Molecular approaches to increasing resistance of wheat (Triticum aestivum L.) towards two insect pests; Cereal aphid (Sitobion avenae F.) and Wheat bulb fly (Delia coarctata Fallen).

PRASHANT SHIVASHARAN PYATI

Molecular approaches to increasing resistance of wheat (*Triticum aestivum* L.) towards two insect pests; Cereal aphid (*Sitobion avenae* F.) and Wheat bulb fly (*Delia coarctata* Fallen).

A thesis submitted by Prashant Shivasharan Pyati in accordance with requirements of Durham University for the degree of Doctor of Philosophy.

School of Biological and Biomedical Sciences

Durham University

March 2010

Abstract

Cereal aphid (*Sitobion avenae*) and wheat bulb fly (*Delia coarctata*) are serious pests of wheat in the UK. At the present, chemical pesticides are used to control these insects, but they are limited in effectiveness, and have undersirable ecological impacts. There is a need to improve wheat genetically to be resistant to such inset pests. The objectives of this work were to investigate digestive biochemistry in the selected insect pests of wheat, and to determine effects of potential endogenous resistance factors in wheat on digestion, nutrition and other insect metabolic processes. The aim was to develop new strategies for crop protection.

Digestive biochemistry in *S. avenae* and *D. coarctata* was studied to characterise gut proteases and their inhibition by host plant proteinase inhibitors (PIs). Investigation of proteolytic digestion in *S. avenae* gut showed that in spite of being a phloem-feeding insect, cereal aphid could digest ingested protein, using cysteine proteases. *D. coarctata* larvae contained mainly serine protease activity. A serine protease (DcSP) and a cysteine protease (DcCathL) from *D. coarctata* gut tissue were expressed as recombinant proteins. Only DcCathL was recovered in active form. DcCathL was insecticidal to *Mamestra brassicae* when injected into hemolymph, causing systemic and extensive melanisation. DcCathL selectively degraded recombinant serpins from *M. brassicae* in *in vitro* assays, and is suggested to interfere with regulation of the proteolytic cascade leading to phenoloxidase activation and melanin production *in vivo*. DcCathL has potential as a biopesticide if it could be made effective when orally delivered. A cationic amino acid transporter from *D. coarctata* gut (DcCAAT) was also cloned as a target for RNA interference.

Potential resistance factors in wheat were characterised by expression as recombinant proteins. Two PIs from wheat (subtilisin/chymotrypsin inhibitor; WSCI, and cysteine proteinase inhibitor; WCPI) were expressed in the yeast *Pichia pastoris*, and purified. WSCI inhibited gut protease activity of both insects in *in vitro* and *in vivo* assays, whereas WCPI only inhibited *S. avenae* gut extract activity. On feeding, WSCI was antimetabolic to both insects, affecting both survival and growth, whereas WCPI was antimetabolic to *S. avenae* only. Wheat Hessian fly responsive (*Hfr*) genes are up-regulated in response to herbivory by Hessian fly (*Mayetiola destructor*). The protein product Hfr-3 was expressed and purified, and showed antimetabolic effects on survival and growth of both *S. avenae* and *D. coarctata*. Both accumulated and induced defence proteins, like WSCI, WCPI and Hfr-3, have the potential to act as endogenous resistance factors in wheat towards a range of insect pests. Developing a wheat variety constitutively expressing these defence proteins by using traditional breeding methods and/or modern biotechnological tools is discussed.

TABLE OF CONTENTS

Declaration and Statement of Copyright

Declaration

No material contained herein has been previously submitted for any other degree. Except where acknowledged, all material is the work of the author.

Statement of Copyright

Date.....

Signed.....

Acknowledgements

I am very grateful for an excellent supervision, constructive criticism and all the support provided by Dr. John Gatehouse (School of Biological and Biomedical Sciences, Durham University) throughout my PhD. Funding by Graduate School, Durham University is also gratefully acknowledged.

Heartfelt thanks to the past and present members of lab 1 who created a friendly working environment and offered endless advice, particularly Dr. Elaine Fitches, Dr. Daniel Price and Dr. Nanasaheb Chougule. My appreciation is also extended to other members of lab 1; Emma, Gareth, Catherine and Dominic who have become good friends over the period for their support and suggestions all the time.

Thanks to Dr. Howard Bell (FERA, york) for providing insect tissues, carrying out feeding bioassays and insect injections.

I would also like to thank other friends from the school namely Vipul, Siddharth, Usha, Natalie, Claire and Sergei for those endless cups of tea, discussions and fun times.

Finally, and by no means least, I am indebted to my family members back in India for their moral support and encouragement throughout the education period. A special thanks to wife Shweta and daughter Anwesha for being very supportive all the time.

Prashant Pyati

List of figures and tables

Chapter 1:

Chapter 2:

Chapter 3:

Chapter 6:

Abbreviations:

Nucleic acid abbreviations:

A: Adenine

T: Thymine

G: Guanine

C: Cytosine

Degenerate Nucleic acid abbreviations:

B: C/G/T

D: A/G/T

H: A/C/T

K: G/T

M: A/C

N: A/T/G/C

R: A/G

S: G/C

V: A/C/G

W: A/T

Y: C/T

Amino acid abbreviations:

Name	**Single letter code**	**Three letter code**
Alanine	A	Ala
Arginine	R	Arg
Asparagine	N	Asn
Aspartic acid	D	Asp
Cystiene	C	Cys
Glutamic acid	E	Glu
Glutamine	Q	Gln
Glycine	G	Gly
Histidine	H	His

Name	Single letter code	Three letter code
Isoleucine	I	Ile
Leucine	L	Leu
Lysine	K	Lys
Methionine	M	Met
Phenylalanine	F	Phe
Proline	P	Pro
Serine	S	Ser
Threonine	T	Thr
Tryptophan	W	Trp
Tyrosine	Y	Tyr
Valine	V	Val

Other abbreviations:

AOX1: Alcohol oxidase 1

BSA: Bovine Serum Albumin

ECL: Enhanced Chemiluminescence

ORF: Open reading frame

PBS: Phosphate buffered saline

PCR: Polymerase chain reaction

RT-PCR: Reverse transcription polymerase chain reaction

RACE: Rapid Amplification of cDNA ends

GSP: Gene specific primer

UTR: Untranslated region

CDS: Coding sequences

MCS: Multiple cloning site

SDS-PAGE: Sodium dodecyl sulphate polyacrylamide gel electrophoresis

TEMED: Tetramethylethylenediamine

WGA: Wheat germ agglutinin

TAE: Tris acetate EDTA buffer

CBB: Coomassie brilliant blue stain

dNTPs: deoxyribonucleotide triphosphate

dATP: deoxyadenosine triphosphate

dTTP: deoxythymine triphosphate

dGTP: deoxyguanine triphosphate

dCTP: deoxycytosine triphosphate

DTT: Dithiothreitol

M-MLV RT: Moloney Murine Leukemia virus reverse transcriptase

TRX: Thioredoxin

rpm: revolutions per minute

PTM: *Pichia* trace metal

BCA: Bicinchonic acid

OD: Optical density

HRP: Horseradish Peroxidase

IgG: Immunoglobulin G

FPLC: Fast Protein Liquid Chromatography

MWCO: Molecular weight cut off

GSSG: Oxidized Glutathione

GSH: reduced Glutathione

bp: base pairs

kD / kDa: KiloDalton

RNAi: RNA interference

dsRNA: double stranded RNA

List of Publications / Sequence submissions to Global database:

Publication:

Pyati P.S., Bell H. A., Fitches E., Price D.R.G., Gatehouse A.M.R. and Gatehouse J.A., (2009). Cathepsin L-like cysteine proteinase (DcCathL) from *Delia coarctata* (wheat bulb fly): Basis of insecticidal activity. Insect Biochemistry and Molecular Biology 39: 535-546.

Sequence Submissions:

FJ 763761 *Delia coarctata* trypsin-like serine proteinase mRNA, complete cds.(DcSP)

FJ 763762 *Delia coarctata* cathepsin L-like cysteine proteinase mRNA, complete cds. (DcCathL)

FJ 763763 *Mamestra brassicae* serine proteinase inhibitor-1A mRNA, complete cds.(MbSpn1A)

FJ 763764 *Mamestra brassicae* serine proteinase inhibitor-1B/C mRNA, complete cds.(MbSpn1B/C)

NCBI Accession numbers for mRNA / Proteins

AY 148483 *Mamestra configurata* serine protease inhibitor serpin 1a mRNA,complete cds. coding for protein **AAN71632**

AY 148484 *Mamestra configurata* serine protease inhibitor serpin 1b mRNA,complete cds. coding for protein **AAN71633**

AY 148485 *Mamestra configurata* serine protease inhibitor serpin 1c mRNA,complete cds. coding for protein **AAN71634**

CAC 87119 *Nilaparvata lugens* trypsin-like protease

CAC 87118 *Nilaparvata lugens* cathepsin B-like protease

P10968 Wheat germ agglutinin Isolectin A (*Triticum aestivum*)

P10969 Wheat germ agglutinin Isolectin B (*Triticum aestivum*)

P02876 Wheat germ agglutinin Isolectin D (*Triticum aestivum*)

AAA34257 Wheat germ agglutinin [*Triticum turgidum* subsp. durum]

CG9413 PA *Drosophila melanogaster* Transmembrane amino acid transporter protein (Isoform A).

CG9413 PB *Drosophila melanogaster* Transmembrane amino acid transporter protein (Isoform B)

AF483596 *Triticum aestivum* hessian fly response gene 1 protein (Hfr-1) mRNA,complete cds

AY587018 *Triticum aestivum* pore-forming toxin-like protein Hfr-2 mRNA, cds.

DQ 462308 *Triticum aestivum* clone UPW1 HFR-3 (Hfr-3) mRNA, complete cds.

AAY45744 WSCI proteinase inhibitor precursor [*Triticum aestivum*].

BAB18766 Cysteine proteinase inhibitor [*Triticum aestivum*].

CG6692 *Drosophila melanogaster* Cysteine proteinase-1

ABV03130 *Sitobion avenae* Cathepsin-B2744

ABV03144 *Sitobion avenae* Cathepsin-B16D

ABV 03134 *Acyrthosiphon pisum* Cathepsin B-84

ABV 03127 *Acyrthosiphon pisum* Cathepsin B-2744

ABV 03124 *Acyrthosiphon pisum* Cathepsin B-16A

Colour plate images courtesy:

Colour plate I

Agroites lineatus (larva)
http://www.maine.gov/agriculture/pi/images/pests/veg/EWW_agriotes-sp01.jpg
Agroites lineatus (adult)
http://molbiol.ru/forums/uploads/a001/b003/post-184-1150328909.jpg
Cephus cinctus
http://entomology.unl.edu/images/smgrains/sawflies/wstem_sawfly2.jpg
Eurygaster integriceps
http://www.agroatlas.ru/content/pests/Eurygaster_integriceps/Eurygaster_integriceps.jpg
Faronta diffusa
http://upload.wikimedia.org/wikipedia/commons/0/02/Faronta_diffusa_larva.jpg
Diuraphis noxia
http://www.sel.barc.usda.gov/aphid/dnoxia.jpg

Colour plate II

Mayetiola destructor
http://www.museum.state.il.us/exhibits/agriculture/images/animals/hessianfly_UKy.jpg
Meromyza americana
http://www.cedarcreek.umn.edu/insects/newslides/029094050000bpo.jpg
Oulema melanopa
http://farm4.static.flickr.com/3077/2901926327_dcef5f7682.jpg
Clorops pumilionis
http://www.bugsandweeds.co.uk/flies/chlorops%20pumilionis.jpg
Sitobion avenae
http://farm3.static.flickr.com/2185/2400138737_7386a43afe.jpg
Sitobion avenae infested wheat
http://www.dewarcropprotection.co.uk/images/sitobion_avenae_large.jpg

Colour plate III

Delia coarctata (adult)
http://www.inra.fr/hyppz/IMAGES/7031562.jpg
Delia coarctata: Eggs, Larva, Pupae, deadheart and devastated field
(Courtesy: Dr. Howard Bell, FERA, York)

Colour plate IV

Paul Sidney, Photograph facility, School of Biological and Biomedical Sciences, Durham University.

Chapter 1

Introduction

1.1 Wheat

Among the food crops, wheat (*Triticum spp.*) is the third important cereal after Maize (*Zea mays*) and Rice (*Oryza sativa*). It is the major cereal crop in many parts of world (Kent and Evers, 1994). The projected world wheat production for 2008-2009 was 656 million metric tonnes (Elton Robinson, 2008). Beginnings of wheat can be traced to a tribe of wild grasses called 'Triticeae' that includes wheat, barley, rye and some other grasses. The seeds of these had a flavour that was pleasing to primitive people and the plants were domesticated for cultivation as food source. The fertile crescent at the core of Western Asia and Northern Africa is the origin of this tribe. Hybridization of wheat occurs naturally and through human assistance by manual pollination. Wheat breeding has been in practice since early agriculture. Breeding was the first step taken towards improvement of wheat; breeding has been practiced since people first selected the biggest seeds that were easiest to thresh and stored them for planting. Favorable characteristics such as disease resistance, large kernels, short straw, and cold hardiness were selected in two, parent wheat plants, which were then crossed to form a hybrid.

Two varieties of wheat, Einkorn and Emmer have been in existence for about 75,000 years and are the earliest ancestors of today's wheat. Einkorn (*T. monococcum*) is considered to be the first wheat gathered and cultivated. All wheat varieties are identified by the presence of chromosome sets (ploidy), which were fourteen in einkorn (2n=14). Einkorn and wild grass crosses produced a tetraploid wheat Emmer (*T. dicoccoides*) with doubled chromosome content (2n=28). Emmer was once the most widely cultivated wheat with twenty-eight-chromosomes. Durum, the hardy wheat (*T. durum*), is a hexaploid (2n=42) and is derived from Emmer wheat crosses. Wheat varieties containing forty-two chromosomes are the most recently evolved and are generally used for cultivation. They are products of crosses between twenty-eight-chromosome wheat and wild fourteen-chromosome wheat or grasses. Modern bread

wheat varieties contain forty-two chromosomes and they came from crosses between emmer and goat grass. The goat grass is the source of glutenin genes that give bread dough the ability to form gluten. Ninety-five percent of wheat grown today is a true breeding hybrid and is hexaploid.

The thousands of wheat varieties grown today came from different crosses involving 14 different species (Mangelsdorf, 1953). These are: wheat containing 14 chromosomes: *Triticum aegilopoides* (wild einkorn), *T. monococcum* (einkorn), 28 chromosomes: *Tritcum dicoccoides* (wild emmer), *T. dicoccum* (emmer), *T. durum*, *T. persicum* (Persian wheat), *T. turgidum* (rivet wheat), *T. polonicum* (Polish wheat), *T. timopheevi* and 42 chromosomes: *Triticum aestivum* (common wheat), *T. sphaerococcum* (shot wheat), *T. compactum* (club wheat) (these three are true bread wheat; account for about 90 percent of all wheat grown today), *T. spelta*, *T. macha* (Macha wheat). Modern bread wheat is thus an allohexaploid containing three distinct but genetically related (homeologous) copies each of the three originally independent haploid genomes (Gill and Gill, 1994). Many varieties of wheat with desirable traits related to yield or resistance are now available worldwide and there are continuous efforts being taken for developing wheat with more desirable traits.

Wheat is one of the most important staple foods, abundant source of energy and proteins for the world population. Two species *T. aestivum* (vulgare) and *T. durum* (hard wheat) are commercially very important. Production of wheat has nearly tripled since 1955 and has grown an average of 2.3 percent annually since 1951. Wheat is grown as both a winter and spring cereal, owing to the number of species and varieties and their adaptability. It is grown in many countries around the world. The life cycle of wheat is of ninety days and it requires dry, sunny weather during the grain filling and ripening period. Cultivation of wheat requires a rainfall between 254 and 762 mm annually and soil preferred ranges from sandy loam to clay. With the availability of different varieties, wheat can be grown in every temperate climate in the world and world wheat production is perennial. Some of the countries that produce wheat on large scale are United States, China and Russia. Countries like India, Pakistan, the EU, Canada, Argentina and Australia also make a significant contribution to global wheat production. Many developing countries are becoming self-sufficient in their wheat needs.

The origin of wheat production today in the UK can be traced back over 10,000 years to two important wheat varieties, the Einkorn and Emmer that grew wild in the Middle East. These varieties were domesticated and were introduced in UK about 6,000 years ago. The wheat varieties that are grown today have repeatedly undergone selection process for higher yields and better disease resistance. Wheat is generally sown in autumn in the UK, due to the temperate climate conditions that allows plants to grow through the winter and produce a higher yield than a spring sown alternative. As of today there are about 2,000,000 hectares of land under wheat cultivation in UK yielding 15 million tones of wheat, worth £1.2 billion. 25% of the wheat is exported to the countries around the world and about 40% is used as animal feed. UK produced wheat is used as major favorite source of food in the form of bread, and starch extracted from wheat is used in confectionaries, soft drinks, alcoholic drinks and convenience foods. Since 2009, wheat is now anticipated to gain an additional market as a feedstock for use in production of bioethanol.

(The information presented in this part of introduction was obtained from http://www.answers.com/topic/the-natural-history-of-wheat, The natural History of Wheat, Food and Culture encyclopedia, World Wheat Facts and Trends, 1998-99, hhttp://www.ukagriculture.com/crops/wheat and http://en.wikipedia.org/wiki/Wheat)

1.2 Impact of insect herbivores on wheat:

Environment poses a major impact on the growth, development and yield of crop plants. Two different types of stresses, biotic and abiotic constantly threaten agricultural industry. Abiotic stresses include high winds, extreme temperatures, drought, edaphic factors like salinity, pH etc. Biotic stress involves infestation by pathogens like bacteria, viruses, fungi, parasites and insect pests. The pest associated losses are estimated at 14% of the total world agricultural production: 52% in wheat, 83% in rice, 59% in maize, 74% in potato, 58% in soybean and 84% in cotton (Oerke *et al.,* 1994, Sharma *et al.,* 2000). Among all the biotic stresses insects are an important factor for the crop related losses. They are major threat to agricultural industry by decreasing the overall production not only by herbivory but also by spreading many different types of diseases. Herbivorous insect pests feed on all parts of the plant such as leaves, stem, inflorescence, young grains, etc. as well as specialist feeders like sap-sucking insects. Wheat has many insect pests worldwide, with a few

of major importance, causing severe damage over large geographical areas; most species are only occasional pests and/or are not geographically widespread. Some of the pests belonging to different orders are discussed here.

The Hemipteran pests of wheat; aphids, include *Sitobion avenae* (Cereal aphid / British grain aphid; Colour plate-II) and *Diuraphis noxia* (Russian wheat aphid; colour plate-I). Aphid infestation in sufficient numbers can cause yellowing and premature death of leaves. They exude drops of sugary liquid known as honeydew, which may cause scorch marks on the foliage and tend to encourage the development of sooty molds. They also act as carriers for several viruses that cause diseases in plants. Sunn pest, (*Eurygaster integriceps*; Colour plate-I) is also an important pest of wheat especially in North Africa and West Asia.

The Coleopteran pest of wheat, Cereal Leaf Beetle **(***Oulema melanopa*; colour plate-II), adults and larvae feed on wheat leaves by chewing out long strips of tissue between the veins of leaves. Adults eat right through the leaf, but larvae eat the upper leaf surface leaving a thin membrane, giving a windowpane effect. Larvae attack the flag leaf, beginning at the tip and moving down the leaf. When damage is extensive the leaves turn whitish and the plant takes on the appearance of frost damage. Young plants may be killed or the yield may be seriously reduced. Yield of wheat can be reduced upto 55% in spring wheat and 23% in winter-sown wheat. Soil inhabiting Wireworm (*Agriotes lineatus*; colour plate-I) larvae may attack wheat as soon as the crop is seeded, eating the endosperm of the kernels and leaving only the seed coat. A common sign of wireworm attack is the wilting and/or dying of a number of adjacent plants, either in a row or patches.

There are many dipteran pests of wheat and Hessian Fly **(***Mayetiola destructor*; colour plate II) is the most serious of them. Severe infestations of Hessian flies result in stunting of the plants, thin stands, lodging, and reduced yield. Injury is caused entirely by the larvae, which suck juices from plant tissues. If infestation occurs during jointing, infested stems often break prior to maturity. When wheat stem maggot **(***Meromyze americana*; colour plate-II) attacks young tillers of wheat in the fall or early spring, the tillers usually die; infested plants show the "white head" condition typically produced by stem-boring insects. Damage by wheat stem sawfly (*Ciphus*

cinctus: colour plate-I) includes premature yellowing of the head and shriveling of the grain. The larvae girdle the stem and, later in the crop cycle, lodging is commonly observed. Gout fly (*Chlorops pumilionis*; colour plate-II) causes late-developing of wheat, small shoots and tillers become stunted, gouty and fail to produce ears. Advanced crops have larvae feeding down one side of the ear to the uppermost node, checking growth and damaging grain, causing upto 50% yield loss. Wheat bulb fly (*Delia coarctata*; colour plate-III) larvae burrow themselves in stem and feed on the youngest leaf roll causing a dead-heart, the plant wilts immediately and shows dieback symptoms. Infestation rate is so high that the effect is clearly visible at field level as dead patches.

Among Lepidoptera, wheat head armyworm (*Faronta diffusa*; colour plate-I) is a minor pest of wheat and only in some parts of the world, occasionally causing noticeable injury.

However, two major pests of wheat in the UK are Wheat bulb fly (*D. coarctata;* Diptera) and Cereal aphid (*S. avenae*; Hemiptera) and need a special attention from the crop protection strategy. There is not enough recent documentation on *D. coarctata* as it was before 1960s. It still remains one of the important pests of wheat in UK and other European countries. Flies are similar to small houseflies that lay eggs in exposed soil in July and August. Eggs need a cold winter to break diapause and hatch from January to March, depending on soil temperatures. Hatched larvae are legless and slowly migrate to wheat plants through soil and burrow in central shoot of wheat. This causes the dead heart of wheat, which is detected only if closely observed as outer leaves look healthy and green. Larvae grow very quickly in March-April (up to 12mm) and move to adjacent leaves causing further damage. Larvae then pupate in soil in early May and there is one generation per year. Estimation of yield loss due to *D. coarctata* is approx. 0.7 tonnes per hectare when 20% of the field plants are attacked. Due to its long and complicated life cycle it is difficult to devise a successful control strategy for this insect. The insecticides could be sprayed as suggested by many insecticide companies at egg hatch stage or deadheart stage depending on level of infestation. Spraying of insecticides such as Chlorpyrifos (crystalline organophosphate that inhibits acetylcholinesterase e.g. Dursban, Spannit) or Dimethoate could show protection, however sprayings are required to be at a specific

stage of infestation to be effective. Some insecticides used against *D. coarctata* like treatment of seeds with Tefluthrin (Pyrethroid ester group insecticide) could also be effective (http://www.sac.ac.uk/mainrep/pdfs/tn552wheatbulbflypests.pdf; http://www.dowagro.com/uk/cereal/wheatbulb.htm). Though the companies that sell these insecticide claim for a good protection of wheat against *D. coarctata*, the are also suggested to minimize the attack of *D. coarctata* by avoiding leaving bare soil in July-August and deep drilling, use of rotational choices, seed treatment, etc. (http://www.dowagro.com/uk/cereal/wheatbulb.htm).

Cereal aphid (*S. avenae*) is a successful pest of wheat due to its specialist feeding behaviour, they feed on phloem sap rich is sucrose and free amino acids. It is a major pest of wheat and other cereal crops in the spring. Adult aphids are 2 to 2.8 mm long and there is a colour variation from yellow to green, red and purple. Nymphs feed on young plant parts including filling grains. Heavy infestations can cause a reduction of the number of grains per ears and thus a noticeable reduction of the yield. Secretion of honeydew is another problem, which provides a suitable growth conditions for Sooty moulds. Another problem associated with aphids is that they act as vectors for many viruses such as barley yellow dwarf virus (BYDV).

