. barberi* and distinguished by a red head and three to four yellow bands crossing the green wing covers (EPPO (1), 2002).

1.2.2. Current *Diabrotica* control technologies.

Yearly crop rotation with alternating years of corn and soybean cultivation was the most effective and widespread control method for *Diabrotica* species prior to chemical pesticide usage. *D. barberi* and *D. virgifera* biotypes have adapted to utilise non-corn host plants and biotypes undergoing an extended two year egg diapause have also been observed (Landis *et al.*, 1992; Levine *et al.*, 1992). Modelling the expansion of a

soybean adapted *D. virgifera* biotype in the Midwest predicts a spread of 10-30 km per year dependant on prevailing weather conditions (Onstad *et al.*, 1999). *D. virgifera* biotypes developing on soybean have also been discovered in Ottawa, Canada (Meloche *et al.*, 2001).

The development of extended diapause and *Diabrotica* species adapting to develop on soybean would be expected to reduce the effectiveness of crop rotation as a control method for *Diabrotica*. Other cultural control methods in current usage include early ploughing and/or discing, delayed planting, trap plantings of suitable host crops, employing cloth row protectors and implementing heavy seeding rates to encourage good stand growth (Sorenson, WWW. Resource; Day, 1996).

Disadvantages of cultural control methods are that they are labour intensive and are economically unattractive compared to annual maize crops. Cultural methods also fail to provide season long protection against *Diabrotica* feeding, primarily protecting the crop at the most vulnerable point of crop cultivation to reduce yield loss. Season long protection is a desirable trait due to the migratory patterns of *Diabrotica* adults and high-altitude dispersal of adult beetles.

Insecticide application both ground and aerial, is the major method of corn rootworm control. *D. undecimpunctata* exhibit variable vulnerability to different classes of insecticides. Tefluthrin a pyrethroid ester was the most potent and dieldrin a cyclodiene the least effective ovicidal agent over combined egg age, terbufos an aliphatic organothiophosphate and carbofuran a benzofuranyl methyl carbamate exhibited relative inactivity at older egg ages (Michaelides *et al.*, 1997a). Tefluthrin also provided increased maize protection against larval *D. undecimpunctata* compared to

terbufos at sub-lethal dosages (Michaelides *et al.*, 1997b). In cases of heavy infestation multiple applications are necessary to achieve pest control and reduce yield loss due to pest feeding and pathogen transmission.

Innate Coleopteran insect resistance to Cry III endotoxins is present in corn rootworm species. *Bacillus thuringiensis* (Bt) endotoxins have been shown to be less effective inducing mortality against *D. undecimpunctata* compared to Colorado potato beetle (*Leptinotarsa decemlineata*) due to inefficient binding of toxin to brush border basal membrane receptors (Rupar *et al.*, 1991; Slaney *et al.*, 1992). Bt endotoxins that in combination exhibit toxicity to *D. virgifera* have recently been described in the literature (Mollenbeck *et al.*, 2001; Ellis *et al.*, 2002). *D. virgifera* alimentary tract epithelium cells showing symptomatic damage (see section 1.9.) when larvae are fed on a diet of transgenic corn roots expressing the 14 kDa and 44 kDa proteins in combination. Toxicity to *D. undecimpunctata* however remains to be reported.

Adaptation to crop rotation, pesticide resistant biotypes, yearly maize monoculture and lack of commercially available resistant cultivars or a transgenic crop expressing an effective transgene providing resistance has lead to corn rootworms becoming one of the most destructive insect pests in the American Midwest (see Levine and Oloumi-Sadeghi, 1991 for review).

1.3. Insect pesticide and xenobiotic detoxification.

Detoxification, decreased uptake and the development of insensitivity of targets to chemical pesticides is a major problem in commercial agriculture. Over 500 species of

insect are now resistant to chemicals (Moberg, 1990). Many pesticides used to control insect populations have become increasingly redundant, indicated by reduced pest lethality, rising populations on commercial crops and reduced yields. The economic cost of insecticide resistance has been estimated to total $1.4 billion per annum in the United States and has lead to increased pesticide application with an inherent ecological cost (Pimentel *et al.*, 1992).

In regions of extensive pesticide application bacterial metabolism further reduces the active concentration of pesticide in the soil and consistency in target insect lethality (Felshot *et al.*, 1989). Exposure to sub lethal doses of pesticides assists pests to adapt and develop natural resistance by endogenous oxidative, conjugative and hydrolytic pathways mediated by target P450 mixed function oxidases, glutathione *S*-transferases and esterases (Scott, 1999, Hemmingway, 2000). The development of 'super pests' resistant to one or more pesticide classes is a major concern to agri-ecosystems, a view generally supported (Utton, 2002).

The ability of resistant strains to emerge and become the dominant population is a result of natural selection and proliferation. Insects have multiple pathways and enzymes whose function is detoxification of deleterious compounds. Enzyme families identified thus far that perform a detoxifying role in insects include the glutathione *S*-transferases and P450 mixed function oxidases. Previous research has elucidated the activity of these enzymes and has focused on their role in endogenous pathway mediated resistance to pesticides and plant allelochemicals.

1.3.1. Glutathione *S*-transferases.

Glutathione *S*-transferases (GSTs) are found in all eukaryotic organisms in which they perform a varied and primarily cytosolic function. GST roles include intracellular transport, protection from oxidative damage and detoxification of xenobiotics such as pesticides and host plant allelochemicals (Snyder *et al.*, 1995; Feng *et al.*, 1999). GSTs catalyse the conjugation of reduced tripeptide glutathione (GSH) in nucleophilic attack reactions on hydrophobic electrophilic substrates.

Insect GSTs are dimeric proteins and are divided into two serologically distinct classes GST 1 and GST 2, each with two subunits involved in glutathione and hydrophobic substrate binding (Grant and Matsumura, 1989; Fournier *et al.*, 1992). Classification of insect GST enzymes into two groups is undoubtedly an over simplification pending further evaluation of the diversity of these enzymes (Hemmingway, 2000). Class 1 GST enzymes share higher phylogenetic identity with theta and alpha, whereas class 2 GST enzymes show higher homology to sigma, pi and mu classes of mammalian GSTs (Hemingway, 2000).

Characterisation of a *D. melanogaster* GST gene cluster revealed the genes are closely associated in the genome (within 60 kb) and separated by AT rich sequences indicating each gene is controlled by individual regulatory sequences (Toung *et al.*, 1993). *Musca domestica* GST 1 was located in haemolymph cells and GST 2 distributed in indirect flight muscles and the central nervous system demonstrating organ specific regulation of GST isoforms (Franciosa and Berge, 1995).

GST activity was found to be highest in pupae and larvae of *Tenebrio molitor* (Coleoptera) indicating life cycle stage dependent regulation of GST activity (Kostaropoulos *et al.*, 1996). Diapausing *Choristoneura fumiferana* (Lepidoptera) larvae showed induction of GST transcription relative to feeding larvae (Feng *et al.*, 1999). These observations support the environment protection and developmental regulation hypotheses of GST action (Gilliot, 1980; Hazleton and Lang, 1983). Further to these natural roles, a significant body of research has linked pesticide detoxification in insects with GST family enzymes.

It is the role of GSTs in insecticide detoxification that has driven much of the research in insect GST enzymes since GST expression has been linked to insecticide metabolism (Clark, 1989). *Musca domestica* strains resistant to organophosphorus and carbamate insecticides were shown to have upregulated GST 1 transcription and exhibited increased metabolism of organochlorine substrates 1-chloro-2,4-dinitrobenzene (CDNB) and 1,2-dichloro-4-nitrobenzene (DCNB) compared to susceptible strains (Fournier *et al.*, 1992).

E. coli expressed *Drosophila* GSTs revealed differing substrate preferences for the substrate CDNB between GST isoforms GST D1 and D21. *In vitro* only GST D1 demonstrated dechlorinase activity metabolising 1,1,1-trichloro-2,2-bis-(p-chlorophenyl)ethane (DDT) an organochlorine insecticide to 1,1-dichloro-2,2-bis-(p-chlorophenyl)ethane (DDE) (Tang and Tu, 1994).

Inhibition of *Anopheles dirus* (Diptera) GST conjugation to CDNB substrate has been demonstrated in the presence of a range of organophosphate and pyrethroid insecticides, indicating GST enzymes are binding to these insecticides (Hemingway *et*

al., 1991; Prapanthadra *et al.*, 1998). Research demonstrates a similar action of Coleopteran GSTs the pyrethroid insecticide decamethrin inhibiting *Tenebrio molitor* GST conjugation of CDNB (Kostaropoulos *et al.*, 2001).

Purification of four GSTs from *Plutella xylostella* (Lepidoptera) and their subsequent characterisation revealed GST-3 and 4 exhibited substrate preferences for DCNB and organophosphates such as methyl parathion a phenyl organothiophosphate (Chiang and Sun, 1993; Ku *et al.*, 1994). A *Plutella xylostella* strain resistant to methyl parathion also showed higher activity of the GST 3 isoform compared to susceptible *Plutella xylostella* (Chiang and Sun. 1993). Subsequent molecular isolation and *E. coli* expression of *P. xylostella* GST-revealed similar biochemical properties as purified GST-3 and upregulated gene expression in methyl parathion resistant FS strain (Huang *et al.*, 1998).

Research conducted in this field shows the ability of insect encoded GSTs to interact with pyrethroid, organophosphate class pesticides and dechlorinate chlorinated hydrocarbon insecticides. GST isoforms also show enhanced expression in insecticide resistant strains compared to susceptible populations further indicating the role of GST enzymes in the *in vivo* resistance of insects to chemical pesticides.

1.3.2. Cytochrome P450 mixed function oxidases.

The P450 mixed function oxidases (CYP) family are also known as P450 cytochrome oxidases and cytochrome P450 monooxygenases, members of the CYP superfamily are found throughout nature. Current classification divides the CYP superfamily into 70

families and 127 subfamilies according to sequence similarities (Nelson, 1996). Members of a family possess equal or greater than 40 % sequence identity, a subfamily possessing 46 % identity or above (Nerbert *et al.*, 1991).

Insect CYPs have been shown to function in juvenile hormone, ecdysteroid, and pheromone metabolism also functioning in detoxification of exogenous substrates such as plant allelochemicals and pesticides (see Scott, 1999; Feyereisen, 1999 for reviews). Estimations of the number of CYP genes possessed by insect genomes vary but they are known to be numerous, 80 CYP genes have been identified in the genome of the nematode *C. elegans* (Consortium, 1998 cited in Scott, 1999). The focus of invertebrate CYP research is CYP mediated metabolism of xenobiotics such as insecticides and plant allelochemicals.

The classic chemical reaction of cytochrome P450 mixed function oxidases is the insertion of an oxygen atom into the substrate bound in the enzyme:substrate complex. Insertion of the oxygen occurs via reduction of Fe^{3} in the heme binding loop to Fe^{2} by an electron donor such as NADPH, cytochrome P450 reductase or cytochrome b_5 (Guzov *et al.*, 1996), followed by binding of O_2. Oxygen is split and the oxygen atom is inserted into the substrate. The ability of CYP enzymes to catalyse this and a wide range of other chemical reactions is fundamental to their role in the metabolism of plant allelochemicals and resistance to pyrethroid pesticides. Insertion of oxygen mediated by CYP enzymes is required for enhancement of organophosphate pesticide activity, oxygen replacing the less reactive sulphur (Feyereisen, 1999).

Plant CYP enzymes synthesize toxic allelochemicals to deter insect herbivores, insects employ their own CYP's to detoxify these toxins thus leading to animal-plant

warfare being suggested as the driving force behind CYP evolution (Gonzalez and Nerbert, 1990). *In vivo* bioassays with *Drosophila* host species dependent on alkaloid producing cacti demonstrated increased larval mortality with increased concentration of piperonyl butoxide (PBO) a CYP inhibitor. The same researchers also demonstrated *in vitro* inhibition of cactus alkaloid metabolism by PBO and *Drosophila* CYP activity induction by cactus tissue and phenobarbital a general CYP inducer (Frank and Fogleman, 1992).

Induction of P450 gene transcription by xanthotoxin and the tomato allelochemicals 2-undecanone and 2-tridecanone was demonstrated by artificial diet in *Manduca Sexta* larvae (Stevens *et al.*, 2000). Xanthotoxin also induced mRNA transcription of cytochrome P450 (CYP6B8) in *Helicoverpa zea*, as did phenobarbital to a lesser extent (Li *et al.*, 2000). Similar results showing induction of CYP gene transcription have been obtained in Sonoran desert *Drosophila* larvae on exposure to cactus alkaloids (Danielson *et al.*, 1997).

Analysis of CYP gene expression in southern house mosquito showed elevated mRNA transcription in permethrin a pyrethroid ester insecticide resistant strains and elevated CYP activity in gut tissue compared to carcass (Kasai *et al.*, 2000). Similar research showed elevated expression levels of a CYP gene (CYP6D1) in a permethrin resistant *Musca domestica* strain (LPR) compared to a susceptible strain (Liu and Scott, 1998). *D. virgifera* resistant to carbaryl a carbamate insecticide showed increased metabolism and hydrolysis in resistant York and Phelps populations (Scharf *et al.*, 1999). Increased *N*-demethylation of carbaryl and expression levels of CYP isoforms were subsequently found in the resistant *D. virgifera* in comparison to a susceptible

population (Scharf *et al.*, 2000).

Elevated CYP activity and induction of CYP gene transcription in response to plant allelochemicals and insecticides has been demonstrated in a range of insects. The number of CYP genes in the insect genome estimated to be one hundred (Feyereisen, 1999), the range of catalysable reactions and inducibility by xenobiotics indicates CYP enzymes are a numerous and varied means to combat extrogeneous toxins such as plant allelochemicals and chemical pesticides.

1.4. Biotechnology and crop protection.

Pesticidal transgene expression offers a significant advantage over conventional spray pesticide usage. Pesticide toxicity is dependent on environmental, cultural and operational conditions. Biological interactions between pesticides soil bacteria and target detoxification systems can also adversely affect pesticide persistence and thereby crop protection. Insecticidal transgene expression in transgenic crops overcomes environmental, cultural, operational and bacterial degradation factors effecting conventional pesticide efficacy.

Conventional plant breeding techniques have been in use for centuries and have led to the development of today's elite commercial crop strains currently grown. The number of plants that a conventional breeder can select attributes from and breed into a target commercial crop is limited by plant species cross-pollination specificity. Recombinant DNA technology allows the transfer and expression of agronomically beneficial genes from distantly related plants and from unrelated organisms into

economically important crops.

Recombinant DNA technology offers a means to improve agricultural performance of crops and reduce reliance on ecologically damaging chemical pesticides for control of pathogens and herbivorous insects. Transgenic crop strategies to develop enhanced resistance to herbivorous insect pests currently include the use of proteinase inhibitors, *Bacillus thuringiensis* toxins and chitinolytic enzymes (for reviews see Gatehouse *et al.*, 1993; Jouanin *et al.*, 1998; Hilder and Boulter, 1999).

1.4.1. *Bacillus thuringiensis* toxins.

Bacillus thuringiensis soil bacterium synthesise protein crystalline multimers that on ingestion exhibit toxicity to many orders of insects. The δ-endotoxins are solubilised in the insect alimentary tract and are subsequently cleaved by luminal proteases to produce the active toxin (Cry). Cry protein activity by binding and lysis of the alimentary tract epithelium disrupts the digestive system and induces rapid mortality of the target insect.

Classification of Cry proteins was originally proposed by toxic specificity to insect orders and structural similarity (Hofte and Whitely, 1989), further classifications are available in the literature. Cry I toxins are specific for Lepidoptera, Cry II toxins are specific for Lepidoptera and Diptera, Cry III are specific for Coleoptera and Cry IV are specific for Diptera. Crystal structure studies of Cry IIIa at 2.5 Å resolution revealed a pore-forming domain I, a receptor binding domain II and domain III involved in protein stability and mediating proteolytic insertion of domain I into alimentary tract epithelial cells (Li *et al*, 1991).

Pore formation is mediated by binding of domain II to cellular surface receptors in the case of gypsy moth (*L. dispar*) Cry IA(c) was preferentially bound by alimentary tract brush border membrane aminopeptidase-N, Cry IA(a), Cry IA(b) and Cry IIIa exhibiting little or no binding (Valaitas *et al.*, 1995). Proteolytic activation by cleavage of domain III leads to the insertion of domain I into the membrane of alimentary tract epithelial cells. Insertion results in cell surface lesions and increased permeability of alimentary tract epithelial cells. Subsequent cellular rupturing results in disruption of nutrient absorption and mortality by starvation and cellular damage (Endo and Nishiitsutsuji-Uwo, 1980; Knowles, 1994).

Bt strains have been widely used in biopesticide spray formulations for over twenty years with resistance development by crop pests being reported during the last decade. Field resistance to Bt sprays has been demonstrated in Central American populations of diamond back moth (*Plutella xylostella*) (Tabashnik *et al.*, 1990). Resistance to Bt toxin Cry 1A (b) in a *Plutella xylostella* strain from the Philippines was associated with reduced binding to brush border membranes (Ferre *et al.*, 1991).

Plutella xylostella populations in Florida also exhibited reduced binding of toxin Cry 1A(b) to membrane receptors resulting in reduced cellular lysis and mortality (Tang *et al.*, 1996). Confirmation of Bt resistant *Plutella xylostella* populations in Central America lead to recommendation for control strategies to be implemented to manage the development of resistance (Perez and Shelton, 1997).

Reversion of *Plutella xylostella* Cry1A(c) resistant to susceptible phenotype occurred in the absence of selection, however resistant populations were found to rapidly re-adapt on re-exposure and contained highly resistant individuals (Tabashnik *et*

al., 1994). Further *Plutella xylostella* breeding experiments found one recessive gene conferred resistance to four different Bt Cry I toxins (A, Ab, Ac and F), with 21 % percent of a susceptible population being heterozygous for the resistance gene (Tabashnik *et al.*, 1997).

Multiple mechanisms of resistance have been identified in target insects to Bt toxins. Target insect proteases function in a protoxin activation as well as having a defensive role by hydrolytic toxin inactivation (Milne and Kaplan, 1993; Keller *et al.*, 1996). Two resistant strains of the indianmeal moth (*Plodia-interpunctella*) lack a proteolytic activity involved in protoxin cleavage (Oppert *et al.*, 1997). A Bt-resistant European corn borer biotype exhibited 35 % reduced trypsin like activity and reduced hydrolysis of Cry 1A(b) protoxin compared to susceptible strains (Huang *et al.*, 1999). The interactions of pest proteases and Bt protoxins provide an additional means of resistance to the insecticidal activity of Bt toxins.

Resistance development to Bt spray formulations may affect the performance of current GM crops expressing Bt toxins. Proteinase hydrolysis of Bt proteins and receptor-mediated resistance will have been selected for in populations exposed to spray application of Bt formulations. Proteinase failure to activate the Bt protoxins will not be a factor as GM crops express an active form of the toxin. Receptor mediated resistance however may be a factor in GM agri-systems where pests have been previously exposed to Bt spray formulations.

Strategies to delay the onset of resistance have been devised to maintain the efficacy of Bt. Refuges of non-engineered crops provides Bt susceptible populations to reproduce with the resistant individuals, delaying the emergence of the resistant biotype.

Employing refuges, resistance to Bt toxins is predicted within ten years in *Heliothis virescens* populations and within four years in populations of the European corn borer and cotton bollworm (Gould *et al.*, 1997).

Chloroplast expression of Bt toxin (Cry2Aa2) exposed resistant and susceptible tobacco budworm (*Heliothis Virescens*), cotton bollworm (*Helicoverpa zea*) and beet amyworm (*Spodoptera exigua*) to doses sufficient for lethality (Kota *et al.*, 1999). Chloroplast expression shows promise in the immediate future to control leaf-eating pests. However root eating larvae of pests such as corn rootworms would be expected to be largely unaffected by this development.

Adaptation of insects to Bt toxins mirrors the development of resistance to chemical pesticides. Reliance on heavy usage of any single compound be it chemical or biological, produces a more stringent selection criteria advancing those individual insects able to adapt and reproduce in the presence of the toxin. Current opinion favours the development of integrated pest management strategies involving multiple pest control strategies reducing dependency on any single control agent.

1.4.2. Chitinases.

Chitinases degrade chitin a structural polysaccharide (β-(1-4)-*N*-acetyl-glucosamine) similar in structure to cellulose. Chitin forms a major part of the cuticle and to a lesser extent the peritrophic matrix (PM) in most insect species. The latter is a semi-permeable matrix lining the guts of most insects at some point in development (Tellam, 1996). The first chitinase sequence of insect origin was isolated from *Manduca sexta* (Kramer *et*

al., 1993) and was followed by examples from *Bombyx mori* (Koga *et al.*, 1997), *Phaedon cochleariae* (Girard and Jouanin., 1999a) and *Spodoptera litura* (Shinoda *et al.*, 2001).

PMs are semi-permeable and porous, composed of chitin, proteins and proteoglycans forming a sieve like structure with considerable strength (Peters, 1992; Tellam, 1996; Lehane, 1997). The membrane functions as a physical and chemical barrier, protecting the alimentary canal epithelial cells of the insect from pathogens such as bacteria, viruses and parasites. The PM also serves to compartmentalise ingested food, the digestive processes and the alimentary tract epithelial cells. The role of chitin as a structural component of the PM, has led to research in chitinase enzymes as a means of disrupting these functions.

Insect peritrophic membranes are classed as type 1 or type 2. Type 1 PM are secreted by all epithelial cells lining the insect midgut synthesis is induced by feeding, the PM enveloping the ingested meal. Type 2 PM are continually synthesised from the specialised cardia found in the insect midgut to form a tube into which food is passed on consumption (Tellam, 1996).

Chitinase sequences have been isolated and employed to produce transgenic plants to investigate chitinase toxicity towards herbivorous insect pests and fungal pathogens (see Kramer and Muthukrishnan, 1997 for review). Bacterial and plant chitinases when fed at 1-2 % levels in artificial diet exhibited no observable toxic effects on the merchant grain beetle (*Oryzaephilis mercator*) (Coleoptera) compared to larvae reared on control diet (Kramer and Muthukrishnan, 1997).

Recent research however shows *Trichoplusia ni* (Lepidoptera) larvae fed

transgenic tomato expressing microbial chitinolytic enzymes (endochitinase and chitobiosidase) showed increased mortality and decreased weight gain. Membrane permeability was increased in insects fed on chitinolytic enzyme expressing transgenic plants relative to those fed on control plants, FITC dextran marker molecules 50 % larger passing through the peritrophic membrane (Gongora *et al.*, 2001).

Baculovirus mediated expression of a tobacco hornworm (*Manduca Sexta*) chitinase (Kramer *et al.*, 1993) has been shown to enhance insecticidal activity of viral entomopathogens against fall amyworm (*Spodoptera frugiperda*). Larvae injected with virus expressing the chitinase gene showed decreased survival times approximately 20 hours shorter on injection compared to larvae injected with wild type virus lacking the *Manduca Sexta* chitinase gene (Gopalakrishnan *et al.*, 1995).

A *Manduca sexta* chitinase (Kramer *et al.*, 1993) purified from transgenic tobacco leaves and incorporated in artificial diet at a 2 % level resulted in complete mortality of *Oryzaephilis mercator* (Coleoptera) larvae, albeit with a restrictive test sample size of 7 larvae (Wang *et al.*, 1996). Tobacco budworm (*Heliothis virescens*) larvae exhibited reduced growth and feeding on the chitinase expressing transgenic tobacco, larvae of *Manduca sexta* however exhibited no growth reduction on the transgenic tobacco line. Complimenting chitinase transgene expression with exposure to sub lethal doses of Bt endotoxins resulted in synergistic toxicity, reducing both *Manduca sexta* and *Heliothis virescens* growth (Ding *et al.*, 1998).

The potential of insect and bacterial chitinolytic enzymes appears to be greater than that of chitinases encoded by plants for insect herbivore control. Degradation of the peritrophic membrane by chitinolytic enzymes reduces its efficiency as a defensive

barrier exposing the alimentary tract epithelial cells directly to the alimentary tract contents. The cited research indicates the potential not only of chitinases, but of the peritrophic matrix as a target for control strategies.

1.4.3. Proteinase inhibitors.

Insects like mammals cannot synthesise their full requirement of amino acids, therefore they require the diet on which they feed to provide essential amino acids and supplement their own production. Essential amino acids such as His, Leu, Ile, Lys, Val, Phe, Met, Thr, Trp and Asn are defined as those that cannot be synthesised, or produced at a sufficient rate to meet the body's requirements. These amino acids are obtained from their diet through proteolytic digestion by extracellular proteases in the gut. The free amino acids are then absorbed across the lumen wall and into the body mass for utilisation.

Supplementing artificial diet containing soybean trypsin inhibitor with the essential amino acid methionine alleviated nutritional stress and restored larval *Spodoptera exigua* weight gain to control levels (Broadway and Duffey, 1986). Reduced amino acid availability and digestive enzyme hyperproduction combine to produce a critical shortage in free amino acids available for cellular function. The *in vivo* effects of the reduction in amino acid availability are typically target mortality, reduced weight gain and stunted larval development and induction of adaptive responses (Jongsma and Bolter, 1997).

The majority of transgenic proteinase inhibitor research has focused on the production of serine proteinase inhibitors in plants for protection against Lepidopteran pests. Transformation of tobacco plants with cowpea trypsin inhibitor (CPTI) reduced *Heliothis virescens* weight gain, survivability and leaf damage (Hilder *et al.*, 1987). *Manduca sexta* feeding on tobacco leaves expressing tomato proteinase inhibitor II and potato proteinase inhibitor II showed retarded growth and caused reduced feeding damage (Johnson *et al.*, 1989).

Numerous further studies have shown the ability of proteinase inhibitors to reduce insect survival and weight gain when ingested in artificial diet or expressed in transgenic plants (for review see Jongsma and Bolter, 1997). Variable effectiveness has been observed with proteinase inhibitor and target insect combinations, barley trypsin inhibitor proving ineffective against *Spodoptera exigua* (Lara *et al.*, 2000). Soybean trypsin Inhibitor (SBTI) exhibited no significant affect on *Helicoverpa armigera* larvae (Nandi *et al.*, 1999). Tomato moth larvae exhibited only marginal toxic effects when fed transgenic potato expressing soybean trypsin inhibitor (SKTI) (Gatehouse *et al.*, 1999)

Responses to the presence of proteinase inhibitor proteins in the insect diet have been characterised as being three fold. In response to high levels of proteinase inhibitor in the diet, insects synthesise inhibitor insensitive proteases. Inhibitor insensitive proteinase induction has been observed in response to proteinase inhibitors in *Helicoverpa zea* and *Lymantria dispar* (Broadway, 1995), *Spodoptera exigua* (Jongsma *et al.*, 1995), *Spodoptera littoralis* (De Leo *et al.*, 1998) and *Helicoverpa armigera* (Harsulkar *et al.*, 1999).

Leptinotarsa decemlineata (Coleoptera) fed on potato leaves induced with methyl jasmonate showed minimal weight retardation, decreased overall proteolysis but increased inhibitor insensitive cysteine proteinase activity (Bolter and Jongsma, 1995). *Baris coerulescens* showed no adverse effects when fed transgenic oilseed rape expressing oryzacystatin I, midgut cysteine proteinase activity was reduced with serine proteinase increasing twofold becoming the primary proteolytic activity (Bonade-Bottino *et al.*, 1999).

Chronic ingestion of proteinase inhibitor proteins induced twofold upregulated activity of existing trypsin activity in *Helicoverpa zea* and *Spodoptera exigua* (Broadway and Duffey, 1986). *Spodoptera littoralis* fed transgenic tobacco expressing a trypsin inhibitor showed overproduction of existing proteolytic activity on low expressing plants. Larvae also exhibited increased weight gain, faster development and caused greater damage to leaf material (De Leo *et al.*, 1998). *Psylliodes chrysocephala* (Coleoptera) exhibited a two fold increase in cysteine and serine proteinase activity and increased weight gain when fed on transgenic oilseed rape expressing oryzacystatin I at < 1 % soluble protein (Girard *et al.*, 1998a).

Proteolytic cleavage of oryzacystatin I and Bowman-Birk inhibitor by leucine aminopeptidase and serine proteinases resulted in ineffective control of *Phaedon cochleariae* by these inhibitors (Girard *et al.*, 1998b). Demonstrating the requirement to ensure proteinase inhibitors are resistant to cleavage by endogenous target insect proteolytic activities. Research has shown that insects possess the ability to adapt to the ingestion of proteinase inhibitors. Therefore, the toxicity 'escape' mechanisms need to be addressed in future proteinase inhibitor strategies.

Factors effecting the success of proteinase inhibitors in controlling target pests include a) identification of inhibitors effective against the initial proteolytic compliment identified in the target insect alimentary tract b) resistance to proteolytic cleavage by endogenous proteolytic activities c) incorporation of inhibitors effective against the induced insensitive proteinases and d) achieving high level expression in transgenic plants. Greater control of target pests could be achieved by expressing a mix of inhibitors with differing proteinase specificities in transgenic plants (Ortego *et al.*, 1998).

Many natural proteinase inhibitors exhibit inhibitory activity against a range of proteolytic enzymes examples being potato multicystatin (Walsh and Strickland, 1993; Waldron *et al.*, 1993) and the 20-24 kDa proteinase inhibitors of potato (Walsh *et al.*, 1991). Protein engineering of inhibitory domains to inhibit primary, and response proteinases of a target insect would theoretically produce more effective proteins. Optimisation of transgene expression would also theoretically result in higher levels of multivalent inhibitor protein expression and increased toxicity by proteolysis inhibition to target pests.

1.5. Plant proteinase inhibitors.

Proteinase inhibitor proteins are found across the plant kingdom in all plant tissues and are found in high concentrations in seeds and tubers (Ryan, 1981). Inhibitors have been classified into families dependent on their inhibitory specificity and mode of inhibition (Ryan, 1989 and 1990). Proteinase inhibitors are believed to perform a tissue protective

role preventing ingested plant material protein hydrolysis thereby reducing the nutrient value of the host plant to pathogens such as herbivorous insects (Broadway and Duffey, 1986).

Research has elucidated the role of the phytohormones [*S*-(*Z*,*E*)]-5-(1-hydroxy-2,6,6-trimethyl-4-oxo-2-cyclohexen-1-yl)-3-methyl-2,4-pentadienoic acid (abscisic acid) and [1*R*-[1α,2β(*Z*)]]-3-oxo-2-(2-pentenyl)cyclopentaneacetic acid (jasmonic acid) in the wound response pathway of inhibitor gene induction. Wounded soybean stems accumulated methyl jasmonate a jasmonic acid derivative and methyl jasmonate application to soybean cells increased proteinase inhibitor gene transcription (Creelman *et al.*, 1992). Wounding and abscisic acid (ABA) application to potato plants resulted in induction of potato proteinase inhibitor gene expression (Peña-Cortés *et al.*, 1989). Exposure of tomato plants by direct spraying and airborne methyl jasmonate induced proteinase inhibitor I and II protein accumulation (Farmer and Ryan, 1990).

ABA deficient potato mutants failed to accumulate proteinase inhibitor transcripts on wounding, application of ABA and methyl jasmonate restoring mRNA accumulation (Hildmann *et al.*, 1992). Wounded potato roots failed to induce proteinase inhibitor 2 (Pin2) mRNA transcription but resulted in leaf transcription of Pin2. Root treatment with jasmonic acid induced Pin2 genes in both roots and leaves, inhibition of jasmonic acid biosynthesis reduced induction of Pin2(L), but not Pin2(R), identifying organ specific pathways exist for ABA and jasmonic acid inhibitor gene activation (Dammann *et al.*, 1997).

A model has been proposed for linking the components of the wound response pathway (Vick and Zimmerman, 1987; Farmer and Ryan, 1992a and b; Peña-Cortés *et*

al., 1995, Dammann *et al.*, 1997). The model proposes mechanical damage induces systemin release an inducer of Pin2 expression (Pearce *et al.*, 1991) and electrical signals increase the endogenous cellular ABA levels activating jasmonic acid synthesis. Jasmonic acid biosynthesis is proposed to originate by the release of the fatty acid linolenic acid (18:3) from the plasma membrane by a lipase. Insertion of oxygen into the carbon 12 double bond by a lipoxygenase, possibly a P450 mixed function oxidase is followed by allene oxide synthase, allene oxide cyclase, reductase and three rounds of β-oxidation to generate jasmonic acid (12:1).

Inhibitor genes are activated at sites distant (systemic response) from the wound site (local response) by transport of systemin a proline rich 18-amino acid oligopeptide via the phloem (McGurl *et al.*, 1992) and electrical signals (Peña-Cortés *et al.*, 1995). Leaf distal inducer signals appears not to stimulate root transcription of proteinase inhibitor genes possibly due to signals in leaves not reaching roots or root tissue insensitivity (Dammann *et al.*, 1997).

Previous research has elucidated components of the wound response pathway of plants. Accumulation of plant phytohormones promoting a local proteinase inhibitor gene induction response and the systemic induction of proteinase inhibitor genes in distal tissues. The evidence supports the view of proteinase inhibitors as anti-nutritional proteins synthesised upon insect feeding as part of the defence mechanism against insect herbivores (Broadway and Duffey, 1986).

1.5.1. The 20-24 kDa proteinase inhibitors of potato.

A wide range of proteinase inhibitors have been characterised in the family *Solanaceae* with inhibitory activity against most classes of proteolytic activity. Inhibitors of serine and aspartic proteinases (Keilova and Tomasek, 1977; Mares *et al.*, 1989), cysteine proteinases (Rodis and Hoff, 1984), and carboxypeptidase (Rancour and Ryan, 1968) have all been identified in potato. A family of 20-24 kDa tuber and leaf proteinase inhibitor proteins have been shown to possess multivalent inhibitory activity against serine, cysteine and cathepsin D-like aspartic proteinases.

Isoforms of the 20-24 kDa family isolated from potato tuber inhibited cathepsin D-like but not pepsin or cathepsin E aspartic proteinase activity and were also shown to possess trypsin inhibitory activity (Keilova and Tomasek, 1977). Potato proteinase inhibitor sequence analysis revealed close homology to soybean Kunitz trypsin and winged bean trypsin inhibitors, particularly in respect to the putative trypsin inhibitory domain (P-V-R-I) (Mares *et al.*, 1989; Ritonja *et al.*, 1990). Purification of further proteins of this family demonstrated *in vitro* chymotrypsin, subtilisin and cysteine proteinase inhibitory activity (Suh *et al.*, 1991; Walsh and Twitchell, 1991).

The 20-24 kDa potato family has been referred to as potato Kunitz proteinase inhibitors (PKPI) and divided into three subclasses (Ishikawa *et al.*, 1994a); PKPI-A (cathepsin D and trypsin inhibitors) (Mares *et al.*, 1989; Ritonja 1990), PKPI-B including PKI-1 (trypsin, chymotrypsin and subtilisin inhibitors) (Walsh and Twitchell, 1991), PKPI-C including PKI-2 (trypsin, chymotrypsin, subtilisin and papain inhibitors) (Walsh and Twitchell, 1991). Immunoblot analysis of the 20-24 kDa potato protein

family revealed expression in tubers and leaves but not in stems, stolons or roots (Suh *et al.*, 1990).

Expression studies demonstrated PKPI-A and B protein isoforms are constitutively expressed in the potato tuber, PKPI-A transcripts were also expressed in young leaf and flower buds (Ishikawa *et al.*, 1994a; Kreft *et al.*, 1997). Localised and systemic PKPI-A mRNA transcription induction occurred in mature leaves upon wounding, induction also occurring with ABA and methyl jasmonate treatment (Hildmann *et al.*, 1992; Ishikawa *et al.*, 1994a and b). PKPI-B expression was found to be limited to potato tuber even under induction (Herber *et al.*, 1994; Ishikawa *et al.*, 1994a).

PKPI gene family induction on wounding similar to Pin2 induction indicates a role in protection of plant tissue from pathogens and/or insect herbivore feeding. Research into the viability of this family of proteinase inhibitors for use as insecticidal proteins is limited to *in vitro* studies employing multiple native PKPI isoforms isolated from potato (Wilhite *et al.*, 2000). *In vivo* bioassay by artificial diet or by a transgenic plant strategy has been limited due to no reports in the literature of individual PKPI protein isoform expression and purification utilising a recombinant expression system.

1.5.2. The Cystatin superfamily.

The cystatin superfamily is a diverse protein family all linked by the ability to inhibit cysteine proteinase mediated peptide bond hydrolysis. Family 1 the stefins lack disulphide bonds, family 2 contain disulphide bonds and family 3 compromise the

kininogens. The cystatins contain a highly conserved peptide motif typically QVV(A/S)G. Examples of this super family include the human stefin A and B proteins, and the phytocystatins soyacystatin, oryzacystatins I and II. Inhibition of cysteine proteinase activity is achieved by tight and reversible binding to cysteine proteinases by complimentary binding of the inhibitor to the active site cleft (Barret *et al.*, 1986, Bode *et al.*, 1988).

Cystatin inhibitors of cysteine proteinases are widespread throughout the plant kingdom, phytocystatins further to the conserved QVVAG motif present in all cystatins possess a specific conserved motif [LVI]-[AGT]-[RKE]-[FY]-[AS]-[VI]-X-[EDQV]-[HYFQ]-N at the N-terminal (Margis *et al.*, 1998). The first phytocystatins were isolated and characterised from the endosperm of rice (Abe *et al.*, 1987; Kondo *et al.*, 1990), further examples being isolated from potato (Hildmann *et al.*, 1992: Waldron *et al.*, 1993).

One of the first cystatin clones to be isolated from plant material was oryzacystatin I isolated from rice and found to be expressed transiently in seeds (Abe *et al.*, 1987). Oryzacystatin I inhibits papain preferentially to cathepsin H, oryzacystatin II also isolated from rice inhibits cathepsin H preferentially to papain neither of the rice cystatins inhibits cathepsin B or L (Kondo *et al.*, 1990). The rice cysteine proteinase inhibitor oryzacystain I has been characterised structurally and found like other cystatins to compromise of an alpha helix and five antiparallel beta sheets (Nagata *et al.*, 2000).

Evaluation of truncated oryzacystatin proteins revealed the N-terminal twenty one amino acid s including a conserved N-terminal glycine and the C-terminal eleven

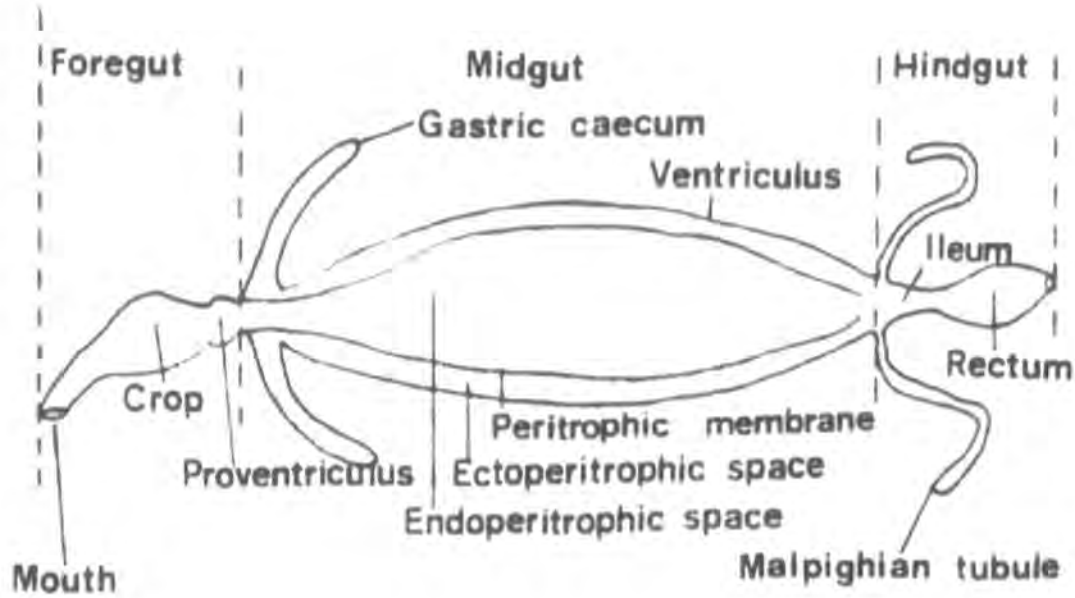
amino acids are not essential for inhibition of papain like proteinases (Abe *et al.*, 1988). Recent research has shown the presence of additional N-terminal sequence to that previously reported adding 38 amino acids, including a previously unidentified 5' hydrophobic signal sequence (Womack *et al.*, 2000). Oryzacystatin I is the most widely employed cysteine proteinase inhibitor for transgenic research into proteinase inhibitor mediated pest control strategies of *Cucujiformia*.

1.6. Insect digestive physiology.

The digestive physiology of insects has been extensively reviewed (Terra and Ferreira, 1994; Lehane and Billingsley, 1996), and is summarised here. The alimentary tract of insects is split into three sections the foregut, midgut and hindgut (see figure 1.6a). The foregut comprises the mouth, crop and proventriculus controlling diet entry to the midgut. The foregut is lined with cuticle and is the site of initial digestion, the crop can also act as a storage organ.

The midgut ventricular is tubelike in structure and lined by a single layer of epithelial cells regulating gut conditions, absorption of nutrients and digestive enzyme secretion. Digestive enzyme synthesis and secretion occurs constitutively and is regulated by prandial and paracrine mechanisms the latter being associated with cellular accumulation of secretory organelles (Lehane, 1995).

Sac-like gastric caeca branch from the anterior midgut epithelial wall. The midgut is lined with the peritrophic membrane a sieve like matrix comprised of chitin, protein and proteoglycans. The peritrophic membrane separates the midgut epithelium



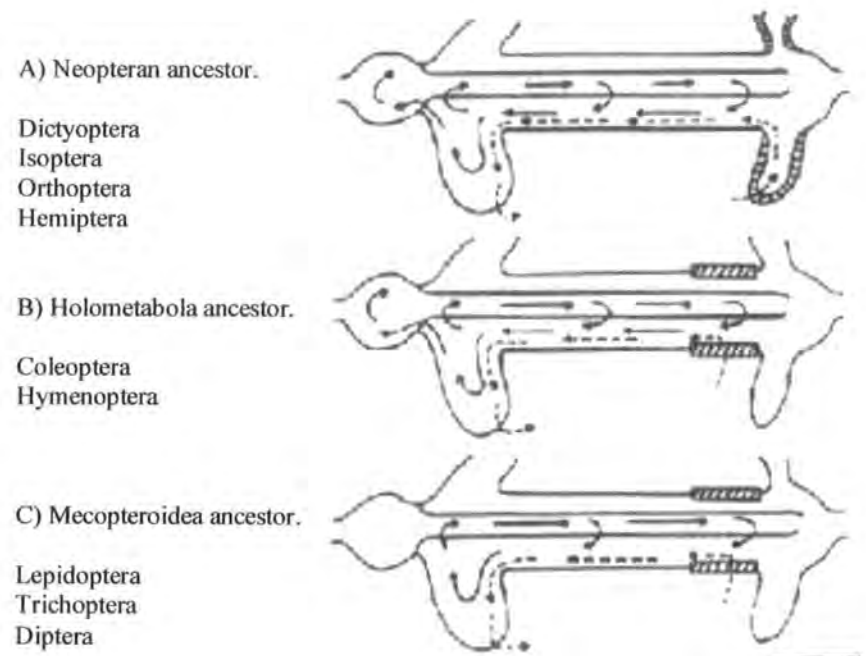


Figure 1.6a. Standardised diagram of the insect alimentary tract (top) (from Terra and Ferreira, 1994) and fluid fluxes in putative insect ancestor midguts (bottom) (from Terra, 1990). Circulation of digestive enzymes is indicated by solid arrows, circulation of fluid is indicated by dashed lines. In the bottom figure (A) represents circulation in Neopteran ancestors, (B) represents circulation in Holometabola ancestors of Coleoptera such as *Diabrotica* species and (C) represents circulation in Panorpoid ancestors. Shaded areas indicate areas of fluid secretion (Terra, 1990; Terra and Ferreira, 1994).

from ingested food protecting epithelial cells from the diet contents and compartmentalising digestion into the endo and ectoperitrophic spaces. The hindgut is separated from the midgut by the sphincter and comprises the malpighian tubules, ileum and rectum. The hindgut serves as an area of water and ion absorption prior to excretion of undigested diet.

A counter flow of fluid from posterior to anterior in the ectoperitrophic space of insect alimentary tracts is generated by secretion of fluid from the posterior midgut and absorption in the anterior midgut. Fluid fluxes vary between insects Coleopterans such as *Diabrotica* exhibiting a fluid flow similar to that of their Holometabolan ancestors (see figure 1.6a). Ectoperitrophic fluid flow allows recycling of the digestive enzymes passed from the endoperitrophic to the ectoperitrophic space in the posterior midgut (Terra, 1990).

1.7. Transport and localisation of proteinases.

Secretory, lysosomal, transmembrane and endoplasmic reticulum (ER) proteins are directed by signal peptides and inserted into the lumen of the rough ER during initial translation. In the ER pregenerated dolichol-P-P-oligosaccharides (2 *N*-acetylglucosamine, 9 mannose, 3 glucose) are selectively ß-linked to the amide nitrogen of Asparagine (Asn) in potential *N*-glycosylation sites (N-(X≠P)-S/T). Three glucose and one mannose residue are removed from the oligosaccharide ER prior to transport to the pre-Golgi apparatus.

In a pre-Golgi apparatus *N*-acetylglucosamine-1-phosphate is selectively transferred to mannose residues on *N*-linked high mannose type oligosaccharides attached to proteins destined for transport to the lysosomes. In the Golgi apparatus the *N*-acetylglucosamine monosaccharide is cleaved leaving a phosphate label on the mannose residue attached to high mannose type *N*-linked oligosaccharides. Mannose 6-phosphate receptors (MPR) in the Golgi apparatus then recognise phosphorylated high mannose type oligosaccharides and transport lysosomal proteins to lysosomes via a specialised endosome (Kornfield and Mellman, 1989).

Non-lysosomal glycoproteins exhibit low affinity for *N*-acetylglucosamine-1-phosphotransferase and oligosaccharides attached to non-lysosomal proteins are poorly phosphorylated by this phosphotransferase preventing lysosomal transport (Reitman and Kornfield, 1981). Transport of proteins to the lysosomes is known to be dependent on the possession and phosphorylation of *N*-linked high mannose type oligosaccharides proteins either not possessing *N*-linked oligosaccharides or not undergoing phosphorylation of *N*-linked oligosaccharides are not transported to the lysosomes.

Pepsinogen a non lysosomal aspartic proteinase mutated to contain cathepsin D a lysosomal aspartic proteinase glycosylation sites at Asn_{70} and Asn_{199} was glycosylated but secreted demonstrating that the phosphorylation determinants were not the oligosaccharides themselves (Faust *et al.*, 1987). Proteolytic cleavage of human lysosomal cathepsin D revealed a structural site possessed by the intact natively folded protein determined *N*-linked oligosaccharide phosphorylation at positions Asn_{70} and Asn_{199} (Lang *et al.*, 1984).

Glyco-pepsinogen cathepsin D chimeras demonstrated that Lys_{203} and amino acids 265-292 were required for efficient phosphorylation of attached *N*-linked oligosaccharides and protein retention by *Xenopus* oocytes (Baranski *et al.*, 1991). Further glycopepsinogen:cathespsin D chimeras showed recognition domains on either lobe were sufficient for phosphorylation of both Asn_{70} and Asn_{199} oligosaccharides (Baranski *et al.*, 1992; Cantor *et al.*, 1992). Shuffle mutagenesis identified two amino acids assisting in phosphorylation of the N-terminal Asn_{70} oligosaccharide His_{77} and Lys_{34}, mutation of His_{77} to Arg_{77} caused negligible phosphorylation loss (Dustin *et al.*, 1995).

N-linked oligosaccharide phosphorylation of cathepsin L and other lysosomal proteins was found to have a similar K_m and was inhibited by denaturing indicating a similar interaction of *N*-acetylglucosamine-1-phosphotransferase with lysosomal proteins was occurring determined by the native protein structure (Cuozzo and Sahagian, 1994). Site-directed mutagenesis of either cationic amino acids Lys_{54} or Lys_{99} of cathepsin L to Ala reduced phosphorylation by >70 % (Cuozzo *et al.*, 1995). Similar research demonstrated mutagenesis of Lys_{203} and Lys_{293} of human cathepsin D reduced phosphorylation of the *N*-linked high mannose type oligosaccharides by 42 % (Lys_{203}), 13 % (Lys_{293}) and 69 % (Lys_{203} and Lys_{293}) (Cuozzo *et al.*, 1998). Mutation of additional lysines further reduced phosphorylation rates indicating a cumulative effect was occurring.

Amino acids Lys_{54} and Lys_{99} were conserved with Asn_{221} of cathepsin L but do not show conservation with variation of *N*-linked glycosylation site positions other than Asn_{221}. Lys_{203} and Lys_{293} were conserved with Asn_{70} and Asn_{199} of cathepsin D and in

the presence of an Asn_{337} but not with oligosaccharides positioned at other sites in aspartic proteinases indicating a link between lysine phosphorylation determinant position and oligosaccharide phosphorylation target position (Cuozzo *et al.*, 1998).

In mammalian cells lysosomal cathepsin D is *N*-glycosylated with high mannose oligosaccharides, which are phosphorylated for subsequent recognition by mannose 6-phosphate receptors that direct the sequences to the lysosomes. Conserved lysines have been shown to determine phosphorylation of *N*-linked oligosaccharides as part of a protein surface recognition domain, phosphorylated high mannose type oligosaccharides then mediate transport to lysosomes via MPR receptors (Baranski *et al.*, 1990 and 1991, Cuozzo *et al.*, 1995 and 1998).

Insect *N*-glycosylation patterns have been studied (see Altman, 1996 for review) by baculovirus expression of recombinant mammal proteins for use as an expression system with similar post-translational modification characteristics. Baculovirus mediated expression of IgG in *Trichoplusia Ni* cells demonstrated high mannose type *N*-glycans predominated (49.3 %) in intracellular mouse IgG isolated from cell lysate, secreted IgG containing only hybrid, complex and paucimannosidic type *N*-glycans (Hsu *et al.*, 1997).

Inhibition of an insect specific beta-*N*-acetylglucosaminidase activity with 2-acetamido-1,2-dideoxynojirimycin allows galactosylation and subsequent sialylation to create complex type *N*-glycans (Watanbe *et al.*, 2002). This research demonstrated insect cells possessed the ability to generate sialylated *N*-glycans and that beta-*N*-acetylglucosaminidase activity reduces sialylated complex type *N*-glycan formation and promotes construction of the paucimannosidic type *N*-glycans observed in insect cells.

Little published research is available that describes lysosomal transport mechanisms of insects, to the authors knowledge insect equivalents of the mammalian *N*-acetylglucosamine-1-phosphate transferase and MPR proteins have yet to be identified or characterised. Baculovirus expression of IgG has however described a similar relationship with intracellular retention of proteins carrying high mannose type oligosaccharides (Hsu *et al.*, 1997). The mechanism by which intracellular retention of high mannose type *N*-glycan bearing proteins are retained in the intracellular environment remains to be elucidated.

1.8. Proteolytic enzymes.

Insect dietary protein digestion is performed by secreted peptidases in the endoperitrophic space (Terra and Ferreira. 1994; Lehane and Billingsley, 1996). Peptidase (E.C. 3.4) family enzymes are alternatively known as proteases and peptide bond hydrolases. Enzymes of this class are all linked by a common ability to cleave the C-N peptide bond linking amino acids. Peptidases are found in all phylogenetic kingdoms performing a variety of functions including developmental regulation, zymogen activation and nutrient digestion (see Barrett *et al.*, 1998 for review).

The proteases are split into two distinct classes the endopeptidases (EC 3.4.21-24) and the exopeptidase (EC 3.2.4.11-19). Endopeptidases (proteinases) catalyse the hydrolysis of peptide bonds within a polypeptide chain. The proteinases are classified by the catalytic mechanism of peptide bond cleavage determined by the active site amino acid. Endopeptidases are divided into the following sub-families; serine (EC

3.4.21), cysteine (EC 3.4.22), aspartic (EC 3.4.23) which are all found in insect digestive systems, and metalloproteinases (EC 3.4.24).

Exopeptidases catalyse peptide bond cleavage adjacent to either the N or C-terminals of a polypeptide chain. Carboxypeptidases (EC 3.4.16-18) sequentially cleave individual amino acids from the C terminal, aminopeptidases (EC 3.4.11) sequentially cleave amino acids from the N terminal. Exopeptidases have been characterised in numerous insect digestive systems (see Terra and Ferreira, 1994 for review). In insect digestive systems endopeptidases hydrolyse proteins to polypeptide chains providing N and C terminals for exopeptidase activity that hydrolyses the peptide chains to the constituent amino acids for absorption by gut epithelial cells.

1.8.1. Cysteine Proteinases.

An active site cysteine in the catalytic triad Cys-His-Asn defines the cysteine proteinases. Cysteine proteinases in conjunction with the aspartic proteinases are known as acid proteinases, the optima of such enzymes being at low pH. Members of the cysteine proteinase family include papain, the mammalian lysosomal cathepsins L, B, H, K, F, W and S (Wiederlanders *et al.*, 1992; Velasco *et al.*, 1994; Shi *et al.*, 1994; Santamaria *et al.*, 1999; Turk *et al.*, 2000). Cysteine proteinases are characterised by pH optima at mildly acidic pH, activation by addition of reducing agents and inhibition by trans-epoxysuccinyl-L-leucylamide (4-guanidino)-butane (E-64) and members of the cystatin superfamily.

The cysteine thiolate ion stabilised by the imidazolium ring of the histidine

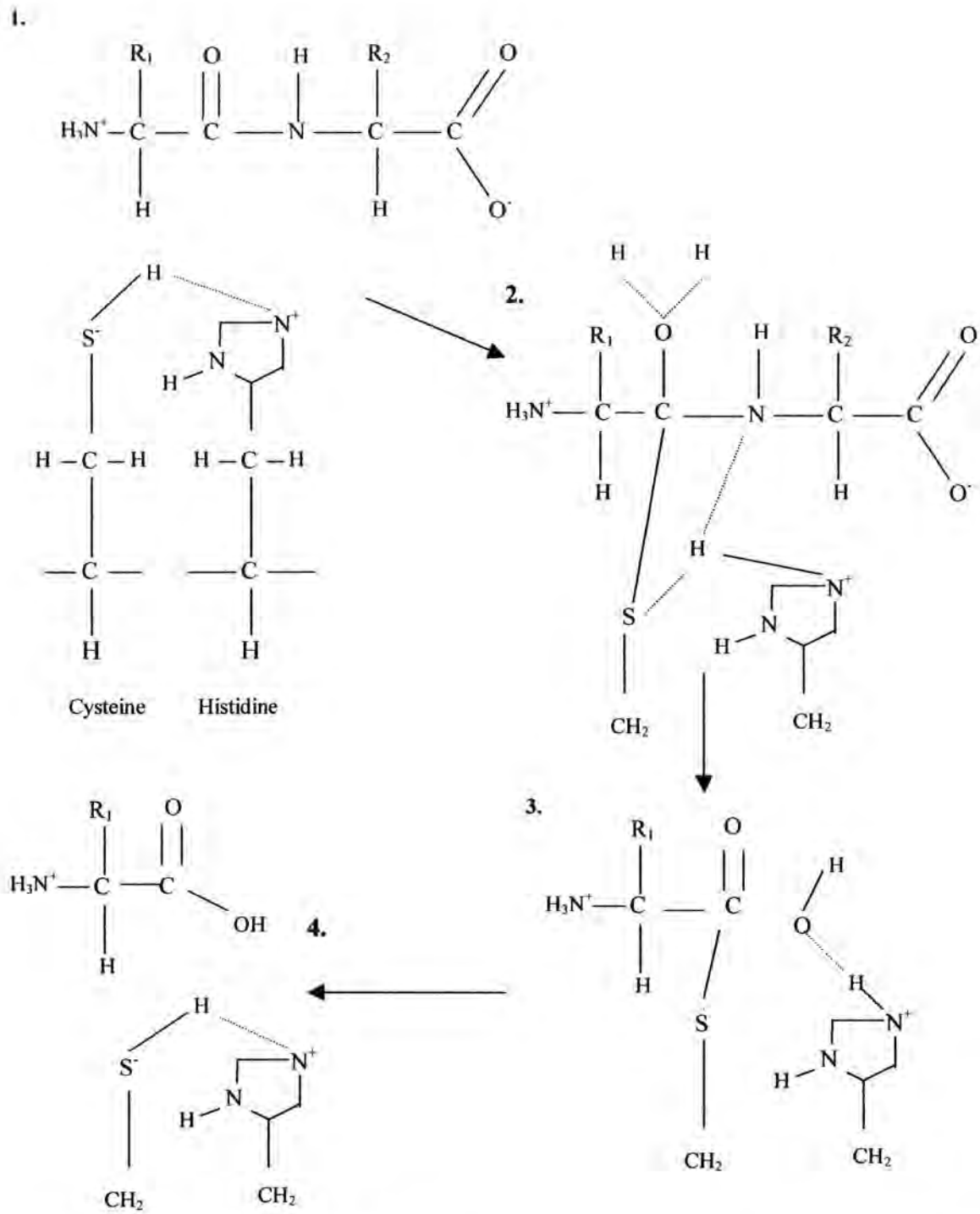


Figure 1.8.1a. Diagrammatical representation of cysteine proteinase mediated cleavage of a peptide bond. **1.** Substrate (dipeptide), catalytic cysteine and histidine. **2.** Formation of the tetrahedral intermediate; nucleophilic attack by the ionised sulphur of the sulphydryl group forming a bond to the substrate carbon. **3.** Cleavage of the peptide bond and exit of the amino terminal. **4.** Cleavage of the C-S bond by an hydroxyl ion then releases the peptide C-terminal which exits and the active site is restored. Adapted from (Dunn, 1996).

acts as the nucleophile in peptide bond cleavage. The Asn and Glu amino acids complete the catalytic site functioning to hydrogen bond the amino acid substrate. The peptide bond is cleaved by nucleophillic attack on the carbonyl group performed by the negatively charged sulphur of the thiolate ion. The oxygen of the carbonyl group is hydrogen bonded by the active site Asn and Gln. Protonation of the amino terminal nitrogen forms the tetrahedral intermediate resulting in peptide bond cleavage. The newly formed amino terminal is released and replaced by water, hydrolytic cleavage of the acyl-enzyme intermediate releases the newly formed carboxy terminal (see figure 1.8.1a.).

In Insects cysteine proteinase enzymes are associated with the alimentary tracts of Dipterans such as *Musca domestica* (Matsumoto *et al.*, 1995). Digestive cysteine proteinase activity is widespread throughout the Coleopteran series Cucujiformia (Murdock *et al.*, 1987; Wolfson and Murdock, 1990) of which *D. undecimpunctata* is a member. Cysteine proteinase activity is present in many other Cucujiformia crop pest species including rice weevil and red flour beetle (Liang *et al.*, 1991), Colorado potato beetle (Michaud *et al.*, 1993), *Baris coerulescens* (Bonade-Bottino *et al.*, 1999), corn rootworms (Purcell *et al.*, 1992; Orr *et al.*, 1994), maize weevil (Matsumoto *et al.*, 1997 and 1998) and alfalfa weevil (Wilhite *et al.*, 2000).

Cysteine proteinases perform negligible digestion in the mammalian digestive tract, enzymes of the aspartic (pepsin) and serine proteinase (trypsin/chymotrypsin) classes performing nutritional peptide hydrolysis. Ingestion of cysteine proteinase inhibitors would therefore not be expected to exhibit deleterious effects upon ingestion by humans. However, toxic effects would be expected and have been shown to occur in

targeted pests employing cysteine proteinases as digestive enzymes (discussed previously).

1.8.2. Aspartic Proteinases.

Aspartic proteinases contain a diad of aspartic acids that catalyse peptide bond cleavage. The aspartic proteinases are ubiquitous and include the mammalian pepsins, cathepsins E and cathepsins D (see Chitpinityol and Crabbe, 1998 for review). Aspartic proteinases form a mirrored bilobed tertiary structure the two similar halves each contain catalytic aspartic acids, Asp_{32} and Asp_{215} in pig pepsin (Tang *et al.*, 1973). Aspartic proteinase activity is diagnosed by optimal activity at acidic pH and inhibition by pepstatin A.

The catalytic aspartic acids are typically found in a characteristic amino acid sequence D*-(S/T)-G (*catalytic). Cleavage of peptide bonds requires interaction with a water molecule, which when polarised by the catalytic aspartic acid diad acts as the attacking nucleophile on the carbon of the peptide bond. Several models are available to explain the catalytic mechanism of aspartic proteinase mediated peptide bond cleavage (Jaskolski *et al.*, 1991), primarily differing on the position and role of hydrogen atoms. The general acid-general base mechanism is described (see figure 1.8.2a) in which the catalytic mechanism depends upon proton transfer as proposed by Veerapandian *et al* (1992) and supported by theoretical studies (Beveridge, 1998).

Peptide bond cleavage is initiated by a water molecule hydrogen bonded to the oxygen's of protonated N-terminal and ionic C-terminal aspartic acids. The scissile bond carbon undergoes nucleophilic attack by the oxygen of the bound water polarised

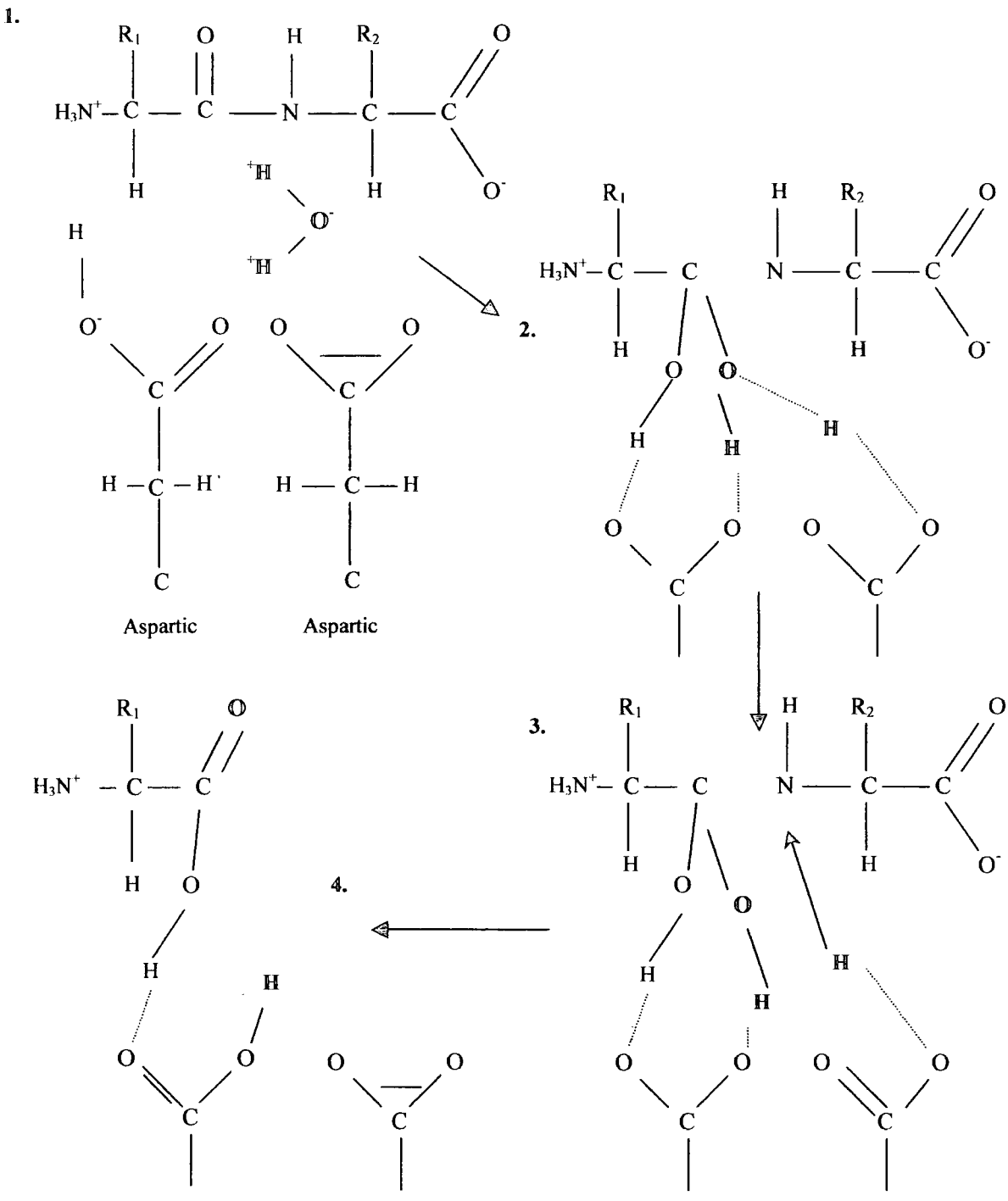


Figure 1.8.2a. Diagrammatical representation of aspartic proteinase mediated cleavage of a peptide bond. **1.** Substrate (dipeptide), catalytic aspartic acids the first of which is protonated the second is ionised. **2.** Formation of the tetrahedral intermediate 1 nucleophilic attack by the oxygen of the polarised water on the substrate carbon. **3.** Tetrahedral intermediate 2 Protonation of the peptide bond N, cleavage of the peptide bond. **4.** Simultaneous release of the newly formed amino acid N-terminal and C-terminals that exit. Adapted from (Veerapandian *et al*, 1992).

by Asp_{215} (pig pepsin numbering), the carbonyl oxygen is protonated by the hydrogen of Asp_{32}. The hydrogen of the hydroxyl group linked to Asp_{215} is donated to the amino terminal nitrogen cleaving the peptide bond and releasing the N-terminal, which exits. Peptide bond cleavage simultaneously releases the C-terminal peptide.

A *Musca domestica* aspartic proteinase was shown to be cathepsin D-like by preferable hemoglobin substrate digestion relative to albumin and failure to hydrolyse synthetic substrates of pepsin (Lemos and Terra, 1991). A diagnostic cathepsin D inhibitor has to date not been commercially available, aspartic proteinase activity defined by pepstatin A inhibition which is an effective inhibitor of cathepsin D-like, cathepsin E-like and pepsin-like aspartic proteinase hydrolytic activity.

Further aspartic proteinases from Coleopteran series Cucujiformia have shown similar cathepsin D-like aspartic proteinase activity in many of the families (Terra and Ferreira, 1994). Pepstatin A and specific cathepsin D inhibitors isolated from potato resulted in similar inhibition levels of alfalfa weevil midgut aspartic proteinase activity, further indicating the aspartic proteinase activity is cathepsin D-like in the Cucujiformia series (Wilhite *et al.*, 2000).

1.8.3. Serine Proteinases.

The serine proteinases are characterised by the essential catalytic serine amino acid, defining the class. Variation in active site amino acids is present within the family in respect to order and number. The family includes the trypsins, chymostrypsins and bacterial serine proteinases such as subtilisin. The protein structure between the families

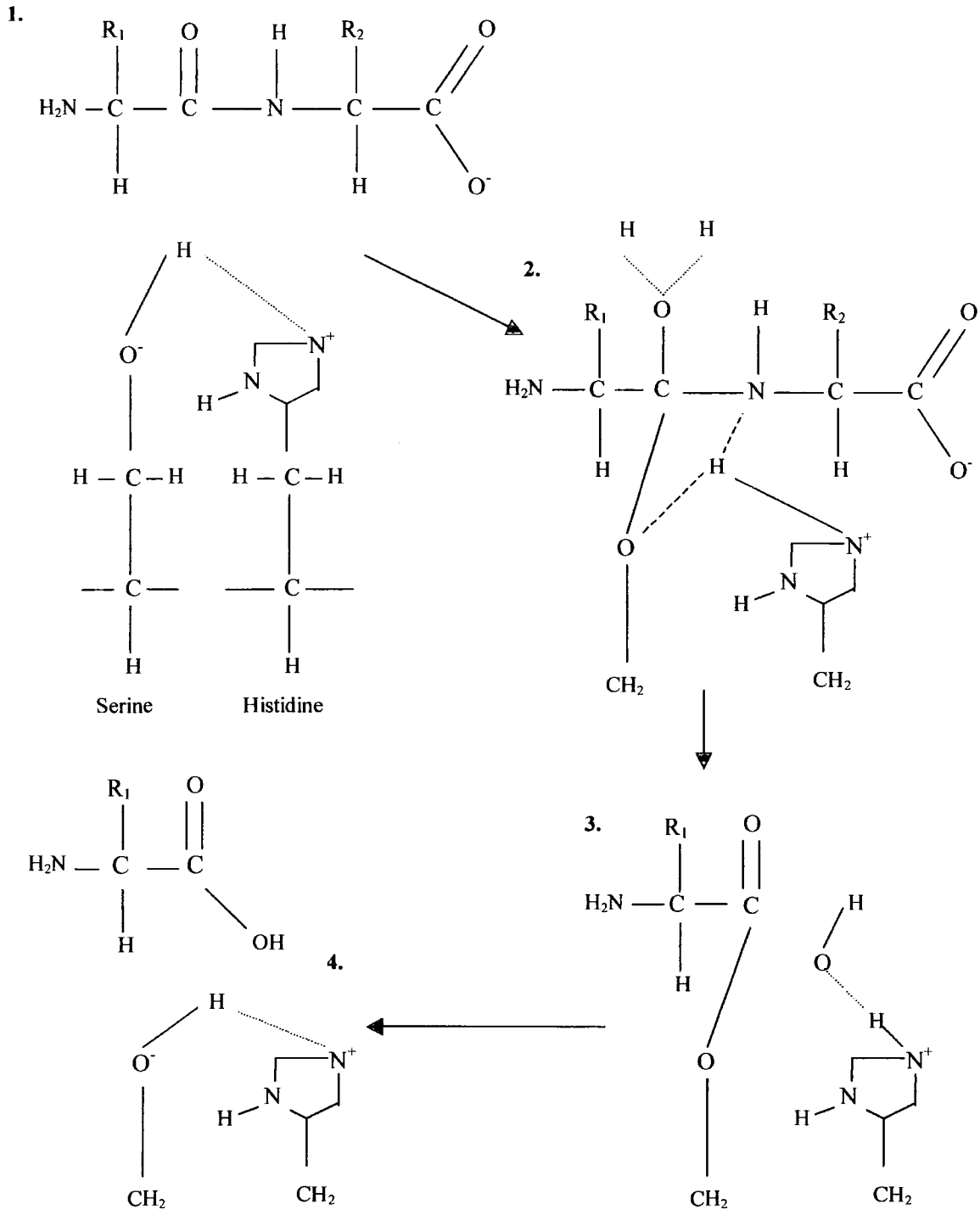


Figure 1.8.3a. Diagrammatical representation of serine proteinase mediated cleavage of a peptide bond. **1.** Substrate (dipeptide), catalytic serine and histidine. **2.** Formation of the tetrahedral intermediate; nucleophilic attack by the oxygen of the serine hydroxyl group forming a bond to the substrate carbon. **3.** Cleavage of the peptide bond and exit of the amino terminal and replacement with H_2O. **4.** Cleavage of the C-O bond by an hydroxyl ion then releases the peptide C-terminal which exits and the active site is restored. Adapted from (Dunn, 1996).

shows variation, the catalytic site geometry remains similar and the hydrolysis mechanism is the same (Barrett *et al.*, 1998). Serine proteinases show optimal peptide bond hydrolysis at alkaline pH typically between pH 8-pH 10.

Diisopropylphospoflouridate (DIFP) is a diagnostic serine proteinase inhibitor that irreversibly binds the catalytic serine amino acid. The ketones tosyl-L-phenylalanine chloromethylketone (TPCK) tosyl-L-lysine chloromethylketone (TLCK) also inactivate serine proteinases by alkylation of the catalytic site histidine. The serine hydroxyl group functions as the attacking nucleophile. The histidine stabilises the hydroxyl group and acts as a proton acceptor/donor. Two further amino acids typically Gly and a Ser hydrogen bond the carbonyl oxygen during peptide bond cleavage.