(The information presented in this part of introduction was obtained from http://wheat.pw.usda.gov/ggpages/wheatpests.html,http://www.sac.ac.uk/mainrep/pdfs/tn552wheatbulbflypests.pdf, http://greengenes.cit.cornell.edu/wpest.html and http://www.dowagro.com/uk/cereal/wheatbulb.htm

Colour Plate-I

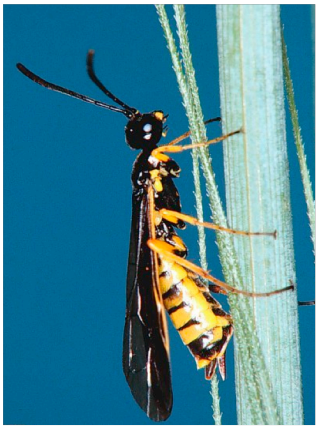
Insect Pests of Wheat (various scales)

Wireworm
(*Agriotes lineatus,* Coleoptera)

Larvae

Adult

Wheat stem sawfly
(*Cephus cinctus*, Diptera)



Sunn pest
(*Eurygaster integriceps,* Hemiptera)

Wheat head armyworm
(*Faronta diffusa*, Lepidoptera)

Russian wheat aphid
(*Diuraphis noxia,* Hemiptera)

Colour Plate-II

Insect Pests of Wheat (various scales)

Insect pests of wheat

Hessian fly
(*Mayetiola destructor, Diptera)*

Wheat stem maggot
(*Meromyza americana*, Diptera)

Cereal leaf beetle
(*Oulema melanopa*, Coleoptera)

Gout fly
(*Chlorops pumilionis*, Diptera)

Cereal aphid / Grain aphid
(*Sitobion avenae,* Hemiptera)

Colour Plate-III

Insect Pests of Wheat (various scales)

Wheat bulb fly (*Delia coarctata* Fallen) (Diptera)

Adult

Eggs

Larva burrowed in stem

Pupae

Devastated wheat field

Deadheart caused by WBF

Increase in global population has been estimated to be 9-10 billion by 2050. (http://www.un.org/esa/population/). Availability of limited cultivable land along with many other agricultural related problems have revealed the possibility of scarcity of food in future. Increasing the food crop related production is an immediate priority as a step to face this problem. One of the practical means of achieving the greater yields in the available cultivable land is to minimise the insect pest-associated losses. Use of chemical insecticides for control of pests is widely and routinely used method for control of insect pests. However use of insecticides is needed be minimised to avoid development of insecticide resistance in insects and environmental contamination (Clark and Yamaguchi, 2002). Integrated Pest Management (IPM) is an effective and environmentally sensitive approach to pest management. It has shown many alternative approaches of controlling insect pests by reducing the use of insecticides and pesticides. This would also take care of emerging problems such as soil erosion and other related consequences.

Improvement of crop plants for pest-resistance using traditional breeding alone or in combination with modern biotechnological tools has become an important approach to overcome insecticide related problems, which could make agriculture business sustainable and cost effective. Traditional breeding methods include transfer of economically important traits by crossing related methods. Use of biotechnological tools leads to development of transgenic crops with modified genome necessary for better yield and / or pest resistance. While devising crop protection strategy against insect pests, it is necessary to understand the interaction between crop plant and insect pests. Plant-insect interaction studies involve understanding the type of stress laid upon by insect and improving crop plants to confer tolerance or resistance to overcome particular type of stress. Study of endogenous defence mechanism in crop plants and / or investigation of potential targets in insect pests are very important. These endogenous defence mechanisms could be improved by using modern biotechnological tools to make the agriculture business sustainable and cost effective. This particular field of plant–insect interactions has been become an important area of research from agricultural point of few since last few decades. Some of the endogenous resistance factors from plants, in general and wheat along with insect gut proteases as potential targets in insects are discussed here.

1.3 Endogenous defence mechanism of plant against insect herbivores

Insects having the highest number of species in animal kingdom, have co-existed with plants for as long as 350 million years (Gatehouse, 2002). Insects have been notoriously problematic to agricultural industry due to the "evolutionary arms race". Co-evolutionary theory by Ehlrich and Raven (1964) suggests that every plant species is preyed on by at least one insect species. Increase in species diversity in both feeding herbivores and host plants is greatly influenced by insects' feeding on plants (Harborne, 1988). Some of the interactions between plants and insects are mutually beneficial like pollination, however the most important interaction involves insect predation of plants (Gatehouse, 2002). Over the period of their co-evolution plants have developed some strategies themselves to resist feeding insects in the form of accumulation of high levels of toxic chemicals that function in defence mechanism.

The overall mechanism of plant defence against insect herbivores has been reviewed by Gatehouse (2002). Accordingly, such defence is divided basically into two types, 'static' or constitutive and 'active' or induced defences. Static type of plants defence involves synthesis and accumulation of defence chemicals by plant during its normal growth and development (in absence of infestation by herbivore). These accumulated compounds enable plants to deter or kill the insects when they are attacked. Chemicals such as secondary metabolites are highly accumulated in plants and play an important role in defence mechanism (Bennett & Wallsgrove, 1994). The toxicity of these compounds results in lethal effects on herbivores such as membrane disruption, inhibition of transport or signal transduction, inhibition of metabolism and even disruption of hormonal control of developmental processes. Plants are also able to defend themselves against insects by using 'active' defence mechanism. In this, the defence chemicals are not accumulated and stored during the plants' growth and development but are synthesized specially in response to herbivory by insects (Harborne, 1988). These compounds involve different types of proteins which themselves can act as toxins or can disrupt insect metabolism (Ryan, 1978). Attack by the herbivore induces the 'active' mechanism systemically, so that the defence compounds are produced throughout the plant and not only restricted to the site of herbivory. Production of proteinase inhibitors in potato (*Solanum tuberosum*) or tomato (*Lycopersicon esculentum*) leaves in response to feeding by a lepidopteran pest (*Manduca sexta*) is a classic example of the active type of plant defence (Ryan,

1978). However, the end products of both these mechanisms, the defensive compounds themselves are often same in constitutive and induced responses. Proteinase inhibitors (PIs) produced by potato in response to herbivory are also accumulated in the tubers constitutively (Garcia-Olmedo *et al.,* 1987). Similarly tobacco species (*Nicotiana tabacum; N. attenuata*) accumulate PIs in its tissues before insect attack, although herbivory induces the synthesis of increased levels of PIs (Van Damme *et al.,*2001). Two defence proteins that are either involved in 'static' or 'active' or both type of defence are PIs and sugar-binding lectins. These two have been anticipated as important endogenous defence proteins and are extensively studied from the crop protection point of view.

1.3.1 Plant Secondary Metabolites

Plant secondary metabolites (PSMs) are organic compounds, which are not required for normal plant growth and development. They are often produced as by-products during synthesis of primary metabolic products (Herbert, 1989). The structures for about 50,000 plant secondary metabolites have been elucidated which is believed to be only a small fraction of all PSMs existing in nature (De Luca & St Pierre, 2000). It is now established that PSMs provide plants with their specific odours, tastes and colours and in addition they can be involved in storage of nitrogen, protection from UV, attraction of pollinating organisms, as well as serving as tools for transport (Wink, 1999), protection against osmotic and other environmental stresses, allelopathic interactions with other plants, etc. One of the major and important roles played by secondary metabolites is in plant defense mechanisms (Bennett and Wallsgrove, 1994; Theis & Lerdau, 2003; Mao *et al.,* 2007; Chen, 2008). They are important components of the age-long 'Chemical warfare' fought between plants and their pests and pathogens. Though vast numbers of PSMs have been reported and extensive literature on potential defence roles of many PSMs through bioassays is available, definitive proof of the defence function of a particular PSM still remains undetermined in most cases (Chen, 2008). Also the mode of action of all of the PSMs has not been understood yet.

There are many different classes of secondary metabolites produced in plants. Some major classes of secondary metabolites that have been shown to play important roles in plant defences against insect herbivores are cyanogenic glucosides, Glucosinolates,

non-protein amino acids, alkaloids, phenolics, terpenes, sesquiterpenes & sterols, phytoalexins, salicylic acid, methyl jasmonate (Bennett and Wallsgrove (1994), flavones, flavonoid, isoflavoinoid, thioglucosidase, hydroxamic acids, nitrogen compounds, etc. (Chen, 2008). Some examples of PSMs are tannins (polyphenol), maysin and apimaysin (flavones), isoorientin (flavonoids), glyceollin (isoflavonoid), lignin, sinalbin and sinigrin (thioglucosides), dhurrin (cyanogenic glucosides), DIMBOA & DIBOA (hydroxamic acids), monoterpenes and diterpenes, saponin and tomatine (glycosylated triterpenes), etc. (Chen, 2008).

In addition to these chemicals there is an enormous range of other compounds present in the plant kingdom with a varied distribution. Some of them are species-specific and others have almost universal distribution. A single metabolite or a class of metabolite present in plant may not form a defence system of that particular plant and several other defence-related compounds may be found at the same time. Metabolites like tannins, polyphenols, proteases and chitinases are widely distributed in plants in which other PSMs such as cyanogenic glucosides, glucosinolates, alkaloids are also present. In addition to these, plant might possess physical defence mechanism in the form of secondary thickenings, thorns, barbs, cuticular waxes, leaf hairs, etc. (Royle, 1976, Koliatakudy and Koller, 1983). However, the distribution of secondary metabolites within a plant is not uniform and may vary according to the plant's requirement from a defence point of view. Upregulated production and accumulation of secondary metabolites in response to herbivory, infection, wounding, etc. is now established to be part of the integrated defence mechanism of plant.

Members of Graminae e.g. wheat, maize and rye exclusively possess a secondary metabolite, Hydroxamic acid (Hx) that functions as natural insecticide protecting them against insect pests (Niemeyer and Perez, 1995). Hydroxamic acids occur in plants as glycosides, which are hydrolysed to the aglucone when the tissue is injured (Hofman and Hofmanova, 1969). An aglucone isolated from wheat, DIMBOA (2,4-dihydroxy-7-methoxy-1, 4-benzoxazin-3-one), plays an important role in cereals to confer resistance against aphids. DIMBOA is the major defense compound active against insect pests of wheat and maize (Gianoli and Niemeyer, 1998). Genes responsible for accumulation of DIMBOA are located on wheat chromosome 4A, on which other defense proteins such as Hfr-1 are present (Niemeyer and Jerez, 1997).

The role of Hx in cereal plant defense has been established by studies on accumulation of Hx (Niemeyer *et al.,* 1989), excretion of Hx (Givovich *et al.,* 1992) and enzymatic detoxification of Hx by aphids (Leszczynski *et al.,* 1992). It is suggested that the resistance of cereals containing Hx could be due to the inhibitory effect of Hx on aphid detoxification enzymes, which decreases the capacity of the aphid to detoxify the allelochemicals (Mukanganyama *et al.,* 2003).

1.3.2 Proteinase inhibitors:

Proteinase inhibitors (PIs) are proteins which act as inhibitors of specific proteases by interacting at the active site of enzymes. PIs can regulate the activity of specific corresponding proteases and play key regulatory roles in many biological processes. In plants, PIs are often present in storage tissue like tubers and seeds. Their expression may be induced in other tissue like leaves, stems and phloem by herbivory, pathogenesis or wounding (Ryan, 1990). On the basis of sequence homologies of their inhibitor domains, plant PIs have been classified into 48 families (Rawlings *et al.,* 2004; Habib and Fazili, 2007). Serpins, Kunitz, Bowman-Birk, Squash, Mustard, Cystatin, Kininogen, Potato Type-I and Potato type-II are the few major families of PIs to which many plant PIs belong. The defensive capabilities of plant PIs rely on inhibition of proteases present in insect guts or secreted by micro-organisms, causing a reduction in the availability of amino acids necessary for their growth and development (Lawrence and Koundal, 2002).

The contribution of plant PIs to plant defences by inhibiting the proteases of feeding herbivores affecting their growth and development was reported for the first time by Lipke *et al.,* (1954). A proteinaceous fraction from soybean (*Glycine max*) contained an inhibitor of the digestive proteolytic activity from the red flour beetle (*Tribolium castaneum*). This was the first description of inhibition of an insect enzyme by plant PI and discovery of a novel PI in soybeans with a potential to control insect pests. It is now well established that PIs are produced in host plant in response to herbivory by insects, these PIs are one of the most extensively studied classes of proteinaceous defence compounds with respect to crop protection strategies.

Insect guts are thought to be the most vulnerable and potential target for the control of insect pests. Controlling insect pests by inhibiting their digestive proteases has been

implicated in many crop plants by expressing the exogenous PIs. They have been incorporated into transgenic plants to enhance resistance (Hilder *et al.,* 1993). Plant PIs have been transferred into crop plants like tobacco and tomato and tested for their potential in defense against various insect species (De Leo *et al.,* 1998, 2001; Charity *et al.,* 1999). Herbivory of *Manduca sexta* on tobacco plants was found to be reduced by expressing a recombinant potato proteinase inhibitor (PIN) in the leaves (Johnson *et al.*, 1989). A review by Jouanin *et al.,* (1998) reveals information on different plant PIs studied and their application in plant protection.

Helicoverpa armigera, known by many names (Cotton Bollworm, Corn Earworm, Tobacco Budworm or Old World Bollworm) is a polyphagous and devastating pest worldwide, feeding on wide range of crop plants including cotton, maize, tomato, chickpea, pigeon pea (Patankar *et al.,* 2001). It has been used as a model pest to study the effect of several host or non-host plant proteinase inhibitors (Harsulkar *et al.,* 1999; Telang *et al.,* 2003; Tamhane *et al.,* 2005). *Alocasia macrorrhiza* proteinase inhibitor was over-expressed in tobacco and showed growth retardation in *Helicoverpa armigera* feeding on transgenic plants (Wu *et al.,* 1997). A multi-domain PI from *Nicotiana alata* was over-expressed in tobacco plant and showed increased mortality and delayed growth of *H. armigera* feeding on leaves of transgenic plant (Charity *et al.,* 1999). Two Kunitz-type proteinase inhibitors (WCI2 and WCI5; both chymotrypsin inhibitors) were recombinantly expressed in *E. coli* and purified PIs showed growth retardation of *H. armigera* when fed in artificial diet (Telang *et al.,* 2009a). An inhibitor belonging to Squash family (McTI-II) was recombinantly produced in *P. pastoris* and fed in the artificial diet showed antimetabolic effects on growth and development of *H. armigera* (Telang *et al.,* 2009b).

However, this strategy has not worked for all the transgenic crops. Polyphagous insects like *Helicoverpa zea* and *Spodoptera exigua* have been shown to adapt to the presence of proteinase inhibitors in their diet, by switching to the production of proteinases that are resistant or insensitive to plant proteinase inhibitors (Jongsma *et al.,* 1995; Broadway *et al.,* 1995; 1997 and Mazumdar-Leighton & Broadway, 2001). Also the insect guts can synthesize large number of protease isoforms and the secretion of these depends on several other factors such as the diet on which they are feeding. Insects feeding on artificial diet containing proteinase inhibitors have been

shown to secrete proteases that are insensitive to these proteinase inhibitors (Bown *et al.,* 1997 and Chougule *et al.,* 2005). There are some examples showing a non-significant effect of exogenous PI genes in host plant, soybean trypsin inhibitor highly expressed in tobacco plants failed to confer resistance against *H. armigera* (Nandi *et al.,* 1999). Under these circumstances, PIs could be used with additional defense proteins or compounds by 'stacking' method, which would make insects more vulnerable. Stacking of genes is using more than one resistance genes for plant transformation in combination to make plants completely resistant to insect pests.

Wheat expresses several proteinase inhibitors in its leaves and food grains, which are both serine as well as cysteine type PIs. Two PIs from wheat have been studied in detail for their characterisation by recombinant expression and their inhibition of insect gut proteases. An inhibitor of bacterial subtilisin and animal chymotrypsin (WSCI) belonging to potato inhibitor-I family was purified from endosperm of hexaploid wheat and its complete amino acid sequence has been revealed by Poerio *et al.,* (2003). This inhibitor consisting of 72 amino acid residues with molecular mass of 8.1263 kDa showed sequence similarity to barley subtilisin-chymotrypsin iso-inhibitors of CI-2 type and with maize subtilisin-chymotrypsin inhibitor (MPI). This inhibitor was effective against subtilisin from *Bacillus licheniformis* and *B. subtilis* along with different forms (α, β, λ and δ) of bovine and porcine chymotrypsin (Poerio *et al.,* 2003). However WSCI did not show inhibitory activity against bovine trypsin, porcine pancreatic or human leukocyte elastase. WSCI inhibited protease activity present in crude extracts of midguts dissected from coleopteran larvae of *T. molitor* and from lepidopteran larvae of two insects (*H. armigera* and *Plodia interpunctella*) (Poerio *et al.,* 2003). It is now established that coleopteran insects possess mainly cysteine type digestive proteases (Terra *et al.,* 1996) in their gut and inhibition of coleopteran gut extracts, suggesting that WSCI might also be inhibiting activity of cysteine type proteases.

In plants a class of cysteine proteinase inhibitors called cystatins or phytocystatins regulates activity of cysteine proteases. Cystatins are proteins ranging from molecular mass of 12 to 16kDa and are characterised by absence of disulphide bonds. Presence of the conserved LARFAVDEHN sequence, in the N-terminal region is typical to phytocystatins and absent in animal cystatins. Many cDNA sequences of

phytocystatins cloned from seeds, fruits, leaves and roots have been reported (Abe *et al.,* 1987; Ojima *et al.,* 1997; Ryan *et al.,* 1998). These phytocystatins reversibly bind to cysteine peptidases thereby blocking their activity. Cystatins form a tripartite wedge-shaped structure and directly interact with the active site clefts of cysteine proteases (Corre-Menguy *et al.,* 2002). Three important regions involved in the mechanism of inhibition are; a conserved Glycine (Gly) residue located within the N-terminal region, a central loop segment containing the highly conserved QxVxG motif and a tryptophan residue located in the C-terminal region (Bode *et al.,* 1988; Machleidt *et al.,* 1989).

Phytocystatins play an important role in plant defense mechanisms by inhibition of digestive peptidases present in insect guts. Cystatin over-expression in response to herbivory by insect pests shows their involvement in plant endogenous resistance mechanisms (Botella *et al.,* 1996). Transgenic plants over-expressing exogenous cystatin genes have shown increased resistance to insect pests, where OC-I expressing poplar (*Populus tremula*) showed resistance towards *Chrysomela tremulae* larvae (Leple *et al.,* 1995). Many isoforms of wheat cystatins have been cloned so far. Four types of cDNAs coding for cystatins WC1, WC2, WC3 and WC4 were cloned from wheat seed. All four were different cDNA forms of cysteine proteinase inhibitors and showed 47-68% amino acid sequence similarity to other plant cystatins (Kuroda *et al.,* 2001). Another wheat cDNA encoding cystatin (WC5) was cloned from wheat embryos and recombinant WC5 inhibited the protease activity of seed protein extracts (Corre-Menguy *et al.,* 2002).

Inhibition of insect gut protease activity by WSCI and cystatin like wheat PIs suggests their potential role in plant defense which can be utilized to confer resistance to wheat against insect pests. These PIs could be potential antimetabolic endogenous defense proteins, which could be tested against two wheat pests under consideration, *D. coarctata* and *S. avenae.*

1.3.3 Lectins:

Lectins are sugar-binding proteins found in all plants, animals, bacteria. In plants they act as storage as well as defence compounds. Lectins are known to cause agglutination of cells and/or precipitation of glycoconjugates (Giovanini *et al.,* 2007).

Lectins were first isolated from plants with sugar specific agglutination properties (Boyd and Shapleigh, 1954) and hence are also called 'agglutinins'. Plant lectins are classified into four main groups according to their Carbohydrate Recognition Domains (CRDs), (1) Legume lectins from members of the Leguminoseae family, (2) Chitin-binding lectins possessing one or more hevein domains (hevein domain is a 43 amino acid chitin-binding polypeptide present in latex of rubber tree; *Hevea brasiliensis*), (3) type-2 RIPs (Ribosome Inactivating Proteins) and (4) mannose-binding lectins from monocots and further into 3 small groups (5) Jacalin-related lectins (found in seeds of Jackfruit), (6) Cucurbitaceae phloem lectins (chitin binding lectins found only in phloem of members of Cucurbitaceae family) and (7) Amaranthin group (found in *Amaranthus spp.*). These seven groups include virtually all the plant lectins reported so far (Van Damme *et al.,* 1998). Plant lectins contain at least one noncatalytic domain that binds reversibly to specific mono or oligosaccharides (Peumans and Van Damme, 1998). Chitin binding lectins contain one or more hevein domains and are ubiquitous in plants. They bind to N-acetylglucosamine and oligomers and polymers of GlcNAc (Murdock and Shade, 2002). Many plant lectins-agglutinins have been isolated, cloned and characterised so far e.g. lectins from snowdrop (*Galanthus nivalis;* GNA), Common bean (*Phaseolus vulgaris*; PHA), Garlic (*Allium sativum*; ASA), Wheat (*Triticum aestivum*; WGA) by Van Damme *et al.,* (1987), Rigas *et al.,* (1955), Van Damme *et al.,* (1992), Raikhel and Wilkins, (1987) respectively.

Two major roles of lectins in plants have been understood so far. They act as stores of proteins that can be mobilised for plant growth and development, and secondly they are plant defences against feeding herbivores and pathogens. Lectins are usually accumulated during the normal growth and development or reproductive phase of the plant life cycle, mobilized and are used later (Murdock and Shade 2002). Sometimes lectins occur in very high abundance e.g. a low molecular weight mannose binding lectin makes up 75% of the proteins content in the nectar of leek (*Allium porrum*) flowers (Peumans *et al.,* 1997). Lectins are abundantly found in the seeds as well, which suggest their possible role as storage proteins or protection of seeds from herbivores and pathogens. Legume seeds contain 1 to 10% lectins as total soluble proteins (Puemans and Van Damme, 1998).

The role of lectins in plant defence against feeding insects is now established and there are many views on the mode of action of lectins on insects. According to Murdock and Shade, (2002), there is no obvious correlation between the sugar specificity of lectins and their toxicity when fed to insects, however some trends have been observed. Lectins that bind to N-acetyl glucasamine (GlnAc) and its oligomers often retard growth and development of certain beetles after feeding (Murdock *et al.,* 1990). Lectins bind to the midgut tissue of insects, however no correlation was found between the ability of lectin to bind to insect midgut and their toxicity as there are several lectins that bind to midgut tissues, which are still not toxic. All lectins that are toxic to insects exert their toxicity via binding to specific carbohydrate moities. The prerequisite for action of lectins is that they survive the hostile proteolytic conditions of gut environment in insect. Also a lectin toxic to one insect may not necessarily be toxic to other insects. Three possible modes of action of lectins have been suggested that affect the overall growth of insects (Murdock and Shade, 2002). First is causing difficulty in food recognition by insect. Insects recognise the quality of food by sensory receptors located on tips of feet, the tarsi, antennae and mouthparts. Binding of lectins in diet to the carbohydrate moities associated with membrane of chemosensory sensillae could block the access of food chemical signals to their actual receptor proteins. Also lectins could disrupt the integrity of sensory membranes thereby affecting the detection of food. Secondly, lectins may act on the peritrophic matrix (PM) of insects. PM is a protective envelope secreted by the epidermal cells of the midgut and contains proteins, glycoproteins, chitin and glucosamineglycans. Lectins could bind to the surfaces of epithelial cells in insect midgut, which possibly is the third site of action (Powell *et al.,* 1998; Zhu-Salzmann *et al.,* 1998 and Harper *et al.,* 1995). These cells perform the very important function of secreting digestive enzymes and absorption of chemical products of digestion.

GNA was found to bind strongly to brush border membrane vesicles made from the midgut of European corn borer (*Ostrinia nubilalis*) but was not toxic to the insects suggesting that binding alone is not sufficient to cause disruption of physiological function (Harper *et al.,* 1995). However some molecular variants of single lectin bound to the brush border membrane vesicles caused mortality also, suggesting molecular variants of lectins specifically bind to cell surface carbohydrate molecules depending on insect species or particular stage of development. This was supported

by creating site-specific mutants of recombinant N-acetylglucosamine-specific lectin GSII from African legume (*Griffonia simplicifolia*) which showed both types of activity (Zhu-Salzmann *et al.,* 1998).

Among the plant lectins, a lectin from snowdrop bulbs (GNA; *Galanthus nivalis* agglutinin) is an extensively studied anti-insect lectin. GNA is a mannose-binding lectin, a homotetramer of approx. 50kD molecular mass. Each homotetramer contains a non-covalently bonded protomer with three highly homologous mannose-binding sites. The number of subunits per molecule in mannose binding lectins is suggested to be correlated with insecticidal properties (Murdock and Shade, 2002). GNA was highly insecticidal and the most effective plant lectin against many insects belonging to different insect groups. Some examples include, rice brown plant hopper (*Nilaparvata lugens*) and rice green leafhopper (*Nephottetix cinciteps*) (Powell *et al.,* 1995), pea aphid (*Acyrthosiphon pisum; Hemiptera*) sugarcane whitegrub (*Antitrogus sanguineus;* Coleoptera) (Allsopp and McGhie, 1996), potato aphid (*Aulocothum solani*; Hemiptera), (Down *et al.,* 1996) and tomato moth (*Lacanobia oleracea*; Lepidoptera) (Fitches and Gatehouse, 1998). The insecticidal effects of GNA on these insects were due to antifeedant and antinutritive properties that caused mortality, retarded growth or affected fecundity (Murdock and Shade, 2002). GNA was found to have much lower effects on mammals as compared to other plant lectins. The antifeedant and antinutritive qualities of GNA have made it a potential candidate to be considered as a defence protein to be introduced into other crop plants. Many crop plants have been transformed with GNA e.g. potato (*Solanum tuberosum*) (Gatehouse *et al.,* 1996), rice (*Oryza sativa*) (Rao *et al.,* 1998), wheat (*Triticum aestivum*) (Stoger *et al.,* 1999) and tobacco (*Nicotiana tabacum*) (Wang and Guo, 1999). All the plants expressing GNA showed different levels of resistance against the target pests as compared to non-expressing control plants.

Wheat germ agglutinin (WGA), a chitin binding lectin from wheat has been extensively studied. Feeding WGA to European corn borer larvae (*O. nubilais*) disrupted the chitin-meshwork of PM causing large holes in the envelope. WGA also affected the microvillar structure of the midgut (Harper *et al.,* 1998). WGA has a very high affinity for oligomers of GlcNAc (N-Acetylglucosamine). WGA may bind to the nascent chitin oligosaccharide chains required for the formation of chitin polymers,

which is the major component of PM. Binding of WGA to glycoproteins may disrupt the assembly of glycoprotein-chitin linkages needed for normal PM. Disruption of chitin network results in hyper secretion of PM by the microvilli followed by disintegration of this tissue, which is lethal to the insect (Murdock and Shade 2002). Wheat also expresses lectin like proteins in response to herbivory by insects.

Three lectin like defence proteins (Hessian fly responsive; Hfrs) are specifically overexpressed in response to herbivory by Hessian fly (*Mayetiola destructor*) in wheat and are important insect specific defence proteins. The genetic interactions between these two operate on a gene-for-gene basis (Hatchett and Gallun, 1970; Gallun, 1978) similar to plant and microbial pathogen interactions. Gene for gene interaction displays another important endogenous resistance mechanism of plants in response to herbivory.

1.3.4 Gene for gene interaction between wheat and Hessian fly:

Hessian fly (*Mayetiola destructor* Say; Diptera) is a significant and devastating pest of cereal crops including wheat, barley and rye. The female fly lays a large number of eggs in the leaf sheaths of cereal crops. Emerged larvae feed on sap and weaken the plants leading to failure in grain filling. Wheat is one of the preferred hosts of Hessian fly and the interactions between wheat and Hessian fly have been studied extensively. Gene-for-gene interaction and its genetic basis were first reported in rust (*Melampsora lini*) and its host, flax (*Linum usitatissimum*) (Flor, 1955). The inheritance of both resistance in host and ability of parasite to cause disease is controlled by a pair of corresponding genes, the 'resistance' (R) gene in host for resistance and 'avirulence' (Avr) gene in parasite. A host plant producing a specific R gene product becomes resistant to a pathogen that produces the corresponding Avr gene product. The wheat-Hessian fly interaction is very specific i.e. though wheat possesses one or more resistance genes, the plant becomes resistant only when its resistance gene is able to mediate recognition of the specific biotype of Hessian fly larva that has infested it (Williams *et al.,* 2002). Accordingly, a wheat genotype can be resistant by participating in an incompatible interaction with a biotype of larvae that is avirulent on that wheat genotype. However, the same wheat genotype can become susceptible by participating in an incompatible interaction with a different larval biotype that is virulent on that wheat genotype.

Such specific and complex interactions between host and pest are important in understanding molecular analysis of signaling pathways and gene regulation. More than 29 resistance genes that protect wheat against 16 biotypes of the Hessian fly have been characterised in wheat and related species (Delibes *et al.,* 1997; Ratcliffe and Hatchett, 1997). Studies of changes in wheat gene expression during infestation by Hessian fly larvae have shown three important genes that are up regulated. These were called as Hessian fly responsive genes (Hfrs). Three Hfrs [Hfr-1 (Williams *et al.,* 2002), Hfr-2 (Puthoff *et al.,* 2005) and Hfr-3 (Giovanini *et al.,* 2007)] genes have been cloned and well characterised. All three Hfr genes code for defence proteins and have at least one domain coding for lectin / agglutinin like protein. Taking into consideration the role of lectin/ agglutinin like proteins in plants' defense mechanism against feeding herbivores, Hfrs play important role in wheat's defense against Hessian fly.

1.4 Increasing resistance of wheat against insect pests:

1.4.1 Wheat Breeding:

Biotechnology of wheat started in the early stages of wheat cultivation where breeding was the favoured and feasible method for improvement in wheat quality. Wheat breeders have been successful in introducing desirable traits that increased yield and minimized crop losses due to biotic or abiotic stresses. Conventional breeding utilizes domestic crop cultivars and related genera as a source of genes for improvement of existing cultivars. The process of breeding involves the transfer of a set of genes related to beneficial traits from the donor to the recipient. However, conventional breeding techniques are based on processes of crossing, back crossing and selection, which is time consuming and laborious. The process of obtaining the necessary breed is challenging due to fast co-evolution of pathogenic microorganisms and pests. The green revolution, which allowed an immense increase in wheat production, in early 60s was only possible due to conventional breeding, coupled with improved farm management practices (Patnaik and Khurana, 2001).

1.4.2 Genetic modification of wheat using biotechnological tools:

In recent years, with availability of different biotechnological tools, conventional wheat breeding has been coupled with *in vitro* technologies. Biotechnological methods are comparatively faster than traditional breeding methods. In contrast to

conventional breeding methods, modern Biotechnological methods can transfer defined genes from any organism, thereby increasing the gene pool available for improvement of crop plants. Genetic improvement of crop plants by introduction of agronomically important genes and making them Genetically Modified (GM), using modern biotechnological tools such as plant transformation is a promising method. Two plant transformation methods commonly used are Particle Bombardment and *Agrobacterium* mediated transformation. Using these methods, development of many transgenic crop plants expressing insecticidal toxins or proteins has been achieved.

Transgenic crops expressing δ-endotoxin or 'cry' proteins (A class of crystalline proteins produced by strains of *Bacillus thuringiensis; Bt*) were the first successful steps in this direction. Crop plants expressing Bt toxins have now been commercialised all over world and shown resistance against many lepidopteran pests (Shelton *et al.,* 2002). Use of other potential insecticidal proteins either on their own or in combination with Bt proteins was thought to be a possible solution to avoid this problem. However, development of transgenic crops with resistance involves deep understanding of the interactions between crop plants and insect pests. Studies of this kind involve identification of potential targets in insect pests. Some of the insecticidal proteins that have been tested against various insect pests are lectins (Gatehouse *et al.,* 1997), chitinases (Ding *et al.,* 1998), cholesterol oxidases (Purcell *et al.,* 1993; Cho *et al.,* 1995), avidin (Kramer *et al.,* 2000) and protease inhibitors (Hilder *et al.,* 1987). Inhibition of insect gut digestive proteases using plant proteinase inhibitors or use of lectin like proteins to cause damage to insect gut, thereby reducing insect survival or fecundity are two most studied approaches.

Many insecticidal proteins/toxins that have been shown to have potential in crop protection are only effective in insect hemolymph and not in gut. When given in food, these proteins are unable to cross the insect gut wall to reach their site of action and undergo proteolytic degradation in the gut. It has been suggested that this problem could be overcome by use of the 'Fusion protein technology' in which a carrier protein like a plant lectin can be used (Fitches *et al.,* 2004b). Lectins are known to bind the insect gut wall and be able to enter the hemolymph. Insecticidal proteins or toxins linked to lectins thus can be delivered to insect hemolymph. Most of these crop

protection strategies that could lead to development of GM crops are under constant pressure from environmental considerations.

Many novel heterologous DNAs encoding highly potential antimicrobial peptides, defence related proteins and enzymes for the production of anti microbial compounds in crop plants have greatly enhanced the possibility of engineering crop plants for resistance to pests and pathogens (Shah, 1997; Sharma *et al.,* 2000). Though plant genetic transformation by introduction of foreign genes is obtained and is successful in most of the crop plants, wheat was the last cereal to be genetically transformed (Patnaik and Khurana 2001). The large genome size (approx. 1700mb) of wheat makes the genetic improvement of this crop by any method genetically challenging. Many agronomically important genes have been introduced into what by genetic transformation such as, quality improvement, engineering of male sterility, transposon tagging, resistance to draught and stress, resistance against fungal pathogens and insect resistance (Patnaik and Khurana, 2001).

It has been possible to introduce genes conferring resistance to wheat against many different types of biotic stresses (pathogens and insect pests). Most of the genetic improvement of wheat approaches concentrated on fungal pathogens, as they are the major threat to wheat. For developing resistance against insect pests in wheat many genes encoding lectin, chitinase, ribosome inactivating proteins (RIP) and proteinase inhibitors have been successfully introduced into wheat. Use of proteinaceous defence compounds like lectins and proteinase inhibitors have shown great effect by decreasing the fecundity of insects feeding on transgenic seeds. Many such genes related to increase in developing resistance in wheat against pathogens and insect pests, have been reviewed by Patnaik and Khurana, (2001).

Wheat proteinase inhibitors (discussed earlier) as endogenous resistance factors could be important candidates to confer resistance on wheat against many of its insect pests. However, it is first extremely necessary to understand the digestive physiology of the relevant insect pest. Insects are diverse groups of animals and within this group are representatives can eat virtually anything organic. Tissues of plant origin include wood, leaves, flowers, roots, tubers, nectar, seeds and fruits while tissues of animal origin include flesh, animal wastes, blood and other insects. Fungi and bacteria are

also eaten by some insects. Insects are required to digest different polymers in their ingested diet, such as starch, cellulose, hemicellulose and proteins (Dadd, 1985). Protein diet however is essential for the better growth and survival of insects. Insects require same 10 essential amino acids to complete their life cycle as other animals (Murdock and Shade 2002) that come from dietary protein. Insects use their major digestive organ the 'Midgut' for proteolytic digestion. Insects possess a wide diversity of digestive proteases as they feed on all different kinds of protein diet. It is now established that insects contain both serine and cysteine proteases and they are involved in gut proteolytic digestion. However it is important to identify types of protease(s) in gut and their role in protein digestion, which will help in selecting the particular PIs. Recombinant expression of these proteases would make available the active proteases that can be characterised for their properties. With *in vitro* inhibition assays with respective recombinant host plant proteinase inhibitors would assess the validity of host PIs as potential candidates for insect control.

1.4.3 Insect proteases as potential targets for crop protection strategy

Proteases are enzymes that catalyze hydrolytic cleavage of specific peptide bonds in proteins. These enzymes are widely distributed in nearly all plants, animals and microorganisms (Joanitti *et al.,* 2006; Valueva and Mosolov, 2004; Christeller, 2005; Haq *et al.,* 2004; Supuran *et al.,* 2002; Habib *et al.,* 2007) and in higher organisms nearly 2% of the genes code for these enzymes (Barrett *et al.,* 2001). Apart from their potential role in digestion and development, proteases play crucial role in many biological processes. These include mediation of signal initiation, transmission and termination in many cellular events such as inflammation, apoptosis, blood clotting, hormone processing pathways, immune responses, etc. (Habib *et al.,* 2007; Ivanov *et al.,* 2006; Barrett *et al.,* 2003).

Proteases have been classified based on two points of views, first the wideness of divergence in their catalytic activity and second to what extent peptidases from separate evolutionary lines have converged in properties. Accordingly they are classified into following, serine type proteases, cysteine type proteases, aspartic type proteases, metallo-type proteases and unknown catalytic type proteases (Rawlings and Barrett, 1993). Two types of proteases serine and cysteine, which are shown to be involved in insect gut protein digestions are discussed here.

Serine proteases (SPs) or serine endopeptidases are proteases containing Serine as one of the active site residue in the reactive site. There are two major types of SPs, Trypsin-like (EC 3.4.21.4) and Chymotrypsin-like (EC 3.4.21.1). Chymotrypsin like SPs are one of the thoroughly understood families of enzymes, members of this S1 family of SPs (Rawlings and Barrett, 1993) have several important roles in physiological functions in higher animals including digestion, blood coagulation, cellular immunity, fertilisation and embryonic development. Study of bovine chymotrypsin by X-ray crystallography revealed the presence of an active centre containing His57, Asp102 and Ser195 (Bovine chymotrypsin numbering), known as the 'Catalytic triad' that is responsible for the acyl transfer mechanism of catalysis (Ross *et al.,* 2003). Substrate binding clefts present near the active site determine the substrate specificity of these proteases. SPs are commonly synthesized as zymogens that are converted to the active enzyme by proteolytic cleavage at a particular peptide bond. There is a cascade pathway for the activation of these SPs in which one protease activates the zymogen of another to mediate a rapid local reaction. The blood clotting cascade in human plasma and the phenoloxidase pathway in the insect immune response are two well studies examples of this type of cascade. Cysteine proteinases are peptidases having a cysteine residue at the active site, which provides the catalytic nucleophile for peptide bond hydrolysis. Cathepsin L (EC 3.4.22.15) is a cysteine proteinase with a broad specificity for cleavage of peptide bonds, which is normally located in lysosomes, and is ubiquitous in eukaryotic cells. It is synthesized as a pre-proenzyme by ribosomes bound to the ER, and may be modified by glycosylation in the Golgi network prior to transport. The propeptide acts as an intramolecular inhibitor, and is necessary for proper folding, stability and transport to the Golgi apparatus (Tao *et al.,* 1994).

1.4.3.1 Insect serine proteases

A large gene family of Serine Proteases (SPs) is present in *Drosophila melanogaster* and they are classified according to their sequence identities and conserved reactive site residues. Similar families of SPs are present in other Dipteran insects, which contain several isoforms of SPs involved in a whole array of functions. The presence of 306 SP-related genes has been reported from *Anopheles gambiae* (Holt *et al.,* 2002; Zdobnov *et al.,* 2002; Christophides *et al.,* 2002). The presence of three important conserved regions- TAAHC, DIAL and GDSGGP is a hallmark for identification of

these SPs. GD**S**GGP possesses a reactive Serine residue (Ser195; chymotrypsin numbering) that attacks and hydrolyzes the scissile peptide bond of a substrate by an acylation-deacylation mechanism and this region is highly conserved. TAA**H**C contains another essential residue His (His57 chymotrypsin numbering), which abstracts the proton from the serine hydroxyl side chain and allows a nucleophilic attack on the protein substrate and TAAHC region shows less conservation than GDSGGP. Another region, DI**A**L containing Asp (Asp102 in chymotrypsin) helps in the stabilisation of protonated histidine and this region is least conserved among the three conserved regions.

Expression of recombinant serine proteases in *E.coli* has been attempted by several workers and the efforts to express bovine trypsin or trypsinogen in the cytoplasm of *E. coli* produced the insoluble aggregates i.e. inclusion bodies (IBs) (Greaney and Rosteck, 1994). Human trypsinogen expressed in *E. coli* also resulted in formation of IBs accompanied with a low yield of 30–60 mg L^{-1}, (Kopetzki *et al.*, 1999). Bovine trypsinogen was recombinantly expressed as IBs at a level of 40 mg L^{-1} (Peterson *et al.*, 2001). These IBs are functionally inactive and refolding or renaturing of IBs to get the active enzyme is often unsuccessful. Expression of rat trypsin in *E. coli* using *pho* leader sequence (Vasquez *et al.*, 1989) enabled to successful secretion of the active trypsin to the periplasm, however the yield was very low, even in an optimized fed batch process, (56 mg L^{-1} ; Yee and Blanch, 1993). The resulting low yield was thought to be due to the damage of host cells by the proteolytic activity of the product. So further efforts were made to express the mutant protease in an inactive form and recombinant expression was followed by later activation.

1.4.3.2 Insect Cysteine proteases:

The biochemistry of protein digestion in insect guts is complex, and variable between species, but in Coleoptera and Hemiptera, cysteine proteinases are important digestive enzymes (Terra *et al.,* 1996). Although cysteine proteinases have been identified in gut tissue in a number of Dipteran insect species, their role in digestion of dietary proteins remains to be established. A cathepsin L-like cysteine proteinase from *Drosophila melanogaster* Cp-1 (CG6692), was reported as present in alimentary organs (Matsumoto *et al.,* 1995), and was suggested to play a role in digestion, but the encoding gene is also expressed throughout the insect (FlyAtlas;

http://www.flyatlas.org). A cathepsin B-like cysteine proteinase whose expression was upregulated after feeding was cloned from gut tissue of tsetse fly (*Glossina morsitans morsitans*; Yan *et al.,* 2002), although a role in digestion was not established. Cysteine proteinase activity, and a specific cathepsin L-like proteinase designated DrCP1 was present in gut tissue of the crucifer root fly (*Delia radicum*), but like *D. melanogaster* CP-1, DrCP1 was also present in other tissues, and was not considered to act as a digestive enzyme (Hegedus *et al.,* 2002). Several studies have concluded that the major role for cathepsin L-like cysteine proteinases in insects is that of tissue remodeling, based on a seminal study of the enzyme in the flesh fly, *Sarcophaga peregrina* (Homma *et al.,* 1994). A similar enzyme in *D. radicum*, DrCP1, shows elevated expression levels associated with moulting and metamorphosis, and was suggested to be involved in all aspects of tissue reconstruction (Hegedus *et al.,* 2002). In addition, cathepsins have been proposed to be important in embryogenesis (Yamamoto and Takahashi, 1993) and in reproductive processes (Matsumoto *et al.,* 1997). In *Helicoverpa armigera*, a significant role for cathepsin L in cuticle degradation during insect moulting has been proposed (Liu *et al.,* 2006), in agreement with earlier studies in nematodes and crustaceans, where inhibition of enzyme activity (Richer *et al.,* 1993) or down-regulation of enzyme expression by RNAi (Hashmi *et al.,* 2002) has shown that cathepsin L-like cysteine proteinases were needed for successful moulting. The requirement for cathepsin L-like activity in tissue remodeling in insects is thought to involve degradation of basement membrane by limited proteolysis (Homma *et al.,* 1994). Cathepsin like cysteine proteases have been recombinantly expressed using different expression systems and were proteolytically active (Phillip *et al.,* 2007 and Renard *et al.,* 2000).

Insects belonging to the Orders Lepidoptera, Coleoptera and Diptera are thus known to use these serine and cysteine type proteases for gut proteolytic digestion. However insects belonging to Hemiptera feed on plant phloem and are thought not to require any protein digestion in their gut. Due to their unique feeding behaviour interactions between aphids and plants are different than in case of other herbivorous insects. It is necessary to investigate whether these insects can utilize gut proteases for digestion at least to a limited extent, which would allow the use of plant proteinase inhibitors against these insect pests. Cereal aphid (*Sitobion avenae*) is a major pest of wheat. The most widely used control strategy against aphids is use of chemical insecticides,

which has serious drawbacks like environmental and food chain contamination. Additionally, ever increasing use of chemical insecticides is causing an emerging problem of developing resistance by aphids. As a step towards environmentally friendly crop protection strategy, a small number of natural aphid-resistance genes have been characterised so far (Deraison *et al.,* 2004; Pitrat and Lecoq, 1982; Khush & Brar, 1991; Porter *et al.,* 2000). Two such genes have been successfully cloned, the *Mi* gene provides resistance to tomato against a nematode *Macrosiphium euphorbiae* (Rossi *et al.,* 1998) and *Vat* genes in watermelon against *Aphis gossipii* (Brotman *et al.,* 2002) and most of these genes are specific to single aphid species.

1.4.4 Phloem feeding insects (aphids):

Phytophagous insects are dependent on utilisable nitrogen as a dietary component for their growth and fecundity. Availability of nitrogen (quantity) and composition (quality) of nitrogen are two important considerations for the success of insects feeding on plants (Douglas, 2005). Hemipteran insects, including aphids are the only group of animals that can survive on the grossly unbalanced amino acid supply from the plant phloem (Douglas 1993, Douglas, 2006a; Sandstrom and Moran, 1999, Sandstrom *et al.,* 2000). Feeding on plant phloem poses two major problems for aphids. First the 'nitrogen barrier' which is low nitrogen quality in the form of essential amino acids and secondly, the 'sugar barrier' which is very high concentration of sugars. (Douglas, 2005 and 2006b).

It has been shown that the essential amino acid content of phloem sap is insufficient to support the observed growth rate of the aphids (Douglas, 2005). To satisfy the requirement of essential amino acids for survival has compelled these insects to form a symbiotic association with micro-organisms e.g. *Buchnera* which are situated in the specialised organs called as bacteriocytes in these insects (Buchner, 1965). *Buchnera aphidicola*, a γ-proteobacterium, supplies its host with the essential amino acids and there are strong dietary, metabolic and genomic evidences for this (Douglas, 1988; Febvay *et al.,* 1999; Shigenobu *et al.,* 2000). It is now widely accepted that symbiotic bacteria complement the essential amino acids supply in the phloem sap (Douglas 2003; Dale and Moran, 2006; Gunduz and Douglas, 2009). Due to their specialist feeding habit and surviving on minimal or unbalanced supply of free amino acids from plant phloem, aphids have become a class of insect pests that need special

attention. Their feeding on phloem sap and not requiring any additional nitrogenous nutritional resources raises many questions. In spite of association with symbiotic bacteria are they required to utilise phloem peptides or proteins for their nutritional fulfilment, whether they have specific proteases in their guts, which are devoted for gut proteolytic digestion to digest the ingested protein, if so then what kind of proteases do they have in their guts? Before answering these questions, it is necessary to understand the composition of plant phloem and presence of proteins that could be ingested and used for gut proteolytic digestion by aphids.

1.4.4.1 Plant phloem sap:

Plant phloem sap is like a pre-digested food for the insects that feed on it. It consists of high concentrations of sugars that provide an abundant source of carbon and energy, and nitrogen predominantly in the form of free amino acids (Douglas, 2005). It is also free of toxins and feeding deterrents, hence making it a poorly defended, nutrient rich food source for those animals with special feeding devices like aphids with stylets. Sugars derived from photosynthesis occur dominantly in plant phloem, especially in the form of sucrose, which is a chemically stable disaccharide and has low viscosity. The concentration of sugars in phloem sap is very high and can reach values upto 1M, which poses osmotic pressures 2-5 times greater than that in insects' body. However, low nitrogen quality is the major concern for phloem feeding insects. The ratio of essential:non-essential amino acids in plant phloem is 1:4 to 1:20 which is considerably lower than in animal protein (1:1) (Douglas, 2005). Apart from this, phloem sap contains many different proteins in varying concentrations 0.3 to 9.8 mg ml^{-1} and the content of protein in phloem varies in different plant species (Zeigler, 1975). In plants belonging to Cucurbitaceae, concentrations of phloem proteins have been reported to be as high 10 to 40 (Thompson and Schulz, 1999) or 60 mg ml^{-1} (Cronshaw and Sabnis, 1990). These proteins include small peptides, several enzymes, proteinase inhibitors, lectins, etc. (Kehr, 2006). Protein concentrations in excess of about 5 mg/ml in phloem sap could provide an adequate source of amino acids and nitrogen for phloem-feeding insects (Deraison *et al.,* 2004). The presence of proteinase inhibitors in the phloem sap might be related to and play an important role in plants' defence against phloem feeding insects by inhibiting their gut proteases. Many proteinase inhibitors from plant phloem have been reported e.g. an aspartic protease inhibitor from *Cucumis melo* (Christeller *et al.,* 1998; Walz *et al.,* 2004), a

Chymotrypsin inhibitor from *C. melo* (Murray and Christeller, 1995; Walz *et al.,* 2004) and cystatin from *Brassica napus*, *C. melo, C. sativus, Ricinus communis* (Barnes *et al.,* 2004; Haebel and Kehr, 2001; Schobert *et al.,* 1998; Walz *et al.,* 2004), serpin-1 from *C. melo* (Walz *et al.,* 2004; Yoo *et al.,* 2000) and Trypsin inhibitor from *C. melo* and *C. sativus* (Murray and Christeller, 1995; Walz *et al.,* 2004). The high abundance of different types of PIs in phloem sap is consistent with growing evidence that phloem feeders contain digestive proteases (Kehr, 2006) and therefore may digest ingested proteins.