The catalytic mechanism of serine proteinases is dependent on the serine oxygen of the hydroxyl group functioning as the attacking nucleophile (see figure 1.8.3a). A proton is donated to the imidazole group of the histidine forming an imidazolium ion and enhancing the nucleophilic attack. The serine oxygen forms a covalent bond to the scissile carbon of the substrate protonation of the scissile nitrogen cleaves the peptide bond and newly formed amino terminal nitrogen exits the active site. A water molecule replaces the departing amino terminal, protones the histidine imidazolium ring and cleaves the peptide-enzyme ester bond reforming the serine O-H by and releasing the newly formed C-terminal of the peptide chain.

Serine proteinases are found throughout the insect orders as a digestive enzymes in the alimentary tract. Trypsin and chymotrypsin members of this proteinase family have been identified as being dietary digestive proteinases in most insect species studied (Terra and Ferreira, 1994). The order Lepidoptera contains some of the most damaging

pest species of commercial crops such as *Helicoverpa armigera, Manduca sexta, Plutella xylostella and Heliothis virescens* members of this order are known to predominately employ serine proteinases for nutritional digestion (Applebaum, 1985).

1.8.4. Exopeptidases.

Exopeptidases sequentially liberate individual amino acids from peptide chains, included in this family are the amino and carboxypeptidases. These peptidases are subdivided based on the catalytic nucleophile and substrate specificity. Insect exopeptidase classification is determined by the their mammalian equivalents aminopeptidases A, B, N, leucine aminopeptidase and carboxypeptidases A, B, serine and cysteine. Aminopeptidases and metallocarboxypeptidases (A and B) contain bound metal ions in the active site typically Mg^{2+} or Zn^{2+}, chelating agents such as EDTA and phenanthroline resulting in loss of enzymatic activity.

The serine and cysteine carboxypeptidases lack a bound metal ion possessing a serine or cysteine respectively in the catalytic site. These two classes are distinguishable from aminopeptidases and carboxypeptidases A and B by their insensitivity to metal chelating agents and sensitivity to inhibitors of the serine and cysteine proteinases. Exopeptidase activity has been characterised in many species in insect orders (see Terra and Ferreira, 1994 for review) including major crop pest species such as the bollworm (Bown *et al.*, 1998), western spruce budworm (Valaitas *et al.*, 1999), and corn rootworm (Edmonds *et al.*, 1996).

1.9. Plant inhibitors of *D. undecimpunctata* digestive proteinases.

Previous *in vitro* inhibition studies of *D. undecimpunctata* alimentary tract proteolytic activity show broad agreement. Cysteine proteinases are the major proteolytic activity, approximately 65 % to 70 % of proteolysis being arrested by inhibitors of this class of proteinase (Purcell *et al.*, 1992: Orr *et a*l., 1994; Edmonds *et al.*, 1996: Fabrick *et al.*, 2002). The same authors found minor inhibition with serine proteinases inhibitors such as soybean Kunitz trypsin inhibitor, potato carboxypeptidase inhibitor also being a relatively ineffective inhibitor. A secondary aspartic proteinase activity has also been identified by 34 % inhibition *D. undecimpunctata* larval proteolytic activity by pepstatin A (Edmonds *et al.*, 1996).

In vivo bioassays incorporating oryzacystatin I, potato multicystatin or a potato cysteine proteinase inhibitor (PCPI-10) reduced larval *D. undecimpunctata* survival by 50 % to 60 % (Orr *et al.*, 1994; Edmonds *et al.*, 1996; Fabrick *et al.*, 2002). This research shows the effectiveness of cysteine proteinase inhibitors when employed in artificial diet as insecticidal proteins to the larvae of *D. undecimpunctata.* Identification of a secondary aspartic proteinase activity in *D. undecimpunctata* (Edmonds *et al.*, 1996) and other Coleoptera (Wilhite *et al.*, 2000) raises the possibility of employing an aspartic proteinase inhibitor to achieve more effective control of rootworm larvae.

Plant encoded cathepsin D-like aspartic proteinase inhibitors have been extensively investigated in their role as wound responsive genes involved in plant tissue defence. Inspite of this role little published research is available describing the effectiveness of cathepsin D-like aspartic proteinase inhibitors as insecticidal toxins by

in vitro inhibition of insect herbivore proteolytic activity or *in vivo* toxicity by artificial diet bioassay against Coleopteran pests of crops such as the corn rootworms.

Previous work conducted in this field clearly shows the potential of proteinase inhibitors a large body research is available showing the effectiveness of these inhibitors when expressed in plants in enhancing endogenous resistance to herbivorous insect pests. In light of this previous research it has been proposed to transform maize with proteinase inhibitors to enhance resistance to corn rootworms by disruption of larval proteolytic digestion.

1.10. Project Aims.

The aim of this project is to generate transgenic maize with enhanced resistance to corn rootworm by the use of proteinase inhibitors using the polyphagous *D. undecimpunctata* as a model corn rootworm (See appendix A.5. for grant proposal and code). The research goals are to be achieved by conducting biochemical and molecular characterisation of the proteolytic activities present in the alimentary tracts of *D. undecimpunctata*.

Plant proteinase inhibitors of *D. undecimpunctata* digestive enzymes are to be identified and characterised *in vitro* in respect to activity against *D. undecimpunctata* digestive proteinases. The effectiveness of proteinase inhibitor proteins under *in vivo* conditions is to be assessed by artificial diet bioassay or by the production of transgenic plant material. The toxicity 'escape' mechanisms if induced by *D. undecimpunctata* due to the presence of proteinase inhibitors in the diet are also to be characterised.

Chapter 2. Materials and Methods.

2.1. Chemical and solution abbreviations.

BP blue	-	Bromophenol blue.
Carb	-	Carbenicillin.
CSPD	-	disodium 3-(4-methoxyspiro(1,2-dioxetane-3,2'-(5'-chloro)tricyclo[3,3.1.1^{3,7}]decan}-4-yl) phenyl phosphate.
DEPC	-	Diethyl pyrocarbonate.
d.H_2O	-	Deionised water.
DMSO	-	Dimethyl sulfoxide.
DTT	-	Dithiothreitol.
E-64	-	*trans*-Epoxysuccinyl-L-leucylamido(4-guanidino)-butane.
EDTA	-	Ethylenediaminetetra-acetic acid disodium salt.
EtBr	-	Ethidium Bromide.
EtOH	-	Ethanol.
HCl	-	Hydrochloric acid.
HGT	-	High gelling temperature.
IPTG	-	Isopropyl-β-D-thiogalactoside.
Kan	-	Kanamycin.
LB	-	Luria-Bertani medium.
MOPS	-	3-(*N*-Morpholino)propanesulphonic acid.
Ni-NTA	-	Nickel-nitrilotriacetic acid.
Pepstatin A	-	Isovaleryl-Val-Val-Sta-Ala-Sta(3S,4S)-4-amino-3-hydroxy-6-methylheptanoic acid.
SB	-	Sample buffer.
SDS	-	Sodium dodecyl sulphate.
SSC	-	Saline sodium citrate.
TAE	-	Tris-acetate.
TEMED	-	*N,N,N,N*, Tetramethyl-ethylenediamine.
Tris	-	Tris (hydroxymethyl) aminomethane.
X-gal	-	5-bromo-4-chloro-3-indolyl beta-d-galactopyranoside.

2.2. Materials and software.

Chemical reagents were provided by B.D.H. Chemicals Ltd, Poole, Dorset, UK or Sigma Chemicals Co (www.sigmaaldrich.com/Brands/Sigma.html), except those stated below.

3MM paper. - Schleicher & Schuell, (www.s-und-s.de/english-index.html).
Agarose. - Life technologies Inc. (Gibco BRL), (www.lifetech.com).
Agarose HGT. - FMC Bioproducts, (www.bioproducts.com).
Bacto-Agar. - Difco Laboratories, (www.vgdllc.com/DIFCO.htm).
Bodipyl FL. - Molecular Probes (www.probes.com).
Developer/fixer. - Photosol Ltd, Basildon, Essex, UK.
Digoxigenin-11-dUTP/CSPD/anti-DIG-AP fragments. - Boehringer Mannheim GMBH, (www.roche.com).
DNA size markers. - MBI Fermentas, (www.fermentas.com).
Ecolite™ Scintillation fluid. - ICN biomedicals, (www.icnbiomed.com).
Hybond-N nylon membranes. - Amersham, (www.apbiotech.com/uk).
MAP (Huang, 1994). - http://genome.cs.mtu.edu/map.html.
Megalign and Editseq software. - DNASTAR, (www.dnastar.com).
Petri-dishes. - M&Q plastic products, (www.sterilyn.com).
Pipette tips, eppendorfs and microtitre plates. - Greiner Labortechnik Ltd, Stonehouse, Gloucestershire, UK.
Protparam (Expasy tools). - http://ca.expasy.org/tools/protparam.html.
Restriction enzymes. - Promega (www.promega.com), MBI Fermentas (www.fermentas.com).
Sequencher 4.0. - Genecodes, (www.genecodes.com).
SignalP (Nielsen *et al.*, 1997). - www.cbs.dtu.dk/services/SignalP-2.0/
Smart cDNA Library Kit. -Clontech Laboratories Inc., (www.clontech.com).
Zero Blunt/TOPO TA cloning kits. - Invitrogen BV, (www.invitrogen.com).

2.3. Frequently used media and solutions.

Bacterial culture solutions.

2× YT broth (1 L):	NaCl	10 g
	Peptone	16 g
	Yeast Extract	10 g
	NaOH to pH 7.5	
	Autoclave.	
LB broth (1 L):	NaCl	10 g
	Peptone	10 g
	Yeast Extract	5 g
	NaOH to pH 7.0	
	Autoclave.	
LB plates (40):	NaCl	10 g
	Peptone	10 g
	Yeast Extract	5 g
	NaOH to pH 7.0	
	Bacto agar	10 g
	Autoclave.	

RNA gel buffers.

10× MOPS/EDTA (1 L):	0.5 M MOPS	104.63 g
	10 mM EDTA	3.72 g
	NaOH to pH 7.0	
Buffer A:	2.9 × MOPS/EDTA	
RNA gel dyes:	37 % formaldehyde	2 µl
	deionised formamide	5 µl
	Gel dyes (500 µl 10× MOPS/EDTA, 500 µl glycerol)	7 µl
	0.25 % Bp blue/Xylene cyanol ff	

DNA gel buffers.

50× TAE (1 L):	2 M Tris-acetate	242.0 g
	1 M Acetic acid glacial	57.1ml
	50 mM EDTA (pH 8.0)	18.61g
6× Loading buffer:	(Promega)	

DNA gel slice purification buffers.

50× TAE (Ultrafree-DA) (1 L):	2 M Tris-acetate	242.0 g
	2 M Acetic acid glacial	57.1 ml
	0.1 mM Na_2EDTA	1.86 g
Prep-A-Gene® binding buffer (25 ml):	6 M $NaClO_4$	21.07 g
	50 mM Tris pH 8.0	0.151 g
	10 mM EDTA pH 8.0	0.093 g
Prep-A-Gene® wash buffer (25 ml):	600 mM NaCl	1.40 g
	40 mM Tris	0.120 g
	50 % EtOH (100 %)	12.5 ml
	4 mM EDTA pH 7.5	0.037 g

SDS-PAGE buffers.

Volumes for 2× 15 % SDS-PAGE gels.

Resolving gel (15 ml):	(30 %/0.8 %) acrylamide/Bis	7.5 ml
	1 M Tris-HCl pH8.8	3.75 ml
	d.H_2O	3.04 ml
	10 % SDS	150 µl
	2 % Ammonium persulfate	560 µl
	0.0005 % TEMED	7.5 µl
Stacking gel (10 ml):	(30 %-0.8 %) acrylamide-Bis	1.5 ml
	1 M Tris-HCl pH 6.8	1.25 ml
	d.H_2O	6.65 ml
	10 % SDS	100 µl
	2 % Ammonium persulfate	500 µl
	0.00075 % TEMED	7.5 µl
10× Reservoir buffer (1 L):	1.92 M Glycine	144.00 g
	0.25 M Tris.HCl pH 8.3	30.30 ml
	1 % SDS	10 g
2× SB buffer (50 ml):	100 mM DTT	0.772 g
	2 % SDS	1.00 g
	80 mM Tris-HCl, pH 6.8	0.485 g
	0.006 % bromophenol blue	
	15 % glycerol.	7.5 ml
SDS stain (1 L):	40 % (v/v) Methanol	400 ml
	7 % (v/v) Glacial acetic acid	70 ml
	0.05 % (w/v) Kenacid blue	0.50 g

SDS destain (1 L):	40 % Methanol	400 ml
	7 % Glacial acetic acid	70 ml

cDNA library storage and screening solutions.

Denaturation buffer (1 L):	1.5 M NaCl	87.66 g
	0.5 M NaOH	20.0 g
Renaturation buffer (1 L):	1.5 M NaCl	87.66 g
	0.5 M Tris-HCl, pH 8.0	60.56 g
20× SSC (1 L):	3 M NaCl	175.30 g
	0.3 M $Na_3C_6H_5O7.2H_2O$	88.20 g
	pH 7.0	
	Autoclave.	
Lambda buffer (10×) (1 L):	1.0 M NaCl	58.44 g
	0.1 M $MgSO_4.7H_2O$	24.65 g
	0.35 M Tris-HCl, pH 7.5	42.39 g
	Autoclave.	
1× Lambda buffer:	1:10 10× dilution of Lambda buffer	
	(0.01 %) Gelatin	

Protein purification buffers under denaturing conditions (Qiagen).

Buffer B (1 L):	8 M Urea	484.40 g
	0.1 M NaH_2PO_4	13.80 g
	Tris-HCl	1.22 g
	pH 8.0	
Buffer C (1 L):	8 M Urea	484.40 g
	0.1 M NaH_2PO_4	13.80 g
	Tris-HCl	1.22 g
	pH 6.3	
Buffer D (1 L):	8 M Urea	484.40 g
	0.1 M NaH_2PO_4	13.80 g
	Tris-HCl	1.22 g
	pH 5.9	

Renaturation Buffer (1 L):	6 M Urea	363.30 g
	500 mM NaCl	29.22 g
	20 mM Tris-HCl	2.42 g
	20 % glycerol	100 ml
	pH 7.4	
Renaturation Buffer (1 L):	1 M Urea	60.54 g
	500 mM NaCl	29.22 g
	20 mM Tris-HCl	2.42 g
	20 % glycerol	100 ml
	pH 7.4	
Elution buffer (1 L) as above with	250 mM imidazole	17.02 g

Protein purification buffers under native conditions (Qiagen).

Lysis buffer (1 L):	50 mM NaH_2PO_4 pH 8.0	6.90 g
	300 mM NaCl	17.52
	10 mM imidazole	0.68 g
Wash buffer (1 L):	50 mM NaH_2PO_4 pH 8.0	6.90 g
	300 mM NaCl	17.52 g
	20 mM imidazole	1.36 g
Lysis buffer (1 L):	50 mM NaH_2PO_4 pH 8.0	6.90 g
	300 mM NaCl	17.52 g
	250 mM imidazole	17.02 g

2.4. Methods.

All standard methods are those practised in Durham University, Department of Biological Science or adapted protocols from Molecular Cloning: A laboratory manual Sambrook *et al.*, (1989).

2.5. Larval *D. undecimpunctata* culture.

Corn rootworm eggs were either obtained from French Agricultural Research Inc. or Syngenta PLC. Culturing was performed as previously described on maize seedlings (Edmonds, Thesis 1996; Edmonds *et al.*, 1996). G. Davison performed insect culturing all life cycle stages were cultured at 25°C. Briefly, eggs were maintained at in the dark at in closed ventilated containers prior to hatching. Hatchlings were transferred to maize beds that were replaced weekly to provide fresh diet. Adults were transferred to ventilated cages, maintained in 16 hours light and 8 hours dark light regime and provided with a cabbage leaf and artificial diet (Branson *et al.*, 1975).

2.6. Biochemical methodology.

2.6.1. Alimentary tract extraction.

Ten *D. undecimpunctata* larval guts were dissected be removal of head and tail section. Alimentary tracts were pulled free from the carcass with watchmakers forceps, excess

fat body was removed and the alimentary tracts suspended in 1 ml of chilled deionised water. The suspended guts were homogenised in a teflon homogeniser, lipid content was reduced by extracting with four volumes of chloroform. The preparation was vortexed and centrifuged at 10 000g in a Beckman Ja20 rotor (4°C 20 min). The aqueous layer was aliquoted into 250 µl samples, frozen in liquid nitrogen and stored at -20°C.

2.6.2. Proteolysis assays system.

Larval alimentary tract proteolysis assays were performed in 96 well Labsystems Cliniplate microtitre plates. BODIPYL® FL labelled casein substrate (Molecular Probes) was diluted 200 µg in 2 ml of dH_2O (0.1 µg/µl). Each assay contained 5 or 10 µl (variable dilutions of neat extract employed) of the gut enzyme preparation and 10 or 15 µl of substrate. Total volume was made to 200 µl per assay by buffer containing 0.05 % (v/v) Brij 35, 1 mM EDTA, 1mM L-cysteine. Assays were initiated by the addition of substrate, a 5 minute incubation period was allowed to allow reaction equilibrium to be reached.

Proteolytic cleavage and release of the intramolecularly quenched BODIPYL® FL green fluorescent labelled casein was performed at 30°C in a Labsystems Fluoroskan Ascent fluorimeter. Filters were set to 458 nm and 538 nm for excitation and emission respectively, 30 readings (2 min intervals) were taken with a prior 25mm 5 second shake at 120 rpm. Ascent research edition V 1.3.3. software was employed to collect and collate the generated data.

2.6.3. pH dependent proteolysis assays.

General conditions were as described previously. Three different buffers were employed, acetate buffer (pH 3.5-5.5), Bis-Tris propane (bis[2-Hydroxyethyl]imino-tris[hydroxymethyl]methane) (pH 6-9.5) and CAPS (3-[cyclohexylamino]-1-propanesulfonic acid) (pH 10-10.5). Stock solutions (0.1 M) were adjusted to the required pH with NaOH or HCl. Assays were performed with 5 µl of 1:2.5 diluted alimentary tract extract and 10 µl of substrate (0.1 µg/µl).

2.6.4. Proteolysis inhibition assays.

Inhibition assays were performed under identical conditions to those already described employing 0.1 M acetate buffer and 0.05 or 0.0125 gut equivalents. Assays were performed at pH 4.5 or pH 5.0 to match the estimated pH of larval *D. undecimpunctata* guts (Edmonds, unpublished). A pre-incubation period of 5 minutes was observed to allow inhibition equilibrium to be reached, prior to the addition of substrate. Inhibitor volumes were limited to 10 µl per assay.

Isovaleryl-Val-Val-Sta-Ala-Sta(3S,4S)-4-amino-3-hydroxy-6-methylheptanoic acid (Pepstatin A) was prepared as a 1 mM stock in DMSO(9):acetic acid(1) and heated to 50°C, final dilutions were performed in d.H_2O. *trans*-Epoxysuccinyl-L-leucylamido(4-guanidino)-butane (E-64), soybean Kunitz trypsin inhibitor (SKTI) and Oryzacystatin 1(rOzc) were dissolved in deionised water. rASPI 1 and 2 were solubilised in 0.1 M acetic acid as described in Chapters 5 and 6.

2.7. RNA extraction and analysis.

2.7.1. Larval gut dissection.

Heads and rear 1-2 mm of *D. undecimpunctata* larvae were detached, the gut was then pulled free and any attached fat body removed. Larval guts were placed in chilled Buffer A (300 mM Mannitol, 17 mM Tris, 5 mM EDTA) before freezing in liquid nitrogen (10 guts in 50 µl). The dissected guts were stored at -80°C prior to RNA extraction.

2.7.2. Total RNA extraction and purification.

Total RNA purification was performed with Tri-Reagent (Sigma) according to technical bulletin MB-205 instructions, utilising a ratio of 1 ml Tri-Reagent to 10 guts (50 µl). Quantitative spectrophotometric analysis indicated a 260/280 purity of 1.70 and a yield of 2.92 mg of total RNA from two hundred larval guts. The sample was diluted to a concentration of 5 µg/ml and aliquoted into eppendorff tubes for storage in liquid nitrogen or at –80°C.

2.7.3. mRNA extraction and purification.

mRNA was isolated from 800 µg of RNA with the Poly-A-Tract system IV (Promega) according to manufacturers instructions (TM021). Spectrophotometric analysis

indicated a 260/280 ratio of 2.7 and a yield 9.2 µg. mRNA was precipitated overnight at -20°C in diethylpyrocarbonate (DEPC) treated 3 M sodium acetate pH 4.8 and 100 % ethanol. Following precipitation, the mRNA was centrifuged (10,000g at 4°C for 1 hour), the precipitate was dried under vacuum and resuspended at 0.5 µg/µl in DEPC treated water. The sample was aliquoted in 1.5 µg fractions and stored in liquid nitrogen.

2.7.4. RNA electrophoresis.

RNA gels were run under denaturing conditions according to the recommendations of Sambrook *et al* (1989).Gels contained the following reagents 1.4 g Seakem® HGT agarose (FMC Bioproducts), 63 ml d.H_2O, 9.3 ml 10× MOPS/EDTA pH 7.0 and 17 ml 37 % formaldehyde. Sample volumes of less than 4.5 µl were combined with 4.4 µl buffer A and 11.6 µl of a mix of 37 % formaldehyde (89 µl) in deionised formamide (250 µl). Sample was heated at 70°C for 10 minutes and cooled on ice, the RNA gel was pre-run for 10 minutes at 85V in 1× MOPS/EDTA running buffer within a ScotLab gel apparatus.

Gel loading buffer (1.5 µl) was added to each sample and the samples loaded, electrophoresis was performed 80V. Gels were stained in a solution of 5 µg/ml ethidium bromide (EtBr) for 40 min followed by destaining in multiple washes of d.H_2O.

2.8. Polymerase Chain Reaction (PCR).

2.8.1. PCR primer design.

The EMBL or Genbank databases were searched for target genes using a proteinase family name or inhibitor class. The amino acid code of chosen sequences were transferred into Megalign via EditSeq (DNASTAR). A Clustal alignment was performed and conserved motifs identified for PCR primer design. Translation of amino-acid to DNA was performed using the standard genetic code.

2.8.2. Reverse-transcriptase PCR.

First strand synthesis was performed using 0.5 μg of mRNA, mRNA was combined with 1 μl poly-Tx primer (20 μM), 1μl of RNase inhibitor and DEPC treated water added to a total volume of 22 μl. Reactions were incubated at 70°C for 10 minutes then incubated for 5 minutes at room temperature followed by incubation on ice for a further 5 minutes. 8 μl of MMLV-RT 5×buffer, 2 μl dNTP's (10mM), 1 μl of MMLV-RT and 7 μl DEPC treated water was added to a total volume of 45 μl. The first strand cDNA reaction was then incubated for 1 hour at 37°C to allow copy strand synthesis.

An Applied Biosystems GeneAmp® PCR System 2400 system was employed for thermal cycling. Mastermix incorporated 41.6 μl 25 mM magnesium chloride, 65 μl 10× Taq buffer ($MgCl_2$ free), 13 μl 10 mM dNTP's and 400.4 μl of DEPC water (total 520 μl/ 13 reactions). Individual PCR reactions contained 40 μl of mastermix, 2.5 μl

(20 μM) of each primer, 3 μl of first strand reaction and 2 μl of DEPC water in a total reaction volume of 50 μl. Thermal cycling was as standard and performed for 40 cycles. PCR reactions were run on a 1.5 % agarose gel, amplification products were excised with a scalpel and stored at -20°C.

2.8.3. *Taq* polymerase PCR.

Taq (*Thermus aquaticus*) PCR was employed to amplify desired sequences from cDNA libraries and plasmids. A standard PCR mix was employed with variations dependent on template and yield requirement. cDNA library samples were preboiled for 4 min 30 sec prior to use as template for PCR reactions. PCR reactions were performed in 0.2 ml PCR tubes (Greiner) utilising and Applied Biosystems GeneAmp® PCR System 2400. Samples were frozen until analysed by agarose gel electrophoresis.

Standard PCR mix per reaction.

10× PCR buffer ($MgCl_2$ free).	5 μl
$MgCl_2$ (25mM).	3.2 μl
dNTP's (10 mM).	1 μl
d.H_2O.	34 μl
Primer 1 (100 mM).	0.4 μl
Primer 2 (100 mM).	0.4 μl
Taq DNA polymerase.	1 μl (0.5 units/μl)
Template.	5 μl

Standard PCR conditions.	94°C	5 minutes.	Initial denaturing
	94°C	30 seconds.	Cyclic denaturing
	47°C	30 seconds.	Annealing
	72°C	45 seconds.	Extension
	72°C	7 minutes.	Final Extension.

2.8.4. Advantage 2 PCR.

PCR reactions requiring high fidelity amplification products were performed with the Advantage 2 polymerase mix (Clontech) with minor modifications. Thermal cycling was limited to 10-17 cycles to ensure high fidelity PCR amplification. Samples were frozen at -20°C until analysed by agarose gel electrophoresis.

Advantage 2 (Clontech) PCR mix per reaction.

38µl	d.H_2O.
5µl	10× Advantage 2 PCR buffer.
1µl	50× Advantage 2 dNTP mix.
1µl	primer 1 (20µm).
1µl	primer 2 (20µm).
1µl	50× Advantage 2 polymerase mix.
3µl	template.

Advantage 2 PCR conditions.	95°C	20 seconds.	Initial denaturing.
	95°C	5 seconds.	Cyclic denaturing.
	50°C	30 seconds.	Annealing.
	68°C	6 minutes.	Extension.
	68°C	6 minutes.	Final Extension.

2.8.5. Cloning PCR products.

TAQ polymerase amplified PCR products were cloned into pCR®2.1-TOPO® using the TOPO TA Cloning® kit (Version L 012601 25-0184) (Invitrogen). Advantage 2 polymerase mix (Clontech) amplified PCR products were cloned into pCR®-Blunt II-TOPO® using Zero Blunt™ TOPO™ (Version D 190305 25-0215) (Invitrogen).

2.9. DNA manipulation and sequence analysis.

2.9.1. DNA electrophoresis.

DNA samples to be electrophoresed were combined at a ratio of 1:5 with 6× DNA loading dye (MBI Fermentas). Gels contained a final concentration of 1× TAE, 0.15 μg/ml EtBr and 1-2 % agarose depending on the DNA length to be electrophoresed. Gels were run in 1× TAE buffer at a maximum of 100V using a ScotLab gel tank and DNA marker standards λDNA/Eco471 (AvaI1) or ΦX174 DNA/*Hae* III (see appendix section A.1.).

2.9.2. Visualisation and photography of EtBr stained gels.

A Chromato - Vue® transluminator model 61 (Ultra violet products Inc.) U.V. illuminator was utilised to visualise stained gels. Photographs were taken with an MP-4 land camera (Polaroid) loaded with 667 black and white film (Polaroid).

2.9.3. Silica fines DNA gel slice purification.

DNA purification from agarose gels was performed by a standard protocol utilising Prep-A-Gene® Binding and wash buffers (Bio-Rad Laboratories). Gel slices were heated to 65°C in 1 ml of Prep-A-Gene® binding buffer until the slice had melted. Silica fines was added (20 μl) and the sample was allowed to cool to room temperature

with occasional agitation (1× 5 minutes).

The sample was then centrifuged (6500 rpm, 45 sec). After removal of the supernatant 1 ml Prep-A-Gene® wash buffer was added and the silica fines resuspended. Two spins (6500 rpm, 30 sec) and resuspensions replacing the supernatant with fresh 70 % EtOH were performed. DNA pellets were air dried and 20 µl $d.H_2O$ added. Mixing was performed for 5 min, the silica fines pelleted by centrifugation and the supernatant containing DNA transferred to a fresh tube. Sample was stored at -20°C.

2.9.4. Restriction endonuclease digestion of DNA.

Restriction digests were performed at the temperatures suggested by manufacturers and using restriction buffers suggested by the Promega Restriction Enzyme Guide (Promega, www.promega.com/reguide/). Standard digests consisted of the following; 2 µl appropriate buffer, 20 units of each restriction enzyme, 10 µl plasmid sample and $d.H_2O$ to 20 µl. Digests were performed for 4 hours. Restriction reactions to be silica fines gel purified were performed overnight with a 40 µl total reaction volume incorporating 4µl of buffer, 20 units of each endonuclease, 20-25 µl of plasmid preparation and $d.H_2O$.

2.9.5. DNA ligation.

Ligation of vector and insert DNA reactions were performed overnight according to Promega instructions (Protocols and Applications guide, Third edition). Inserts and

vectors were digested with appropriate restriction endonucleases and silica fines gel purified. Ligation reactions were performed with 2 µl of T4 DNA ligase (Promega), 2 µl of 10× T4 DNA ligase buffer and a 2:1 ratio of insert to vector gel purified samples d.H_2O was added to a final volume of 20 µl. Reactions were incubated overnight at room temperature, T4 DNA ligase was inactivated by heat treatment at 65°C for 10 min in a waterbath. Chemically competent TOPO cells were transformed by standard protocols utilising 1 µl of the ligation reaction.

2.9.6. DNA sequencing.

DNA sequencing was performed by the departmental sequencing service (Durham University Sequencing Laboratory, Department of Biological Sciences, University of Durham, U.K.). Reactions were performed using a Perkin Elmer thermocycler 1 (Warrington, U.K.) employing the dideoxy terminator chemistry method (Sanger *et al.*, 1977) and Amplitaq®DNA polymerase FSv (Perkin Elmer). Sequence data was obtained using Applied Biosystems models 373 and 377 automated sequencers.

2.9.7. DNA sequence data analysis.

Sequence data returns were analysed employing Sequencher™ 3.0 software package (Gene Codes Corporation). Identification of sequenced clones was performed utilising the online blastx translated nucleotide for PCR fragments or the blastp protein to protein homology search engine for cDNA clones (www.ncbi.nlm.nih.gov/BLAST/).

2.10. Bacterial culture and plasmid isolation.

2.10.1. Overnight cultures.

Single colonies were transferred with a sterile toothpick from LB plates with the appropriate antibiotic (50μg/ml) and cultured overnight in 10 ml LB broth, at 37°C or 31°C (BM25.8 cells, Clontech).

2.10.2. Plasmid mini-preparations.

Plasmid mini-preps were performed using Wizard SV mini-prep kit (Promega, cat# A1460) or GenElute plasmid miniprep kit (Sigma, cat# PLN-350), following kit manufacturer instructions. Plasmid preparations were resuspended in 100 μl of d.H_2O and stored at –20°C.

2.10.3. Glycerol stocks.

Bacterial growths were performed overnight in 10 ml of autoclaved LB with the appropriate antibiotic at 37°C or 31°C (BM 25.8). A 10 μl sample and 1 ml of autoclaved 100 % glycerol was added to 3 ml stock bottles (Samco) with 1 ml of autoclaved 2× YT broth. The glycerol stock was then vortexed and stored at –80°C.

2.11. cDNA library construction and screening.

2.11.1. cDNA library generation and evaluation.

A *D. undecimpunctata* larval alimentary tract cDNA library was constructed using the SMART™ cDNA Library Construction Kit with minor modifications to the protocol supplied with the kit (Clontech, PT3000-1). Synthesis of the first cDNA strand was performed as per manufacturer's instructions using Superscript II reverse transcriptase (Gibco BRL, Life technologies) employing 1 µg of mRNA. Subsequent cDNA synthesis was performed by LD PCR (long distance polymerase chain reaction) using 2 µl of the first strand reaction and 20 thermal cycles.

Proteinase K digestion, DNA precipitation, Sfi I digestion and cDNA size fractionation on Chromaspin-400 columns were all conducted as per kit manufacturer instructions. Ligation of DNA into the λTriplEx2 phage vector was performed in two separate reactions containing 2 µl (a)/ 3µl (b) cDNA, 2 µl vector, 1 µl 10× ligation buffer, 1µl ATP (10mM), 1 µl T4 DNA ligase, 3 µl (a)/ 2 µl (b) d.H_2O. Gigapack III Gold packaging extract (Stratagene, cat # 200201) was used to package ligated DNA according to Strategenes recommended protocol provided, unused ligation mix was stored at -20°C.

Unamplified library titer and recombination frequency were calculated as follows; XL1-Mrf cells were grown overnight at 31°C in 50 ml of LB, 500 µl 20 % maltose and 500 µl 1 M $MgSO_4$. Cells were harvested and resuspended in 10 mM $MgSO_4$ at an optical density $(O.D.)_{600}$ of 0.5. Individual dilutions of cells (200 µl) were

infected with serial dilutions (1:10, 1:20 and 1:50) of library A and library B. Infections were incubated for 15 min without shaking at 37°C and mixed with 2.5 ml of top agar pre-warmed to 42°C containing 6 μM IPTG and 7 mM X-gal, plated on LB plates and incubated overnight at 37°C to allow plaque growth and blue-white selection.

1,086,000 plaque forming units (pfu) of library B were amplified on ten 25 cm^2 plates, and stored at 4°C and -80°C following SMART™ cDNA Library Construction Kit (Clontech, PT3000-1) protocols.

2.11.2. Digoxigenin labelled cDNA probe generation.

pCR®2.1-TOPO® plasmids containing cysteine and aspartic proteinase sequences to be dUTP-digoxigenin labelled (dUTP-DIG) (Boehringer Mannheim) were restriction digested with EcoR1 and the inserts gel purified. dUTP-DIG labelling reactions were performed using a protocol provided by X. Fossiac (personal communication). Briefly, a 2 μl sample of each purified insert was serially diluted 10^{-1} to 10^{-5} and 25 thermal cycles employed in each PCR reaction with appropriate primers. Thermal cycling was as performed as previously described for standard PCR reactions.

Trial quantitative PCRs were performed to establish the dilution at which no PCR product could be seen following PCR reaction analysis on a 1 % agarose gel. The dilution 10^2 times more concentrated than this concentration was then used in the labelling reactions (ie. if the 10^{-5} plasmid dilution produced no visible PCR result then the 10^{-3} reaction was used in the labelling reaction).

Labelling dUTP-DIG was incorporated at a ratio of 1 μl per 50 μl in each PCR

reaction replacing 1 µl of d.H_2O. Primers forward primer 1 and back primer 1 were employed to amplify inserts F1B1 and F1B1a, forward primer 1 and T7 were employed to amplify insert F173 and forward primer 2 and T7 were utilised to amplify insert F271. Primers RAM and T7 were utilised to amplify fragment R71, RP3 was amplified using primers RAM and poly-TG (See Results Chapter 4 for primer and fragment identities). PCR conditions and cycling were performed identically as the trial quantitative PCR.

Gel analysis of the reactions containing dUTP-dig showed a decrease in amplification product gel mobility relative to control reactions. Decreased mobility confirmed dUTP-dig incorporation into the PCR product as dUTP-dig is a larger nucleotide than dTTP resulting in lower gel mobility of equally sized amplification products.

2.11.3. cDNA library screening.

Previous screening experiments (data not shown) resulted in up to 400 positively hybridised plaques being observed on screens of approximately 50 000 pfu. Therefore, 7500 pfu were plated to provide sufficiently spaced plaques to allow individual plaque isolation from a primary screen. This eliminated the requirement for a secondary screening, while retaining a sufficient number of plaques predicted to be approximately 60 to provide surplus positive hybridised signals.

XL1 Blue MRF- cells were grown as described in cDNA library construction and evaluation. Aliquots of cells (6 ml) were incubated with 1.5 µl of un-amplified

cDNA library A (approx 7500 pfu) for 15 min at 37°C. Cells and phage were then mixed with 45 ml of LB top agar and spread on 25 cm^2 LB plates. Plates were incubated at 37°C. Infection was allowed to occur for 3-6 hours until plaques were approximately 0.5 mm in diameter, plates were then incubated at 4°C overnight to harden.

Plaque lifts were performed with Hybond-N nylon membranes (Amersham), initial plaque lifts were conducted for 30 seconds and a second lift for 2 minutes on each plate. Plates were stored at 4°C until positively hybridising plaque selection. DNA bound to membranes was denatured in denaturation solution for 1 minute followed by a 5 minute wash in renaturation buffer. Membranes were then soaked in 2×SSC for 5 minutes, dried on Whatman 3MM paper and loosely wrapped in aluminium foil.

Crosslinking to bind phage DNA to the nylon membrane was performed with a BioRad GS gene linker UV chamber using program CL (125 mj of UV). A Pre-Pre hybridisation wash to remove bacterial debris was performed in 5×SSC, 0.5 % SDS, 1 mM EDTA pH 8.0 (Sambrook *et al.*, 1989). Pre-Pre hybridisation washing was performed on a Lab shaker (Adolf Kuhner AG Schweiz) set at 100 rpm for 1.5 hours at 37°C in a covered glass dish with 250 ml solution per membrane.

The following washes were all conducted in a Techne HB-1D hybridisation oven employing Techne glass hybridisation cylinders. Prehybridisation was performed with 60 ml of 6×SSC, 0.05 % Blotto, 0.5 % SDS, for 2 hours at 65°C. Probe was denatured by boiling at for 5 minutes and added at a 1 µl:1 ml concentration in 10 ml of fresh hybridisation solution. Hybridisation was conducted overnight at 65°C.

Post-hybridisation washes were all conducted with 125 ml of wash solution. Washing was initiated with a 5 minute wash in 2×SSC, 1 % SDS at room temperature

followed by a repeat for 15 minutes. A further increased stringency wash in 1×SSC, 1 % SDS gradually warmed to 65°C over a 30 minute period was performed. Wash solution was replaced with fresh pre-warmed solution and washing continued for a further 2 hours.

2.11.4. Antibody detection and visualisation of hybridised probe.

Techne HB-1D hybridisation oven and Techne glass hybridisation cylinders were employed for this protocol performed at room temperature. Boehringer Mannheim recommendations were followed for disodium 3-(4-methoxyspiro(1,2-dioxetane-3,2'-(5'-chloro)tricyclo[3,3.1.1$^{3.7}$]decan}-4-yl) phenyl phosphate (CSPD) detection of dUTP-digoxigenin labelled probe bound by anti-digoxigenin conjugated to alkaline phosphatase (AP) with minor modifications due to the 400 cm^2 membrane size employed in this procedure.

Reagent volumes were proportionally increased as required for individual 400 cm^2 nylon membranes. Hybridised filters were washed in 100 ml of wash buffer (0.1 M maleic acid, 0.15 M NaCl, 0.3 % Tween 20, pH 7.5) and blocked with 100 ml of buffer 2 (0.1 M maleic acid, 0.15 M NaCl, 1 % milk powder pH 7.5). Anti-dig-AP was applied at a 1:10 000 antibody dilution in 40 ml of buffer 2. Excess antibody was removed with 2× 200 ml washes in wash buffer and the membranes equlibriated with 40 ml of buffer 3 (0.1 M Tris-HCl, 0.1 M NaCl, pH 9.5) and 2 ml of buffer 3 with CSPD (1:100) were used in the detection procedure.

Membranes were placed in a plastic bag and overlayed with Super RX X-ray

medical film (Fuji) to detect alkaline phosphatase mediated dephosphorylation of CSPD generating the light emitting metastable phenolate ion. Pre-warmed film cassettes (Genetic Research Instrumentation Ltd) were then incubated at 37°C for 15-30 min before removal and developing in a Compact ×4 developer (X-ograph imaging systems). The positions of positive signals were mapped back to the screening plates and the appropriate plaque excised with the wide end of 200 µl pipette. Excised plaques were stored in 1× SM buffer at 4°C.

2.11.5. Excision and circularisation of pTriplEx 2 plasmid.

LoxP site specific Cre recombinase mediated excision and circularisation of the bacterial vector pTriplEx2 (see appendix A.1. for vector and MCS map) from the phagemid vector λTriplEx2 was performed according to Clontech protocols in BM25.8 cells with minor modifications (Clontech, PT3000-1). 50 µl of infected BM25.8 cell suspension were each plated on LB-Carb plates (50 µg/ml). Plates were incubated at 31°C overnight to allow colony formation.

2.12. pET 24a construct preparation, recombinant protein expression and analysis.

2.12.1. pET 24a expression vector construct generation.

Sequences to be expressed were amplified by Advantage 2 PCR mix (Clontech), with

primers adding appropriate restriction sites (Nde 1 and Bam H1) for ligation into the pET 24a vector (see appendix A.1. for vector and MCS map). Inserts were ligated in frame with the His-Tag (6× His) present on the pET 24a vector to allow for one step purification. Transformation of chemically competent BL21 DE3 cells with constructs was performed according to Novagen protocols (pET System Manual (TB055), 2000) with an extended 42°C (2min) heat shock (H.S. Wilkinson personal communication).

2.12.2. Recombinant protein expression analysis.

Small scale expression analysis was performed on each pET construct following Novagen recommended protocols (pET system Manual (TB055), 2000) given in detail below. BL21 DE3 containing pET 24a constructs were grown in 3 ml LB-Kan at 37°C until an $O.D._{600}$ of 0.5, 3 ml cultures were transferred to 100 ml LB-Kan and grown to an O.D. of 0.75. The 100 ml culture was split equally into separate flasks and 50ml induced with 1 mM IPTG, induced and non-induced bacterial growths were cultured for a further 4 hours.

Total cell protein (TCP) samples were prepared by centrifugation of 1 ml of culture at 10 000g, the cells were lysed with a 27-gauge needle and the supernatant was retained for media fraction preparation. Media proteins were precipitation of 1 ml of supernatant by trichloroacetic acid (TCA) spun at 14 000g, washed with acetone and air dired. Periplasmic fractions were prepared by suspension of the culture pellet in 30 mM Tris-HCl, 20 % sucrose, 0.5 M EDTA the suspension was stirred (10 min) and spun at 10 000g for 10 min at 4°C. The pellet was resuspended in chilled 5 mM $MgSO_4$, stirred

on ice, spun at 10 000g for 10 min at 4°C and 1 ml of the supernatant TCA precipitated.

Soluble cellular proteins were extracted by resuspension of the cell pellet in 0.25 M NaCl 20 mM Tris-HCl pH 7.5, lysozyme treatment, sonication (Soniprep 150, MSE) and centrifugation at 14 000g for 10 min. Pelleted insoluble proteins were prepared by a repeated wash in 20 mM Tris-HCl pH 7.5 and centrifugation at 10 000g for 5 min. The pellet was solubilised in 1.5 % SDS. Collected fractions were mixed 1:1 with 100 µl of 2× SB buffer, heated at 70°C for 3 minutes and stored at -20°C prior to SDS-PAGE analysis.

2.12.3. Large Scale recombinant protein expression.

Large scale recombinant protein production was performed as detailed in Novagen protocols (pET System Manual, TB055) described below. PET 24a constructs in BL21 DE3 were grown in 3 ml LB-Kan starter cultures at 37°C until an $O.D._{600}$ of 0.5. Started cultures were transferred to 1 L LB-Kan and cultured at 37°C for insoluble protein production or 25°C for soluble protein production until an $O.D._{600}$ of 1.0. Cultures were induced with 1 mM IPTG for 4-6 hours or overnight and harvested by centrifugation at 6500 g for 15 minutes at 4°C. Spun cells were stored overnight at -80°C as a dry cell pellet.

Culture lysis under denaturing and native conditions were prepared as detailed in Qiagen protocols. Under denaturing conditions (protocol 9, The QIAexpressionist™, 1998) cell pellets were thawed on ice, resuspended in Buffer B (5 ml per gram wet weight), stirred for 60 min and cell debris removed by centrifugation at 10 000g for 30

min. Under native conditions (protocol 8, The QIAexpressionist™, 1998) pellets were thawed on ice and resuspended in lysis buffer (5 ml per gram wet weight), lysozyme was added at (1 mg/ml) and incubated for 30 min on ice. DNase 1 and RNase 1 were added at 5 μg/ml and 10 μg/ml respectively and incubated on ice for 15 min. Lysate was centrifuged to remove cell debris at 10 000g for 30 min at 4°C.

2.12.4. Ni-NTA chromotography.

Loading and washing of protein sample under both native and denaturing conditions was performed with Qiagen recommended buffers and protocols for FPLC purification without modification (protocols 12 and 15, The QIAexpressionist™, 1998) and detailed below. Under denaturing conditions lysate was applied to Ni-NTA column equilibrated with Buffer B at a flow rate of 0.5 ml/min, the column was washed with Buffer B (50 ml) until the A_{280} was stable. Non-specific binding protein was removed by washing (0.5-1 ml/min) in Buffer C (50 ml) until the A_{280} was stable.

Refolding of denatured protein was performed on the column, using a manual gradient by a glass bridge and two 75 ml beakers (Pyrex). Qiagen recommended renaturing buffers (The QIAexpressionist™, 1998) were employed without protease inhibitors. Renaturing was performed over a period of 3-4 hours depending on flowthrough rate achieved, protein was eluted by washing in renaturation buffer containing of 250 mM imidazole (60 ml).

Under native conditions lysate was applied to Ni-NTA column equilibrated with lysis buffer at a flow rate of 1 ml/min, the column was washed with lysis buffer (50 ml)

until the A_{280} was stable followed by a wash in wash buffer (50 ml) again until the A_{280} was stable. Protein was eluted with a 60 ml wash in elution buffer. A C10/10 column and peristaltic pump P-1 (Pharmacia biotech) were employed for chromotography. A Uvicord II (8300) (LKB – Bromma) was utilised to monitor protein flowthrough and elution. Eluted protein sample was collected in a 15 ml falcon tube (Greiner Labortechnik).

2.12.5. Oryzacystatin purification.

rOzc was purified as the published protocol (Edmonds *et al.*, 1996) with minor modifications. The oryzacystatin 1 coding sequence cloned into the pUC 8 expression vector maintained in DH5ά cells (pHEV I) was cultured for 8 hours at 37°C in 10 ml of LB-carb. Cultures were transferred into 1 L volumes of LB-carb and induced overnight with 0.5 mM IPTG at 30°C differing from the published culture temperature of 37°C. Employing a lower culture temperature increased soluble rOzc protein yields from 5-6 mg/L to a yield of 27 mg/L of recombinant protein.

Cultures were harvested by centrifugation at 15900 g for 10 min at 4°C, heat treated at 82°C for 15 min, ammonium sulphate precipitated to 65 % saturation and the precipitate collected by centrifugation. Precipitated protein was dialysed against 20 mM Tris pH 8.0 overnight, applied to a Sephadex G50 column equilibrated with 20 mM Tris pH 8.0 and eluted with a manual salt gradient employing 0.5 M NaCl, 20 mM Tris pH 8.0 as previously described (Edmonds *et al.*, 1996).

2.12.6. Dialysis of protein samples.

Sample fractions were pipetted at approximately 0.1 mg/ml into preboiled for 10 min with 1 spatula of ammonium bicarbonate dialysis tubing (Medicell MWCO 12-140000 daltons) and dialysed overnight at 4°C in 6 L of d.H_2O, followed by 2 further washes for 3 hours with changes of d.H_2O at room temperature. Sample was then transferred into convex flasks and frozen in liquid nitrogen.

2.12.7. Protein freeze drying.

Samples were freeze dried in an Edwards Modulyo freeze drier, protein sample was weighed and stored at 4°C.

2.12.8. SDS-PAGE analysis.

Resolving gels at 15 % were made with 7.5 ml acrylamide/bis-acrylamide (30 %/0.8 %), 12 % resolving gels were made with 6.25 ml acrylamide/bis-acrylamide (30 %/0.8 %) with an additional 1.25 ml of d.H_2O. Resolving gels were overlayed with acetone allowed to polymerise for 45 min prior to stacking gel application. Samples to be electrophoresed were mixed 2:1 with 2× SB buffer, heated for 3 minutes at 70°C and stored at -20°C. Samples and SDS 7 protein size marker (see appendix section A.1.) were loaded with protein loading tips (Greiner Labortechnik).

Electrophoresis was performed in an ATTO AE-6450 SDS-PAGE gel apparatus in 1× reservoir buffer. Samples were run at 50v through the stacking gel and 100v through the resolving gel. Gel staining was performed in 25 ml SDS stain with gentle agitation overnight. Destaining was performed overnight with two 25 ml washes of SDS destain.

Chapter 3. Biochemical characterisation of *D. undecimpunctata* alimentary tract proteolytic activities.

3.0. Introduction.

Characterisation of *D. undecimpunctata* digestive proteolysis was performed to identify the major proteolytic activities present in the alimentary tract. Previous research (Orr *et al.*, 1994; Edmonds *et al.*, 1996) identified cysteine and aspartic proteinases as being the major proteolytic activities present in *D. undecimpunctata* larval midgut extracts by oryzacystatin I, potato multicystatin and pepstatin A inhibition. The research described in this chapter was performed to verify previous published research by independently establishing the major digestive proteolytic activities in *D. undecimpunctata* by proteolysis pH optima and inhibition studies.

3.1. Results: pH optima of proteolytic activities in *D. undecimpunctata* alimentary extracts.

Proteolytic rate assays over a pH range of pH 3.5 to pH 10.5 (see figure 3.1a) revealed proteolysis over the entire pH spectrum the highest proteolytic activity was observed between pH 4.0 and pH 6.5. The major peak of proteolytic activity 0.095 rate of fluorimeter units per minute (rfu/min) occurred at pH 5.5. Two secondary peaks occurred at pH 8 (0.02 rfu/min) and pH 10 (0.029 rfu/min). The majority of proteolytic activity occurring in the acidic range indicating that cysteine and/or aspartic proteinases

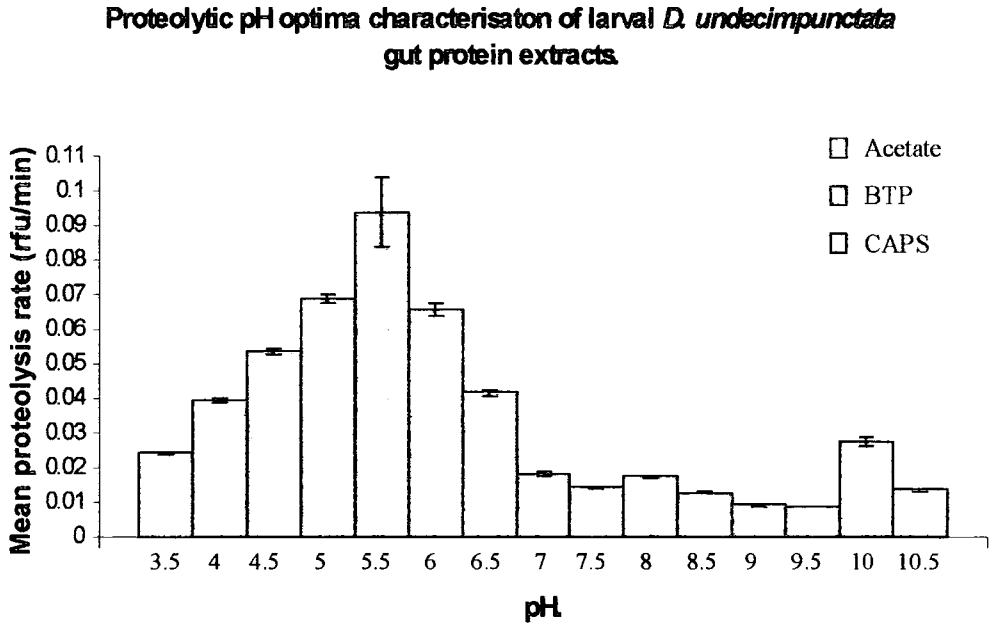


Figure 3.1a. pH dependant proteolysis rates of larval *D. undecimpunctata* alimentary tract extracts. pH values are indicated under each column. Acetate; acetate buffer, BTP; Bis-Tris propane buffer, CAPS assay buffer. Rfu/min; rate of fluorimeter units.

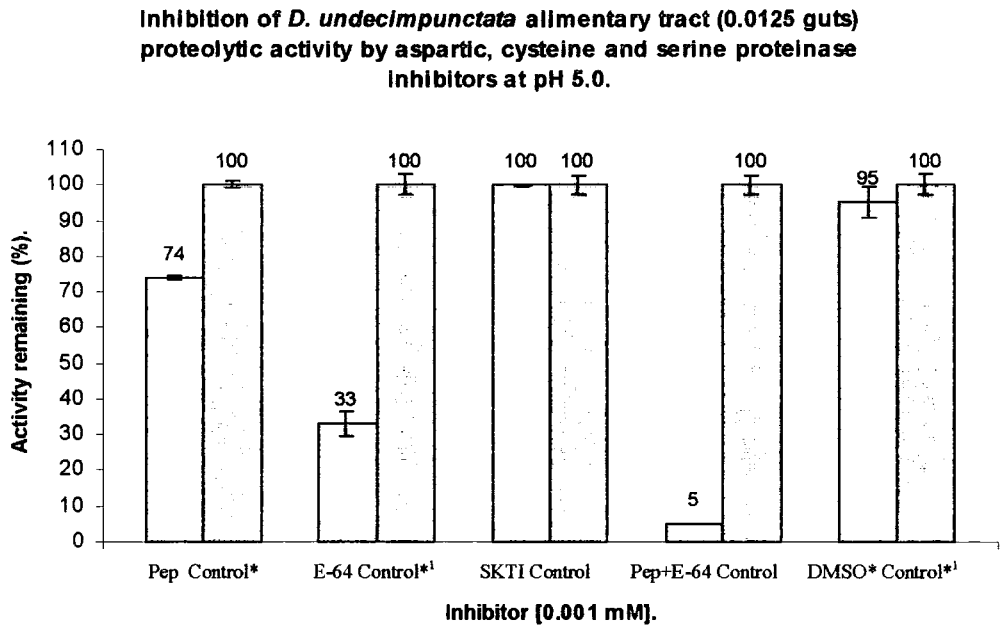


Figure 3.2a. Inhibition of larval *D. undecimpunctata* proteolytic activities (0.0125 guts) by Pepstatin (Pep), E 64 (chemical inhibitors) and soybean Kunitz trypsin inhibitor (SKTI). Assays were performed with 0.0125 gut equivalents at pH 5.Control uninhibited assays; Pep control (10 µl DMSO), E-64 and SKTI control (10 µl water), Pep+E64 (10 µl DMSO and 10 µl water). DMSO*; uninhibited with 10 µl DMSO, Control*[1]; uninhibited with 10 µl water.

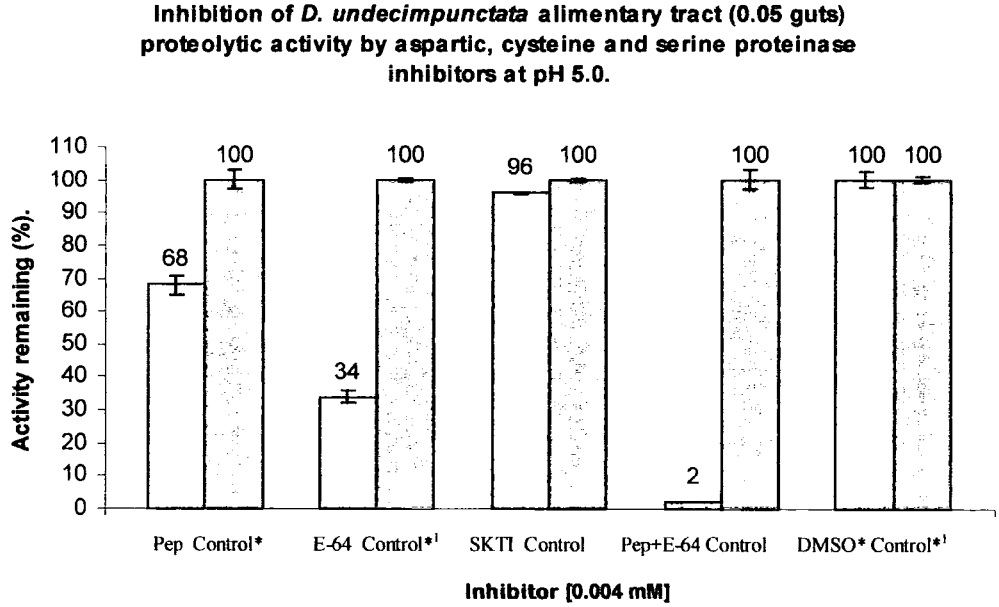


Figure 3.2b. Inhibition of larval *D. undecimpunctata* proteolytic activities (0.05 guts) by Pepstatin (Pep), E 64 (chemical inhibitors) and SKTI. Assays were performed with 0.05 gut equivalents at pH 5.Control uninhibited assays; Pep control (10 µl DMSO), E-64 and SKTI control (10 µl water), Pep+E64 (10 µl DMSO and 10 µl water). DMSO*; uninhibited with 10 µl DMSO, Control*[1]; uninhibited with 10 µl water.

are the major digestive peptidases present as metalloproteinase, aminopeptidase, carboxypeptidase A and B activity would be inhibited by addition of 1 mM EDTA to the assay buffer (see Chapter 2) and serine proteinases show optimal proteolytic activity at alkaline pH.

The secondary peaks at pH 8 or pH 10 implies serine proteinase activity is also present in *D. undecimpunctata* alimentary tracts. Physiological variation of pH in alimentary tracts of insects of all orders is well documented (for review see Terra and Ferreira, 1994). Spatial regulation of physiological pH in the *D. undecimpunctata* digestive tract may occur as observed in other Coleoptera (Terra and Cristofoletti, 1996), alkaline pH allowing serine proteinase mediated proteolytic digestion.

3.2. Results: Inhibition of *D. undecimpunctata* larval alimentary tract proteolysis by chemical and plant protein proteinase inhibitors.

The major peak of activity was further characterised at pH 5.0 by inhibition assays with the diagnostic cysteine proteinase inhibitor E-64 and aspartic proteinase inhibitor pepstatin A and serine proteinase inhibitor soybean Kunitz trypsin inhibitor (SKTI) a trypsin and chymotrypsin inhibitor (see figures 3.2a and 3.2b). Inhibition assays with E-64, pepstatin A and SKTI were performed with two dilutions of gut extract equating to 0.05 gut equivalents and 0.0125 gut equivalents per assay.

Incorporating E-64 a cysteine proteinase class specific inhibitor (Hanada *et al.*, 1978; Barrett *et al.*, 1982), at 10^{-6} M (0.0125 gut equivalents) and 4×10^{-6} M (0.05 gut equivalents) resulted in a reduction of proteolysis to approximately 34 % of controls.

Control rates were 0.184 ± 0.004 rfu/min employing 0.0125 gut equivalents and 0.39 ± 0.01 rfu/min employing 0.05 gut equivalents.

Pepstatin A an aspartic proteinase inhibitor (Barrett, 1977), reduced proteolysis to 74 % and 68 % of controls at 10^{-6} M and 4.0×10^{-6} M when assayed against 0.0125 and 0.05 gut equivalents respectively. The 8 % variation in the level of inhibition observed may indicate incomplete inhibition of aspartic proteinase activity by pepstatin A at 10^{-6} M against 0.0125 gut equivalents. A 5 % inhibitory effect on proteolytic digestion was observed with addition of the DMSO:acetic acid pepstatin A solvent. The additional molarity of acetic acid added (0.0068 M) is unlikely to alter the pH of the assay. This would indicate that DMSO inhibits proteolytic activity, this effect was not observed in assays incorporating 0.05 gut equivalents.

Assays incorporating the plant proteinase inhibitor SKTI at 10^{-6} M (0.0125 gut equivalents) and 4×10^{-6} M (0.05 gut equivalents) showed no observable specific inhibition. SKTI inactivity as an inhibitor of proteolysis at pH 5.0 against 0.0125 gut equivalents would be expected as serine proteinase activity would not be expected at this pH. At 0.05 gut equivalents 4 % inhibition was observed with 4×10^{-6} M SKTI this inhibition is most likely due to the additional extract content of the assay relative to the substrate.

Joint E-64 and pepstatin A inhibition assays performed with 0.0125 and 0.05 gut equivalents reduced proteolysis by 96-98 % compared to uninhibited controls. Inhibition rates by E-64 and pepstatin A indicates protein hydrolysis is performed almost exclusively by cysteine and aspartic class proteinase enzymes at pH 5.0. This

data indicates that at pH 5.0 proteinases of classes other than cysteine or aspartic make a negligible contribution to protein hydrolysis.

3.3. Discussion.

Biochemical characterisation of proteolysis in dissected larval *D. undecimpunctata* alimentary tracts was performed to verify the major classes of proteinase activity previously observed (Edmonds *et al.*, 1996). Proteolysis in cusp 3-4th instar larvae were chosen as this is the most destructive stage of larval development, with respect to plant feeding. Alimentary tracts were dissected to reduce contamination of the digestive proteolytic activity by lysosomal or intracellular proteinases present in the fat body.

Aminopeptidase and carboxypeptidase A and B activity has been inhibited in the assays performed in this work by addition of 1 mM EDTA. Thus allowing the protein hydrolases to be studied rather than the peptidases which hydrolyse the polypeptides generated by proteinase digestion of proteins. Relative cysteine and aspartic proteolytic activity will vary with the *in vivo* pH conditions, depending on the individual optimum pH for activity possessed by the cysteine and aspartic proteinase isoforms present. The *in vivo* pH of *D. undecimpunctata* larval guts has been estimated at between pH 4 and 5 (Edmonds Thesis, 1996), *in vitro* data showed optimal substrate hydrolysis at pH 5.5 a compromise pH of 5.0 was employed to reflect both the *in vitro* and *in vivo* data.

Previous research has demonstrated similar findings to those presented here (Purcell *et al.*, 1992, Orr *et al.*, 1994; Edmonds *et al.*, 1996) approximately 65 % inhibition of *D. undecimpunctata* larval extract proteolysis was observed with E-64, and

between 20 and 34 % inhibition by pepstatin A in agreement with the findings presented herein. Alfalfa weevil midgut preparations were inhibited by 65 % with E-64, and 50 % by pepstatin A (Wilhite *et al.*, 2000) providing further evidence for possibly incomplete aspartic proteinase inhibition by pepstatin A in this work.

The peak of proteolytic activity observed at pH 10 maybe due to serine proteinase activity as previously discussed, a possible alternative is that this hydrolytic activity represents cysteine proteinase activity. The cysteine amino acid side chain has a p*K* of 8.33, at pH 10 the thiol group of the cysteine may become partially ionised. At acidic pH the catalytic site histidine acts as a proton acceptor stabilising the thiolate ion at pH 10 the catalytic site histidine would be neutral. If the catalytic site cysteine becomes partially or fully ionised in some or all of the *D. undecimpunctata* cysteine proteinases present at pH 10.0 then nucleophilic activity would be expected to be restored and cysteine proteinase mediated peptide bond cleavage occur.

Chapter 4. Molecular characterisation of larval *D. undecimpunctata* cysteine and aspartic proteinases.

4.0. Introduction.

The biochemical characterisation described in Chapter 3 confirmed the majority of larval *D. undecimpunctata* alimentary tract proteolytic activity occurs in the acidic range and is inhibited by cysteine and aspartic proteinase class specific inhibitors. Molecular isolation of alimentary tract aspartic and cysteine proteinase sequences was undertaken to characterise the digestive proteinases of *D. undecimpunctata* at the protein sequence level.

The aims of this Chapter were threefold; 1) clone PCR generated fragments of cysteine and aspartic proteinase sequences from *D. undecimpunctata*, 2) construct a 3rd-4th instar *D. undecimpunctata* larval alimentary tract cDNA library and 3) screen the alimentary tract cDNA library using the aspartic and cysteine proteinase fragments and obtain full-length clones encoding cysteine and aspartic proteinases transcribed by *D. undecimpunctata* alimentary tract cells. Isolation of cDNA clones allowed the proteinases sequences to be characterised at the molecular level.

4.1. Results: Isolation and identification of aspartic and cysteine proteinase encoding sequence fragments from larval *D. undecimpunctata*.

An alignment of aspartic proteinases from insects and other species revealed conserved amino acid motifs within the sequences (see appendix A.6). Degenerate PCR primers

Ram:
Amino-acid sequence. KFDGILG (20-mer)
Nucleotide sequence. 5′-AARTTYGAYGGXATHYTXGG-3′

Jam:
Amino-acid sequence. PLWILGD (20-mer)
Nucleotide sequence. * 5′-TCXCCXARDATCCAXARXGG-3′

Poly-TA/C/G (standard primer):
Nucleotide sequence. 5′-TTTTTTTTTTTTTTTTTTTTTTTTTA/C/G-3′ (26-mer)

T7 (standard primer):
Nucleotide sequence. 5'-GTAATACGACTCACTATAGGGCG-3' (23-mer)

Figure 4.1a. Aspartic proteinase specific Ram and Jam, poly-Tx and T7 PCR primers. Oligonucleotide primer lengths, amino acid sequence and degenerate nucleotide sequence are shown. * indicates the nucleotide sequence has been reversed and complimented.

```
        [1                                                    50]
mLAP    MLIKSIIALV CLAVLSQADF VRVQLHKTES ARQHFRNVDT EIKQLRLKYN

        [51                                                  100]
mLAP    AVSGPVPEPL SNYLDAQYYG AITIGTPPQS FKVVFDTGSS NLWVPSKECS

        [101                                                 150]
mLAP    FTNIACLMHN KYNAKKSSTF EKNGTAFHIQ YGSGSLSGYL STDTVGLGGV

        [151                  (Ram)                          200]
mLAP    SVTKQTFAEA INEPGLVFVA AKFDGILGLG YSSISVDGVV PVFYNMFNQG
                              ------->

        [201                                                 250]
mLAP    LIDAPVFSFY LNRDPSAAEG GEIIFGGSDS NKYTGDFTYL SVDRKAYWQF

        [251                                                 300]
mLAP    KMDSVKVGDT EFCNNGCEAI ADTGTSLIAG PVSEVTAINK AIGGTPIMNG

        [301                                                 350]
mLAP    EYMVDCSLIP KLPKISFVLG GKSFDLEGAD YVLRVAQMGK TICLSGFMGI

        [351     (Jam)                        377]
mLAP    DIPPPNGPLW ILGDVFIGKY YTEFDMGNDR VGFATAV // poly-Tx/T7
                <-------                      <-------
```

Primer	Jam	poly-Tx
Ram	528 bp	618 + 3'UTR bp

Figure 4.1b. Predicted amino acid sequence of Aed-asp a lysosomal aspartic proteinase identified in *Aedes aegypti* (Cho and Raikhel, 1992). Amino acids in bold show the positions of the specific Ram and Jam degenerate primers in the sequence, the poly-Tx (poly A tail) and T7 (present in the λTriplEx 2 vector) are also indicated. Primer directed strand synthesis is indicated by arrowheads. Also shown is the expected size of amplification products from each primer pair (forward primers to T7 not shown as they would be as poly-Tx primers with addition of polyA tail and λTriplEx 2 vector sequence).

Forward primer 1 (F1):
Amino-acid sequence. VDWRTK (19-mer)
Nucleotide sequence. 5′-GTXGAYTGGMGXACXAARG-3′

Forward primer 2 (F2):
Amino-acid sequence. CGSCWAF (20-mer)
Nucleotide sequence. 5′-TGYGGNXXXTGYTGGGCNTT-3′

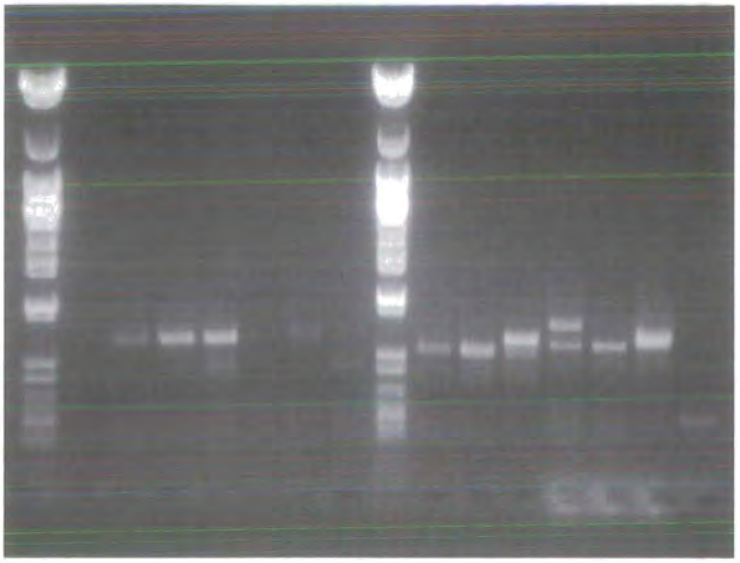
Backward primer 1(B1):
Amino-acid sequence. WLVKNSW (21-mer)
Nucleotide sequence.* 5′-CCAXXXRTTYTTNACNARCCA-3′

Figure 4.1c. Cysteine proteinase specific (F1, F2 and B1), poly-Tx and T7 PCR primers. Oligonucleotide primer lengths, amino acid sequence and degenerate nucleotide sequence are shown. * indicates the nucleotide sequence has been reversed and complimented.

```
            [1                                                  (F1)  50]
Dros_cys    VGIDRDYNVH IFNFAFAFSA ADESFKGVTF ISPAHVTLPK SVDWRTKGAV
                                                          ------>

            [51       (F2)                                       100]
Dros_cys    TAVKDQGHCG SCWAFSSTGA LEGQHFRKSG VLVSLSEQNL VDCSTKYGNN
                    ------>

            [101                                                   150]
Dros_cys    GCNGGLMDNA FPYIKDNGGI DTEKSYPYEA IDDSCHFNRA QVGATDRGFT

            [151                                                   200]
Dros_cys    DIPQGDEKKM PEAVATVGPV SVAIDASHES FQFYSEGVYN EPQCDAQNLD

            [201                  (B1)                             250]
Dros_cys    HGVLVVGFGT DESGEDYWLV KNSWGTTWGD KGFIKMLRNK ENQCGIASPS
                                <------

                          255]
Dros_cys    SYPLV //  poly-Tx/T7
                      <------
```

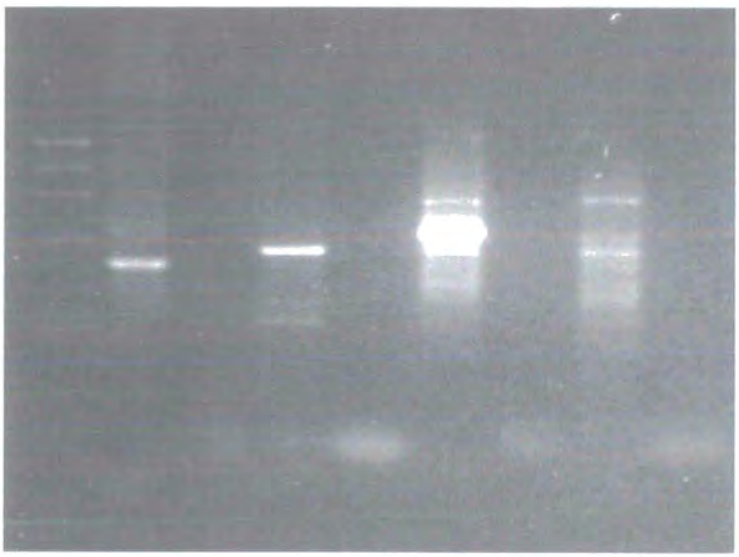
Primer	B 1	poly-Tx
F 1	546 bp	642 + 3' UTR bp
F 2	498 bp	594 + 3' UTR bp

Figure 4.1d. Amino acid sequence of Dros_cys a putative digestive cysteine proteinase fragment identified in *Drosophila melanogaster* (Matsumoto *et al.*, 1995). The positions of the specific degenerate primers F1, F2 and B1 are shown in bold the standard T7 (λTriplEx 2 priming site) and polyTx (polyadenylation priming site) primers are also shown. The direction of primed strand synthesis is indicated by arrowheads. The table below shows the expected size of amplification products (bp) from each primer pair (forward primers to T7 not shown as they would be as poly-Tx primers with addition of polyA tail and λTriplEx 2 vector sequence).



M 1 2 3 4 C1 C2 C3 M C4 C5 C6 5 C7 C8 6

Figure 4.1e. 1 % agarose gel analysis of PCR reactions for aspartic and cysteine proteinase fragments utilising the larval *D. undecimpunctata* alimentary tract cDNA library as template. C; denotes PCR products amplified using cysteine proteinase specific PCR primers. Lane guide (left to right): M) λDNA/Eco471 (AvaII), 1) no primer control , 2) Ram-polyTA, 3) Ram-polyTC, 4) Ram-polyTG, C1) F1-polyTA, C2) F1-polyTC, C3) F1-polyTG, M) λDNA/Eco471 (AvaII) size marker, C4) F2-polyTA, C5) F2-polyTC, C6) F1-polyTG, 5) RAM-T7, C7) F1-T7, C8) F2-T7, 6) Ram-Jam.