1.4.4.2 Proteases in phloem feeding insect guts:

It has been established that aphids contain proteases in the gut, which might be involved in digestion, and the majority of them are of cysteine type proteases (Terra, 1990). The cotton melon-aphid (*Aphis gossypii*) digestive tract proteinase activity was found to be of the cysteine type. A gut specific cathepsin-L like cysteine proteinase (AgCatL) has been cloned and characterised (Deraison *et al.,* 2004). Pea aphid (*Acyrthosiphon pisum*) has been shown to contain a family of genes coding for cathepsin-B like cysteine proteinases and many of them were found to be over-expressed in gut (Rispe *et al.,* 2007). A Cathepsin-L like protease from the midgut of blood feeding bug, *Rhodnius prolixus* involved in the insect digestion was reported by Houseman and Downe (1980) which was the first demonstration of a cysteine proteinase being involved in gut protein digestion (Foissac *et al.,* 2002). Recently several proteases have been reported from aphid guts e.g. aminopeptidase, cathepsin-L like cysteine proteinase (Rahbe *et al.,* 1995; Cristofoletti *et al.,* 2003 and Kutsukake *et al.,* 2004). It is now established that most of the proteases involved in insect digestion to be cysteine type proteases (Silva and Terra, 1994).

1.4.4.3 Protein digestion in phloem feeding insects:

Aphids uptake peptides and proteins present in phloem sap without any restrictions. It was long been a debate whether aphids can carry out gut proteolysis to fulfil the need of amino acids required. Though many proteases with their possible role in protein digestion from aphids have been reported, there are only a few reports that actually support the hypothesis of gut proteolytic digestion. With recent findings it is now established that aphids can carry out proteolytic digestion in their guts to generate the amino acids required for their growth and fecundity, though some believe that they

have only limited capacity to do this (Foissac *et al.,* 2002 and Habibi *et al.,* 2002). According to Rahbe *et al.,* (1995), in pea aphid *Acyrthosiphon pisum,* lectins fed in artificial diet were recovered in honeydew without any degradation by gut proteases suggesting that they this insect has only limited capacity to digest ingested protein, however this study provides evidence of protein ingestion and excretion via honeydew by aphids. The reason that lectins were detected in honeydew without any degradation could be due to fact that lectins are proteins which are hard to digest, even for herbivorous Lepidopteran pests, which have a whole array of digestive proteases (Gatehouse *et al.,* 1994).

Tracking experiments through digestive system using green fluorescent protein along with Casein by Habibi *et al.*, (2002) showed that Cotton bug (*Lygus hesperus*) is able to digest ingested protein. Feeding ^{35}S labelled cotton leaf proteins to Silverleaf Whitefly (*Bemesia argentifolii*) and tracking them showed that leaf proteins are not only ingested by the aphid but also digested using gut proteases producing free amino acids that were either excreted via honeydew or used for *de novo* protein synthesis (Salvucci *et al.,* 1998). These evidences of presence of proteases in the aphid gut and their possible role in digestion of ingested proteins opens a new field in aphid nutrition research. Though cathepsin B-like proteases have been cloned from *S. avenae*, there are no reports on proteolytic digestion in the gut of this wheat pest. The establishment of proteolytic digestion in *S. avenae* gut, and study of interaction between *S. avenae* gut proteases and wheat proteinase inhibitors could help in devising control strategies against this pest.

1.4.5 Insecticidal proteases:

Many insecticidal proteases have been identified as potential candidates in development of insect resistant crop plants. However it is necessary to understand the mode of action of these proteases and to find out the targets of these proteases in insect. A cathepsin L-like cysteine proteinase (ScathL) from flesh fly (*Sarcophaga peregrina*) functions in basement membrane (BM) remodelling during insect development. ScathL was recombinantly expressed in *Pichia pastoris,* and when injected into *Lacanobia oleracea* (Lepidoptera) 5th instar larval hemolymph caused extensive melanisation of the injected larvae leading to death (Phillip *et al.,* 2007). Mortality of injected larvae was associated with systemic melanisation, which was a

result of the cysteine proteinase activity of ScathL. Recombinant ScathL was also found to be insecticidal against *H. virescens* and pea aphids (*Acyrthosiphon pisum*) (Li *et al.,* 2008).

Melanisation is a feature of the innate immune response of insects and there is a complex cascade of proteases that controls this mechanism. During melanisation, melanin, a phenolic pigment ubiquitously found in animals, is produced and deposited on invading pathogens and parasites. Melanin also plays an important role in healing of wounds caused by microbial infestation or by accident. Melanisation occurs as a result of conversion of enzymatically inactive prophenoloxidase (ProPO) in the hemolymph to active phenoloxidase (PO; EC 1.14.18.1) (Ashida and Yoshida, 1988). Phenoloxidase is a copper containing oxidoreductase enzyme that catalyzes the oxidation of phenolic compounds to quinones. Quinones after polymerization produce an insoluble pigmented polymer, melanin. Melanin is used in microbial encapsulation or wound healing / plugging of infection site to avoid excess fluid loss from the wound site.

This trigger to melanin formation is the conversion of prophenoloxidase to phenoloxidase, which is carried out by a serine proteinase usually referred to as PPAE (prophenoloxidase activating enzyme) (Satoh *et al.,* 1999; Kanost *et al.,* 2004; Lee *et al.,* 1998). PPAE itself is activated from its zymogen form through a stepwise process involving other serine proteinases, the whole process forming an irreversible activation cascade. The activity of the proteases in this cascade is controlled by specific serine proteinase inhibitors of the serpin family. Serpins are a special class of protease inhibitors containing 350-500 amino acid residues that fold into a conserved structure and employ a unique suicide substrate-like inhibitory mechanism. The presence of an extended reactive site loop near the carboxyl terminus helps serpins to interact with the active site of serine proteases (Kanost *et al.,*2004). When a susceptible proteinase cleaves a specific bond in the reactive site loop (designated P1-P1' bond), the serpin undergoes a dramatic conformational change, in which the exposed reactive site loop sequence inserts into a β-Sheet in the protein (Gettins, 2002). The X-ray crystallography structure of *M. sexta* serpin-1K showed that the reactive loop exists in an extended conformation from residues P3 to P3', ready to bind to target proteinase active site (Li *et al.,* 1999). Serpins irreversibly inhibit the

activating serine proteinases once the encapsulation or wound healing is complete.

Insects contain a gene family encoding serpins, in which alternative splicing generates additional variants. Variants of serpin-1 are constitutively expressed in the fat bodies and other tissues in insects. However, serpins-1b/1c in *M. configurata* were only detected in hemolymph suggesting their possible interaction with hemolymph serine protease and are found to be actively involved in immune related cascades (Chamakhah *et al.,* 2003). A schematic representation of the melanisation cascade in insects is shown in figure 1.1.

The activation of melanisation cascade is triggered as a result of injury or exposure to certain microbial cell wall components, such as peptidoglycan, glucan, and lipopolysaccharide (LPS), through pattern recognition proteins. Pattern recognition proteins containing a protease domain form the first protease of the process and play an important role in initiating the phenoloxidase cascade. A cDNA encoding a modular protein (HP14) has been cloned from *M. sexta* hemolymph containing five low density lipoprotein receptor class 'A' repeats, a Sushi domain, a unique Cys-rich region and a proteinase-catalytic domain. Its expression was found to increase in 24 hours after the insect is infected with bacteria. Its interaction with peptidoglycan resulted in proteolytic processing of its purified zymogen and generation of peptidase activity, suggesting that HP14 is pattern recognition protein that binds to bacteria, auto-activates and triggers the prophenoloxidase system in the hemolymph of *M. sexta* (Ji *et al.,* 2004; Ma and Kanost, 2000).

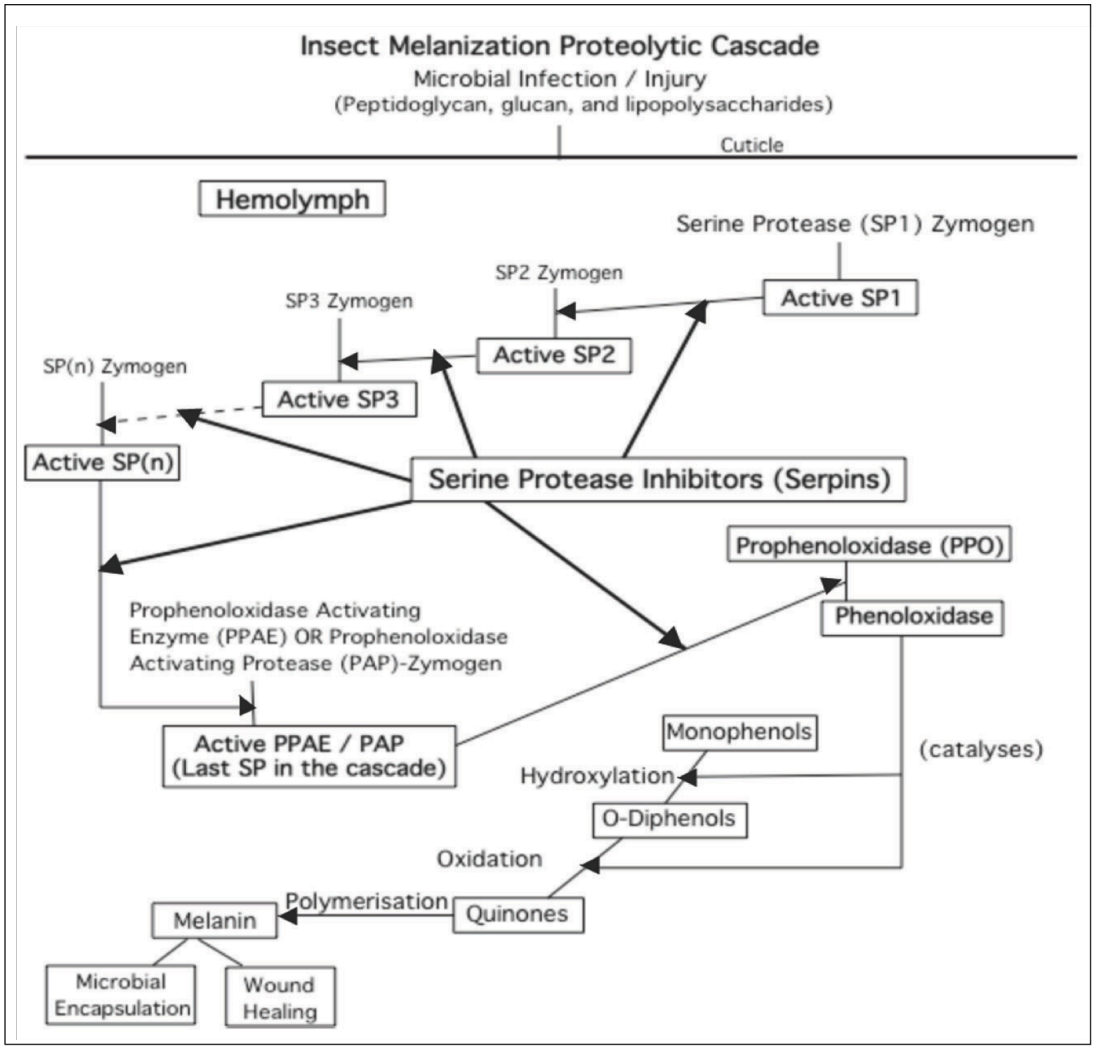


Figure 1.1: Schematic representation of insect melanisation cascade. [composed on the basis of Ashida and Brey, (1998), Hopkins and Kramer, (1992), Gillespie and Kanost, (1992), Jiang *et al.,* (1998), Lee *et al.,* (1998),Yoshida *et al.,* (1996), Ma and Kanost, (2000), Ennio De Gregorio *et al.,* (2002), Christensen *et al.,* (2005), Nappi *et al.,* (2005), Kanost, (1990)].

The initial zymogen is converted to active protease and acts as activator for the next serine protease in the cascade. A serine protease PAP-1 from *M. sexta,* has been shown to cleave ProPO to generate active PO. *M. sexta* contains a single copy of PAP-1, which is constitutively expressed in the fat body, trachea and nerve tissue of fifth instar larvae (Zou *et al.,* 2005).

The proteolytic activation cascade must be tightly regulated to avoid systemic activation of PO causing extensive damaging melanisation. This negative regulation required for homeostasis is carried out by serpins (Janciauskiene, 2001). *Manduca*

sexta is shown to have 12 naturally occurring reactive site variants of serpin-1 genes. All of these variants are cloned and characterised, of which serpin-1J is involved in the melanisation process. Serpin 1J is shown to regulate a prophenoloxidase-activating serine proteinase (PAP-3) in *M. sexta* hemolymph (Jiang *et al.,* 2003). Serpin-1J does not inhibit the actual phenoloxidase directly but instead it inhibits activation of its zymogen (Jiang and Kanost, 1997).

The involvement of serpins in negative regulation of PPO activation in Toll pathway has been well documented in *Drosophila melanogaster* (Levashina *et al.,* 1999). A study of loss-of-function mutation in the gene encoding a blood serine protease inhibitor, Spn43Ac, showed that Spn43Ac negatively regulates the Toll signaling pathway. Serpin 27A (Spn27A) normally functions as a negative regulator of phenoloxidase activation, and mutant larvae lacking Spn27A show a melanotic phenotype and excessive melanisation in response to immune challenge (Nappi *et al.,* 2005). Lepidopteran insects contain a complex family of serpins, with alternate splicing used to generate multiple inhibitors with differences in specificity towards proteinases from single genes (Jiang and Kanost, 1997). In the model species, *Manduca sexta*, different serpins inhibit different proteinases in the prophenoloxidase activation pathway, thereby regulating the response (Zhu *et al.,* 2003; Tong *et al.,* 2005). Serpins thus play an important role throughout the melanisation cascade by inhibiting particular serine proteases at any step, stopping the immune cascade to keep the melanisation process restricted to the wound or infestation site. These serpins in hemolymph could be the possible targets of cathepsin L-like cysteine proteases and loss of function of serpins might lead to uncontrolled and systemic melanisation.

1.4.6 RNAi mediated crop protection strategy:

All eukaryotic organisms possess common and normal machinery for sequence specific gene silencing that is triggered by the presence of double stranded RNA (dsRNA). This process is called RNA interference (RNAi) in animals and post-transcriptional gene silencing in plants (Eamens *et al.,* 2008). RNA interference (RNAi) is normal and conserved phenomenon of controlling gene expression. It involves two types of small RNA molecules, microRNA (miRNA) and small interfering RNA (siRNA), which bind to mRNA molecules that will make specific proteins by translation. Binding of these small RNA molecules can increase or

decrease the activity of RNA molecules thereby controlling the translation and overall gene expression. The process of RNAi can be induced in an organism to inhibit the expression of specific genes, which is called 'Gene Knock Down'. In this process a double stranded RNA is prepared by *in vitro* transcription and is introduced into the organism.

RNA interference (RNAi) was first demonstrated to cause potent and specific genetic interference in *Caenorhabditis elegans.* The effect was seen either by injecting a double stranded RNA or by feeding bacteria expressing RNAi or simply feeding dsRNA directly to the worm (Fire *et al.,* 1998). The overall process of RNAi occurs in following steps, (1) injected dsRNA is cleaved into small interfering RNAs (siRNAs) duplexes (approx. 21 nucleotide long) by Dicer RNaseIII-type enzymes, (2) siRNAs duplexes are incorporated into a multiprotein RNA-inducing silencing complex (RISC), (3) the antisense strand guides RISC to its homologous target mRNA, (4) this leads to endonucleolytic cleavage and degradation of target mRNA thereby silencing this gene. The fruit fly; *Drosophila melanogaster* has been an ideal insect system for RNAi studies, where RNAi is used for loss of function mutations to identify the functions of various genes (Kennerdell and Carthew, 1998, 2000; Misquitta and Paterson, 1999). Stable and functional RNA interference has been seen in Coleopteran insects like Red flour beetle (*Tribolium castaneum*) (Hebert *et al.,* 2008; Lindbo and Dougherty, 2005). Several reports of successful gene knock-down of crucial genes in insects by RNAi are available e.g. *Anopheles gambiae, Drosophila melanogaster, Manduca sexta, Periplanata Americana, Oncopeltus fasciatus* (Blandin *et al.,* 2002; Hughes and Kaufman, 2000; Kennerdell and Carthew, 1998; Marie *et al.,* 2000; St. Johnston, 2002; Vermehren *et al.,* 2001).

Downregulation of the expression of specific genes by RNAi mediated gene knock down in the target insect is an emerging and potential tool for crop protection strategy. However, there are many concerns for this approach, as the ingestion of dsRNA has not been very effective as compared to injecting a dsRNA, and injecting is not practically feasible at field level. The *in planta* expression of dsRNA in the host has been suggested as a solution to this problem and has shown success to some extent. The effects of dsRNA would only be expected in cells exposed to the nucleic acids. Insect midgut and associated structures are thought to be the targets for this, as

they are not covered with chitin exoskeleton. The degradation of dsRNA in the insect midgut is another major concern, suggesting that it might be necessary to feed continuous and high levels of dsRNA to target insects (Price and Gatehouse, 2008). Apart from these considerations, according to some recent reports, viable levels of insect resistance can be achieved by producing sufficient levels of dsRNAs in plants (Baum *et al.,* 2007, Mao *et al.,* 2007).

Many plants have been modified for improvement of quality using RNAi technology for increasing nutritive value, elimination of allergens, to create male sterility, decrease toxic compounds in food crops, etc. (Auer and Frederick, 2009). One of the interesting examples is reduction of terpenoid gossypol in the seeds of cotton. Gossypol plays an important role in cotton as feeding deterrent for herbivorous insects. A cadenine synthase gene in the gossypol synthesis pathway was knocked down using RNAi, the use of a seed specific promoter ensuring that the gene was only silenced in seeds but leaves could synthesize normal amounts of terpenoid for protection against insects (Sunilkumar *et al.,* 2006).

1.5 Aims and objectives of the project:

The aims of this project were to explore the potential endogenous resistance factors in the form of defence proteins in wheat (*Triticum aestivum*) that are expressed either systemically or in response to herbivory by insects. Also to study the digestive biochemistry of two insect pests wheat bulb fly (*Delia coarctata*) and cereal aphid (*Sitobion avenae*) and to discover the potential targets in these insects which can be utilised for devising a suitable crop protections strategy for wheat.

The objectives were:

1. To investigate the digestive biochemistry in *D. coarctata* and *S. avenae* guts; to identify the protease activity types of proteases; to clone protease encoding genes; to express recombinant proteins, to purify and biochemically characterise; similarly with wheat protease inhibitors.

2. To investigate the proteolytic digestion of ingested protein in the gut of phloem feeding insect *S. avenae*.

3. To investigate the basis of insecticidal effect of recombinant cysteine proteinase from *D. coarctata* (DcCathL) to lepidopteran insects. To clone and undertake recombinant expression of serpins (MbSpn-1a and MbSpn-1b/c) from *Mamestra brassicae*, to purify and to check vulnerability of recombinant serpins to DcCathL activity.

4. To assess the potential of endogenous resistance factors in wheat such as protease inhibitors or proteins specifically expressed in response to herbivory like Hessian fly responsive (Hfrs). To clone the respective genes, recombinant expression, purification, characterisation and study their antimetabolic / toxic effects on *D. coarctata* and *S. avenae.*

5. To clone cationic amino acid transporter gene (DcCAAT) from the gut of *D. coarctata*, preparation of double stranded RNA by *in vitro* transcription, injection or feeding dsRNA to *D. coarctata* to knock out DcCAAT.

Chapter 2

Materials and Methods

2.1 Materials:

All chemicals and reagents were supplied by Sigma Chemical Company (St. Louis, USA) or VWR (BDH) Chemical Company (Poole, Dorset, UK) otherwise unless stated. Chemicals and reagents were of analytical grade, or best commercially available.

2.2 Recipes of Commonly used media, buffers and reagents:

(A) Bacterial Culture Media

LB Broth:	1% (w/v) NaCl, 1% (w/v) tryptone (Merck), 0.5% (w/v) yeast extract (Merck), prepared in distilled water.
LB agar:	1.5% Bacto agar (Difco) added to LB broth.
LSLB broth:	0.5% (w/v) NaCl, 1% (w/v) tryptone, 0.5% (w/v) yeast extract, prepared in distilled water.
LSLB agar:	1.5% Bacto agar added to LSLB broth.

(B) Yeast Culture Media:

YPG:	1% (w/v) yeast agar, 2% (w/v) tryptone, 4% (v/v) glycerol prepared in distilled water.
YPG Agar:	1.5% Bacto agar added to YPG media.

Fermentation Media: Basal salt media as per Higgins, D. R. and Creggs J M. (1998)

(D) Agarose Gel Electrophoresis:

TAE (50X):	2M Tris/Acetic acid pH 7.7, 50mM EDTA.
	DNA loading buffer: 10mM Tris/HCl pH 8.0, 10mM EDTA, 30% (w/v) glycerol, 0.1% (v/v) Fast Orange G, prepared in distilled water.

(E) Protein Gel Electrophoresis (SDS-PAGE):

5X SDS sample buffer:	0.5M Tris/HCl (pH 6.8), 50% (v/v) glycerol, 5% (w/v) SDS, 0.005% (w/v) bromophenol blue.
Acrylamide:	30% (w/v) acrylamide: 0.8% (w/v) bis-acrylamide stock solution (37.5:1) (Protogel, National diagnostics).
Resolving buffer:	3.0 M Tris/HCl pH 8.8.
Stacking buffer:	0.5 M Tris/HCl pH 6.8.
Reservoir buffer (10X):	0.25M Tris/HCl pH 8.3, 1.92 M Glycine, 1% (w/v) SDS.
(CBB) Stain:	40% (v/v) Methanol, 7% (v/v) glacial acetic acid, 0.05% (w/v) Coomasie Brilliant Blue (CBB).
Destain:	40% (v/v) Methanol, 7% (v/v) glacial acetic acid.
Stacking gel mixture:	2.5% Protogel (37.5 : 1 acrylamide : bisacrylamide; National Diagnostics), 125mM Tris-HCl (pH 6.8), 0.1% (w/v) SDS, 0.1% (w/v) ammonium persulphate, 0.0075% (v/v) *N, N, N', N'*-tetramethylethylenediamine (TEMED)
Resolving gel mixture:	(12.5% or 15% or 17.5 % (w/v) Protogel, 375mM Tris-HCl (pH 8.8), 0.1% (w/v) SDS, 0.075% (w/v) ammonium persulphate, 0.05% (v/v) *N, N, N', N'*-teretramethylethylenediamine (TEMED)

(F) Silver staining reagents:

Fixative:	40% (v/v) Ethanol, 10% (v/v) glacial acetic acid, 50% distilled water.
Wash solution:	30% Ethanol, 70% distilled water.
Thiosulphate reagent:	0.02% (w/v) sodium thiosulphate in distilled water.
Silver nitrate reagent:	0.2% (w/v) silver nitrate, 0.02% (v/v) formaldehyde (37% solution) in distilled water.
Developer:	3% (w/v) sodium carbonate, 0.05% (v/v) formaldehyde (37% solution), 0.0005% (w/v) sodium thiosulphate in distilled water.
Stop reagent:	0.5% (w/v) Glycine in distilled water.

(G) Protein Molecular weight marker:

SDS7 (Sigma)	66 kDa Bovine albumin
	45 kDa Egg albumin
	36 kDa Glyceraldehyde-3-phosphate
	29 kDa Carbonic anhydrase bovine erythrocytes
	24 kDa PMSF-treated trypsinogen
	20 kDa Soybean trypsin inhibitor
	14 kDa α-lactalbumin

(H) Western blotting:

Bjerrum & Schafer-Neilson buffer:	48 mM TrisHcl, 39 mM Glycine, 20% (v/v) methanol, 0.0375% SDS, pH 9.2
Ponceau stain:	0.1% Ponceau S, 5% acetic acid in distilled water.
PBS (10X):	0.015 M KH_2PO_4, 0.08 M Na_2HPO_4, 1.37 M NaCl in distilled water.
Blocking solution:	5% Non-fat milk powder, 1X PBS, 0.1% Tween-20
Anti-Sera solution:	5% Non-fat milk powder, 1X PBS, 0.1% Tween-20
PBST:	1X PBS, 0.1% Tween-20

(I) Chemiluminescent detection reagents:

Solution A:	100 mM Tris/HCl pH 8.0, 0.2mM coumaric acid, 1,25 mM luminol in 50 ml distilled water.
Solution B:	10% H_2O_2 (30% solution) in distilled water

(J) DNA molecular weight marker: Lambda DNA digested with Eco471 (AvaII)

(K) 2X PCR Mix:	200µl (10X) PCR buffer $(NH4)_2SO_4$(-$MgCl_2$), 40µl dNTPs (10mM each), 120µl $MgCl_2$ (25mM), 20µl BSA (10mg/ml) and nuclease free water to 1ml
(L) Protein extraction buffer:	50mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)} + 0.1% Triton-X100 pH 7.4

2.3 Artificial rearing of insects:

2.3.1 Wheat Bulb Fly (Delia coarctata):

Wheat bulb fly cultures were maintained at insect rearing facilities at FERA (The Food and Environment Research Agency), York. Adult flies were fed with sucrose, milk, aqueous honey solution, yeast extract and meat juice (aqueous extract of fatty beef mince) and maintained at 15°C, 60% (RH) relative humidity, 16:8 LL:DD on moist sand. Eggs were collected from the sand and stored in moist vermiculite at 15°C for 60–70 days under a 16-h photoperiod followed by further 60–70 days at 5°C in constant darkness. Egg hatch was initiated through removing eggs from the 5°C incubator and placing them at 10°C in constant darkness. On hatching, larvae were immediately transferred to potted wheat seedlings (var. Claire) and were allowed to develop on wheat plants until required.

2.3.2 Cereal Aphids (Sitobion avenae F.):

Parthenogenetic females of the grain aphid *S. avenae* (F.) were maintained on oat (*Avena sativa* L. cv. Coastblack) seedlings maintained at 18°C in a long-day regime of 16 h of light and 8 h of dark. For artificial diet feeding trials, 2-day old parthenogenetic apterous aphids were transferred to sterile liquid diet containing vitamins, minerals, 150 mM amino acids and 500 mM sucrose (as described below). Feeding chambers containing 10 aphids were maintained under the same environmental conditions as cultures on plants. Aphids were maintained on artificial liquid diet for 24 hours and day-3 aphids were used for feeding bioassays.

Aphid liquid diet (Kunkel H, 1976; Prosser and Douglas, 1992)

Stock solutions of amino acids, vitamin and minerals were prepared in advance and kept frozen at -20°C until use.

Amino acids stock (150mM):

Alanine	50.8mg
Asparagine	213.9mg
Aspartate/Aspartic acid	189.7mg
Cysteine	42.5mg
Glutamic acid	123.6mg
Glutamine	241.1mg

Glycine	9.0mg
Proline	65.6mg
Serine	59.9mg
Tyrosine	10.9mg
Arginine	300.2mg
Histidine	182.4mg
Isoleucine	114.1mg
Leucine	114.1mg
Lysine	158.9mg
Methionine	42.5mg
Phenylalanine	47.1mg
Threonine	103.6mg
Tryptophan	58.2mg
Valine	101.9mg

Dissolved in 50ml distilled water, made aliquots of 5ml

Minerals stock:

$FeCl_3.6H_2O$	11mg
$CuCl_2.4H_2O$	2mg
$MnCl_2.6H_2O$	4mg
$ZnSO_4$	17mg

Dissolved in 10ml distilled water, made aliquots of 500µl

Vitamins stock:

Biotin	0.1mg
Patnothenate	5mg
Folic acid	2mg
Nicotinic acid	10mg
Pyridoxine	2.5mg
Thiamine	2.5mg
Choline	50mg
Myo-inositol	50mg

Dissolved in 5 ml distilled water, made aliquots of 500µl

On the day of diet preparations, the stocks of amino acids, minerals and vitamins were thawed; 100µl mineral stock, 500µl vitamin stock and 5ml of amino acid stock were mixed together. Sucrose mix was prepared fresh as follows and added to above mixture and dissolved properly:

Ascorbic acid	10mg
Citric acid	1mg
$MgSO_4.7H_2O$	20mg
Sucrose	1.7g (500mM)

Phosphate solution was prepared by dissolving 150 mg $K_2HPO_4.3H_2O$ or 115mg K_2HPO_4 in 1ml sterile distilled water and was added to diet mixture. pH of the mixture was checked to be around 7.0 to confirm that all the components are in right proportion. Distilled water was added to make final volume 10ml and the diet was filter sterilised using a 0.2 µm filter and frozen at -20°C until use.

2.3.3 (Tomato moth) Lacanobia oleracea and (Cabbage moth) Mamestra brassicae:

Artificial rearing of *L. oleracea* and *M. brassicae* was performed as described in Chougule *et al.,* (2008) on a standard diet prepared in-house as described in Bown *et al.,* (1997). Insects were maintained at 25°C, 40% relative humidity and 16h:8h light:dark regime. Larval stage insects were fed continuously with artificial diet which was prepared with following ingredients per litre of diet: 13.3g bacto-agar, 3.57 g ascorbic acid, 1.77 g sorbic acid, 2.93 g methyl-4-hydroxybenzoate, 6.33 g Vitmix (Vanderzant modification vitamin mixture for insects, ICN Biomedicals Ltd, Thame, Oxon., U.K.), 0.123 g ampicillin or 1.00 g aureomycin (Cyanamid Ltd, Gosport, Hants., U.K.), 3.67 ml formaldehyde solution (37%), 6.67 g Wesson salts (salt mix W, ICN Biomedicals Ltd.), 74.0 g haricot bean meal, 59.0 g wheat germ, 30.0 g soyabean meal, 20.0 g casein, 33.3 g yeast, distilled water to 1000 ml. Fifth instar larval stage was identified based on size and moulting stage of larva and were selected for injections.

2.4 Standard Molecular Biological Techniques:

All standard molecular biological techniques were based on protocols from Molecular Cloning: A Laboratory Manual (Sambrook and Russell, 2001) unless otherwise stated.

Commercially available kits were used according to manufacturers instructions unless otherwise stated.

2.4.1 Oligonucleotides:

Oligonucleotides were synthesized by Sigma Genosys Service. Primers were resuspended in sterile distilled water to a concentration of 100 pmol /µl and stored at -20°C. In standard PCR reactions the primers were used at final concentration of 0.2 to 1.0 µM depending on primer types. Melting temperature for the oligos was calculated by the following formula T_m = 69.3+0.41x(%G+C)-650/(nA+nT+nG+nC). The annealing temperature (T_a) of a particular oligonucleotide was generally 3°C below the calculated T_m or modified depending on type of polymerase being used.

2.4.2 Bacterial Culture:

Liquid bacterial cultures were grown in 5 ml Luria-Bertani (LB) or LSLB (Low Salt LB) media using single colonies picked from LB-agar plates. Liquid cultures were grown at 37°C overnight (approx. 16 hours) on a rotary shaker with shaking at 220 rpm. Where appropriate, antibiotics for selection were added to culture media. For LB-agar or LSLB-agar plates, antibiotics were added when the media had cooled to about 50°C. Low salt media were used with Zeocin (Invitrogen) as an antibiotic selection.

2.4.3 Electro-competent Cells:

Electro-competent cells of *E. coli* TOP10 strain were prepared as described in Sambrook and Russell, (2001) and were used throughout for transformation.

2.4.4 Transformation of E. coli:

Transformation of electro-competent cells was performed as per standard protocol (Sambrook and Russell, 2001). 50µl competent cell aliquot was mixed with an appropriate amount of ligation reaction (0.5 to 1µl) and transformed using the following conditions: Biorad Electroporator, electrical pulse set to 25µF, capacitance set to 2.5 kV and 200 ohm resistance. The pulse was delivered to the cells at with above settings and time constant of 4-5 milliseconds with field strength of 12.5

kV/cm. Transformation mix was incubated for 1 hour at 37°C and plated onto the respective media containing antibiotic. The plates were incubated overnight at 37°C.

2.4.5 DNA sequencing:

DNA sequencing reactions were carried out using BigDye Terminator with AmpliTaq DNA polymerase (ABI Biosciences). Reaction products were analysed on automated sequencers (ABI Prism 373 STRETCH and ABI Prism 377 XL), DBS Genomics, Dept. at Durham University, School of Biological and Biomedical Sciences. Expression constructs were completely sequenced on both strands of the DNA by using primers directed against the determined sequence to complete overlaps.

2.4.6 Sequence Analysis:

Contiguous sequences were produced in Sequencher ™ Version 4.5 (Gene Codes Corporation) on a Macintosh computer. Edited nucleotide or predicted amino acid data was used in BLAST similarity searches (Altschul *et al.,* 1990) against the GenBank database (http://www.ncbi.nlm.nih.gov/BLAST/) and identification of sequence features in encoded polypeptides was performed using different tools available on the ExPAsy proteomics server (http://www.expasy.ch/tools/). Some amino acid multiple sequence alignments were performed on BoxShade server (http://www.ch.embnet.org/software/BOX_form.html).

2.4.7 Glycerol Stocks:

Single colonies of recombinant *E. coli* or *P. pastoris* containing DNA plasmids were inoculated onto LB broth or YPG respectively containing appropriate antibiotic selection and grown overnight at 37°C for *E. coli* or 30°C for *P. pastoris.* The culture was centrifuged at 5000 rpm for 10 minutes at RT. 750µl of respective media and 250 µl of 60% sterile Glycerol was added to cell pellet, vortexed briefly and frozen at -80°C.

2.4.8 Preparation of total RNA:

Total RNA from insect tissue (either whole insect or dissected guts) was prepared using Tri Reagent (Sigma Chemical Company, T-9424). Larvae were dissected under a dissecting microscope and flushed with 0.9% saline. Tissue was weighed and added

to Tri Reagent, homogenised and processed as specified by the manufacturer's instructions. RNA was quantified using a Nanodrop (NanoDrop ND-1000 Spectrophotometer; Bosch Institute) and integrity was checked by denaturing formaldehyde / agarose gel electrophoresis. RNA samples were snap frozen and stored at -80°C.

2.4.9 Formaldehyde Gel Electrophoresis of RNA:

RNA samples were analysed by denaturing formaldehyde agarose gel electrophoresis, using a modified protocol (Fourney *et al.,* 1988). Gels contained agarose up to concentration of 1.5% (w/v) in 1X MOPS/EDTA (20mM MOPS, 5mM sodium acetate, 1mM EDTA, pH 7.0), and formaldehyde at a final concentration of 2% (v/v). Gels were allowed to polymerise for 1 hour after pouring and RNA samples were quantified (5-10μg) for loading. Diluted RNA samples in 5μl were mixed with 25μl loading buffer (0.75 ml deionised formamide, 0.15 ml 10X MOPS, 0.24 ml formaldehyde, 0.1ml deionised RNAse free water, 0.1ml glycerol, 0.08 ml 10% (w/v) bromophenol blue) and denatured by incubating at 65°C for 15 minutes. 1μl of Ethidium bromide was added (10mg/ml) was added to each sample. Denaturing gels were submerged in 1X MOPS/EDTA and samples were loaded and run at room temperature at 50-100V. An RNA size marker (Promega) was run alongside RNA samples. After electrophoresis RNA was visualised and imaged on a transilluminator without further staining or destaining.

2.4.10 DNA amplification by Polymerase Chain Reaction (PCR):

PCR reactions (25 to 50μl) were set up on ice in thin walled, 0.2ml PCR tubes. A typical reaction of 50μl consisted of 0.2 mM each of dATP, dTTP, dGTP and dCTP in 1X PCR mix, DNA template (5-100ng) and 1.25 units of *Taq* polymerase. For multiple PCR reactions a master mix containing all common components was prepared and dispensed into individual tubes. PCR thermo-cycling was determined empirically for each primer combination and type of polymerase being used. PCR reactions were performed on Perkin Elmer 2400 thermal cycler. Where high fidelity PCR was required, Advantage-2 polymerase mix (Clontech), proof reading polymerase Phusion (Fermentas) or KOD polymerase (Novagen) were used. High

fidelity PCRs were set up by using buffers supplied by manufacturers and following suggested protocols.

2.4.11 Reverse Transcription-PCR (RT-PCR):

In a standard procedure total RNA was reverse transcribed into first strand cDNA using Moloney Murine Leukemia Virus Reverse Transcriptase (M-MLV RT, Promega). In a typical RT-PCR reaction, 2μg of total RNA was mixed with 0.5μg of Poly-$T_{(24)}$ primer in a 15μl volume. RNA secondary structure was melted by heating to 70°C for 5 minutes, and then placed on ice to prevent secondary structure reforming. The reaction was completed on ice by adding, 5μl of 5X reaction buffer (250 mM Tris-HCl, pH 8.3, 375 mM KCl, 15 mM $MgCl_2$, 50mM DTT); dNTPs (final concentrations, 0.5mM) and 200 Units of M-MLV RT (where 1 U catalyzes the incorporation of 1nmol of dNTP into an acid insoluble form in 10 minutes at 37°C), plus additional nuclease free water up to a volume of 25μl. Reaction tubes were incubated at 42°C for 60 minutes for production of first strand cDNA. Reaction products were diluted 10-fold, snap frozen and stored at -80°C. 5μl aliquots were used as template in a standard 50μl PCR reaction.

2.4.12 Rapid Amplification of cDNA Ends (RACE) PCR:

RACE experiments were performed on 1μg total RNA using SMART RACE cDNA amplification kit (Clontech), according to manufacturers instructions. RACE reactions were used to retrieve the complete 5' and 3' end of partial cDNA sequences including any untranslated region (UTR). Gene specific primers (GSP) from partial cDNA sequences for a particular gene were designed; a sense primer was used in 3' RACE experiments and antisense primer in 5' RACE experiments. cDNA synthesis was primed by 3'-RACE CDS primer (5'- AAGCAGTGGTATCAACGCAGAGTAC $(T)_{30}$ N_{-1}N-3'; N=A, T, G OR C; N_{-1}= A, G or C; universal sequence is underlined). This primer anneals to the poly-A^+ tail of mRNA. First strand cDNA was produced using PowerScript reverse transcriptase (Clontech). After reverse transcription RACE ready cDNA was amplified by PCR between the 3' RACE gene specific primer (GSP) and the universal sequence incorporated by the 3' RACE CDS primer. Amplification between these two primers ensured that the 3' end of the gene was fully represented, including and 3' UTR. For 5' RACE, cDNA synthesis was primed by 5' RACE CDS

primer (5'-$T_{(25)}$ N_{-1}N-3'; N=A, T, G or C and N_{-1}= A, G or C). This primer anneals to the poly-A^+ tail of mRNA. First strand cDNA was produced using PowerScript reverse transcriptase (Clontech), which adds 3 to 5 additional dC residues to the 3' end of the nascent first strand cDNA. Subsequently, a stretch of complementary dG in the SMART IIA primer base pairs and serves as an extended template for PowerScript reverse transcriptase. After reverse transcription first strand cDNA was amplified by PCR between the 5' RACE GSP and the universal sequence incorporated by cDNA extension across the SMART IIA oligonucleotide. Amplification between these two primers ensured that 5' end of the gene was fully represented, including any 5' UTR. 3' and 5' RACE products were analysed by gel electrophoresis and products of the predicted size were purified from gel, cloned and sequenced using standard protocols. When RACE experiments failed to generate specific product, a nested PCR approach was used. Initial RACE amplification reaction products were diluted and reamplified using a pair of internal gene specific primers. (Based on PhD Thesis by Dr. Daniel Price, Durham university, 2004).

2.4.13 Sub-cloning of PCR products:

PCR products (amplified with proofreading polymerases) were purified from agarose gels using QiAquick Gel Extraction Kit, Qiagen) and cloned into pJET 1.2 cloning vector (Fermentas) according to manufacturers instructions. The PCR products amplified with *Taq* Polymerase (products with sticky ends) were blunted using blunting enzyme provided with kit, before ligation into pJET 1.2 vector. Ligation reactions were then transformed into TOP10 *E. coli* electro-competent cells. Transformants were plated on LB-carbenicillin (50μg/ml). Several independent recombinant clones were screened by colony PCR with appropriate primer sets using standard PCR components and *Taq* polymerase, by extending the initial denaturation step to 10 minutes. PCR positive clones were picked and grown overnight at 37°C in 5ml LB broth containing 50μg/ml carbenicillin. Plasmid DNA from positive transformants (see below) and checked by restriction digestion with XhoI and XbaI restriction enzymes for release of insert. Plasmid DNA was sequenced using pJET (M13 forward and reverse) sequencing primers. Sequences were analysed using Sequencher software (Gene Codes Corp.) and sequence comparison was done using BLAST (www.ncbi.nlm.nih.gov/blast) server.

2.4.14 Isolation of Plasmid DNA:

Clones were grown overnight in 5ml LB or LSLB media containing appropriate antibiotics at 37°C, centrifuged, and plasmid DNA was isolated from cell pellet by using the Wizard *Plus* SV Miniprep DNA Purification system (Promega) as per manufacturers instructions. Plasmid DNA was eluted in 100µl of elution buffer and stored at -20°C.

2.4.15 Restriction Endonuclease Digestion of DNA:

Restriction enzyme digestion was carried out using commercially available restriction endonuclease enzymes (Fermentas, Promega, NEB or Roche) with suggested buffers. For double digestions buffer compatibility was followed as suggested by manufacturer. Typically analytical digests were carried out in a total volume of 20µl containing 5µl plasmid DNA and sub-cloning digestion reactions were 50µl containing 25µl of plasmid DNA. The digestions were incubated at 37°C for 1 to 3 hours or sometimes overnight for complete digestion.

2.4.16 Agarose Gel Electrophoresis:

DNA was separated in gels containing Agarose (Melford), typically 1% (w/v) in 1X TAE buffer containing 0.5µg/ml Ethidium Bromide to allow visualization under UV transilluminator (UVB, 300 λ nm). DNA samples were prepared before loading to contain 1X DNA sample loading buffer. Gels were run at 50-100 V in 1X TAE buffer. Eco471 (PvuII) digested λDNA was used as a molecular weight marker. Gels were photographed under UV light on a Gene Flash (Syngene) Bio Imager.

2.4.17 Purification of DNA from Agarose gel:

Gels were visualized on UV Transilluminator and bands of interest were chopped from the agarose gel using sterile blades, agarose blocks containing DNA were weighed and DNA was purified from agarose gels using a QiAquick gel extraction kit (Qiagen) according to the manufacturers instructions.

2.4.18 DNA Ligation:

After restriction digestion, DNA from ligation (insert and vector) were separated by agarose gel electrophoresis and purified from gel. DNA fragments with compatible

ends were ligated in a 10µl reaction volume containing commercially available T4 DNA Ligase and Ligase buffer (Promega). For a typical insert and vector ligation, vector and insert DNA were added in approximate 1:3 and 1:1 ratios and reaction made up to 10µl with sterile distilled water. Ligation reactions were generally incubated at 10°C overnight or 22°C for minimum of 3 hours, unless otherwise specified.

2.4.19 SDS-PAGE Electrophoresis:

Protein samples along with molecular weight marker (SDS7; see recipe) were denatured and reduced before loading by diluting in 1X sample loading buffer and boiling for 10 minutes. Denatured proteins were separated according to their size by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) according to Laemmli *et al.,* (1970). Minigels (9x10 cm) were run in 1X reservoir buffer at constant voltage (100-150 V) in ATTO-AE450 apparatus. Following electrophoresis, gels were either stained or transferred to nitrocellulose membrane for western blots.

2.4.20 Staining and destaining SDS-PAGE gels with Coomassie Blue:

After SDS-PAGE, proteins in the gel were visualized by staining with CBB stain for minimum 3 hours, followed by destaining with destain until the background was clear. Both staining and destaining were carried out at room temperature with gentle agitation.

2.4.21 Silver staining of SDS-PAGE gels:

Silver staining of SDS-PAGE gels was carried out using a modified protocol based on an original protocol by Blum *et al.,* (1987). All the reagent solutions (see recipe) used were 50 ml/gel unless otherwise specified. The gel was incubated in fixative for 1 hour followed by wash with wash solution for 40 minutes with one change in between. Gel was then washed with distilled water for 20 minutes followed by one-minute incubation in sensitizer. The gel was washed with distilled water three times (20 seconds each). Gel was incubated in silver nitrate reagent for 20 minutes and 2 washes of distilled water were given (20 seconds each). The gel was then incubated in developer solution until the expected bands were seen (3 to 10 minutes, avoiding the development of background) followed by two washes of 100 ml distilled water. The

gel was incubated in stop solution for 5 minutes and washed two times with distilled water.

2.4.22 Western blotting:

Proteins were transferred to nitrocellulose (Hybond ECL, Amersham) membranes followed by electrophoresis by electro-blotting, using a standard semi-dry transfer method. Gels to be blotted were equilibrated in Bjerrum and Schafer-Neilson buffer (see recipe) by soaking for 30 minutes at room temperature. Nitrocellulose membrane and 3MM blotting papers (Whatman) were cut to the same dimension as gels and pre-wet in same buffer. The blot was set up on an ATTO AE-6675 blotting apparatus as shown in Figure 2.1. Electroblotting was conducted at constant current set to 125-150 mA (2.0 mA/cm^2) for 60 minutes. Efficiency of transfer was checked by staining the membrane with Ponceau S stain, molecular weight marker bands were marked with pencil and the membrane was destained with distilled water.

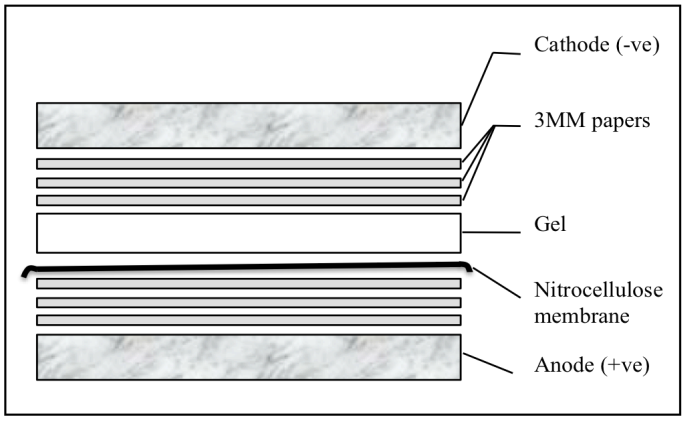


Figure 2.1: Standard setup for western blot to transfer proteins from acrylamide gels to nitrocellulose membranes

2.4.23 Chemiluminescent detection of membranes:

Non-specific protein binding sites on the membrane were blocked by incubating in 50ml blocking solution for 1 hour at room temperature with three changes and gentle agitation. After blocking membranes were reacted with the appropriate dilution (generally 1:3000 unless otherwise specified) of primary antibodies (Anti His or Anti Myc; Cell Signalling) for 2 hours at room temperature or overnight at 4°C with gentle agitation. Unbound primary antibody was removed by washing with antisera for 30 minutes with three changes at room temperature. The membrane was transferred to

antisera containing secondary antibody (Goat Anti Mouse IgG (H+L)-HRP conjugate) in appropriate dilution (generally 1:3000 unless otherwise specified) and incubated for 1-2 hours at room temperature with gentle agitation. The membrane was washed with 1X PBST for 30 minutes with three changes to remove unbound secondary antibody. Excess Tween 20 was removed by washing the membrane briefly with distilled water. Enhanced Chemiluminescence (ECL) reagents (Amersham) were used to detect specifically bound secondary antibodies. Solution A and B were prepared as described in recipe. Solution A (5ml) was mixed with solution B (15µl) freshly before detection. Specific antibody binding was visualised by exposing membranes to photosensitive film (Fuji-RX). Exposed films were washed and developed with an automatic developer (X-ograph Imaging Systems Compact X4).

2.4.24 Ethanol Precipitation of DNA:

$1/10^{th}$ Volume of 5M Ammonium acetate and double volume of absolute ethanol was added to the DNA solution for precipitation and incubated overnight at -20°C. Solution was centrifuged at 14,000 rpm for 15 minutes at 4°C. The pellet was washed with 70% ethanol twice. The DNA pellet was air dried and re-suspended in appropriate amount of sterile distilled water, sometimes dissolution was done by incubating DNA at 37°C for 15-30 minutes.

2.5 Transformation of P. pastoris with expression construct plasmids:

Plasmid DNA was prepared from clones verified for correct sequence by mini-prep kit. Plasmid DNAs from 3x 5 ml cultures were pooled and linearized with BlnI restriction enzyme (Roche) overnight, following standard restriction digestion protocol. Plasmid was checked on agarose gel with undigested plasmid for confirmation of linearisation. Digested DNA was precipitated by following ethanol precipitation protocol (see above), resuspended in 10-15 µl sterile distilled water and transformed into competent cells of protease deficient *P. pastoris* (SMD1168; Invitrogen) as per manufacturers instructions. The transformants were selected on (25 µg/ml) Zeocin in YPG-agar plates by growing for 2-3 days. A control plasmid pGAPZαB without insert was always included as a control during transformation.