M 1 2 3 4 5 6 7 8

Figure 4.1f. 1 % agarose gel analysis of RT-PCR products from *D. undecimpunctata* alimentary tract mRNA. Lane guide (left to right): M) ΦX174 DNA/*Hae* III size marker, 1) F2-B1, 2) F2-B1 control, 3) F1-B1, 4) F1-B1 control, 5) F2-polyTG, 6) F2-polyTG, 7) F1-polyTG, 8) F1-polyTG control. All controls were template minus.

```
RP3  : 2   NMINQGVLDAPVFSFYLSQTANG-DKGELLLGGSDSKYYKGDFTYTKVSTQLYWQTNLQG  60
           NM NQG++DAPVFSFYL++  +  + GE++ GGSDS  Y GDFTY  V  + YWQ  +
mLAP : 195 NMFNQGLIDAPVFSFYLNRDPSAAEGGEIIFGGSDSNKYTGDFTYLSVDRKAYWQFKMDS 238

RP3  : 61  VSVGSRSVCKSGCEAVIDTGTSLIYGPTDDVDVVNSAIGATYDYSVGLYTVNCNTDLNKL 118
           V VG    C +GCEA+ DTGTSLI GP  +V  +N AIG T     G Y V+C+  + KL
mLAP : 255 VKVGDTEFCNNGCEAIADTGTSLIAGPVSEVTAINKAIGGT-PIMNGEYMVDCSL-IPKL 298

RP3  : 121 PNVSFTFGGKKFDIPASAYIIK----DSGYCISSFVAQEF-FLSGFEWLVGDSFLKTVYS 178
           P +SF  GGK FD+  + Y+++         C+S F+  +    +G  W++GD F+   Y+
mLAP : 313 PKISFVLGGKSFDLEGADYVLRVAQMGKTICLSGFMGIDIPPPNGPLWILGDVFIGKYYT 358

RP3  : 176 EFDFGNNRIGFA 187
           EFD GN+R+GFA
mLAP : 373 EFDMGNDRVGFA 384
```

Identities = 87/192 (45 %), Positives = 120/192 (62 %), Gaps = 8/192 (4 %)
Score = 171 bits (432), Expect = 7e-42

Figure 4.1g. Alignment search result for the amino acid sequence predicted by fragment RP3. The amino acid sequence shows closest homology to a lysosomal aspartic proteinase of *Aedes aegypti* (yellow fever mosquito), (Cho and Raikhel, 1992). Alignment performed by blastp (www.ncbi.nlm.nih.gov/BLAST/), + indicates conservative substitution.

```
R71  : 1   KFDGILGMGYPEISVNGITPVFNTLVEQGAVKEPVFSFYLNRDPDGDVGGELLLGGSDSN  60
           KFDGILGM YP ISV+G+ PVF+ ++ Q  V++ VFSFYLNR+PD   GGELLLGG+D
Cat.D: 179 KFDGILGMAYPRISVDGVPPVFDMMMSQKKVEKNVFSFYLNRNPDTQPGGELLLGGTDPK 238

R71  : 61  YYKGDFTYIDVSAKGYWQIVMDSLNVGSSLKLCSDGCQVIVDTG--TSLIAGPSAEVEKL 118
           YY GDF Y+D+S + YWQI MD +++GS L LC  GC+ IVDTG  TSLI GP+AEV+ L
Cat.D: 239 YYTGDFNYVDISRQAYWQIHMDGMSIGSGLSLCKGGCEAIVDTGTSTSLITGPAAEVKAL 298

R71  : 119 HQEIGAFSFLNGEYIIDCNKVDQLPEISFVFGGKEFKLSGNDYILKQSNGLIDICISGFM 178
            + IGA   + GEY++DC KV  LP ISF  GGK + L+G  YILK+S G  DIC+SGFM
Cat.D: 299 QKAIGAIPLMQGEYMVDCKKVPTLPTISFSLGGKVYSLTGEQYILKESQGGHDICLSGFM 358

R71  : 179 GLDLDTRTHVEWILGDVFIGKFYTEFDFGNNRVGLAGA 216
           GLD+       WILGDVFIG++YT FD  NNRVG A A
Cat.D: 359 GLDIPPPAGPLWILGDVFIGQYYTVFDRENNRVGFAKA 396
```

Identities = 130/218 (59 %), Positives = 159/218 (72 %), Gaps = 2/218 (0 %)
Score = 270 bits (689), Expect = 1e-71

Figure 4.1h. Alignment search result for the amino acid sequence predicted by fragment R71. The amino acid sequence shows closest homology to a cathepsin D of *Danio rerio* (Zebrafish) (Riggio *et al.*, 2000). Alignment performed by blastp (www.ncbi.nlm.nih.gov/BLAST/), + indicates conservative substitution.

```
F271    : 1   CGGCWAFSTTGALEGQLAIHKNQAIPLSEQELMDCDTGNSA-CFGGNPDVAFEYIESNGI  59
              CG CWAFS TG+ EG  A    + + LSEQ+L+DC T  SA C GG+ D  F+Y+  +G+
Hyp-cys: 133 CGSCWAFSITGSTEGAYARKSGKLVSLSEQQLIDCCTDTSAGCDGGSLDDNFKYVMKDGL 192

F271    : 60  SSESQYEYTQQKGECRKVENKPVSSISGWLGVPS-DEDALMEAVAQYGPVSVSVFANNDW 118
               SE  Y Y  + G C+      V+ +S +  +P+ DEDAL+EAVA  GPVSV + A+
Hyp-cys: 193 QSEESYTYKGEDGACKYNVASVVTKVSKYTSIPAEDEDALLEAVATVGPVSVGMDASY-L 251

F271    : 119 SLYGGGIFEHASCR-GHPNHAVLAVGYTQKS----WIVKNSWGAAWGEDGYIQLSLVNNQ 173
              S Y  GI+E   C     NHA+LAVGY  ++    WI+KNSWGA+WGE GY +L+   NQ
Hyp-cys: 252 SSYDSGIYEDQDCSPAGLNHAILAVGYGTENGKDYWIIKNSWGASWGEQGYFRLARGKNQ 311

F271    : 174 CNITFASQIPLL 185
              C I+  +  P +
Hyp-cys: 312 CGISEDTVYPTI 323
```

Identities = 86/192 (44 %), Positives = 117/192 (60 %), Gaps = 8/192 (4 %)
Score = 158 bits (399), Expect = 5e-38

Figure 4.1i. Alignment search result for the amino acid sequence predicted by fragment F271. The amino acid sequence shows closest homology to a cathepsin L-like cysteine proteinase isolated from the midgut of *Hypera postica* (alfalfa weevil), (Wilhite *et al.*, unpublished). Alignment performed by blastp (www.ncbi.nlm.nih.gov/BLAST/), + indicates conservative substitution.

```
F173    : 1   SETQYQYTGRDGSCKNVQNKQLSSISGYVEL-SETESALVSALASVGPVSIAVDADPWQF  59
              SE  Y Y G DG+CK       ++ +S Y  + +E E AL+ A+A+VGPVS+ +DA
Hyp-cys: 194 SEESYTYKGEDGACKYNVASVVTKVSKYTSIPAEDEDALLEAVATVGPVSVGMDASYLSS 253

F173    : 60  YSGGVFNNRNCGTA-LNHGVLAVGYT----KDVFIVKNSWGTSWGEQGYIRISRGHNLCG 114
              Y  G++ +++C  A LNH +LAVGY     KD +I+KNSWG SWGEQGY R++RG N CG
Hyp-cys: 254 YDSGIYEDQDCSPAGLNHAILAVGYGTENGKDYWIIKNSWGASWGEQGYFRLARGKNQCG 313

F173    : 115 LNLMNSYPKL 124
              ++    YP +
Hyp-cys: 314 ISEDTVYPTI 323
```

Identities = 59/130 (45 %), Positives = 83/130 (63 %), Gaps = 6/130 (4 %)
Score = 119 bits (299), Expect = 1e-26

Figure 4.1j. Alignment search result for the amino acid sequence predicted by fragment F173. The amino acid sequence shows closest homology to a cathepsin L-like cysteine proteinase isolated from the midgut of *Hypera postica* (alfalfa weevil), (Wilhite *et al.*, unpublished). Sequence shown does not represent the full obtained clone. Alignment performed by blastp (www.ncbi.nlm.nih.gov/BLAST/), + indicates conservative substitution.

```
F1B1    : 1   VDWRTKGAVTAVKDQGHCGSCWSFSATGSLEGQHFRKTGKLVSLSEQNLVDCSTKFGNNG  60
              VDWR KGAVT VKDQGHCGSCWSFSATGSLEGQHFRKTGKLVSLSEQNLVDCS ++GNNG
Sit-cys: 125 VDWRDKGAVTEVKDQGHCGSCWSFSATGSLEGQHFRKTGKLVSLSEQNLVDCSGRYGNNG 184

F1B1    : 61  CNGGLMDNAFRYIKANGGIDTEASYKYKAEDEKCHYNPKKIGATDKGFVDIESGNEDALQ 120
              CNGGLMDNAFRYIK NGGIDTE SY Y AEDEKCHY  +  GATDKGFVDIE  NED L+
Sit-cys: 185 CNGGLMDNAFRYIKDNGGIDTEKSYPYLAEDEKCHYKAQNSGATDKGFVDIEEANEDDLK 244

F1B1    : 121 AAVATIRPVSVAIDASHETFQLYNQGVYYEPECSSDELDHGVLVVGYGTEND-QDYWLVK 179
              AAVAT+ PVS+AIDASHETFQLY+ GVY +PECSS ELDHGVLVVGYGT +D QDYWLVK
Sit-cys: 245 AAVATVGPVSIAIDASHETFQLYSDGVYSDPECSSQELDHGVLVVGYGTSDDGQDYWLVK 304

F1B1    : 180 NPW 182
              N W
Sit-cys: 305 NSW 307
```

Identities = 153/183 (83 %), Positives = 162/183 (87 %), Gaps = 1/183 (0 %)
Score = 323 bits (828), Expect = 8e-88

Figure 4.1k. Alignment search result for the amino acid sequence predicted by fragment F1B1. The amino acid sequence shows closest homology to a cathepsin L-like cysteine proteinase isolated from *Sitophilus zeamais* (maize weevil), (Matsumoto *et al.*, 1997). Alignment performed by blastp (www.ncbi.nlm.nih.gov/BLAST/), + indicates conservative substitution.

```
F1B1a   : 5   EGVDDKCRFDVSKVAAKISNFTYIKKNDEEDLKNAVFAKGPISVAIDAS-SKFQLYVAGI  63
              E VD +CRF    V A  + +  IK   E DLK AV   GPISVAIDAS S FQLY  G+
Boo-cys: 207 EAVDGECRFKKEDVGATDTGYVEIKAGSEVDLKKAVATVGPISVAIDASHSSFQLYSEGV 266

F1B1a   : 64  LDDTECSNEFDSLNHGVLVVGYGTENGKDYWLVKNPW 100
               D+ ECS+E   L+HGVLVVGYG + GK YWLVKN W
Boo-cys: 267 YDEPECSSE--DLDHGVLVVGYGVKGGKKYWLVKNSW 301
```

Identities = 56/97 (57 %), Positives = 64/97 (65 %), Gaps = 3/97 (3 %)
Score = 105 bits (263), Expect = 1e-22

Figure 4.1l. Alignment search result for the amino acid sequence predicted by fragment F1B1a. The amino acid sequence shows closest homology to a cathepsin L-like cysteine proteinase isolated from *Boophilus microplus* (southern cattle tick), (Renard *et al.*, 2000). Alignment performed by blastp (www.ncbi.nlm.nih.gov/BLAST/), + indicates conservative substitution.

```
R71    : 85  NVGSSLKLCSDGCQVIVDTGTSLIAGPSAEVEKLHQEIGA-FSFLNGEYIIDCN-KVDQL 142
             +VGS   +C  GC+ ++DTGTSLI GP+ +V+ ++  IGA + +  G Y ++CN  +++L
RP3    : 62  SVGSR-SVCKSGCEAVIDTGTSLIYGPTDDVDVVNSAIGATYDYSVGLYTVNCNTDLNKL 120

R71    : 143 PEISFVFGGKEFKLSGNDYILKQSNGLIDICISGFMGLDLDTRTHVEWILGDVFIGKFYT 202
             P +SF FGGK+F +  + YI+K S      CIS F+  +    +  EW++GD F+   Y+
RP3    : 121 PNVSFTFGGKKFDIPASAYIIKDSG----YCISSFVAQEF-FLSGFEWLVGDSFLKTVYS 175

R71    : 203 EFDFGNNRVGLAGAV 217
             EFDFGNNR+G A
RP3    : 176 EFDFGNNRIGFAELA 190
```

Identities = 91/196 (46 %), Positives = 130/196 (65 %), Gaps = 9/196 (4 %)
Score = 177 bits (448), Expect = 1e-43

Figure 4.1m. Alignment and comparison of PCR amplified aspartic proteinase fragments R71 and RP3 obtained by direct PCR on the larval *D. undecimpunctata* alimentary tract cDNA library. Alignment performed by the BLAST 2 server (Tatusova and Madden, 1999), + indicates conservative substitution.

```
2b   VDWRTKGAVTAVKDQGHCGSCWSFSATGSLEGQHFRKTGKLVSLSEQNLVDCSTKFGNNG  60
F271 -----------------CGGCWAFSTTGALEGQLAIHKNQAIPLSEQELMDCDT--GNSA  41
F173 ------------------------------------------------------------  0
3c   ------------------------------------------------------------  0

2b   CNGGLMDNAFRYIKANGGIDTEASYKYKAEDEKCHYNPKKIGATDKGFVDIESGNEDALQ 120
F271 CFGGNPDVAFEYIESNG-ISSESQYEYTQQKGECRKVENKPVSSISGWLGVPS-DEDALM  99
F173 --------------------SETQYQYTGRDGSCKNVQNKQLSSISGYVELSE-TESALV  39
3c   ----------------------GLEDEGVDDKCRFDVSKVAAKISNFTYIKKNDEEDLK  37

2b   AAVATIRPVSVAIDASHETFQLYNQGVYYEPECSS--DELDHGVLVVGYGTENDQDYWLV 178
F271 EAVAQYGPVSVSVFANND-WSLYGGGIFEHASCRG---HPNHAVLAVGY----TQKSWIV 151
F173 SALASVGPVSIAVDAD-P-WQFYSGGVFNNRNCGT---ALNHGVLAVGY----TKDVFIV  90
3c   NAVFAKGPISVAIDASSK-FQLYVAGILDDTECSNEFDSLNHGVLVVGYGTENGKDYWLV  96

2b   KNPW                                                          182
F271 KNSWGAAWGEDGYIQLSLVNNQCNITFASQIPLL.TTKYLFK.NIFMP              199
F173 KNSWGTSWGEQGYIRISRGHNLCGLNLMNSYPKL.MMKYK.NSIIKIHGKKKKKKKK    147
3c   KNPW                                                          100
```

F271 v F173	Identities = 62/129 (48 %)	Positives = 87/129 (67 %)	Gaps = 1/129 (0 %)
F271 v 2b	Identities = 68/165 (41 %)	Positives = 97/165 (58 %)	Gaps = 10/165 (6 %)
F271 v 3c	Identities = 32/91 (35 %)	Positives = 49/91 (53 %)	Gaps = 8/91 (8 %)
F173 v 2b	Identities = 38/102 (37 %)	Positives = 60/102 (58 %)	Gaps = 8/102 (7 %)
F171 v 3c	Identities = 34/95 (35 %)	Positives = 53/95 (55 %)	Gaps = 9/95 (9 %)
2b v 3c	Identities = 50/94 (53 %)	Positives = 65/94 (68 %)	Gaps = 3/94 (3 %)

Figure 4.1n. Alignment and comparison of PCR amplified cysteine proteinase fragments 3c, 2b, F173 and F271. Cysteine proteinase fragments obtained by RT-PCR and direct PCR on the larval cDNA library. Amino acids in bold and underlined indicate those showing homology across all four sequences. Identities calculated by BLAST 2 server (Tatusova, and Madden, 1999). Alignment performed using the multiple alignment program (MAP) (Huang, 1994).

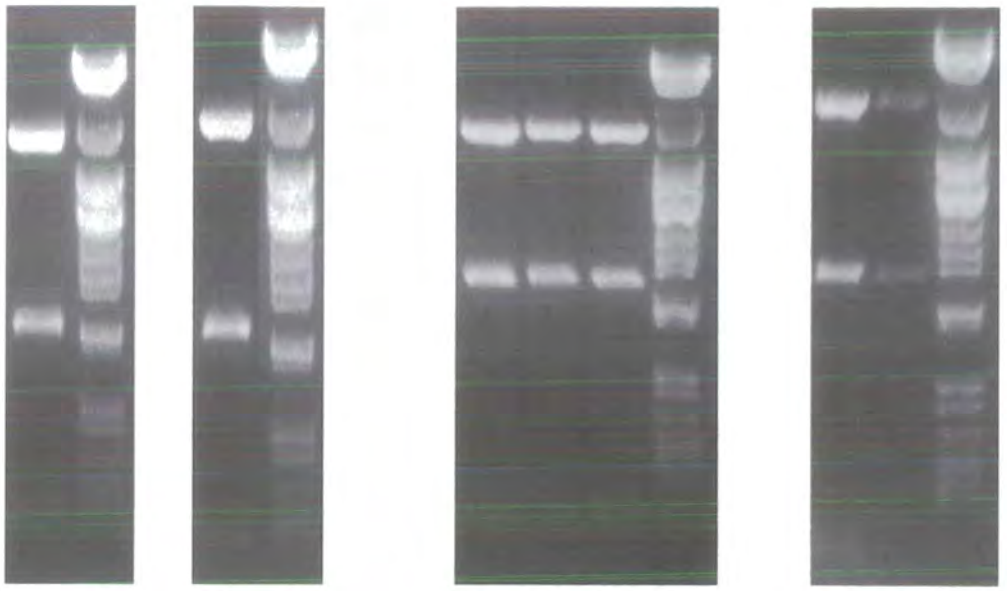


Figure 4.2a. EcoR1 and Sal 1 digested pTriplEx2 plasmids containing *D. undecimpunctata* aspartic proteinase sequences. Digests correspond to cDNA clones (left to right); A20 and A16 (Du-asp 2) showing an insert length of approximately 1280 bp, A72, A62, A42, C51 and C52 (Du-asp 1); showing an insert length of approximately 1250 bp. M: λDNA/Eco 471 (AvaII) DNA marker.

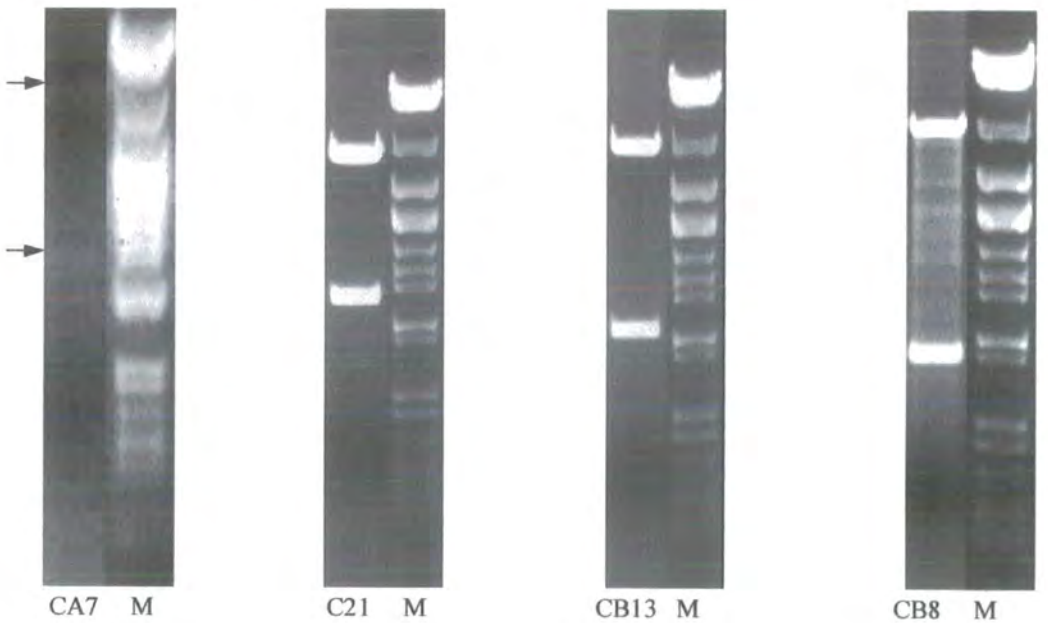


Figure 4.2b. EcoR1 and Sal 1 digested pTriplEx2 plasmids containing *D. undecimpunctata* cysteine proteinase sequences. Digests correspond to cDNA clones (left to right); CA7 (Du-cys 1) 1250 bp insert, C21 (Du-cys 2) 1100 bp insert, CB13 (Du-cys 3) 1000 bp insert and CB8 (Du-cys 4) 950 bp insert. Arrows indicate linearised plasmid and digested insert for plasmid CA7. M: λDNA/Eco 471 (AvaII) DNA marker.

```
1                                                                    66
DFI MMYAKFLVVLVVAALASS EIIRIPLKKVPDNGQRRLRNVARSGLRNKFGGQ GAVPLVNEYDLD
5'] [     Leader     ] [          Pro-peptide          ] [
68                                                                  135
YYGEIGVGTPAQKFNVIFDTGSTDLWVPSSKCQATSAQNGCKNHKQYDASKSSTYQKDGRSFHIEYGSG

136                                                                 204
ALSGFLSKDSVEVAGLAITDQTFAEATVETAPAFVSSTFDGILGLAYPDLSAQNIPPVFQNMIKQGVLD
                          Mature protein
205                                                                 273
APVFSFYLSQTANGDKGELLLGGSDSKYYKGDFTYTKVSTQLYWQTNLQGVSVGSRSVCKSGCEAVIDT

274                                                                 342
GTSLIYGPTDDVDVVNSAIGATYDYSVGLYTVNCNTDLNKLPNVSFTFGGKKFDIPASAYIIKDSGYCI

343                                                                 410
SSFVAQEFFLSGFEWLVGDSFLKTVYSEFDFGNNRIGFAELA.IIITLSIFYY.ILL.QKKKKKKKKK
                                          ][    3'UTR     ][poly-a ]
```

Figure 4.2.1a. Annotated predicted amino acid consensus sequence of clones A72, A62, A42 and C52. Putative catalytic aspartic acids are in bold and underlined. Signal peptide identified by SignalP server (Nielsen *et al.*, 1997), pro-peptide and mature chain junction positioned by similarity to mLAP (Cho and Raikhel, 1992).

```
Du-asp 1 : MMYAKFLVVLVVAALASSEIIRIPLKKVPDNGQRRLRNVARS--GLRNKFG---GQGAVP 58
           M+    + ++ +A L+ ++ +R+ L K  ++ ++  RNV      LR K+    G    P
Aed-asp  : MLIKSIIALVCLAVLSQADFVRVQLHKT-ESARQHFRNVDTEIKQLRLKYNAVSGPVPEP 59

Du-asp 1 : LVNEYDLDYYGEIGVGTPAQKFNVIFDTGSTDLWVPSSKCQATSAQNGCKNHKQYDASKS 118
           L N  D  YYG I +GTP Q F V+FDTGS++LWVPS +C  T+    C  H +Y+A KS
Aed-asp  : LSNYLDAQYYGAITIGTPPQSFKVVFDTGSSNLWVPSKECSFTNI--ACLMHNKYNAKKS 117

Du-asp 1 : STYQKDGRSFHIEYGSGALSGFLSKDSVEVAGLAITDQTFAEATVETAPAFVSSTFDGIL 178
           ST++K+G +FHI+YGSG+LSG+LS D+V + G+++T QTFAEA  E    FV++ FDGIL
Aed-asp  : STFEKNGTAFHIQYGSGSLSGYLSTDTVGLGGVSVTKQTFAEAINEPGLVFVAAKFDGIL 177

Du-asp 1 : GLAYPDLSAQNIPPVFQNMIKQGVLDAPVFSFYLSQTAN-GDKGELLLGGSDSKYYKGDF 237
           GL Y  +S   + PVF NM  QG++DAPVFSFYL++  +  + GE++ GGSDS  Y GDF
Aed-asp  : GLGYSSISVDGVVPVFYNMFNQGLIDAPVFSFYLNRDPSAAEGGEIIFGGSDSNKYTGDF 237

Du-asp 1 : TYTKVSTQLYWQTNLQGVSVGSRSVCKSGCEAVIDTGTSLIYGPTDDVDVVNSAIGATYD 297
           TY  V  + YWQ  +  V VG    C +GCEA+ DTGTSLI GP  +V  +N AIG T
Aed-asp  : TYLSVDRKAYWQFKMDSVKVGDTEFCNNGCEAIADTGTSLIAGPVSEVTAINKAIGGT-P 296

Du-asp 1 : YSVGLYTVNCNTDLNKLPNVSFTFGGKKFDIPASAYIIK----DSGYCISSFVAQEF-FL 352
              G Y V+C+  + KLP +SF  GGK FD+  + Y+++        C+S F+  +
Aed-asp  : IMNGEYMVDCSL-IPKLPKISFVLGGKSFDLEGADYVLRVAQMGKTICLSGFMGIDIPPP 355

Du-asp 1 : SGFEWLVGDSFLKTVYSEFDFGNNRIGFAELA 384
           +G  W++GD F+   Y+EFD GN+R+GFA
Aed-asp  : NGPLWILGDVFIGKYYTEFDMGNDRVGFATAV 387
```

Identities = 172/389 (44 %), Positives = 244/389 (62 %), Gaps = 16/389 (4 %)
Score = 330 bits (845), Expect = 2e-89

Figure 4.2.1b. Alignment search result for the consensus amino acid sequence predicted by cDNA clones A72, A62, A42 and C52 (Du-asp 1). The amino acid sequence shows closest homology to a lysosomal aspartic proteinase of *Aedes aegypti* (Cho and Raikhel, 1992). Alignment performed by blastp (www.ncbi.nlm.nih.gov/BLAST/), + indicates conservative substitution.

```
1                                                                    67
IVPCIEEQGSVK MLSKLLMLVFAAAVATA EIARIPLKKFEKPRDSLNRNGLKETLLAKYS GNFNDD
[  5'UTR   ] [     Leader    ] [        Pro-peptide          ] [
68                                                                  136
SAVVLTNYLDQYYGEIGIGTPAQSFNVIFDTGSSNLWVPSSKCGFLEVACLLHNKYDSDKSSTYVKNDT

137                                                                 205
RFAIRYGSGDVAGVVSQDVVEVGGLQAKDQLFAEATQEPGLAFLVGKFDGILGMGYPEISVNGITPVFN
                          Mature Chain
206                                                                 274
TLVEQGAVKEPVFSFYLNRDPDGDVGGELLLGGSDSNYYKGDFTYIDVSAKGYWQIVMDSLNVGSSLKL

275                                                                 343
CSNGCQVIVDTGTGSLIAGPSAEVEKLHQEIGAFSFLNGEYIIDCNKVDQLPEISFVFGGKEFKLSGNDY

344                                                                 411
ILKQSNGLIDICISGFMGLDLDTRTHVEWILGDVFIGKFYTEFDFGNNRVGLAEAV. IYLIGSLQ
                                                          ] [ 3'UTR ]
```

Figure 4.2.1c. Annotated predicted amino acid sequence of cDNA clone A20. Putative catalytic aspartic acids are in bold and underlined. Clones A20 and A16 are identical sequences apart from A16 containing an extra 9 amino acids and two substitutions Ser_{276} and Arg_{329}. Signal peptide identified by SignalP server (Nielsen *et al.*, 1997), pro-peptide and mature chain junction positioned by similarity to mLAP (Cho and Raikhel, 1992).

```
Du-asp 2  : RIPLKKFEKPRDSLNRNG--LKETLLA----KYSGNF---NDDSAVVLTNYLDAQYYGEI 71
            RIPLKKF   R +L+ +G  L+E + +    KY+  F   ND +   L NYLDAQYYGEI
Zeb-asp   : RIPLKKFRTLRRTLSDSGRSLEELVSSSNSLKYNLGFPASNDPTPETLKNYLDAQYYGEI 79

Du-asp 2  : GIGTPAQSFNVIFDTGSSNLWVPSSKCGFLEVACLLHNKYDSDKSSTYVKNDTRFAIRYG 131
            G+GTP Q+F V+FDTGSSNLWVPS  C   ++ACLLH+KY+  KSSTYVKN T+FAI+YG
Zeb-asp   : GLGTPVQTFTVVFDTGSSNLWVPSVHCSLTDIACLLHHKYNGGKSSTYVKNGTQFAIQYG 139

Du-asp 2  : SGDVAGVVSQDVVEVGGLQAKDQLFAEATQEPGLAFLVGKFDGILGMGYPEISVNGITPV 191
            SG ++G +SQD   +G +  + Q+F EA ++PG+AF+  KFDGILGM YP ISV+G+ PV
Zeb-asp   : SGSLSGYLSQDTCTIGDIAVEKQIFGEAIKQPGVAFIAAKFDGILGMAYPRISVDGVPPV 199

Du-asp 2  : FNTLVEQGAIKEPVFSFYLNRDPDGDVGGELLLGGSDSNYYKGDFTYIDVSAKGYWQIVM 251
            F+ ++ Q  +++ VFSFYLNR+PD   GGELLLGG+D  YY GDF Y+D+S + YWQI M
Zeb-asp   : FDMMMSQKKVEKNVFSFYLNRNPDTQPGGELLLGGTDPKYYTGDFNYVDISRQAYWQIHM 259

Du-asp 2  : DSLSVGSSLKLCSNGCQVIVDTG--TSLIAGPSAEVEKLHQEIGAFSFLNGEYIIDCNRV 309
            D +S+GS L LC  GC+ IVDTG  TSLI GP+AEV+ L + IGA   + GEY++DC +V
Zeb-asp   : DGMSIGSGLSLCKGGCEAIVDTGTSTSLITGPAAEVKALQKAIGAIPLMQGEYMVDCKKV 319

Du-asp 2  : DQLPEISFVFGGKEFKLSGNDYILKQSNGLIDICISGSMGLDLDTRTHVEWILGDVFIGK 369
              LP ISF  GGK + L+G  YILK+S G  DIC+SG MGLD+       WILGDVFIG+
Zeb-asp   : PTLPTISFSLGGKVYSLTGEQYILKESQGGHDICLSGFMGLDIPPPAGPLWILGDVFIGQ 379

Du-asp 2  : FYTEFDFGNNRVGLAEA 386
            +YT FD  NNRVG A+A
Zeb-asp   : YYTVFDRENNRVGFAKA 396
```

Identities = 213/377 (56 %), Positives = 272/377 (71 %), Gaps = 11/377 (2 %)
Score = 438 bits (1126), Expect = e-122

Figure 4.2.1d. (Previous page). Alignment search result for the amino acid sequence predicted by cDNA clone A20 (Du-asp 2). Du-asp 2 shows closest homology to a cathepsin D of *Danio rerio* (Zebrafish), (Riggio *et al.*, 2000). Alignment performed by blastp (www.ncbi.nlm.nih.gov/BLAST/), + indicates conservative substitution.

```
Du-asp 1:    MYAKFLVVLVVAALASSEIIRIPLKKVPDNGQRRLRNVARSGLRNKFGGQ----GAVPLV 57
             M +K L+++  AA+A++EI RIPLKK          RN  +  L  K+ G      AV L
Du-asp 2:    MLSKLLMLVFAAAVATAEIARIPLKKFEKPRDSLNRNGLKETLLAKYSGNFNDDSAVVLT 60

Du-asp 1:    NEYDLDYYGEIGVGTPAQKFNVIFDTGSTDLWVPSSKCQATSAQNGCKNHKQYDASKSST 117
             N  D  YYGEIG+GTPAQ FNVIFDTGS++LWVPSSKC        C  H +YD+ KSST
Du-asp 2:    NYLDAQYYGEIGIGTPAQSFNVIFDTGSSNLWVPSSKCGFLEV--ACLLHNKYDSDKSST 118
                                                 1--------1

Du-asp 1:    YQKDGRSFHIEYGSGALSGFLSKDSVEVAGLAITDQTFAEATVETAPAFVSSTFDGILGL 177
             Y K+   F I YGSG ++G +S+D VEV GL   DQ FAEAT E   AF+   FDGILG+
Du-asp 2:    YVKNDTRFAIRYGSGDVAGVVSQDVVEVGGLQAKDQLFAEATQEPGLAFLVGKFDGILGM 178

Du-asp 1:    AYPDLSAQNIPPVFQNMIKQGVLDAPVFSFYLSQTANGD-KGELLLGGSDSKYYKGDFTY 236
              YP++S   I PVF  +++QG +  PVFSFYL++  +GD  GELLLGGSDS YYKGDFTY
Du-asp 2:    GYPEISVNGITPVFNTLVEQGAIKEPVFSFYLNRDPDGDVGGELLLGGSDSNYYKGDFTY 238

Du-asp 1:    TKVSTQLYWQTNLQGVSVGSR-SVCKSGCEAVIDTGTSLIYGPTDDVDVVNSAIGATYDY 295
               VS + YWQ  +  +SVGS   +C +GC+ ++DTGTSLI GP+ +V+ ++  IGA + +
Du-asp 2:    IDVSAKGYWQIVMDSLSVGSSLKLCSNGCQVIVDTGTSLIAGPSAEVEKLHQEIGA-FSF 297
                                  2---2

Du-asp 1:    SVGLYTVNCNTDLNKLPNVSFTFGGKKFDIPASAYIIKDSG----YCISSFVAQEF-FLS 350
               G Y ++CN  +++LP +SF FGGK+F +  + YI+K S      CIS  +  +    +
Du-asp 2:    LNGEYIIDCNR-VDQLPEISFVFGGKEFKLSGNDYILKQSNGLIDICISGSMGLDLDTRT 356
                    3-------------------------------------3

Du-asp 1     GFEWLVGDSFLKTVYSEFDFGNNRIGFAE 379
               EW++GD F+   Y+EFDFGNNR+G AE
Du-asp 2     HVEWILGDVFIGKFYTEFDFGNNRVGLAE 385
```

Identities = 190/389 (48 %), Positives = 253/389 (64 %), Gaps = 15/389 (3 %)
Score = 355 bits (910), Expect = 8e-97

Figure 4.2.1e. Alignment of *D. undecimpunctata* aspartic proteinase sequences 1 and 2. Homology calculated by BLAST 2 server (Tatusova and Madden, 1999) + indicates conservative substitution. Signal peptides are shown in italics, putative catalytic aspartic acids are underlined and in bold, cysteines involved in di-sulphide bridge formation are in bold, numbers indicate which bridge the pairs of cysteines form (Karlsen *et al.*, 1997). Predicted mature protein start positions are shown by white font in a grey box.

```
1                                                                    67
ERIQFCYK MVSQFLKIKMKFLILIFLTTAAIQA LNDKEKWVQFKVKHNKSYINYVEEQTRFTIFQEN
[ 5'UTR] [    Leader sequence    ] [
68                                                                  136
LKKIENHNKKYYNGESSFMLGVTKFADLTEKEFMDMLVLSKNSRPNRPHVTHLLAPLGD LPSAFDWRN
                          Propeptide                          ] [
137                                                                 205
KGAVTEVKDQGMCGSCWTFSTTGSVEGAHFLKTGNPVSLSEQNLVDCANDNCYGCGGGWMDKALEYIEK
                               Mature chain
206                                                                 274
GGIMSEKDYPYEGVDDKCRFDVSKVAAKISNFTYIKKNDEEDLKNAVFAKGPISVAIDASSKFQLYVAG

275                                                                 343
ILDDTECSNEFDSLNHGVLVVGYGTENGKDYWIVKNSWGVNWGMDGYIRMSRNKNNQCGITTDGVYPNV

344                                 377
. RHAFRIFHSDVLLI..F.NNFLVIRLFFI.NKVE
] [            3'UTR               ]
```

Figure 4.2.2a. Annotated predicted amino acid sequence of clone Ca7. Putative catalytic amino acids are in bold and underlined, sequence features are annotated underneath. Signal peptide identified by SignalP server (Nielsen *et al.*, 1997), pro-peptide and mature chain positioned by similarity to database sequences and the frequently conserved Pro 2 in mature proteins of family C1 of which cathepsin L's are members (Barrett *et al.*, 1998).

```
Du-cys 1 : MKFLILIFLTTAAIQALNDKE---KWVQFKVKHNKSYINYVEEQTRFTIFQENLKKIENH 74
           MK  I   L   AI A   +E   K+  FK++H K+Y+N  EE  RF IF +N++ IE H
Hyp-cys  : MKVAIFFSLLVVAISASISEELGAKFQAFKLEHGKTYLNQAEESKRFNIFTDNVRAIEAH 60

Du-cys 1 : NKKYYNGESSFMLGVTKFADLTEKEFMDMLVLSKNSRPNRPHVTHLLAPLGDLPSAFDWR 134
           N  Y  G+ S+  G+ KF D++++EF  ML LS + +P     +++   + ++PS+ DWR
Hyp-cys  : NALYEQGKVSYKKGINKFTDMSQEEFKTMLTLSASRKPTLETTSYVKTGV-EIPSSVDWR 119

Du-cys 1 : NKGAVTEVKDQGMCGSCWTFSTTGSVEGAHFLKTGNPVSLSEQNLVDCANDNCYGCGGGW 194
            +G VT VKDQG CGSCW FS TGS EGA+  K+G  VSLSEQ L+DC  D   GC GG
Hyp-cys  : KEGRVTGVKDQGDCGSCWAFSITGSTEGAYARKSGKLVSLSEQQLIDCCTDTSAGCDGGS 179

Du-cys 1 : MDKALEYIEKGGIMSEKDYPYEGVDDKCRFDVSKVAAKISNFTYIKKNDEEDLKNAVFAK 254
           +D   +Y+ K G+ SE+ Y Y+G D  C+++V+ V  K+S +T I   DE+ L  AV
Hyp-cys  : LDDNFKYVMKDGLQSEESYTYKGEDGACKYNVASVVTKVSKYTSIPAEDEDALLEAVATV 239

Du-cys 1 : GPISVAIDASSKFQLYVAGILDDTECSNEFDSLNHGVLVVGYGTENGKDYWIVKNSWGVN 314
           GP+SV +DA S    Y +GI +D +CS     LNH +L VGYGTENGKDYWI+KNSWG +
Hyp-cys  : GPVSVGMDA-SYLSSYDSGIYEDQDCSPA--GLNHAILAVGYGTENGKDYWIIKNSWGAS 296

Du-cys 1 : WGMDGYIRMSRNKNNQCGITTDGVYPNV  342
           WG  GY R++R K NQCGI+ D VYP +
Hyp-cys  : WGEQGYFRLARGK-NQCGISEDTVYPTID 324
```

Identities = 155/328 (47 %), Positives = 209/328 (63 %), Gaps = 8/328 (2 %)
Score = 312 bits (799), Expect = 5e-84

Figure 4.2.2b. Alignment search result for the amino acid sequence predicted by cDNA clone Ca7 (Du-cys 1). The amino acid sequence shows closest homology to a cathepsin L-like cysteine proteinase isolated from the midgut of *Hypera postica* (alfalfa weevil), (Wilhite *et al.*, unpublished). Alignment performed by blastp (www.ncbi.nlm.nih.gov/BLAST/), + indicates conservative substitution.

```
1                                                                    57
KSFVVRPST..FC.IGYI MKFLLLFVAFFVGSQA ISFVDLVQGEWTAFKMTHRKSYESPTEEKFRMK
[     5'UTR      ] [    Leader    ] [
68                                                                  136
IFMENSHKVAKHNKLFALGLVSYKLGLNKYADMLHHEFIATLNGFNKTKNMLRQSELEDSVTFIKPANV
                              Pro-peptide
137                                                                 205
E LPGEVDWRPKGAVTGVKDQGHCGSCWSFSATGSLEGQHFRKTGKLVSLSEQNLVDCSTKFGNNGCNG
] [                              Mature chain
206                                                                 274
GLMDNAFRYIKANGGIDTEASYKYKAEDEKCHYNPKKIGATDKGFVDIESGNEDALQAAVATIGPVSVA

275                                                                 343
IDASHETFQLYNQGVYYEPECSSYELDHGVLVVGYGTENNQDYWLVKNSWGPSWGLDGYIKMARNRNNS

344                     367
CGIATQASYPLV .VYRLI.CSV.TK
           ] [   3'UTR    ]
```

Figure 4.2.2c. Annotated predicted amino acid sequence of clone C21. Putative catalytic amino acids are in bold and underlined, sequence features are annotated underneath. Signal peptide identified by SignalP server (Nielsen *et al.*, 1997), pro-peptide and mature chain positioned by similarity to database sequences and the conserved Pro 2 in mature proteins of family C1 of which cathepsin L's are members (Barrett *et al.*, 1998).

```
Du-cys 2 : MKFLLLFVAFFVGSQAISFVDLVQGEWTAFKMTHRKSYESPTEEKFRMKIFMENSHKVAK 75
           MK  L+  A  +  QA+SF DLVQ +W++FKM H K+Y+S TEE+FRMKIFMEN+HKVAK
Sit-cys  : MKLFLILAAVVISCQAVSFYDLVQEQWSSFKMQHSKNYDSETEERFRMKIFMENAHKVAK 60

Du-cys 2 : HNKLFALGLVSYKLGLNKYADMLHHEFIATLNGFNKTKN-MLRQSELEDSVTFIKPANVE 134
           HNKLF+ G V +KLGLNKYADMLHHEF++TLNGFNKTKN +L+ S+L D+V FI PANV+
Sit-cys  : HNKLFSQGFVKFKLGLNKYADMLHHEFVSTLNGFNKTKNNILKGSDLNDAVRFISPANVK 120

Du-cys 2 : LPGEVDWRPKGAVTGVKDQGHCGSCWSFSATGSLEGQHFRKTGKLVSLSEQNLVDCSTKF 194
           LP  VDWR KGAVT VKDQGHCGSCWSFSATGSLEGQHFRKTGKLVSLSEQNLVDCS ++
Sit-cys  : LPDTVDWRDKGAVTEVKDQGHCGSCWSFSATGSLEGQHFRKTGKLVSLSEQNLVDCSGRY 180

Du-cys 2 : GNNGCNGGLMDNAFRYIKANGGIDTEASYKYKAEDEKCHYNPKKIGATDKGFVDIESGNE 254
           GNNGCNGGLMDNAFRYIK NGGIDTE SY Y AEDEKCHY  +  GATDKGFVDIE  NE
Sit-cys  : GNNGCNGGLMDNAFRYIKDNGGIDTEKSYPYLAEDEKCHYKAQNSGATDKGFVDIEEANE 240

Du-cys 2 : DALQAAVATIGPVSVAIDASHETFQLYNQGVYYEPECSSYELDHGVLVVGYGT-ENNQDY 313
           D L+AAVAT+GPVS+AIDASHETFQLY+ GVY +PECSS ELDHGVLVVGYGT ++ QDY
Sit-cys  : DDLKAAVATVGPVSIAIDASHETFQLYSDGVYSDPECSSQELDHGVLVVGYGTSDDGQDY 300

Du-cys 2 : WLVKNSWGPSWGLDGYIKMARNRNNSCGIATQASYPLV 351
           WLVKNSWGPSWGL+GYIKMARN++N CG+A+QASYPLV
Sit-cys  : WLVKNSWGPSWGLNGYIKMARNQDNMCGVASQASYPLV 338
```

Identities = 261/338 (77 %), Positives = 295/338 (87 %), Gaps = 2/338 (0 %)
Score = 551 bits (1421), Expect = e-156

Figure 4.2.2d. Alignment search result for the amino acid sequence predicted by cDNA clone C21 (Du-cys 2). The amino acid sequence shows closest homology to a cathepsin L-like cysteine proteinase isolated from Maize weevil (Matsumoto *et al.*, 1997). Alignment performed by blastp (www.ncbi.nlm.nih.gov/BLAST/), + indicates conservative substitution.

```
1                                                                    69
QWTSFKATHNKSYSVVEDKLRFAVFHENLRKIEEHNAKYENGEETYYLAVNQFADWSSAEFKALLNSQM
[                                Pro-peptide
70                                                                  137
INRPELSFIETFEA DPNLKADSVDWRNKADLGVKNQGSCGSCWAFSATGALEGQLAIHKNQHVQLSEQ
             ] [
138                                                                 206
ELVDCDTTNSGCNGGLMTNAFAYVRSHGLASEKQYAYTARDGSCKKVQNKQVSSISGYVNVAKTESALA
                           Mature chain
207                                                                 275
SALASVGPISIAVDADTWQFYGGGIFNNKNCGTTLNHGVLAVGYTKDVFIVKNSWGTSWGELGYIRISR

276                                              323
GHNLCGLNQMNSYPKL .MIINK.YICVYIYVYIYICIYICIYMHKQIN
               ] [             3'UTR             ]
```

Figure 4.2.2e. Annotated predicted amino acid sequence of clone Cb13. Putative catalytic amino acids are in bold and underlined, sequence features are annotated underneath. Signal peptide identified by SignalP server (Nielsen *et al.*, 1997), pro-peptide and mature chain positioned by similarity to known sequences and the conserved Pro 2 in mature proteins of family C1 of which cathepsin L's are members (Barrett *et al.*, 1998).

```
Du-cys 3 : WTSFKATHNKSY-SVVEDKLRFAVFHENLRKIEEHNAKYENGEETYYLAVNQFADWSSAE 60
           W SFK  H K Y + +E+++RF+VF  NL+ I EHNAKYE G   Y +AVNQFAD +  E
Dv.Cal 1 : WESFKVQHGKVYKNPIEERVRFSVFQANLKTINEHNAKYEQGLVGYTMAVNQFADMTPEE 80

Du-cys 3 : FKALLNSQMINRPELSFIETFEADPNLKADSVDWRNK-ADLGVKNQGSCGSCWAFSATGA 119
           FKA L  Q  N P++      +       DSVDWR K A LGVK+QG CGSCWAFSATG+
Dv.Cal 1 : FKAKLGMQAKNMPKIKKSRHVKNVNAEVPDSVDWRQKGAVLGVKDQGQCGSCWAFSATGS 140

Du-cys 3 : LEGQLAIHKNQHVQLSEQELVDCDTT--NSGCN-GGLMTNAFAYVRSHGLASEKQYAYTA 176
           LEGQ  I   +   LSEQEL+DC     N  C+ GGLMT AF +V  +G+ SE  Y Y A
Dv.Cal 1 : LEGQNYIVNGKSEPLSEQELLDCSVEYGNGDCDEGGLMTLAFEFVEENGIVSEASYPYEA 200

Du-cys 3 : RDGSCKKVQNKQVSSISGYVNVAKTESALASALASVGPISIAVDADTWQFYGGGIFNNKN 236
             G C+   +K V  I GY  V  +E AL  A+ +VGPIS A+ A+  QF+  GI+++ N
Dv.Cal 1 : IQGDCRTTNDKAVLHIQGYNEVYPSEEALRQAVGTVGPISAAIWAEPIQFFSSGIYDDPN 260

Du-cys 3 : C---GTTLNHGVLAVGYTKD----VFIVKNSWGTSWGELGYIRISRGHNLCGLNQMNSYP 289
           C      L+HG+L VGY ++     +IVKNSWG +WGE GY R+ R   LCGL QM SYP
Dv.Cal 1 : CLNYVEYLDHGILVVGYGEENGTPYWIVKNSWGATWGEEGYFRLKRNIALCGLAQMASYP 320

Du-cys 3 : KL 291
            L
Dv.Cal 1 : VL 322
```

Identities = 149/302 (49 %), Positives = 187/302 (61 %), Gaps = 12/302 (3 %)
Score = 267 bits (682), Expect = 1e-70

Figure 4.2.2f. Alignment search result for the amino acid sequence predicted by cDNA clone Cb13 (Du-cys 3). The amino acid sequence shows closest homology to a cathepsin L-like cysteine proteinase isolated from the midgut of *Diabrotica virgifera virgifera* (Western corn rootworm) (Koiwa *et al.*, 2000). Alignment performed by blastp (www.ncbi.nlm.nih.gov/BLAST/), + indicates conservative substitution.

```
1                                                                    69
FQKNLRKIVEHNSKYEKGEKAYFLKITKFADWTDKELNAILNPKIVAKAQHKNTKTFVR DPNLTRPAS
[                          Pro-Peptide                          ] [
70                                                                  138
IDWRDKAVLAVKDQANCGSCWAFSTTGALEGQLAIHKNQAIPLSEQELMDCDTGNSACFGGNPDVAFEY

139                                                                 207
IESNGISSESQYEYTQQKGECRKVENKPVSSISGWLGVPSDEDALMEAVAQYGPVSVSVFANNDWSLYG
                        Mature chain
208                                                                 275
GGIFEHASCRGHPNHAVLAVGYTQKSWIVKNSWGAAWGEDGYIQLSLGNNQCNITFASQIPLL .TTKY
                                                                ] [
276                                    314
LFK.NIFMLKRCFLSVFRFLCICTVYLYKRE KKKKKKKKK
            3'UTR              ] [poly-a ]
```

Figure 4.2.2g. Annotated predicted amino acid sequence of clone Cb8. Putative catalytic amino acids are in bold and underlined, sequence features are annotated underneath. Signal peptide identified by SignalP server (Nielsen *et al.*, 1997), pro-peptide and mature chain positioned by similarity to database sequences and the conserved Pro 2 in mature proteins of family C1 of which cathepsin L's are members (Barrett *et al.*, 1998).

```
Du-cys 4 : ------------------------------------------------------------   0
Dv.Cal 1 : ----------------LFIAFAAVILSAGALSLNQHWESFKVQHGKVYKNPIEERVRFSV  44

Du-cys 4 : FQKNLRKIVEHNSKYEKGEKAYFLKITKFADWTDKELNAILNPKIVAKAQHKNTKTFVRD  60
           FQ NL+ I EHN+KYE+G   Y + + +FAD T +E  A L  +     + K ++  V++
Dv.Cal 1 : FQANLKTINEHNAKYEQGLVGYTMAVNQFADMTPEEFKAKLGMQAKNMPKIKKSR-HVKN 103

Du-cys 4 : PNLTRPASIDWRDK-AVLAVKDQANCGSCWAFSTTGALEGQLAIHKNQAIPLSEQELMDC 119
            N   P S+DWR K AVL VKDQ  CGSCWAFS TG+LEGQ  I   ++ PLSEQEL+DC
Dv.Cal 1 : VNAEVPDSVDWRQKGAVLGVKDQGQCGSCWAFSATGSLEGQNYIVNGKSEPLSEQELLDC 163

Du-cys 4 : --DTGNSAC-FGGNPDVAFEYIESNGISSESQYEYTQQKGECRKVENKPVSSISGWLGVP 176
             + GN  C  GG   +AFE++E NGI SE+ Y Y   +G+CR   +K V  I G+  V
Dv.Cal 1 : SVEYGNGDCDEGGLMTLAFEFVEENGIVSEASYPYEAIQGDCRTTNDKAVLHIQGYNEVY 223

Du-cys 4 : SDEDALMEAVAQYGPVSVSVFANNDWSLYGGGIFEHASCRGHP---NHAVLAVGYTQKS- 232
             E+AL +AV   GP+S +++A      +  GI++  +C  +    +H +L VGY +++
Dv.Cal 1 : PSEEALRQAVGTVGPISAAIWA-EPIQFFSSGIYDDPNCLNYVEYLDHGILVVGYGEENG 282

Du-cys 4 : ---WIVKNSWGAAWGEDGYIQLSLGNNQCNITFASQIPLL 269
              WIVKNSWGA WGE+GY +L      C +   +  P+L
Dv.Cal 1 : TPYWIVKNSWGATWGEEGYFRLKRNIALCGLAQMASYPVL 322
```

Identities = 119/280 (42 %), Positives = 167/280 (59 %), Gaps = 13/280 (4 %)
Score = 214 bits (545), Expect = 1e-54

Figure 4.2.2h. Alignment search result for the amino acid sequence predicted by cDNA clone Cb8 (Du-cys 4). The amino acid sequence shows closest homology to a cathepsin L-like cysteine proteinase isolated from the midgut of *Diabrotica virgifera virgifera* (Western corn rootworm) (Koiwa *et al.*, 2000). Alignment performed by blastp (www.ncbi.nlm.nih.gov/BLAST/), + indicates conservative substitution.

```
Du-cys 1  MVSQFLKIKMKFLILIFLTTAAIQALNDKE----KWVQFKVKHNKSYINYVEEQTRFTIF   56
Du-cys 2  ---------MKFLLLFVAFFVGSQAISFVDLVQGEWTAFKMTHRKSYESPTEEKFRMKIF   51
Du-cys 3  ----------------------------------QWTSFKATHNKSY-SVVEDKLRFAVF   25
Du-cys 4  ------------------------------------------------------------F    1
                   MKFL L         QA           W  FK  H KSY    E   R   F   60

Du-cys 1  QENLKKIENHNKKYYNGESSFMLGVTKFADLTEKEFMDMLV-LSKNSRPNRPH-----VT  110
Du-cys 2  MENSHKVAKHNKLFALGLVSYKLGLNKYADMLHHEFIATLNGFNKTKNMLRQSELEDSVT  111
Du-cys 3  HENLRKIEEHNAKYENGEETYYLAVNQFADWSSAEFKALL-----NSQMINRPELSFIET   80
Du-cys 4  QKNLRKIVEHNSKYEKGEKAYFLKITKFADWTDKELNAIL-----NPKIVAKAQHKNTKT   56
            N  KI  HN      G     L     AD    E    L                  T

Du-cys 1  HLLAPLGDLPSAFDWRNKGAVTEVKDQGMCGSCWTFSTTGSVEGAHFLKTGNPVSLSEQN  170
Du-cys 2  FIKPANVELPGEVDWRPKGAVTGVKDQGHCGSCWSFSATGSLEGQHFRKTGKLVSLSEQN  171
Du-cys 3  FEAPPNL-KADSVDWRNK-ADLGVKNQGSCGSCWAFSATGALEGQLAIHKNQHVQLSEQE  138
Du-cys 4  FVRDPNLTRPASIDWRDK-AVLAVKDQANCGSCWAFSTTGALEGQLAIHKNQAIPLSEQE  117
                     DWR K A   VK Q  CGSCW FS TG  EG           LSEQ
                                    1-----------------------------

Du-cys 1  LVDCAND-NCYGCGGGWMDKALEYIE-KGGIMSEKDYPYEGVDDKCRFDVSKVAAKISNF  228
Du-cys 2  LVDCSTKFGNNGCNGGLMDNAFRYIKANGGIDTEASYKYKAEDEKCHYNPKKIGATDKGF  231
Du-cys 3  LVDCDT--TNSGCNGGLMTNAFAYVRSHG-LASEKQYAYTARDGSCKKVQNKQVSSISGY  195
Du-cys 4  LMDCDT--GNSACFGGNPDVAFEYIESNG-ISSESQYEYTQQKGECRKVENKPVSSISGW  174
          L DC        C GG    A  Y    G       Y Y      C     K
          ----------1
             2-------------------------------------------2

Du-cys 1  TYIKKNDEEDLKNAVFAKGPISVAIDASSK-FQLYVAGILDDTECSNEFDSLNHGVLVVG  287
Du-cys 2  VDIESGNEDALQAAVATIGPVSVAIDASHETFQLYNQGVYYEPECSSY--ELDHGVLVVG  289
Du-cys 3  VNVAK-TESALASALASVGPISIAVDADT--WQFYGGGIFNNKNCGTT---LNHGVLAVG  249
Du-cys 4  LGVPS-DEDALMEAVAQYGPVSVSVFANND-WSLYGGGIFEHASCRGH---PNHAVLAVG  228
                E  L  A    GP S    A       Y         C        H VL VG
                                                     3---------------

Du-cys 1  YGTENGKDYWIVKNSWGVNWGMDGYIRMSRNKNNQCGITTDGVYPNV   333
Du-cys 2  YGTENNQDYWLVKNSWGPSWGLDGYIKMARNRNNSCGIATQASYPLV   335
Du-cys 3  Y----TKDVFIVKNSWGTSWGELGYIRISRG-HNLCGLNQMNSYPKL   290
Du-cys 4  Y----TQKSWIVKNSWGAAWGEDGYIQLSLG-NNQCNITFASQIPLL   269
          Y          VKNSWG  WG  GYI       N C       P
          ------------------------------------3
```

Cys 1 v Cys 2	Identities = 165/338 (48 %), Positives = 210/338 (61 %), Gaps = 15/338 (4 %)
Cys 1 v Cys 3	Identities = 133/304 (43 %), Positives = 174/304 (56 %), Gaps = 13/304 (4 %)
Cys 1 v Cys 4	Identities = 120/274 (43 %), Positives = 157/274 (56 %), Gaps = 16/274 (5 %)
Cys 2 v Cys 3	Identities = 137/309 (44 %), Positives = 190/309 (61 %), Gaps = 20/309 (6 %)
Cys 2 v Cys 4	Identities = 113/286 (39 %), Positives = 169/286 (58 %), Gaps = 17/286 (5 %)
Cys 3 v Cys 4	Identities = 147/269 (54 %), Positives = 192/269 (70 %), Gaps = 02/269 (0 %)

Figure 4.2.2i. Alignment of *D. undecimpunctata* cysteine proteinases 1,2,3 and 4. Homology calculated by BLAST 2 server (Tatusova and Madden, 1999). Alignment performed by MAP (Huang, 1994). Conserved cysteines (C) that form di-sulphide bridges are marked by bond (1,2 or 3) and connected by dashed line to the second cysteine of the pair (Cygler and Coulombe, 1999). Cathepsin L-like cysteine proteinase defining ERFNIN motif is shown in bold (Karrer *et al.*, 1993), putative catalytic amino acids are in bold and underlined. Predicted mature protein start is shown by white font in a grey box.

were designed to two motifs, KFDGILG (5'-3') (Ram) and PLWILGD (3'-5') (Jam) chosen as suitable sites by amino acid composition. A mosquito lysosomal aspartic proteinase (mLAP) sequence was employed as a primer design template upon which the nucleotides of the Ram and Jam aspartic proteinase specific primers were based (see figures 4.1a and b).

PCR was performed on samples of the constructed *D. undecimpunctata* cDNA library (see section 4.2.) using multiple primer combinations, the forward (5'-3') Ram primer was paired with the Jam, poly-TA/C/G (polyadenylation site) and T7 (pTripleEx2 vector site) (3'-5') primers. A standard PCR was conducted on cDNA library samples, employing 30 thermal cycles. Each reaction was then analysed by gel electrophoresis on a 1 % agarose gel (see figure 4.1.e.). Three amplified products of approximately 800 bp (Ram-polyTG (RP)), 800 bp (Ram-T7(R7)) and 900 bp (Ram-T7(R7)) were selected for TOPO TA cloning, sequenced and termed R71, R73 and RP3.

Three PCR product clones (RP3, R73 and R71) encoding two differing aspartic proteinase fragments were isolated from the cDNA library. Two clones RP3 and R73, encoded almost identical aspartic proteinase fragments that shared 98 % amino acid identity over 145 amino acids. The longer clone (RP3) exhibited closest homology (see figure 4.1g.) to a cathepsin D identified in yellow fever mosquito (Cho and Raikhel, 1992).

R71 representing the second class of aspartic proteinase possessed highest homology (59 % identity, see figure 4.1h) to a zebrafish aspartic proteinase (Riggio *et al.*, 2000). R71 possessing closest homology to a zebrafish sequence may be due to the limited number of invertebrate cathepsin D sequences present on public databases or may be indicative of a conserved enzyme. R71 and RP3 showed only 46 % amino acid

identity over the obtained fragment length indicating that different genes or gene families encoded the isolated transcripts (see figure 4.1m.).

The obtained aspartic proteinase sequences were accepted as candidates for the pepstatin A inhibited proteolytic activity (see Chapter 3) as the PCR primers Ram and Jam were designed to generic sequence regions conserved in aspartic proteinases. The primer designs were therefore not believed to incorporate specificity within the aspartic proteinase family. The clones RP3 and R71 thus probably represent the most frequently transcribed aspartic proteinases in *D. undecimpunctata* alimentary tract epithelial cells.

An alignment of cysteine proteinase sequences from insects and other species revealed conserved amino acid motifs (see appendix A.6). Degenerate PCR primers were designed to three motifs VDWRTK (5'-3') (F1), CGSCWAF (5'-3') (F2), and WLVKNSW (3'-5') (B1) (see figure 4.1c) chosen as suitable sites by amino acid composition. A *D. melanogaster* sequence was used as a primer design template upon which the nucleotides of the primers were based (see figures 4.1d).

The following primer combinations; F1-B1, F2-B1, F1-polyTG and F2-polyTG were utilised employing 40 thermal cycles for the RT-PCR reaction. Reactions were analysed by 1 % agarose gel electrophoresis (see fig 4.1f). Two amplified fragments of 540 bp and 300 bp were selected from the F1-B1 RT-PCR reaction for further characterisation. PCR on the alimentary tract cDNA library was performed under standard conditions employing 30 thermal cycles. Primer combinations utilised in the direct PCR on the cDNA library were as follows; F1-T7, F2-T7, F1-polyTA, F1-polyTC, F2-polyTA and F2-polyTC (see figure 4.1e). Two PCR products were of approximately 800 bp (F1-T7) and 850 bp (F2-T7) were chosen for cloning and sequencing.

Four clones encoding differing cysteine proteinase fragments were identified by BLAST homology search. Clones F271 (see figure 4.1i.) and F173 (see figure 4.1j.) were obtained in the direct PCR on the cDNA library with primer pairs F2 and T7 and F1 and T7 respectively. Clones F1B1 (see figure 4.1k.) and F1B1a (see figure 4.1l.) were obtained from the RT-PCR reaction with primer pair F1 and B1. No two of the obtained clones showed greater than 53 % identity over the obtained sequence length indicating all the mRNA sequences originated from different genes or gene families (see figure 4.1n).

The cysteine proteinase fragments obtained by PCR amplification showed greatest homology to cathepsin L-like invertebrate cysteine proteinases. F271 and F173 exhibits 44 % and 45 % amino acid identity to a cathepsin L-like cysteine proteinase of alfalfa weevil, F1B1 and F1B1a exhibiting 83 % and 57 % amino acid identity to cathepsin L-like cysteine proteinases of maize weevil and southern cattle tick respectively.

The PCR primers F1, F2 and B1 were designed to generic regions conserved in cysteine proteinases to ensure no selectivity was incorporated in PCR amplification. Cathepsin L-like cysteine proteinases have been previously demonstrated to be present in the digestive tract of *D. virgifera* by affinity purification (Koiwa *et al.*, 2000). Previous identification and non-selective design of the PCR primers for a particular family of cysteine proteinases indicated the sequence fragments obtained were viable candidates for the E-64 inhibited proteolytic activity identified in *D. undecimpunctata* larval alimentary tracts (see Chapter 3).

4.2. Results: Isolation of cathepsin D and L cDNA transcripts from the larval *D. undecimpunctata* alimentary tract cDNA library.

A cDNA library was constructed using the SMART cDNA library construction kit (Clontech) using dissected alimentary tracts of cusp 3rd-4th instar larval *D. undecimpunctata* as starting material. This instar is the most destructive in respect to plant damage. Un-amplified library titre and recombination frequency were calculated by serial dilution and IPTG/X-gal blue-white selection. A minimum of two dilutions were used for calculation of titer and recombination frequency. Library A contained 5430 pfu/µl and was 94 % recombinant and library B contained 9075 pfu/µl for and was 91 % recombinant.

All four differing cysteine proteinase fragments (2b, 3c, F173 and F271) and both differing aspartic proteinase fragments (R71 and RP3) were used to PCR generate dig-dUTP labelled probes. Each labelled probe was used in an individual screen of approximately 7500 plaque forming units (pfu). Following screening each filter showed approximately 30 positive signals, the positions were mapped back to the agar plates and the corresponding plaques excised. pTriplEx2 plasmid was recovered in BM25.8 cells. Plasmid samples were restriction digested with EcoR1 and Sal1 to obtain insert length (see figures 4.2a and b).

Utilising known aspartic (1.5 kb) and cysteine proteinase (1.1 kb) sequences as a guide inserts of approximately 1.5 kb or 1.1 kb were sequenced. Insert sequencing was performed with poly-Tx primers and a 5' sequencing primer provided with the SMART cDNA library kit (Clontech). Sequences were identified by BLAST search and the full clones arranged employing Sequencher™ 3.0 (Gene Codes Corporation).

4.2.1. Results: Identification and characterisation of cathepsin D-like aspartic proteinase cDNA clones.

Two differing aspartic proteinase encoding cDNAs were obtained from the screening procedure. *D. undecimpunctata* aspartic proteinase 1 (Du-asp 1) corresponding to fragment RP3 and *D. undecimpunctata* aspartic proteinase 2 (Du-asp 2) corresponding to fragment R71. Eleven clones from the RP3 and R71 screens of approximately 7500 pfu were sequenced, 10 clones from the RP3 screen were Du-asp 1 type sequences whereas 4 of 11 clones from the R71 screen were Du-asp 2 sequences.

Du-asp 1 was isolated at a ratio of 2.5:1 relative to Du-asp 2 out of a total of 14 aspartic proteinase sequences returned (two Du-asp 1 type sequences were obtained in the R71 screen). Frequency of clone isolation indicates that Du-asp 1 mRNA is more abundant in alimentary tract cells than that of Du-asp 2, due to a lower frequency of Du-asp 2 type clone return from an equal number of clones sequenced. Relative frequency of clone isolation indicates to the author that either, the Du-asp 1 gene/s mRNA transcription is upregulated in alimentary tract cells relative to Du-asp 2 gene/s or the *D. undecimpunctata* genome contains a greater copy number of Du-asp 1-like genes relative to Du-asp 2-like genes.

Du-asp 1, cDNA clones A72, A62, A42, C52 and C51 (see figure 4.2.1a.) corresponded to fragment RP3 (see figure 4.1g.) and showed closest homology to an aspartic proteinase of yellow fever mosquito (see figure 4.2.1b.). Sequencing 2.4 kb of DNA flanking the mLAP gene revealed binding sites of a *Drosophila* homologue of the vertebrate transcription factor CAAT/enhancer binding protein (C/EBP) (Dittmer *et al.*,

1997). The *Drosophila* C/EBP homologue is expressed in the fat body, Malpighian tubules and midgut (Montell *et al.*, 1992).

Du-asp 1 encodes a full length transcript containing 5' untranslated region, hydrophobic leader sequence, stop codon and 3' untranslated region (see figure 4.2.1a). The mature sequence encodes a 330 amino acid protein with a molecular weight of 35.5 kDa, GRAVY value of –0.092 and a theoretical pI of pH 4.44.

Du-asp 2, cDNA clones A16 and A20 (see figure 4.2.1c) corresponded to fragment R71 (see figure 4. 1h.) and most closely resembles a cathepsin D identified in zebrafish (see figure 4.2.1d). Northern blotting revealed expression of the transcript in all organs tested including brain, liver, muscle and female ovary (Riggio *et al.*, 2000). Du-asp 1 encodes a full length transcript containing 5' untranslated region, hydrophobic leader sequence, stop codon and 3' untranslated region (see figure 4.2.1c). The mature sequence encodes a 339 amino acid protein with a molecular weight of 36.6 kDa, GRAVY value of 0.016 and a theoretical pI of pH 4.27. Du-asp 2 isoform A16 contains an additional 9 amino acids $G_{78}T_{79}I_{80}G_{81}R_{82}F_{83}L_{84}L_{85}R_{86}$ at the N-terminal not possessed by Du-asp 2 isoform A20.

Alignment of Du-asp 1 and Du-asp 2 isoform A20 (see Figure 4.2.1e) reveals conservation of cysteines forming putative di-sulphide bonds and putative catalytic aspartic acids Asp_{82} and Asp_{269} (Du-asp 1) and Asp_{85} and Asp_{272} (Du-asp 2). Du-asp 1 and 2 show only 48 % amino acid identity indicating transcription of at least two aspartic proteinase genes or gene families by *D. undecimpunctata* alimentary tract cells. Phylogenetic analysis of the Du-asp 1 and 2 sequences against known vertebrate and other invertebrate aspartic proteinases would further elucidate the phylogeny of the *D. undecimpunctata* cathepsin D-like aspartic proteinases

4.2.2. Results: Identification and characterisation of cathepsin L-like cysteine proteinase cDNA clones.

Four cysteine proteinase encoding clones were isolated from *D. undecimpunctata* alimentary tract cDNA library. *D. undecimpunctata* cysteine proteinase 1 (Du-cys 1) corresponded to fragment F1B1a and showed closest homology to a cathepsin L-like cysteine proteinase isolated from the midgut of alfalfa weevil (Wilhite *et al.*, unpublished) (see figure 4.2.2b.). Du-cys 1, clone CA7 (see figure 4.2.2a) encodes a full length transcript containing 5' untranslated region, hydrophobic leader sequence, stop codon and 3' untranslated region (see figure 4.2.2a). The mature sequence encodes a 216 amino acid protein with a molecular weight of 23.7 kDa, GRAVY value of –0.430 and theoretical pI of pH 4.75.

D. undecimpunctata cysteine proteinase 2 (Du-cys 2) corresponded to fragment F1B1 and showed closest homology to a cathepsin L-like cysteine proteinase isolated from maize weevil that is expressed in the gastric caeca and other tissues indicating at least a partial role in extracellular digestion (Matsumoto *et al.*, 1997) (see figure 4.2.2d.). Du-cys 2, clone C21 (see figure 4.2.2c) encodes a full length transcript containing 5' untranslated region, hydrophobic leader sequence, stop codon and 3' untranslated region (see figure 4.2.2c). The mature sequence encodes a 217 amino acid protein with a molecular weight of 23.6. kDa, GRAVY value of –0.514 and theoretical pI of pH 5.22.

D. undecimpunctata cysteine proteinases 3 and 4 (Du-cys 3 and Du-cys 4) corresponding to fragments F173 and F271 respectively show closest homology to a cathepsin L-like cysteine proteinase isolated by soyacystatin affinity chromatography

from midguts of *Diabrotica virgifera virgifera* (Koiwa *et al.*, 2000) (see figures 4.2.2f. and 4.2.2h.). Du-cys 3 and Du-cys 4, clones CB13 and CB8 encode partial transcripts containing propeptide, stop codon and 3' untranslated region (see figures 4.2.2e and g). Mature Du-cys 3 encodes a 208 amino acid protein with a molecular weight of 22.2 kDa, GRAVY value of –0.303 and theoretical pI of pH 8.17. Mature Du-cys 4 encodes a 205 amino acid protein with a molecular weight of 22.2 kDa, GRAVY value of –0.291 and theoretical pI of pH 4.64.

Multiple alignment of the *D. undecimpunctata* cathepsin L-like cysteine proteinases Du-cys 1-4 show the conservation of the putative catalytic site Gly_{147}, Cys_{153}, His_{294} and Asn_{314} (alignment numbering, see figure 4.2.2i). Di-sulphide bridge forming cysteine pairs and the presence of the propeptide ERFNIN sequence possessed by cathepsin L-like, but not cathepsin B-like cysteine proteinases is shown (Karrer *et al.*, 1993). Non of the Du-cys sequences showed >54 % amino acid identity (see figure 4.2.2i), indicating transcription of at least four differing genes or gene families in the alimentary tracts of *D. undecimpunctata*. A full phylogenetic characterisation is required to establish the phylogeny of these sequences in respect to each other, vertebrate and invertebrate cysteine proteinases.