2.6 Screening of transformants:

Putative transformant clones were grown in McCartney bottles containing liquid YPG (10ml) and Zeocin (25 μg/ml) for 2-3 days at 30°C with shaking (230 rpm). Cultures were centrifuged at 5,000 rpm at room temperature and 30μl supernatant was used for SDS-PAGE followed by western blotting. A supernatant from a control (pGAPZαB only) was always included during screening as an expression background.

2.7 Over-expression of recombinant proteins in P. pastoris: (Fermentation):

Selected clone expressing a recombinant protein was grown on fresh YPG-agar plate with Zeocin (25 μg/ml) and three 100ml YPG liquid cultures (in 250 ml baffled flasks) were inoculated from master plate to initiate the starter culture. The culture was grown for 2 to 3 days at 30°C with shaking (230 rpm). A 5 litre, laboratory fermentor (BioFlo 110; New Brunswick Scientific) was used to grow the selected *P. pastoris* clone and produce recombinant protein, the methodology was similar to Fitches *et al.,* (2004). 3 litres of basal salt media was prepared (Higgins and Cregg, 1998), and sterilised by autoclaving. Sterile minimal media was supplemented with PTM1 salts (Cino, 1999). Fermentation was carried out at 30°C with a constant dissolved oxygen level of 30% at pH 4 to 5 and continuous agitation. When cell density reached a level where 30% dissolved oxygen could not be sustained by agitation alone, a constant feed of 50% glycerol (v/v with distilled water) was initiated (5 to 10ml / hour). When the glycerol feed was exhausted the run was continued until the dissolved oxygen level began to rise (3-4 days after inoculation of starter culture), indicating growth had stopped and the run was terminated.

2.8 Processing of fermentor supernatant:

Culture from fermentor was centrifuged to separate the *P. pastoris* cells at 8000 rpm, 4°C for 30 minutes. The supernatant was filtered through 2.7 μM, followed by 0.7μM glass microfiber filters (Whatman, UK) using a vacuum manifold and then diluted with desired buffer for loading onto purification columns. An aliquot of fermentor supernatant was always saved prior to dilution as load control for analysis on SDS-PAGE gels. Purification of proteins was performed by using column chromatography either by ion-exchange (S-sepharose) or metal affinity (Ni-NTA) depending on type

of protein and tags therewith. Purification of each protein is described in detail elsewhere.

2.9 Dialysis of protein samples:

Eluted and pooled fractions of proteins (S-sepharose or Ni-NTA) were dialysed against distilled water (unless and otherwise specified) with a pinch of 50mM ammonium hydrogen bicarbonate. Dialysis was carried out at 4°C overnight with constant stirring and three changes. Dialysis tubing was prepared by boiling in a solution of 2% (w/v) sodium bicarbonate and 1mM EDTA (pH 8.0) for about 20 minutes. Dialysed protein samples were frozen on the walls of freeze-drying flasks by shelling in liquid Nitrogen and lyophilized on a vacuum freeze-dryer overnight.

2.10 Total protein extraction from insect guts and estimation:

S. avenae or *D. coarctata* guts were dissected under dissecting microscope in 0.9% saline and were washed with protein extraction buffer. Guts were then homogenised in extraction buffer (10µl of extraction buffer per mg tissue) with a micropestle. The homogenate was incubated at 4°C for 30 minutes and then centrifuged at 14,000 rpm at 4°C for 10 minutes. Total protein in the supernatant was quantified using BCA protein quantification kit (Pierce) and BSA (Bovine serum albumin; 1.2-10µg/ml) as a standard. Concentrations of unknown protein were predicted using the standard curve. BCA reagent was prepared by mixing Solution B with A (1:50) freshly before use. In microtitre plates 10 µl of each standard or unknown sample was added to separate wells (in duplicate) and then mixed with 200µl of BCA Reagent. The plate was incubated for 30 to 40 minutes at 37°C. Absorbance was then read measured at 562 nm using VERA max microplate reader (Molecular Devices). For some protein samples, protein estimation was done by stained gel. Protein samples were loaded onto SDS-PAGE gel in different dilutions along with standard BSA dilutions. Gels were stained as per standard staining procedure and concentration of unknown protein sample was determined from BSA standard protein bands.

2.11 Determination of pH optima for gut proteolytic activity:

Total protein extracted from dissected guts of *S. avenae* (~25µg) or *D. coarctata* (~10µg) were used for determination of optimal pH for the activity of gut proteases. A

standard substrate (EnzChek; Molecular Probes) assay consisted of 1μg fluorescein-labelled casein in a total volume of 200μl assay buffer. Assay buffer generally used was EnzChek digestion buffer (10mM Tris, pH 7.8) unless otherwise specified. Assays were carried out in microtitre plates, and fluorescence was monitored continuously using a Fluoroskan Ascent (Labsystems) fluorimeter (485nm excitation, 538nm emission), for at least 30 min. The average reaction rate for each assay was measured; all assays were the mean of three replicates unless otherwise specified. For determination of optimal pH for insect gut proteases, substrate EnzChek and AMT (Acetate, MES and Tris) buffer system were used to detect the gut proteolytic activity either in presence or absence of (1mM) glutathione as a reducing agent. Gut proteolytic activity in terms of relative fluorescence units per minute were measured. An additional buffer system containing four buffers with different buffering range [Citrate phosphate buffer (pH 4 to 8), MES (2-(*N*-morpholino) ethanesulfonic acid) (pH 5.5 to 6.5), Bis-Tris propane (pH 6.5 to 9.5) and CAPS (*N*-cyclohexyl-3-aminopropanesulfonic acid) (pH 9.5 to 11.0)] and EnzChek as substrate was used for *D. coarctata* gut assay.

2.12 Cloning of D. coarctata serine proteinase (DcSP):

Degenerate PCR:

The amino acid and nucleic acids sequences for the serine proteases of Dipteran insects, Fruit fly (*Drosophila melanogaster*), Hessian fly (*Mayetiola distructor*) and House fly (*Musca domestica*) were aligned using ClustalW server (http://www.ebi.ac.uk/Tools/clustalw2/) and highly conserved regions (N-terminal HVCGGSII and C-terminal CNGDSGG) with minimum degeneracy were located for degenerate primers. The following degenerate primers were designed 5' CAYTTHTGYGGWGGTTCDATWAT (forward primer for both trypsin like and chymotrypsin like protease), 5' WCCNCCRCTRTCBCCWTGRCA (reverse primer for chymotrypsin like protease) and 5' WCCNCCRGARTCBCCWTGRCA (reverse primer for trypsin like protease). Degenerate PCR was performed using cDNA prepared from total RNA as template and standard PCR reaction components with *Taq* polymerase by following a thermal cycle: Step-I: 94°C for 2 minutes, Step-II: 94°C for 30 seconds, 45°C for 30 seconds, 72°C for 45 seconds (25 cycles) and Step-

III: 72°C for 5 minutes. PCR products (~500bp) were subcloned and sequenced by following standard methods.

RACE (3' and 5') for DcSP:

RACE was done to obtain full-length coding sequence of a chymotrypsin-like serine protease (DcSP). Following primers were designed 5' GSP: 5' CAG GCA CCC GCG CCT GAA GGA TG and 3' GSP 5' GGC GAG CGT TAT GTC TTA ACC GCT GCT. 3' and 5' RACE were performed as described in standard RACE procedures.

UTR PCR:

Following UTR primers were designed to obtain a full coding DNA sequence of DcSP 5' CGG GGA AAT TCA TTT TTA TAC G (forward) and 5' CTA TCA GTG TAT CAA AGA GTG (reverse). PCR was performed using WBF gut cDNA as template and KOD polymerase (Novagen) as per manufacturers instructions and following thermal cycle; Step-I: 95°C for 2 minutes, Step-II: 94°C for 30 seconds, 50°C for 30 seconds, 72°C for 1 minute (25 cycles) and Step-III: 72°C for 10 minutes. The PCR product (approx. 800bp) was subcloned and sequenced as described elsewhere.

2.13 Sequence analysis of DcSP

Predicted DcSP protein sequence from the cloned cDNA sequence was compared against the global database using BlastP software running on the NCBI Blast server (www.ncbi.nlm.nih.gov/blast) in pairwise comparisons to identify sequences similar to DcSP. Multiple sequence alignments were carried out using the Clustal algorithm running under ClustalW software (Thompson *et al.,* 1997). Prediction of sequence features in DcSP was performed using SignalP (signal peptide prediction; www.cbs.dtu.dk/services/SignalP) and comparison to the InterPro domain database using InterProScan (domains and signatures; http://www.ebi.ac.uk/Tools/ InterProScan) supplemented by manual inspection.

2.14 Recombinant expression DcSP:

2.14.1 Expression construct for DcSP in pET32a

Following primers were designed for cloning the mature peptide of DcSP in pET32a *E. coli* expression vector (Novagen) to produce fusion of DcSP:Thioredoxin using

BamHI and HindIII restriction enzyme sites (in bold) 5' C **GGATCC** AGT CCA ACT ACTG (forward) and 5' TA **AAGCTT** AAT CAT ATT CTC TCT AAT CC (reverse). Phusion polymerase and gut cDNA were used for the PCR. The reaction was carried out as per manufacturers instructions, following a thermal cycle; Step-I: 98°C for 30 seconds, Step-II: 98°C for 10 seconds, 65°C for 30 seconds, 72°C for 45 seconds (25 cycles) and Step-III: 72°C for 7 minutes. The PCR product (approx. 700bp) was subcloned and sequenced as described elsewhere. Plasmid containing DcSP with BamHI and HindIII sites was digested with respective enzymes and ligated into vector pET32a digested with same enzymes using T4 DNA ligase (Promega). The ligation reaction was introduced into *E. coli* (TOPO) electrocompetent cells. Transformants were selected on LB-agar plates containing (50μg/ml) carbenicillin and colonies were screened for presence of DcSP coding sequence by colony PCR with cloning primers. Plasmid DNA was prepared from positive clones and was sequenced with T7 promoter and terminator primers. Sequence confirmed clone was used for induction using IPTG for DcSP:Thioredoxin (Trx) expression.

2.14.2 Recombinant expression and purification of DcSP:Thioredoxin fusion:

Plasmid was introduced into chemical competent cells of *E. coli* (BL21DE Origami; expression strain) using standard chemical transformation method and clones were screened by PCR as described earlier. Positive clone was grown overnight in 500ml LB in 1litre baffled flask at 37°C with shaking (250rpm). 0.1mM IPTG was added to the cells when the OD_{600} reached 0.42 and an aliquot of un-induced sample was taken before addition of IPTG. The cells were allowed to grow for 18 hours post induction and the culture was centrifuged at 8,000 rpm at 4°C for 20 minutes. 50mM Tris buffer containing 0.5M NaCl (pH 8) was added to the cell pellets (both induced and uninduced) and were sonicated at 26 microns for 5 seconds followed by chilling on ice for 2-3 minutes and repeating the cycle for 10 times. The lysate was centrifuged at 12,000 rpm for 10 minutes at 4°C. Supernatant was treated as soluble fraction and to the pellet 6M Urea was added. Pellet was vortexed thoroughly and re-suspended in 6M Urea and treated as insoluble fraction.

2.14.3 Confirmation of DcSP: Thioredoxin expression:

Expression of DcSP was confirmed by analysing the protein samples by CBB stain and western blotting followed by SDS-PAGE. 3µl of 5X SDS sample loading buffer (see recipe) was added to 15µl of each fraction Un-induced Soluble (US), Induced Soluble (IS), Un-induced Insoluble (UI) and Induced Insoluble (II), boiled for 10 minutes and were loaded onto SDS-PAGE (15% acrylamide) and the gel was stained with CBB as described elsewhere. The same amounts of samples were loaded onto another gel for western blot and transferred to nitrocellulose membrane (Hybond ECL, Amersham) and probed with primary Anti $(His)_6$ antibody (Cell signalling) followed by secondary antibody (Goat Anti Mouse IgG (H+L)-HRP conjugate) with. Western blot and Chemiluminescent detection of immunoreactive bands was performed as per standard western blotting procedures described elsewhere.

2.14.4 Purification of DcSP: Thioredoxin by Ni-NTA column chromatography:

5 ml Ni-NTA column (Amersham Biosciences) was used for purification of DcSP: Thioredoxin fusion protein. Induced insoluble fraction was equilibrated in a binding buffer (BB) to a final concentration of 20mM Tris-C, 500mM NaCl and 5mM DTT (pH8.0) and loaded onto Ni column pre-equilibrated with BB at the rate of 3ml/min at room temperature. Column was washed with 10-fold column volume of BB until the OD came to a baseline on FPLC recorder and then with BB containing 20mM Imidazole followed by elution with BB containing 200mM Imidazole. All fractions Load, Wash, 20mM Imidazole wash and elution were checked by SDS-PAGE followed by CBB stain as described earlier. Eluted DcSP:Trx in 200mM Imidazole fraction was dialysed against 20mM Tris pH 7.8 as described elsewhere and buffer exchanged to the final concentration of 1mg/ml with same buffer.

2.15 Characterisation of purified DcSP

2.15.1 Enterokinase digestion of DcSP:Trx

Enterokinase (Roche) stock (0.3mg/ml) was prepared in distilled water and pilot experiments were carried out with 25µg (25ul of mg/ml) DcSP:Trx with 0.6 µl enterokinase (2µg with 0.3mg/ml). Reaction aliquots (5µg) were taken at different time intervals 1, 3, 6, 8 and 24 hours and reaction in aliquote was stopped by adding 5µl of 2X SDS-PAGE loading buffer, samples were boiled for 10 minutes and stored

at -20°C until gel run. For the detection of self-degradation / autocatalysis, a control reaction without enterokinase was also incubated along with samples. All the reaction aliquots were analyzed by SDS-PAGE followed by CBB staining. The maximum cleaved product mature DcSP was obtained 8 hours time point. Based on the pilot experiment, major enterokinase digestion experiment was setup with 2mg of DcSP:Trx and digested product was purified by standard Ni-NTA column chromatography. In pET32a enterokinase cleavage site is between S-tag and the recombinant protein, accordingly Trx had a (His)6 tag and mature DcSP did not have any tags for purification as N-terminal (His)6 was not present in the construct. Enterokinase digested DcSP:Trx mixture was loaded onto Ni-NTA column. Thioredoxin bound to Ni-column matrix and mature DcSP was collected in the flow through. Eluted DcSP in binding buffer was dialyzed against 20mM Tris buffer pH 7.8 and concentrated using centrifugal concentrator (30 kD MWCO). Concentrated DcSP was stored at -20°C until further use.

2.15.2 Refolding of DcSP by dilution method:

As purified DcSP did not show any proteolytic activity with EnzChek substrate, *in vitro* refolding of DcSP done by dilution method. In dilution based refolding method, protein refolding is initiated by reduction of target protein in denaturing conditions and oxidative protein refolding is done by altering the redox environment to enable disulphide bond formation. 500µg (mg/ml) purified DcSP was prepared for denaturant conditions by adding 5mM DTT. Refolding buffer / Oxidation buffer consisted of 100mM Tris-Cl (pH 8.0), 0.78mM glutathione disulphide / oxidised (GSSG) and 7.8mM reduced glutathione (GSH). 500µl of DcSP was added to 50ml oxidation buffer drop-wise (50µl) at 4°C at 30 minutes interval, with continuous stirring to avoid aggregation. Refolded DcSP was recovered from oxidation buffer by dialysis against 20mM Tris-Cl (pH 7.8) and concentrated using centrifugal concentrator. Refolded DcSP was analysed by SDS-PAGE followed by CBB staining.

2.15.3 DcSP activity assay:

Activity of refolded and concentrated DcSP (0.5µg to 100µg) was checked using EnzChek as substrate and assays were performed as per standard EnzChek assay

described elsewhere. Bovine trypsin (50ng to 1μg) was used as a positive control for the assays.

2.16 Cloning and sequence analysis of D. coarctata cathepsin L (DcCathL):

2.16.1 Cloning of DcCathL

Amino acid sequences of Cathepsin L-like cysteine proteases from Dipteran insects, Flesh fly (*Sarcophaga peregrina*), fruit fly (*D. melanogaster*), Cabbage root fly (*Delia radicum* L.) were aligned using ClustalW server and highly conserved regions were located. Degenerate primers were designed for N-terminal region GS**C**WAF (with C as active site residue belonging to catalytic triad) and C-terminal DYWLVK**N** (with N as active site residue belonging to catalytic triad). The following degenerate primers, (5' GGH TCH TGY TGG KCH TTY) and (5' YTT VAY BAR CCA RTA RTC) were synthesized using Sigma Genosys service. cDNA prepared from 3rd instar WBF gut (as described earlier) was used for amplification of partial sequence of DcCathL using Phusion polymerase (New England Biolabs) as per manufacturers instructions and following thermal cycle: Step I: 94°C for 2 minutes, step II: 94°C for 30 seconds, 40°C for 30 seconds, 72°C for 60 seconds (25 cycles) and step III: 72°C for 5 minutes. The resulting PCR product (approx. 500bp) was subcloned and sequenced as described elsewhere. From the sequence information, following gene specific primers were designed and synthesized for 3' and 5' RACE, 3' RACE (5' -GTC GCC GTG GCC ATT GAT GCC TCC AAC G) and 5' RACE (5' -CCA ACA CAC CAT GAT CCA AGT TAT CGG AAC). RACE reactions were performed using SMART^{TM} RACE cDNA amplification Kit (Clontech) as described earlier. Both RACE products (approx 1000bp) were subcloned and sequenced as described earlier. 3' and 5' sequences were aligned in Sequencher project to obtain a complete cDNA sequence and based on sequence information, following UTR primers were designed, Forward: 5' GAT CAA ATT GCA TTT ACT ATC TTC and reverse: 5' C TAA ATG AAC TAG TTA GCA CAA CTC. PCR amplification of DcCathL complete coding sequence was done from gut cDNA using KOD polymerase and reaction carried out as described for previous PCR except from the annealing temperature was 52°C. PCR product subcloned and sequenced as described earlier.

2.16.2 Sequence analysis of DcCathL:

Predicted DcCathL protein sequence from the cloned cDNA sequence was compared against the global database using BlastP software running on the NCBI Blast server (www.ncbi.nlm.nih.gov/blast) in pairwise comparisons to identify sequences similar to DcCathL. Multiple sequence alignments were carried out using the Clustal algorithm running under ClustalX software (Thompson *et al.,* 1997). Phylogenetic trees were constructed using the Neighbour-Joining method, and were verified by bootstrap trials (n = 1000). Trees were plotted using njplot software (http://pbil.univ-lyon1.fr/software/njplot.html). Prediction of sequence features in DcCathL was performed at SignalP (signal peptide prediction; www.cbs.dtu.dk/services/SignalP) and comparison to the InterPro domain database using InterProScan (domains and signatures; http://www.ebi.ac.uk/Tools/InterProScan) supplemented by manual inspection.

2.17 Expression construct, recombinant expression and purification of DcCathL

2.17.1 Expression construct for DcCathL:

From the sequence information of complete coding sequence of DcCathL, following cloning primers were designed by introducing PstI and NotI restriction sites (in bold) for cloning into expression vector pGAPZαB, forward: 5'A**CT GCA G**CA ATT TCA ATT ACA GAT and reverse 5' GCT G**GC GGC CGC** AAC AGT GGG AAA AC. Complete coding sequence with cloning restriction sites was amplified from gut cDNA by PCR using above primers and KOD polymerase as described earlier with annealing temperature 50°C and final extension for 7 minutes. The PCR product was digested with PstI and NotI enzymes, purified from gel and was ligated using T4 DNA ligase into pGAPZαB digested with same enzymes. The plasmid containing expression cassette was introduced into electro-competent cells of *E. coli* (TOP10) and selection of transformant was done using LSLB-agar plates containing 25 μg/ml Zeocin. Transformant colonies were screened by colony PCR using 3' and 5' pGAP primers (forward: 5' GTC CCT ATT TCA ATC AAT TGA A and 3' AOX primer: 5' GCA AAT GGC ATT CTG ACA TCC available from Invitrogen). Plasmid DNA was prepared from PCR positive clone and was sequenced with 3' and 5' *AOX1* primers for confirmation of correct open reading frame. Sequence confirmed plasmid (~25 μg) was linearised with BlnI (Roche), ethanol precipitated and transformed into

competent cells of proteinase deficient strain (SMD1168) of *Pichia pastoris* using a *Pichia* cloning kit (Invitrogen) as per manual instructions. Transformants were selected by plating on YPG-agar plates containing 25 μg/ml Zeocin.

2.17.2 Recombinant expression and purification of DcCathL:

Putative transformants were screened by western blot using anti Myc and anti His antibodies and high expressing clone was selected for over expression of DcCathL by fermentation using Laboratory fermentor BioFlo 110 (5L vessel, New Brunswick Scientific) as described in standard procedures. Fermentor supernatant was processed as described before and then diluted 10-fold with 50mM Sodium-acetate buffer, pH 4.0 (Loading buffer). Diluted supernatant was loaded overnight onto S-sepharose column (Pharmacia) pre-equilibrated with loading buffer at 5ml/min at 4°C. The column was washed with 5 to 10 column volume of binding buffer until the OD_{280} with 0.5 sensitivity reached to a baseline. Protein was eluted with a 0 to 1.0 M gradient of NaCl in loading buffer using a gradient program of FPLC unit (Pharmacia). Recombinant DcCathL was eluted at approx. 0.3 to 0.4 M NaCl concentration. Peak fractions were analyzed by SDS-PAGE (15% acrylamide gel) followed by CBB staining (as described earlier), confirmed for the presence of recombinant DcCathL and then pooled. Pooled DcCathL fractions were concentrated using a centrifugal concentrator (10,000 MWCO; Vivascience) and dialyzed into 30mM sodium-acetate buffer, pH 5.5. Samples were then quantified using BSA as standard by SDS-PAGE followed by CBB staining and stored frozen at -20°C until further use.

2.18 Assays of DcCathL hydrolytic activity and biochemical characterisation:

2.18.1 Standard DcCatL activity assay

The synthetic substrate Z-Phe-Arg-AMC (Z = carbobenzoxy, AMC = 7-amino-4-methyl Coumarin; Bachem) a substrate for the fluorimetric assay of cathepsin B, L as well as papain was purchased from Bachem and 1mM stock was prepared in dimethylformamide. DcCathL did not show any activity without activation with reducing agent either DTT or reduced glutathione. This activation was done by incubating DcCathL with DTT (0.1mM) or glutathione (1mM) at room temperature. Different concentrations of glutathione (0, 0.25, 0.5, 0.75, 1.0, 1.25 and 1.5 mM) were

tried to optimize the amount of glutathione required for DcCathL activation, which was 1mM. DcCathL showed decrease in activity when incubated for longer periods with glutathione. To optimize the time required for activation of DcCathL with glutathione, the incubation reaction aliquots were analyzed for DcCathL activity for 0, 1.0, 2.5, 5.0, 5.5 and 10 minutes.

A standard activity assay of 100 µl consisted of 79 µl assay buffer (30 mM sodium-acetate, pH 5.5), 10 µl glutathione stock (10 mM in assay buffer), 1 µl DcCathL enzyme (1mg/ml), premixed to activate the enzyme, and 10 µl substrate stock. Assays were performed in black 96 well plates (Greiner, Bio-one) at 25°C, and read using Fluoroskan Ascent microtiter plate flourimeter (Labsystems) with excitation and emission filters set to 355 and 460 nm respectively. The plate was shaken for 5 seconds and fluorescence was read immediately for every 15 seconds for the first 5 minutes of reaction. The average reaction rate (AR) for each assay was measured; all assays were performed in triplicate (technical repeats) unless otherwise specified.

2.18.2 Determination of Km of DcCathL:

Km of DcCathL was determined using Z-Phe-Arg-AMC as substrate and same assay buffer as above. Different concentrations 1.25 X 10^{-5} M to 1 X 10^{-4} M of Z-Phe-Arg-AMC as a substrate were used and assays were performed as described earlier. The Michelis-Menten constant and Km curve was obtained by analyzing the assay data in Prism software on Macintosh computer.

2.18.3 Determination of optimal pH for activity of DcCathL:

To determine the optimal pH for activity of DcCathL, the isoionic strength AMT buffer system described by Ellis and Morrison, (1982) was used. Buffer contained 0.05 M acetic acid (AnalaR, VWR), 0.05 M 4-morpholinethane sulfonic acid (MES; Sigma Chemical co.) and 0.1 M Tris (Melford). Buffers were brought to desired pH by adding NaOH or HCl.