4.3. Discussion.

The bi-lobed tertiary structure of aspartic proteinases hypothesised to be due to a gene duplication and fusion event (Tang *et al.*, 1978) and the presence of the active site cleft

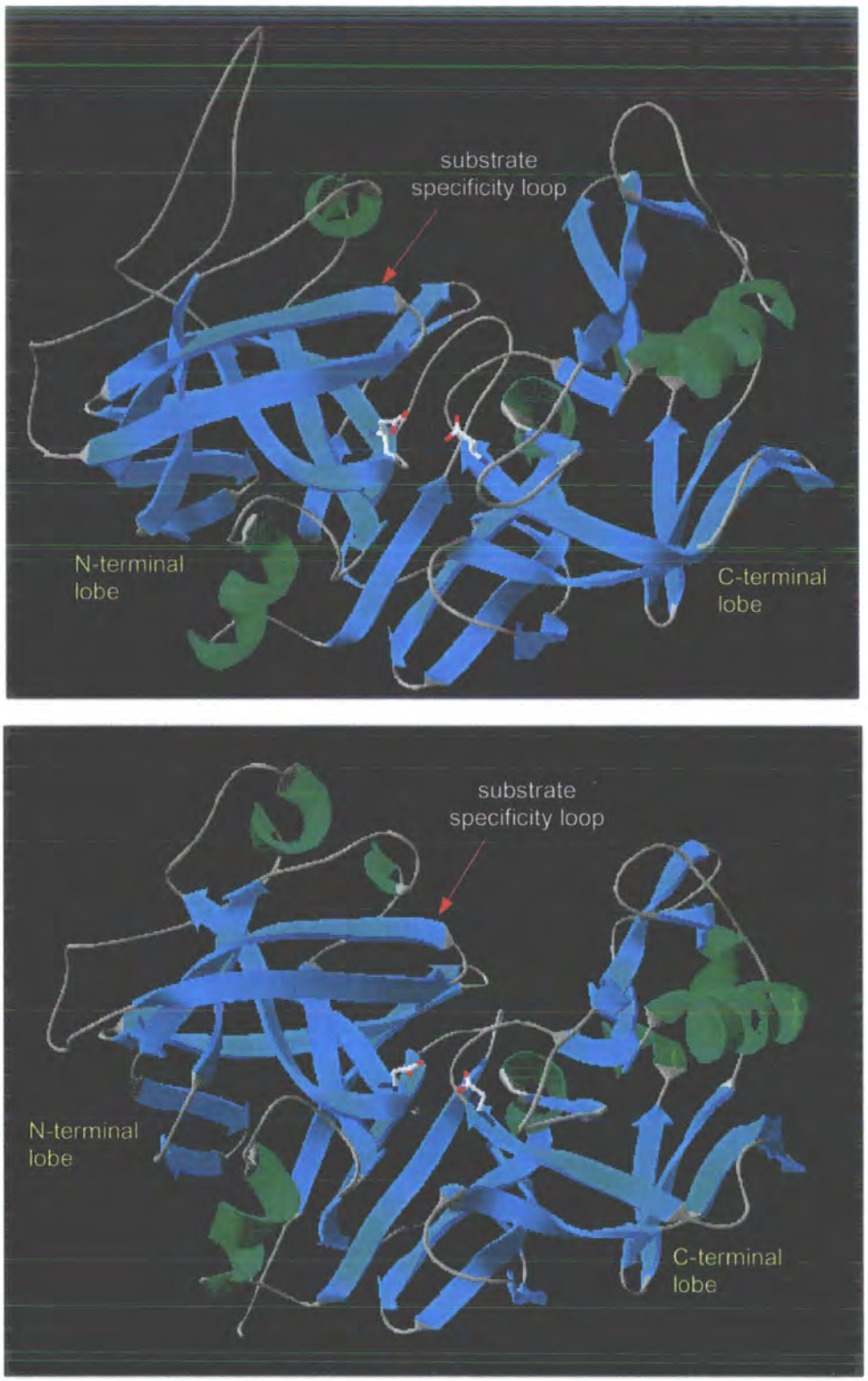
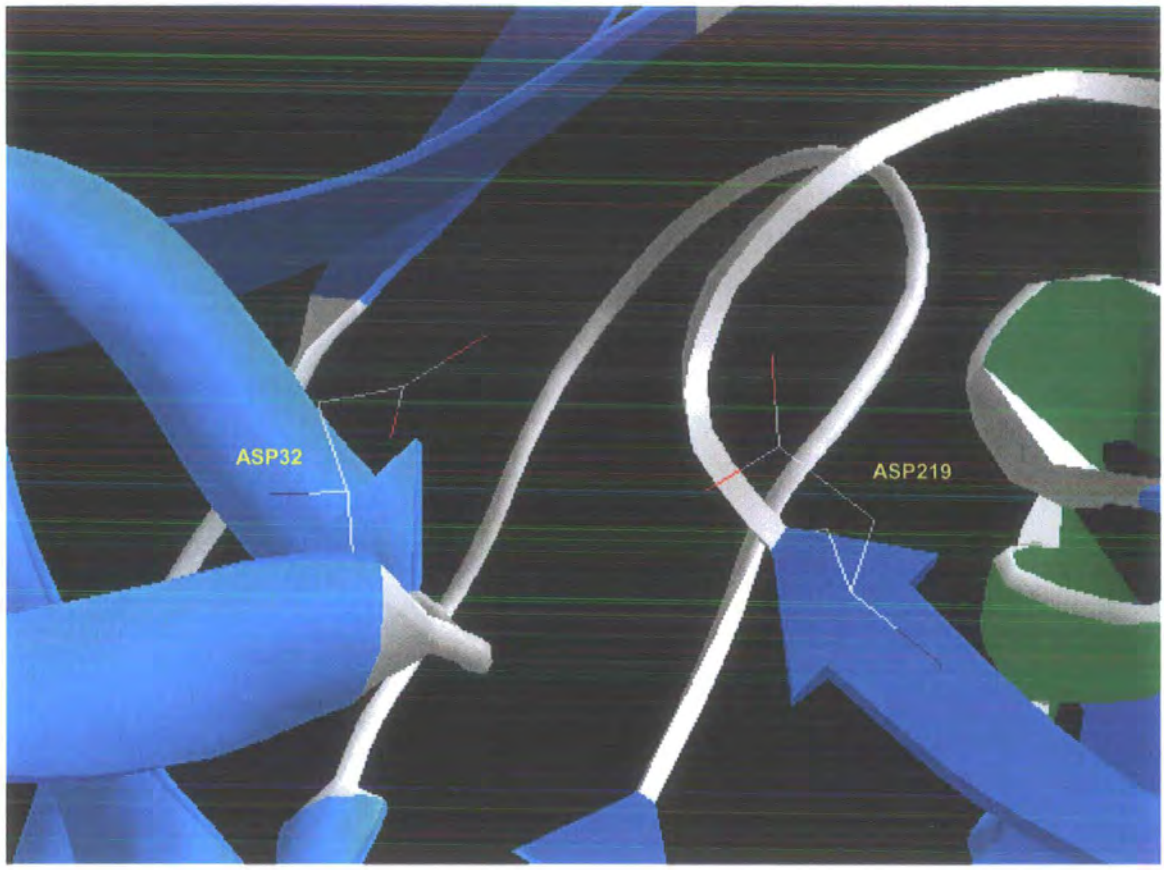


Figure 4.3a. 3D view of Du-asp 1 (top) and Du-asp 2 (bottom) cathepsin D-like aspartic proteinases pictures created with Swiss PDB viewer (Deepview) Version 3.7b1 (Guex *et al.*, unpublished). The SWISS-MODEL (Guex and Peitsch, 1997) predicted bi-lobed protein structure of aspartic proteinases is shown, each lobe providing a catalytic aspartic acid to the active site. The putative catalytic aspartic diad sidechains are shown in solid 3D illustrating the positions, α helices are depicted by green coils, β sheets are shown in blue and strands are in grey.



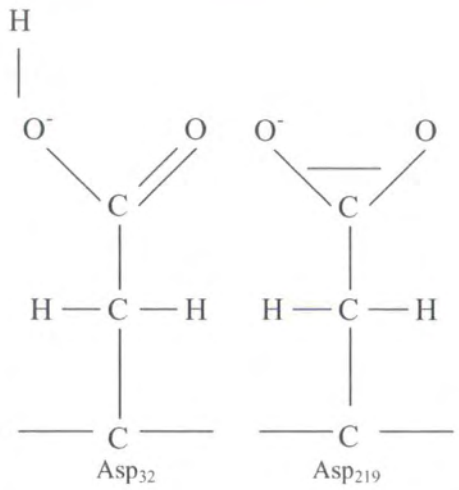


Figure 4.3b. 3D view of the active site of Du-asp 1 created with Swiss PDB viewer (Deepview) Version 3.7b1 (Guex *et al.*, unpublished). The catalytic diad Asp_{32}-Asp_{219} are named, numbered and side chains included to illustrate orientation. Green coils indicate α helices, blue ribbons indicate β sheets and strands are shown in grey. Model created using SWISS-MODEL (Guex and Peitsch, 1997) utilising an Atlantic Cod stomach aspartyl protease model as a template (Karlsen *et al.*, 1997) (See appendix A5 for amino acid alignment). Below main picture is a line drawing of the catalytic aspartic acids shown according to Veerapandian *et al* (1992).

is predicted by SWISS-MODEL (Guex and Peitsch, 1997) homology modelling of Du-asp 1 and 2 (see figure 4.3a). The N and C-terminal lobes of Du-asp 1 providing the aspartic acids Asp_{82} (N-terminal) and Asp_{269} (C-terminal) respectively to the putative catalytic active site. The Asp_{85} and Asp_{272} constitute the putative catalytic active site amino acids of Du-asp 2. The position of the antiparallel ß-pleated sheet substrate specificity loops R_{123}-L_{138} in Du-asp 1 and D_{123}-V_{139} (coding sequence numbering) are also indicated.

Figure 4.3b. shows a 3D Swiss-model homology prediction of the Du-asp 1 catalytic active site, modelled employing cod aspartic proteinase as a template (Karlsen *et al.*, 1997). The catalytic site Asp_{32}, and Asp_{219} sidechains (SWISS-MODEL numbering) are shown including the orientation in the active site. The presence and relative positions of these amino acids further define Du-asp 1 and homologous Du-asp 2 enzymes as aspartic proteinases by catalytic active site orientation and presence. Isolation of Du-asp 1 and Du-asp 2, clones of the former isolated at a ratio of 2.5:1 to the latter from the alimentary tract cDNA library only indicates a possible role in extracellular digestion. The regulation of individual genes or gene families by mRNA transcription, protein translation and subsequent secretion will determine the relative contributions of Du-asp 1 and 2 to *in vivo* dietary digestion in the alimentary tract environment.

No defining research is available in published literature regarding aspartic proteinase localisation and role within the alimentary tract of Cucujiformia. Midgut cathepsin D-like aspartic proteinase activity has been previously identified in *Musca domestica* larvae (Lemos and Terra., 1991). Aspartic proteinase activity has also been

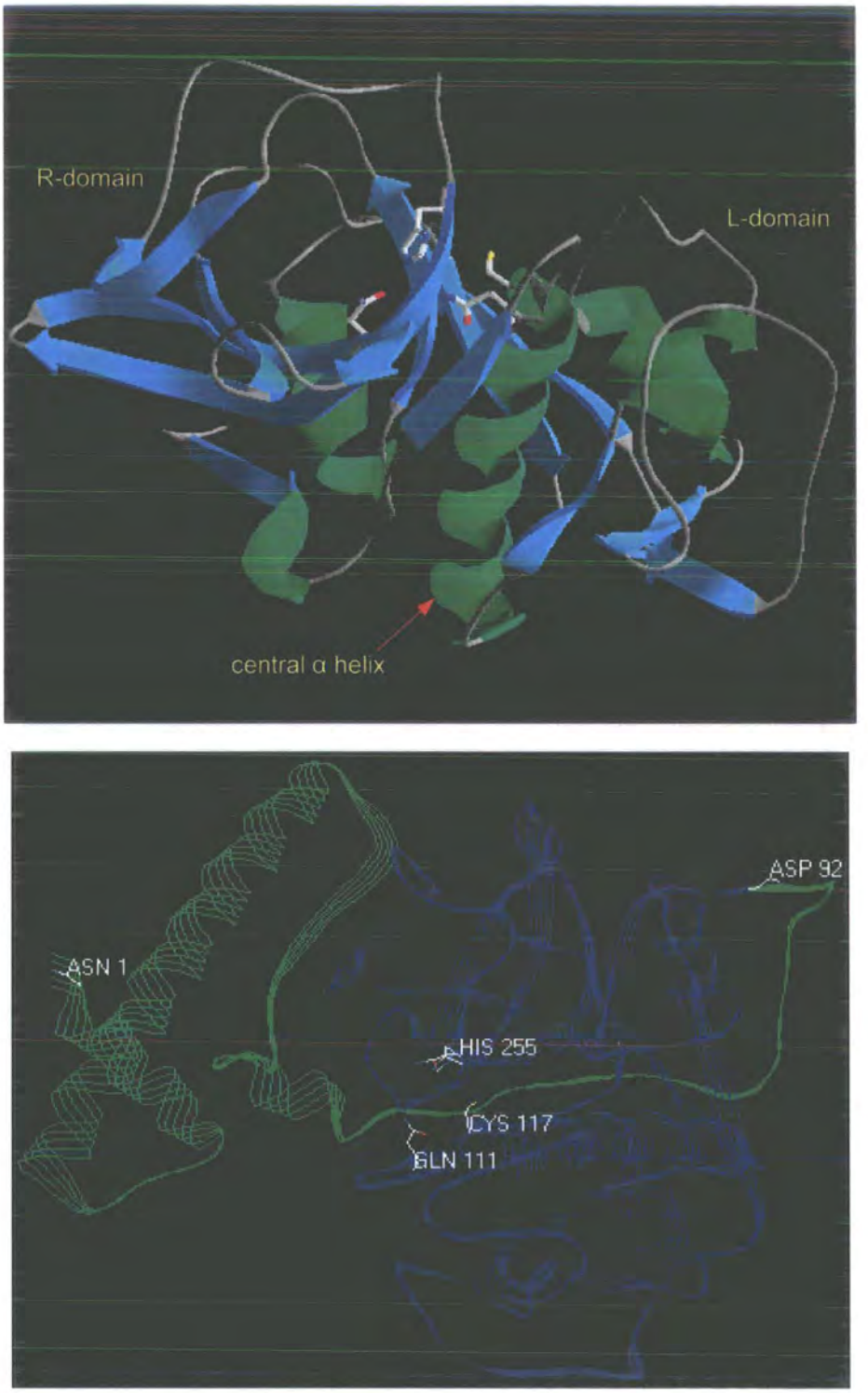
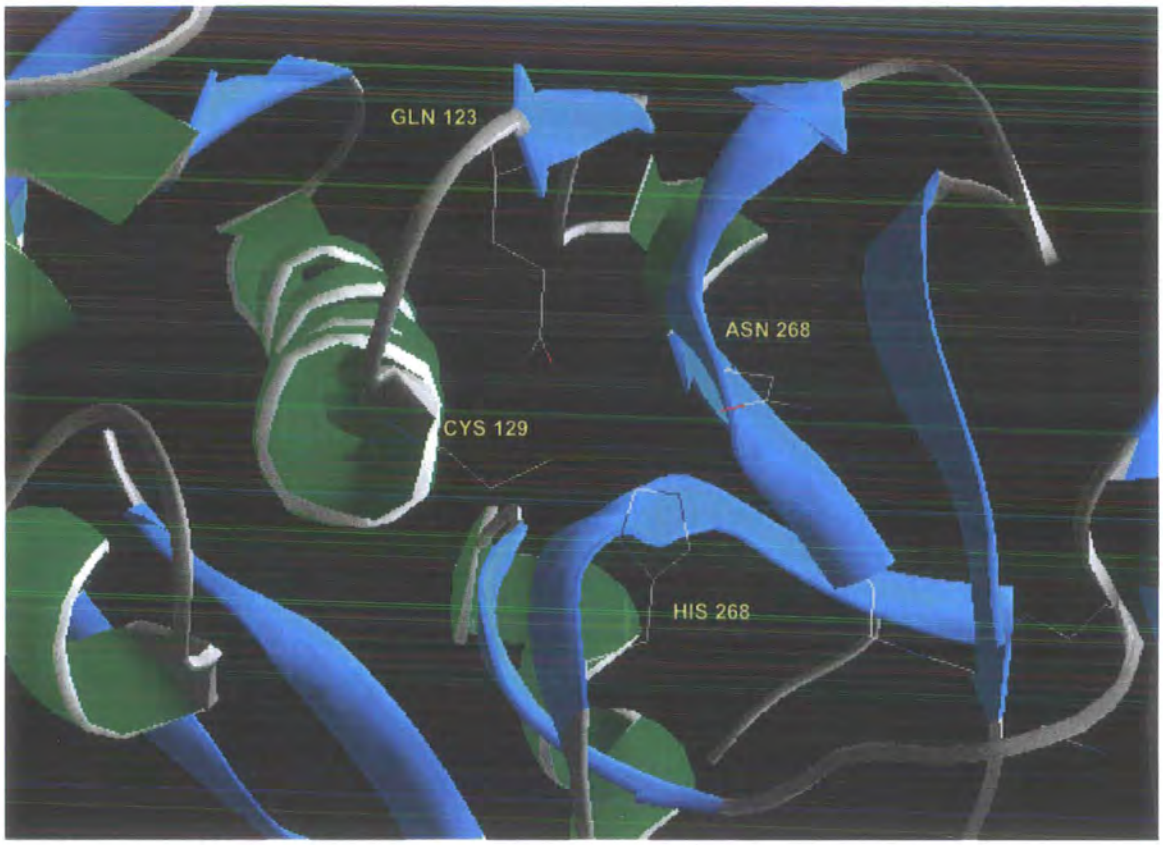


Figure 4.3c. 3D view of Du-cys 1 (top) The putative catalytic tetrad Gln_{123}-Cys_{129}-His_{268}-Asn_{288} amino acids are shown in solid 3D illustrating relative positions, α helices are depicted by green coils, β sheets are shown in blue and strands are in grey. Wireframe model of the Du-cys 1 pro-peptide (shown in green) and mature protein (shown in blue), positions of the putative catalytic Gln_{111}, Cys_{117}, His_{255} are indicated in white with sidechains, putative pro-peptide start Asn_1 and end Asp_{92} are indicated in white. Protein modelled with SWISS-MODEL (Guex and Peitsch, 1997), pictures created with Swiss PDB viewer (Deepview) Version 3.7b1. (Guex *et al.*, unpublished).



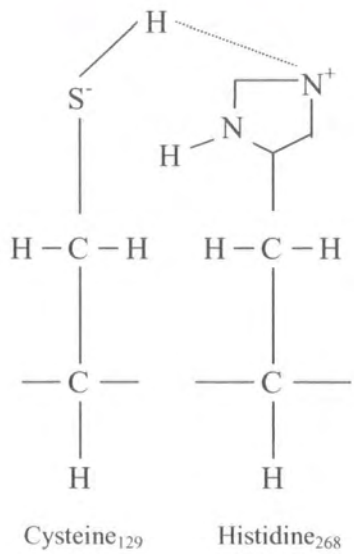


Figure 4.3d. 3D view of the catalytic active site of Du-cys 1 picture created with Swiss PDB viewer (Deepview) Version 3.7b1 (Guex *et al.*, unpublished). The catalytic tetrad Gln_{123}-Cys_{129}-His_{268}-Asn_{288}-is clearly visible illustrating the amino acid and sidechain positions. Below main picture is a line drawings of the catalytic cysteine and histidine. Amino acids are named, numbered and side chains included to illustrate orientation. Model created using SWISS-MODEL (Guex and Peitsch, 1997) (see section A.5. for amino acid alignment) utilising human procathepsin K as a model template (Sivaraman *et al.*, 1999).

identified in midguts of alfalfa weevil (Wilhite *et al.*, 2000) and previously in *D. undecimpunctata* by pepstatin A inhibition (Edmonds *et al.*, 1996). However, extracellular dietary digestion by aspartic proteinases in the alimentary tracts of Cucujiformia remains to be demonstrated.

Figure 4.3c shows a SWISS-MODEL (Guex and Peitsch, 1997) generated prediction of the Du-cys 1 mature chain 3D structure (top) and the mature chain including the pro-peptide (bottom). The L and R-domains are indicated and the L-domain central α-helix terminating at the catalytic cysteine is shown. The pro-peptide of human cathepsin L and *Bombyx mori* cysteine proteinase is necessary for folding, stability and exit of the protein from the endoplasmic reticulum (Tao *et al.*, 1994; Yamamoto *et al.*, 1999). The pro-peptide also regulates the activity of the mouse cathepsin L pro-enzyme, presence of the pro-peptide inactivating the enzyme (Mason *et al.*, 1989). A similar role for the pro-peptide of Du-cys 1 is proposed due to a similar pro-peptide tertiary structure being predicted by homology modelling relative to other cathepsin L-like cysteine proteinase pro-peptides.

Figure 4.3d shows an enlarged view of the catalytic active site of Du-cys 1, modelled with SWISS-MODEL (Guex and Peitsch, 1997) employing human procathepsin K as a template (Sivaraman *et al*, 1999). The catalytic site glutamine$_{123}$, cysteine$_{129}$, histidine$_{268}$ and aparagine$_{288}$ (SWISS-MODEL numbering) including the orientation in the active site are shown. The presence and relative positions of these amino acids further define Du-cys 1 and homologous Du-cys enzymes as cysteine proteinases by catalytic active site presence and orientation.

Glycosylation of cathepsin L is not required for hydrolytic activity (Smith *et al.*, 1989) the mature, non-glycosylated *D. undecimpunctata* cysteine proteinases have

predicted molecular weights of 22-24 kDa. This size corresponds to a minor band of proteolytic activity at 24 kDa activated by 5 mM cysteine identified in *D. undecimpunctata* alimentary tracts by in gel assay hydrolysis assay (Edmonds *et al.*, 1996). A major activity at 35 kDa identified in the same work, may represent Du-cys 1-4 type proteinases possessing the propeptide assisting in refolding (Tao *et al.*, 1994) but in a non-inhibitory confirmation or other class/es of proteolytic activity.

Isolation of cDNAs encoding cathepsin L-like enzymes from the alimentary tract of *D. undecimpunctata* is only indicative of a possible role in extracellular digestion. The regulation of individual genes or gene families by gene transcription, protein translation and subsequent secretion will determine the relative contributions of Du-cys 1-4 to *in vivo* dietary digestion in the alimentary tract environment.

Similar cathepsin L-like sequences have been identified in the midguts of other coleopteran species (Wilhite *et al.*, unpublished; Matsumoto *et al.*, 1997; Girard and Jouanin, 1999b) or have been identified by specific inhibition (Koiwa *et al.*, 2000). Du-cys 1-4 are all on the basis of previous research considered potential candidates for the alimentary tract proteolytic activity previously characterised by E-64, oryzacystatin and potato multicystatin inhibition (Purcell *et al.*, 1992; Orr *et al.*, 1994; Edmonds *et al.*, 1996).

Cathepsin D, cathepsin E and pepsin-like are all members of the aspartic proteinase family. These related aspartic proteinases possess sequences features promoting differential cellular localisation be it lysosomal (cathepsins D), intracellular (cathepsin E) or excretory (pepsin) (Barrett *et al.*, 1998). Mammalian cysteine proteinases, such as cathepsins L, B, K, H and S are lysosomal proteinases, having a

limited extracellular digestive role (Wiederlanders *et al.*, 1992; Velasco *et al.*, 1994; Shi *et al.*, 1994; Barrett *et al.*, 1998; Santamaria *et al.*, 1999; Turk *et al.*, 2000).

Cysteine and aspartic proteinases predominate in the alimentary tracts of *D. undecimpunctata* and other members of Coleopteran series Cucujiformia (Murdock *et al.*, 1987; Terra and Ferreira, 1994). The alimentary tract specific cysteine and aspartic proteinases of Cucujiformia have evolved from intracellular proteinases to perform a dietary digestive role. This evolution is theorised to be an adaptive response to encountering serine proteinase inhibitors in their diet (Murdock *et al.*, 1987).

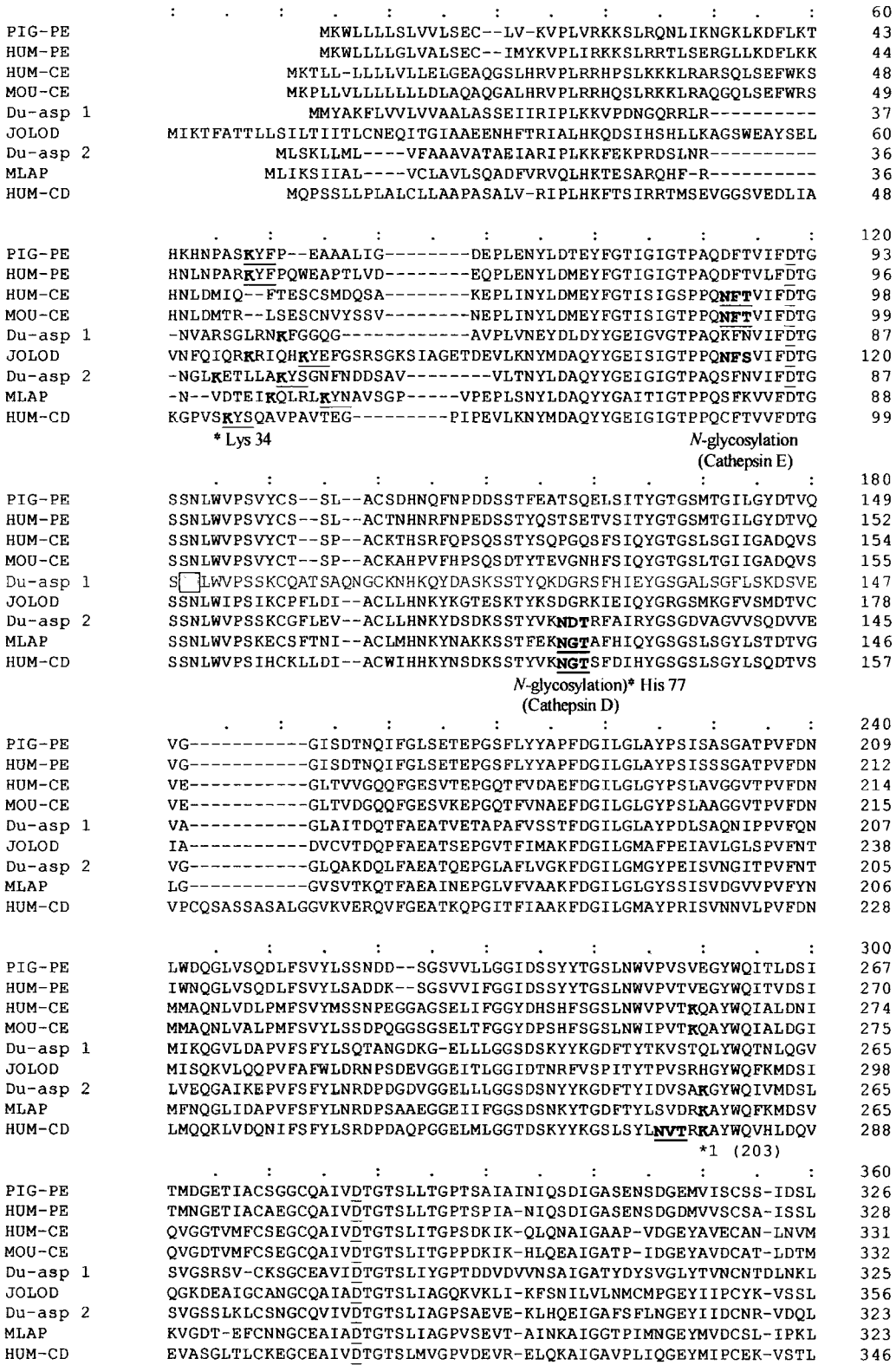
An equivalent evolution has occurred in Cucujiformia and mammals, in respect to adaptation of ancestral lysosomal aspartic proteinases to a intracellular such as cathepsins E or extracellular digestive role as the pepsins. *D. undecimpunctata* cysteine and aspartic proteinase sequences have been aligned with sequences of insect and mammalian proteinases with known *in vivo* localisation profiles. The alignments have been used to identify the presence or absence of conserved *N*-glycosylation sites and lysines and by comparison further refine the functional roles of the cathepsin D and L-like proteinases isolated from *D. undecimpunctata*.

The pepsins (HUM-PE, PIG-PE) contain no potential *N*-glycosylation sites and are secreted *in vivo*. Asn_{70} is a highly conserved *N*-glycosylation site in vertebrate and invertebrate lysosomal cathepsins D both as a single site and with mammalian cathepsins D in combination with Asn_{199} (Cuozzo *et al.*, 1998; Jolodar and Miller, 1998). Asn_{27} is highly conserved in vertebrate and invertebrate cathepsins E (Jolodar and Miller, 1998). The high level of conservation supports the hypothesis that Asn_{70} is a feature of lysosomal cathepsins D and Asn_{27} is a feature of intracellular cathepsins E.

mLAP is a native 80 kDa dimer glycoprotein consisting of two identical 40 kDa subunits. The mosquito protein is localised to the lysosomes *in vivo*, established by density gradient centrifugation of organelles and immunoblot analysis (Cho *et al.*, 1991). mLAP shows highest mature chain amino acid identity (60 %) to human cathepsin D-type aspartic proteinases and contains a single *N*-glycosylation site Asn_{70} (Cho and Raikhel, 1992) homologous to the N-terminal site of human cathepsin D. mLAP upon alignment possesses a conserved lysine at position 203 homologous to the primary human cathepsin D phosphorylation determinant (Cuozzo *et al.*, 1998) and a cationic arginine at position 293 the secondary *N*-linked oligosaccharide phosphorylation determinant.

Du-asp 2 possesses high mature chain amino acid identity (56 %) to mLAP and similarly shows highest (55 %) identity to human cathepsin D-type aspartic proteinases possessing 45 % identity to cathepsin E and 44 % with pepsin. Du-asp 2 possesses a potential *N*-glycosylation site at Asn_{70} a site conserved in mammalian cathepsins D and the lysosomal aspartic proteinase mLAP. mLAP and Du-asp 2 possess lysines at positions 203 and lysine (Du-asp 2) or arginine (mLAP) at position 293 homologous to the primary human cathepsin D phosphorylation determining lysines.

Du-asp 2 shows conservation of amino acid sequence, *N*-glycosylation site position and cationic phosphorylation determinants with human and mosquito lysosomal cathepsins D. Conservation of these features indicates an *N*-linked oligosaccharide ß-linked to Asn_{70} of Du-asp 2 would be efficiently phosphorylated and transported via a putative phosphate receptor mediated targeting pathway to lysosomes. Therefore, Du-asp 2 is predicted to be a ubiquitously expressed lysosomal cathepsin D similar to mLAP that performs a minimal dietary protein digestive role within the


* Lys 34
N-glycosylation
(Cathepsin E)
N-glycosylation)* His 77
(Cathepsin D)
*1 (203)

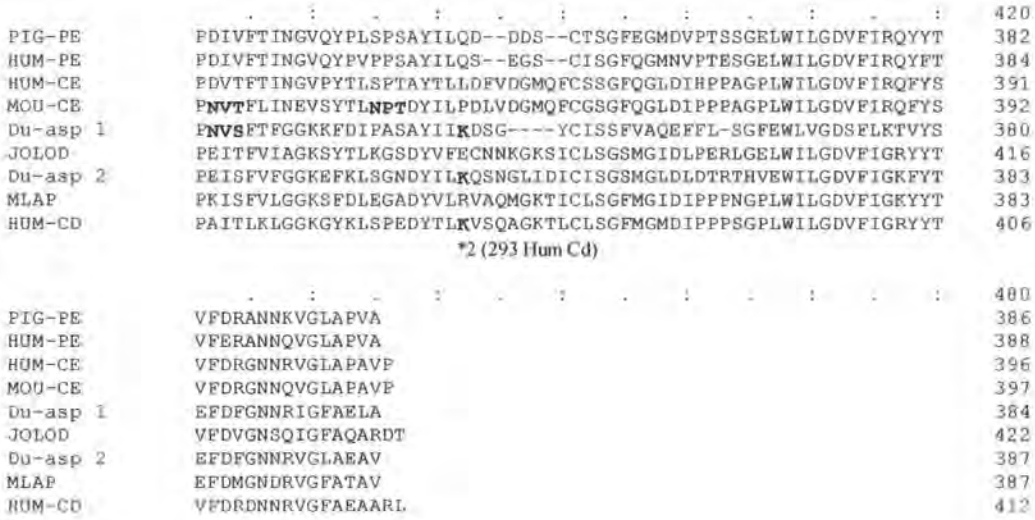


Figure 4.3e. Alignment of *D. undecimpunctata* aspartic proteinases with mammalian pepsins, nematode and insect aspartic proteinases. Du-asp 1 and 2 are aligned with a non-lysosomal cathepsin D (JOLOD) (Jolodar and Miller, 1998), a lysosomal cathepsin D (mLAP) (Cho and Raikhel, 1992) and mammalian aspartic proteinases; PIG-PE; pig pepsin (Moravek and Kostka, 1974), HUM-PE; human pepsin A (Evers *et al.*, 1989), HUM-CE; human cathepsin E (Azuma *et al.*, 1989), MOU-CE; *Mus musculus* procathepsin E (Tatnell *et al.*, 1997), HUM-CD; human cathepsin D sequence (NCBI Annotation Project, 2001). Catalytic aspartic acids are underlined. *N*-glycosylation sites in bold and underlined indicate sites known to be glycosylated *in vivo*, potential *N*-glycosylation sites are in bold. Du-asp 1 Thr_{122} and Asp_{123} (white text on a grey background, alignment numbering) are divergent from the conserved $S_{122}N_{123}$ in all other aspartic proteinases presented. Positions of the primary oligosaccharide phosphorylation determinants Lys_{203} (*1) and Lys_{293} (*2) (HUM-CD; mature chain numbering) are indicated (Cuozzo *et al.*, 1998). *Lys_{34} and *His_{77} denote N-terminal phosphorylation determinants (Dustin *et al.*, 1995).

alimentary tract.

Du-asp 1 exhibits 48 % mature chain amino acid identity to mLAP and 44 % identity to human cathepsin D-type aspartic proteinases, 16 % and 11 % lower amino acid identities than mLAP or Du-asp 2 share with human cathepsin D. Du-asp 1 possesses 42 % identity to human cathepsin E and 39 % to pepsin. Du-asp 1 does not possess the *N*-glycosylation site at Asn_{70} possessed by Du-asp 2, mLAP and human lysosomal cathepsin D (see figure 4.3e.). Du-asp1 shows lower relative amino acid homology to human cathepsin D than lysosomal mLAP or Du-asp 2. Du-asp 1 therefore exhibits significant differences in respect of *N*-glycosylation site position and amino acid conservation to known vertebrate and invertebrate lysosomal aspartic proteinases.

Cathepsins E are non-lysosomal intracellular aspartic proteinases with a limited tissue distribution including gastrointestinal tracts (Azuma *et al.*, 1989). Mammalian cathepsins E possess a common *N*-glycosylation site at Asn_{27} (mature chain numbering) (Yasuda *et al.*, 1999), which is shared with some nematode aspartic proteinases (Jolodar and Miller, 1998; Tcherepanova *et al.*, 2000) such as cathepsin E-like Asp 1 which is specifically expressed in intestinal cells (Tcherepanova *et al.*, 2000).

Cathepsins E of rat and mouse also contain C-terminal *N*-glycosylation sites at Asn_{305} (Yasuda *et al.*, 1999), which upon alignment is homologous to the Du-asp 1 single site at Asn_{313} (see figure 4.3e). In relation to C-terminal *N*-glycosylation site homology Du-asp 1 is most similar to cathepsins E while lacking the highly conserved Asn_{27}. Lack of a highly conserved cathepsin E defining *N*-glycosylation site and pro-peptide cysteine promoting covalent dimer formation (Finley and Kornfield, 1994) is predicted to lead to increased secretion of Du-asp 1 and performance of dietary digestion. Isolation of Du-asp 1 cDNAs at a ratio of 2.5:1 relative to Du-asp 2 cDNAs

indicates increased frequency of Du-asp 1 gene/s mRNA transcription in *D. undecimpunctata* alimentary tract cells or a higher copy number of Du-asp 1 genes in the *D. undecimpunctata* genome.

Pepsins, cathepsins E, JOLOD and Du-asp 1 do not show conservation of lysines at position 203, as expected due to lack of potential *N*-glycosylation sites or differential positioning relative to cathepsins D (see figure 4.3e.). The putative lysine phosphorylation determinants for the C-terminal potential *N*-glycosylation site of Du-asp 1 remain unresolved, as lysine 203 and 293 do not show conservation in aspartic proteinases with glycosylation sites at positions other than Asn_{70}, Asn_{199} or Asn_{337} (Cuozzo *et al.*, 1998). Further research would establish if this site is glycosylated/phosphorylated *in vivo* and if a glycosylated/phosphorylated oligosaccharide at this position is capable of promoting lysosomal transport or intracellular retention of Du-asp 1.

In vivo investigations into the role of Du-asp 1 are required to establish if the protein is membrane bound as other cathepsins E or present as a soluble protein and if Du-asp 1 is retained in the intracellular environment or secreted. Sequence analysis and frequency of clone isolation relative to Du-asp 2 indicates Du-asp 1 does not function as a lysosomal cathepsin D. Du-asp 1 exhibits features associated with pepsin and cathepsin E-like aspartic proteinases and a digestion specific role is proposed for this aspartic proteinase similar to the pepsins or cathepsins E.

A cathepsin L-like cysteine proteinase of *Phaedon cochleariae* (Coleoptera) is highly expressed in midgut material (2.6 %/76 cDNAs) and contains no potential *N*-glycosylation sites (Girard and Jouanin, 1999b). Cathepsins L-like proteinases of and alfalfa weevil (Wilhite *et al.*, unpublished) isolated from midgut material also lack

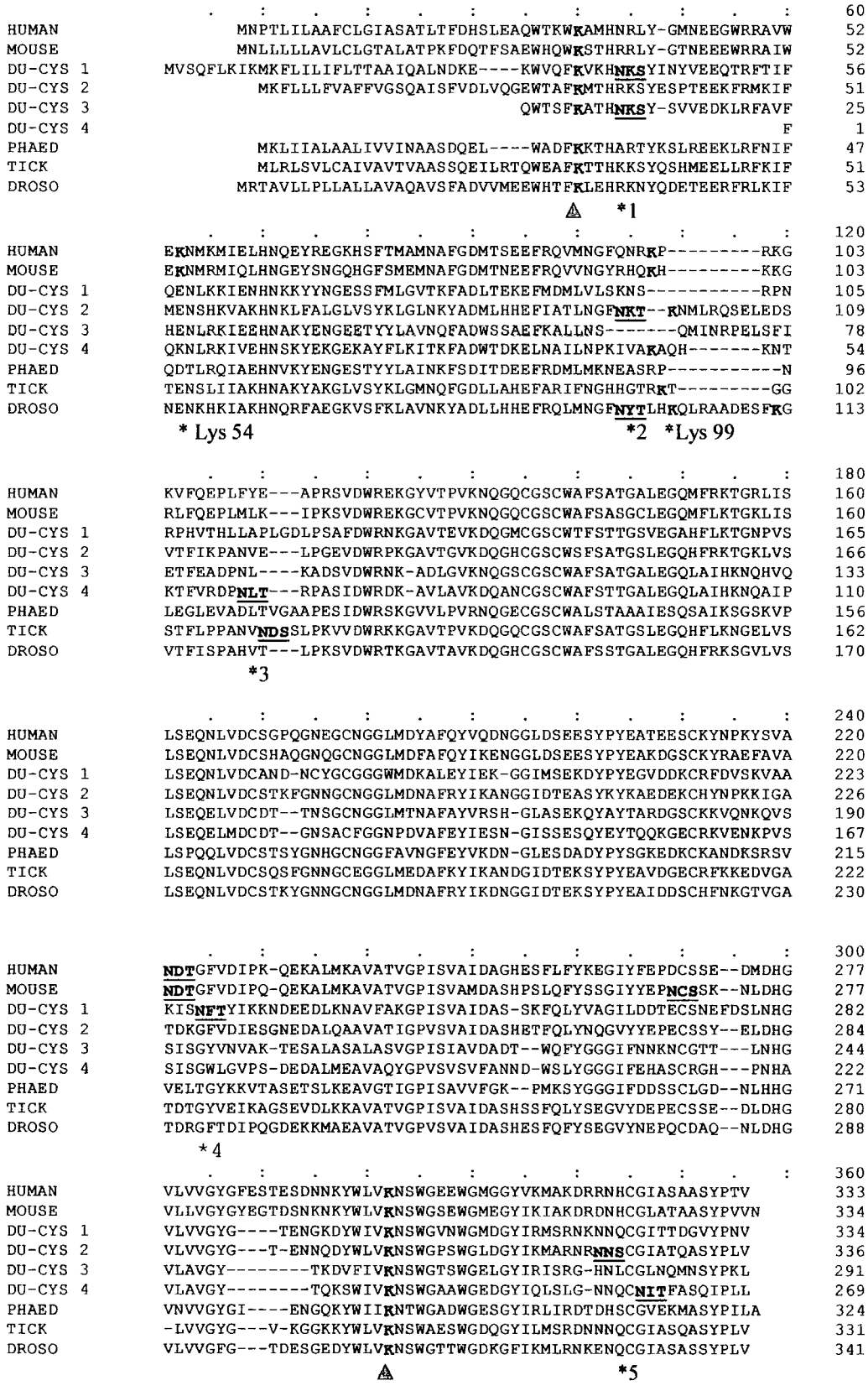

HUMAN
MOUSE
DU-CYS 1
DU-CYS 2
DU-CYS 3
DU-CYS 4
PHAED
TICK
DROSO
*1
* Lys 54
*2
*Lys 99
*3
*4
*5

Figure 4.3f. (previous page). Alignment of *D. undecimpunctata* cathepsin L sequences with invertebrate cysteine proteinases and cathepsin Ls of mouse and human. Amino acids in bold indicate conserved lysines (K) not adjacent to catalytic amino acids. Amino acids in bold and underlined indicate potential *N*-glycosylation sites, *Lys 54 and *Lys 99 (human cathepsin L numbering) the major determinants of oligosaccharide phosphorylation are shown underneath the alignment. Aligned sequences; cattle tick (TICK) (Renard *et al.*, 2000), thiol protease isolated from larval guts of mustard beetle (PHAED) (Girard and Jouanin, unpublished), human cathepsin L (HUMAN) (Machleidt *et al.*, 1986), a transformed mouse fibroblast major excreted cathepsin L (MOUSE)) (Troen *et al.*, 1987) and DROSO a digestive cysteine proteinase isolated from Drosophila and predominately expressed in the midgut (Matsumoto *et al.*, 1995). (*X) regions in which *N*-glycosylation site conservation across two sequences is observed. Δ Indicates conserved Lys 33 and 299 essential for human pro-enzyme folding (Cuozzo *et al.*, 1995).

potential *N*-glycosylation sites. The cathepsin L-like cysteine proteinase sequences of *Sitophilus* and alfalfa weevil would be expected to be constitutively secreted digestive proteinases due to lack of *N*-glycosylation sites by which lysosomal transport is putatively achieved.

The cysteine proteinases of mammals such as cathepsin S, L, H, W, F, O and K are lysosomal proteins, secretion of mouse cathepsin L occurs under specific circumstances such as cell transformation or growth induction (Troen *et al.*, 1987; Prence *et al.*, 1990). The human cysteine proteinase cathepsin L has been extensively characterised in respect to lysosomal transport, lysines 54 and 99 being the primary phosphorylation determinants of an Asn_{221} oligosaccharide that is sufficient to determine lysosomal transport (Kane, 1993; Cuozzo *et al.*, 1995 and 1998).

Mouse and human cathepsin L share a conserved potential *N*-glycosylation site at Asn_{221}, mouse cathepsin L also possesses a site at Asn_{268}. Alignment of *D. undecimpunctata* and invertebrate cathepsin L-like sequences with mammal cathepsin L cysteine proteinases revealed a highly heterogeneous pattern of *N*-glycosylation sites (see figure 4.3f.). Significantly all the obtained *D. undecimpunctata* cysteine proteinases contain potential *N*-glycosylation sites at one or more of the sites 1*, 2*, 3*, 4* or 5* indicating all these sequences may be subject to lysosomal transport.

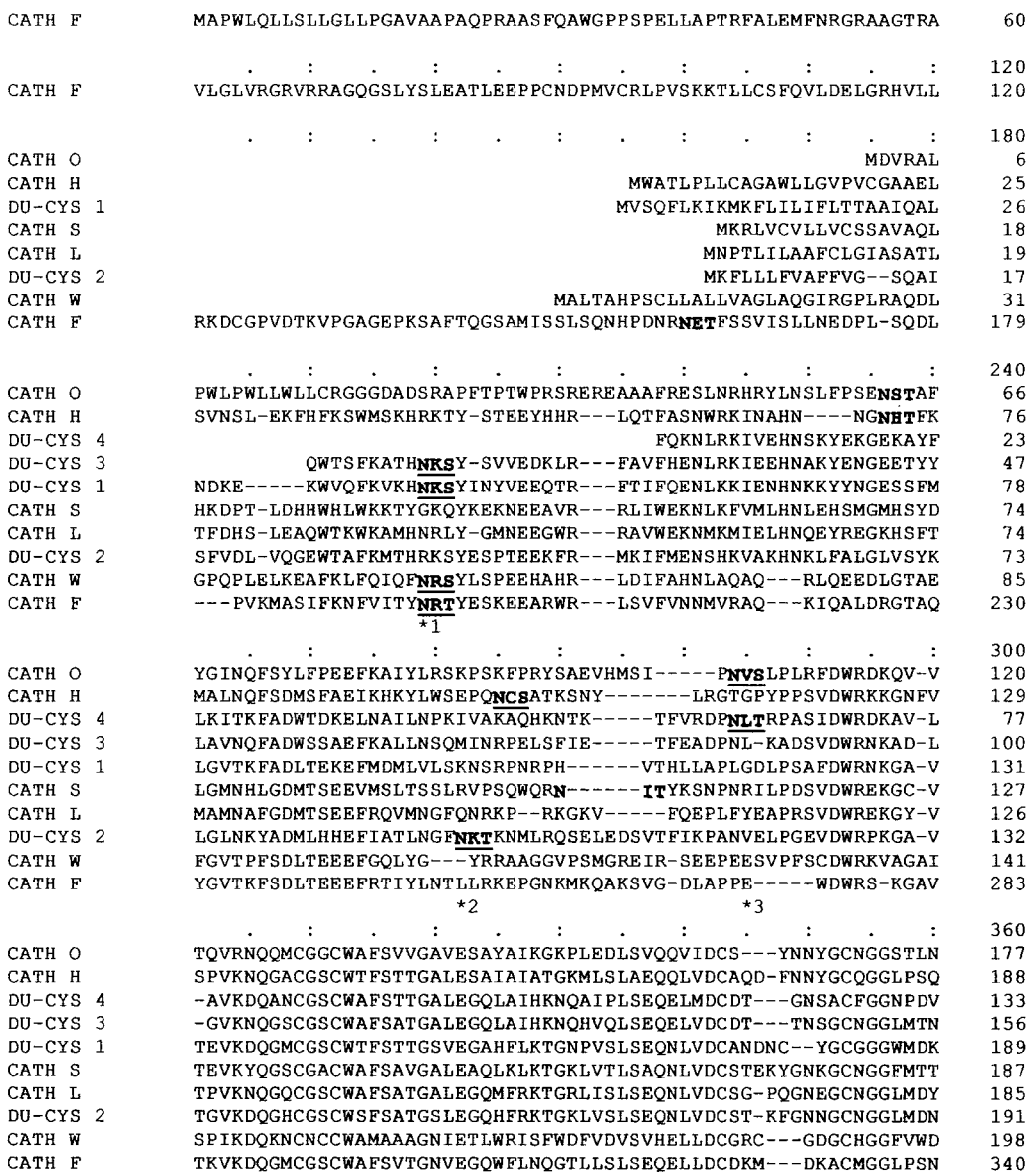


Figure 4.3g. Alignment of *D. undecimpunctata* cysteine proteinases Du-cys 1-4 with human cathepsins O, H, S, L, W and F. Potential *N*-glycosylation sites that show positional conservation in human cathepsins and Du-cys 1-4 are shown in bold and underlined, other potential *N*-glycosylation sites are shown in bold. Aligned sequences; cathepsin O (Velasco *et al.*, 1994), cathepsin H (Ritonja *et al.*, 1998), cathepsin (Shi *et al.*, 1994), cathepsin L (Machleidt *et al.*, 1986), cathepsin W (Wex *et al.*, 1998) and cathepsin F (Santamaria *et al.*, 1999).

Further alignment of the N-terminals of the Du-cys 1-4 sequences with human cathepsins F, O, H, S, L, W and F revealed homology in the positions of potential *N*-glycosylation site between Du-cys sequences and other human cathepsins (see figure 4.3g). Du-cys 1 and 3 share homologous potential *N*-glycosylation sites possessed by cathepsin F and W (*1), Du-cys 1 also possesses a site in close proximity (*4) to the C-terminal sites of cathepsin L and F. Du-cys 2 and 4 possess N-terminal potential *N*-glycosylation sites with in close proximity or at homologous positions to cathepsin H and O respectively.

The reason for the conservation of these potential *N*-glycosylation sites by human and invertebrate cysteine proteinases is unclear and the role of these sites in lysosomal transport or another pro-peptide function requires further research. However, due to similar pro-peptide tertiary structure and conserved potential N-glycosylation positions the mammalian cathepsin L and Du-cys cathepsin L-like cysteine proteinase pro-peptide are predicted to perform similar roles *in vivo*.

Du-cys 2 shares a potential *N*-glycosylation site (*2) with a *Drosophila* cysteine proteinase predominately expressed in the midgut (Matsumoto *et al.*, 1995) and *Sitophilus* cysteine proteinase (77 % amino acid identity with Du-cys 2) expressed in gastric caeca and other organs (not aligned see figure 4.2.2d, Matsumoto *et al.*, 1997). By *N*-glycosylation site and primary sequence homology Du-cys 2 is predicted to exhibit similar predominant expression profile in the alimentary tract as is exhibited by the *Sitophilus* and *Drosophila* cysteine proteinases.

Due to presence, heterogeneity of *N*-glycosylation sites and lack of invertebrate cysteine proteinase expression data in the literature no conclusions can be satisfactorily drawn regarding individual Du-cys proteinase transport to either secretory or lysosomal

pathways. A single clone of each Du-cys sequence was obtained in the library screening procedure and therefore no conclusions can be made regarding mRNA population frequency of Du-cys 1-4 in the alimentary tract and thereby relative contributions to dietary digestion.

Expression in alimentary tract cells can only be deduced at this stage by isolation from alimentary tract material. The relative contributions of Du-cys 1-4 to dietary digestion will depend not only on transcription regulation but also spatial induction within alimentary tract organs and the role of specific potential *N*-glycosylation sites in promoting lysosomal transport. Transcription regulation of Du-cys 1-4 in the alimentary tract and transport be it to lysosomes or the secretory pathway remains to be determined experimentally.

Platelet derived growth factor (PDGF) specifically up regulates a mouse cathepsin L mRNA transcription, cellular secretion occurs due to saturation of lysosomal transport receptors (Prence *et al*, 1990). The midgut epithelium of *Tenebrio molitor* (Coleoptera) undergoes a four-day renewal cycle, stem cells differentiating and replacing old cells (Billingsley and Lehane, 1996). Growth stimulated alimentary tract stem cells may exhibit up regulation of Du-cys enzymes and subsequent secretion due to saturation of lysosomal transport receptors as is found with mouse cathepsin L. A primary role of cellular recycling rather than dietary digestion however is predicted for Du-cys enzymes regulated in this manner.

Cleavage of pro-peptides will be a significant factor in the localisation of the Du-cys enzymes due to possession of pro-peptide *N*-glycosylation sites. Autocatalytic activation of *Sitophilus* cathepsin L-like cysteine proteinases has been shown to be pH dependent, cleavage occurring at acidic pH (Matsumoto *et al.*, 1998). Du-asp 1 type

aspartic proteinases may hydrolytically cleave the cathepsin L-like cysteine proteinases pro-peptides by a similar mechanism as cathepsin D proteolysis cleaves the propeptide of rat cathepsin B activating this cysteine proteinase (Rowan *et al.*, 1992).

Du-cys 1-4 pro-peptide cleavage prior to ER exit is considered unlikely as the pro-peptide of mouse cathepsin L is required for folding and ER exit (Tao *et al.*, 1994) and the pro-peptide of Du-cys 1 shows amino acid and structural homology to mammalian cathepsin L pro-peptides (see figure 4.3e). Cleavage of Du-cys 1-4 pro-peptides by autolysis or other proteolytic activity would remove the only potential *N*-glycosylation site of Du-cys 3 and the pro-peptide potential *N*-glycosylation sites of Du-cys 1,2 and 4. If cleavage occurs post ER exit and prior to recognition by lysosomal transport factors increased secretion and performance of dietary digestion would be predicted.

Chapter 5. Isolation and expression of aspartic and serine proteinase inhibitors (ASPI) from potato.

5.0. Introduction.

Biochemical identification of cysteine and aspartic proteinase activity in alimentary tract extracts by pepstatin A and E-64 inhibition led to the molecular isolation and characterisation of cathepsin L-like and cathepsin D-like proteinases from the *D. undecimpunctata* alimentary tract cDNA library. Plant derived proteinase inhibitors were identified that would possess inhibitory activity against the *D. undecimpunctata* cathepsin L and cathepsin D-like proteinases. A recombinant form of the oryzacystatin I gene had already been shown to be an effective inhibitor of *D. undecimpunctata* cysteine proteinases (Edmonds *et al.*, 1996).

The 20-24 kDa family of inhibitor proteins identified in potato inhibit cathepsin D-like aspartic proteinase activity, but significantly not pepsin or cathepsin E proteolytic activity (Keilova and Tomasek, 1976 and 1977). Members of this protein family also inhibit trypsin, chymotrypsin, subtilisin and yeast proteinase A (Keilova and Tomasek, 1977; Suh *et al.*, 1991; Walsh and Twitchell, 1991; Cater *et al.*, 2002). The potato 20-24 kDa family of proteinase inhibitors were selected for isolation and bacterial expression as potential inhibitors of the *D. undecimpunctata* alimentary tract cathepsin D-like and putative serine proteinase hydrolytic activities.

APIPO:
a.a. sequence. MMKCLFL (21-mer)
ntd. sequence. 5′-ATGATGAAGTGTTTATTTTTG-3′

```
            [1                                                    50]
CLONE 4     MMKCLFLLCL CLVPIVVFSS TFTSQNLIDL PSESPLPKPV LDTNGKELNP
            ---------->
            [51                                                  100]
CLONE 4     NSSYRIISIG RGALGGDVYL GKSPNSDAPC PDGVFRYNSD VGPSGTPVRF

            [101                                                 150]
CLONE 4     IPLSGGIFED QLLNIQFNIP TVRLCVSYTI WKVGINAYLR TMLLETGGTI

            [151                                                 200]
CLONE 4     GQADSSYFKI VKSSILGYNL LYCPITRPIL CPFCRDDDFC AKVGVVIQKG

                        220]
CLONE 4     KRRLALVNEN PLDVNFKEV. // T7 promoter
                                  <----------
```

APIPO-T7 = 660 + 3' UTR, polyA tail + 60 pBluescript® phagemid = 720 bp +

Figure 5.1a. Amino acid of aspartic/serine proteinase inhibitor (CLONE 4) isolated from potato (Hildmann *et al.*, 1992) and amino acid/nucleotide sequence of APIPO PCR primer. Expected PCR product size is shown below. Residues in bold indicate primer positions, arrowheads indicates direction of primed synthesis. Double vertical lines indicate sequence not shown including 3' UTR, poly-A tail and flanking region of the pBluescript® phagemid vector (Stratagene) prior to the T7 promoter primer site.

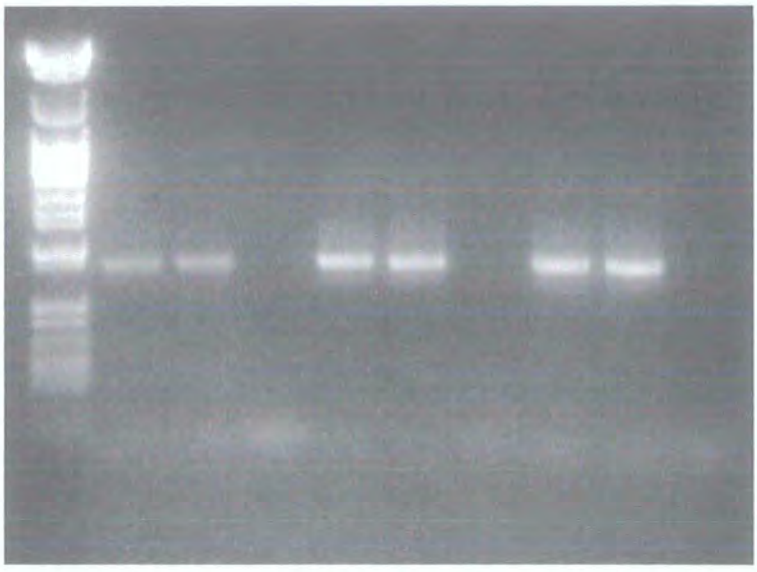


Figure 5.1b. Agarose gel electrophoresis of PCR reactions using a potato cDNA library and the primers APIPO and T7 to isolate aspartic proteinase inhibitors. ASPI# refers to reaction number, cont are no template controls. Lane guide (left to right): M) λDNA/Eco471(AvaII) size marker, 2) ASPI 1, 2) ASPI 1, 3) ASPI 1 control, 4) ASPI 2, 5) ASPI 2 , 6) ASPI 2 control , 7) ASPI 3, 8) ASPI 3, 9) ASPI 3 control.

```
ASPI 1: 1    MMKCLFLLCLCLFPIVVFSSTFTSQNPINLPSDATPVLDVTGKELDPRLSYHIISTFWGA 60
ASPI 3: 1    MMKCLFLLCLCLFPIVVFSSTFTSQNPINLPSDATPVLDVTGKELDPRLSYHIISTFWGA 60

ASPI 2: 1    MMKCLFLLCLCLVPIVVFSSTFTSKNPISLPSDATPVLDVAGKELDSRLSYRIISTFWGA 60

Homologous: MMKCLFLLCLCL PIVVFSSTFTS+NPI+LPSDATPVLDV GKELD RLSY IISTFWGA
             [      Leader     ][pro-pept][          Mature chain


ASPI 1: 61   LGGDVYLGKSPNSDAPCANGIFRYNSDVGPSGTPVRFIGSSSHFGQGIFENELLNIQFAI 120
ASPI 3: 61   LGGDVYLGKSPNSDAPCANGIFRYNSDVGPSGTPVRFIGSSSHFGQGIFENELLNIQFAI 120

ASPI 2: 61   LGGDVYLGKSPNSDAPCPDGVFRYNSDVGPSGTPVRFIPLSG----GIFEDQLLNIQFNI 116

Homologous: LGGDVYLGKSPNSDAPC +G+FRYNSDVGPSGTPVRFI  S     GIFE++LLNIQF I
                              1-------------------------------------------


ASPI 1: 121  STSKLCVSYTIWKVGDYDASLGTMLLETGGTIGQADSSWFKIVKSSQFGYNLLYCPVTST 180
ASPI 3: 121  STSKLCVSYTIWKVGDYDASLGTMLLETGGTIGQADSSWFKIVQSSQFGYNLLYCPVTST 180
                                                                 B-----
ASPI 2: 117  ATVKLCVSYTIWKVGNLNAYFRTMLLETGGTIGQADSSYFKIVKLSNFGYNLLYCPITPP 176

Homologous: +T KLCVSYTIWKVG+ +A   TMLLETGGTIGQADSS+FKIV+ S FGYNLLYCP+T
             -----1                                                B-----


ASPI 1: 181  MSCPFSPDDQFCLKVGVVHQNGKRRLALVKDNPLDVSFKQVQ 222
ASPI 3: 181  MSCPFSSDDQFCLKVGVVHQNGKRRLALVKDNPLDVSFKQVQ 222
             --B--------B
ASPI 2: 177  FLCPFCRDDNFCAKVGVVIQNGKRRLALVNENPLDVLFQEV  217
             --B--B-----B
Homologous:    CPF  DD FC KVGVV QNGKRRLALV +NPLDV F++V
                                                    ]
```

ASPI 1 vs ASPI 2	Identities = 176/221 (79 %), Positives = 190/221 (85 %), Gaps = 4/221 (1 %)
ASPI 1 vs ASPI 3	Identities = 220/222 (99 %), Positives = 221/222 (99 %), Gaps = 0/222 (0 %)
ASPI 2 vs ASPI 3	Identities = 175/221 (79 %), Positives = 190/221 (85 %), Gaps = 4/221 (1 %)

Figure 5.1c. Annotated predicted amino acid sequence of ASPI 1-3 and relative sequence homologies Alignment and homologies calculated by BLAST 2 server (Tatusova and Madden, 1999). Signal peptide identified by SignalP server (Nielsen *et al.*, 1997), mature protein chain positioned by previous identification (Ishikawa *et al.*, 1994a). Residues in bold and italics indicate a putative vacuolar targeting site similar to NPIRLP identified previously (Matsuoka and Nakamura, 1991; Chrispeels and Raikhel, 1992). Cysteine residues involved in disulfide bridge formation are indicated bold, linked by dashed lines with numbers denoting bond (Mares *et al*, 1989), cysteine residues in bold, linked and denoted by B indicate putative di-sulphide bridge confirmations.

```
ASPI 1 :   MKCLFLLCLCLFPIVVFSSTFTSQNPINLPSDATPVLDVTGKELDPRLSYHIISTFWGAL 61
           MKCLFLLCLCLFPIVVFSSTFTSQNPINLPSDATPVLDVTGKELDPRLSYHIISTFWGAL
Pot.AI :   MKCLFLLCLCLFPIVVFSSTFTSQNPINLPSDATPVLDVTGKELDPRLSYHIISTFWGAL 60

ASPI 1 :   GGDVYLGKSPNSDAPCANGIFRYNSDVGPSGTPVRFIGSSSHFGQGIFENELLNIQFAIS 121
           GGDVYLGKSPNSDAPCANGIFRYNSDVGPSGTPVRFIGSSSHFGQGIFENELLNIQFAIS
Pot.AI :   GGDVYLGKSPNSDAPCANGIFRYNSDVGPSGTPVRFIGSSSHFGQGIFENELLNIQFAIS 120

ASPI 1 :   TSKLCVSYTIWKVGDYDASLGTMLLETGGTIGQADSSWFKIVKSSQFGYNLLYCPVTSTM 181
           TSKLCVSYTIWKVGDYDASLGTMLLETGGTIGQADSSWFKIV+SSQFGYNLLYCPVTSTM
Pot.AI :   TSKLCVSYTIWKVGDYDASLGTMLLETGGTIGQADSSWFKIVQSSQFGYNLLYCPVTSTM 180
Pot.AIa:   TSKLCVSYTIWKVGDYDASLGTMLLETGGTIGQADSSWFKIVKSSQFGYNLLYCPVTSTM 180

ASPI 1 :   SCPFSPDDQFCLKVGVVHQNGKRRLALVKDNPLDVSFKQVQ 222
           SCPFS DDQFCLKVGVVHQNGKRRLALVKDNPLDVSFKQVQ
Pot.AI :   SCPFSSDDQFCLKVGVVHQNGKRRLALVKDNPLDVSFKQVQ 221
            C   P+DQFC KVGVV QNGKRRLALV +NPLDV+FK+V
Pot.AIb:   -CLRCPEDQFCAKVGVVIQNGKRRLALVNENPLDVNFKEV
```

Identities = 219/221 (99 %), Positives = 220/221 (99 %)
Score = 452 bits (1164), Expect = e-126

Figure 5.1d. Alignment search result for clone ASPI 1. The obtained sequence shows closest homology to an aspartic/serine proteinase inhibitor from potato (Pot.AI), (Banfalvi *et al.*, 1996). Two amino-acid differences can be seen (in bold and underlined) Pot.AIa (Identities = 215/221 (97 %), Positives = 216/221 (97 %) (Ishikawa *et al.*, 1994a) is a further aspartic proteinase inhibitor sequence from potato showing the lysine as opposed to the glutamine in position 154. Pot.AIb (Identities = 137/186 (73 %), Positives = 152/186 (81 %), Gaps = 5/186 (2 %) (Girard *et al.*, unpublished) is a sequence from black nightshade showing the proline residue at position 187 instead of the serine residue more commonly found. Alignment performed by blastp (www.ncbi.nlm.nih.gov/BLAST/).

```
ASPI 3 :   MKCLFLLCLCLFPIVVFSSTFTSQNPINLPSDATPVLDVTGKELDPRLSYHIISTFWGAL 61
           MKCLFLLCLCLFPIVVFSSTFTSQNPINLPSDATPVLDVTGKELDPRLSYHIISTFWGAL
Pot.AI :   MKCLFLLCLCLFPIVVFSSTFTSQNPINLPSDATPVLDVTGKELDPRLSYHIISTFWGAL 60

ASPI 3 :   GGDVYLGKSPNSDAPCANGIFRYNSDVGPSGTPVRFIGSSSHFGQGIFENELLNIQFAIS 121
           GGDVYLGKSPNSDAPCANGIFRYNSDVGPSGTPVRFIGSSSHFGQGIFENELLNIQFAIS
Pot.AI :   GGDVYLGKSPNSDAPCANGIFRYNSDVGPSGTPVRFIGSSSHFGQGIFENELLNIQFAIS 120

ASPI 3 :   TSKLCVSYTIWKVGDYDASLGTMLLETGGTIGQADSSWFKIVQSSQFGYNLLYCPVTSTM 181
           TSKLCVSYTIWKVGDYDASLGTMLLETGGTIGQADSSWFKIVQSSQFGYNLLYCPVTSTM
Pot.AI :   TSKLCVSYTIWKVGDYDASLGTMLLETGGTIGQADSSWFKIVQSSQFGYNLLYCPVTSTM 180

ASPI 3 :   SCPFSSDDQFCLKVGVVHQNGKRRLALVKDNPLDVSFKQVQ 222
           SCPFSSDDQFCLKVGVVHQNGKRRLALVKDNPLDVSFKQVQ
Pot.AI :   SCPFSSDDQFCLKVGVVHQNGKRRLALVKDNPLDVSFKQVQ 221
```

Identities = 221/221 (100 %), Positives = 221/221 (100 %)
Score = 456 bits (1173), Expect = e-127

Figure 5.1e. Alignment search result for clone ASPI 3. The obtained sequence shows identity to an aspartic and serine proteinase inhibitor from potato (Banfalvi *et al.*, 1996). Alignment performed by blastp (www.ncbi.nlm.nih.gov/BLAST/).

```
ASPI 2 : 1   MMKCLFLLCLCLVPIVVFSSTFTSKNPISLPSDAT---PVLDVAGKELDSRLSYRIISTF 57
             MMKCLFLLCLCL+PIVVFSSTFTS+N I LPS++     PVLD  GKEL+   SYRIIS
P.In   : 1   MMKCLFLLCLCLLPIVVFSSTFTSQNLIDLPSESPLPKPVLDTNGKELNPDSSYRIISIG 60

ASPI 2 : 58  WGALGGDVYLGKSPNSDAPCPDGVFRYNSDVGPSGTPVRFIPLSGGIFEDQLLNIQFNIA 117
              GALGGDVYLGKSPNSDAPCPDGVFRYNSDVGPSGTPVRFIPLSGGIFEDQLLNIQFNI
P.In   : 61  RGALGGDVYLGKSPNSDAPCPDGVFRYNSDVGPSGTPVRFIPLSGGIFEDQLLNIQFNIP 120

ASPI 2 :118  TVKLCVSYTIWKVGNLNAYFRTMLLETGGTIGQADSSYFKIVKLSNFGYNLLYCPITPPF 177
             TVKLCVSYTIWKVGNLNAYFRTMLLETGGTIGQADSSYFKIVKLSNFGYNLLYCPITPPF
P.In   :121  TVKLCVSYTIWKVGNLNAYFRTMLLETGGTIGQADSSYFKIVKLSNFGYNLLYCPITPPF 180

ASPI 2 :178  LCPFCRDDNFCAKVGVVIQNGKRRLALVNENPLDVLFQEV 217
             LCPFCRDDNFCAKVGVVIQNGKRRLALVNENPLDVLFQEV
P.In   :181  LCPFCRDDNFCAKVGVVIQNGKRRLALVNENPLDVLFQEV 220
```

Identities = 200/220 (90 %), Positives = 205/220 (92 %), Gaps = 3/220 (1 %)
Score = 404 bits (1037), Expect = e-112

Figure 5.1f. Alignment search result for clone ASPI 2. The obtained sequence shows closest homology to an aspartic and serine proteinase inhibitor from potato (P.In) (Strukelj *et al.*, 1990). Underlined residues are those which are not homologous between the two sequences. Alignment performed by blastp (www.ncbi.nlm.nih.gov/BLAST/).

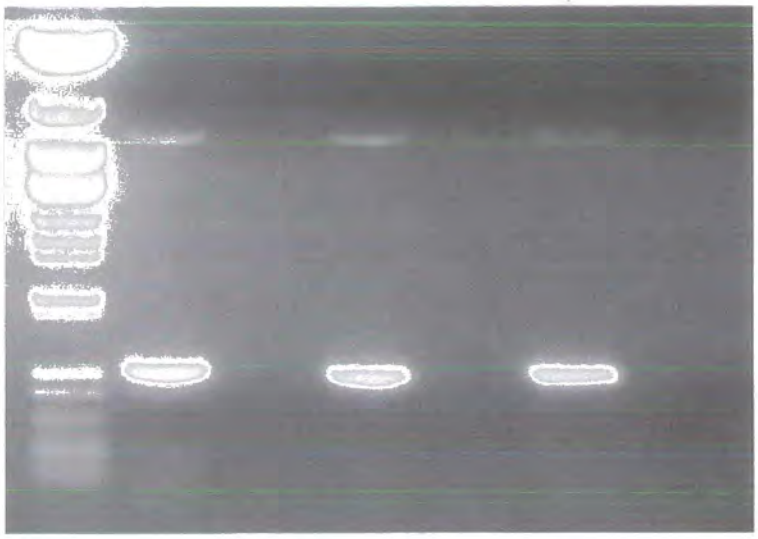
```
ASPI 2: 2   MKCLFLLCLCLVPIVVFSSTFTSKNPISLPSDATPVLDVAGKELDSRLSYRIISTFWGAL 61
            MKCLFLLCLCLVPIVVFSSTFTSKNPI+LPSDATPVLDVAGKELDSRLSYRIISTFWGAL
PIG    : 1  MKCLFLLCLCLVPIVVFSSTFTSKNPINLPSDATPVLDVAGKELDSRLSYRIISTFWGAL 60

ASPI 2: 62  GGDVYLGKSPNSDAPCPDGVFRYNSDVGPSGTPVRFIPLSGGIFEDQLLNIQFNIATVKL 121
            GGDVYLGKSPNSDAPC +G+FRYNSDVGPSGTPVRF     GIFE++LLNIQF I+T KL
PIG    : 61 GGDVYLGKSPNSDAPCANGIFRYNSDVGPSGTPVRFSHFGQGIFENELLNIQFAISTSKL 120
```

Identities = 178/216 (82 %), Positives = 191/216 (88 %)
Length = 217 Score = 330 bits (846), Expect = 1e-89

```
ASPI 2: 58   WGALGGDVYLGKSPNSDAPCPDGVFRYNSDVGPSGTPVRFIPLSGGIFEDQLLNIQFNIA 117
              GALGGDVYLGKSPNSD PCPDGVFRYNSDVGPSGT VRFIPLSGGIFEDQLLNIQFNIA
Cd.in : 61   RGALGGDVYLGKSPNSDGPCPDGVFRYNSDVGPSGTFVRFIPLSGGIFEDQLLNIQFNIA 120

ASPI 2: 118  TVKLCVSYTIWKVGNLNAYFRTMLLETGGTIGQADSSYFKIVKLSNFGYNLLYCPITPPF 177
             TVKLCVSYTIWKVGNLNAYFRTMLLETGGTIGQADSSYFKIVKLSNFGYNLLYCPITPPF
Cd.in : 121  TVKLCVSYTIWKVGNLNAYFRTMLLETGGTIGQADSSYFKIVKLSNFGYNLLYCPITPPF 180
```

Identities = 199/220 (90 %), Positives = 204/220 (92 %),
Length = 220 Score = 360 bits (924), Expect = 9e-99

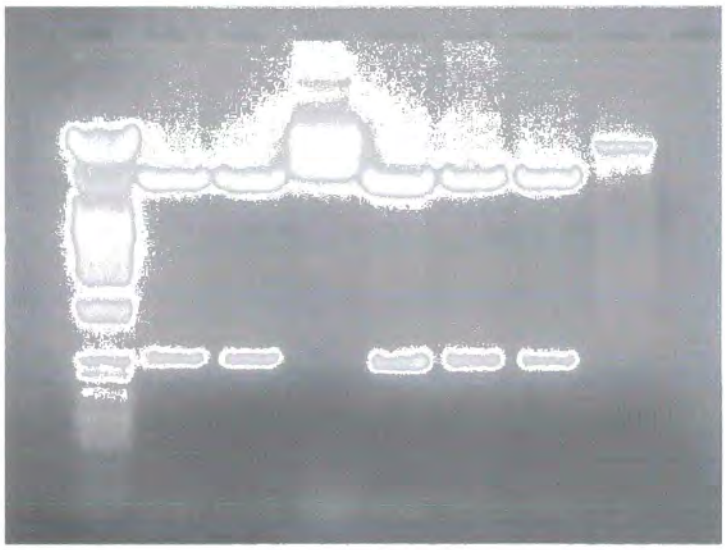
Figure 5.1g. Further alignment of ASPI 2 with two further aspartic proteinase inhibitors (PIG) (Maganja *et al.*, 1992) and (Cd.in) (Strukelj *et al.*, 1992) isolated from potato. Amino-acid residues differing from the closest matching sequence (Figure 3.4.1e) are shown to be present in the above inhibitor sequences, all variable residues are found in the second (PIG) or third (Cd.in) sequences except for Ser_{29} which is typically Asp_{29} or Asn_{29}). Residues underlined and in bold do not show identity to other ASPI sequences. Residues in bold do not show identity to ASPI 2 in the top alignment but show homology in the second alignment displayed. Alignment performed by blastp (www.ncbi.nlm.nih.gov/BLAST/).