2.18.4 Inhibition of DcCathL activity by E-64:

To study inhibition of cysteine protease like activity of DcCathL, an irreversible inhibitor of cysteine proteases E-64: (2S,3S)-3-(N-{(S)-1-[N-(4-guanidinobutyl)

carbamoyl]3-methylbutyl} carbamoyl) oxirane-2-carboxylic acid) (Sigma Chemical Co.) was used. Different concentrations (0.1–10 mM) of E-64 were prepared in assay buffer. The inhibitor was added to the premix of DcCathL and glutathione in assay buffer and allowed to form enzyme-inhibitor complex by incubating for 5 minutes at room temperature. The mixture was then added to substrate and read as described before.

2.18.5 Effect of ionic strength on DcCathL activity:

To study the effect of ionic strength on the activity of DcCathL, a standard assay buffer (30 mM sodium-acetate, pH 5.5) was added with NaCl with varying concentrations (0, 0.1, 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 1.8 and 2M) and assay was performed as described earlier.

2.18.6 Injections of L. oleracea and M. brassicae with DcCathL:

Fifth instar (day1) larvae of *Lacanobia oleracea* or *Mamestra brassicae* were weighed and were injected with 0.5–10 μg of DcCathL dissolved in 5 μl 30 mM sodium-acetate buffer pH 5.5. DcCathL was pre-activated by addition of 1 mM glutathione to the buffer. Larvae were gassed for 30 seconds with CO_2 and then injected on the ventral side just below cuticle into the hemocoel using a 10 μl syringe (Hamilton, 701N). Control larvae were injected with the same volume of buffer with glutathione (1mM). Larvae were allowed access to unlimited artificial diet and surviving larvae were weighed every 24 h.

2.19 Serpins from M. brassicae:

2.19.1 Cloning of MbSpn1A and MbSpn1B/C from M. brassicae:

The sequences of Serpin1A, Serpin1B and Serpin 1C of *Mamestra configurata* available at NCBI (accession numbers; 1A: **AY148483**, 1B: **AY148484**, 1C: **AY148485**) were used to design the gene specific primers for *M. brassicae* serpins. Following primers were synthesized; MbSpn1A; (1) forward: 5’ ATG AAG CTC TTC ATA TGC ATA TTG and (2) reverse: 5’ CTA AGA CTG GTA GAC TCC ATT AAA C) and MbSpn1BC; (3) forward: 5’ ATG AAG CTC TTC ATA TGC ATA TTG and (4) reverse: 5’ CTA AGA GTA AAG TAC TCC GCT G. *M. brassicae* cDNA was prepared from whole insect total RNA using MMLV-RT as described in

standard procedures. Serpin coding sequences were amplified using Phusion polymerase (New England Biolabs) using the PCR cycle: Step I: 98°C for 30 seconds, Step II: 98°C for 10 seconds, 60°C for 20 seconds, 72°C for 30 seconds (20 cycles) and Step III: 72°C for 7 min. The PCR product (approx. 1300 bp) was subcloned and sequenced as described earlier.

2.19.2 Sequence analysis of MbSpn1A and MbSpn1B/C:

Protein sequences predicted from sequenced clones of MbSpn1A and MbSpn1b/C were compared against the global database using BlastP software running on the NCBI Blast server (www.ncbi.nlm.nih.gov/blast) in pairwise comparisons to identify sequences similar to *M. brassicae* serpins. Multiple sequence alignments were carried out using the Clustal algorithm running under ClustalX software (Thompson *et al.,* 1997).

2.19.3 Expression construct, recombinant expression and purification of MbSpn1A and MbSpn1B/C:

The complete coding sequences of MbSpn1A and MbSpn1B/C with mature peptides (1131 bp; 377 amino acids) were cloned into the pGAPZαB yeast expression vector, excluding C-terminal Myc and His tags and introducing an N-terminal $(His)_6$ tag by amplification from the initial PCR products using following primers: GGG AATT CAT CAT CAT CAT CAT CAT GAA GAA TCA AAC GTA (forward for both MbSpn1A and MbSpn1B/C); TA GCGGCCGC TCA AGA CTG GTA GAC TCC ATT (reverse for MbSpn1A); TA GCGGCCGC CTA AGA GTA AAG TAC TCC GCT (reverse for MbSpn1B/C). Cloning, screening, linearization and *P. pastoris* transformation of serpins was carried out as described for DcCathL above except that EcoRI and NotI sites were used for cloning. The recombinant clones were screened by western blot using Anti-(His)6 antibodies (1:5000 dilution) as described earlier. High serpin expressing *Pichia* clones were used for large-scale production of recombinant protein. Clone was inoculated on 1L YPG medium in baffled flask (5L) and the culture was grown for 120 hours at 30°C with shaking at 250 rpm. The culture supernatant was processed and filtered as described above for DcCathL, diluted with an equal volume of 2X Binding Buffer (BB; 20 mM Sodium Phosphate and 0.5 M NaCl, pH 7.4), and loaded onto a 1 ml Ni-NTA column (Amersham Biosciences) pre-

equilibrated with BB. The column was washed with 5 column volumes of BB, then BB containing 20 mM Imidazole (Sigma). Serpins were eluted with BB containing 500 mM imidazole. Purified recombinant serpins were dialysed into 10 mM Tris-Cl, pH 7.0, concentrated using a centrifugal concentrator (Vivascience; 10,000 MWCO) and frozen at -20°C until further use.

2.19.4 Inhibition of standard proteases by recombinant MbSpn1A and MbSpn1B/C:

Recombinant MbSpn1A and MbSpn1B/C were assayed for their inhibition of standard target serine proteases. Synthetic fluorimetric substrates Z-Arg-Arg-AMC (substrate for Trypsin like serine proteases) and Suc-Ala-Ala-Pro-Phe-AMC (for Chymotrypsin like activity) were purchased from Bachem and 1mM stock was prepared in dimethylformamide (DMS; Sigma). A standard assay of 190 µl assay buffer (10mM Tris-Cl, pH 8.0), 0.5µg bovine chymotrypsin (≥ 40 BTEE units/mg; Sigma) was mixed with 0.125 to 2µg MbSpn1A and incubated at 37°C for 30 minutes to allow formation of enzyme inhibitor complex. The mixture was then added to 10µl Z-Arg-Arg-AMC at final concentration of 50µM and residual activity was measured for first 5 minutes of reaction in terms of RFU/min as described earlier for DcCathL assays. Assay for MbSpn1B/C (0.1 to1µg) was carried out in similar way using trypsin (≥ 10,000 BTEE units/mg; Sigma) and 50µM Suc-Ala-Ala-Pro-Phe-AMC as substrate. Inhibitory properties of both serpins (0 to 2µg) on DcCathL (1µg) were assayed using Z-Phe-Arg-AMC as substrate and performing the assay as described above.

2.19.5 Assay of general proteolytic activity in M. brassicae haemolymph:

Haemolymph was extracted from 5th instar larvae of *M. brassicae* as described earlier (Fitches *et al.,* 2002). Larvae were chilled on ice, blotted with ethanol, dried and hemolymph was extracted by piercing the cuticle with a fine needle without going deep enough to pierce the gut. The exuding liquid was collected into pre-chilled 1.5 ml eppendorf tubes containing phenylthiocarbamide (0.1mg powder / tube) as an anticoagulant / antioxidant. Prior to use collected hemolymph was centrifuged (12,000 g, 4°C for10 minutes) to pellet cells. The cell-free supernatant was assayed for its proteolytic activity using a general protease substrate, fluorescein-labelled casein (EnzChek; Molecular Probes). A 200 µl assay mixture in 30 mM sodium-acetate

buffer pH 6.5, 1 mM glutathione (as a reducing agent), contained 10 µl haemolymph (approx. 40 mg total protein), 1 µg EnzChek substrate (10 µl) and varying amounts of DcCathL (1–8 µg). Control reactions contained the same amounts of hemolymph or DcCathL separately to allow measurement of their individual protease activities. Assays were carried out at 25°C and fluorescence was monitored continuously for 30 min, with excitation and emission filters 485 and 538 nm respectively. The average reaction rate for each assay was measured; all assays were the mean of 3 replicates.

2.19.6 Digestion of recombinant serpins / Myoglobin with DcCathL in vitro:

To study whether recombinant DcCathL is able to hydrolyze and degrade recombinant serpins and a neutral protein Horse skeletal myoglobin (Sigma), serpins and myoglobin were digested *in vitro* with active DcCathL. MbSpn1A, MbSpn1B/C or myoglobin (50µg each) were digested separately with 5 µg DcCathL (activated with 1 mM glutathione) in 30 mM sodium-acetate buffer pH 5.5 in a total volume of 100 µl. Reaction aliquots of 10 µl were collected at different time points 5, 10, 20, 30, 60, 120 and 240 minutes. The reaction in each aliquote was stopped adding 1 µl of 0.5M NaOH, as DcCathL is not active above pH 8.0 and samples were frozen. Samples containing either serpin or myoglobin without DcCathL were incubated as controls to observe self-degradation. Reaction aliquots at different time points were analysed on 15% SDS-PAGE gels and stained with Coomassie Blue as described earlier.

2.19.7 Effect of DcCathL on M. brassicae hemolymph proteins in vitro:

To study the effect of DcCathL on degradation of hemolymph proteins *in vitro*, hemolymph digestion reactions were set up. In a total volume of 240 µl, containing 30 mM sodium-acetate buffer pH 6.5, 40 µl *M. brassicae* hemolymph proteins (approx. 160 mg total protein) were incubated in the presence or absence of 8 µl (16 µg) of recombinant DcCathL at 25°C. Aliquots of 30 µl (each with approx. 20 µg total hemolymph proteins $\pm$ 2 µg DcCathL) were collected after intervals of 4, 8 and 26 hours, and frozen to quench the reaction. The reaction aliquots were analysed by SDS-PAGE (17.5% acrylamide), and gels were stained with Coomassie blue for visualization of total protein. To determine whether DcCathL degraded recombinant MbSpn1A or MbSpn1B/C in the presence of *M. brassicae* haemolymph proteins, the

240 µl digestion reactions contained in addition 16 µl (16 µg) recombinant MbSpn1A or MbSpn1B/C. Aliquots of 30 µl (each with approx. 20 µg total hemolymph proteins + 2 µg recombinant serpins + 2 µg DcCathL) were collected at intervals of 0, 30, 60, 120, 240, 360, 480 minutes and after overnight incubation. The reaction aliquots were analysed by SDS-PAGE (17.5% acrylamide), and gels were stained with Coomassie blue for visualization of total proteins or subjected to western blotting and probed using anti-$(His)_6$ antibodies as described earlier.

2.19.8 Co-injections of DcCathL and recombinant serpins in M. brassicae:

To study the *in vivo* effect of recombinant serpins, 5th instar *M. brassicae* larvae were injected with DcCathL and serpins together in 30mM sodium-acetate buffer, pH 5.5. A mixture of DcCathL + MbSpn1A (5µg each) and DcCathL + MbSpn1B/C (5µg each) were prepared and injected along with a buffer control as described earlier for DcCathL. The larvae were observed (4 hours and 24 hours post injection) for response to DcCathL in presence or absence of recombinant serpins, in the form of fluid loss (not measured quantitatively, but only visual observation in the shrinkage of larvae), mortality and degree of melanization. Melanization was scored on a scale of 0 to 4, where 0 was no melanization, 1 was melanization at injection site, 2 was melanization spreading from injection site and 4 was larvae completely melanised. Data were subjected to statistical analysis using the Prism software (GraphPad Software Inc).

2.20 Characterisation of S. avenae gut proteolytic activity:

2.20.1 Determination of pH optima for gut protease activity:

AMT buffer system (Ellis and Morrison, 1982) was used to determine the optimum activity of the gut proteases. 10ug total protein of CA gut extract was used for the activity assays using 1µg EnzChek substrate and buffers of pH 5.0, 5.5, 6.0, 6.5, 7.0, 7.5, 8.0, 8.5 and 9.0 (in absence or presence of 1mM Glutathione as reducing agent). Assay method was similar to standard EnzChek assay described elsewhere.

2.20.2 Inhibition of S. avenae gut protease activity with standard proteinase inhibitors:

The following synthetic protease inhibitors Trans-epoxysuccinyl-L-leucylamido-(4-guanidino)butane (E-64), 4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride

(AEBSF), Chymostatin, Nα-$_P$-Tosyl-$_L$-Lysine Chloromethyl Ketone TLCK, N-Tosyl-$_L$-Phenylalanine Chloromethyl Ketone (TPCK) were purchased from Sigma Chemical company. Inhibitors were added to the standard assay mixture at following final concentration, E-64 (1μM), AEBSF (0.1mM), Chymostatin (10μM), TLCK (10μM) and TPCK (10μM) and residual protease activity was measured using EnzChek substrate. Assay methods were similar to standard EnzChek assays discussed elsewhere.

2.20.3 Activity assays of S. avenae gut extract with specific substrates:

Synthetic fluorimetric substrates from Bachem Ltd. were used to study the type of protease activity in *S. avenae* gut extract. 1mM stocks of following substrates were prepared in Dimethylformamide (Sigma Chemical Co.), Z-Phe-Arg-AMC (for cathepsin like Cysteine proteases), Z-Arg-Arg-AMC (for Trypsin and Cathepsin like activity), Suc-Ala-Ala-Pro-Phe-AMC (for Chymotrypsin like activity). In an assay of 200 μl substrates were used at the final concentration of 0.1mM. Method of assay was similar to standard synthetic fluorimetric substrate assay as discussed elsewhere.

2.21 Wheat Subtilisin/Chymotrypsin Inhibitor (WSCI)

2.21.1 Cloning and expression construct for WSCI:

The following gene specific primers were designed based on Subtilisin/ chymotrypsin inhibitor of wheat (accession no. **AAY45744**), forward 5' ATG AGT TCT GTG GTG AAG AAG and reverse: 5' GCC GAC CCT GGG GAC CTG and were synthesized by Sigma Genosys service. Wheat (c.v. Claire) cDNA was prepared from leaf total RNA using qScript^{TM} cDNA SuperMix (Quanta Biosciences) as per manufacturers instructions. Coding sequence of Subtilisin/chymotrypsin inhibitor was amplified from above cDNA using Phusion polymerase (New England Biolabs) and following PCR conditions: Step-I: 98°C for 30 seconds, Step-II: 98°C for 10 seconds, 64°C for 10 seconds, 72°C for 30 seconds (25 cycles) and step-III 72°C for 5 minutes. The PCR product (approx. 255 bp) was subcloned and sequenced as described earlier. Following cloning primers were designed to clone the coding sequence of mature WSCI into pGAPZαB *Pichia pastoris* expression vector (Invitrogen) using the restriction sites PstI and XbaI (in bold) Fwd: 5' TA **CTG CAG** CA ACC GAT ACT GGT GAC and Rev 5' TA **TCT AGA** CC GAC CCT GGG GAC CTG. PCR was

carried out using Phusion polymerase (New England Biolabs) and PCR cycle: Step-I: 98°C for 30 seconds, Step-II: 98°C for 10 seconds, 69°C for 20 seconds, 72°C for 20 seconds (15 cycles) and step-III 72°C for 7 minutes. The PCR product and pGAPZαB vector were digested with PstI and XbaI, separated on 1% agarose gel and were gel purified using QiAquick gel extraction kit (Qiagen). Digested vector and PCR product were ligated using T4 DNA ligase (Promega) and transformed into TOPO cells of *E.coli*. Resultant clones were screened and sequenced using pGAPZαB sequencing primers (*AOX1*). Sequence confirmed plasmid was linearized using BlnI (Roche) and transformed into protease deficient strain (SMD1168) of *P.pastoris* using a *Pichia* cloning kit (Invitrogen) as described earlier.

2.21.2 Recombinant expression and purification of WSCI:

The putative transformants expressing WSCI were screened by western blot using anti Myc and anti $(His)_6$ antibodies (both 1:3000 dilution) as described earlier. The selected high expressing clone was used for scale up expression by Bench-top fermentation (5 l vessel, BioFlo 110; New Brunswick scientific) as previously described. Recombinant WSCI was purified by cation exchange chromatography using S-sepharose column (1.6cm dia., 40ml volume; Amersham Pharmacia. Fermentor supernatant was processed as described earlier and was diluted 5X in loading buffer (50mM Sodium Acetate pH 4.0) and loaded onto S-sepharose column pre-equilibrated with same buffer. Column was washed with 5 volumes of loading buffer and bound protein was eluted with 0-1.0M gradient of NaCl in loading buffer. Recombinant WSCI was eluted at approx. 0.6 to 0.7 M NaCl. Peak fractions were analysed on SDS-PAGE, pooled and pooled fractions were dialysed against DW as described earlier using 12kD MWCO dialysis membrane. Dialysed samples were freeze dried and were quantified (either by BCA estimation of gel quantification) before use as per standard procedures.

2.21.3 Assays for inhibition activity of recombinant WSCI:

1mM stocks of synthetic fluorimetric substrates Suc-Ala-Ala-Pro-Phe-AMC {(N-Succinyl- L-alanyl- L-alanyl- L-prolyl- L-phenylalanine-4-methylcoumaryl-7-amide) for chymotrypsin like activity} and Z-Gly-Gly-Leu-AMC {(N-benzyloxycarbonylglycyl-glycyl-leucyl) amino-4-methylcoumarin, for Subtilisin like

activity} (both BACHEM) were prepared in dimethylformamide (DMF; Sigma Chemical Co.)

2.21.4 Chymotrypsin inhibition assays with WSCI:

A standard assay (10mM Tris-Cl pH 8.0 as assay buffer) of 100µl consisted of 500ng chymotrypsin (which was standardised for optimum activity) and 50µM of respective substrate with varying amounts of WSCI (50ng to 4µg). Assays were performed in black 96 well (flat bottom) plates (Greiner Bio-one) at 25°C and read using Fluoroskan Ascent microtitre plate flourimeter (Labsystems; excitation and emission filters set to 355 and 460nm resp.). 50µl chymotrypsin and (1 to 40µl) WSCI were mixed together in buffer (to final volume 100µl) and incubated at room temperature with slow shaking for 20 minutes to form the enzyme inhibitor complex. 10µl substrate was then added to the plate wells and mixed with enzyme-inhibitor mixture. The plate was shaken for 5s and fluorescence was read immediately for every 30 seconds for the first 5 minutes of reaction. The average reaction rate (AR) for each assay was measured; all assays were the mean of three replicates unless otherwise specified.

2.21.5 Subtilisin inhibition assays with WSCI:

Assays for Subtilisin inhibition with WSCI were performed similar to Chymotrypsin (as above) apart from, the concentration of Subtilisin used per assay was 12µg, substrate concentration was 100µM and WSCI concentrations used were 3µg to 36µg.

2.21.6 Inhibition assays of S. avenae gut extracts with WSCI:

Quantified total gut protein was used for the gut protease inhibition assays. Assays were performed using 10mM Tris buffer (pH 7.2) and EnzChek substrate. In a standard assay of 200µl, 10µl gut extract (3µg) was inhibited with (0, 0.25, 0.5, 0.75, 1, 2, 3 and 4µg) of WSCI and residual protease activity was detected using 2µg substrate. The mixture of buffer, gut extract (proteases) and inhibitor was mixed together and incubated at RT for 15 minutes to form enzyme inhibitor complex and then added to flat bottom black welled plate (Greiner-BioOne) containing substrate. The fluorescence was measured in terms of RFU/min for the first 30 minutes of reaction using Fluoroskan Ascent microtiter plate flourimeter (Labsystems) with

excitation and emission filters set to 485 and 538 nm resp. All assays were performed in triplicate unless and otherwise specified.

2.22 Wheat Cysteine Proteinase Inhibitor (WCPI):

2.22.1 Cloning, recombinant expression and purification of WCPI:

Following gene specific primers were synthesized based on wheat cysteine proteinase inhibitor (accession no. **BAB18766**), forward 5' ATG GAG ATG TGG AAA TAT CGG and reverse: 5' TGC GCT TGG AAC GTC CTC. Coding sequence of WCPI (approx. 426 bp) was amplified by following exactly the same PCR conditions as described above for WSCI, cloned and sequenced. Following primers were designed to clone the coding sequence of WCPI into pGAPZαB *Pichia pastoris* expression vector using the restriction sites PstI and XbaI (in bold) Fwd: 5' TA **CTG CAG** CA CAG ACG CAG AGC GC and Rev 5' TA **TCT AGA** GC GCT TGG AAC GTC CTC. Methods for preparation of expression construct for WCPI in pGAPZαB, screening for transformants, plasmid linearization, *P. pastoris* transformation, screening by western blot, fermentation, purification, SDS-PAGE electrophoresis and western blot (Anti-Myc and Anti-His antibodies) confirmation were similar to WSCI as discussed above.

2.22.2 Inhibition of Papain by WCPI:

Inhibition of Papain a cysteine proteinase from papaya (Sigma Chemical Co.) was studied using a synthetic flourimetric substrate Z-Phe-Arg-AMC (where Z = carbobenzoxy-, AMC = 7-amino-4-methyl Coumarin). 1mM stock of substrate was prepared in dimethylformamide (DMF; Sigma). In a standard assay mixture (using 200µl of 30mM sodium acetate buffer pH5.5; assay buffer), 1µg of papain (100 ng/µl) was inhibited with 0.01-1µg of WCPI (10 ng/µl) in presence of 1mM Glutathione (10µl of 10mM stock) as reducing agent (all solutions prepared and diluted in assay buffer). Reaction volume was adjusted to 190µl with assay buffer and mixture was incubated at 25°C with shaking for 10 minutes to allow formation of enzyme-inhibitor complex. This assay mixture was added to 10µl substrate (50µM) in flat 96-welled black plate (Greiner, Bio-one) and read using Flouroskan Ascent microtiter plate flourimeter (Labsystems,) with excitation and emission filters set to 355 and 460nm resp.). The plate was shaken for 5 seconds and fluorescence was read immediately for

every 15 seconds for the first five minutes of reaction. The average reaction rate (AR) for each assay was measured; all assays were the mean of 3 replicates unless otherwise specified.

2.22.3 Inhibition of S. avenae gut extracts with WCPI:

Assay method was similar to WSCI using EnzChek as substrate and similar concentrations of inhibitor.

2.23 S. avenae Inhibitor Feeding Bioassays:

Following diet combinations were used for the inhibitor feeding bioassays to assess the importance of presence or absence of free amino acids in diet while feeding on protease inhibitors. (1) Control diet (+AA: Vitamins + Minerals + Amino acids) (see recipe), (2) ½ AA diet: 50% (Vitamins + Minerals + 50% of original Amino acids) (3) No protein diet (-AA): No free amino acids (Vitamins + Minerals). Inhibitors WSCI and WCPI (1mg/ml) were added to the all the above combinations of diet. Adult aphids were isolated 48 hours before starting feeding assay, placed in Perspex aphid feeding chambers covered by two sheets of Parafilm with 100µl artificial diet sandwiched between two layers as shown in Figure 2.2 and were left to produce nymphs overnight. Nymphs were fed for another 24 hours on normal diet and then exposed to the diet containing inhibitors. Ten aphids were fed per cage on 100µl diet and 3 cages were used per combination of diet. Aphids were monitored for the survival and growth. Aphid images were taken using Openlab software (Modular Imaging Software) and measurements of length and width were done using 'ImageJ' image processing program (Wayne Rasband, National Institute of Health, USA). For feeding assays with recombinant proteins for *D. coarctata* purified recombinant proteins (2µl; with desired concentration) were injected into approx. 2 cm long wheat stems and neonates of *D. coarctata* were exposed to feed on injected stems and survival was monitored for one week following feeding with one change of stem in between. For control, neonates were fed on stems injected with distilled water.

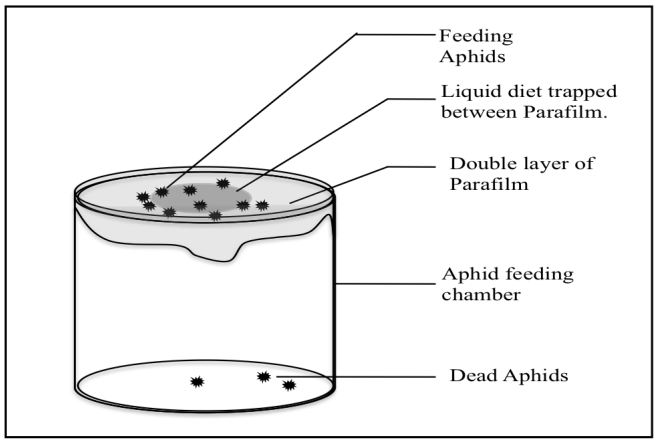


Figure 2.2: Setup for feeding aphids on liquid diet.