M 1 2 3 4 5 6

5' ASPI 1+2+3 GCCG**CA^TATG**TCAACTTTCACTTCCMAAAATCC
3' ASPI 1+3 CGGC**G^GATCC**TGGACTTGCTTGAAGGAGAC
3' ASPI 2 CGGC**G^GATCC**ACTTCCTGGAATAAGACATC

Figure 5.2a. 1 % agarose gel electrophoresis of Advantage 2 PCR reactions, amplifying ASPI sequences 1, 2 and 3 for ligation into pET 24a. Primer nucleotide sequences, Nde 1 (CA^TATG) and BamH1 (G^GATCC) restriction site positions are illustrated. ASPI # refers to sequence number, controls are template minus. Lane guide (left to right): M) λDNA/Eco471 (AvaI1) size marker, 1) ASPI1, 2) ASPI1 control, 3) ASPI2, 4) ASPI 2 control, 5) ASPI3, 6) ASPI 3 control.



M 1 2 3 4 5 6 7

Figure 5.2b. 1 % agarose gel analysis of Nde 1/ BamH1 restriction digested pCR®-Blunt II-TOPO ASPI clones and pET 24a. Restricted ASPI sequences migrating at approximately 600 bp, linearised pCR®-Blunt II-TOPO migrating at approximately 3.6 kb and linearised pET 24a migrating according to the marker larger than it's 5.3 kb actual length. Lane guide (left to right): M) λDNA/Eco471 (AvaI1) size marker, 2) ASPI1, 3) ASPI1a, 3) ASPI2, 4) ASPI 2a, 5) ASPI3, 6) ASPI 3a, 7) pET24a.

Figure 5.2c. 1 % agarose gel analysis of Nde 1/ Xho1 restriction digested pET 24a plasmids containing rASPI constructs in BL21 (DE3) expression cells. rASPI sequences migrating at approximately 600 bp and linearised pET 24a migrating according to the marker larger than it's 5.3 kb actual length. Lane guide (left to right): Lane guide (left to right): M) λDNA/Eco471 (AvaII) size marker, 1) rASPI 1, 2) rASPI 1, 3) rASPI 2, 4) rASPI 2, 5) rASPI 3, 6) rASPI 3.

```
          1
rASPI 1:  PIVSG.QFPSRNNFV.L.EGDIHMSTFTSQNPINLPSDATPVLDVTGKELDPRLSYH
rASPI 3:  PIVSG.QFPSRNNFV.L.EGDIHMSTFTSKNPINLPSDATPVLDVTGKELDPRLSYH
rASPI 2:  --VSG.QFPSRNNFV.L.EGDIHMSTFTSKNPISLPSDATPVLDVAGKELDSRLSYR
          [       pET 24a          ][[ pro-pept ]][

rASPI 1:  IISTFWGALGGDVYLGKSPNSDAPCANGIFRYNSDVGPSGTPVRFIGSSSHFGQGIF
rASPI 3:  IISTFWGALGGDVYLGKSPNSDAPCANGIFRYNSDVGPSGTPVRFIGSSSHFGQGIF
rASPI 2:  IISTFWGALGGDVYLGKSPNSDAPCPDGVFRYNSDVGPSGTPVRFIPLSG----GIF
                              rASPI coding sequence

rASPI 1:  ENELLNIQFAISTSKLCVSYTIWKVGDYDASLGTMLLETGGTIGQADSSWFKIVKSS
rASPI 3:  ENELLNIQFAISTSKLCVSYTIWKVGDYDASLGTMLLETGGTIGQADSSWFKIVQSS
rASPI 2:  EDQLLNIQFNIATVKLCVSYTIWKVGNLNAYFRTMLLETGGTIGQADSSYFKIVKLS


rASPI 1:  QFGYNLLYCPVTSTMSCPFSPDDQFCLKVGVVHQNGKRRLALVKDNPLDVSFKQVQD
rASPI 3:  QFGYNLLYCPVTSTMSCPFSSDDQFCLKVGAVHQNGKRRLALVKDNPLDVSFKQVQD
rASPI 2:  NFGYNLLYCPITPPFLCPFCRDDNFCAKVGVVIQNGKRRLALVNENPLDVLFQEV-D
                              248                                  ][
rASPI 1:  PNSSSVDKLAAALEHHHHHH.
rASPI 3:  PNSSSVDKLAGALEHHHHHH.
rASPI 2:  PNSSSVDKLAAALEHHHHHH.
            pet 24a          ]
```

Figure 5.2d. Sequence of rASPI constructs in pET 24a. Protein translation start site is in bold, underlined and indicated by a pointed arrow, the stop codon is indicated by a closed diamond. The histidine tag employed in the subsequent purification is shown in bold. Alignment performed by MAP (Huang, 1994). ASPI 3 PCR generated mutation Ala_{197} is indicated in bold and underlined.

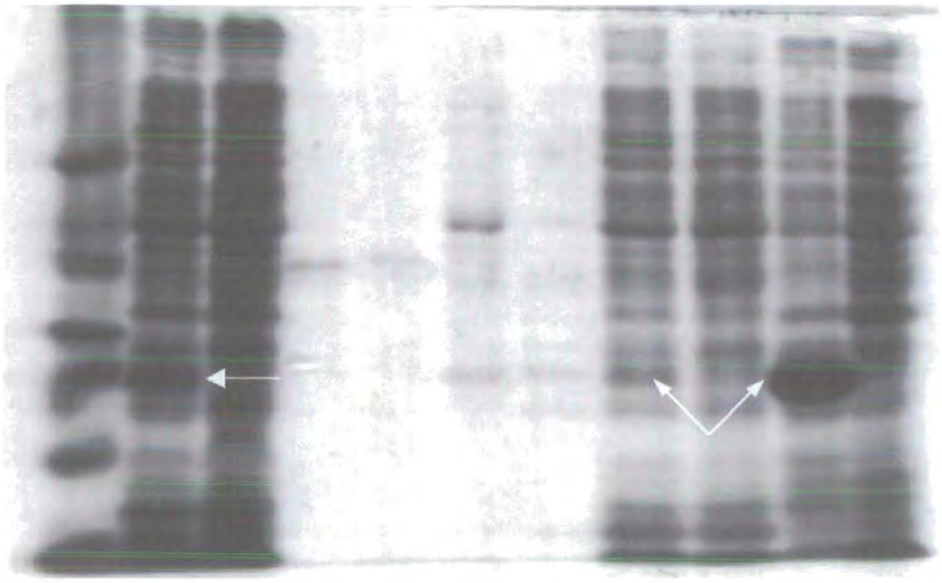


Figure 5.3a. 15 % SDS-PAGE analysis of recombinant rASPI 1 expression in BL21 (DE3) cells. Lanes denoted (I) are cultures induced by addition of 1 mM IPTG, Lanes denoted (U) = not induced. White arrows pointing to TCP (induced), insoluble whole cell extract (induced), and soluble whole cell extract (induced) indicate a protein band corresponding to the expected size of rASPI 1 (24 kd).

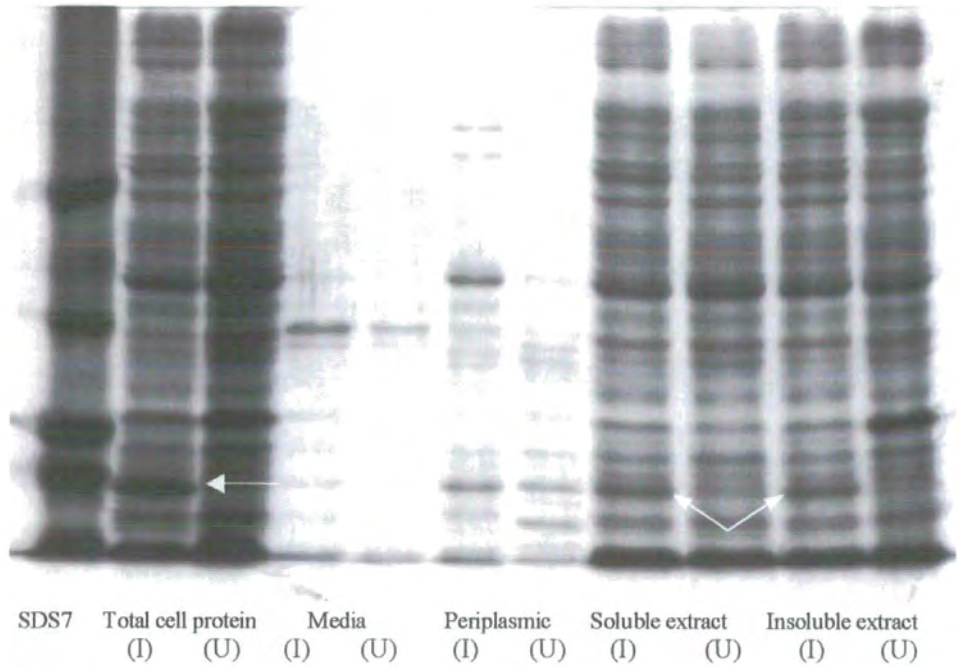


Figure 5.3b. 12.5 % SDS-PAGE analysis of recombinant rASPI 2 expression in BL21 (DE3) cells. Lanes denoted (I) indicate cultures induced by addition of 1 mM IPTG, lanes denoted (U) = non-induced culture. White arrows pointing to TCP (induced), insoluble whole cell extract (induced) and soluble whole cell extract (induced) indicate a protein band corresponding to the expected size of rASPI 2, (24 kd).

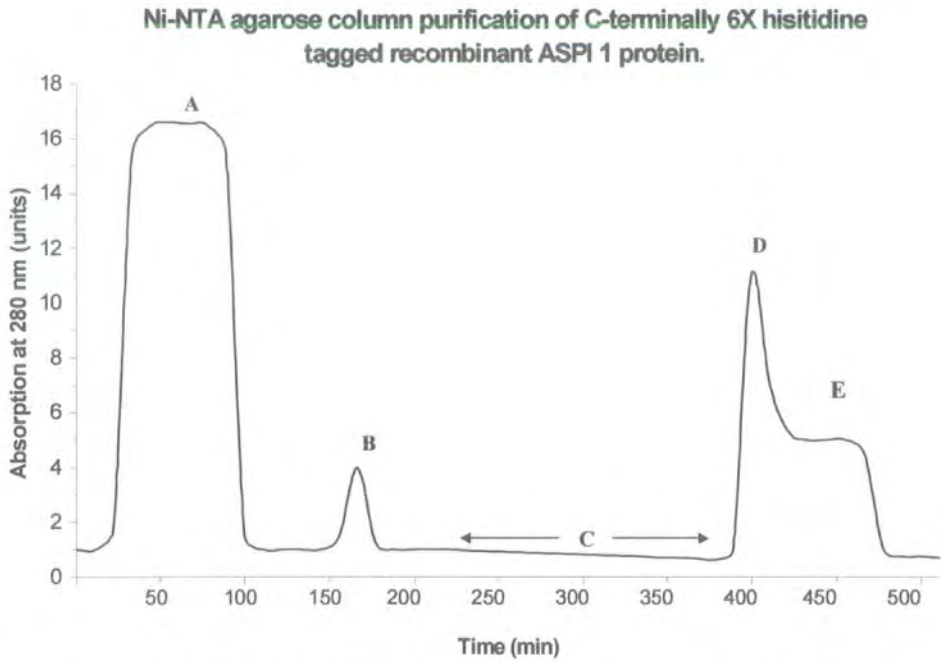


Figure 5.4a. Elution profile of Ni-NTA purified rASPI proteins. Feature guide; A: loading of bacterial lysate in buffer B, B: protein eluted during the buffer C wash, C: renaturing of bound protein sample with a 6 M to 1 M urea gradient in renaturing buffer, D: elution of protein sample by addition of 250 mM imidazole to 1 M renaturing buffer, E: continuous peak due to imidazole absorption at 280 nm.

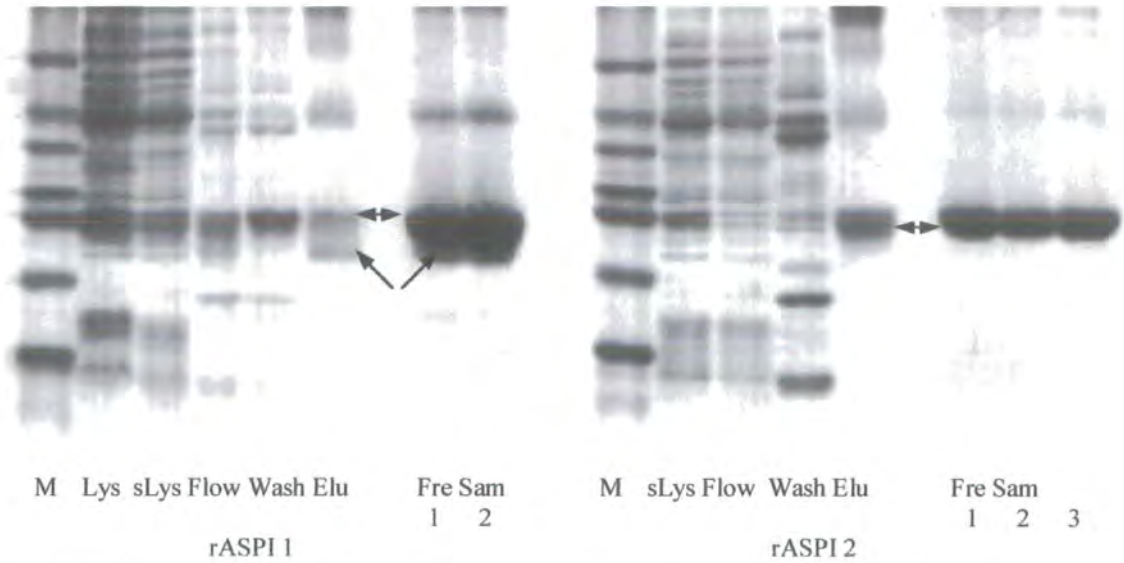


Figure 5.4b. 15 % SDS-PAGE analysis of large-scale (1 litre) purification of rASPI 1 and rASPI 2. Capitals in brackets indicate the corresponding absorption peak in figure 5.4a Lane guide (left to right): M) SDS 7 marker, Lys) bacterial lysate, sLys) spun bacterial lysate, Flow) flow-through (A), Wash) buffer C wash (B), Elu) eluted peak (D), Fre Sam) freeze dried sample (numbers indicate individual preparations). Double arrows indicate purified rASPI 1 and 2 respectively, single arrows indicate a second smaller version of the rASPI 1 protein at approximately 22.5 kDa. The ASPI 2 wash lane is from a different purification of ASPI 2 performed with fresh Ni-NTA agarose column.

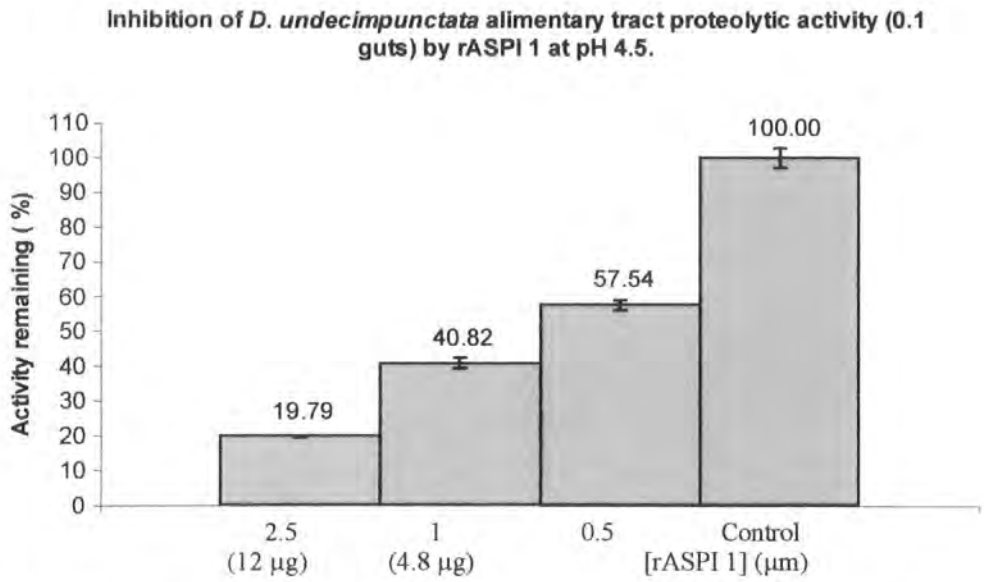


Figure 5.5a. Inhibition of *D. undecimpunctata* alimentary tract proteolytic activity by rASPI 1 (pH 4.5). Each point represents a minimum of 2 readings, vertical bars indicate standard error. Bracketed numbers indicate inhibitor protein content in micrograms per assay.

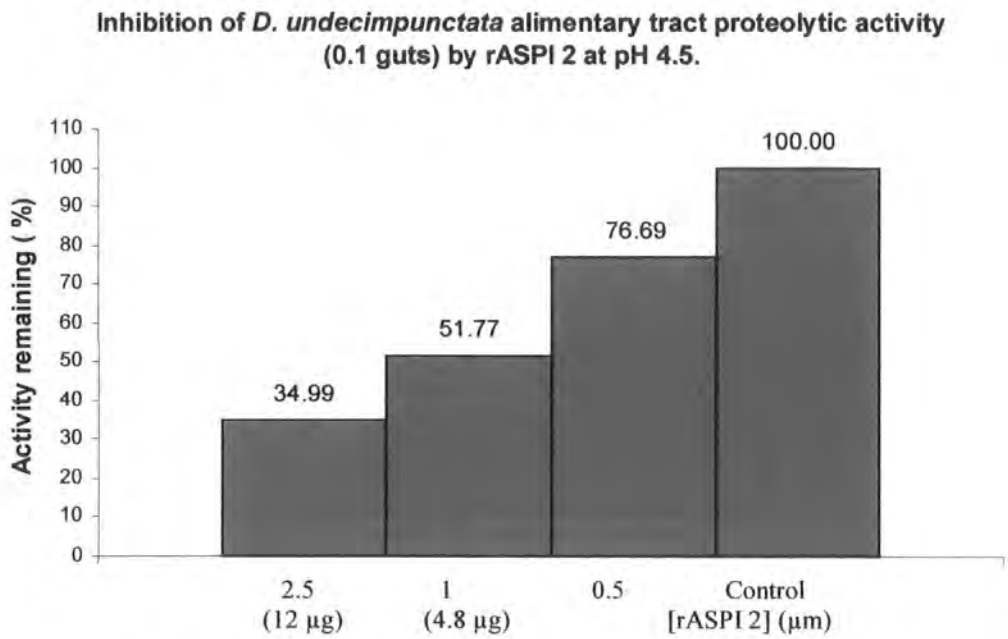


Figure 5.5b. Inhibition of *D. undecimpunctata* alimentary tract proteolytic activity by rASPI 2 (pH 4.5). Bracketed numbers indicate inhibitor protein content in micrograms per assay.

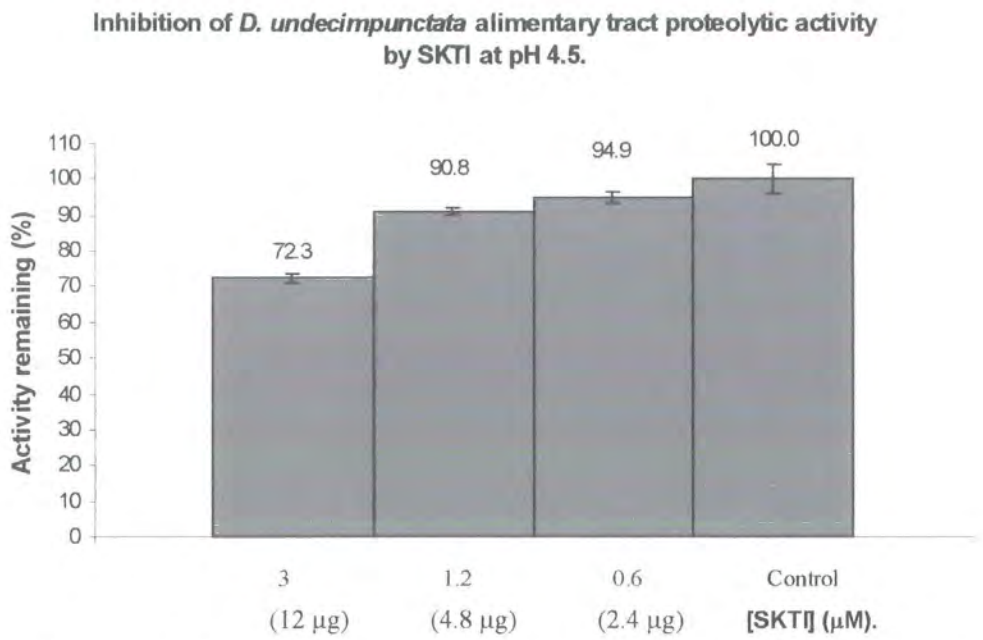


Figure 5.5c. Inhibition of *D. undecimpunctata* alimentary tract proteolytic activity by SKTI (pH 4.5). Each point represents a minimum of 2 readings, vertical bars indicate standard error. Bracketed numbers indicate inhibitor protein content in micrograms per assay.

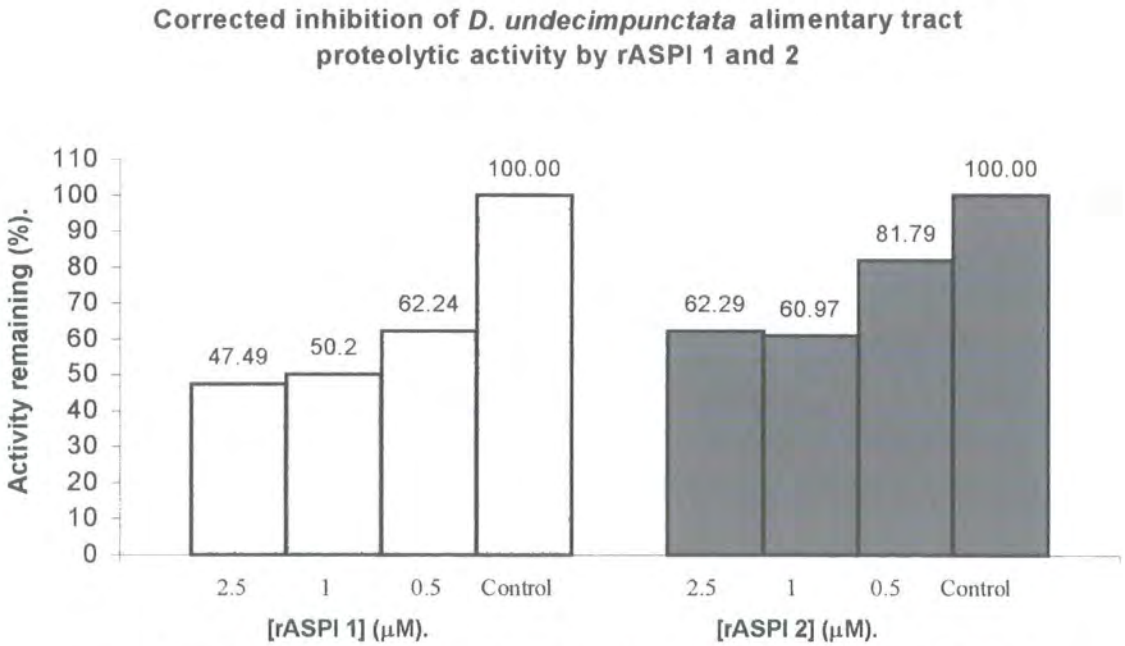


Figure 5.5d. Corrected Inhibition of *D. undecimpunctata* alimentary tract proteolytic activity by rASPI 1 and rASPI 2. Figures represent addition of non-specific inhibition observed in SKTI assays to the inhibitory activity observed in rASPI 1 and 2 assays respectively. Corrected figures remove the effects of non-specific inhibition representing only specific inhibition of alimentary tract proteolysis by rASPI 1 and 2.

5.1. Results: Molecular isolation and identification of potato aspartic and serine proteinase inhibitors.

The Entrez (http://www.ncbi.nlm.nih.gov/entrez/) database was searched for aspartic proteinase inhibitors and sequences were aligned in Megalign (DNASTAR) (see Appendix A.6.). A 5' primer (APIPO) was designed to the translation start codon of a previously identified potato aspartic proteinase inhibitor (Hildmann *et al.*, 1992), (see figure 5.1a). APIPO in conjunction with a T7 promoter primer site on the pBluescript® phagemid vector PCR amplified full-length clones of aspartic proteinase inhibitors from a 'Désiré' potato tuber λZap-cDNA® library (Stratagene) provided by D. Bown.

A standard *Taq* polymerase PCR was performed in triplicate, with duplicate positives for each of the triplicates on the potato tuber cDNA library employing 4 µl of cDNA library as template and 17 thermal cycles. Fragments of the expected size 0.9 kb were amplified in each positive reaction (see figure 5.1b.). PCR products were TOPO TA cloned, plasmids from positive colonies were isolated and EcoR1 restriction digested to verify insert size. One plasmid from each cloning was sequenced using the M13 forward and reverse primers on the pCR®2.1-TOPO®cloning vector.

Sequenced clones were identified by BLAST search and termed aspartic and serine proteinase inhibitor (ASPI) 1, ASPI 2 and ASPI 3 by homology to known cathepsin D, trypsin and chymotrypsin inhibitors from potato (Keilova and Tomasek, 1976 and 1977; Mares *et al.*, 1989; Strukelj *et al.*, 1990 and 1992, Walsh and Twitchell, 1991; Suh, 1991). ASPI 1 encodes a complete coding sequence including leader sequence, pro-peptide, mature protein and translation stop codon (see figure 5.1c). ASPI

1 showed 99 % amino acid identity to Pot.AI (Banfalvi *et al.*, 1996) a PKPI-B family member (Ishikawa *et al.*, 1994a), possessing only two divergent residues Lys_{164} and Pro_{186} (see figure 5.1d).

Further alignment of ASPI 1 with related inhibitor proteins Pot.AIa (Ishikawa *et al.*, 1994a) and Pot.AIb (Girard *et al.*, unpublished) showed homologous Lys_{164} and Pro_{186} were possessed by these sequences (see figure 5.1d). It is plausible that Lys_{164} and Pro_{186} represent genuine differences between ASPI 1 and Pot.AI rather than being PCR generated mutations. ASPI 1 coding sequence encodes a 222 amino acid protein with a molecular weight of 24.2 kDa, pI of pH 6.26 and GRAVY value of 0.119, prior to post-translational modifications.

ASPI 2 encodes a hydrophobic leader sequence, pro-peptide, mature chain, stop codon (see figure 5.1c) and showed closest homology to a PKPI-A family member (Ishikawa *et al.*, 1994a) possessing 90 % amino acid identity with P.In isolated from potato (see figure 5.1f) (Strukelj *et al.*, 1990). ASPI 2 encodes a 217 amino acid protein with a molecular weight of 23.9 kDa, pI of pH 6.26 and GRAVY value of 0.284 prior to post-translational modifications. The N-terminal residues of the mature ASPI 2 protein Pro_{27} to $Tryp_{58}$ are divergent from other PKPI-A sequences and most closely resembles the N-terminal sequence found in PKPI-B sequences such as PIG (Maganja *et al.*, 1992) and Cd.In (Strukelj *et al.*, 1992) (see figure 5.1f and g). ASPI 2 from Gly_{58} to the stop codon most closely resembles the sequences of PKPI-A family member P.In (Strukelj *et al.*, 1990) (see figure 5.1f).

The ASPI 2 N-terminal sequence divergence from other PKPI-A family members maybe the result of a gene splicing event or a PCR generated anomaly. Annealing of an

incomplete PKPI-B PCR product to PKPI-A template would result in a similar gene fusion. The unresolved Ser_{29} amino-acid mutation in ASPI 2 which is typically Asp_{29} or Asn_{29} in other PKPI sequences is discussed later. If the observed ASPI 2 sequence hybrid is a natural phenomenon then further investigation would be warranted. Studies in respect to inhibitory capabilities and expression localisation within potato plants may reveal differential expression compared to that observed for other PKPI-A sequences.

ASPI 3 shows 100 % identity amino acid identity with Pot.AI (Banfalvi *et al.*, 1996) (see figure 5.1e) and encodes the entire translated sequence including leader sequence, pro-peptide, mature chain and stop codon (see figure 5.1c). Prior to post-translational modifications ASPI 3 encodes a 222 amino acid protein with a molecular weight of 24.3 kDa, pI of pH 5.85 and a GRAVY value of 0.125.

The putative di-sulphide bridges present in rASPI proteins were identified by cysteine residue positional homology to soybean trypsin inhibitor (STI) known disulphide bond forming cysteine pairs (Mares *et al.*, 1989). ASPI 1 and 3 possess the conserved di-sulphide bridge forming $Cys_{77\text{-}126}$ similar to STI and putative bridge/s between $Cys_{175\text{-}183}$, $Cys_{175\text{-}192}$ and/or $Cys_{183\text{-}192}$. ASPI 2 possesses almost homologous conserved residues $Cys_{77\text{-}122}$ and possible di-sulphide bridge forming cysteine pairs between $Cys_{171\text{-}179}$, $Cys_{171\text{-}182}$, $Cys_{171\text{-}188}$, $Cys_{179\text{-}182}$, $Cys_{179\text{-}188}$, and $Cys_{182\text{-}188}$.

5.2. Results: Construction of ASPI-pET24a expression constructs.

The leader sequence of each ASPI was determined by employing SignalP 2.0 (Nielsen *et al.*, 1997). PCR primers 5' ASPI 1+2+3, 3' ASPI 1+3 and 3'ASPI 2 (see figure 5.2a)

were designed respectively to the ASPI 1, 2 and 3 pro-peptide start and coding sequence stop codons. Nde 1 and BamH1 restriction sites were included in the 5' and 3' primers for ligation into the pET 24a expression vector (see appendix A.1. for pET 24a vector and MCS map). A standard Advantage 2 PCR was performed employing 20 thermal cycles. Amplification products were silica fines purified and cloned in TOPO-Blunt, and restriction digested with Nde 1 and BamH1 to verify insert size (see figure 5.2b).

XL1-Blue cells previously transformed with the pET 24a plasmid (provided by A. P. Fordham-Skelton) were cultured overnight and plasmid isolated and restriction digested with Nde 1 and BamH1 (see figure 5.2b). Nde 1 and BamH1 restricted ASPI 1-3 and pET 24a expression vector were gel purified, ligated at room temperature overnight and transformed into TOP10 cells and plated on LB-Kan. Plasmid samples were digested with Nde 1 and Xho 1 (30 µl of plasmid sample, 10 units Nde 1 and Xho 1, 4 µl Buffer D, 4 µl d.H_2O) and analysed by 1 % agarose gel electrophoresis to verify insert size.

Chemically competent BL21 (DE3) cells were transformed with the ASPI-pET 24a constructs and the reactions plated on LB-Kan. BL21 (DE3) cells were cultured and plasmid purified. Restriction digest with Nde 1 and Xho 1 confirmed the transformation of the rASPI-pET 24a constructs into BL21 (DE3) (see figure 5.2c) which were re-termed rASPI 1, rASPI 2 and rASPI 3 respectively. Insert sequencing within the pET 24a vector was performed to assess insert sequence integrity post Advantage 2 PCR amplification. Sequence analysis of the ASPI 1 and 2 showed no PCR mutations were incorporated in the amplification step, the ASPI 3 clone showed a Val_{197} to Ala_{197} non-

conservative mutation in the amino-acid sequence occurred in the amplification reaction (see figure 5.2d) and was therefore discontinued at this stage.

5.3. Results: Expression analysis of rASPI 1 and 2 recombinant proteins.

To verify construct transcription by bacteriophage T7 RNA polymerase under control of the lac operator test recombinant rASPI 1 and 2 expressions were performed using Novagen recommended protocols (pET system manual, 2000), inducing rASPI 1 and 2 cultures at 37°C. Cellular fraction samples were mixed 1:1 with 2× SB buffer and SDS-PAGE analysed (see figures 5.3a and b). Induced rASPI 1 and 2 test cultures produced an additional protein when induced with 1 mM IPTG, this protein was not present in the non-induced cultures. The additional protein band, in each case was approximately the predicted size for the expression products of rASPI 1 (24.48 kDa) and rASPI 2 (24.16 kDa) respectively.

The expression analysis revealed the majority of rASPI 1 and 2 proteins were present in the soluble and insoluble cytoplasmic fractions of induced cultures, no determinable quantity being visible in the media or periplasmic space fractions. Approximately 95 % of rASPI 1 target protein was retained in the cytoplasm as insoluble aggregate whereas rASPI 2 recombinant protein was distributed at a ratio of 50:50 between both soluble and insoluble cytoplasmic fractions.

Growth, induction and preparation of cellular fractions were identical for both rASPI constructs. SDS-PAGE analysis of recombinant rASPI protein expression revealed differing yields and partitioning characteristics into the soluble and insoluble

cellular fraction. rASPI 1 construct produced a visibly greater yield of recombinant protein than rASPI 2 (see figures 5.3a and b). Relative yields of target protein can be determined by numerous factors such as individual host bacteria and plasmid copy number.

Protein partitioning into the soluble and insoluble cellular fractions is affected by salt concentration in the lysis buffer, expression rate and target sequence hydropathicity (pET system manual (TB055), 2000). Lower induction temperatures reduce target protein expression rate and insoluble aggregate formation (Schein and Noteborn, 1989). Protparam calculated Grand average of Hydropathicity (GRAVY) values for rASPI 1 and 2 are –0.181 and –0.32 respectively. GRAVY values indicate that rASPI 2 is more likely to form insoluble aggregates upon expression than rASPI 1, the opposite is observed in the expression analysis. Formation of insoluble aggregates upon expression is therefore sequence hydropathicity independent. Increased aggregation is therefore most likely due to a higher expression rate of rASPI 1 relative to rASPI 2, the higher yielding construct forming a greater percentage of insoluble aggregates within the cytoplasm.

5.4. Results: Large scale expression and purification of rASPI 1 and 2 proteins.

Large scale protein expression was performed maintaining the expression analysis growth temperature of 37°C, cultures were induced with 1 mM IPTG for a period of 4 hours. Maintaining a higher growth temperature and shortened induction time maximised the yield of insoluble protein preventing interactions with expression host

proteins such as aspartic proteinases, subtilisin (serine proteinase) and cysteine proteinases. Interactions that could possibly leading to protein degradation or inactivation.

Figure 5.4a shows the U.V. absorption profile at 280 nm over the Ni-NTA chromatography process. Peak A represents protein flowthrough from the denatured bacterial lysate, peak B represents the removal of non specifically bound proteins to the matrix by the wash step in buffer C. The timespan denoted by C represents the 6 M to 1 M urea refolding gradient. Peak D represents elution of His tagged protein, continuous peak E is due to imidazole present in the elution buffer absorption at 280 nm.

Throughout the purification process samples were taken and mixed for subsequent analysis by SDS-PAGE to monitor purification of the rASPI 1 and 2 proteins (see figures 5.4b). The SDS-PAGE analysis of rASPI 1 and 2 purification prepared under denaturing conditions show protein samples from lysate (Lys), lysate centrifuged at 10 000g for 30 min (sLys), flow-through of non-bound protein (Flow), protein eluted with a wash in buffer C (Wash), purified protein eluted by addition of 250 mM imidazole (Elu) and multiple freeze-dried samples that were dialysed in d.H_2O and stored at 4°C (Fre Sam).

Freeze-dried samples of both rASPI proteins show a high level purification up to approximately 98 % of the eluted sample being the target protein (see figure 5.4b). Yields from 1 litre cultures varied with individual cultures but were typically between 17-23 mg/litre for rASPI 1 and 11-15 mg/litre for rASPI 2. The variation in yield between two such similar proteins may be linked to the relative solubility of rASPI 1 and 2 in the expression host cytoplasm. The relative yields, if linked to the solubility may

indicate interaction of refolded rASPI proteins in the periplasmic space with endogenous expression host systems. A putative interaction of active rASPI proteins having a deleterious effect on the expression host cells and reducing yield.

Visible in the SDS-analysis in the eluted and all freeze dried samples of the large scale purification of rASPI 1 is a slightly smaller protein band (approximately 22.5 kDa) relative to the main purified protein band (24 kDa) (figure 5.4b), indicated by single headed arrows). This band may represent a bacterially cleaved form of the rASPI 1 protein or a post transcriptionally modified version of the mRNA signal. The frequency of cleavage also matches approximately the rASPI 1 ratio of soluble to insoluble protein observed in the SDS-PAGE expression analysis. The site of cleavage would appear to be located on the N-terminal of the protein, as His tagged purification was successful.

The proposed explanation is the BL21 (DE3) cells are recognising the plant mature protein cleavage signal and processing rASPI 1 to the mature form of the protein (see figure 5.1c and 5.2d). Cleavage of the rASPI construct pro-peptide sequence Met_1-Asn_{11} which (molecular weight of 1239.3 kDa) would leave a smaller mature protein with an N-terminal Leu_1 (Ishikawa *et al.*, 1994a) and a molecular weight of 23.2 kDa. The predicted processed ASPI 1 protein is approximately 700 kDa larger than the 22.5 kDa protein observed in the SDS-PAGE, it is the possible that ASPI 1 maybe cleaved at a residue C-terminally distal to the predicted mature chain start Leu_1.

A similar smaller band is not present in the rASPI 2 freeze dried or eluted samples indicating cleavage does not occur with this protein or mRNA signal. The lack of an apparently cleaved form of ASPI 2 indicates that the cleavage site rASPI 1 resides in the N-terminal in an area of rASPI 1 and 2 amino acid code divergence. The

unresolved typically Asn_{29} or Asp_{29} in other PKPI family sequences to Ser_{29} mutation in ASPI 2 referred to in figure 5.1g is proposed to prevent the bacterial system from recognising and cleaving the pro-peptide to generate a mature version of the rASPI 2 protein with an N-terminal Leu_1. Further possible cleavage sites in the N-terminal at amino acids residues differing between rASPI 1 and 2 have been discounted due to cleavage product size heterogeneity to the observed 22.5 kDa processed rASPI 1 protein.

The freeze dried preparations of rASPI 1 and 2 were purposely overloaded for SDS-PAGE analysis to detect any contaminating proteins A protein band of approximately 48-50 kDa is visible this may be a dimer version of the respective ASPI proteins. The band however might also be a bacterial protein with a histidine rich motif. Further analysis is required to verify which of these hypotheses is supported. The level of dimerisation or a contaminating bacterial protein is approximately 5 % when compared to target protein present as a monomer.

5.5. Results: Inhibitory activity evaluation of rASPI 1 and 2 proteins.

Gene transcription of PKPI isoforms such as Clone 4 are induced in planta by the accumulation of jasmonic and abscisic acids in the cellular environment (Hildmann *et al.*, 1992). rASPI protein refolding was performed *in vitro* at room temperature by addition of protein sample at 50 μM to 0.1 M acetic acid, 0.1 M acetic acid pH 4.5 having been found to be relatively ineffective in promoting refolding. Under these *in vitro* conditions oxidation of the cysteine residues to form the rASPI protein di-sulphide bridges completed the refolding process.

Initial tests of refolded rASPI 1 and 2 proteins for inhibitory activity against the aspartic proteinase activity present in larval *D. undecimpunctata* alimentary tract extracts were performed with 0.1 gut equivalents with a control hydrolytic activity of 0.114 rfu/min ± 0.016. Control SKTI assays were performed to provide a negative control for which no or limited inhibition would be expected at pH 4.5. Serine proteinases exhibit optimal proteolytic activity at alkaline pH and would not be expected to be catalytically active at pH 4.5.

Addition of rASPI 1 and 2 proteins at 2.5-0.5 μM resulted in marked inhibition of the proteolytic activity of *D. undecimpunctata* alimentary tract extracts (see figures 5.5a and b). The inhibition levels achieved were greater than expected based on previous research employing pepstatin A (Edmonds *et al.*, 1996). The SKTI assays also resulted in a reduction of proteolytic activity (see figure 5.5c), the inhibition observed was directly proportional to the amount of additional protein added. The SKTI assays were performed with equal μg:μg quantities of inhibitor added as the rASPI assays to allow the non-specific inhibition to be quantitated.

The inhibition observed in the SKTI assays is proposed to be due to the additional protein content in the assays and is discussed fully in Chapter 6. The inhibition percentage calculated at each concentration of SKTI was added to the inhibition percentage observed in the rASPI 1 and 2 assays to produce corrected inhibition figures (see figure 5.5d). The corrected inhibition percentages represent the removal of the non-specific inhibitory effect of the additional protein content from the rASPI assays, representing only specific inhibition of cathepsin D-like proteolytic activity present in *D. undecimpunctata* alimentary extracts.

Maximal inhibition of approximately 50 % and 40 % was observed employing rASPI 1 and rASPI 2 respectively, lower maximal inhibition by rASPI 2 may represent incomplete inhibition of *D. undecimpunctata* alimentary tract aspartic proteinases by rASPI 2. Maximal inhibition by rASPI 1 and 2 was approximately 18 % and 9 % greater than inhibition levels achieved with pepstatin A in Chapter 3 (4 μM pepstatin A/0.05 gut equivalents) and 16 % and 6 % greater than observed in previous research with 9.72 μM pepstatin A (Edmonds *et al.*, 1996). The lower inhibition levels obtained in the previous research is believed to be due to incomplete inhibition of the aspartic proteinase activity present in *D. undecimpunctata* alimentary tracts with the concentration of pepstatin A employed.

Amended figures for rASPI 1 showed inhibition of approximately 38 % of extract proteolytic activity at 0.5 μM compared to 18 % inhibition with rASPI 2. Maximum inhibition was achieved in this work at 1 μM, a higher concentration than the maximum inhibition of 50 % with 0.46 μM of native PKPI isoform protein extract (Wilhite *et al.*, 2000) lower rASPI activity maybe due to numerous factors. The artificial refolding strategy is sensitive to the concentration of rASPI protein added, attempting refolding at lower concentrations resulted in higher inhibition due to reduced aggregate formation (see Chapter 6). Relative inhibition may also be affected by the differing assay systems employed and the inhibitor isoforms present in the previous research. The aspartic proteinases present in midguts of alfalfa weevil and *D. undecimpunctata* may also exhibit differential sensitivity to inhibition by PKPI proteins.

The assay data presented in this section demonstrates that biologically active PKPI family proteins have been successfully expressed purified and artificially refolded.

The closest available assay comparison was performed with native PKPI isoforms isolated from potato assayed at pH 5.0 against alfalfa weevil gut extracts, 51 % inhibition of azocasein digestion was observed with a PKPI isoform concentration of 0.46 μM (Wilhite *et al*, 2000). The observed inhibition in this work at pH 4.5 of 38 % at a 0.5 μM concentration of ASPI 1 raises the possibility that the entire protein compliment of the sample may not be natively folded.

5.6. Discussion.

This work represents the first reporting of the successful expression and purification of biologically active PKPI-A and PKPI-B protein isoforms. Previous to this work characterisation of the properties of this family of proteins was performed by protein isolation from tuber material (Keilova and Tomasek, 1976 and 1977, Walsh and Twitchell, 1991; Suh, 1991; Wilhite *et al.*, 2000, Cater *et al.*, 2002). Heterologous expression of individual members of this protein family is a major step forward allowing further characterisation at many levels of members of this protein family.

Successful refolding of rASPI proteins was performed in acetic acid an organic acid. PKPI gene transcription is well documented to be induced by increased levels of jasmonic and abscisic acid in the cellular environment. Jasmonic and abscisic acid, organic acid phytohormones may perform a dual role both in PKPI gene induction and creating a relatively oxidising cellular environment promoting di-sulphide bond formation and thereby refolding of the PKPI proteins. A dual role in both gene

transcription induction and translated protein refolding for wound response inducers is a concept that may be repeated in other similar systems.

Inhibition by a specific cathepsin D-like aspartic proteinase inhibitors provides evidence that the aspartic proteinase activity inhibited by pepstatin (Chapter 3) and the cathepsin D-like by sequence homology clones described in Chapter 4 will also show cathepsin D-like proteolytic specificity similar to that observed in *Musca domestica* (Lemos and Terra, 1991) and alfalfa weevil (Wilhite *et al.*, 2000).

During preparation of this manuscript Cater *et al* (2002) have reported the successful expression of a tomato proteinase inhibitor (TI) in *Pichia pastoris* using the pPICZαC expression vector. Evaluation of the inhibitory activity of the recombinant tomato protein revealed 9 fold lower K_i of inhibition of yeast proteinase A than human lysosomal cathepsin D at pH 4.7. Comparison of the tomato inhibitor with PKPI isolated from potato revealed the potato preparation possessed 52 fold and 150 fold lower K_i than recombinant TI against cathepsin D and yeast proteinase A respectively.

The respective K_i figures obtained indicate the potato extract contains more potent aspartic proteinase inhibitors than the recombinant TI. The TI and extracted native potato inhibitor proteins showed greater inhibitory activity against yeast proteinase A than human cathepsin D. This may indicate these inhibitor proteins may be part of a defence response to plant fungal pathogens such as *Botrytis cinerea* (Cater *et al.*, 2002). The possibility of an anti-fungal and insect defensive role for the PKPI proteins warrants further research to evaluate their ability to enhance crop resistance *in vivo* by a transgenic plant strategy.

Chapter 6. Biochemical characterisation of larval *D. undecimpunctata* cysteine and aspartic proteinase interactions with plant proteinase inhibitors.

6.0. Introduction.

Biochemical characterisation of the interactions between the cysteine and aspartic proteinase activities identified in *D. undecimpunctata* was continued employing plant derived proteinase inhibitors. The availability of inhibitor protein genes encoding effective proteinase inhibitors is a prerequisite to the production of transgenic material expressing proteinase inhibitors for *in vivo* evaluations of plant resistance to pests. The objective of this chapter is to further characterise the inhibition of *D. undecimpunctata* proteinases by oryzacystatin I and rASPI 1 to evaluate their potential as biopesticides in a transgenic crop strategy.

6.1. Results: Inhibition assays.

Differing gut extract samples were utilised for the assays presented in this section, although samples were prepared at the same time as those employed in section 3.1. Uninhibited rates of substrate hydrolysis by 0.0125 gut equivalents showed variation in the assays presented. Uninhibited control rates were 0.202 ± 0.01 rfu/min for the assays presented in figures 6.2a and 6.3b whereas uninhibited control rates for assays 6.2b, 6.3a, 6.4a and 6.5a were 0.129 ± 0.004 rfu/min. The variation in control uninhibited substrate hydrolysis represents a 27 % loss of proteolytic activity in alimentary tract extracts. The reduction in proteolytic activity is most likely due to experimental error

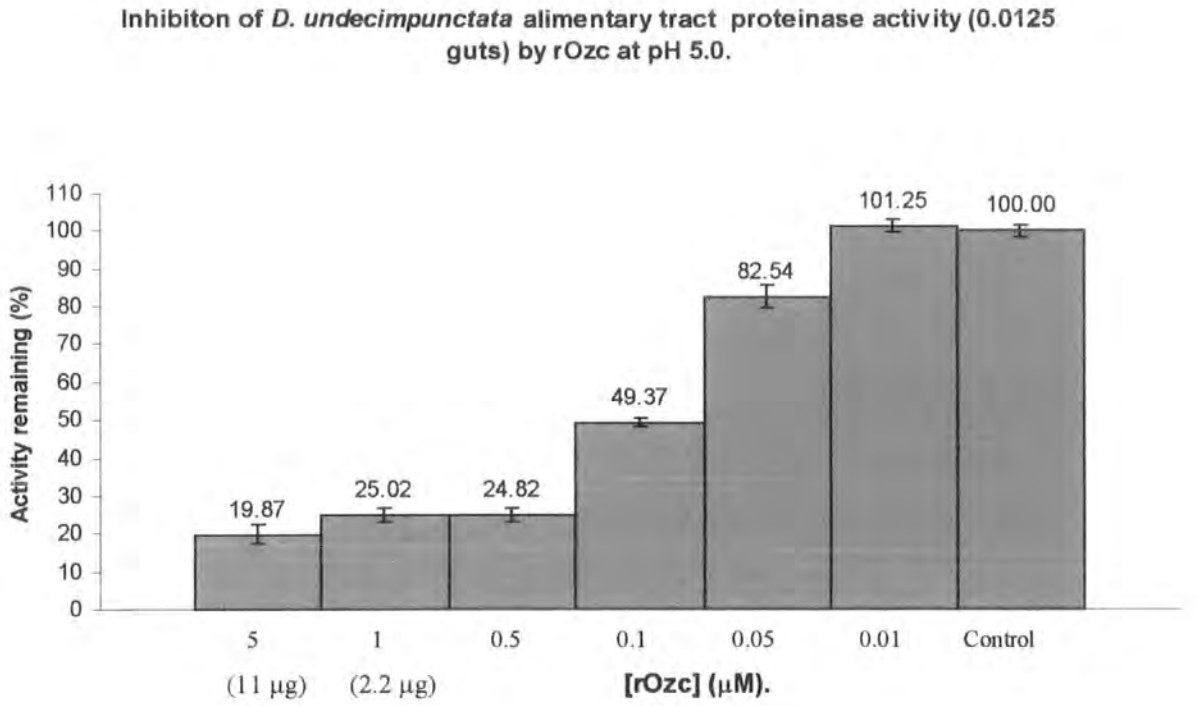


Figure 6.2a. Inhibition of *D. undecimpunctata* alimentary tract proteolytic activity by rOzc (pH 5.0). Columns represent a minimum of 2 readings, vertical bars indicate standard error. Bracketed numbers indicate inhibitor protein content per assay, controls are uninhibited assays.

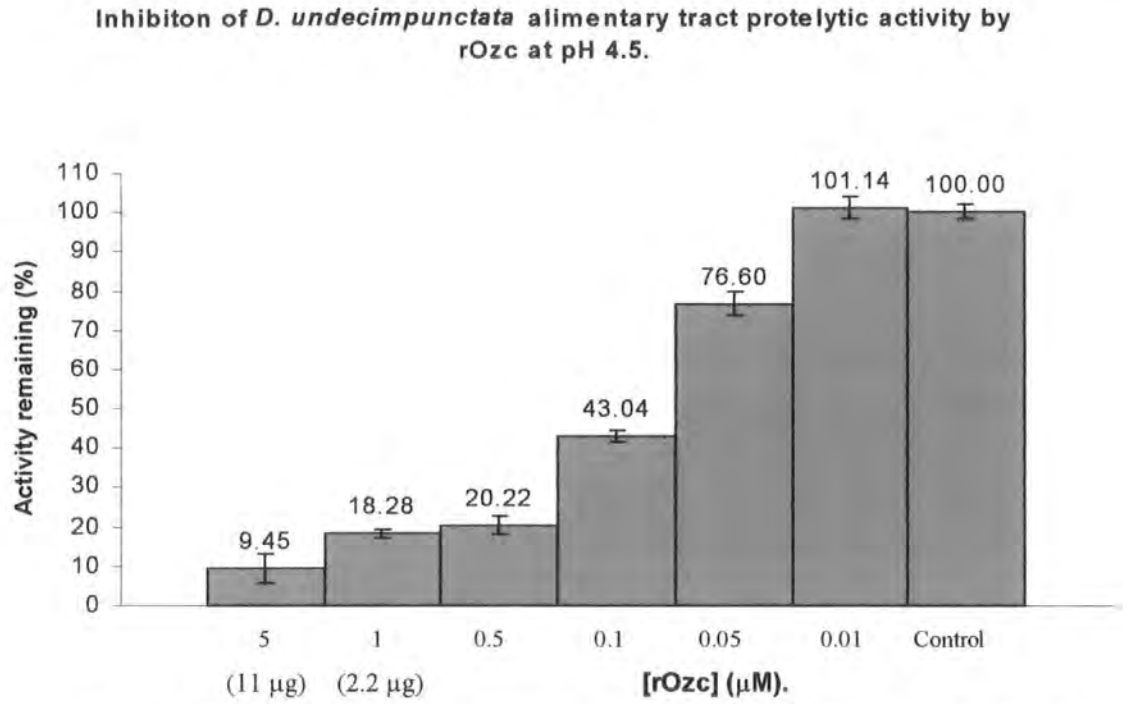


Figure 6.2b. Inhibition of *D. undecimpunctata* alimentary tract proteolytic activity by rOzc (pH 4.5). Columns represent a minimum of 2 readings, vertical bars indicate standard error. Bracketed numbers indicate inhibitor protein content per assay, controls are uninhibited assays.

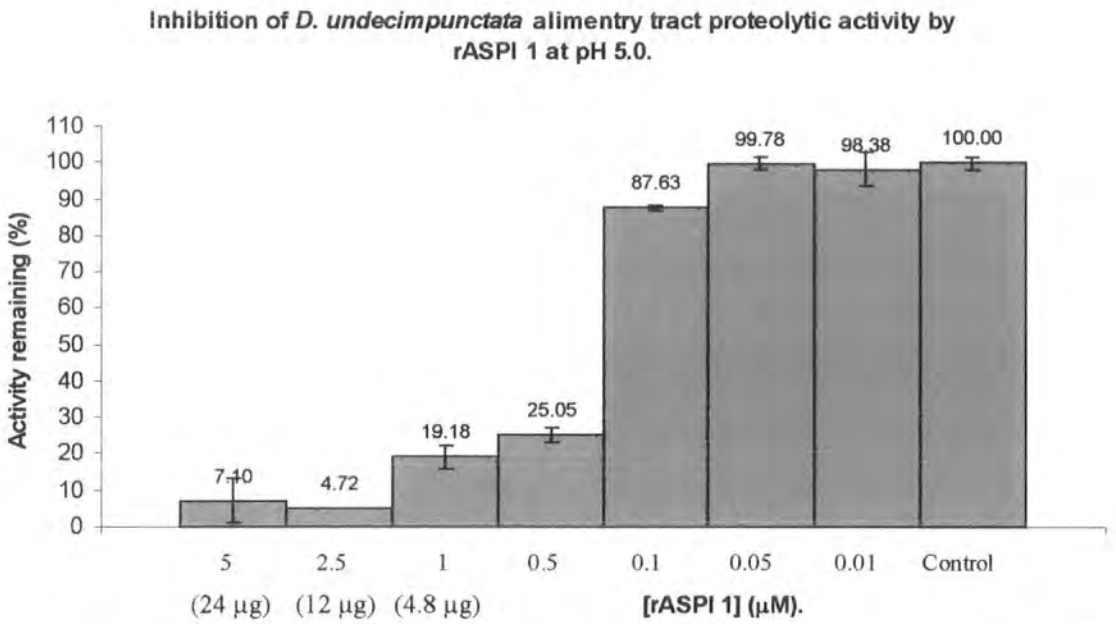


Figure 6.3a. Inhibition of *D. undecimpunctata* alimentary tract proteolytic activity by rASPI 1 (pH 5.0). Columns represent a minimum of 2 readings, vertical bars indicate standard error. Bracketed numbers indicate inhibitor protein content per assay.

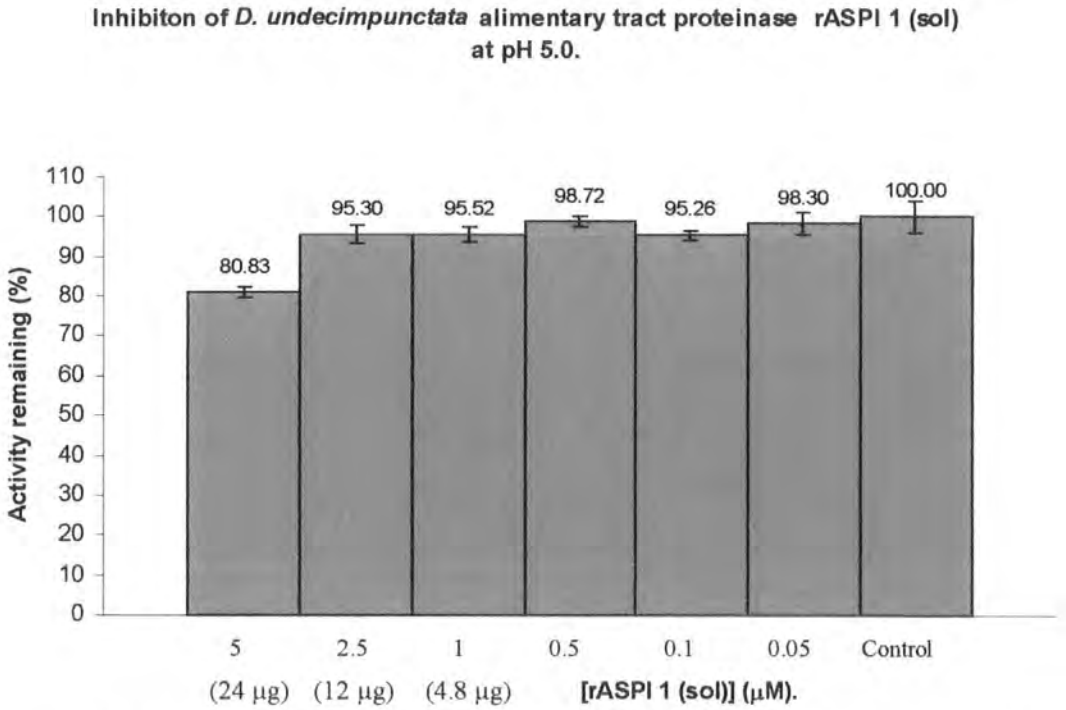


Figure 6.3b. Inhibition of *D. undecimpunctata* alimentary tract proteolytic activity by rASPI 1 (sol), (pH 5.0). Each point represents a minimum of 2 readings, vertical bars indicate standard error. Bracketed numbers indicate inhibitor protein content per assay.

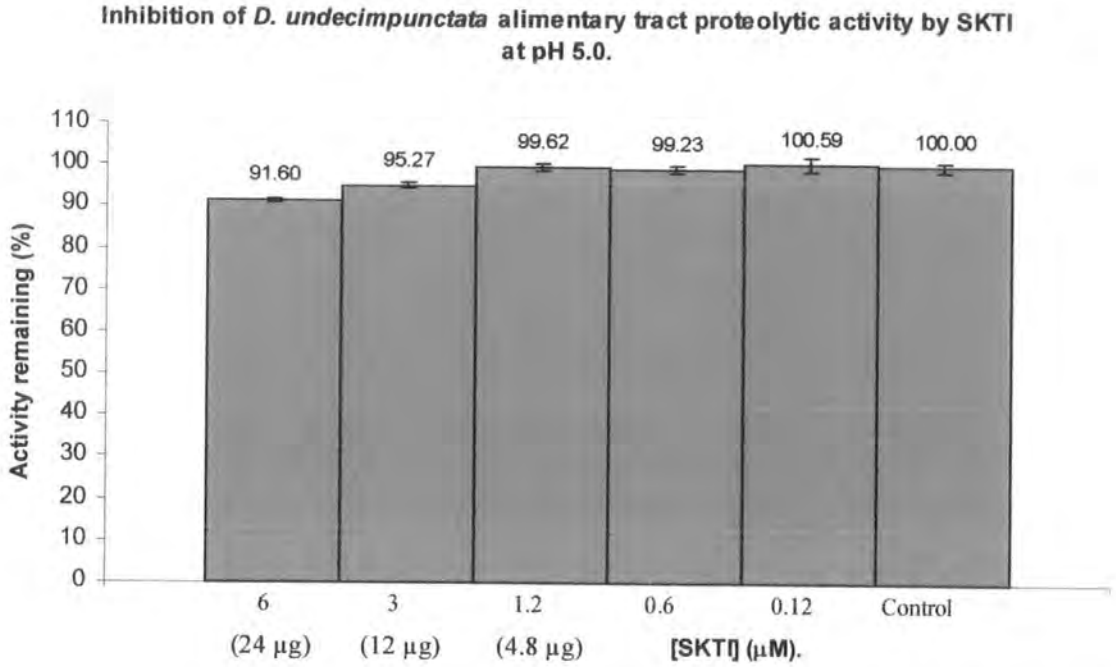


Figure 6.4a. Inhibition of *D. undecimpunctata* alimentary tract proteolytic activity SKTI, (pH 5.0). Each point represents a minimum of 2 readings, vertical bars indicate standard error. Bracketed numbers indicate inhibitor protein content per assay, controls are uninhibited assays.

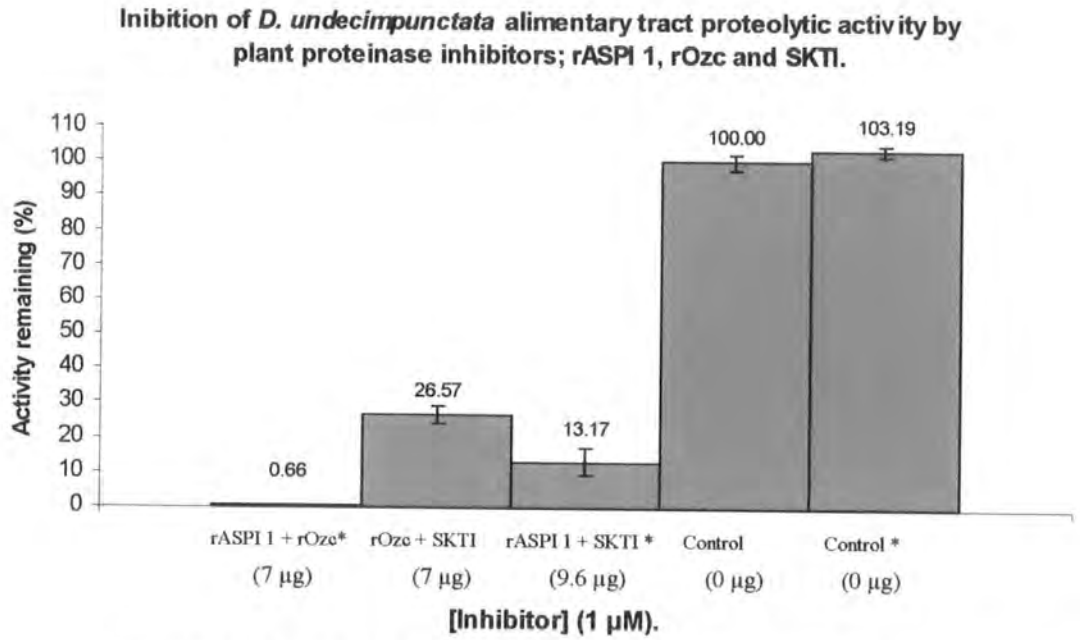


Figure 6.5a. Inhibition of *D. undecimpunctata* alimentary tract proteolytic activity by rASPI 1 rOzc and SKTI (pH 5.0). Each point represents a minimum of 2 readings, vertical bars indicate standard error. Bracketed numbers indicate inhibitor protein content per assay, controls are uninhibited assays.

resulting in proteinase inactivation.

Assays performed with >7 μg of inhibitor protein showed increased inhibition beyond the levels observed at lower microgram quantities when inhibition had reached a plateau (see figures 6.2-5.a-b). The increased levels of inhibitory activity observed with >7 μg of inhibitor protein were characterised as being inhibitor independent, as increased inhibition was observed with rOzc, rASPI, rASPI 1 (sol), and SKTI.

Figure 6.4a. illustrates the inhibition observed with SKTI a trypsin and to a lesser extent chymotrypsin and subtilisin inhibitor. At the pH conditions employed for the assay little or no serine proteinase activity would be expected, as is supported by inhibition not being observed at lower SKTI concentrations. Approximately 9 % inhibition was observed at 6 μM (24 μg protein per assay) and 5 % at 3 μM (12 μg per assay) no discernible inhibition was observed at 1.2 μM (4.8 μg per assay).

Figures 6.2a and b show rOzc inhibition of alimentary tract extracts, at a 5 μM (11 μg protein) concentration an additional 6 % inhibition is observed relative to a rOzc concentration of 1 μM (2.2 μg protein). rASPI (sol) also showed enhanced inhibition at higher protein concentrations, 18 % inhibition was observed at 5 μM concentration (24 μg of protein) and 3 % at 2.5 μM concentration (12 μg of protein). Additional inhibition was also observed at higher concentrations with rASPI following a similar pattern with a single anomaly at 2.5 μM.

Specific inhibition was unlikely to be the cause due to the effect being observed with both rOzc, rASPI (sol), rASPI and SKTI. The most plausible explanation for this observation is the addition of inhibitor at levels greater than 2.5-5 μM, approximately 10 μg per assay results in non-specific inhibition due to the additional protein content in

each assay. Non-specific inhibition has been characterised at approximately 9 % with addition of 24 μg of inhibitor protein and approximately 5 % with addition of 10 μg inhibitor protein to an assay.

6.2. Results: rOzc inhibition assays.

Inhibition with the recombinant oryzacystatin I protein over a concentration gradient resulted in approximately 75 % maximal inhibition of alimentary tract extract at 0.5 μM (see figures 6.2a and b). Inhibition of proteolysis by rOzc rapidly declined at concentrations below 0.5 μM no inhibition being observed at 0.01 μM. The level of inhibition was approximately 8 % greater than that observed with E-64 a general cysteine proteinase inhibitor (see Chapter 3). Higher inhibition with oryzacystatin I a preferential inhibitor of papain-like proteinases indicates proteinase inactivation in the gut samples assayed as described in section 6.1.

Assay pH was altered to 4.5 and the assay repeated as before. The inhibition observed with rOzc at pH 4.5 was approximately 5-6 % less than at pH 5.0. This experiment was performed to assay any differences in hydrolytic class activity with variations of assay pH. Due to the minor variability and the reduced proteolytic activity between midgut preparations it is believed that conclusions cannot be drawn on the basis of these results in respect to the relative contributions of proteinase classes at pH 5.0 and pH 4.5.

6.3. Results: rASPI 1 and rASPI 1 (sol) inhibition assays.

rASPI 1 maximally inhibits proteolytic activity by 75 % at 0.5 μM (see figure 6.3a), additional inhibition is observed at higher concentrations in line with the non specific inhibition discussed previously. Maximal inhibition by rASPI 1 in these assays was greater than previously observed by rASPI 1 and 2 (50 % and 40 % inhibition respectively, see Chapter 5). This anomaly is believed to be due to the lower proteolytic activity of the alimentary tract extracts used in these assays and described earlier.

Figure 6.3a. shows inhibition of *D. undecimpunctata* alimentary tract samples by rASPI refolded as described in Chapter 5 with minor modifications. rASPI 1 protein employed in these assays was refolded at each individual concentration for assay instead of employing serial dilutions of a 50 μM stock solution. *In vitro* re-folding was performed by adding sample cumulatively to the final concentration required for assay. Cumulative addition of rASPI 1 protein resulted in increased levels of natively refolded rASPI protein and reduced aggregate formation.

The amended rASPI 1 refolding strategy resulted in maximal inhibition of approximately 75 % at 0.5 μM and approximately 12.5 % inhibition at 0.1 μM, no inhibition was observed at a 0.05 μM rASPI 1 concentration. Allowing for inherent assay error (2-3 %) this is approximately a 1:1 ratio of inhibition (rASPI 1 final concentration; 0.5/5 = 0.1, inhibition; 74.95/5 = 14.99) indicating the majority of the inhibitors in the rASPI samples prepared in this manner are refolded in the native conformation, assuming a 1:1 enzyme-inhibitor complex.

rASPI (sol) protein produced under native conditions as soluble protein exhibited no inhibitory activity against alimentary tract proteolysis upto a concentration of 5 μM (see figure 6.3b) confirming results obtained by rASPI 2 construct protein production under native conditions (D.P. Bown personal communication). The lack of inhibitory activity of rASPI 1 produced as soluble protein compared to artificially refolded rASPI 1 protein is due to the *E. coli* cytoplasm being a relatively reducing environment (Pollitt and Zalkin, 1983) preventing thiol oxidation and di-sulphide bond formation. di-sulphide bond formation occurring in the relatively oxidising periplasmic space.

Cytoplasmic localisation of the rASPI proteins (see Chapter 5) would lead to inefficient formation of di-sulphide bridges preventing tertiary structure being achieved by the rASPI proteins. Inhibition levels at 5 μM (24 μg) rASPI (sol) were approximately 11 % higher than expected due to additional protein content quantitated by the SKTI assays (see figures 6.3b and 6.4a). The additional 11 % inhibition observed may indicate a proportion of the protein is transported into the periplasmic space of the BL21 (DE3) expression cells. In the relatively oxidising periplasmic space the di-sulphide bridges may form and the rASPI 1 protein achieve the native tertiary conformation.

6.4. Results: SKTI inhibition assays.

Inhibition assays at pH 5.0 with SKTI a 20 kDa serine proteinase inhibitor were performed to ascertain the levels of non-specific inhibition resulting from the addition

of inhibitor protein to the assays (see figure 6.4a). At SKTI concentrations of less than 1.2 μM (4.8 μg) no inhibition occurred, at a concentration of 3 μM (12 μg) approximately 5 % inhibition of proteolytic activity was observed. The highest level of non-specific inhibition observed was approximately 9 % at 6 μM SKTI (24 μg).

The level of non-specific inhibition observed in these assays performed with 0.0125 gut equivalents was significantly lower than those observed in the assays described in Chapter 5 that were performed with 0.1 gut equivalents. Lower non-specific inhibition is to be expected as the assays presented in this chapter contain a lower total protein content due to the addition of 4 fold less alimentary tract extract. Lowering the alimentary tract protein content in the assays conversely increases the ratio of BODIPYL® FL labelled casein substrate to additional protein in each assay.

6.5. Results: Dual rASPI 1, rOzc and SKTI inhibition assays.

Figure 6.5a shows inhibition of alimentary extract by a combination of plant inhibitors, all inhibitors were added at individual final concentrations of 1 μM. Upon addition rOzc and rASPI 1 less than 1 % of control substrate hydrolysis remained. These two plant inhibitors therefore, inhibit 95-99 % of the proteolytic activity in the *D. undecimpunctata* alimentary tract larval extracts used in these assays. A proportion of the observed inhibition will be due to protein content, the level of inhibition due to this cause can be estimated at less than 5 % by comparison with previous data. This result does not necessarily represent inhibition of the entire cysteine and aspartic proteinase compliment in alimentary extract due to the reasons discussed in section 6.1.

Addition of rOzc and SKTI at 1 μM concentrations showed comparable inhibition levels (74.5 %) to inhibition assays incorporating 1 μM of rOzc (75 %), when assayed at pH 5.0. Comparable inhibition levels indicates that SKTI has little or no specific inhibitory activity at pH 5.0 as would be predicted, protein content inhibition was also within margin of error values. Joint rASPI 1 and SKTI inhibition assays resulted in 13 % inhibition compared to 19 % observed with 1 μM rASPI 1 alone. The additional 6 % inhibition observed is within the bounds of that expected with the additional protein content (9.2 μg).

The assay data presented in cannot be seen as representative of the full in vivo compliment proteinase compliment of *D. undecimpunctata* alimentary tracts. Inactivation of a proportion of the overall hydrolytic activity, equalling approximately 27 % of unaffected alimentary tracts was observed. These assays can however be utilised to gauge the effectiveness of plant proteinase inhibitors of the remaining digestive compliment. Further assays with fresh extract would elucidate the combined inhibitory effects of rASPI and rOzc against the full compliment of cysteine and aspartic proteinases present in the alimentary tracts of larval *D. undecimpunctata.*

6.6. Discussion.

Inhibition studies have shown the PKPI family proteins are effective inhibitors of cathepsin D, trypsin, chymotrypsin, subtilisin, papain but not inhibitors of cathepsin E or pepsin like aspartic proteinases (Keilova and Tomasek, 1976 and 1997; Walsh and Twitchell, 1991; Suh *et al.*, 1991). Studies with native PKPI proteins extracted from

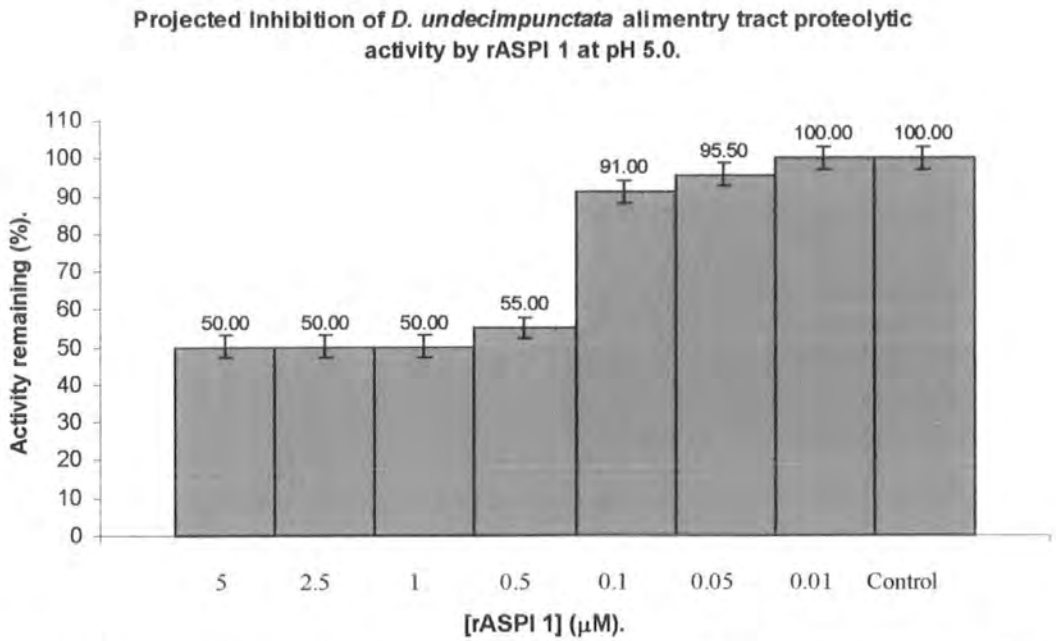


Figure 6.6a. Projected inhibition of *D. undecimpunctata* alimentary tract proteolytic activity by rASPI 1 at pH 5.0. Vertical bars represent a predicted standard error of 3 %.

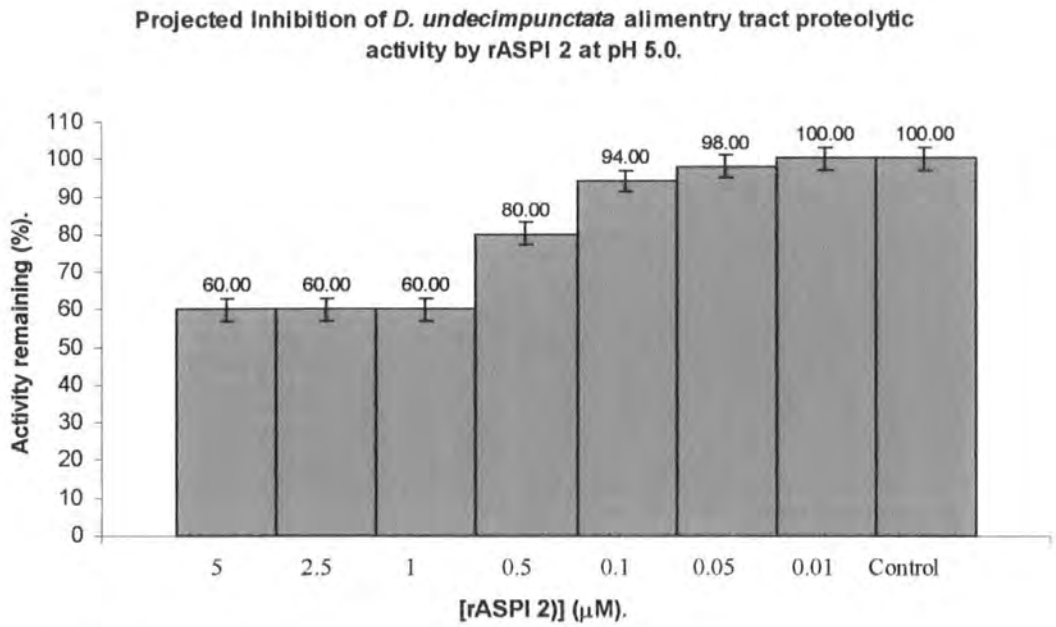


Figure 6.6b. Projected inhibition of *D. undecimpunctata* alimentary tract proteolytic activity by rASPI 2 at pH 5.0. Vertical bars represent a predicted standard error of 3 %.

potato have also shown this family of proteins are effective inhibitors of alfalfa weevil cathepsin D-like midgut aspartic proteinase activity (Wilhite *et al.*, 2000).

Figures 6.6a and b illustrate the predicted inhibition by rASPI 1 and 2 of *D. undecimpunctata* larval alimentary tract proteolytic activity (0.0125 gut equivalents). The projected maximal inhibition figures of 50 % for rASPI 1 and 40 % for rASPI 2 are drawn from the initial test assays employing inhibitor refolded at a concentration of 50 μM as described in Chapter 5. The rASPI induced inhibition levels at lower micromolar concentrations are based upon the results obtained by the amended refolding procedure with the alimentary tract preparations described in Section 6.3. A 1:1 inhibitor proteinase complex is assumed due to an approximate 5 fold reduction of inhibition being observed with a 5 fold reduction of rASPI 1 protein inhibitor concentration as described in Section 6.3.

PKPI proteins are also effective inhibitors of *D. undecimpunctata* and *D. virgifera* (data not shown) alimentary tract aspartic proteinase proteolytic activity. Maximal inhibition of *D. undecimpunctata* alimentary tract aspartic proteinase activity occurring at 0.5 μM rASPI 1 similar to the maximal inhibition at 0.5 μM of the cysteine proteinase activity by rOzc. rOzc however is a more effective inhibitor at concentrations lower than 0.5 μM compared to rASPI 1 protein. rASPI 1 used in combination with a cysteine proteinase inhibitor such as oryzacystatin I have been shown to arrest both cysteine and aspartic proteinase mediated hydrolytic digestion of the general protein substrate BODIPYL® FL labelled casein.

Dual expression of a rASPI 1 and oryzacystatin I in plant material would be expected to produce a transgenic plant resistant to corn rootworm feeding by virtue of

inhibiting the cathepsin L and cathepsin D-like proteinase activities in the larval alimentary tract. The rASPI proteins ability to inhibit the putative serine proteinase proteolytic activity identified in *D. undecimpunctata* larval midguts by pH optima and SKTI inhibition (Edmonds *et al.*, 1996) would require further experimentation. Inhibition of the serine proteinase activity present in *D. undecimpunctata* would be predicted as PKPI proteins are known to be inhibitors of trypsin, chymotrypsin and the bacterial serine proteinase subtilisin (Walsh and Twitchell, 1991).

Chapter 7. Conclusions and future work.

7.1. Conclusions.

The proteolytic activities present in larval *D. undecimpunctata* alimentary have been characterised by biochemical and molecular studies. Cysteine and aspartic proteolytic activity has been identified by pH optima and inhibition employing class specific proteinase inhibitors, an as yet uncharacterised proteinase activity has been identified by pH optima at pH 10. The *in vitro* data presented in this work supports earlier research, with both *D. undecimpunctata* and other coleopterans that the majority of alimentary tract endopeptidase proteolytic digestion is performed by cysteine proteinases with secondary cathepsin D and serine proteinase activities (Purcell *et al.*, 1992; Orr *et al.*, 1994; Edmonds *et al.*, 1996).

Cathepsin L and D-like cDNA clones have been isolated and characterised, lysine and *N*-glycosylation site positional homology to known lysosomal proteinases have been used to predict the localisation of the proteins in the *in vivo* conditions of the alimentary tract based upon a putative glycosylation-phosphorylation pathway. The aspartic proteinase activity characterised by pepstatin A and rASPI protein inhibition is predicted to be due to Du-asp 1 like proteinases. This is hypothesised due to Du-asp 1 cDNA clones being isolated at a frequency of 2.5:1 relative to Du-asp 2 and inefficient lysosomal transport due to the lack of a highly conserved *N*-glycosylation site present in lysosomal cathepsins D and Du-asp 2. Du-asp 1 localisation be it intra or extracellular remains to be

demonstrated experimentally, but is predicted to be extracellular as either a soluble or alimentary tract epithelial cell membrane bound proteinase.

The annotated sequence of Du-asp 1, analysis as a putative cathepsin E-like aspartic proteinase and in addition *D. undecimpunctata* alimentary tract mRNA samples have been provided to D.P. Bown for expression in *Pichia pastoris*. To the authors knowledge at the time of leaving the yeast expression of the Du-asp 1 sequence was successful. Recombinant Du-asp 1 protein had been purified and initial characterisations of proteolytic activity and inhibition had commenced. The characterisation of rASPI 1 and 2 protein inhibition of Du-asp 1 and Du-asp 2 may provide additional evidence for the role of Du-asp 1 as the major aspartic proteinase activity in the larval alimentary tract of *D. undecimpunctata.*

No conclusions can be drawn regarding the relative contributions of the obtained Du-cys proteinases to dietary digestion, other than isolation from an alimentary tract library indicates the genes are transcribed in the gut. Alimentary tract specific induction of individual Du-cys isoforms, the role of conserved *N*-glycosylation sites and the role of the proteinases in the alimentary tract *in vivo* remains to be elucidated.

A range of additional protein sequences (see appendix A.2.) has been obtained from the larval *D. undecimpunctata* alimentary tract cDNA library constructed in this work. These proteins have been characterised at the sequence level including a serine carboxypeptidase, a gut-specific chitinase, two glutathione *S*-transferases, P450 mixed function oxidase and an encapsulation response protein. Further clones encoded putative proteins whose function cannot be assigned by homology to previously characterised sequences (see appendix A.3.).

PKPI sequences of both PKPI-A and B classes have been isolated by PCR from a potato tuber cDNA library. The amplified cDNAs of rASPI 1 and 2 have been ligated into the pET 24a expression vector, expressed and artificially refolded. Biochemical assays have demonstrated purified PKPI proteins possesses inhibitory activity against larval *D. undecimpunctata* alimentary tract aspartic proteinase proteolytic activity. Dual assays incorporating oryzacystatin I and rASPI 1 protein have shown increased inhibition of *D. undecimpunctata* proteolytic activity, relative to single rASPI or rOzc inhibitor assays. Further assays are required to establish the effectiveness of rASPI and rOzc against the full range of proteolytic activities present in *D. undecimpunctata* alimentary tracts.

7.2. Future work.

7.2.1. *D. undecimpunctata* aspartic and cysteine proteinases.

Further research is required to establish the tissue distribution and factors effecting transcription regulation of the cathepsin L and D-like cysteine proteinase isoforms in the *in vivo* conditions of the alimentary tract. Northern blot analysis was planned for this work but due to failure to obtain required insect material was not performed. Northern blot analysis would reveal which of the obtained cathepsin D and L sequences showed upregulated mRNA expression in the alimentary tract cells relative to other cell types such as head and fat body.

Proteinase genes showing upregulated transcription in alimentary tract tissue would indicate a specific role in extracellular digestion for these gene products. Northern Blot analysis alone would not elucidate if the *D. undecimpunctata* proteinases are secreted, or the mechanism by which the sequences were secreted, constitutively or due to stem cell proliferation. Creation of a *D. undecimpunctata* cell line (Lynn and Stoppleworth, 1984) would allow the cellular localisation characteristics of the cathepsin sequences to be studied.

Polyacrylamide in-gel analysis of *D. undecimpunctata* proteolytic activities showed E-64 inhibited cysteine proteinase activity at 24 kDa and 35 kDa (Edmonds *et al.*, 1996) within 1-2 kDa of the predicted mature (22-23 kDa) and propeptide (33-35 kDa) forms of Du-cys 1-4 lacking *N*-linked oligosaccharides. Glycosylation of potential sites *in vivo* could be established by SDS-PAGE of purified recombinant cathepsins. Digestion of purified cathepsins followed by SDS-PAGE would further establish the glycosylation *in vivo* relative to potential sites on Du-cys 1-4 N and C-terminals.

Deleting potential *N*-glycosylation sites by site directed mutagenesis followed by expression and localisation studies would establish the ability of *N*-linked oligosaccharides to determine cellular localisation *in vivo*. Expression studies would identify if secretion of the *D. undecimpunctata* cathepsin L-like cysteine proteinases is constitutive or determined by cell growth as with mouse cathepsin L (Prence *et al.*, 1990). Factors inducing cysteine proteinase gene upregulation could also be investigated and the effect of upregulation on cellular localisation of the cathepsin sequences isolated in this work. Of particular interest would be the cellular localisation of the proteinases under transcription induced and non-induced conditions.

Mutational studies of lysines similar to those performed in mammal cathepsins (Cuozzo *et al.*, 1995 and 1998) followed by determination of phosphorylation would provide evidence for the presence of a putative conserved lysine structure and the position of primary lysines determining phosphorylation of each glycosylation position. 3-D modelling of invertebrate cathepsin L and D enzymes would elucidate if the structural motif observed in mammalian cathepsin L and D sequences (Cuozzo *et al.*, 1998) is mirrored in invertebrate sequences.

Such a program would be the first such study undertaken in insects to investigate specific insect proteinase gene transcription in response to cellular growth cycles. The characterised proteinase sequence would also warrant further research focusing on the *N*-glycosylation sites and lysine amino acid role in putatively determining *in vivo* alimentary tract localisation and hence relative contribution to dietary digestion.

7.2.2. Additional sequences isolated from the *D. undecimpunctata* alimentary tract cDNA library.

A range of further sequences were obtained from the *D. undecimpunctata* cDNA library these were characterised at the sequence level by BLAST homology search (see Appendix A.2. and A.3.). Isolation from alimentary tract material of chitinase, glutathione *S*-transferase, P450 mixed function oxidase, serine carboxypeptidase and an encapsulation response protein indicate expression of these sequences in the alimentary tract cells. Northern Blot analysis would elucidate which of the sequences perform a gut-

specific function, further research would establish if the proteins are present in the cellular, alimentary tract or hemolymph environments.

Expression of the sequences in preferentially a baculovirus or bacterial expression system would allow the characterisation of the protein properties. In respect of the glutathione *S*-transferases and the P450 mixed function oxidase interactions between the proteins and plant allelochemicals or pesticides could be investigated and the mechanism of potential detoxification elucidated be it by metabolism or sequestration. The serine carboxypeptidase cellular localisation and hence the role of the protein in dietary or intracellular digestion also warrants further research.

The chitinase sequence isolated from *D. undecimpunctata* alimentary tract exhibits high amino acid identity and sequence features possessed by a gut-specific chitinase of *Phaedon cochleariae* (Girard and Jouanin, 1999). Research is required to confirm the *in vivo* role of this protein however due to protein similarity the *D. undecimpunctata* chitinase is predicted to perform a similar gut-specific role. The putative encapsulation response protein identified by sequence homology to a *Tenebrio molitor* 56 kDa protein (Cho *et al.*, 1999) requires further sequencing to obtain the 3' end of the clone amino acid code prior to further investigation of the properties of this protein.

7.2.3. PKPI proteins.

Further *in vitro* characterisation of the PKPI proteins and *in vivo* bioassays were planned for this work but due to time and other restrictions these attempts had to be abandoned. *In*

vivo assessment of rOzc and rASPI on larval development, mortality and growth by artificial diet bioassays was hindered by problems associated with artificial refolding of milligram quantities of PKPI proteins by the technique described. rASPI sequence insertion into the pET 32 expression vector and utilising a thioredoxin reductase deficient expression strain for expression may solve these problems. A new vector system would be needed as pET 24a and Novagen trxB$^-$ share kanamycin resistance. A pET 32 trxB$^-$ combination would allow di-sulphide bridge formation within the cytoplasm of *E.coli* and theoretical recovery of sufficient quantities of natively refolded PKPI proteins.

Recognition of the plant cleavage signal by the BL21 DE3 cells could be established by site directed mutation converting Ser_{29} in rASPI 2 to an Asn_{29}. On expression analysis if a cleaved form of the mutated rASPI 2 protein is observed recognition of the plant mature chain cleavage signal by the bacterial expression cells and the role of Ser_{29} in blocking the cleavage recognition signal between Asx_{29} and Leu_{30} would be established. N-terminal sequencing of the processed mature protein would provide the definitive evidence.

The full inhibitory range of rASPI 1 a member of the PKPI-B sub-family of PKPI proteins has as yet to be characterised, cathepsin D-like inhibition only being confirmed in Chapters 5 and 6. Characterisation of PKPI-A group potato aspartic and serine proteinases has shown multiple inhibitory activities against trypsin, chymotrypsin and subtilisin, the PKPI-C class also possessing papain inhibitory activity (Keilova and Tomasek, 1976 and 1977, Walsh and Twitchell, 1991, Suh *et al.*, 1991; Valueva *et al.*, 1997). Further characterisation of the inhibitory capabilities of rASPI 1 and rASPI 2

against vertebrate and invertebrate proteinases would elucidate the full inhibitory spectrum of these proteins.

Analysis of the *in vitro* inhibitory capabilities and structure of the ASPI proteins would be desirable to establish the presence of one or multiple inhibition sites. Crystallisation and X-ray diffraction studies with the rASPI proteins refolded by the method described in this work would elucidate the rASPI protein structure providing a 3D model, in conjunction with multiple sequence alignment this may allow individual inhibitory domains to be identified and the inhibition mechanism to be understood. Analysis of the domains could be used to engineer the properties of the natural protein, altering and/or extending the inhibitory specificity of the native protein.

Incorporation of a cysteine proteinase inhibiting domain such as that possessed by oryzacystatin 1 to one of the rASPI proteins could be achieved by alteration of a native inhibitory site. This alteration would produce a theoretically effective toxin capable of inhibiting the major cathepsin L and D-like proteolytic activities identified in larval *D. undecimpunctata*. Retaining a serine (trypsin, chymotrypsin and subtilisin) inhibitory domain/s may also enable inhibition of the as yet uncharacterised proteolytic activity identified at pH 10 (see Chapter 3).

Expression of the rASPI proteins allows the characterisation of single protein isoforms of this family to be performed without the requirement to isolate and purify protein from potato material. The capability to express and purify members of this family of proteins in a bacterial expression system is a significant step forward in their usage. Protein analysis can be used to elucidate the structure and domains responsible for inhibitory activity, furthering our understanding of proteinase and inhibitor interactions.

Site directed mutagenesis can now be employed to alter the properties of the PKPI proteins, bacterial expression and refolding allowing rapid recombinant protein isolation and evaluation of the mutated isoforms.

7.2.4. Proteinase inhibitors as insecticidal proteins.

The production of transgenic maize material expressing a combination of oryzacystatin and PKPI proteins was unsuccessful due to failure of calli material to proliferate in all but control experiments. Further preparatory research regarding the properties of the PKPI proteins is required before such experiments can be performed successfully. Problems to be addressed include *in vivo* di-sulphide bond formation and PKPI gene transcription induction repression by auxin (Ishikawa *et al.*, 1994b).

Due to the *in vivo* properties of the cathepsin D inhibitors the author considers that protein extraction from transgenic plant material and *in vitro* evaluation of the inhibitory properties is necessary to establish successful expression of active PKPI transgenes. *In vivo* bioassays employing only purified active rASPI proteins would not be expected to show high levels of toxicity to *D. undecimpunctata* larvae even if Du-asp 1 is a soluble excreted proteinase due to cysteine proteinase mediated digestive proteolysis. Dual bioassays incorporating rASPI 1 protein and a cysteine proteinase inhibitor such as oryzacystatin I would be expected to show increased toxicity relative to single rASPI or oryzacystatin I bioassays. The *in vivo* effectiveness of the rASPI inhibitors if Du-asp 1 is an alimentary tract epithelial cell membrane bound proteinase will be determined by rASPI protein ability to transverse the larval peritrophic membrane.

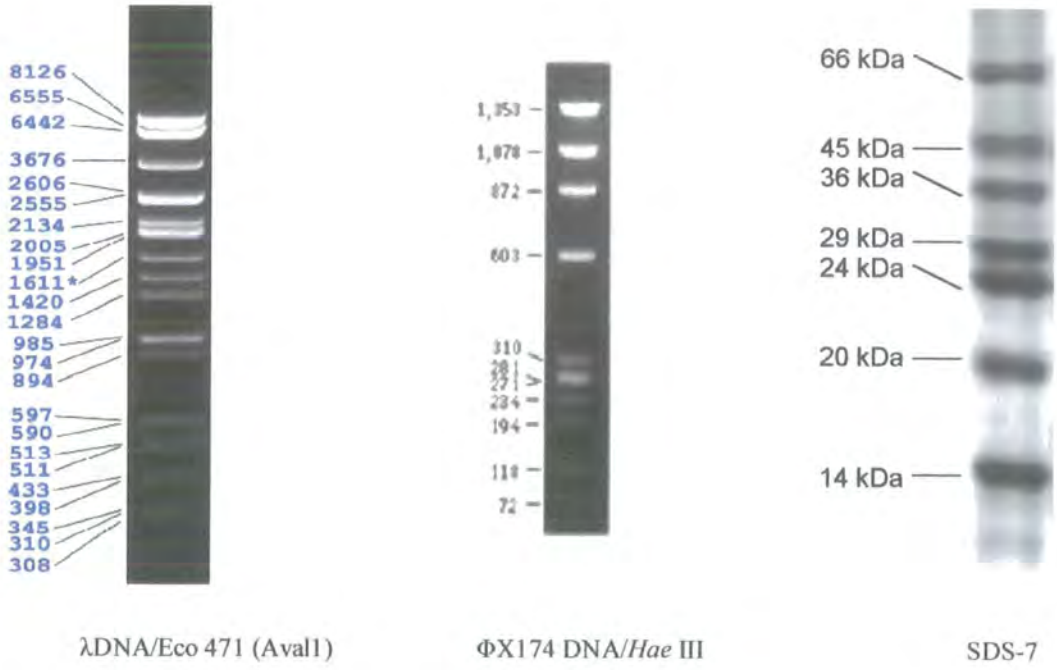
Previously, single inhibitor proteins have been employed by bioassay or transgenic plants with inhibitory activity towards only a percentage of target pest proteinases (reviewed in Jongsma and Bolter, 1997) Incorporation of oryzacystatin in artificial diet (0.10 mM) reduced larval survival of *D. undecimpunctata* by 50 % compared to controls (Edmonds *et al*, 1996). This work shows that ingestion of a cysteine proteinase inhibitor alone causes a high level of larval mortality. Inhibition of the cysteine proteinase activity by oryzacystatin I, the aspartic proteinase and putative serine activities by rASPI proteins would be expected to further reduce larval endopeptidase protein hydrolysis. Reduced dietary protein hydrolysis would theoretically lead to amino acid depravation and increased mortality *in vivo*.

Further research could be performed to develop single inhibitor proteins with cathepsin L-like, cathepsin D-like and serine proteinase inhibitor domains such proteins may also prove an effective toxin against other major Coleopteran and fungal crop pests. Potential insect herbivore targets including other Western corn rootworm, Northern corn rootworm, Mexican corn rootworm, alfalfa weevil, Colorado potato beetle, maize weevil, cabbage seed weevil and other coleopteran pests of agricultural crops

The main concerns in transgenic crop production are consumer safety, environmental impact, protection levels and increased crop yield. The rASPI proteins are natural plant proteins isolated from potato with cathepsin D-like aspartic proteinase inhibitory specificity making them attractive proteins with which to engineer transgenic plants for herbivore resistance. Commercial crops containing engineered versions of the native PKPI and oryzacystatin I proteins would theoretically have great potential to safely enhance crop resistance to herbivorous pests.

Appendices.

A.1. Commercial size markers and vectors.



λDNA/Eco 471 (AvaII) ΦX174 DNA/*Hae* III SDS-7

Figure A.1a. λDNA/Eco 471 (AvaII), ΦX174 DNA/*Hae* III DNA markers and SDS-7 protein size marker.SDS-7 protein bands 66 kDa (bovine albumin), 45 kDa (egg albumin), 36 kDa (rabbit muscle glyceraldehydes-3-phosphate Dehydrogenase), 29 kDa (bovine carbonic anhydrase), 24 kDa (bovine pancreas trypsinogen), 20 kDa (soybean Trypsin inhibitor), 14 kDa (bovine milk α-lacalbumin).

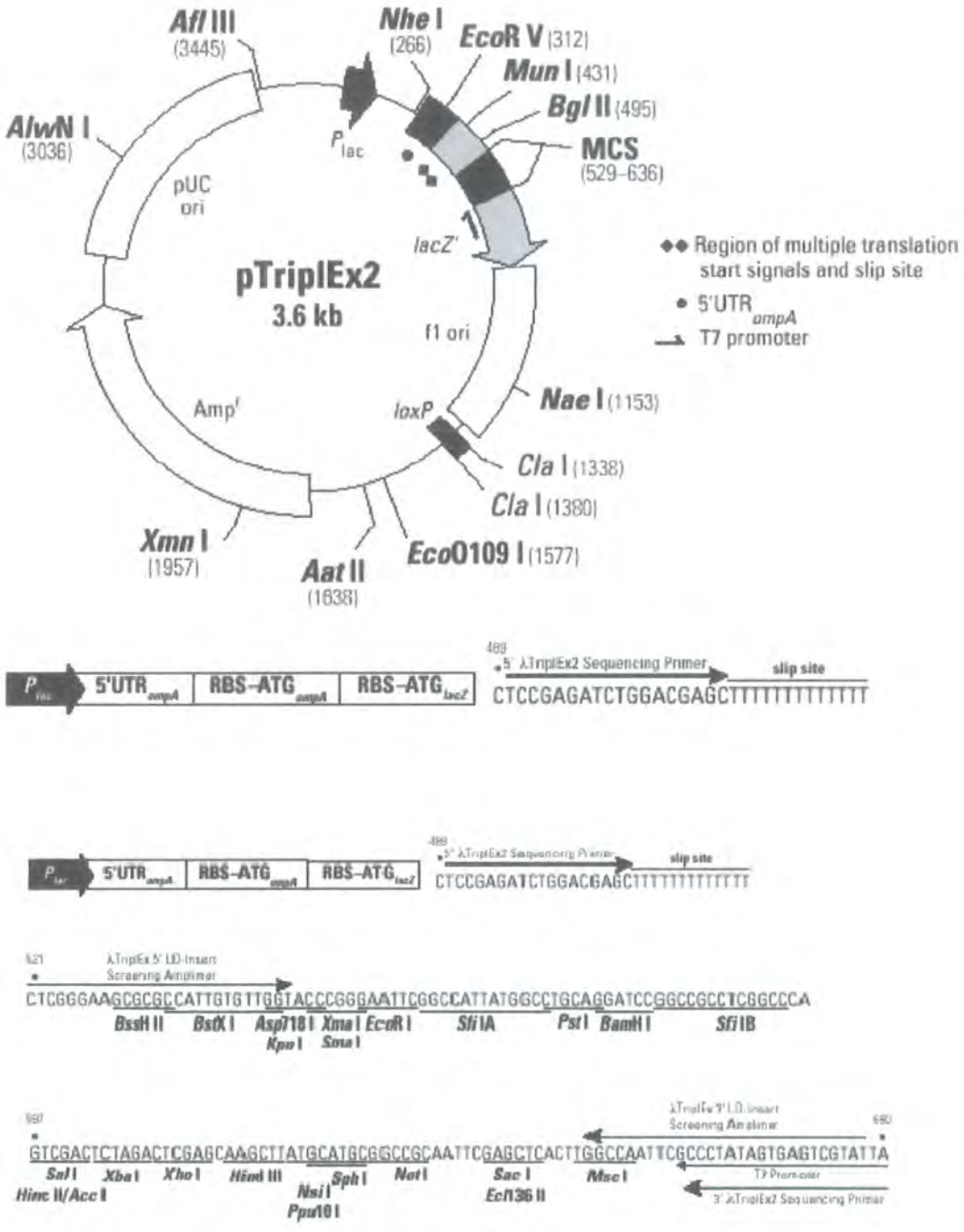


Figure A.2a. pTriplEx 2 plasmid map (top) and multiple cloning site (bottom) (Clontech (Pro9517), 2000).

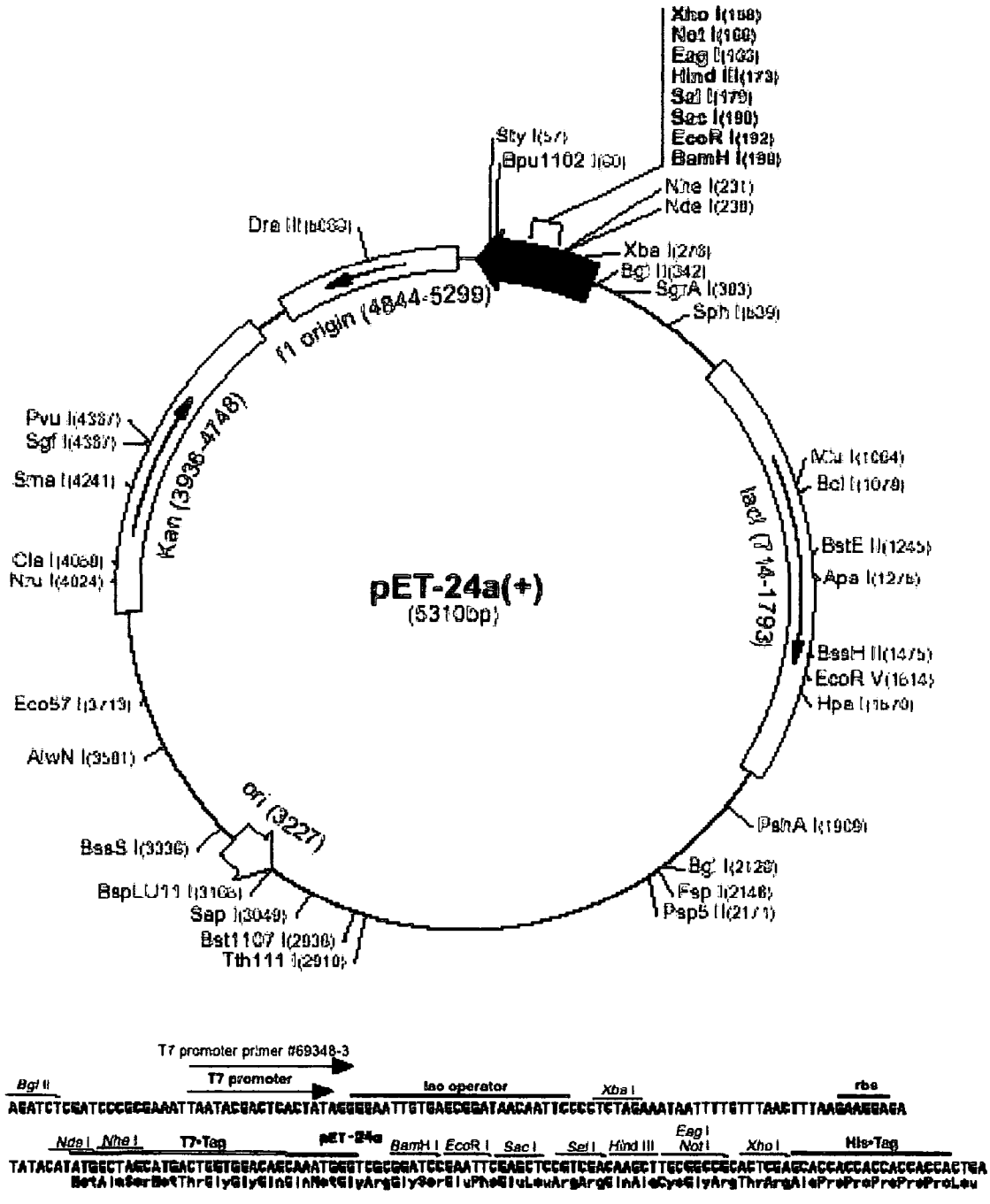


Figure A.3a. pET 24a plasmid map (top) and multiple cloning site (bottom) (Novagen (TB070), 1998).

A.2. Results: Additional cDNA sequences identified in the *D. undecimpunctata* alimentary tract library.

A.2.0. Introduction.

A range of sequences were obtained from the *D. undecimpunctata* alimentary tract cDNA library during the endoproteinase screening process. 1 % agarose gel electrophoresis of EcoR1 and Sal 1 digested pTriplEx2 plasmids revealing inserts of the expected size range for aspartic (approximately 1.5 kb) and cysteine proteinases (approximately 1.1 kb), were sequenced as putative aspartic or cysteine proteinases clones. Plasmids possessing inserts outside this size range were sequenced as possible clones of interest due to isolation from the *D. undecimpunctata* alimentary tract cDNA library (see figures A.2a and b).

Sequencing a range of *D. undecimpunctata* cDNA clones revealed a number of clones encoded proteins showing amino acid code similarity to sequences previously submitted to the NCBI Genbank database. Additional clones encoded putatively full length transcripts unidentifiable by BLAST sequence homology searches (see appendix 3). Complete sequences of all cDNA clones are currently not available, all the sequences were present in the cDNA library at a frequency of at least 1 in approximately 7500 cDNA clones.

A.2.1. Results: *D. undecimpunctata* carboxypeptidase (EC 3.4.16.1).

Blast homology search of clone A4 (Du-carb) (see figure A.2.1a.) showed 38 % amino

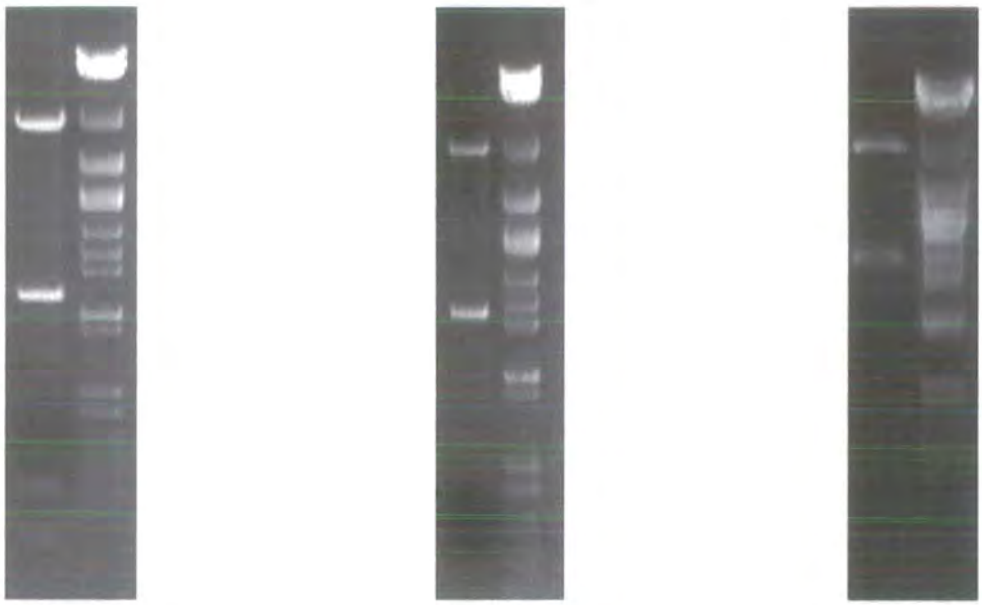


Figure A.2a. EcoR1 and Sal 1 digested pTriplEx2 plasmids. Digests correspond to cDNA clones (left to right); A4 1630 bp (*D. undecimpunctata* serine type carboxypeptidase), CA5 1400 bp (*D. undecimpunctata* chitinase) and A14 1650 bp (*D. undecimpunctata* early stage encapsulation response protein). M: λDNA/Eco 471 (AvaII) DNA marker. Nucleotide sequences are presented in section A.3.

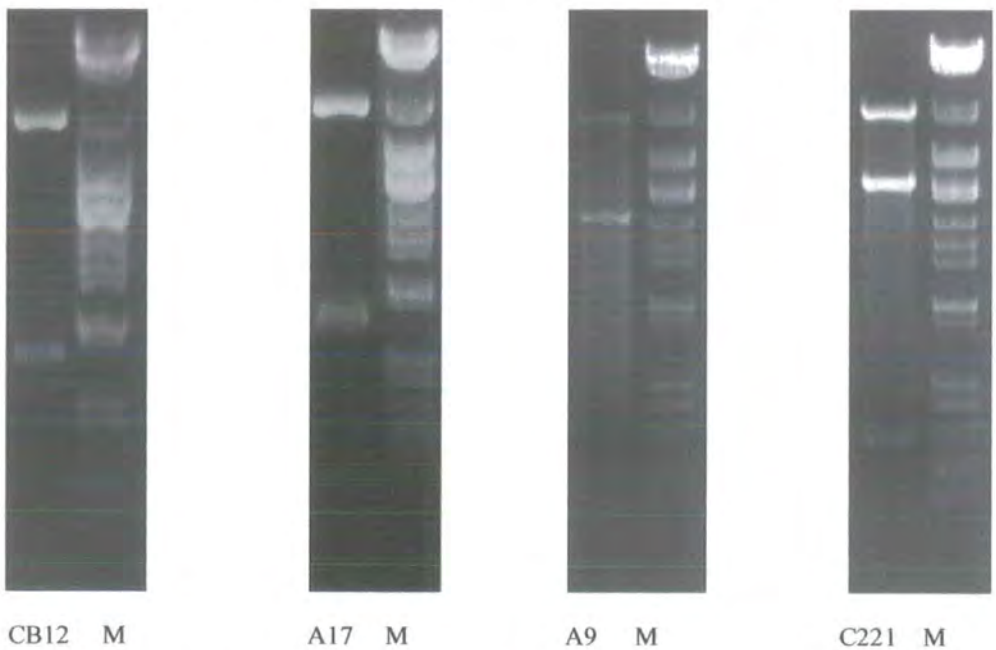


Figure A.2b. EcoR1 and Sal 1 digested pTriplEx2 plasmids. Digests correspond to cDNA clones (left to right); CB12 800 bp (*D. undecimpunctata* glutathione *S*-transferase 1), A17 800 bp (*D. undecimpunctata* glutathione *S*-transferase 2), A9 1700 bp (*D. undecimpunctata* P450 mixed function oxidase), C221* 2500 bp (*D. undecimpunctata* homologue of a *C. elegans* slit protein) (* see appendix A.3.). M: λDNA/Eco 471 (AvaII) DNA marker. Nucleotide sequences are presented in section A.3.

```
1                                                                    67
TIIPGIFKAVT.KC.ALK MKAILSISFVLFILCSIES KSLNVKHSRKPIFKDLYGKLTPEVRGNPGE
[      5'UTR      ] [       Leader      ] [
68                                                                  136
PLILTDLIKAGKLDEAQNKALVQGLDTDVQSYSGYFIVDEQHNSNIFFWFFPSQSDASSDPVVLWLQGG

137                                                                 205
PGSTSMFGLFQENGPLLLKDGELGVRPTSWNRNHSVIYIDQPVGTGWSFTDGGYANDQNKVAADLYEAL

206                                                                 274
QQFFHSIYQHQDRDFVTGESYGGKYVPAISYEIHQRNEDAELKINLKGLAIGNGFTDPIVQMRYSEYVY
                                  Propeptide
275                                                                 343
QHGLIDLNTKNSLHNLENKAVQAIQSEEYYDAFLLWDDIVNTIQDFTELDIYNYLTLKGSAADEEELNE

344                                                                 412
FFSKEEVRKAIHVGATEFGDGQVYTHLEKDIPKSEAYKVAELLKYYKVLIYNGQTDLIVPYALTVNYLQ

413                                                                 481
RLKFPGSEEYAEADRTVWHDDNDRSVVAGYAKTAGKLTEIMVRNSGHMVPTDQPEWALQLINKFTRNEP

482         495
FN .I.LKFCICRKK
[    3'UTR    ]
```

Figure A.2.1a. Annotated predicted amino-acid sequence of cDNA clone A4. Signal peptide identified by SignalP server (Nielsen., *et al* 1997). Sequence domains in bold show conservation in serine carboxypeptidases, catalytic serine is in bold and underlined (Galijart *et al.*, 1990). Potential *N*-glycosylation sites are indicated in bold, underlined and in italics. UTR; untranslated region.

```
Du-carb  :  ISFVLFILCSIESKSLNVKHSRKPIFKDLYGKLTPEVRGNPGEPLILTDLIKAGKLDEAQNKALV-QGLD 92
            + F L +L +    + +      P  K + G  +P   G  GEPL LT L++ GK++EA+NKA V   +
Mos-carb :  VKFHLLVLIAFTCYTCSDATLWNPYKKLMRGSASPPRPGESGEPLFLTPLLQDGKIEEARNKARVNHPML 51

Du-carb  :  TDVQSYSGYFIVDEQHNSNIFFWFFPSQSDASSDPVVLWLQGGPGSTSMFGLFQENGPL-LLKDGELGVR 161
            + V+SYSG+  VD +HNSN+FFW+ P++++    P+++WLQGGPG++S+F G+F+ENGP + ++  +  R
Mos-carb :  SSVESYSGFMTVDAKHNSNLFFWYVPAKNNREQAPILVWLQGGPGASSLFGMFEENGPFHIHRNKSVKQR 131

Du-carb  :  PTSWNRNHSVIYIDQPVGTGWSFTDG--GYANDQNKVAADLYEALQQF---FHSIYQHQDRDFVTGESYG 238
              SW++NH +IYID PVGTG+SFTD   GY+ ++  V  +L + +QQF   F ++ +H    +++GESYG
Mos-carb :  EYSWHQNHHMIYIDNPVGTGFSFTDSDEGYSTNEEHVGENLMKFIQQFFVLFPNLLKHP--FYISGESYG 211

Du-carb  :  GKYVPAISYEIHQRNEDAELKINLKGLAIGNGFTDPIVQMRYSEYVYQHGLIDLNTKNSLHNLENKAVQA 295
            GK+VPA  Y IH  N  ++ KINL+GLAIG+G+TDP+ Q+ Y EY+Y+ GLIDLN +         A+
Mos-carb :  GKFVPAFGYAIH--NSQSQPKINLQGLAIGDGYTDPLNQLNYGEYLYELGLIDLNGRKKFDEDTAAAIAC 277

Du-carb  :  IQSEEYYDAFL----LWDDIVNTIQDFTEL----DIYNYLTLKGSAADEEELNEFFSKEEVRKAIHVGAT 361
             + ++   A      L+D +      F ++      YN++     +  +  L EF S  EVRK IHVG
Mos-carb :  AERKDMNSANRLIQGLFDGLDGQESYFKKVTGFSSYYNFIKGDEESKQDSVLMEFLSNPEVRKGIHVGEL 347

Du-carb  :  EF----GDGQVYTHLEKDIPKSEAYKVAELLKYYKVLIYNGQTDLIVPYALTVNYLQRLKFPGSEEYAEA 423
            F    G  + V   L +D   + A  V++LL +Y+VL YNGQ D+I  Y +TV++L ++ F G  EY  A
Mos-carb :  PFHDSDGHNKVAEMLSEDTLDTVAPWVSKLLSHYRVLFYNGQLDIICAYPMTVDFLMKMPFDGDSEYKRA 417

Du-carb  :  DRTVWHDDNDRSVVAGYAKTAGKLTEIMVRNSGHMVPTDQPEWALQLINKFT 475
            +R ++  D +   +AGY K AG+L E+++RN+GHMVP DQP+WA  +I  FT
Mos-carb :  NREIYRVDGE---IAGYKKRAGRLQEVLIRNAGHMVPRDQPKWAFDMITSFT 466
```

Identities = 190/472 (40%), Positives = 289/472 (60%), Gaps = 26/472 (5%)
Score = 353 bits (905), Expect = 3e-96

Figure A.2.1b. Alignment search result for cDNA clone A4 (Du-carb). Du-carb exhibits closest homology to a Serine carboxypeptidase of *Aedes aegypti* (Cho *et al.*, 1991). Alignment performed by blastp (www.ncbi.nlm.nih.gov/BLAST/), + indicates conservative substitution.

```
1                                                                    67
ES.Y..VLK MKVLVLLSLLSTFLICQTSG AKKNVVCYFASWTIYRAGNGAFDVSNIDPSLCTHINFA
[ 5'UTR ] [      Leader       ] [
68                                                                  136
FLGLNLDGSIHILDSWESNDPGGHEGFKRLVGLKETNPDLKVSVSMGGWNEGSEHYSAVASDPAKRVKL

137                                                                 205
AGEVLAFIENWGFDGFDLDWEYPGLRGGNETIDKEDFVELVKALSDVLQPKGYLLSIATAGAVEKIDLG
                          Pro-peptide
206                                                                 274
YDVPVINELVDMLNVMVFDFHGAFENFVGHVSPLFPAQIDYEYEANSTYNVDAGIQHWILSGADPAKIN

275                                                                 343
LGIVTYGRTFTLADKTNTSLYADVIGGGNKGPYTGQSGYLGYNEICEFHLNSTYVWDDQQKVPHRYWDD

344                                                                 411
QWVGFEDERSIREKVNYARDNELGGMMVWSLDYDDFRGTCGEKYILLKTINKIHDPQF .ICKF.LCVV
                                                          ] [    3'UTR
412    424
GYSRKLS.IL
        ]
```

Figure A.2.2a. Annotated predicted amino-acid sequence of cDNA clone CA5. Signal peptide identified by SignalP server (Nielsen *et al.*, 1997). Sequence in bold indicates the conserved region II of glycosyl hydrolase family 18 chitinases. Amino acid tryptophan$_{145}$ shown in bold and underlined is required for chitinolytic activity, but not sugar binding (Huang *et al.*, 2000). Potential *N*-glycosylation sites are shown in bold, underlined and italics. UTR; untranslated region.

```
Du-chit  : LKMKVLVLLSLLSTFLICQTSGAKKNVVCYFASWTIYRAGNGAFDVSNIDPSLCTHINFA 64
           +K +V +   +L+      +  + +N+VCYFASWT+YR GNG FDVSNI+P LCTHINFA
Phae-chi : MKNQVFISFCVLTLIFSSISIVSGRNIVCYFASWTVYRPGNGLFDVSNIEPDLCTHINFA 60

Du-chit  : FLGLNLDGSIHILDSWESNDPGGHEGFKRLVGLKETNPDLKVSVSMGGWNEGSEHYSAVA 124
           F+GL+ DG+I+I+D WES+D G + GF+ L+ L+ ++P LKV VSMGGWNEG++++S VA
Phae-chi : FIGLHEDGTINIIDKWESDDDGKYHGFRNLLDLRNSHPSLKVLVSMGGWNEGTKNFSKVA 120

Du-chit  : SDPAKRVKLAGEVLAFIENWGFDGFDLDWEYPGLR-GGNETIDKEDFVELVKALSDVLQP 183
           +DP  R  LA  V AFI   GFDGFD+DWEYPG R G N TIDK++FV L++ LS VL P
Phae-chi : ADPVLRKTLANNVGAFIRQVGFDGFDIDWEYPGSREGSNVTIDKDNFVALLEDLSAVLHP 180

Du-chit  : KGYLLSIATAGAVEKIDLGYDVPVINELVDMLNVMVFDFHGAFENFVGHVSPLFPAQIDY 243
           KG LL+ A AG VE+IDLG+DVP +NE++DM+NVMV+DFHG FE FVGH+SPL  + +DY
Phae-chi : KGKLLTAAVAGGVERIDLGFDVPKVNEILDMINVMVYDFHGHFEPFVGHLSPLHASSLDY 240

Du-chit  : EYEANSTYNVDAGIQHWILSGADPAKINLGIVTYGRTFTLADKTNTSLY---------- 292
           E   N+T  V  GI++WI  GA P KIN+GI TYGR+FTL D  NT LY
Phae-chi : ENGRNATMTVATGIKYWIYKGASPEKINMGIATYGRSFTLKDPNNTQLYAPNVGGPTTFS 300

Du-chit  : --ADVIGGGNKGPYTGQSGYLGYNEICEFHLNSTYVWDDQQKVPHRYWDDQWVGFEDERS 350
             A  +GGG  GPYT Q G+LGYNEICE H + TY WDD+QKVPHR   DQWVG+ED  S
Phae-chi : CNAPNVGGGRSGPYTRQEGFLGYNEICELHSDWTYHWDDEQKVPHRTSGDQWVGYEDPAS 360

Du-chit  : IREKVNYARDNELGGMMVWSLDYDDFRGTCGEKYILLKTI 390
           ++ KV +A    LGGMM+W+ D DDF G CG+ Y LLKT+
Phae-chi : LKYKVEFAVSKNLGGMMMWAFDTDDFGGHCGDTYPLLKTL 400
```

Identities = 227/400 (56%), Positives = 288/400 (71%), Gaps = 14/400 (3%)
Score = 475 bits (1223), Expect = e-133

Figure A.2.2b. Alignment search result for cDNA clone CA5 (Du-Chit). Du-chit exhibits closest homology to a gut-specific chitinase of *Phaedon cochleariae* (Mustard beetle) (Girard and Jouanin, 1999a). Alignment performed by blastp (www.ncbi.nlm.nih.gov/BLAST/), + indicates conservative substitution. Amino acids in bold and underlined indicate conserved potential *N*-glycosylation sites.

```
1                                                                    68
AGFESIYY..LVI MAHTYKVGYFNFTARGEPIRMLLSYGEIPFEDERISREDWPKIKPTTPLGQLPVL
[   5'UTR   ] [
   f
69                                                                  138
EIDGKQIPHSISICRYLASVVKLDGKDAKENLRIDVAVETLLDLQKLAFEYVFESDAAKKDEKNKKLEE
                        Mature Chain

139                                                                 207
SLPLFLGKLEEQANKNGGYTALDRLSWADIIFICIYEGLMNILGKNTFDSYPSLQKIKKNVLDVKGIRN


208                                          256
WIKNRPETPSTMGPLKFEYDLKQEV .V.LLYIVNIY.NDYILLKIKIK
                          ] [        3'UTR        ]
```

Figure A.2.3a. Annotated predicted amino-acid sequence of cDNA clone CB12. UTR; untranslated region. Amino acids in bold, underlined and in italics indicate a potential *N*-glycosylation site. Amino acids in bold and underlined amino acids show high conservation in all GST enzymes, amino acids underlined and in italics show high conservation in GST 2 class enzymes (Feng *et al.*, 1999).

```
Du-glu 1:  YKVGYFNFTARGEPIRMLLSYGEIPFEDERISREDWPKIKPTTPLGQLPVLEIDGKQIPH 75
           YKV YFN  A GEP+R LLSYG +PF+D RI+RE+WP +KPT P+ Q+PVLE+DGK++
Mos-glu :  YKVYYFNVKALGEPLRFLLSYGNLPFDDVRITREEWPALKPTMPMRQMPVLEVDGKRVHQ 78

Du-glu 1:  SISICRYLASVVKLDGKDAKENLRIDVAVETLLDLQKLAFEYVFESDAAKKDEKNKKL-E 134
           S+++CRY+A  + L G +  E L+ID  V+T+ D +       +E D   K++K   L
Mos-glu :  SLAMCRYVAKQINLAGDNPLEALQIDAIVDTINDFRLKIAIVAYEPDDMVKEKKMVTLNN 138

Du-glu 1:  ESLPLFLGKLEEQANKNGGYTALDRLSWADIIFICIYEGLMNILGKNTFDSYPSLQKIKK 194
           E +P +L KL   A +N G+  L + +WAD+ F  I + L  +   N  +++P+LQ++ +
Mos-glu :  EVIPFYLTKLNVIAKENNGHLVLGKPTWADVYFAGILDYLNYLTKTNLLENFPNLQEVVQ 198

Du-glu 1:  NVLDVKGIRNWIKNRPET 212
            VLD + ++ +I  RP T
Mos-glu :  KVLDNENVKAYIAKRPIT 216
```

Identities = 82/198 (41%), Positives = 126/198 (63%), Gaps = 1/198 (0%)
Score = 166 bits (419), Expect = 3e-40

Figure A.2.3b. Alignment search result for cDNA clone CB12 (Du-glu 1). Du-glu 1 exhibits closest homology to a glutathione *S*-transferase (Class Sigma) of *Anopheles gambiae* (African malaria mosquito) (Reiss and James, unpublished). Alignment performed by blastp (www.ncbi.nlm.nih.gov/BLAST/), + indicates conservative substitution.

1 69
MAPTLYMIAASPAV**R**SVTLCAKALNIELNEKEVNLLTGAHLTPEFL**K**LNP**Q**HTV*P*VLDDDGFVIVD**S**H**A**
[

70 138
INT**Y**LVAKYGKDDSLYPKDIKKRAIVDHRLHYDSSILFVRGLSISRSLLFEGEKEIAPKLLTALEEAYT
Mature chain

139 207
ITEKFLEKNNYIAGDQLTV*AD*FSFITSITSWNVFFNLKEAKYPKIVAWMKRMQALPYYKEANQVGLDQY

208 245
YAIMKDKVPK .STSYLLIATVYFFKISNFHF.IYYFKS
] [3'UTR]

Figure A.2.3c. Annotated predicted amino-acid sequence of cDNA clone A17. UTR; untranslated region. Amino acids in bold and underlined show high conservation in GST enzymes, amino acids underlined and in italics show conservation in GST 2 enzymes (Feng *et al.*, 1999).

```
Du-glu 2:  LYMIAASPAVRSVTLCAKALNIELNEKEVNLLTGAHLTPEFLKLNPQHTVPVLDDDGFVI 64
           LY I  SP VR+  L  KALN+    K VNL    HL+ E+LK NPQHTVP L++DG +I
Fly-glu :  LYGIDPSPPVRACLLTLKALNLPFEYKVVNLFAKEHLSEEYLKKNPQHTVPTLEEDGHLI 65

Du-glu 2:  VDSHAINTYLVAKYGKDDSLYPKDIKKRAIVDHRLHYDSSILFVRGL-SISRSLLFEGEK 123
            DSH I  YLV+KYGKDDSLYPKD+ KRA+VD R+++++ +LF  GL +I+  L F  +
Fly-glu :  WDSHPIMAYLVSKYGKDDSLYPKDLLKRAVVDQRMYFEAGVLFQGGLRNITAPLFFRNQT 125

Du-glu 2:  EIAPKLLTALEEAYTITEKFLEKNNYIAGDQLTVADFSFITSITSWNVFFNLKEAKYPKI 183
           +I    + ++ E+Y   E FL+ N Y+AGD LT+ADFS +TS+TS   F  + ++K+PK+
Fly-glu :  QIPQHQIDSIVESYGFLESFLKNNKYMAGDHLTIADFSIVTSVTSLVAFAEIDQSKFPKL 185

Du-glu 2:  VAWMKRMQALPYYKEANQVGLDQYYAIMKDK 214
            AW+K +Q+LP+Y+EAN  G  Q  A++K K
Fly-Glu :  SAWLKSLQSLPFYEEANGAGAKQLVAMVKSK 216
```

Identities = 104/211 (49%), Positives = 145/211 (68%), Gaps = 1/211 (0%)
Score = 213 bits (543), Expect = 1e-54

Figure A.2.3d. Alignment search result for cDNA clone CB12 (Du-glu 2). Du-glu 2 exhibits closest homology to glutathione *S*-transferase 6B of *Musca domestica* (House fly) (Wei, unpublished). Alignment performed by blastp (www.ncbi.nlm.nih.gov/BLAST/), + indicates conservative substitution.

```
1                                                                   55
DIH—LAK MKVLVLFSLVALSVALPTLLGR TVLVKEDGTVTLFDSRGYQLVILKSV
[5'UTR] [    Leader sequence   ] [

56                                                                  112
TDPRKVELVLRGPNTRTQMFQVGESLRTREIIDRTLVGVSVGENYNQGDILADIFRQ
                    Mature sequence

113                                                                 170
YEGTLDDXKYHSLLNRIQLLVEAGVI***NET***IYDIIRNWDLEYRMQGVSDIVPVQGVRT

171
LHQYAEQVGD
          ]
```

Figure A.2.4a. Annotated predicted amino-acid sequence of the 5' sequences of cDNA clone A14. A potential *N*-glycosylation site is in bold, underlined and in italics, signal peptide identified by SignalP server (Nielsen and Krogh, 1998). UTR; untranslated region.

```
Du-enc:    MKVLVLFSLVALSVALPTLLGRTVLVKEDGTVTLFDSRGYQLVILKSVTDPR---KVELV 57
           MK L+L S V    A      R + V +DG++ +    G ++VI +  T P     +++
Te-enc:    MKFLLLASFVVFCTASTVYRNRNISVNDDGSILITSPTGRRVVISRP-TGPSGQGNLDIS 59

Du-enc:    LRGPNTRTQMFQVGE-SLRTREIIDRTLVGV-SVGENY----------NQGDILADIFRQ 105
           + GPN  T+  Q+ + +L T    +     V S+ E +           QGDIL +I R+
Te-enc:    VAGPNIPTKRIQINDNALYTPGTGEGPWSSVDSLSEEWRSKRGSLKGRTQGDILDEILRE 119

Du-enc:    YEGTLDDXKYHSLLNRIQLLVEAGVINETIYDIIRNWD 143
           Y+G +D+ +Y  LLNRI   V  GV++  IYD++   D
Te-enc:    YQGVVDEAQYEKLLNRINKAVRNGVVSPQIYDLLAGLD 157
```

Identities = 49/158 (31%), Positives = 79/158 (49%), Gaps = 16/158 (10%)
Score = 61.7 bits (149), Expect = 5e-09

Figure A.2.4b. Alignment search result for cDNA clone A14 (Du-enc). Du-enc exhibits closest homology to a 56 kDa early stage encapsulation response protein (Cho *et al.*, 1999). Alignment performed by blastp (www.ncbi.nlm.nih.gov/BLAST/), + indicates conservative substitution.

acid code identity to a serine type carboxypeptidase isolated from *Aedes aegypti* (see figure A.2.1b.). The clone appears to encode a full length carboxypeptidase transcript including 5'untranslated region and translation start methionine. The mosquito carboxypeptidase is specifically expressed in fat body tissue 24-48 hours post food ingestion (Cho *et al.*, 1991). Pro-peptide Du-carb encodes a 446 amino acid protein containing one potential *N*-glycosylation site at Asn_{169} and a molecular weight of 50.7 kDa, GRAVY value of –0.490 and theoretical pI of pH 4.88.

The presence of 2 of 3 conserved sequences (W_{132}LQGGPGST) and (G_{223}ESYGG) (Galijart *et al.*, 1990) including the catalytic Ser_{225} and overall homology, indicates clone A4 encodes a serine type carboxypeptidase. Du-carb represents the first reported serine type carboxypeptidase sequence isolated from a Coleopteran insect. Fat body specific expression, induced by diet ingestion was found in the mosquito carboxypeptidase sequence (Cho *et al.*, 1991). A similar pattern of fat body expression induced by diet ingestion expression may occur for the *D. undecimpunctata* serine type carboxypeptidase.

Carboxypeptidase activity has been found in the endoperitrophic, associated with the peritrophic membrane and ectoperitrophic spaces in Lepidoptera (Bown *et al.*, 1998). This research indicates carboxypeptidases are distributed throughout the alimentary tract, digesting polypeptides and releasing individual amino acids for direct absorption by midgut epithelial cells. Further research (Ortego *et al.*, 1996) indicates compartmentalisation of carboxypeptidase activity, carboxypeptidase A being found in the midgut epithelium cells and carboxypeptidase B being located in the endoperitrophic

space in Lepidoptera. This may reflect differential localisation of carboxypeptidase activity in differing insect species.

Trypsin digested potato multicystatin (T-PMC) was 5-fold less effective against second instar *D. undecimpunctata* compared to neonate larvae (Orr *et al*, 1994). The same research showed co-feeding potato carboxypeptidase inhibitor (PCI) with T-PMC restored larval growth inhibition, indicating carboxypeptidase mediated digestion of PMC fragments. Du-carb, by isolation from alimentary tract material may represent the carboxypeptidase activity responsible for T-PMC digestion further research is required to test this hypothesis.

A carboxypeptidase inhibitor such as PCI would theoretically have a deleterious affect on larval development. Inhibition of carboxypeptidase activity would be expected to reduce digestion of peptide fragments to free amino-acids for absorption by midgut cells, thereby reducing the free amino acid pool available to the insect. Significant mortality would not be expected however as only 5 % of overall proteolysis in *D. undecimpunctata* midgut larval extracts is inhibited by carboxypeptidase inhibitor PCI at 1.59 μM (Edmonds *et al.*, 1996).

A.2.2. Results: *D. undecimpunctata* chitinase (EC 3.2.1.14).

Isolation and characterisation of cDNA clone CA5 (Du-chit) (see figure A.2.2a.) revealed 56 % amino acid identity (see figure A.2.2b.) to a gut-specific chitinase identified in *Phaedon cochleariae* (Coleoptera) (Girard and Jouanin, 1999a). The *D. undecimpunctata* chitinase sequence possesses a conserved region identified in insect chitinase enzymes F_{149}DGFDLDWEYP and a putative catalytic Trp_{156} (Huang *et al.*,

2000). The presence of these sequences provides further evidence that the obtained clone encodes a *D. undecimpunctata* chitinolytic enzyme.

The CA5 cDNA clone sequence includes a 5' untranslated sequence, translation start methionine and mature protein. Du-chit encodes a 372 amino acid protein with a molecular weight of 41.6 kDa, GRAVY value of –0.283 and theoretical pI of pH 4.61. Du-chit possesses four potential *N*-glycosylation sites Asn_{291}, Asn_{165}, Asn_{251}, and Asn_{325} the former three of these sites show positional homology to potential *N*-glycosylation sites possessed by the gut-specific chitinase isolated from *Phaedon cochleariae* (see figure A.2.2b). Du-chit lacks C-terminal serine/threonine and cysteine-rich domains possessed by chitinases showing expression in regions other than the insect midgut and also lacked by the gut-specific chitinase of *Phaedon cochleariae* (Girard and Jouanin, 1999).

A putative role in peritrophic membrane regulation by chitin degradation is proposed for this enzyme by homology, active site presence and tissue type of isolation. *D. undecimpunctata* possess a type 1 peritrophic membrane secreted along the length of the alimentary canal (Ryerse *et al.*, 1994) as opposed to type 2 which are secreted by specialised cardia. The use and potential of chitinases as biopesticides has been discussed in the introduction to this work and has been extensively reviewed (Jongsma and Muthukrishnan, 1997).

A.2.3. Results: *D. undecimpunctata* glutathione *S*-transferases and P450 mixed function oxidases (EC 2.5.1.18 and EC 1.14).

Isolation and characterisation of cDNA clone CB12 (Du-glu 1) (see figure A.2.3a.) revealed close 41 % amino acid code identity (see figure A.2.3b.) to a Sigma class glutathione *S*-transferase identified in *Anopheles gambiae* (Diptera) (Reiss and James, unpublished). The Du-glu 1 sequence possesses a 5' untranslated region, a single *N*-glycosylation position at Asn_{11} and encodes a 218 amino acid, 25.0 kDa protein with a theoretical pI of pH 6.2. Comparison of Du-glu 1with aligned insect GST 1 and 2 GST's (Feng *et al.*, 1999) indicates by amino acid conservation that the protein is a class 2 insect GST.

Clone A17 (Du-glu 2) (see figure A.2.3c.) possesses 49 % amino acid code identity (see figure A.2.3d.) to a glutathione *S*-transferase 6B isolated from *Musca domestica* (House fly) (Wei, unpublished). Clone Du-glu 2 lacks a 5' untranslated region and encodes a 217 amino acid 24.6 kDa protein with a theoretical pI of pH 6.97. Comparison of Du-glu 1with aligned insect GST 1 and 2 GST's (Feng *et al.*, 1999) indicates by amino acid conservation that the protein is also a class 2 insect GST.

Two glutathione *S*-transferases and a P450 mixed function oxidase have been isolated from the *D. undecimpunctata* alimentary tract cDNA library. The P450 mixed function oxidase 5' sequence is not presented in this work due to the quality of sequencing return, this may be due to insufficient quantity or quality of plasmid DNA due to the pTriplEx2/BM 25.8 host combination employed. These enzyme families perform multiple functions in the cellular environment, including regulating the

endogenous cellular processes and detoxification of xenobiotics (discussed previously). High expression levels of these would be expected in alimentary tract epithelial cells to detoxify xenobiotics and pesticides such as allelochemicals ingested with the plant diet.

A.2.4. Results: *D. undecimpunctata* early stage encapsulation response protein.

The *D. undecimpunctata* clone A14 encodes a putative full length transcript of approximately 1.6 kb (see figure A.2a.). Sequence data is only available for the 5' terminal of the obtained cDNA. The fragment sequenced shows 31 % amino acid code identity to the 5' terminal of a 56 kDa early stage encapsulation response protein (ESERP) isolated from *Tenebrio molitor* (Cho *et al.*, 1999) (see figure A.2.4b). The sequenced fragment includes a 5' untranslated region, leader sequence and 150 amino acids of the 5' propeptide and includes a single potential *N*-glycosylation site (see figure A.2.4a).

Phagocytosis and encapsulation are a part of the insect cellular defence immune system against foreign bacteria, viral particles and parasitoid eggs. Foreign bodies too large for individual hemocytes to phagocytose are encapsulated by hemocyte cells. Encapsulation of foreign targets involves a pathway initiated by recognition of foreign targets by adhesion molecules, that then trigger capsule formation by hemocytes. Microscopic studies indicate the end immune response is encasement in multilayered, overlapping hemocytes forming a capsule surrounding the foreign body (Pech and Micahel, 1996).

Immuno-blotting localised the 56 kDa ESERP to hemocyte membranes (Cho *et al.*, 1999). The encapsulation response of *Tenebrio molitor* (Coleoptera) to DEAE-Sepharose beads was shown to be inhibited by addition of antibodies to the 56 kDa and a 48 kDa encapsulation response proteins (Cho *et al*, 1999). This research indicates the protein is a hemocyte membrane receptor active in a signal recognition pathway inducing encapsulation of foreign bodies as part of the insect immune response.

A.2.5. Discussion.

A range of cDNA clones have been obtained from the *D. undecimpunctata* alimentary tract cDNA library, sequenced and identified by Blast search. Initial characterisation has identified transcription of genes putatively involved in peritrophic membrane degradation, xenobiotic detoxification, immune response to invasive pathogens and proteolysis. Four of these clones encode proteins involved directly or indirectly with the protection of the insect tissue from damage. High expression of genes performing a tissue defensive role would be expected in the alimentary tract of insects physiologically a major area of pathogen and xenobiotic entry into the insect body.

Clones CA6, CB4, CA9 and A18 contained clones encoding open reading frames with start codons (see appendix A.3.4) however the putative function of these sequences was unassignable by Blast homology search employing the 5' sequence ends of the clones. Further sequencing from the 3' direction is needed to obtain the entire cloned sequences to enable further identification of the clones obtained and to obtain reliable sequence data of the putative *D. undecimpunctata* P450 mixed function oxidase.

A.3.1. Nucleotide and predicted amino acid sequences of aspartic proteinase encoding cDNA clones.

```
gacttcatcatgatgtacgctaagtttttggttgtgttggtggtagctgctctagctagc
 D  F  I  M  M  Y  A  K  F  L  V  V  L  V  V  A  A  L  A  S
agtgaaattatcagaattcctttgaaaaaagtacctgataatggacaaagaaggttgcgt
 S  E  I  I  R  I  P  L  K  K  V  P  D  N  G  Q  R  R  L  R
aatgtagccagaagcggactacgcaataaatttggaggacaaggagcagtacctttagtc
 N  V  A  R  S  G  L  R  N  K  F  G  G  Q  G  A  V  P  L  V
aacgaatatgaccttgattattatggagaaatcggtgtaggaacacctgcacaaaaattc
 N  E  Y  D  L  D  Y  Y  G  E  I  G  V  G  T  P  A  Q  K  F
aacgtcatttttgatactggatctaccgatctttgggtcccttcaagcaaatgccaagcc
 N  V  I  F  D  T  G  S  T  D  L  W  V  P  S  S  K  C  Q  A
acctccgcacaaaacggttgcaaaaaccataaacaatacgatgctagcaaatcttcaaca
 T  S  A  Q  N  G  C  K  N  H  K  Q  Y  D  A  S  K  S  S  T
taccagaaagatggcagatccttccacattgaatacggatctggtgctttgagtggtttc
 Y  Q  K  D  G  R  S  F  H  I  E  Y  G  S  G  A  L  S  G  F
ttatctaaagacagcgtcgaggttgccggtcttgctataaccgaccaaacttttgctgaa
 L  S  K  D  S  V  E  V  A  G  L  A  I  T  D  Q  T  F  A  E
gccactgtcgaaactgccccagctttcgtaagcagcacattcgacggtatcttaggattg
 A  T  V  E  T  A  P  A  F  V  S  S  T  F  D  G  I  L  G  L
gcttatccagatctgtcagcacaaaacattcctccagttttccaaaacatgatcaaacaa
 A  Y  P  D  L  S  A  Q  N  I  P  P  V  F  Q  N  M  I  K  Q
ggagttcttgacgcaccagtattttccttttatctgagccaaactgccaatggtgacaaa
 G  V  L  D  A  P  V  F  S  F  Y  L  S  Q  T  A  N  G  D  K
ggagagctgttgctcggtggatcagattcaaaatactacaaaggagatttcacctacaca
 G  E  L  L  L  G  G  S  D  S  K  Y  Y  K  G  D  F  T  Y  T
aaagtatcaacacaactttattggcaaaccaatttacaaggtgtttctgtaggcagtcgt
 K  V  S  T  Q  L  Y  W  Q  T  N  L  Q  G  V  S  V  G  S  R
agtgtatgcaaatctggatgtgaagctgtcatagatactggcacttccctcatctacgga
 S  V  C  K  S  G  C  E  A  V  I  D  T  G  T  S  L  I  Y  G
ccaaccgatgacgttgacgttgttaactcggctattggtgccacatatgactactctgtt
 P  T  D  D  V  D  V  V  N  S  A  I  G  A  T  Y  D  Y  S  V
ggtctgtacaccgttaattgcaacaccgatttaaacaaacttcctaatgtatcatttact
 G  L  Y  T  V  N  C  N  T  D  L  N  K  L  P  N  V  S  F  T
ttcggaggcaagaaatttgacattcctgcttcagcgtacatcatcaaagactctggttat
 F  G  G  K  K  F  D  I  P  A  S  A  Y  I  I  K  D  S  G  Y
tgcatctccagtttcgtcgcccaggagttcttcctcagtggatttgaatggttggttgga
 C  I  S  S  F  V  A  Q  E  F  F  L  S  G  F  E  W  L  V  G
gattcattcctgaaaaccgtttacagcgagttcgattttggcaacaaccgtattggattt
 D  S  F  L  K  T  V  Y  S  E  F  D  F  G  N  N  R  I  G  F
gctgaacttgcttaaataattataactttatcaatattttattattaaatattgctgtag
 A  E  L  A  -  I  I  I  T  L  S  I  F  Y  Y  -  I  L  L  -
```

Figure A.3.1a. Nucleotide and predicted amino acid sequence of *D. undecimpunctata* aspartic proteinase 1 (clones A72, A62, A42 and C52).

```
attgttccttgtattgaagaacagggtagcgtcaaaatgttatccaaattattaatgtta
 I  V  P  C  I  E  E  Q  G  S  V  K  M  L  S  K  L  L  M  L
gtatttgctgcagctgttgccactgctgaaatcgctagaataccactgaaaaagtttgaa
 V  F  A  A  A  V  A  T  A  E  I  A  R  I  P  L  K  K  F  E
aagccgagggattcacttaacagaaatggactaaaagaaacattattagccaaatacagt
 K  P  R  D  S  L  N  R  N  G  L  K  E  T  L  L  A  K  Y  S
ggaaatttcaatgacgatagcgctgtcgttttaactaactatttagatggaacgatcggc
 G  N  F  N  D  D  S  A  V  V  L  T  N  Y  L  D  G  T  I  G
cgatttttgttaagagctcaatattatggagaaattggtattggaactcctgctcaaagc
 R  F  L  L  R  A  Q  Y  Y  G  E  I  G  I  G  T  P  A  Q  S
tttaatgtaattttcgacactggttcatctaatctctgggtaccctccagcaagtgtgga
 F  N  V  I  F  D  T  G  S  S  N  L  W  V  P  S  S  K  C  G
tttttggaagttgcttgtttacttcataacaaatacgattccgataaatcttctacctat
 F  L  E  V  A  C  L  L  H  N  K  Y  D  S  D  K  S  S  T  Y
gtaaaaaacgacaccagattcgcaattagatatggttctggagatgttgcaggggttgtt
 V  K  N  D  T  R  F  A  I  R  Y  G  S  G  D  V  A  G  V  V
tcccaagatgtggttgaggttgggggtttacaagctaaggatcaactatttgctgaagct
 S  Q  D  V  V  E  V  G  G  L  Q  A  K  D  Q  L  F  A  E  A
acccaggaacccggtttagctttccttgtgggaaaattcgatgggattttaggaatgggg
 T  Q  E  P  G  L  A  F  L  V  G  K  F  D  G  I  L  G  M  G
taccccgaaatttccgttaatggaattactccagtatttaatacactggtagaacaaggt
 Y  P  E  I  S  V  N  G  I  T  P  V  F  N  T  L  V  E  Q  G
gctgttaaagaacccgttttctccttttatttgaacagagaccctgatggagacgttggt
 A  V  K  E  P  V  F  S  F  Y  L  N  R  D  P  D  G  D  V  G
ggtgaattgttgctgggtggatcagactcaaactattacaaaggggactttacttatata
 G  E  L  L  L  G  G  S  D  S  N  Y  Y  K  G  D  F  T  Y  I
gatgtttcagctaaaggctattggcaaattgttatggacagtcttaacgtgggatcttcc
 D  V  S  A  K  G  Y  W  Q  I  V  M  D  S  L  N  V  G  S  S
ttaaaactctgttcaaatggatgtcaagttatcgtggacaccggtaccagtttgatcgcg
 L  K  L  C  S  N  G  C  Q  V  I  V  D  T  G  T  S  L  I  A
ggacctagcgccgaagtcgaaaagttacatcaagaaataggagctttttcatttcttaat
 G  P  S  A  E  V  E  K  L  H  Q  E  I  G  A  F  S  F  L  N
ggtgaatatataattgattgtaataaagttgatcaactacccgaaattagttttgttttc
 G  E  Y  I  I  D  C  N  K  V  D  Q  L  P  E  I  S  F  V  F
ggaggaaaagaattcaaactttctggaaacgattacattttaaagcaatctaatggtctt
 G  G  K  E  F  K  L  S  G  N  D  Y  I  L  K  Q  S  N  G  L
attgacatctgtatttctggatttatgggattagacctcgacactagaactcatgtggaa
 I  D  I  C  I  S  G  F  M  G  L  D  L  D  T  R  T  H  V  E
tggatcctaggagatgtttttattggaaaattctacacagaatttgattttggtaataat
 W  I  L  G  D  V  F  I  G  K  F  Y  T  E  F  D  F  G  N  N
cgtgtaggattagctgaagcagtttaaatatacctaataaaaagtcaataaa
 R  V  G  L  A  E  A  V  -  I  Y  L  I  K  S  Q  -
```

Figure A.3.1b. Nucleotide and predicted amino acid sequence of *D. undecimpunctata* aspartic proteinase 2 (cDNA clone A16).

```
gttccttgtattgaagaacagggtagcgtcaaaatgttatccaaattattaatgttagta
 V  P  C  I  E  E  Q  G  S  V  K  M  L  S  K  L  L  M  L  V
tttgctgcagctgttgccactgctgaaatcgctagaataccactgaaaaagtttgaaaag
 F  A  A  A  V  A  T  A  E  I  A  R  I  P  L  K  K  F  E  K
ccgagggattcacttaacagaaatggactaaaagaaacattattagccaaatacagtgga
 P  R  D  S  L  N  R  N  G  L  K  E  T  L  L  A  K  Y  S  G
aatttcaatgacgatagcgctgtcgttttaactaactatttagatgctcaatattatgga
 N  F  N  D  D  S  A  V  V  L  T  N  Y  L  D  A  Q  Y  Y  G
gaaattggtattggaactcctgctcaaagctttaatgtaattttcgacactggttcatct
 E  I  G  I  G  T  P  A  Q  S  F  N  V  I  F  D  T  G  S  S
aatctctgggtaccctccagcaagtgtggatttttggaagttgcttgtttacttcataac
 N  L  W  V  P  S  S  K  C  G  F  L  E  V  A  C  L  L  H  N
aaatacgattccgataaatcttctacctatgtaaaaaacgacaccagattcgcaattaga
 K  Y  D  S  D  K  S  S  T  Y  V  K  N  D  T  R  F  A  I  R
tatggttctggagatgttgcaggggttgtttcccaagatgtggttgaggttgggggttta
 Y  G  S  G  D  V  A  G  V  V  S  Q  D  V  V  E  V  G  G  L
caagctaaggatcaactatttgctgaagctacccaggaacccggtttagctttccttgtg
 Q  A  K  D  Q  L  F  A  E  A  T  Q  E  P  G  L  A  F  L  V
ggaaaattcgatgggattttaggaatggggtaccccgaaatttccgttaatggaattact
 G  K  F  D  G  I  L  G  M  G  Y  P  E  I  S  V  N  G  I  T
ccagtatttaatacactggtagaacaaggtgctattaaagaacccgttttctccttttat
 P  V  F  N  T  L  V  E  Q  G  A  I  K  E  P  V  F  S  F  Y
ttgaacagagaccctgatggagacgttggtggtgaattgttgctgggtggatcagactca
 L  N  R  D  P  D  G  D  V  G  G  E  L  L  L  G  G  S  D  S
aactattacaaaggggactttacttatatagatgtttcagctaaaggctattggcaaatt
 N  Y  Y  K  G  D  F  T  Y  I  D  V  S  A  K  G  Y  W  Q  I
gttatggacagtcttagcgtgggatcttccttaaaactctgttcaaatggatgtcaagtt
 V  M  D  S  L  S  V  G  S  S  L  K  L  C  S  N  G  C  Q  V
atcgtggacaccggtaccagtttgatcgcgggacctagcgccgaagtcgaaaagttacat
 I  V  D  T  G  T  S  L  I  A  G  P  S  A  E  V  E  K  L  H
caagaaataggagctttttcatttcttaatggtgaatatataattgattgtaatagagtt
 Q  E  I  G  A  F  S  F  L  N  G  E  Y  I  I  D  C  N  R  V
gatcaactacccgaaattagttttgttttcggaggaaaagaattcaaactttctggaaac
 D  Q  L  P  E  I  S  F  V  F  G  G  K  E  F  K  L  S  G  N
gactacattttaaagcaatctaatggtcttattgacatctgtatttctggatccatggga
 D  Y  I  L  K  Q  S  N  G  L  I  D  I  C  I  S  G  S  M  G
ttagacctcgacactagaactcatgtggaatggatcctaggagatgtttttattggaaaa
 L  D  L  D  T  R  T  H  V  E  W  I  L  G  D  V  F  I  G  K
ttctacacagaatttgattttggtaataatcgtgtaggattagctgaagcagtttaaata
 F  Y  T  E  F  D  F  G  N  N  R  V  G  L  A  E  A  V  -  I
tacctaaaaagggtcata
 Y  L  K  R  V  I
```

Figure A.3.1c. Nucleotide and predicted amino acid sequence of *D. undecimpunctata* aspartic proteinase 2a (cDNA clone A20).

A.3.2. Nucleotide and predicted amino acid sequences of cysteine proteinase encoding cDNA clones.

```
gaacgtattcaattttgttataaaatggtttcccagtttttaaaaattaagatgaagttt
 E  R  I  Q  F  C  Y  K  M  V  S  Q  F  L  K  I  K  M  K  F
ctaatattaattttcctgacaactgccgctatccaggcattaaacgataaagaaaaatgg
 L  I  L  I  F  L  T  T  A  A  I  Q  A  L  N  D  K  E  K  W
gtacaatttaaggttaaacataacaaaagttacataaactacgttgaagaacagacccgt
 V  Q  F  K  V  K  H  N  K  S  Y  I  N  Y  V  E  E  Q  T  R
ttcacaatttttcaagaaaatctgaaaaaaatcgaaaatcacaataaaaaatattacaat
 F  T  I  F  Q  E  N  L  K  K  I  E  N  H  N  K  K  Y  Y  N
ggagagtcttcttttatgttgggtgttaccaaatttgctgacttgaccgaaaaagaattt
 G  E  S  S  F  M  L  G  V  T  K  F  A  D  L  T  E  K  E  F
atggacatgttggtgctttctaaaaattcaaggcctaacagacctcatgttacacactta
 M  D  M  L  V  L  S  K  N  S  R  P  N  R  P  H  V  T  H  L
ctagctccactaggagatctaccttcagcatttgattggagaaataagggagctgtaaca
 L  A  P  L  G  D  L  P  S  A  F  D  W  R  N  K  G  A  V  T
gaggtaaaagatcagggaatgtgtggctcttgttggaccttcagtacaactggatcagta
 E  V  K  D  Q  G  M  C  G  S  C  W  T  F  S  T  T  G  S  V
gaaggagcacatttccttaaaactggaaatccggtatctttaagtgaacaaaatctagtt
 E  G  A  H  F  L  K  T  G  N  P  V  S  L  S  E  Q  N  L  V
gattgcgctaatgacaattgttatggatgcggaggaggttggatggacaaagctctggaa
 D  C  A  N  D  N  C  Y  G  C  G  G  G  W  M  D  K  A  L  E
tatattgaaaaaggggaataatgtcagaaaaggattatccttacgaaggcgtggacgat
 Y  I  E  K  G  G  I  M  S  E  K  D  Y  P  Y  E  G  V  D  D
aaatgtagatttgatgtttccaaagttgctgccaagatcagcaactttacttatattaag
 K  C  R  F  D  V  S  K  V  A  A  K  I  S  N  F  T  Y  I  K
aaaaatgatgaagaagatcttaaaaacgcagtttttgcaaaaggtcctatctccgtagct
 K  N  D  E  E  D  L  K  N  A  V  F  A  K  G  P  I  S  V  A
attgatgcttcttctaagtttcagctatatgtggcaggaatacttgatgatacggaatgc
 I  D  A  S  S  K  F  Q  L  Y  V  A  G  I  L  D  D  T  E  C
tctaatgaatttgactcattgaatcatggtgtacttgttgttggctatggtacagaaaat
 S  N  E  F  D  S  L  N  H  G  V  L  V  V  G  Y  G  T  E  N
ggaaaagattactggatcgtcaaaaactcctggggagtaaactggggaatggacggatat
 G  K  D  Y  W  I  V  K  N  S  W  G  V  N  W  G  M  D  G  Y
attcggatgagcagaaataagaataaccaatgtggaattactactgacggcgtatatcct
 I  R  M  S  R  N  K  N  N  Q  C  G  I  T  T  D  G  V  Y  P
aacgtttaaagacatgcttttagaatttttcattcagatgttttactgatttgatagttt
 N  V  -  R  H  A  F  R  I  F  H  S  D  V  L  L  I  -  -  F
taaaataatttttttagtgattagattattctttatttaaaataaagtagaa
 -  N  N  F  L  V  I  R  L  F  F  I  -  N  K  V  E
```

Figure A.3.2a. Nucleotide and predicted amino acid sequence of *D. undecimpunctata* cysteine proteinase 1 (cDNA clone CA7).

```
aaatcatttgttgtcagaccttcaacgtaatagttttgttaaataggatacatcatgaag
 K  S  F  V  V  R  P  S  T  -  -  F  C  -  I  G  Y  I  M  K
tttttgctacttttttgttgcatttttttgtgggcagccaggctatttcctttgttgatttg
 F  L  L  L  F  V  A  F  F  V  G  S  Q  A  I  S  F  V  D  L
gtccaaggagaatggacagctttcaagatgacacacagaaaaagttatgaaagcccaacc
 V  Q  G  E  W  T  A  F  K  M  T  H  R  K  S  Y  E  S  P  T
gaggaaaagttccgtatgaagatctttatggaaaattcacacaaagttgccaaacacaac
 E  E  K  F  R  M  K  I  F  M  E  N  S  H  K  V  A  K  H  N
aagctcttcgctttgggattagtttcttataaattaggactcaacaaatacgccgatatg
 K  L  F  A  L  G  L  V  S  Y  K  L  G  L  N  K  Y  A  D  M
ctccaccatgaattcatcgccaccctcaacggcttcaacaaaacaaagaatatgttgagg
 L  H  H  E  F  I  A  T  L  N  G  F  N  K  T  K  N  M  L  R
cagtcagagctagaagactcggtcacattcataaaacccgctaacgtagagttacccggc
 Q  S  E  L  E  D  S  V  T  F  I  K  P  A  N  V  E  L  P  G
gaagttgactggagaccaaagggtgccgttactggtgttaaggaccaagggcattgcgga
 E  V  D  W  R  P  K  G  A  V  T  G  V  K  D  Q  G  H  C  G
tcgtgctggagtttcagtgctactggatctctcgaaggacaacattttaggaaaaccgga
 S  C  W  S  F  S  A  T  G  S  L  E  G  Q  H  F  R  K  T  G
aaattggtctccctttccgaacaaaacttagtcgactgctctactaaattcggcaacaac
 K  L  V  S  L  S  E  Q  N  L  V  D  C  S  T  K  F  G  N  N
ggatgcaatggaggtttaatggataatgccttccgttatattaaagccaacggtggtatc
 G  C  N  G  G  L  M  D  N  A  F  R  Y  I  K  A  N  G  G  I
gacactgaagcaagttataaatacaaggccgaagacgagaagtgccattacaatcccaag
 D  T  E  A  S  Y  K  Y  K  A  E  D  E  K  C  H  Y  N  P  K
aaaataggagccactgacaaaggtttcgtagatatcgaatctggaaacgaagatgccctc
 K  I  G  A  T  D  K  G  F  V  D  I  E  S  G  N  E  D  A  L
caggccgctgtcgccaccatcggacccgtatccgtagccattgacgctagccacgagacc
 Q  A  A  V  A  T  I  G  P  V  S  V  A  I  D  A  S  H  E  T
ttccaactctacaatcagggtgtatattatgaacctgaatgcagctcatacgaattagat
 F  Q  L  Y  N  Q  G  V  Y  Y  E  P  E  C  S  S  Y  E  L  D
cacggtgtcctcgttgtcggttatggtactgaaaacaaccaagactattggctcgtcaag
 H  G  V  L  V  V  G  Y  G  T  E  N  N  Q  D  Y  W  L  V  K
aattcctggggtccatcttggggtctagacggttacatcaaaatggccagaaacaggaac
 N  S  W  G  P  S  W  G  L  D  G  Y  I  K  M  A  R  N  R  N
aactcctgtggtattgcaactcaagccagctatcctttagtttaagtttatagattaata
 N  S  C  G  I  A  T  Q  A  S  Y  P  L  V  -  V  Y  R  L  I
taatgttcagtgtaaacaaaaa
 -  C  S  V  -  T  K
```

Figure A.3.2b. Nucleotide and predicted amino acid sequence of *D. undecimpunctata* cysteine proteinase 2 (cDNA clone C21).

```
caatggaccagttttaaggcaacacacaacaaatcctactccgttgttgaggacaaactt
 Q  W  T  S  F  K  A  T  H  N  K  S  Y  S  V  V  E  D  K  L
cgtttcgcagtcttccacgaaaacttgcgcaaaattgaagaacacaatgccaaatacgag
 R  F  A  V  F  H  E  N  L  R  K  I  E  E  H  N  A  K  Y  E
aacggagaagagacctactaccttgccgtcaatcaatttgccgattggtccagcgctgaa
 N  G  E  E  T  Y  Y  L  A  V  N  Q  F  A  D  W  S  S  A  E
tttaaagctttgttgaattctcagatgattaatagaccagaactatcctttatcgaaaca
 F  K  A  L  L  N  S  Q  M  I  N  R  P  E  L  S  F  I  E  T
ttcgaagcagatcccaacttaaaagcagattctgttgactggagaaataaagctgatttg
 F  E  A  D  P  N  L  K  A  D  S  V  D  W  R  N  K  A  D  L
ggagtcaaaaatcaaggaagctgtggctcttgctgggcttttagtgccaccggagccctc
 G  V  K  N  Q  G  S  C  G  S  C  W  A  F  S  A  T  G  A  L
gaaggtcaacttgctattcacaaaaatcaacatgttcaactgagtgagcaagaattggta
 E  G  Q  L  A  I  H  K  N  Q  H  V  Q  L  S  E  Q  E  L  V
gactgtgatacgacaaactctggttgtaatggtggtttaatgacaaatgcctttgcctat
 D  C  D  T  T  N  S  G  C  N  G  G  L  M  T  N  A  F  A  Y
gttagaagccacggtcttgcatcagaaaaacaatatgcatacacagctagagacggtagt
 V  R  S  H  G  L  A  S  E  K  Q  Y  A  Y  T  A  R  D  G  S
tgcaagaaagtacaaaacaaacaagtctcttccatcagcggatacgtaaatgtcgccaaa
 C  K  K  V  Q  N  K  Q  V  S  S  I  S  G  Y  V  N  V  A  K
actgaaagtgcattggctagtgctcttgctagtgtaggtccaatatccattgctgtcgat
 T  E  S  A  L  A  S  A  L  A  S  V  G  P  I  S  I  A  V  D
gcagacacatggcagttttatggaggtggaattttcaacaacaaaaattgtggaaccacc
 A  D  T  W  Q  F  Y  G  G  G  I  F  N  N  K  N  C  G  T  T
cttaaccacggagttcttgctgtgggatacactaaagatgtcttcatcgttaaaaattca
 L  N  H  G  V  L  A  V  G  Y  T  K  D  V  F  I  V  K  N  S
tggggaaccagttggggtgaacttggttatatcagaattagccgtggtcacaacttatgc
 W  G  T  S  W  G  E  L  G  Y  I  R  I  S  R  G  H  N  L  C
ggtcttaaccaaatgaacagttaccctaaattgtaaatgataataaataaatgatatata
 G  L  N  Q  M  N  S  Y  P  K  L  -  M  I  I  N  K  -  Y  I
tg
```

Figure A.3.2c. Nucleotide and predicted amino acid sequence of *D. undecimpunctata* cysteine proteinase 3 (cDNA clone CB13).

```
tttcaaaaaaatcttcgtaaaattgttgaacataattcaaaatatgagaagggagaaaag
 F  Q  K  N  L  R  K  I  V  E  H  N  S  K  Y  E  K  G  E  K
gcttactttcttaaaataaccaaatttgcggactggaccgataaagaattgaatgctata
 A  Y  F  L  K  I  T  K  F  A  D  W  T  D  K  E  L  N  A  I
ttaaacccaaaaatagttgctaaagcacagcataaaaacaccaaaacgttcgtacgtgat
 L  N  P  K  I  V  A  K  A  Q  H  K  N  T  K  T  F  V  R  D
ccaaacttaactcgtccagcgagcatcgactggagagacaaagctgttttggctgtaaaa
 P  N  L  T  R  P  A  S  I  D  W  R  D  K  A  V  L  A  V  K
gatcaggcaaactgtggttcttgttgggcatttagtactacaggtgcacttgaaggtcaa
 D  Q  A  N  C  G  S  C  W  A  F  S  T  T  G  A  L  E  G  Q
cttgccattcataaaaatcaggcaataccttaagcgaacaagaattaatggactgtgaa
 L  A  I  H  K  N  Q  A  I  P  L  S  E  Q  E  L  M  D  C  E
tctgggaatagcgcatgctttggaggtaatcctgacgtcgcatttgaatacattgagtca
 S  G  N  S  A  C  F  G  G  N  P  D  V  A  F  E  Y  I  E  S
aatggtattagttcagaaagtcaatacgagtatacacaacaaaaaggtgaatgtagaaaa
 N  G  I  S  S  E  S  Q  Y  E  Y  T  Q  Q  K  G  E  C  R  K
gtggaaaataaaccagtgtctagcatttcaggatggttgggagtaccttcagatgaagat
 V  E  N  K  P  V  S  S  I  S  G  W  L  G  V  P  S  D  E  D
gcacttatggaagctgttgctcaatatggtccagtatctgtttctgtgtttgctaacaac
 A  L  M  E  A  V  A  Q  Y  G  P  V  S  V  S  V  F  A  N  N
gattggagtttgtatggaggtggcatatttgaacacgcaagctgcagaggacatcctaat
 D  W  S  L  Y  G  G  G  I  F  E  H  A  S  C  R  G  H  P  N
caygctgtgttggctgttggctatacacaaaaaagttggatagttaaaaattcctgggga
 X  A  V  L  A  V  G  Y  T  Q  K  S  W  I  V  K  N  S  W  G
gcagcttggggagaagatggctacattcaattatccctcggtaataatcaatgcaatatt
 A  A  W  G  E  D  G  Y  I  Q  L  S  L  G  N  N  Q  C  N  I
actttcgccagtcaaattcccttattataaacgactaaatatttgtttaaataaaacatt
 T  F  A  S  Q  I  P  L  L  -  T  T  K  Y  L  F  K  -  N  I
tttatgctaaaacgttgtttttatcagtttttaggtttttatgtatttgtactgtatat
 F  M  L  K  R  C  F  L  S  V  F  R  F  L  C  I  C  T  V  Y
ttgtataaacgtgaaaaaaaaaaaaaaaaaaaaaaaaaaaa
 L  Y  K  R  E  K  K  K  K  K  K  K  K  K
```

Figure A.3.2d. Nucleotide and predicted amino acid sequence of *D. undecimpunctata* cysteine proteinase 4 (cDNA clone CB8).

A.3.3. Nucleotide and predicted amino acid sequences of serine type carboxypeptidase, chitinase, early stage encapsulation response protein and glutathione S-transferase encoding cDNA clones.

```
actattattccaggtatctttaaagcagtgacatagaagtgttaggctttaaagatgaag
 T  I  I  P  G  I  F  K  A  V  T  -  K  C  -  A  L  K  M  K
gctattctaagtattagttttgtacttttatactttgttcaatagaatcaaagtctctt
 A  I  L  S  I  S  F  V  L  F  I  L  C  S  I  E  S  K  S  L
aatgtcaagcattctagaaagccaatatttaaagacttatacggaaaactaaccccggaa
 N  V  K  H  S  R  K  P  I  F  K  D  L  Y  G  K  L  T  P  E
gtaagaggaaatcctggagagcccttgatattaactgacttaattaaagcaggaaaactg
 V  R  G  N  P  G  E  P  L  I  L  T  D  L  I  K  A  G  K  L
gatgaagctcagaataaagcactcgtacaaggattggacacggatgttcaaagttactcc
 D  E  A  Q  N  K  A  L  V  Q  G  L  D  T  D  V  Q  S  Y  S
ggatactttattgtagatgaacagcataactctaatatcttcttctggtttttcccttca
 G  Y  F  I  V  D  E  Q  H  N  S  N  I  F  F  W  F  F  P  S
caaagtgatgccagttcggatccggtagttctatggctccaaggaggaccaggatctacg
 Q  S  D  A  S  S  D  P  V  V  L  W  L  Q  G  G  P  G  S  T
tccatgtttggacttttcaggaaaatggaccacttttattaaaggatggtgagcttggt
 S  M  F  G  L  F  Q  E  N  G  P  L  L  L  K  D  G  E  L  G
gttcgaccgacgtcttggaatagaaatcactcagttatatacattgatcagccagttgga
 V  R  P  T  S  W  N  R  N  H  S  V  I  Y  I  D  Q  P  V  G
actggatggagttttactgatggaggatacgccaatgatcaaaataaagtagctgctgac
 T  G  W  S  F  T  D  G  G  Y  A  N  D  Q  N  K  V  A  A  D
ttgtacgaagctttgcagcagttttttcactctatttatcaacaccaggatagagacttc
 L  Y  E  A  L  Q  Q  F  F  H  S  I  Y  Q  H  Q  D  R  D  F
gttacaggtgaatcttatggaggaaagtatgttcccgctataagttatgaaattcaccaa
 V  T  G  E  S  Y  G  G  K  Y  V  P  A  I  S  Y  E  I  H  Q
aggaacgaggacgcggagctgaagattaatcttaaagggttggctataggtaatggtttc
 R  N  E  D  A  E  L  K  I  N  L  K  G  L  A  I  G  N  G  F
actgatcctatcgtacaaatgcgatattctgagtacgtctaccaacatggacttattgat
 T  D  P  I  V  Q  M  R  Y  S  E  Y  V  Y  Q  H  G  L  I  D
ttgaatacgaaaaattcgcttcataacttagaaaataaagctgtacaagctattcaatca
 L  N  T  K  N  S  L  H  N  L  E  N  K  A  V  Q  A  I  Q  S
gaggagtactatgatgcctttctcttatgggatgatatcgttaacactatacaagacttt
 E  E  Y  Y  D  A  F  L  L  W  D  D  I  V  N  T  I  Q  D  F
acggaattagatatatacaattatctgacacttaaaggaagtgctgccgatgaagaagaa
 T  E  L  D  I  Y  N  Y  L  T  L  K  G  S  A  A  D  E  E  E
ttaaacgaatttttcagtaaggaagaagtaagaaaggcaattcatgttggcgccactgaa
 L  N  E  F  F  S  K  E  E  V  R  K  A  I  H  V  G  A  T  E
ttcggtgatggtcaagtttatactcatctcgaaaaagatattcccaaatcggaggcatat
 F  G  D  G  Q  V  Y  T  H  L  E  K  D  I  P  K  S  E  A  Y
aaagtcgctgaactcttaaaatattataaagttttaatctataatggacaaacagatctt
 K  V  A  E  L  L  K  Y  Y  K  V  L  I  Y  N  G  Q  T  D  L
atagttccttatgctttgacagtaaactatttgcaaaggctaaaattccctggctccgag
 I  V  P  Y  A  L  T  V  N  Y  L  Q  R  L  K  F  P  G  S  E
gaatacgcagaagcagacaggactgtatggcatgatgataatgatcgaagtgtggttgct
 E  Y  A  E  A  D  R  T  V  W  H  D  D  N  D  R  S  V  V  A
gggtatgcgaaaacagctggaaagttaacagaaattatggttaggaattctggtcatatg
 G  Y  A  K  T  A  G  K  L  T  E  I  M  V  R  N  S  G  H  M
gtgccgacagatcaaccagaatgggctttacaactcattaataaatttacaagaaatgaa
 V  P  T  D  Q  P  E  W  A  L  Q  L  I  N  K  F  T  R  N  E
ccttttaactaaatataacttaaattttgctaagtgcattcaaaaaa
 P  F  N  -  I  -  L  K  F  C  -  V  H  S  K
```

Figure A.3.3a. (previous page) Nucleotide and predicted amino acid sequence of *D. undecimpunctata* serine type carboxypeptidase (cDNA clone A4).

```
gagtcttaatactagtaagtgctcaaaatgaaggtgctagtgttactctcgttgctatct
 E  S  -  Y  -  -  V  L  K  M  K  V  L  V  L  L  S  L  L  S
acatttcttatttgccaaacatcaggtgcgaaaaaaaacgtcgtttgttatttcgccagt
 T  F  L  I  C  Q  T  S  G  A  K  K  N  V  V  C  Y  F  A  S
tggactatttatagagcaggaaatggtgctttcgatgtcagtaatatagatccatcgctg
 W  T  I  Y  R  A  G  N  G  A  F  D  V  S  N  I  D  P  S  L
tgtacacacattaattttgccttccttggtctaaacttagatggttctattcacattttg
 C  T  H  I  N  F  A  F  L  G  L  N  L  D  G  S  I  H  I  L
gattcttgggagtcaaatgatcctggtggtcatgagggttttaagcgtctcgtaggtctt
 D  S  W  E  S  N  D  P  G  G  H  E  G  F  K  R  L  V  G  L
aaagagaccaatcctgaccttaaagtaagtgtaagtatgggcggttggaacgaaggctcc
 K  E  T  N  P  D  L  K  V  S  V  S  M  G  G  W  N  E  G  S
gagcactattcagcagtagcatcagatccagccaaaagggtaaagcttgcaggtgaggtt
 E  H  Y  S  A  V  A  S  D  P  A  K  R  V  K  L  A  G  E  V
ttagcttttatcgaaaattggggcttcgatggttttgatttggattgggaatatccagga
 L  A  F  I  E  N  W  G  F  D  G  F  D  L  D  W  E  Y  P  G
ttacgaggaggaaacgaaactattgacaaagaagactttgtcgaacttgtaaaagctctt
 L  R  G  G  N  E  T  I  D  K  E  D  F  V  E  L  V  K  A  L
agtgacgttcttcaacccaagggttacttgctcagtatagccactgcaggcgccgttgaa
 S  D  V  L  Q  P  K  G  Y  L  L  S  I  A  T  A  G  A  V  E
aaaattgaccttggatatgatgtcccagttataaatgagttggtggatatgcttaatgtt
 K  I  D  L  G  Y  D  V  P  V  I  N  E  L  V  D  M  L  N  V
atggtttttgattttcatggagcatttgagaattttgtaggacatgtttcaccactgttc
 M  V  F  D  F  H  G  A  F  E  N  F  V  G  H  V  S  P  L  F
ccagctcaaattgattacgaatatgaagctaatagtacatacaatgtagatgcaggaatc
 P  A  Q  I  D  Y  E  Y  E  A  N  S  T  Y  N  V  D  A  G  I
caacattggatattgagtggtgcagaccccgcaaaaataaacctcggtattgtcacttat
 Q  H  W  I  L  S  G  A  D  P  A  K  I  N  L  G  I  V  T  Y
ggaagaacatttaccttagctgataaaaccaatacttctctttatgcagatgtcatcggt
 G  R  T  F  T  L  A  D  K  T  N  T  S  L  Y  A  D  V  I  G
ggtggtaataaagggccatatactggacaatctggatatttaggatataatgaaatatgc
 G  G  N  K  G  P  Y  T  G  Q  S  G  Y  L  G  Y  N  E  I  C
gaatttcatttaaattctacgtatgtctgggatgatcagcagaaagtacctcacagatat
 E  F  H  L  N  S  T  Y  V  W  D  D  Q  Q  K  V  P  H  R  Y
tgggacgatcaatgggttggatttgaagacgaaagatcaataagagaaaaagtaaactac
 W  D  D  Q  W  V  G  F  E  D  E  R  S  I  R  E  K  V  N  Y
gctcgggacaatgaattgggaggaatgatggtgtggtcattagattacgacgatttcaga
 A  R  D  N  E  L  G  G  M  M  V  W  S  L  D  Y  D  D  F  R
ggaacttgtggagaaaaatatattttattaaagactataaacaaaattcatgatccacag
 G  T  C  G  E  K  Y  I  L  L  K  T  I  N  K  I  H  D  P  Q
ttttgaatttgtaaattttagttatgtgtagtaggttattctagaaaattgtcatagata
 F  -  I  C  K  F  -  L  C  V  V  G  Y  S  R  K  L  S  -  I
ctt
 L
```

Figure A.3.3b. Nucleotide and predicted amino acid sequence of *D. undecimpunctata* chitinase (cDNA clone CA5).

```
gacattcactaataacttgcgaagatgaaggtattagttctttttttctttagtagcactc
 D  I  H  -  -  L  A  K  M  K  V  L  V  L  F  S  L  V  A  L
agcgttgctctaccaacactccttggtagaaccgtcttagtcaaagaagatggaactgta
 S  V  A  L  P  T  L  L  G  R  T  V  L  V  K  E  D  G  T  V
accctcttcgactctagaggataccaactcgtaatcctcaagagtgttaccgatcccaga
 T  L  F  D  S  R  G  Y  Q  L  V  I  L  K  S  V  T  D  P  R
aaagtcgaacttgttctcaggggtccaaatacaagaacccagatgttccaagtcggggaa
 K  V  E  L  V  L  R  G  P  N  T  R  T  Q  M  F  Q  V  G  E
tctcttagaactagagaaattatcgaccgtactctggttggcgtttctgttggtgaaaac
 S  L  R  T  R  E  I  I  D  R  T  L  V  G  V  S  V  G  E  N
tacaaccaaggtgatattcttgccgatattttccgccagtatgaaggtacacttgatgat
 Y  N  Q  G  D  I  L  A  D  I  F  R  Q  Y  E  G  T  L  D  D
nttaaataccatagcctccttaacagaattcaattgttggttgaagctggagtcatcaac
 X  K  Y  H  S  L  L  N  R  I  Q  L  L  V  E  A  G  V  I  N
gaaaccatctacgatattatcagaaactgggatcttgaataccgtatgcaaggagttagt
 E  T  I  Y  D  I  I  R  N  W  D  L  E  Y  R  M  Q  G  V  S
gacattgtccccgtccaaggagtacgtaccctccaccaatatgctgaacaagtaggaga
 D  I  V  P  V  Q  G  V  R  T  L  H  Q  Y  A  E  Q  V  G
```

Figure A.3.3c. Nucleotide and predicted amino acid sequence of *D. undecimpunctata* early stage encapsulation response protein (cDNA clone A14).

```
gcaggttttgaaagcatatattattagtaacttgtaataatggcgcacacttataaagta
  A  G  F  E  S  I  Y  Y  -  -  L  V  I  M  A  H  T  Y  K  V
ggttactttaactttactgcaaggggtgaaccgattcgaatgcttttaagctatggagaa
 G  Y  F  N  F  T  A  R  G  E  P  I  R  M  L  L  S  Y  G  E
attccattcgaagacgaacgtatttccagggaagactggcctaaaataaaaccaactact
 I  P  F  E  D  E  R  I  S  R  E  D  W  P  K  I  K  P  T  T
cctcttggacaacttccagtgttggaaattgatggaaaacaaattccacatagtatctcc
 P  L  G  Q  L  P  V  L  E  I  D  G  K  Q  I  P  H  S  I  S
atttgcagatatttagcttcagtagtgaaacttgatggaaaagatgctaaagaaaattta
 I  C  R  Y  L  A  S  V  V  K  L  D  G  K  D  A  K  E  N  L
agaattgatgttgcagtggaaacccttctagacttgcaaaaacttgcatttgaatatgtg
 R  I  D  V  A  V  E  T  L  L  D  L  Q  K  L  A  F  E  Y  V
tttgaatccgatgctgccaaaaaagatgagaaaaataaaaaacttgaggaatctttacca
 F  E  S  D  A  A  K  K  D  E  K  N  K  K  L  E  E  S  L  P
ctcttccttggaaaattagaagaacaagcaaataagaatggtggctataccgctcttgat
 L  F  L  G  K  L  E  E  Q  A  N  K  N  G  G  Y  T  A  L  D
aggttgtcatgggccgatatcatattcatttgtatctacgaaggacttatgaatattcta
 R  L  S  W  A  D  I  I  F  I  C  I  Y  E  G  L  M  N  I  L
ggtaaaaataccttgactcttatcctagtctacaaaaaattaagaaaaacgtactagac
 G  K  N  T  F  D  S  Y  P  S  L  Q  K  I  K  K  N  V  L  D
gttaaaggtattagaaactggatcaaaaacagacccgaaacaccttctaccatgggacct
 V  K  G  I  R  N  W  I  K  N  R  P  E  T  P  S  T  M  G  P
ttgaaattcgaatatgacttaaaacaagaagtttaagtgtgattactatacattgtaaat
 L  K  F  E  Y  D  L  K  Q  E  V  -  V  -  L  L  Y  I  V  N
atatattaaaatgattatattttgttaaaaataaaaataaaaatgtaca
 I  Y  -  N  D  Y  I  L  L  K  I  K  I  K  M  Y
```

Figure A.3.3d. Nucleotide and predicted amino acid sequence of *D. undecimpunctata* glutathione *S*-transferase 1 (cDNA clone CB12).

```
atggctcccactttgtatatgatcgcagccagccctgctgtaaggtcggtaactctatgc
 M  A  P  T  L  Y  M  I  A  A  S  P  A  V  R  S  V  T  L  C
gccaaggctctcaatatagaacttaacgagaaagaagttaatcttcttactggagcacat
 A  K  A  L  N  I  E  L  N  E  K  E  V  N  L  L  T  G  A  H
cttacaccagagtttctcaagttaaatcctcaacatactgtacctgtattggatgatgat
 L  T  P  E  F  L  K  L  N  P  Q  H  T  V  P  V  L  D  D  D
ggtttcgtaatagttgacagtcatgctattaatacatatttagtagctaaatacggaaaa
 G  F  V  I  V  D  S  H  A  I  N  T  Y  L  V  A  K  Y  G  K
gatgactctctttatccaaaagacattaaaaagagggcgattgtcgatcacagacttcac
 D  D  S  L  Y  P  K  D  I  K  K  R  A  I  V  D  H  R  L  H
tatgattcaagtattttgtttgttagaggactttcgatttctagaagtcttctcttcgaa
 Y  D  S  S  I  L  F  V  R  G  L  S  I  S  R  S  L  L  F  E
ggtgagaaagagattgcccctaaattattaacagcgctcgaggaagcctatactatcaca
 G  E  K  E  I  A  P  K  L  L  T  A  L  E  E  A  Y  T  I  T
gaaaaatttttggaaaagaataactatatcgctggtgaccaactaacagtagccgatttc
 E  K  F  L  E  K  N  N  Y  I  A  G  D  Q  L  T  V  A  D  F
tcttttattacttccattacctcatggaatgtgtttttcaacttgaaggaagctaaatat
 S  F  I  T  S  I  T  S  W  N  V  F  F  N  L  K  E  A  K  Y
cccaaaatagttgcttggatgaagagaatgcaagctctgccttactacaaggaagcaaat
 P  K  I  V  A  W  M  K  R  M  Q  A  L  P  Y  Y  K  E  A  N
caagttggtttggaccagtattatgcaattatgaaagacaaagttccaaagtagagcaca
 Q  V  G  L  D  Q  Y  Y  A  I  M  K  D  K  V  P  K  -  S  T
tcttatttgctgattgccactgtatatttttttaaaatatcaaattttcatttttaaata
 S  Y  L  L  I  A  T  V  Y  F  F  K  I  S  N  F  H  F  -  I
tattattttaaatcgg
 Y  Y  F  K  S
```

Figure A.3.3e. Nucleotide and predicted amino acid sequence of *D. undecimpunctata* glutathione *S*-transferase 2 (cDNA clone A17).

```
ggttgtcagtcagtaacaatggtgtcgaaactgttatttttggtgttagtggcatttgct
 G  C  Q  S  V  T  M  V  S  K  L  L  F  L  V  L  V  A  F  A
gcctacgtgcaaagtgaaattacggtaaatgtttacactagttattgggaatcagaagaa
 A  Y  V  Q  S  E  I  T  V  N  V  Y  T  S  Y  W  E  S  E  E
tcagtgattccatcaatcgagaacttgaaggatcttatcaccaatgtaacctacagggtg
 S  V  I  P  S  I  E  N  L  K  D  L  I  T  N  V  T  Y  R  V
gaaattgaaaaccagaccattcctaaattacctagcggagtgttacaaggtgtggccatc
 E  I  E  N  Q  T  I  P  K  L  P  S  G  V  L  Q  G  V  A  I
accgagctaaaactcaagcatgatcaattaaaagacattgaggccggtgcgttggaggga
 T  E  L  K  L  K  H  D  Q  L  K  D  I  E  A  G  A  L  E  G
gcccgcatcaacagaattgagttcgaccacaatgacatcgaaactatcaaaaaaggcgtc
 A  R  I  N  R  I  E  F  D  H  N  D  I  E  T  I  K  K  G  V
ttcaacgataggtcaatatgggagcttgatctatccaacaaccagatttccaccattgag
 F  N  D  R  S  I  W  E  L  D  L  S  N  N  Q  I  S  T  I  E
gatgatgctttcgacacattggatgtgagtactctacaccttaacgacaacaatcttgaa
 D  D  A  F  D  T  L  D  V  S  T  L  H  L  N  D  N  N  L  E
aaaattacagcaggtacatttcataatactaaagttacaagcttagatttgaacaataac
 K  I  T  A  G  T  F  H  N  T  K  V  T  S  L  D  L  N  N  N
aaaattcaaaaaattgaagttaatgctttaaatcaaatacccagtctatattacgttcaa
 K  I  Q  K  I  E  V  N  A  L  N  Q  I  P  S  L  Y  Y  V  Q
ctgagtaacaacctaattaa
 L  S  N  N  L  I
```

Figure A.3.3f. Nucleotide and predicted amino acid sequence of putative *D. undecimpunctata* slit 3 protein (cDNA clone c221).

A.3.4. Nucleotide and predicted amino acid sequences of cDNA clones encoding proteins with functions unassigned by homology search.

```
ggaaactgaaagtagctgtgatagtgacgagttccaaactgcattaagatgaagtggata
 G  N  -  K  -  L  -  -  -  R  V  P  N  C  I  K  M  K  W  I
atgttaacagttgttgtacttcttggtttggccaatgcagaggaagtttccgaaaagaga
 M  L  T  V  V  V  L  L  G  L  A  N  A  E  E  V  S  E  K  R
gcagttcttagaagaaccaccagacaagccctaggatccatgggaggttcatttttagca
 A  V  L  R  R  T  T  R  Q  A  L  G  S  M  G  G  S  F  L  A
caaactgagaggaatattatcaataattgtaagaaaaatggactatccctttcaggtgcc
 Q  T  E  R  N  I  I  N  N  C  K  K  N  G  L  S  L  S  G  A
gacgatctacaggctacttatagtaaaatgaagacttgtatgtccagcaccaaaatttttt
 D  D  L  Q  A  T  Y  S  K  M  K  T  C  M  S  S  T  K  I  F
gaaactcccaagaatgaatttattagcaaaatggttgattgcagtcggaacgcaattaga
 E  T  P  K  N  E  F  I  S  K  M  V  D  C  S  R  N  A  I  R
gaaactaaagattgtttgtctaatgaccaaaaatatttcccagaattcgttttggatctt
 E  T  K  D  C  L  S  N  D  Q  K  Y  F  P  E  F  V  L  D  L
gcaaaatcgatggttaattatatgtacgatgacaaagatctatttaaagatttggaaatt
 A  K  S  M  V  N  Y  M  Y  D  D  K  D  L  F  K  D  L  E  I
gttggttgcataaacaacttgggtagatatacagtacaaattgagtatattaaatgctta
 V  G  C  I  N  N  L  G  R  Y  T  V  Q  I  E  Y  I  K  C  L
accactgtgtctcaaagaaccggagacgatgaaaatattcctacctcccgacaagaattc
 T  T  V  S  Q  R  T  G  D  D  E  N  I  P  T  S  R  Q  E  F
tgcagaaaattcataccagccactgaatgttttccagacacacttaaacagcactgttca
 C  R  K  F  I  P  A  T  E  C  F  P  D  T  L  K  Q  H  C  S
agtacta
 S  T
```

Figure A.3.4a. Nucleotide and predicted amino acid sequence of *D. undecimpunctata* cDNA clone CA6.

```
gacagtatcgaacaacatcgaaacaaaatggcaaaaatcgtcctcttcttctgcgccatc
 D  S  I  E  Q  H  R  N  K  M  A  K  I  V  L  F  F  C  A  I
attttggcttcagctgttgctttaccaagcaaatcagatgacgaaaatttgatcgacatc
 I  L  A  S  A  V  A  L  P  S  K  S  D  D  E  N  L  I  D  I
gtcaaaaagatcgtccacaagattggagaccttctcgacaaagtttttgccgatggtccc
 V  K  K  I  V  H  K  I  G  D  L  L  D  K  V  F  A  D  G  P
aatttcatcgacactttcgccgacaagtatgacgaggaccttactaaagtgctagattat
 N  F  I  D  T  F  A  D  K  Y  D  E  D  L  T  K  V  L  D  Y
acaagggacgttcttactaaactggttcaagcacttaaacccccttgtagatactattgct
 T  R  D  V  L  T  K  L  V  Q  A  L  K  P  L  V  D  T  I  A
gcacaaagtaatgacgctggaaagaaattggtttcttgtatcgaacaacacgaatctgac
 A  Q  S  N  D  A  G  K  K  L  V  S  C  I  E  Q  H  E  S  D
atcgtcgatatccgcgtcactttcgtagaaggcgctggcaaatgtgtaaaagatagcctc
 I  V  D  I  R  V  T  F  V  E  G  A  G  K  C  V  K  D  S  L
gtcgctttggttaaaactgtcgcacctgtgcttaaagatctatctgatgtccacgttgag
 V  A  L  V  K  T  V  A  P  V  L  K  D  L  S  D  V  H  V  E
gccaaagcagccgccgacgaaattgaccaatgtgaaggaagtgatgcacaaactctcctc
 A  K  A  A  A  D  E  I  D  Q  C  E  G  S  D  A  Q  T  L  L
tgcgttgttcaggtcgcctccgatttactcaaagttttaactgagattcccgatgctgtg
 C  V  V  Q  V  A  S  D  L  L  K  V  L  T  E  I  P  D  A  V
aaaggcgatatccaacccgtcattgatgccgt
 K  G  D  I  Q  P  V  I  D  A
```

Figure A.3.4b. Nucleotide and predicted amino acid sequence of *D. undecimpunctata* cDNA clone CB4.

```
atgtaatgacaactccatttactactcttattttggagaatttgagccaatcatctacacca
   V  M  T  T  P  F  T  T  L  I  L  E  N  L  S  Q  S  S  T  P
gatcaagtagatacctcccaatatacatctttcaaaatcgaaaatgaaactgttggacag
 D  Q  V  D  T  S  Q  Y  T  S  F  K  I  E  N  E  T  V  G  Q
ttcagaatgggccgtaaaaggtgccaagctgaattcctgacagctttaggggggaggattg
 F  R  M  G  R  K  R  C  Q  A  E  F  L  T  A  L  G  G  G  L
ttgagtggagcaggccagggaatatccgcctggatgcaacaacaacaagccgaaaaggac
 L  S  G  A  G  Q  G  I  S  A  W  M  Q  Q  Q  Q  A  E  K  D
cgtcaactccagaaggagctccaacagatggttcaaaatggccaaatgtcgctagccgaa
 R  Q  L  Q  K  E  L  Q  Q  M  V  Q  N  G  Q  M  S  L  A  E
gcccaagctaaatatgccgcagctcaagctcaagccaatttcggctataacacacaactg
 A  Q  A  K  Y  A  A  A  Q  A  Q  A  N  F  G  Y  N  T  Q  L
caacaacaacaagcagacaatgcccagacaagagatcgtctccaaatggaaaacaacatc
 Q  Q  Q  Q  A  D  N  A  Q  T  R  D  R  L  Q  M  E  N  N  I
aaatatcgtggaatccaattggcagggcaaaacaccggaaataataatgtcggatgggga
 K  Y  R  G  I  Q  L  A  G  Q  N  T  G  N  N  N  V  G  W  G
aatgctggttgggacactccaagtgtacgcaaggaccgtagaaataactttaaacagtta
 N  A  G  W  D  T  P  S  V  R  K  D  R  R  N  N  F  K  Q  L
ccttttgctgttcgggtaccccaacaagcatctacacaggcaaccacaatccagcctaaa
 P  F  A  V  R  V  P  Q  Q  A  S  T  Q  A  T  T  I  Q  P  K
aacaaagccagtattatgaaagtgactt
 N  K  A  S  I  M  K  V  T
```

Figure A.3.4c. Nucleotide and predicted amino acid sequence of *D. undecimpunctata* cDNA clone CA9.

```
gatatttcaatgaataaattaactttaagcactgcgaacaatatgaataccaaaagtgta
 D  I  S  M  N  K  L  T  L  S  T  A  N  N  M  N  T  K  S  V
tttacgatgtttttgctgggggtcatttctttgacccatggattcgacacagaatttgtt
 F  T  M  F  L  L  G  V  I  S  L  T  H  G  F  D  T  E  F  V
acttccaatggaaagttgcagttgcctaagctagtaaaatatgaacctgaaaatttttaaa
 T  S  N  G  K  L  Q  L  P  K  L  V  K  Y  E  P  E  N  F  K
aattatgaagattctggaattacagcaaacctaaacgaatatgcagaaaaaatttttctac
 N  Y  E  D  S  G  I  T  A  N  L  N  E  Y  A  E  K  I  F  Y
aatgttggaaggtatgctctaacccatggacttgaccctatcgatattgcaaatataact
 N  V  G  R  Y  A  L  T  H  G  L  D  P  I  D  I  A  N  I  T
gaagattttcaactggcgatgataaagcttacacatggtagacttcaaggaatatcgaca
 E  D  F  Q  L  A  M  I  K  L  T  H  G  R  L  Q  G  I  S  T
cttaaaagatatcaggacgttttagctaattaccaacatagtaccaagactttaactgtt
 L  K  R  Y  Q  D  V  L  A  N  Y  Q  H  S  T  K  T  L  T  V
actttacctatagaatttgctgatcttaagtttacctacgattacgatgtaagaatactt
 T  L  P  I  E  F  A  D  L  K  F  T  Y  D  Y  D  V  R  I  L
ggacccggtccaagcggcagcatggatggcgaaatccaggaatttaagttttacattcaa
 G  P  G  P  S  G  S  M  D  G  E  I  Q  E  F  K  F  Y  I  Q
atttgtttagacctcgaaaaattcattgtatcgattaaacaactaaagacaactgacagt
 I  C  L  D  L  E  K  F  I  V  S  I  K  Q  L  K  T  T  D  S
ggacatatctctgttcaattccatggaaaatattac
 G  H  I  S  V  Q  F  H  G  K  Y  Y
```

Figure A.3.4d. Nucleotide and predicted amino acid sequence of *D. undecimpunctata* cDNA clone A18.

A.3.5. Nucleotide and predicted amino acid sequences of PKPI clones isolated by PCR.

```
atgatgaagtgtttatttttgttatgtttgtgtttgtttcccattgtggtgttttcatca
 M  M  K  C  L  F  L  L  C  L  C  L  F  P  I  V  V  F  S  S
actttcacttcccaaaatcccattaacctacctagtgatgctactccagtacttgacgta
 T  F  T  S  Q  N  P  I  N  L  P  S  D  A  T  P  V  L  D  V
actggtaaagaacttgatcctcgtttgagttatcatattatttccacttttttggggtgcg
 T  G  K  E  L  D  P  R  L  S  Y  H  I  I  S  T  F  W  G  A
ttaggtggtgatgtgtacctaggtaagtccccaaattcagatgccccttgtgcaaatggc
 L  G  G  D  V  Y  L  G  K  S  P  N  S  D  A  P  C  A  N  G
atattccgttacaattcggatgttggacctagcggtacacccgttagattcattggttca
 I  F  R  Y  N  S  D  V  G  P  S  G  T  P  V  R  F  I  G  S
tctagtcattttggacaaggtatctttgaaaatgaactactcaacatccaatttgctatt
 S  S  H  F  G  Q  G  I  F  E  N  E  L  L  N  I  Q  F  A  I
tcaacatcgaaattgtgtgttagttatacaatttggaaagtgggagattacgatgcatct
 S  T  S  K  L  C  V  S  Y  T  I  W  K  V  G  D  Y  D  A  S
ctagggacgatgttgttggagactggaggaaccataggtcaagcagatagcagttggttc
 L  G  T  M  L  L  E  T  G  G  T  I  G  Q  A  D  S  S  W  F
aagattgttaaatcatcacaatttggttacaacttattgtattgccctgttactagtaca
 K  I  V  K  S  S  Q  F  G  Y  N  L  L  Y  C  P  V  T  S  T
atgagttgtccatttcccctgatgatcaattctgtttaaaagttggtgtagttcaccaa
 M  S  C  P  F  S  P  D  D  Q  F  C  L  K  V  G  V  V  H  Q
aatggaaaaagacgtttggctcttgtcaaggacaatcctcttgatgtctccttcaagcaa
 N  G  K  R  R  L  A  L  V  K  D  N  P  L  D  V  S  F  K  Q
gtccagtaataacaaatgtctgcctgctagctagactatatgttttagcagctactatat
 V  Q  -  -  Q  M  S  A  C  -  L  D  Y  M  F  -  Q  L  L  Y
atgttatgttgtaaattaaaataaacacctgctaagctatatctatattttagcatggat
 M  L  C  C  K  L  K  -  T  P  A  K  L  Y  L  Y  F  S  M  D
ttctaaataaattgtctttccttaaaaaaaaggggggggcccggtacccaattcgccctat
 F  -  I  N  C  L  S  L  K  K  R  G  G  P  V  P  N  S  P  Y
agtgagtcgtattac
 S  E  S  Y  Y
```

Figure A.3.5a. Nucleotide and predicted amino acid sequence of potato proteinase inhibitor clone ASPI 1.

```
atgatgaagtgtttatttttgttatgtttgtgtttgtttcccattgtggtgttttcatca
 M  M  K  C  L  F  L  L  C  L  C  L  F  P  I  V  V  F  S  S
actttcacttcccaaaatcccattaacctacctagtgatgctactccagtacttgacgta
 T  F  T  S  Q  N  P  I  N  L  P  S  D  A  T  P  V  L  D  V
actggtaaagaacttgatcctcgtttgagttatcatattatttccacttttggggtgcg
 T  G  K  E  L  D  P  R  L  S  Y  H  I  I  S  T  F  W  G  A
ttaggtggtgatgtgtacctaggtaagtccccaaattcagatgccccttgtgcaaatggc
 L  G  G  D  V  Y  L  G  K  S  P  N  S  D  A  P  C  A  N  G
atattccgttacaattcggatgttggacctagcggtacacccgttagattcattggttca
 I  F  R  Y  N  S  D  V  G  P  S  G  T  P  V  R  F  I  G  S
tctagtcattttggacaaggtatctttgaaaatgaactactcaacatccaattcgctatt
 S  S  H  F  G  Q  G  I  F  E  N  E  L  L  N  I  Q  F  A  I
tcaacatcgaaattgtgtgttagttatacaatttggaaagtgggagattacgatgcatct
 S  T  S  K  L  C  V  S  Y  T  I  W  K  V  G  D  Y  D  A  S
ctagggacgatgttgttggagactggaggaaccataggtcaagcagatagcagttggttc
 L  G  T  M  L  L  E  T  G  G  T  I  G  Q  A  D  S  S  W  F
aagattgttcaatcatcacaatttggttacaacttattgtattgccctgttactagtaca
 K  I  V  Q  S  S  Q  F  G  Y  N  L  L  Y  C  P  V  T  S  T
atgagttgtccattttcctctgatgatcaattctgtttaaaagttggtgtagttcaccaa
 M  S  C  P  F  S  S  D  D  Q  F  C  L  K  V  G  V  V  H  Q
aatggaaaaagacgtttggctcttgtcaaggacaatcctcttgatgtctccttcaagcaa
 N  G  K  R  R  L  A  L  V  K  D  N  P  L  D  V  S  F  K  Q
gtccagtaataa
 V  Q  -  -
```

Figure A.3.5b. Nucleotide and predicted amino acid sequence of potato proteinase inhibitor clone ASPI 3.

```
atgatgaagtgtttatttttgttatgtttgtgtttggttcccattgtggtgttttcatca
M  M  K  C  L  F  L  L  C  L  C  L  V  P  I  V  V  F  S  S
actttcacttccaaaaatcccattagcctacctagtgatgctactccagtacttgacgta
T  F  T  S  K  N  P  I  S  L  P  S  D  A  T  P  V  L  D  V
gctggtaaagaacttgattctcgtttgagttatcgtattatttccacttttggggtgcg
A  G  K  E  L  D  S  R  L  S  Y  R  I  I  S  T  F  W  G  A
ttaggtggtgatgtatacctaggaaagtccccaaattcagatgccccttgtccagatggc
L  G  G  D  V  Y  L  G  K  S  P  N  S  D  A  P  C  P  D  G
gtattccgttacaattccgatgttggacctagcggtacacccgttagattcattccttta
V  F  R  Y  N  S  D  V  G  P  S  G  T  P  V  R  F  I  P  L
tctggaggtatatttgaagatcaactactcaacatacaattcaatattgcaacagtgaaa
S  G  G  I  F  E  D  Q  L  L  N  I  Q  F  N  I  A  T  V  K
ttgtgtgttagttatacaatttggaaagtcggaaatctaaatgcatattttaggacgatg
L  C  V  S  Y  T  I  W  K  V  G  N  L  N  A  Y  F  R  T  M
ttgttggagacgggaggaaccatagggcaagcagatagcagctatttcaagattgttaaa
L  L  E  T  G  G  T  I  G  Q  A  D  S  S  Y  F  K  I  V  K
ttatcaaattttggttacaacttattgtattgccctattactccccctttttctttgtcca
L  S  N  F  G  Y  N  L  L  Y  C  P  I  T  P  P  F  L  C  P
ttttgtcgtgatgataacttctgtgcaaaggtgggtgtagttattcaaaatggaaaaagg
F  C  R  D  D  N  F  C  A  K  V  G  V  V  I  Q  N  G  K  R
cgtttggctcttgtcaacgaaaatcctcttgatgtcttattccaggaagtttagtaacaa
R  L  A  L  V  N  E  N  P  L  D  V  L  F  Q  E  V  -  -  Q
ataatgcctgcagatagactatactatgttttagcctgcctgctggctagctactatgtt
I  M  P  A  D  R  L  Y  Y  V  L  A  C  L  L  A  S  Y  Y  V
atgttatgttgtaaaataaacacctgctaaggtatatctatctatattttagcatggctt
M  L  C  C  K  I  N  T  C  -  G  I  S  I  Y  I  L  A  W  L
tctcaataaattgtctttccttaaaaaaaaaaaaaaaaaaa
```

Figure A.3.5c. Nucleotide and predicted amino acid sequence of potato proteinase inhibitor clone ASPI 2.

A.4. Alignment of Du-asp 1 and Du-cys 1 with Atlantic cod pepsin and human cathepsin L.

```
Du-asp 1:  NEYDLDYYGEIGVGTPAQKFNVIFDTGSTDLWVPSSKCQATSAQNGCKNHKQYDASKSST 117
           NE D +YYG I +GTP + F VIFDTGS++LWV SS C A +    C NH ++   +SST
Cod-Peps:  NEADTEYYGVISIGTPPESFKVIFDTGSSNLWVSSSHCSAQA----CSNHNKFKPRQSST 63

Du-asp 1:  YQKDGRSFHIEYGSGALSGFLSKDSVEVAGLAITDQTFAEATVETAPAFVSSTFDGILGL 177
           Y + G++  + YG+G + G L +D+V V G +  +Q   E+  E  P   ++ FDGILGL
Cod-Peps:  YVETGKTVDLTYGTGGMRGILGQDTVSVGGGSDPNQELGESQTEPGPFQAAAPFDGILGL 123

Du-asp 1:  AYPDLSAQNIPPVFQNMIKQGVLDAPVFSFYLS-QTANGDKGELLLGGSDSKYYKGDFTY 236
           AYP ++A    PVF NM  Q +++  +FSFYLS   ANG   E++LGG D+ +Y G   +
Cod-Peps:  AYPSIAAAGAVPVFDNMGSQSLVEKDLFSFYLSGGGANG--SEVMLGGVDNSHYTGSIHW 181

Du-asp 1:  TKVSTQLYWQTNLQGVSVGSRSVCKSGCEAVIDTGTSLIYGPTDDVDVVNSAIGATYDYS 296
             V+ + YWQ  L G++V  ++    GC+A++DTGTS I  P   +  +   IGA+ +
Cod-Peps:  IPVTAEKYWQVALDGITVNGQTAACEGCQAIVDTGTSKIVAPVSALANIMKDIGASENQ- 240

Du-asp 1:  VGLYTVNCNTDLNKLPNVSFTFGGKKFDIPASAYIIKDSGYCISSFVAQEFFLSGFE-WL 355
           G    NC   +  LP+++FT  G K  +P SAYI  D  +C S   +     +  E W+
Cod-Peps:  -GEMMGNC-ASVQSLPDITFTINGVKQPLPPSAYIEGDQAFCTSGLGSSGVPSNTSELWI 298

Du-asp 1:  VGDSFLKTVYSEFDFGNNRIGFAELA 381
           GD FL+  Y+ +D  NN++GFA  A
Cod-Peps:  FGDVFLRNYYTIYDRTNNKVGFAPAA 324
```

Identities = 130/326 (39%), Positives = 190/326 (57%), Gaps = 11/326 (3%)
Score = 242 bits (618), Expect = 3e-63

Figure A.4a. Alignment of Du-asp 1 with a pepsin aspartic proteinase isolated from Atlantic cod (Karlsen *et al.*, 1997). Alignment performed by Blast 2 server (Tatusova and Madden, 1999).

```
Query: 31  KWVQFKVKHNKSYINYVEEQTRFTIFQENLKKIENHNKKYYNGESSFMLGVTKFADLTEK 90
           +W ++K  HN+ Y    EE  R  ++++N+K IE HN++Y  G+ SF + +  F D+T +
Sbjct: 28  QWTKWKAMHNRLY-GMNEEGWRRAVWEKNMKMIELHNQEYREGKHSFTMAMNAFGDMTSE 86

Query: 91  EFMDMLVLSKNSRPNRPHVTHLLAPLG-DLPSAFDWRNKGAVTEVKDQGMCGSCWTFSTT 149
           EF  ++   +N +P +  V     PL  + P + DWR KG VT VK+QG CGSCW FS T
Sbjct: 87  EFRQVMNGFQNRKPRKGKVFQ--EPLFYEAPRSVDWREKGYVTPVKNQGQCGSCWAFSAT 144

Query: 150 GSVEGAHFLKTGNPVSLSEQNLVDCAN-DNCYGCGGGWMDKALEYI-EKGGIMSEKDYPY 207
           G++EG  F KTG  +SLSEQNLVDC+      GC GG MD A +Y+ + GG+ SE+ YPY
Sbjct: 145 GALEGQMFRKTGRLISLSEQNLVDCSGPQGNEGCNGGLMDYAFQYVQDNGGLDSEESYPY 204

Query: 208 EGVDDKCRFDVSKVAAKISNFTYIKKNDEEDLKNAVFAKGPISVAIDAS-SKFQLYVAGI 266
           E  ++ C+++     A  + F  I K  E+ L  AV   GPISVAIDA    F  Y  GI
Sbjct: 205 EATEESCKYNPKYSVANDTGFVDIPK-QEKALMKAVATVGPISVAIDAGHESFLFYKEGI 263

Query: 267 LDDTECSNEFDSLNHGVLVVGYGTE----NGKDYWIVKNSWGVNWGMDGYIRMSRNKNNQ 322
             + +CS+E   ++HGVLVVGYG E    +   YW+VKNSWG  WGM GY++M++++ N
Sbjct: 264 YFEPDCSSE--DMDHGVLVVGYGFESTESDNNKYWLVKNSWGEEWGMGGYVKMAKDRRNH 321

Query: 323 CGITTDGVYPNV 334
           CGI +   YP V
Sbjct: 322 CGIASAASYPTV 333
```

Identities = 137/312 (43%), Positives = 191/312 (60%), Gaps = 14/312 (4%)
Score = 258 bits (658), Expect = 8e-68

Figure A.4b. Alignment of Du-cys 1 with human cathepsin L (Gal and Gottesman, 1988). Alignment performed by Blast 2 server (Tatusova and Madden, 1999).

A.5. Grant award details.

Type of BBSRC Award: Studentship

Institution: University of Durham

Studentship Reference: 97/A4/D/03736

Title: *Engineering crops for resistance to corn rootworm using inhibitors of insect digestive enzymes*

Supervisor: Gatehouse J. A. , Osborn R. W.

Student: Macgregor J. M.

Start date: 1/10/97

Est. End date: 30/09/00

Duration: 36 months

http://dataserv.bbsrc.ac.uk/cgi-bin/thesoff.cgi?*ID=2&*DB=WORK&*QQ=000005991.DOCN..GRNT.

A.6. Multiple sequence alignments for aspartic, cysteine and PKPI protein PCR primer designs.

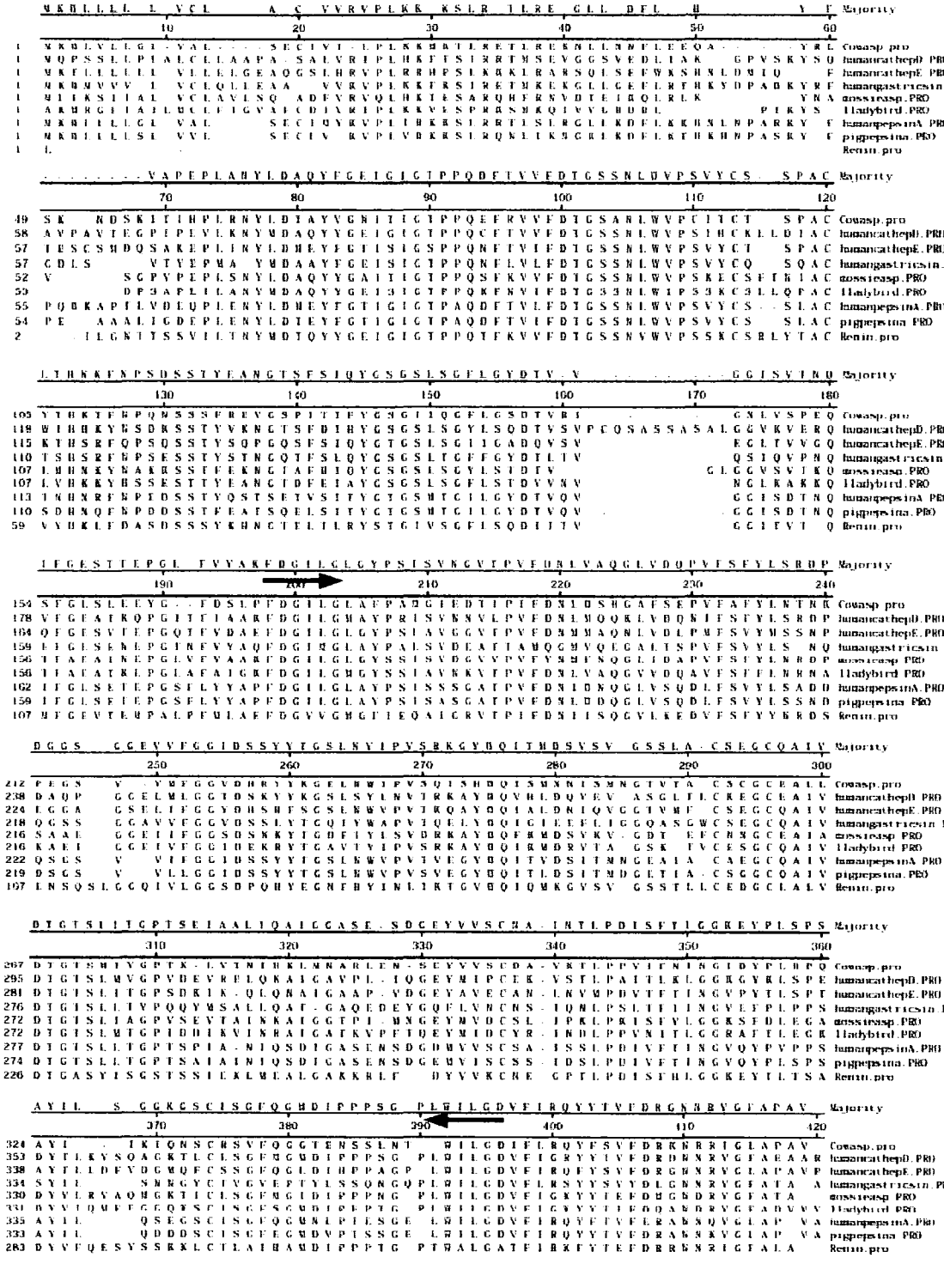
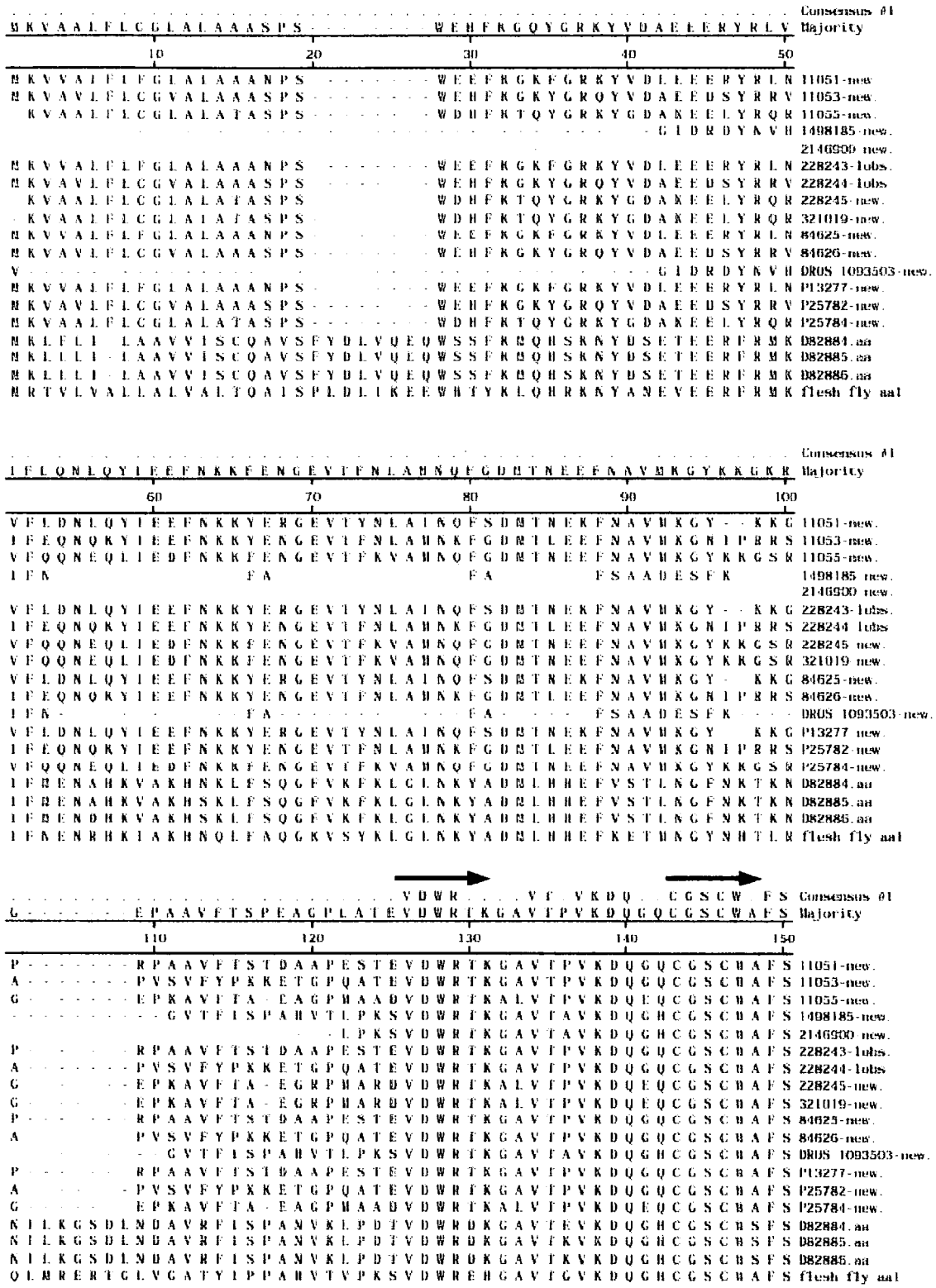
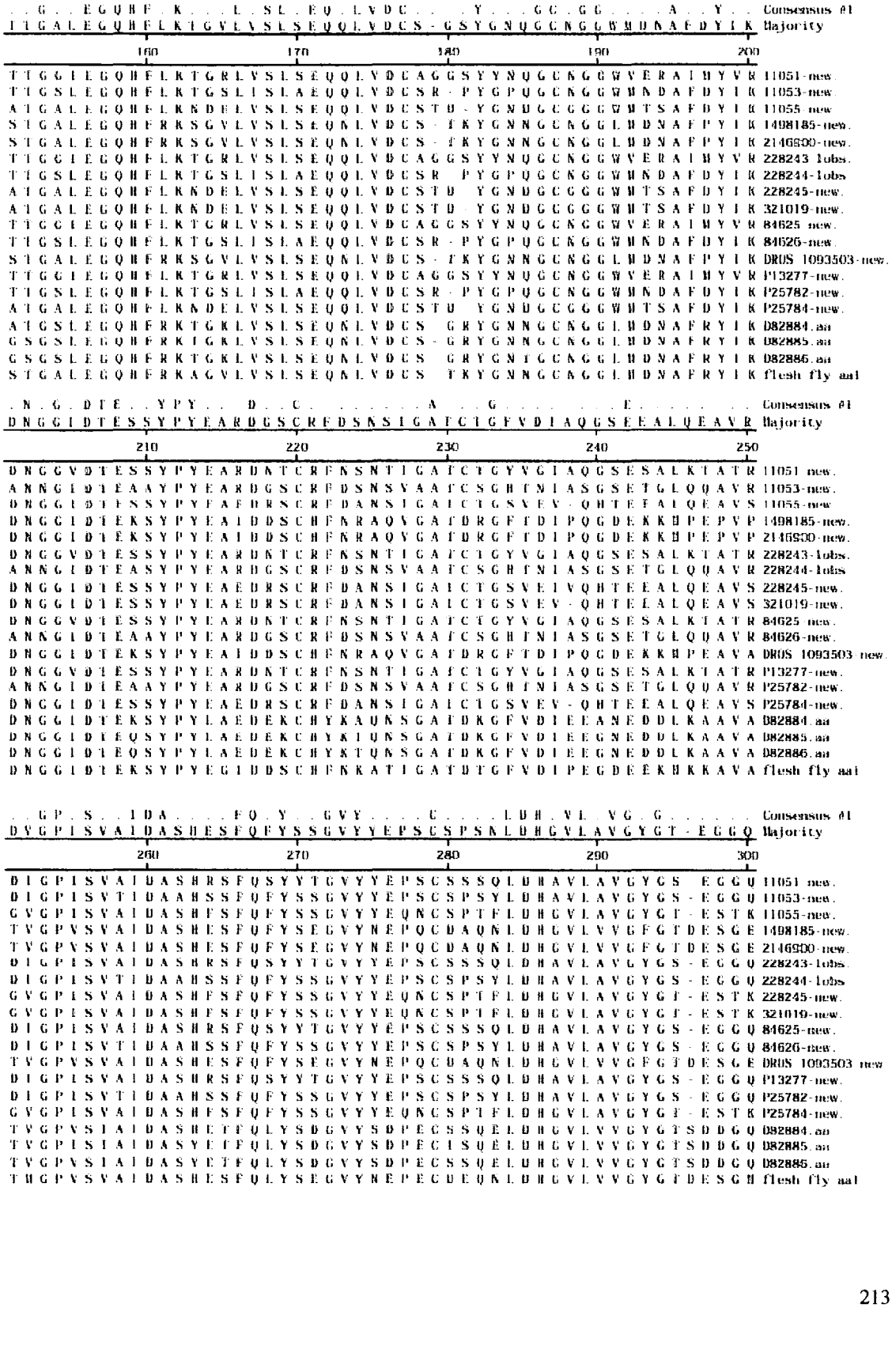
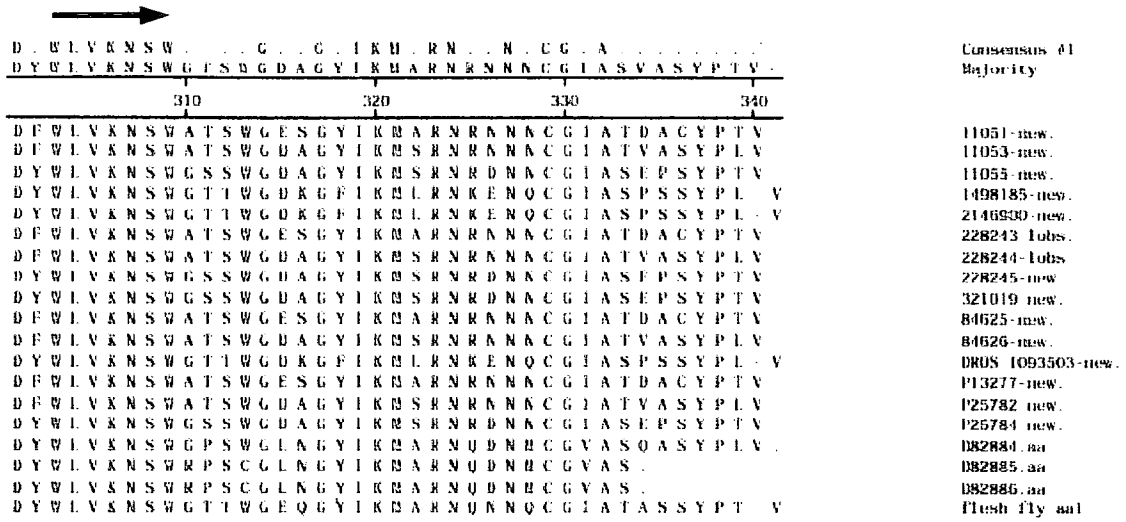


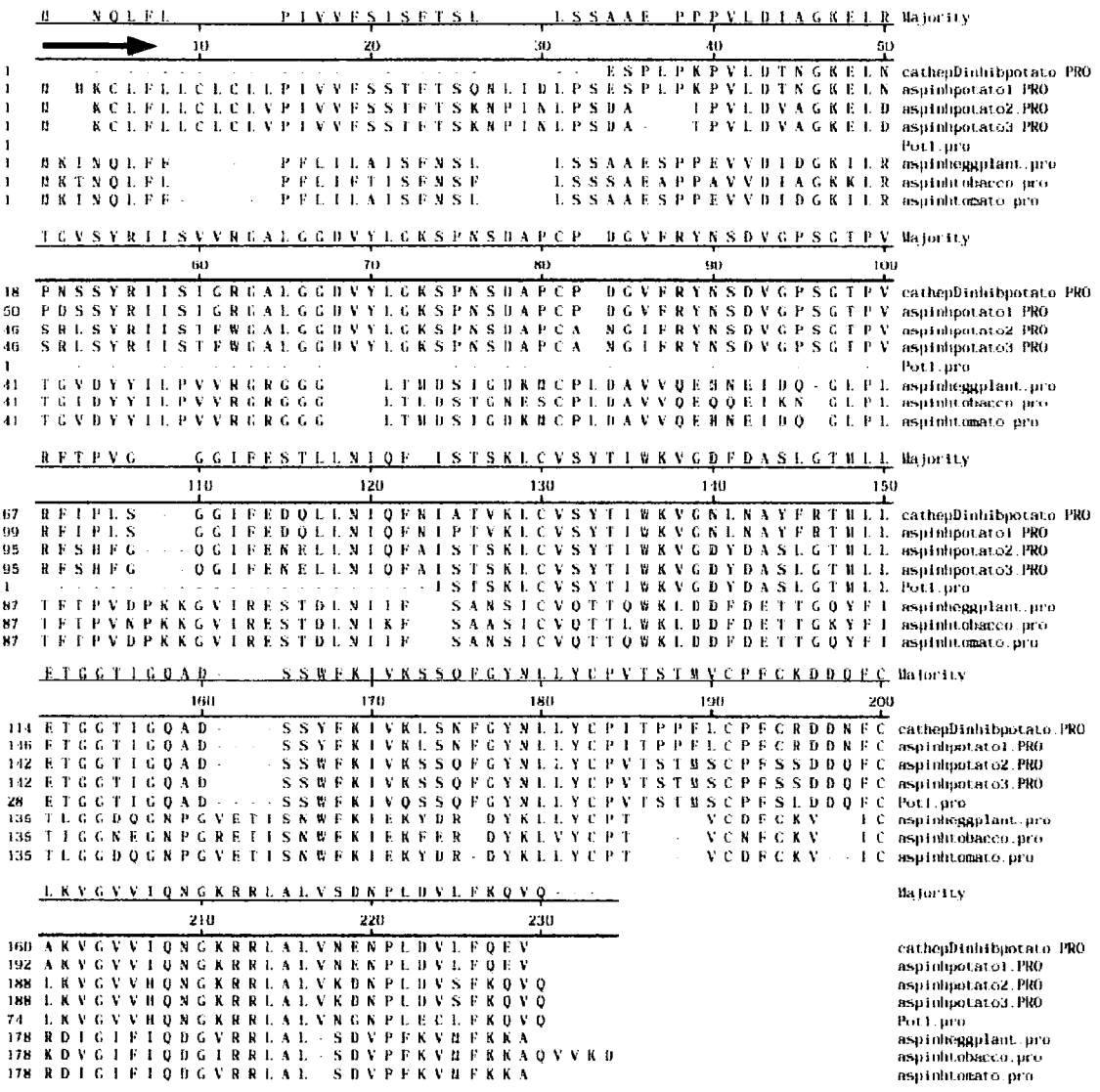
Figure A.6a. Multiple alignment of mammalian and invertebrate aspartic proteinase amino acid sequences. Clustal alignment performed with Megalign (DNASTAR). RAM and JAM PCR primer positions are indicated by underlining, arrowheads indicate direction of primed synthesis.


Consensus #1
Majority
11051-new.
11053-new.
11055-new.
1498185-new.
2146900-new.
228243-lubs.
228244-lubs
228245-new.
321019-new.
84625-new.
84626-new.
DROS 1093503-new.
P13277-new.
P25782-new.
P25784-new.
D82884.aa
D82885.aa
D82886.aa
flesh fly aal


Consensus #1
Majority
160
170
180
190
200
11051-new.
11053-new.
11055-new.
1498185-new.
2146900-new.
228243-1obs.
228244-1obs.
228245-new.
321019-new.
84625-new.
84626-new.
DROS 1093503-new.
P13277-new.
P25782-new.
P25784-new.
D82884.aa
D82885.aa
D82886.aa
flesh fly aa1
Consensus #1
Majority
210
220
230
240
250
11051-new.
11053-new.
11055-new.
1498185-new.
2146900-new.
228243-1obs.
228244-1obs.
228245-new.
321019-new.
84625-new.
84626-new.
DROS 1093503-new.
P13277-new.
P25782-new.
P25784-new.
D82884.aa
D82885.aa
D82886.aa
flesh fly aa1
Consensus #1
Majority
260
270
280
290
300
11051-new.
11053-new.
11055-new.
1498185-new.
2146900-new.
228243-1obs.
228244-1obs.
228245-new.
321019-new.
84625-new.
84626-new.
DROS 1093503-new.
P13277-new.
P25782-new.
P25784-new.
D82884.aa
D82885.aa
flesh fly aa1



A.6b. Multiple alignment of cysteine proteinase amino acid sequences. Clustal alignment performed with Megalign (DNASTAR). The F1, F2 and B1 primer positions are shown overlined direction of primed synthesis is indicated by arrowheads.



A.6c. Multiple alignment of plant aspartic proteinase inhibitors. Clustal alignment performed with Megalign (DNASTAR). The MMKCLFL primer position is underlined the arrowhead indicates the direction of primed synthesis.

References.

Abe, K., Emori, Y., Kondo, H., Arai, S. and Suzuki, K. (1988). The NH2 Terminal-21 amino-acid residues are not essential for the papain-inhibitory activity of Oryzacystatin, a member of the cystatin superfamily – Expression of Oryzacystatin cDNA and its truncated fragments in *Escherichia-coli*. *J. Biol. Chem.* 263, 7655-7659.

Abe, K., Emori, Y., Kondo, H., Suzuki, K. and Arai, S. (1987). Molecular cloning of a cysteine proteinase inhibitor of rice (oryzacystatin). Homology with animal cystatins and transient expression in the ripening process of rice seeds. *J. Biol. Chem.* 262, 16793-16797.

Altman, F. (1996). *N*-glycosylation in insects revisited. *Trends Glycosci. Glyc.* 8, 101-114.

Applebaum, S. (1985). Biochemistry of digestion. (eds. Kerkut G. and Gilbert L.) In; *Comprehensive Insect Physiology, Biochemistry, and Pharmacology.* 4, 279-311. New York: Pergamon Press.

Azuma, T., Pals, G., Mohandas, T., Couvreur, J., Taggart, R. (1989). Human gastric cathepsin E. Predicted sequence, localization to chromosome 1, and sequence homology with other aspartic proteinases. *J. Biol. Chem.* 264, 16748-16753.

Banfalvi, Z., Molnar, A., Molnar, G., Lakatos, L. and Szabo, L. (1996). Starch synthesis, and tuber storage protein genes are differently expressed in *Solanum tuberosum* and in *Solanum brevidens*. *FEBS Lett.* 383, 159-164.

Baranski, T., Cantor, A. and Kornfield, S. (1992). Lysosomal enzyme phosphorylation I. *J. Biol. Chem.* 267, 23342-23348.

Baranski, T., Koelsch, G., Hartsuck, J. and Kornfield, S. (1991). Mapping and molecular modelling of a recognition domain for lysosomal enzyme targeting. *J. Biol. Chem.* 266, 23365-23372.

Barret, A., Rawlings, N., Davies, M., Machleidt, W., Salvesen, G., and Turk, V. (1986). (eds. Barret, A., and Salvesen, G.) In; *Proteinase Inhibitors*. 519-569. Amsterdam: Elsevier/North-Holland Biomedical Press.

Barret, A., Kembhavi, A., Brown, M., Kirschke, H., Knight, C., Tamai, M. and Hanada, K. (1982). L-*trans*-Epoxysuccinyl-L-leucylamido-(4-Guanidino)butane (E-64) and its analogues as inhibitors of cysteine proteinases including cathepsins B, H and L. *Biochem. J.* 201, 189-198.

Barrett, A., Rawlings, N. and Woessner, J (1998). *Handbook of proteolytic enzymes*. Academic Press.

Barrett, A. (1977). Cathepsin D and other carboxyl proteinases. (Ed: Barrett, A.) In; *Proteinases of mammalian Cells and Tissues.* Amsterdam: Elsevier/North-Holland Biomedical Press, 209-248.

Beveridge, A. (1998). A theoretical study of the initial stages of catalysis in the aspartic proteinases. *J. Mol. Struc.* (Theochem) 453, 275-291.

Billingsley, P. and Lehane, M. eds. (1996). In; *Biology of the insect midgut*, 3-15. London: Chapman and Hall.

Bode, W., Engh, R., Musil, D., Thiele, U., Huber, R., Karshikov, A., Brzin, J., Kos, J. and Turk, V. (1988). The 2.0 A X-ray crystal structure of chicken egg white cystatin and it's possible mode of action with cysteine proteinases. *EMBO J.* 7, 2593-2599.

Bolter, C. and Jongsma, M. (1995). Colorado potato beetles (*Leptinotarsa decemlineata*) adapt to proteinase inhibitors induced in potato leaves by methyl jasmonate. *J. Insect Physiol.* 41, 1071-1078.

Bonade-Bottino, M., Lerin, J., Zaccomer, B. and Jouanin, L. (1999). Physiological adaptation explains the insensitivity of *Baris coerulescens* to transgenic oilseed rape expressing oryzacystatin I. *Insect Biochem. Molec. Biol.* 29, 131-138.

Bousquet, Y. (1995). Publication 1861/E. Department of Agriculture and Agri-Food. http://sis.agr.gc.ca/brd/beetles/english/html/bhome_e.html

Bown, D., Wilkinson., H. and Gatehouse, J. (1998). Midgut carboxypeptidase from *Helicoverpa armigera* (Lepidoptera: Noctuidae) larvae: enzyme characterisation, cDNA cloning and expression. *Insect Biochem. Molec. Biol.* 28, 739-749.

Branson, T., Guss, P. Krysan, J., Sutter, G. (1975). Corn rootworms; Laboratory rearing and manipulation. U.S. Dept. Agric., ARS-NC-28.

Broadway, R. and Duffey, S. (1986). Plant proteinase inhibitors: mechanism of action and effect on growth and digestive physiology of larval *Heliothis zea* and *Spodoptera exigua*. *J. Insect Physiol.* 32, 827-833.

Broadway, R. (1995). Are insects resistant to plant proteinase inhibitors. *J. Insect Physiol.* 41, 107-116.

Cantor, A., Baranski, T. and Kornfield, S. (1992). Lysosomal enzyme phosphorylation II. *J. Biol. Chem.* 267, 23349-23356.

Carter, C., Hunt, T., Kline, T., Reagan, T, Barney, W., Baker, J. and Hunt, T. (1996). Insects and related pests of field crops. North Carolina Co-operative Extension ServiceAG-271.http://ipm.ncsu.edu/AG271/contents.html.

Cater, S., Lees, W., Hill, J., Brzin, J., Kay, J. and Phylip, L. (2002). Aspartic proteinase inhibitors from tomato and potato are more potent against yeast proteinase A than cathepsin D. *BBA*. 1596, 76-82.

Chiang, F. and Sun, C. (1993). Glutathione transferase isozymes of Diamondback moth larvae and their role in degradation of some organophosphorus insecticides. *Pestic. Biochem. Physiol.* 45, 7-14.

Chipinityol, S. and Crabbe, M. (1998). Chymosin and aspartic proteinases. *Food Chem.* 61, 395-418.

Cho, W., Deitsch, K. and Raikhel, A. (1991). An extraovarian protein accumulated in mosquito oocytes is a carboxypeptidase activated in embryos. *Proc. Natl. Acad. Sci. USA.* 88, 10821-4.

Cho, W, Dhadialla, T. and Raikhel, A. (1991). Purification and characterization of a lysosomal aspartic proteinase with cathepsin-D activity from the mosquito. *Insect Biochem.* 21, 165-176.

Cho, M., Lee, H., Lee, K., Homma, K., Natori, S. and Lee, B. (1999). Molecular cloning and functional properties of two early-stage encapsulation-relating proteins from the coleopteran insect, *Tenebrio molitor* larvae. *Eur. J. Biochem.* 262, 737-744.

Cho, W. and Raikhel, A. (1992). Cloning of cDNA for mosquito lysosomal aspartic protease. Sequence analysis of an insect lysosomal enzyme similar to cathepsins D and E. *J. Biol. Chem.* 267, 21823-21829.

Chrispeels, M. and Raikhel, N. (1992). Short peptide domains target proteins to plant vacuoles. *Cell.* 68, 613-616.

Clark, A. (1989). The comparative enzymology of the glutathione *S*-transferases from non-vertebrate organisms. *Comp. Biochem. Physiol.* 45, 7-14.

Consortium, C.e. (1998). Genome sequence of the nematode *C. elegans*: A platform for investigating biology. *Science* 282, 2012-2018.

Creelman, R., Tierny, M. and Mullet, J. (1992). Jasmonic acid/methyl jasmonate accumulate in wounded soybean hypocotyls and modulate wound gene expression. *Proc. Natl. Acad. Sci. USA.* 89, 4938-4941.

Cuozzo, J. and Sahagian, G. (1994). Lysine is a common determinant for mannose phosphorylation of lysosomal proteins. *J. Biol. Chem.* 269, 14490-14496.

Cuozzo, J., Tao, K., Wu, Qi-long, Young, W. and Sahagian, G. (1995). Lysine-based structure in the proregion of procathepsin L is the recognition site for mannose phosphorylation. *J. Biol. Chem.* 270, 15611-15619.

Cuozzo, J.W., Tao, K., Cygler, M., Mort, J.S., Sahagian G.G. (1998). Lysine-based structure responsible for selective mannose phosphorylation of cathepsin D and cathepsin L defines a common structural motif for lysosomal enzyme targeting. *J. Biol. Chem.* 273, 21067-76.

Cygler, M.and Coulombe, R. (1999). Crystal Structure of Procathepsin L. Publication; Coulombe R, Grochulski P, Sivaraman J, Menard R, Mort J, and Cygler, M. (1996). Structure of human procathepsin L reveals the molecular basis of inhibition by the prosegment. *EMBO J.* 15, 5492-503.

Damman, C., Rojo, E., and Sanchez-Seiranc, J. (1997). Abscisic acid and jasmonic acid activate wound inducible genes in potato through separate organ-specific signal transduction pathways. *Plant J.* 11, 773-82.

Danielson, P., MacIntyre, R. and Fogleman, J. (1997). Molecular cloning of a family of xenobiotic-inducible drosophilid cytochrome P450s: Evidence of involvement in host-plant allelochemical resistance. *Proc. Natl. Acad. Sci. USA.* 94, 10797-10802.

Day, E. (1996). Cucumber Beetles. Virginia Co-operative extension. Publication 444-222. http://www.ext.vt.edu/departments/entomology/factsheets/cucbeet.html

De Leo, F., Bonade-Bottino, M., Ceci, L., Gallerani, R. and Jouanin, L. (1998). Opposite effects on *Spodoptera littoralis* larvae of high expression level of a trypsin proteinase inhibitor in transgenic plants. *Plant Physiol.* 118, 997-1004.

Ding, X., Gopalakrishnan, B., Johnson, L., White, F., Wang, X., Morgan, T., Kramer, K. and Muthukrishnan, S. (1998). Insect resistance of transgenic tobacco expressing an insect chitinase gene. *Transgen. Res.* 7, 77-84.

Dittmer, N. and Raikhel, A. (1997). Analysis of the mosquito lysosomal aspartic protease Gene: An insect gene housekeeping gene with fat body-enhanced expression. *Insect Biochem. Molec. Biol.* 27, 323-335.

Dunn, B. (1996). Determination of protease mechanism. (eds. Benyan, R. and Bond, J) In; *Proteolytic enzymes: A practical approach.* 57-80. Oxford: IRL/Oxford University Press.

Dustin, M., Baranski, T., Sampath, D. and Kornfield, S. (1995). A novel mutagenesis strategy identifies distantly spaced amino acid sequences that are required for the phosphorylation of both oligosaccharides of procathepsin D by *N*-acetylglucosamine 1-phosphotransferases. *J. Biol. Chem.* 270, 170-179.

Edmonds, H. (1996). Antimetabolic proteins from plants and their potential use in conferring resistance against corn rootworms (Diabrotica SP). Ph.D. Thesis unpublished.

Edmonds, H., Gatehouse, L., Hilder, V. and Gatehouse, J. (1996). The inhibitory effects of the cysteine proteinase inhibitor, oryzacystatin, on digestive proteases and on larval survival and development of the southern corn rootworm (*Diabrotica undecimpunctata howardi*). *Entomol. Exp. Appl.* 78, 83-94.

Endo, Y. and Nishiitsutsuji-Uwo, J. (1980). Mode of action of *Bacillus thuringiensis* delta-endotoxin: histopathological changes in the silkworm (*Bombyx mori*) midgut. *J. Inverteb. Pathol.* 36, 90-103.

Ellis, R., Stockhoff, B., Stamp, L., Schnepf, H., Schwab, G., Knuth, M., Russell, J., Cardineau, G. and Narva, K. (2002). Novel *Bacillus thuringiensis* binary insecticidal crystal proteins active on western corn rootworm, *Diabrotica virgifera virgifera* LeConte. *Appl. Environ. Microbiol.*. 68, 1137-1145.

European and Mediterranean Plant Protection Organisation. (2002). http://www.eppo.org/QUARANTINE/Diabrotica_virgifera/first_reports.html

European and Mediterranean Plant Protection Organisation (1). (2002). http://www.eppo.org/QUARANTINE/Diabrotica_virgifera/key.html

Evers, M., Zelle, B., Bebelman, J., van Beusechem, V., Kraakman, L., Hoffer, M., Pronk, J., Mager, W., Planta, R., Eriksson, A. and Frants, R., (1989). Nucleotide sequence comparison of five human pepsinogen A (PGA) genes: evolution of the PGA multigene family. *Genomics*. 4, 232-239.

Fabrick, J., Behnke, C., Czapla, T., Bala, K., Rao, A., Kramer, K., and Reeck, G. (2002). Effects of a potato cysteine proteinase inhibitor on midgut proteolytic enzyme activity and growth of the southern corn rootworm, *Diabrotica undecimpunctata howardi* (Coleoptera: Chrysomelidae). *Insect Biochem. Molec. Biol.* 32, 405-415.

Farmer, E. and Ryan, C. (1990). Interplant communication: Airborne methyl jasmonate induces synthesis of proteinase inhibitors in plant leaves. *Proc. Natl. Acad. Sci. USA*. 87, 7713-7716.

Farmer, E. and Ryan, C. (1992a). Regulation of expression of proteinase-inhibitor genes by methyl jasmonate and jasmonic acid. *Plant Physiol.* 98, 995-1002.

Farmer, E. and Ryan, C. (1992b). Octadecanoid precursors of jasmonic acid activate the synthesis of wound-inducible proteinase inhibitors. *Plant Cell* 4, 129-134.

Faust, P., Wall, D., Perara, E., Lingpapa, V. and Kornfield, S. (1987). Expression of human cathepsin D in *Xenopus* oocytes: phosphorylation and intracellular targeting. *J. Cell Biology*. 105, 1937-45.

Felsot, A. (1989). Enhanced biodegradation of insecticides in soil: Implications for Agroecosystems. *Ann. Rev. Entomol.* 34, 453-476.

Feng, Q., Davey, K., Pang, A., Primavera, M., Ladd, T., Zheng, S., Sohi, S., Retnakaran A. and Palli, S., (1999). Glutathione *S*-transferase from the spruce budworm, *Choristoneura fumiferana*: identification, characterization, localization, cDNA cloning, and expression. *Insect Biochem. Molec. Biol.* 29, 779-793.

Ferre, J., Real, M., Van rie, J., Jansens, S. and Peferoen, M. (1991). Resistance to *Bacillus thuringiensis* bioinsecticide in a field population of *Plutella xylostella* is due to a change in a midgut membrane receptor. *Proc. Natl. Acad. Sci. USA.* 88, 5119-5123.

Feyereisen, R. (1999). Insect P450 enzymes. *Annu. Rev. Entomol.* 44, 507-533.

Finley, E. and Kornfield, S. (1994). Subcellular localization and targeting of cathepsin E. *J. Biol. Chem.* 269, 31259-31266.

Fournier, D., Bride, J., Poirie, M., Berge, J-B. and Plapp, F. Jr. (1992). Insect glutathione *S*-transferases. Biochemical characteristics of the major forms from Houseflies susceptible and resistant to insecticides. *J. Biol. Chem.* 267, 1840-1845.

Franciosca, H. and Berge, J. (1995). Glutathione *S*-transferases in Housefly (*Musca domestica*): Location of GST-1 and GST-2 families. *Insect Biochem. Molec. Biol.* 25, 311-317.

Frank, M. and Fogleman, J. (1992). Involvement of cytochrome P450 in host-plant utilization by Sonoran Desert *Drosophila*. *Proc. Natl. Acad. Sci. USA.* 89, 11998-12002.

Fuller, B., Boetel, M., Walgenbach, D., Grundler, J., Hein, G., Jarvi, K., Keaster, A., Landis., D, Meinke, L., Oleson, J., Ostlie, K., Tollefson, J., Wedburg, J., Wilde., G. and Evenson., P. (1997). Optimization of soil insecticide rates for managing corn rootworm (Coleoptera: Chrysomelidae) larvae in north central United States. *J. Econ. Entomol.* 90, 1332-1340.

Gal, S. and Gottesman, M. (1988). Isolation and sequence of a cDNA for human pro-cathepsin L. *Biochem. J.* 253, 303-306.

Gatehouse, A., Norton, E., Davison, G., Babbe, S., Newell, C. and Gatehouse, J. (1999). Digestive proteolytic activity in larvae of tomato moth, *Lacanobia oleracea*; effects of plant protease inhibitors *in vitro* and *in vivo*. *J. Insect Physiol.* 45, 545-558.

Gatehouse, A., Shi, Y., Powell, K., Brough, C., Hilder, V., Hamilton, W., Newell, C., Merryweather, A., Boulter, D. and Gatehouse J. (1993). Approaches to insect resistance using transgenic plants. *Phil. Trans. R. Soc. Lond. B.* 342, 279-286.

Gergerich, R., Scott, H. and Fulton, J. (1986). Evaluation of Diabrotica beetles as vectors of plant viruses. In *Methods for the study of Pest Diabrotica*. (eds Kyrsan, J. and Miller, T.) Springer-Verlag, New York.

Gilliot, C. (1980). *Entomology.* New York: Plenum Press.

Girard, C. and Jouanin, L. (1999a). Molecular cloning of a gut-specific chitinase cDNA from the beetle *Phaedon cochleariae. Insect Biochem. Molec. Biol.* 29, 549-556.

Girard, C. and Jouanin, L. (1999b). Molecular cloning of a range of digestive enzymes from a phytophagous beetle, *Phaedon cochleariae. Insect Biochem. Molec. Biol.* 29, 1129-1142.

Girard, C., Le Metayer, M., Zaccomer, B., Bartlet, E., Williams, I., Bonade-Bottino, M., Pham-Delegue, M. and Jouanin, L. (1998a). Growth stimulation of beetle larvae reared on transgenic oilseed rape expressing a cysteine proteinase inhibitor. *J. Insect Physiol.* 44, 263-270.

Girard, C., Le Metayer, M., Bonade-Bottino, M., Pham-Delegue, M., and Jouanin, L. (1998b). High level of resistance to proteinase inhibitors may be conferred by proteolytic cleavage in beetle larvae. *Insect Biochem. Molec. Biol.* 28, 229-237.

Gongora, C., Wang, S., Barbehenn, R. and Broadway, R. (2001). Chitinolytic enzymes from *Streptomyces albidoflavus* expressed in tomato plants: effects on *Trichoplusia ni. Entomol. Exp. Appl.* 99, 193-204.

Gonzalez, F. and Nerbert, D. (1990). Evolution of the P450 gene superfamily: Animal-plant "warfare", molecular drive, and human genetic differences in drug oxidation. *Trends Genet.* 6, 182-186.

Gopalakrishnan, B., Muthukrishnan, S. and Kramer, K.J., (1995). Baculovirus-mediated expression of a *Manduca sexta* chitinase gene: Properties of the Recombinant Protein. *Insect Biochem. Molec. Biol.* 25, 255-265.

Gould, F., Anderson, A., Jones, A., Sumerford, D., Heckel, D.G., Lopez, J., Micinski, S., Leonard, R., Laster, M. (1997). Initial frequency of alleles for *Bacillus thuringiensis* toxins in field populations of *Heliothis virescens*. *Proc. Natl. Acad. Sci. USA.* 94, 3519-3523.

Gould, G. (1969). Insect pests of cucurbit crops in Indiana. *Ind. Acad. Sci. Proc.* 53, 165-171.

Grant, D. and Matsumara, F. (1989). Glutathione *S*-transferase 1 and 2 in susceptible and insecticide resistant *Aedes aegypti*. *Pest. Biochem. Physiol.* 33, 132-143.

Guex, N., Diemand, A., Scwede, T. and Peitsch, M. Swiss PDB Viewer (Deepview). Version 3.7b1. Glaxo Wellcome Experimental Research.

Guex, N. and Peitsch, M. (1997). SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modelling. *Electrophoresis.* 18, 2714-2723.

Guzov, V., Houston, H., Murataliev, M., Walker, F. and Feyereisen, R. (1996). Molecular cloning, overexpression in *E. coli.*, structural and functional characterisation of house fly cytochrome b5. *J. Biol. Chem.* 271, 26637-45.

Hanada, K., Tamai, M., Yamagishi, M., Ohmura, S., Sawada, J., Tanaka, I. (1978). Isolation and characterisation of E-64, a new thiol protease inhibitor. *Agric. Biol. Chem.* 42, 523-528.

Harsulkar, A., Giri, A., Patankar, A., Gupta, V., Sainani, M., Ranjekar, P. and Deshpande, V. (1999). Successive use of non-host plant proteinase inhibitors required for effective inhibition of *Helicoverpa armigera* gut proteinases and larval growth. *Plant Physiol.* 121, 497-506.

Hazelton, G. and Lang, C. (1983). Glutathione *S*-transferase activities in the yellow-fever mosquito [Aedes aegypti (Louisville)] during growth and ageing. *Biochem. J.* 210, 281-287.

Hemmingway, J. (2000). The molecular basis of two contrasting metabolic mechanisms of insecticide resistance. *Insect Biochem. Molec. Biol.* 30, 1009-1015.

Hemmingway, J., Miyamoto, J. and Herath, P. (1991). A possible novel link between organophosphorus and DDT insecticide resistance genes in *Anopheles* supporting evidence from fenitrothion metabolism studies. *Pestic. Biochem. Physiol.* 39, 49-56.

Herber, K., Prat, S., Willmitzer, L. (1994). Cloning and characterization of a cathepsin D inhibitor gene from *Solanum tuberosum* L. Plant Mol. Biol. 26, 73-83

Hilder, V. and Boulter, D. (1999). Genetic engineering of crop plants for insect resistance - a critical review. *Crop protection.* 18, 177-191.

Hilder, V., Gatehouse, A., Sheerman, S., Barker, R., and Boulter, D. (1987). A novel mechanism of insect resistance engineered into tobacco. *Nature.* 330, 160-163.

Hildmann, T., Ebneth, M., Pena-Cortes, H., Sanchez-Serrano, J.J., Willmitzer, L., Prat, S. (1992). General roles of abscisic and jasmonic acids in gene activation as a result of mechanical wounding. *Plant Cell* 4, 1157-1170.

Hofte, H. and Whiteley, H. (1989). Insecticidal crystal proteins of *Bacillus thuringiensis*. *Microbiol. Rev.* 53, 242-255.

Hsu, T., Takahashi, N., Tsukamoto, Y., Kato, K., Shimada, I., Masuda, K., Whiteley, E., Fan, J., Lee, Y. and Betenbaugh, M. (1997). Differential *N*-glycan patterns of secreted and intracellular IgG produced in *Trichoplusia ni* cells. *J. Biol. Chem.* 272, 9062-9070.

Huang F, Zhu K, Buschman L, Higgins R. and Oppert B. (1999). Comparison of midgut proteinases in *Bacillus thuringiensis*-susceptible and -resistant European corn borer, *Ostrinia nubilalis* (Lepidoptera: *Pyralidae*). *Pestic. Biochem. Physiol.* 65, 132-139.

Huang, H., Hu, N., Yao, Y., Wu, C., Chiang, S. and Sun, C. (1998). Molecular cloning and heterologous expression of a glutathione *S*-transferase involved in insecticide resistance from the diamond back moth, *Plutella xylostella*. *Insect Biochem. Molec. Biol.* 28, 651-658.

Huang, X. (1994). On Global Sequence Alignment. *Comput. Appl. Biosci.* 10, 227-235.

Huang, X., Zhang, H., Zen, K., Muthukrishnan, S. and Kramer, K. (2000). Homology modelling of the insect chitinase catalytic domain-oligosaccharide complex and the role of a putative active site tryptophan in catalysis. *Insect Biochem. Molec. Biol.* 30, 107-117.

Ishikawa, A., Ohta, S., Matsuoka, K., Hattori, T. and Nakamura, K. (1994a). A family of potato genes that encode Kunitz-type proteinases inhibitors: structural comparisons and differential expression. *Plant Cell Physiol.* 35, 303-312.

Ishikawa, A., Yoshihara, T. and Nakamura, K. (1994b). Jasmonate-inducible expression of a potato cathepsin D inhibitor-GUS gene fusion in tobacco cells. *Plant Mol. Biol.* 26, 403-414.

Jaskolski, M. Tomasselli, A., Sawyer, T., Staples, D., Heinrikson, R., Schneider, J., Kent, B. and Wlodower, A. (1991). Structure at 2.5-Å Resolution of chemically synthesized human-immunodeficiency-virus type-1 protease complexed with a hydroxyethylene-based inhibitor. *Biochemistry* 30, 1600-1609.

Johnson, H., Navarez, J., An, G., Ryan, C. (1989). Expression of proteinase inhibitors I and II in transgenic tobacco plants: Effects on natural defense against *Manduca sexta* larvae. *Proc. Natl. Acad. Sci. USA* 86, 9871-9875.

Jolodar, A. and Miller, D. (1998). Identification of a novel family of non-lysosomal aspartic proteases in nematodes. *BBA - Proteins and Proteomics* 1382, 13-16.

Jongsma, M. and Bolter, C. (1997). The adaptation of insects to plant proteinase inhibitors. *J. Insect Physiol.* 43, 885-895.

Jongsma, M. Bakker, L., Peters, J., Bosch, D. and Stiekma, W. (1995). Adaptation of *Spodoptera exigua* larvae to plant proteinase inhibitors by induction of gut proteinase activity insensitive to inhibition. *Proc. Natl. Acad. Sci. USA.* 92, 8041-8045.

Jouanin, L., Bonade-Bottino, M., Girard, C., Morrot, G. and Giband, M. (1998). Transgenic plants for insect resistance. *Plant Science.* 131, 1-11.

Kane, S. (1993). Mouse procathepsin L lacking a functional glycosylation site is properly folded, stable and secreted by NIH 3T3 cells. *J. Biol. Chem.* 268, 11456-11462.

Karlsen, S., Hough E. and Olsen, R. (1997). The crystal structure and proposed amino acid sequence of a pepsin from Atlantic Cod (*Gadus morhua*). PDF: ExPDB; 1AM5.

Karrer, K., Peiffer, S. and Ditomas, M. (1993). Two distinct gene subfamilies within the family cysteine protease genes. *Proc. Natl. Acad. Sci. USA.* 90, 3063-3067.

Kasai, S., Weerashinge, I., Shono, T. and Yamakawa, M. (2000). Molecular cloning, nucleotide sequence and gene expression of a cytochrome P450 (CYP6F1) from the pyrethroid-resistant mosquito, *Culex quinquefasciatus Say. Insect Biochem. Molec. Biol.* 30, 163-171.

Keilova, H. and Tomasek, V. (1976). Isolation and properties of cathepsin D inhibitor from potatoes. *Collect Czech. Chem. Commun.* 41, 489-497.

Keilova, H. and Tomasek, V. (1977). Naturally occurring inhibitors of intracellular proteinases. *Acta. Biol. Med. Ger.* 36, 1873-8.

Keller M, Sneh B, Strizhov N, Prudovsky E, Regev A, Koncz C, Schell J, Zilberstein A. (1996). Digestion of delta-endotoxin by gut proteases may explain reduced sensitivity of advanced instar larvae of *Spodoptera littoralis* to CrylC. *Insect Biochem. Molec. Biol.* 26, 365-373.

Knowles, B. (1994). Mechanisms of action of *Bacillus thuringiensis* insecticidal endotoxins. *Adv. Ins. Physiol.* 24, 275-308.

Koga, D., Sasaki, Y., Uchiumi,Y., Hirai, N., Arakane, Y. and Nagamatsu,Y., (1997). Purification and characterization of *Bombyx mori* chitinases. *Insect Biochem. Molec. Biol.* 27, 757-767.

Koiwa, H., Shade, R., Zhu-Salzman, K., D'Urzo, M., Murdock, L., Bressan, R. and Hasegawa. (2000). A plant defensive cystatin (soyacystatin) targets cathepsin L-like digestive proteinases (DvCALs) in the larval midgut of Western corn rootworm (*Diabrotica virgifera virgifera*). *FEBS Lett.* 471, 67-70.

Kondo, H., Abe, K., Nishimura, I., Watanbe, H., Emori, Y. and Arai, S. (1990). Two distinct cystatin species in rice seeds with different specificities against cysteine proteinases. *Proc. Natl. Acad. Sci. USA.* 265, 15832-15837.

Kornfield, S. and Mellman, I. (1989). The Biogenesis of Lysosomes. *Ann. Rev. Cell Biol.* 483-516.

Kostaropoulos, I., Mantzari, A. and Papadopoulos, A., (1996). Alterations of some glutathione *S*-transferase characteristics during the development of *Tenebrio molitor* (Insecta: Coleoptera). *Insect Biochem. Molec. Biol.* 26, 963-969.

Kostaropoulos, I., Papadopoulos, A., Metaxakis, A., Boukouvala, E. and Papadopoulou-Mourkidou, E., (2001). Glutathione *S*-transferase in the defence against pyrethroids in insects. *Insect Biochem. Molec. Biol.* 31, 313-319.

Kota, M., Daniell, H., Varma, S., Garczynski, S., Gould, F. and Moar, W. (1999). Overexpression of the *Bacillus thuringiensis* (Bt) Cry2Aa2 protein in chloroplasts confers resistance to plants against susceptible and Bt-resistant insects. *Proc. Nat. Acad. Sci. USA.* 96, 1840-1845.

Kramer, K. and Muthukrishnan, S. (1997). Insect Chitinases: Molecular biology and potential use as biopesticides. *Insect Biochem. Molec. Biol.* 27, 887-900.

Kramer, K., Corpuz, L., Choi, H. and Muthukrishnan, S., (1993). Sequence of a cDNA and expression of the gene encoding epidermal and gut chitinases of *Manduca sexta*. *Insect Biochem. Molec. Biol.* 23, 691-701.

Kreft, S., Ravnikar, M., Mesko, P., Pungecar, J., Umek, A., Kregar, I. and Strukelj, B. (1997). Jasmonic acid inducible aspartic proteinase inhibitors from potato. *Phytochemistry* 44, 1001-6.

Ku, C., Chiang, F., Hsin, C., Yao, Y. and Sun, C., (1994). Glutathione transferase isozymes involved in insecticide resistance of Diamondback moth larvae. *Pestic. Biochem. Physiol.* 50, 191-197.

Kyte, J. and Doolittle, R. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157, 105-132.

Landis, D., Levine, E. Haas, M., and Meints, V. (1992). Detection of prolonged diapause of Northern corn-rootworm in Michigan (Coleoptera, Chrysomelidae). *Great Lakes Entomol.* 25, 215-222.

Lang, S, Reitman, M., Tang, J., Roberts, M. and Kornfield, S. (1984). Lysosomal enzyme phosphorylation. *J. Biol. Chem.* 259, 14663-14671.

Lara, P., Ortego, F., Gonzalez-Hidalgo, E., Casternera, P., Carbonero, P. and Diaz, I. (2000). Adaptation of *Spodoptera exigua* (Lepidoptera: Noctuidae) to barley trypsin inhibitor BTI-Cme expressed in transgenic tobacco. *Transge. Res.* 9, 168-178.

Lehane, M. and Billingsley, P. eds. (1996). *Biology of the insect midgut.* London: Chapman and Hall.

Lehane, M., Blakemore, D., Williams, S. and Moffatt, M. (1995). Regulation of digestive enzyme levels in insects. *Comp. Biochem. Physiol.* 110B, 285-289.

Lehane, M.J., (1997). Peritrophic membrane structure and function. *Annu. Rev. Entomol.* 42, 525-550.

Lemos, F. and Terra, W. (1991). Properties and intracellular distribution of a cathepsin D-like proteinase active at the acid region of *Musca domestica* midgut. *Insect Biochem.* 21, 457-465.

Levine, E. and Oloumi-sadeghi, H. (1991). Management of diabroticite rootworms in corn. *Annu. Rev. Entomol.* 36, 229-255.

Levine, E., Oloumi-sadeghi, H. and Ellis, C. (1992). Thermal requirements, hatching patterns, and prolonged diapause in Western corn-rootworm (Coleoptera, Chrysomelidae) eggs. *J. Econ. Entomol.* 85, 2425-2432.

Li, J., Carroll, J. and Ellar, D. (1991). Crystal-structure of insecticidal delta-endotoxin from *Bacillus-thuringiensis* at 2.5-A resolution. *Nature.* 353, 815-821.

Li, X., Berenbaum, M. and Schuler. M. (2000). Molecular cloning and expression of CYP6B8: a xanthotoxin inducible cytochrome P450 cDNA from *Helicoverpa zea*. *Insect Biochem. Molec. Biol.* 30, 75-84.

Liang, Y., Brookhart, G., Feng, G., Reeck, G. and Kramer, K. (1991). Inhibition of digestive proteinases of stored grain Coleoptera by oryzacystatin, a cysteine proteinase inhibitor from rice seeds. *FEBS Lett.* 278, 139-142.

Liu, N. and Scott, J. (1998). Increased transcription of CYP6D1 causes cytochrome P450 mediated insecticide resistance in house fly. *Insect Biochem. Molec. Biol.* 28, 531-535.

Lynn, D. and Stoppleworth, A. (1984). Established cell lines from the beetle, *Diabrotica undecimpunctata* (Coleoptera, Chrysomelidae). *In vitro Journal of the Tissue Culture Association.* 20, 365-368.

Machleidt, W., Ritonja, A., Popovic, T., Kotnik, M., Brzin, J., Turk, V., Machleidt, I. and Mueller-Esterl, W. (1986). Human cathepsins B, H and L: characterization by amino acid sequences and some kinetics of inhibition by the kininogens. Turk, V. eds. In; *Cysteine proteinases and their inhibitors*: 3-18.

Maganja, D.B., Strukelj, B., Pungercar, J., Gubensek, F., Turk,V. and Kregar, I. (1992). Isolation and sequence analysis of the genomic DNA fragment encoding an aspartic proteinase inhibitor homologue from potato (*Solanum tuberosum L.*). *Plant Mol. Biol.* 20, 311-313.

Mares, M., Meloun, B., Pavlik, M., Kostka, V. and Baudys, M. (1989). Primary structure of cathepsin D inhibitor from potatoes and its structure relationship to soybean trypsin inhibitor family. *FEBS Lett.* 251, 94-98.

Margis, R., Reis, E. and Villeret, V. (1998). Structural and phylogenetic relationships among plant and animal cystatins. *Arch. Biochem. Biophys.* 359, 24-30.

Mason, R., Wilcox, D., Wikstrom, P. and Shaw, E. (1989). The identification of active forms of cysteine proteinases in Kirsten-virus-transformed mouse fibroblasts by use of a specific radiolabelled inhibitor. *Biochem. J.* 257, 125–129.

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

Matsumoto, I., Keiko, A., Soichi, A. and Yasufumi, E. (1998). Functional expression and enzymatic properties of two *Sitophilus zeamais* cysteine Proteinases showing different autolytic processing profiles *in vitro*. *J. Biochem.* 123, 693-700.

Matsumoto, I., Watanbe, H., Abe, K., Arai, S. and Emori, Y. (1995). A Putative digestive cysteine proteinase from *Drosophila melanogaster* is predominately expressed in the embryonic and larval midgut. *Eur. J. Biochem.* 227, 582-587.

Matsuoka, K. and Nakamura, K. (1991). Propeptide of a precursor to a plant vacuolar protein required for vacuolar targeting. *Proc. Natl. Acad. Sci. USA.* 88, 834-838.

McGurl, B., Pearce, G., Orozco-Cardenas, M., and Ryan, C. (1992). Structure, expression and antisense inhibition of the systemin precursor gene. *Science.* 255, 1570-1573.

McPherson, R.M., Douce, G.K., Riley, D.G., Hudson, R. D., and J. N. All. (1996). Summary of losses from insect damage and costs of control in Georgia 1995. *The University of Georgia Special Publication.* 90.

Meloche, F., Filion, P., Tremblay, G. and Lesage, L. (2001). Advance of *Diabrotica virgifera virgifera* (Coleoptera : Chrysomelidae) in corn fields in Quebec and sampling in soybean plants in Ottawa, Ontario. *Phytoprotection,* 82, 35-38.

Metcalf, R. (1986). Foreword in Krysan, J. and Miller, T. (1986). eds *Methods for the study of pest Diabrotica.* New York: Springer-Verlag. 260.

Michaelides, P. (1997a). Activity of soil insecticides on eggs of *Diabrotica undecimpunctata howardi*: effects on embryological development and influence of egg age. *Pestic. Sci.* 49, 1-8.

Michaelides, P., Cleverly, A. and Wright, D. (1997b). Sub-lethal effects of tefluthrin on *Diabrotica undecimpunctata howardi* Barber: plant protection and larval development. *Crop protection.* 16, 423-429.

Michaud, D., Nguyen-Quoc, B. and Yelle, S. (1993). Selective inactivation of Colorado potato beetle cathepsin H by oryzacystatin I and II. *FEBS Lett.* 331, 173-176.

Milne, R. and Kaplan, H. (1993). Purification and characterisation of a trypsin-like digestive enzyme from spruce budworm (*Choristoneura-Fumiferana*) responsible for the activation of Delta-endotoxin from *Bacillus-thuringiensis. Insect Biochem. Molec. Biol.* 23, 663-673.

Moberg, W. (1990). Understanding and combating agrochemical resistance. (Eds Green, M., LeBaron, H., Moberg, W.) In; *Managing resistance to agrochemicals. Acs. Sym. Ser.* 421, 3-16.

Moellenbeck, D., Peters, M., Bing, J., Rouse, J., Higgins, L., Sims, L., Nevshemal, T., Marshall, L., Ellis, R., Bystrak, P., Lang, B., Stewart, J., Kouba, K., Sondag, V., Gustafson, V., Nour, K., Xu, D., Swenson, J., Zhang, J., Czapla. T., Schwab. G., Jayne, S., Stockhoff, B., Narva, K., Schnepf, H., Stelman, S., Poutre, C., Koziel, M. and Duck, N. (2001). Insecticidal proteins from *Bacillus thuringiensis* protect corn from corn rootworms. *Nat. Biotechnol.* 19, 668-672.

Montell, D., Rørth, P. and Spradling A. (1992). *Slow border cells*, a locus required for a developmentally regulated cell migration during oogenesis, encodes *Drosophila* C/EBP. *Cell* 71, 51-62.

Moravek, L. and Kostka, V. (1974). Complete amino acid sequence of hog pepsin. *FEBS Lett.* 43, 207-211.

Murdock, L., Brookhart, G., Dunn, P., Foard, D., Kelley, S., Kitch, L., Shade, R., Shulke, R., Wolfson, J., (1987). Cysteine digestive proteinases in Coleoptera. *Comp. Biochem. and Physiol.* 87B, 783-787.

Nagata, K., Kudo., N., Abe, K., Arai., S. and Tanokura. M. (2000). Three-dimensional solution structure of oryzacystatin-1, a cysteine proteinase inhibitor of the rice, *Oryza sativa*. japonica. *Biochemistry* 39, 14753-14760.

Nandi, K., Basu, D., Das, S. and Sen S. (1999). High expression of soybean trypsin inhibitor gene in transgenic tobacco plants failed to confer resistance against damage by *Helicoverpa armigera. J. Biosciences* 24, 445-452.

Nelson D., Koymans, L., Kamataki, T., Stegeman, J., Feyereisen, R., Waxman, D., Waterman, M., Gotoh, O., Coon, M., Estabrook., R., Gunsalus, I., Nebert, D. (1996). P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogentics.* 6, 1-42.

Nerbert, D., Nelson, D., Coon, M., Estabrook, R., Feyereisen, R., Fujii-Kuriyama, Y., Gonzalez, F., Guengerich, F., Gunsalus, I., Johnson, E., Loper, J., Sato, R., Waterman, M. and Waxman, D. (1991). The P450 superfamily: update on new sequences, gene mapping, and recommended nomenclature. *DNA Cell Biol,* 10, 1-14.

Nielsen, H and Krogh, A. (1998). Prediction of signal peptides and signal anchors by a hidden Markov model. In; *Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology.* 6, 122—130. AAAI Press, Menlo Park, California.

Nielsen, H., Engelbrecht, J., Brunak, S., and Heijne, G., (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* 10, 1-6.

Novagen Inc, (2000). pET expression System Manual. TB055, 9th edition, 31.

Oerke, E. (1994). Estimated losses due to pathogens animal pests and weeds. In Estimated losses in food and cash crops. (eds. Oerke, E., Dehne, H., Schonbeck, F. and A. Weber) In; *Crop production and Crop protection.* Amsterdam: Elsevier, 830.

Onstad, D., Joselyn, M., Israd, S., Levine, E., Spencer, J., Bledsoe, L. Edwards, C., Di Fonzo, D. and Wilson, H. (1999). Modelling the spread of western corn rootworm (Coleoptera : Chrysomelidae) populations adapting to soybean-corn rotation. *Environ. Entomol.* 28, 188-194.

Oppert, B., Kramer, K., Beeman, R., Johnson, D. and McGaughey, W. (1997). Proteinase mediated resistance to *Bacillus thuringiensis* toxins. *J. Biol. Chem.* 272, 23473-23476.

Orr, G., Strickland, J. and Walsh, T. (1994). Inhibition of Diabrotica larval growth by a multicystatin from potato tubers. *J. Insect Physiol.* 40, 893-900.

Ortego, F., Farinos, G., Ruiz, M., Marco, V. and Castanera, P. (1998). Characterisation of digestive proteinases in the weevil *Aubeonymus mariaefranciscae* and effects of proteinase inhibitors on larval development and survival. *Entomol. Exp. Appl.* 88, 265-274.

Ortego, F., Novillo, C. and Casternera, P. (1996). Characterisation and distribution of digestive proteases of the stalk corn borer, *Sesamia nonagrioides* Lef (Lepidoptera: Noctuidae). *Archives of Insect Biochemistry and Physiology*. 33, 163-180.

Pearce, G., Strydom, D., Johnson., S., and Ryan, C. (1991). A polypeptide from tomato leaves induces wound-inducible proteinase inhibitor proteins. *Science*. 253, 895-898.

Pech, L., and Michael, R. (1996). Granular cells are required for encapsulation of foreign targets by insect haemocytes. *J. Cell Sci.* 109, 2053-2060.

Peña-Cortés, H., Sanchez-Serrano, J., Mertens, R., Willmitzer, L. and Prat, S. (1989). Abscisic acid is involved in the wound-induced expression of the proteinase inhibitor II gene in potato and tomato. *Proc. Natl. Acad. Sci. USA.* 86, 9851-9855.

Peña-Cortés, H., Fisahn, J. and Willmitzer, L. (1995). Signals involved in wound-induced proteinase inhibitor II gene expression in tomato and potato plants. *Proc. Natl. Acad. Sci. USA.* 92, 4106-4113.

Perez, C. and Shelton, A. (1997). Resistance of *Plutella xylostella* (Lepidoptera: Plutellidae) to *Bacillus thuringiensis* Berliner in Central America. *J. Econ. Entomol.* 90, 87-93.

Peters, W. and Latka, I., (1992). Peritrophic Membranes. (Eds. Bradshaw, D., Burggren, W., Heller, H., Ishii, S., Langer, H., Neuweiler, G., Randall, D.) In; *Zoophysiology*. 130. Berlin: Springer-Verlag. 1-238.

Pimentel, D. (1992). Pesticides and World Food Supply. *Acs. Sym. Ser.* 483, 309-326.

Pimentel, D., Acquay, H., Biltonen, M., Rice, P., Silva, M., Nelson, J., Lipner, V., Giodano, S., Horowitz, A. and D'Amore, M. (1992). Environmental and economic cost of pesticide use. *Bioscience* 42, 750-760.

Pinstrup-Andersen, P. and Pandya-Lorch, R. (2001). Emerging food and agriculture concerns in the developing world. *Agric. Rural Develop. J.* 8, 28-30.

Pollitt, S. and Zalkin, H. (1983). Role of primary structure and di-sulfide bond formation in Beta-Lactamase secretion. *J. Bacteriol.* 153, 27-32.

Prapanthadra, L., Ranson, H., Somboon, P. and Hemingway, J. (1998). Cloning, expression and characterisation of an insect class I glutathione *S*-transferase from *Anopheles dirus* species B. *Insect Biochem. Molec. Biol.* 28. 321-329.

Prence, E., Dong, J. and Sahagian, G. (1990). Modulaton of the transport of a lysosomal-enzyme by PDGF. *J. Cell Biol.* 110 (2): 319-326.

Purcell, J., Greenplate, J. and Sammons, R. (1992). Examination of midgut luminal proteinase activities in six economically important insects. *Insect Biochem. Molec. Biol.* 22, 41-47.

Qiagen. (1998). The QIAexpressionistTM; A handbook for high-level expression and purification of 6×His-tagged proteins. Third edition.

Rancour, J. and Ryan, C. (1968). Isolation of a carboxypeptidase B inhibitor from potatoes. *Arch. Biochem. Biophys.* 125, 380-382.

Reitman, M. and Kornfield, S. (1981). Lysosomal enzyme targeting. *J. Biol Chem.* 256, 11977-11980.

Renard, G., Garcia, J.F., Cardoso, F.C., Richter, M.F., Sakanari, J.A., Ozaki, L.S., Termignoni, C. and Masuda, A. (2000). Cloning and functional expression of a *Boophilus microplus* cathepsin L-like enzyme. *Insect Biochem. Molec. Biol.* 30, 1017-1026.

Riggio, M., Scudiero, R., Filosa, S. and Parisi, E. (2000). Sex- and tissue-specific expression of aspartic proteinases in *Danio rerio* (zebrafish). *Gene* 260, 67-75.

Ritonja, A., Popovic, T., Kotnik, M., Machleidt, W. and Turk, V. (1988). Amino acid sequences of the human kidney cathepsins H and L. *FEBS Lett.* 228, 341-345.

Ritonja, A., Krizaj, I., Mesko, P., Kopitar, M., Lucovnik, P., Strukelj, B., Pungercar, J., Buttle, D., Barret, A. and Turk, V. (1990). The amino acid sequence of a novel inhibitor of cathepsin D from potato. *FEBS Lett.* 267, 13-15.

Rodis, P. and Hoff, J. (1984). Naturally occurring protein crystals in the potato. *Plant Physiol.* 74, 907-911.

Rowan, A., Mason, P., Mach, L. and Mort, J. (1992). Rat Procathepsin-B proteolytic processing to the mature form in vitro. *J. Biol. Chem.* 267, 15993-15999.

Rupar, M., Donovan, W., Gene Groat, R., Slaney, A., Mattison, J., Johnson, T., Jean-Francois, C., Cosmao Dumanoir, V. and De Barjac, H. (1991). Two novel strains of *Bacillus thuringiensis* toxic to Coleopterans. *App. and Environ. Micro.* 57, 3337-3344.

Ryan, C. (1981). Proteinase inhibitors. (eds. A. Marcus) In; *The Biochemistry of Plants, A Comprehensive Treatsie*. New York/San-Francisco/London: Academic Press. 6, 351-371.

Ryan, C. (1989). Proteinase inhibitor gene families: Strategies for transformation to improve plant defenses against herbivores. BioEssays. 10, 20-24.

Ryan, C. (1990). Protease Inhibitors in Plants: Genes for improving defenses against insects and pathogens. *Annu. Rev. Phytopathol.* 28, 425-49.

Ryerse, J., Purcell, J. and Sammons, R. (1994). Structure and formation of the peritrophic membrane in the larva of the Southern Corn Rootworm, *Diabrotica-undecimpunctata*. *Tissue & Cell*. 26, 431-437.

Sambrook, J., Fritsch, E. and Maniatas, T. (1989). Molecular cloning: A laboratory manual. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

Santamaria, I., Velasco, G., Pendas, A., Paz, A. and Lopez-Otin, C. (1999). Molecular cloning and structural and functional characterisation of human cathepsin F, a new cysteine proteinase of the papain family with a long propeptide domain. *J. Biol. Chem.* 274, 13800-13809.

Scharf, M., Meinke, L., Wright, R., Chandler, L. and Siegfried, B. (1999). Metabolism of carbaryl by insecticide-resistant and -susceptible western corn rootworm populations (Coleoptera: *Chrysomelidae*). *Pestic. Biochem and Physiol.* 63, 86-96.

Scharf, M., Siegfried, B., Meinke, L., Wright, R. and Chandler, L. (2000). Cytochrome P450-mediated *N*-demethylation activity and induction in insecticide resistant and susceptible western corn rootworm populations (Coleoptera: *Chrysomelidae*). *Pestic. Biochem. Physiol.* 67. 137-143.

Schein, C., and Noteburn, M. (1989). Formation of soluble recombinant proteins in *Escherichia coli* is favoured by low growth temperature. *Bio. Technol.* 6, 291-294.

Scott, J. (1999). Cytochrome P450 and insecticide resistance. *Insect Biochem. Molec. Biol.* 29, 757-777.

Shi, G., Webb, A., Foster, K, Knoll, J., Lemere, C., Munger, J. and Chapman H. (1994). Human Cathepsin S: Chromosomal localization, gene structure, and tissue distribution. *J. Biol. Chem.* 269, 11530-11536.

Shinoda, T., Kobayashi, J., Matsui, M. and Chinzei, Y., (2001). Cloning and functional expression of a chitinase cDNA from the common cutworm, *Spodoptera litura*, using a recombinant baculovirus lacking the virus-encoded chitinase gene, *Insect Biochem. Molec. Biol.* 31, 521-532.

Sivaraman, J., Lalumiere, M., Menard, R. and Cygler, M. (1999). Crystal structure of wild-type human procathepsin K. *Protein Sci.* 8, 283-290.

Slaney, A., Robbins H. and English, L. (1992). Mode of action of *Bacillus thuringiensis* toxin CryIIIA: an analysis of toxicity in *Leptinotarsa decemlineata* (Say) and *Diabrotica undecimpunctata howardi* Barber. *Insect Biochem. Molec. Biol.* 22, 9-18.

Smith, S., Kane, S., Gal, S., Mason, R., and Gottesman, M. (1989). Glycosylation of Procathepsin L does not account for species molecular-mass differences and is not required for proteolytic activity. *Biochem. J.* 262, 931-938.

Snyder, M., Walding, J. and Feyereisen, R. (1995). Glutathione *S*-transferases from larval *Manduca sexta* midgut: Sequence of two cDNAs and enzyme induction. *Insect Biochem. Molec. Biol.* 25, 455-456.

Sorenson, K., Baker, J., Carter, C. and Stephan, D. Insects and related pests of Vegetables. Centre for Integrated Pest Management, North Carolina State University. AG-295. http://ipm.ncsu.edu/AG295/html/cucumber_beetles.htm

Stevens, J, Snyder, M., Koener., J. and Feyereisen, R. (2000). Inducible P450s of the CYP9 family from larval *Manduca sexta* midgut. *Insect Biochem. Molec. Biol.* 30, 559-568.

Stoner, K. (1996). Cucumber beetles; *Acalymma vittatum* (F.) and *Diabrotica undecimpunctata howardi* (Barber). Department of Entomology, Connecticut Agricultural Experiment Station. http://www.caes.state.ct.us/FactSheetFiles/entomology/fsen006f.htm

Strukelj, B., Pungercar, J., Mesko, P., Barlic-Maganja, D., Gubensek, F., Kregar, I. and Turk,V. (1992). Characterisation of aspartic proteinase inhibitors from potato at the gene, cDNA and protein levels. *Biol. Chem. Hoppe-Seyler.* 373, 477-482.

Strukelj, B., Pungercar, J., Ritonja, A., Krizaj, I., Gubensek, F., Kregar, I. and Turk, V. (1990). Nucleotide and deduced amino acid sequence of an aspartic proteinase inhibitor homologue from potato tubers. *Nucleic Acids Res.* 18, 4605.

Suh, S., Peterson, J., Stiekema, W. and Hannapel, D. (1990). Purification and characterization of the 22-kilodalton potato tuber proteins. *Plant Physiol.* 94, 40-45.

Suh, S., Stiekma, W. and Hannapel, D. (1991). Proteinase-inhibitor activity and wound inducible gene expression of the 22-kDa potato-tuber proteins. *Planta* 184, 423-430.

Szalanski, A., Roehrdanz, R. and Taylor, D. (2000). Genetic relationship among *Diabrotica* species (Coleoptera: *Chrysomelidae*) based on rDNA and mtDNA sequences. *Fla. Entomol.* 83, 262-267.

Tabashnik, B., Cushing, N., Finson, N. and Johnson, M. (1990). Field development of resistance to *Bacillus thuringiensis* in diamond back moth (Lepidoptera: *Plutellidae*). *J. Econ. Entomol.* 83, 1671-1676.

Tabashnik, B., Finson, N., Groeters, F., Moar, W, Johnson, M., Luo, K. and Adang, M. (1994). Reversal of resistance to *Bacillus thuringiensis* in *Plutella xylostella*. *Proc. Natl. Acad. Sci. USA*. 91, 4120-4124.

Tabashnik, B., Liu, Y., Finson, N., Masson, L. and Heckel, D. (1997). One gene in diamondback moth confers resistance to four *Bacillus thuringiensis* toxins. *Proc. Nat. Acad. Sci. USA.* 94. 1640-44.

Tang, J., James, M., Hsu, I., Jenkins, J. and Blundell, T. (1978). Structural evidence for gene duplication in the evolution of the acid proteases. *Nature* 271, 618- 621.

Tang, A. and Tu, C. (1994). Biochemical characterization of *Drosophila* glutathione *S*-transferases D1 and D21. *J. Biol. Chem.* 269, 27876-27884.

Tang, J., Sepulveda, P., Marciniszyn, J., Chen, K., Huang, W., Tao, N., Liu, D. and Lanier, J. (1973). Amino acid sequence of porcine pepsin. *Proc. Natl. Acad. Sci. USA.* 70, 3437-3439.

Tang, J., Shelton, A., Van Rie, J., De Roeck, S., Moar, W., Roush, R. and Pefereon, M. (1996). Toxicity of *Bacillus thuringiensis* spore and crystal to resistant diamond back moth *(Plutella xylostella). Appl. Environ. Microbiol.* 62, 564-569.

Tao, K., Stearns, N., Dong, J., Wu, Q. and Sahagian, G. (1994). The proregion of cathepsin L is required for proper folding, stability and ER exit. *Arch. Biochem. Biophys.* 311, 19-27.

Tatnell, P., Lees, W. and Kay, J. (1997). Cloning, expression and characterisation of murine procathepsin E. *FEBS Lett.* 408, 62-66.

Tatusova, T.A., and Madden, T.L. (1999). Blast 2 sequences - a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol. Lett.* 174, 247-250.

Tcherepanova, I., Bhattacharyya, L., Rubin, S. and Freedman, J. (2000). Aspartic proteases from the nematode *Caenorhabditis elegans. J. Biol. Chem.* 275, 26359-26369.

Tellam, R. (1996). The peritrophic membrane. (eds. Lehane, M., Billingsley, P..) In; *Biology of the Insect Midgut*. Chapman and Hall, London, 86-114.

Terra, W. (1990). Evolution of digestive systems of insects. *Annu. Rev. Entomol.* 35, 181-200.

Terra, W. and Cristofoletti, P. (1996). Midgut proteinases in three divergent species of Coleoptera. *Comp. Biochem. Phys.* 113B, 725-730.

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

Toung, Y., Hsieh., T. and Tu, C. (1993). The glutathione *S*-transferases D Genes: A divergently organized, intronless gene family in *Drosophila melanogaster. J. Biol. Chem.* 268, 9737-9746.

Troen, B., Gal, S. and Gottesman, M. (1987). Sequence and expression of the cDNA for MEP (major excreted protein), a transformation-regulated secreted cathepsin. *Biochem. J.* 246, 731-735.

Turk, B., Turk, D. and Turk, V. (2000). Lysosomal cysteine proteinases: more than just scavengers. *BBA*. 1477, 98-111.

Utton, T., (2002). Why the world is losing the battle of the bugs. *Daily Mail* Tuesday 30th July.

Valaitas, A., Lee, M. and Rajamohan, F. (1995). Brush border membrane aminopeptidase-N in the midgut of the gypsy moth serves as the receptor for the Cry IA(c) δ-endotoxin of *Bacillus thuringiensis*. *Insect Biochem Molec Biol.* 25, 1143-1151.

Valaitas, A., Augustin, S. and Clancy, K. (1999). Purification and characterisation of the western spruce budworm larval midgut proteinases and comparison of gut activities of laboratory-reared and field collected insects. *Insect Biochem. Molec. Biol.* 29, 405-415.

Valeuva, T, Revina, T. and Mosolov, V. (1997). Potato tuber protein proteinase inhibitors belonging to the Kunitz soybean inhibitor family. *Biochemistry* (Mosc). 62, 1367-1374.

Veerapandian, B., Cooper, J., Sali, A. and Blundell, T. (1990). X-ray analyses of aspartic proteinases. III. Three dimensional structure of endiothiapepsin complexed with a transition-state isostere inhibitor of renin at 1.6 Å resolution. *J. Mol. Biol.* 216, 1017-1029.

Velasco, G., Ferrando, A., Puente, X., Sanchez, L. and Lopez-Otin, C. (1994). Human Cathepsin O. *J. Biol. Chem.* 269, 27136-27142.

Vick, B, and Zimmerman, D. (1987). Oxidative systems for modification of fatty acids: the lipoxygenase pathway. (eds. Stumpf, P. and Conn, E.) In; *The biochemistry of plants: Lipids*. New York: Academic Press, 53-90.

Waldron, C., Wegrich, L., Owens Merlo, P. and Walsh, T. (1993). Characterisation of a genomic sequence coding for a potato multicystatin, an eight domain cysteine proteinase inhibitor. *Plant Molec. Biol.* 23, 801-812.

Walsh, T. and Strickland, J. (1993). Proteolysis of the 85-kD crystalline cysteine proteinase inhibitor from potato releases functional cystatin domains. *Plant Physiol.* 103, 1227-1234.

Walsh, T. and Twitchell, W. (1991). Two Kunitz-type proteinase inhibitors from potato tubers. *Plant Physiol.* 97, 15-18.

Wang , X., Ding, X., Gopalakrishnan, B., Morgan, T., Johnson, L., White, F., Muthukrishnan, S. and Kramer, K. (1996). Characterisation of a 46 kDa insect chitinase from transgenic tobacco. *Insect Biochem. Molec. Biol.* 26, 1055-1064.

Watanbe, S., Kokuho, T., Takahashi, H., Takahashi, M., Kubota, T. and Inumara, S. (2002). Sialylation of *N*-glycans on the recombinant proteins expressed by a baculovirus-insect cell system under beta-N-acetylglucosaminidase inhibition. *J. Biol. Chem.* 277, 5090-5093.

Wex, T., Levy, B., Smeekens, S., Ansorge, S., Desnick, R. and Bromme, D. (1998). Genomic structure, chromosomal localization, and expression of human cathepsin W. *Biochem. Biophys. Res. Commun.* 248, 255-261.

Wiederlanders, B., Bromme, D., Kirschke, H., Figura, K., Schmidt, B. and Peters C. (1992). Phylogenetic conservation of cysteine proteinases. *J. Biol. Chem.* 267, 13708-13713.

Wilhite S., Thomas, C., Brzin, J., Smigocki, A. (2000). Inhibition of cysteine and aspartyl proteinases in the alfalfa weevil midgut with biochemical and plant derived proteinase inhibitors. *Insect Biochem. Molec. Biol.* 30, 1181-1188.

Wolfson, J. and Murdock, L. (1990). Diversity in digestive proteinase activity among insects. *J. Chem. Ecol.* 16, 1089-1102.

Womack, J., Randall, J. and Kemp, J. (2000). Identification of a signal peptide for oryzacystatin-1. *Planta*. 210, 844-847.

Yamamoto, Y., Watabe, S., Kageyama, T. and Takahashi, S. (1999). Proregion of Bombyx mori cysteine proteinase functions as an intramolecular chaperone to promote proper folding of the mature enzyme. *Arch. Insect. Biochem. Physiol.* 42, 167-178.

Yasuda, Y., Ikeda, S., Sakai, H., Tsukuba, T., Okamoto, K., Nishishita, K., Akamine, A., Kato, Y. and Yamamoto, K. (1999). Role of *N*-glycosylation in cathepsin E: A comparative study of cathepsin E with distinct *N*-linked oligosaccharides and its nonglycosylated mutant. *Eur. J. Biochem.* 266, 383-391.