2.24 Honeydew analysis of S. avenae:

To study whether aphids use their gut proteases for ingested protein digestion in the gut, full-grown aphids were starved for 1 hour and fed with 1mg/ml Ovalbumin (Sigma Chemical Co.) as a digestible protein in a diet without free amino acids (-AA). 5-10 aphids were fed on 100ul diet in 1.5 ml Eppendorf tube as shown in Figure 2.3 and honeydew was collected after 24, 48 and 72 hours by adding (20µl) 1X SDS-PAGE sample loading buffer to the tubes. Collected honeydew was analysed along with fed diet (20µl; as control) on SDS-PAGE followed by silver staining (Blum *et al.,*1987). In other combinations protease inhibitors (WSCI and WCPI) were added to the diet at the final concentration of 1mg/ml.

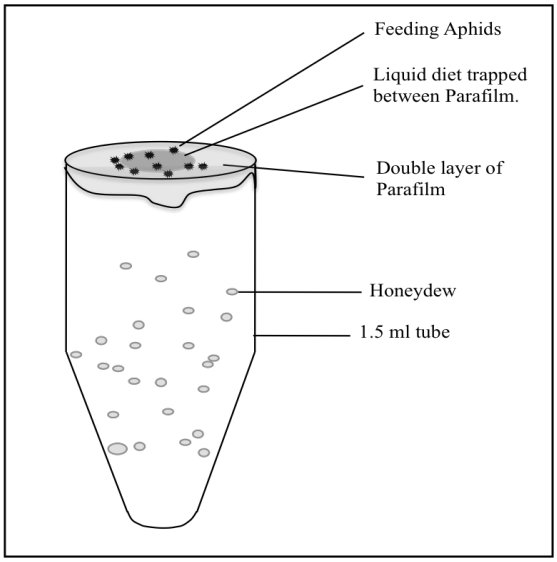


Figure 2.3: Setup for honeydew collection

2.25 Hessian fly Responsive proteins (Hfrs) from wheat:

2.25.1 Cloning and expression constructs Hfrs

Complete coding sequences of Hfr genes were amplified using wheat cDNA prepared from leaf total RNA using MMLV-RT reverse transcriptase as described earlier. Primers were designed for Hfr genes based on Hfr-1, Hfr-2 and Hfr-3 nucleotide sequences available at NCBI with Accession Numbers. **AF483596**, **AY587018** and **DQ462308** respectively. The sequences of Hfr were checked for the presence of restriction sites in the sequences that are being used for cloning. Following primers were synthesized for cloning Hfrs in pGAPZαB: Hfr-1 Forward: 5' TA **CTG CAG** CA ATG TCG CCT CAG, Hfr-1 reverse: 5' TA **TCT AG A**GT TGA AAC GGG TGA ACG, Hfr-2 forward: 5' TA **CTG CAG** CA ATG TTT CCA GTG, Hfr-2 reverse: 5' TA **TCT AG A**AT GGG CTC TTG TGT, Hfr-3 forward: 5' TA **GAA TTC** TT ATG AAG GGC CTC and Hfr-3 reverse: 5' TA **TCT AG A**TT GCA TGC ACC ATT. Restriction enzyme sites (shown in bold) were included in the primers for cloning coding sequences in MCS of expression vector, Hfr-1 and Hfr-2 (PstI and XbaI), Hfr-3 (EcoRI and XbaI). All three Hfr full coding sequences were amplified from wheat cDNA using Phusion polymerase and following PCR conditions: Step-I: 98°C for 2 minutes, Step-II: 98°C for 30 seconds, 63°C for 30 seconds, 72°C for 60 seconds (15 cycles) and Step-III: 72°C for 5 minutes. PCR products, Hfr-1 approx. 1000bp, Hfr-2 approx. 1470bp and Hfr-3 approx. 600 bp were subcloned and sequenced as described earlier. Sequences were analysed in Sequencher software and compared with sequences in database.

2.25.2 Expression constructs for Hfr-1 and Hfr-2

The PCR products were purified from agarose gel as described earlier and restriction digested using respective enzymes (PstI and XbaI). Expression vector pGAPZαB was also digested with respective enzymes and purified from gel. Inserts were ligated into pGAPZαB vector using T4 DNA ligase following standard procedures. Expression cassette was introduced into TOP10 E.coli cells and transformant screening, sequencing were done as described earlier.

2.25.3 Expression construct for Hfr3:

Sequence of Hfr-3 was analysed for a presence of signal peptide using SignalP 3.0

server (http://www.cbs.dtu.dk/services/SignalP/) and a forward primer (TA**GAATTC**AGCTGCAGTCC; EcoRI site in bold) was designed to clone only mature peptide (QLQS.....GACN) excluding the signal peptide (MKGLLLCALALAFAAVTTHA). New mature Hfr-3 coding sequence was amplified using Phusion polymerase and following PCR cycle: Step-I: 98°C for 30 seconds, Step-II: 98°C for 10 seconds, 64°C for 20 seconds, 72°C for 20 seconds (15 cycles) and Step-III: 72°C for 7 minutes. PCR product (534 bp) was purified from gel, subcloned and sequence confirmed as described earlier. Coding sequence of Hfr-3 ws found to have a BlnI restriction site, which is crucial for linearization of the expression vector pGAPZαB with insert in it before transforming *P. pastoris.* This BlnI restriction site was modified by PCR based strategy. Following primers were designed to remove BlnI site (C**C**TAGG was replaced by C**T**TAGG in reverse primer) by maintaining the codon; forward: 5' ACC GGA GTT CTG TGG TAC TGG and reverse 5' CCT AAG CCA CAG TAA CCG TAC. Whole pGAPZαB vector containing Hfr-3 was amplified round the plasmid using above primers and Phusion polymerase. Following PCR conditions were used; Step-I: 98°C for 30 seconds, Step-II: 98°C for 30 seconds, 63°C for 10 seconds, 72°C for 5 minutes (15 cycles) and Step-III: 72°C for 7 minutes. PCR product of (approx 4.1 kb pGAPZαB 3.6 kb + Hfr-3 approx. 534 bp) was purified from agarose gel as described earlier, and was re-ligated using T4 DNA polymerase. The expression cassette containing Hfr-3 was linearised with BlnI, transformed into *P. pastoris* and screening of putative transformants by western blot was done as described earlier.

2.25.4 Recombinant expression and purification of Hfr3

High expressing clone was selected for over-expression of recombinant Hfr-3 in *P. pastoris.* Bench-top fermentation was performed and culture supernatant was processed as described earlier. To the supernatant, equal volume of 2X Binding Buffer (50mM sodium acetate, 0.5M NaCl, pH 6.0) was added to final concentration of 1X and loaded onto Ni-NTA (Amersham Biosciences) column pre-equilibrated with 1X BB. Loading was performed overnight at 3ml/minute at room temperature. Elution of Hfr-3 was performed as described earlier in BB containing 300mM Imidazole.

2.25.5 Feeding bioassays recombinant Hfr-3 and WGA to S. avenae

To study the antimetabolic effects on *S. avenae*, wheat germ agglutinin (WGA) and purified recombinant Hfr-3 were added to aphid diet (+AA) at the concentration of 1mg/ml and aphids were fed up to 1 week. Survival was observed for 5 days after feeding and growth in terms of length and breadth was measured as described earlier. For the dose response effect, different concentrations of recombinant Hfr-3 (0.125, 0.25, 0.5 and 1mg/ml) were added to the liquid diet (+AA). Control diet (+AA) did not contain Hfr-3. Survival of aphids was monitored for upto 12 days.

2.26 Cationic amino acid transporter from D. coarctata (DcCAAT):

2.26.1 Cloning of DcCAAT:

Following degenerate primers were designed based on cationic amino acid transporter nucleotide sequence of *D. melanogaster* (Accession numbers **CG9413 PA** and **CG9413 PB**), for the conserved regions (LLSGVA and WAYDGW) forward: 5'CTC NTC AGT GGG GTG GC and reverse: 5' CCA BCC RTC GTA GGC CC. DcCAAT partial coding sequence was amplified from gut cDNA using Taq polymerase and PCR cycle as Step-I: 94°C for 5 minutes, Step-II: 94°C for 30 seconds, 58°C for 30 seconds, 72°C for 60 seconds (25 cycles) and Step-III: 72°C for 7 minutes. PCR product was subcloned and sequenced as described earlier. Sequence was analysed and confirmed by BLAST search on NCBI server (www.ncbi.nlm.nih.gov/blast). Gene specific primers were designed for 3' RACE (GAG TGG GCA TAT TTT ATG GAC GCA TTT GGAC) and 5' RACE (T AGC TCCCAT CGT GGG TGT GGG TCC AC). 3' and 5' RACE cDNA preparation and RACE were performed similar to DcCathL as described earlier. RACE products purified from gel, subcloned and sequenced as described earlier. Aligning the sequences of degenerate PCR product along with 3' and 5' RACE gave a full coding sequence for DcCAAT.

2.26.2 Construct for DcCAAT RNAi:

RNAi construct was prepared in pJET 1.2 cloning vector (Fermentas). This vector has a T7 promoter (5' TAATACGACTCACTATAG) that can be used for in vitro transcription of sense strand. For RNAi construct reverse primer was modified to contain a T7 promoter sequence for antisense strand. Following primers were designed for approx. 800 bp fragment from the 3' end of DcCAAT (forward: AT

GAA TTC ATG ATT TTA TTT GTT AAC TGT TAT and reverse: AT **TAATACGACTCACTATAG** ATA TTC AAT GCG AGG; T7 sequence in bold). RNAi insert was amplified from parent plasmid DNA using Phusion polymerase and following PCR cycle: Step-I: 98°C for 30 seconds, Step-II: 98°C for 10 seconds, 56°C for 10 seconds, 72°C for 50 seconds (20 cycles) and Step-III: 72°C for 7 minutes. PCR product was ligated into pJET1.2 vector by blunt end ligation, transformed into *E. coli* and PCR screening was performed as described earlier. Positive plasmid was sequence confirmed for insertion of T7 sequence for the antisense strand.

2.26.3 In vitro transcription:

To prepare double stranded RNA, plasmid DNA was quantified using Nanodrop (NanoDrop ND-1000 Spectrophotometer; Bosch Institute) and ~ 10μg plasmid was linearised with either XbaI (for sense strand) or XhoI (for antisense strand). Linearised plasmids were confirmed by 1% agarose gel electrophoresis along with uncut plasmid. Linearization reactions were purified by Phenol : chloroform : isoamylalcohol (25:24:1, Fluka), followed by chloroform. DNA was precipitated as described earlier. Precipitated DNA was washed with 70% ethanol, air-dried, re-suspended in sterile distilled water and quantified. 1 μg linear DNA was used for making double stranded RNA using MEGAscript in vitro transcription kit (Ambion) for each sense and antisense strand in a standard 20μl reaction by following manufacturers instructions.

Chapter 3

Gut proteases from *Delia coarctata*:

3.1 Dipteran gut proteases:

Several proteases have been identified from the Dipteran insects such as *Drosophila melanogaster* and Hessian fly: *Mayetiola destructor*. The proteases reported from *Mayetiola destructor* are mainly serine proteases having either trypsin or chymotrypsin like activity (Zhu *et al.,* 2005) and major gut protease activity in the larval guts seems to be chymotrypsin-like (Shukle *et al.,* 1985). The nature of this proteolytic activity was found to be consistently expressing in all studied larval stages. Expression of these proteases in the midgut has been confirmed using RT-PCR techniques (Mittapalli *et al.*, 2005). However there is no data available on the proteases of *Delia coarctata.* The study of gut proteases of *D. coarctata* is necessary to understand the proteolytic activity in the gut. This study could help in determination of the strategies for control of this severe pest, using the available crop protection strategies.

3.2 Gut proteolytic activity of D. coarctata

Total protein extracted from dissected entire guts of *D. coarctata* showed high proteolytic activity using EnzChek as a general protease substrate. When four buffers with different buffering range were used to find out the types of proteolytic activity in *D. coarctata* gut, a high activity zone around pH 9.5 was observed that could suggest presence of serine type digestive proteases (Figure 3.1). There was another highly active zone seen around pH 7.0 that could be attributed to both serine and cysteine proteases. Use of another buffer system (AMT; Ellis and Morrison 1982) gave more stable and consistent results for gut activity with pH change. Using AMT buffer system proteolytic activity was seen to be highest at pH around 7.0 with no or very little activity below pH 4.0 and above 9.0 (Figure 3.2). The activity of gut proteases was slightly higher in presence of glutathione as a reducing agent, which could be due to complete functionality of Cysteine proteases. The pH of *D. coarctata* gut was found to be around 6.8. Maximum activity at pH 7.0 suggested a suitable gut

environment for the activity of both Serine and cysteine type proteases. The high activity seen around pH 7.0 was considered to be the optimal pH for activity of most gut proteases, as cysteine proteases though highly active in acidic conditions show some activity around neutral pH and serine proteases highly active in alkaline conditions show some activity at neutral pH.

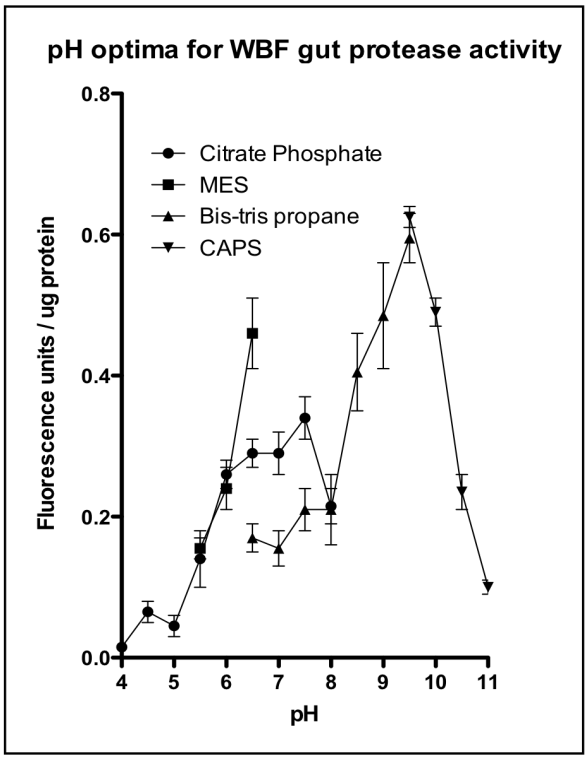


Figure 3.1: Gut proteolytic activity of *D. coarctata.* Different buffers, citrate phosphate, MES, Bis-tris propane and CAPS were used to determine the optimum pH for the activity of gut proteases. Points indicate mean values ± SE (n=5).

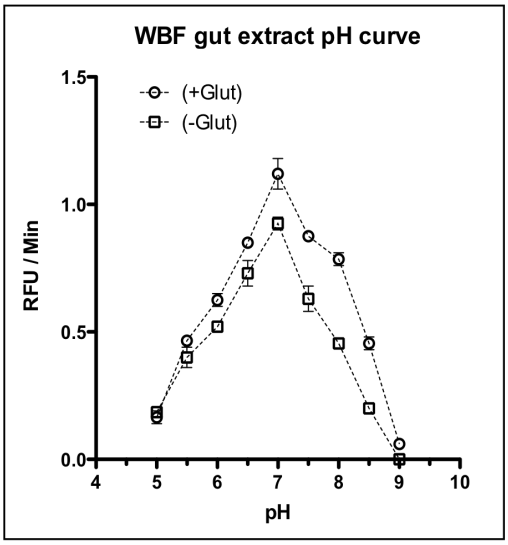


Figure 3.2: Gut proteolytic activity of *D. coarctata.* AMT buffer system (Ellis and Morrison, 1982) was used to determine the optimum pH for the activity of gut proteases. Total proteolytic activity of gut extract in presence or absence of glutathione as reducing agent is shown. Points indicate mean values ± SE (n=3)

In Housefly (*Musca domestica*), it is reported that different regions of the gut (foregut, midgut and hindgut) show morphological differences along with varying pH (Espinoza-Fuentes and Terra, 1987). These different pH zones along the midgut are suggested to be due to fluid fluxes and ion secretions. Accordingly, pH of the foregut and hindgut is neutral to alkaline (8-8.5) and midgut is more acidic. Midgut is further distinguished into fore-midgut and fore-hindgut, which show a neutral pH (6.1 and 6.8 respectively), while the mid-midgut is most acidic (3.1). This acidic zone was suggested to be responsible for killing the bacteria in food using cathepsin D like cysteine protease and an acid lysozyme to release bacterial nutrients for absorption (Espinoza-Fuentes and Terra, 1987; Lemos and Terra. 1991). It is important to note that in *M. domestica* larvae the digestive enzymes are mainly found in the hind-midgut where majority of the digestion should take place and pH of this region is neutral. Presence of different pH zones in the gut could explain the role of different types of proteases in the gut proteolytic digestion. While dissecting the *D. coarctata* guts, tissue was not differentiated as different regions but the extracts were prepared from the entire gut. This was the reason, activity of both serine as well as cysteine type proteases in the extracts was observed.

3.3 Serine Proteinase (DcSP) from D. coarctata gut:

3.3.1 Cloning of DcSP and sequence analysis

A complete coding sequence of DcSP was obtained by PCR based methods. A partial coding sequence of DcSP was amplified from *D. coarctata* gut cDNA using degenerate primers and then the complete coding sequence was obtained by 3' and 5' RACE. The overview of RACE is shown in figure 3.3 (A). Partial coding sequence of DcSP amplified was 474bp long, and 5' and 3' RACE products were approx. 700bp and 900bp respectively. Actual PCR products analysed and cloned from agarose gels are shown in figure 3.3 (B). 3' RACE product had a long untranslated region (UTR) with poly A tail (approx. 780bp) while 5' RACE product had a shorter UTR (approx. 550bp). All RACE products when aligned in 'Sequencher' software and analysed gave a complete coding sequence DcSP with a single open reading frame of 744bp with 248 amino acid residues.

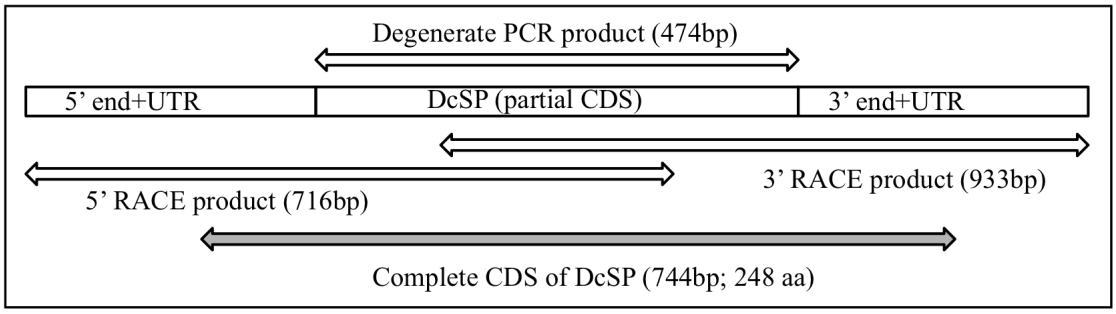


Figure 3.3 (A): Overview of RACE for *D. coarctata* serine protease (DcSP)

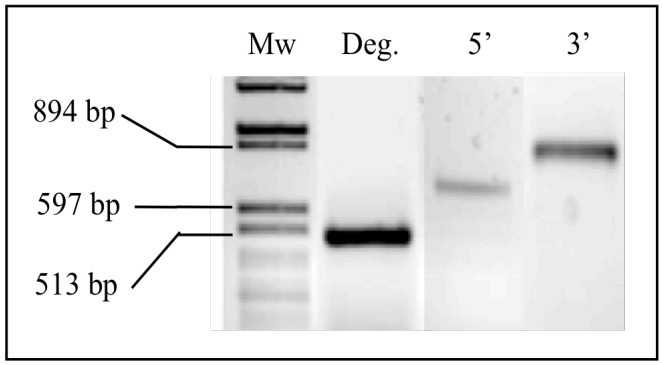


Figure 3.3 (B): Actual PCR amplification products analysed on 1% agarose gel. Lane Deg.: Product of Degenerate PCR, Lane 5': Product of 5' RACE and lane 3': product of 3' RACE.

A complete coding sequence of DcSP with deduced amino acid sequence is shown in figure 3.4 (A). DcSP showed presence of a 19 residues long signal peptide (MSCFSYLLCLSLLLGVAWS) when analysed by 'SignalP' server. Complete

coding sequence of DcSP consists of 229 amino acid residues. Conserved amino acid residues that were used to design degenerate primers were HVCGGSII (on the 5' end) and CNGDSGG (on the 3' end). DcSP showed all the necessary conserved amino acid residues for trypsin / chymotrypsin like serine proteases including catalytic regions of TAA**H**C, **D**LAL and GD**S**GG. The presence of linear order of typical catalytic triad residues in DcSP (H^{66}, D^{113} and S^{204}; DcSP numbering) could confirm it as a member of S1 family of trypsin / chymotrypsin like peptidases (Rawlings and Barrett, 1993). Alignment of DcSP with Dipteran serine proteases is shown in figure 3.4 (B). DcSP showed about 60-70% identical residues to other serine proteases from Dipteran insects especially for conserved regions DIAL, GRVGG, QFPHQISLR, CGGSII, LTAAHCV, GDSGGPA shown in black boxes, which confirms DcSP being a serine protease.

```
1    ATGTCGTGTTTTAGTTATTTGTTGGTATTAAGTTTGCTTCTAGGAGTTGCATGGTCTAGTCCAACTACT
1     M  S  C  F  S  Y  L  L  V  L  S  L  L  L  G  V  A  W  S  S  P  T  T
70   GGTCGTGTTGTAGGTGGTCATGATGCTGCTTCTAAACAATTCCCTCATCAGATCTCRTTACGTCGTGAT
24    G  R  V  V  G  G  H  D  A  A  S  K  Q  F  P  H  Q  I  S  L  R  R  D
139  CATAGTCATGTRTGTGGTGGTTCTATTATAGGCGAGCGTTATGTCTTAACCGCTGCTCATTGTGTTGTT
47    H  S  H  V  C  G  G  S  I  I  G  E  R  Y  V  L  T  A  A  H  C  V  V
208  ATTGAAGGAACTACAGCTTTTCCAGCTGATCGTTTYACAATTCGAGCTGGTACCATCAATCGTATTGCT
70    I  E  G  T  T  A  F  P  A  D  R  F  T  I  R  A  G  T  I  N  R  I  A
277  GGTGGTATTATTGTGCCTGTTAAACGTGTTTTAGTTCATCCGGACTATGTCGTTTACAATGATTTGGCT
93    G  G  I  I  V  P  V  K  R  V  L  V  H  P  D  Y  V  V  Y  N  D  L  A
346  CTTTTGGAATTGGAACAAGAATTAGTCTTCTCCGAAACTATAAATAAGATTGAAATGATGGATAGTGAA
116   L  L  E  L  E  Q  E  L  V  F  S  E  T  I  N  K  I  E  M  M  D  S  E
415  GTTCCTGCCAGCTCGGATGTTATTATTTCGGGTTGGGGTTTAACCGAACATGCCGGTGCTAATTTACCC
139   V  P  A  S  S  D  V  I  I  S  G  W  G  L  T  E  H  A  G  A  N  L  P
484  ATTATTATGCAGTGGTACAAGGTGACTGCTTTAAGCAAAACCGGTTGTGCTAGTAAAATGGGTTTATAT
162   I  I  M  Q  W  Y  K  V  T  A  L  S  K  T  G  C  A  S  K  M  G  L  Y
553  ACCGATGCTATAATTTGCTTAAATCATCCTTCAGGCGCGGGTGCCTGCAATGGTGATTCTGGTGGTCCT
185   T  D  A  I  I  C  L  N  H  P  S  G  A  G  A  C  N  G  D  S  G  G  P
622  GCTACTTATAATGGAAAACTRGTTGGTGTTGCTGGTTTTGTTGTCATTAAATGTGGTTCTTCCAGACCT
208   A  T  Y  N  G  K  L  V  G  V  A  G  F  V  V  I  K  C  G  S  S  R  P
691  GATGGTTATGCTAAAGTTGCTTATAATATTGATTGGATTAGAGAGAAATATGATTTAA
230   D  G  Y  A  K  V  A  Y  N  I  D  W  I  R  E  N  M  I  .
```

Figure 3.4 (A): Nucleotide and deduced amino acid sequence of *D. coarctata* serine protease (DcSP). Conserved regions with residues shown in bold (HVCGGSII and CNGDSGG) were used for design of degenerate primers. Nucleotide sequence shown in bold was used for 3' RACE. Nucleotide sequence shown in bold and underlined was used for 5' RACE. Predicted signal peptide is shown in dark background. The conserved sequences for catalytic regions of serine protease are underlined. Key amino acid residues H66, D113 and S204 (DcSP numbering) in catalytic domain are encircled.

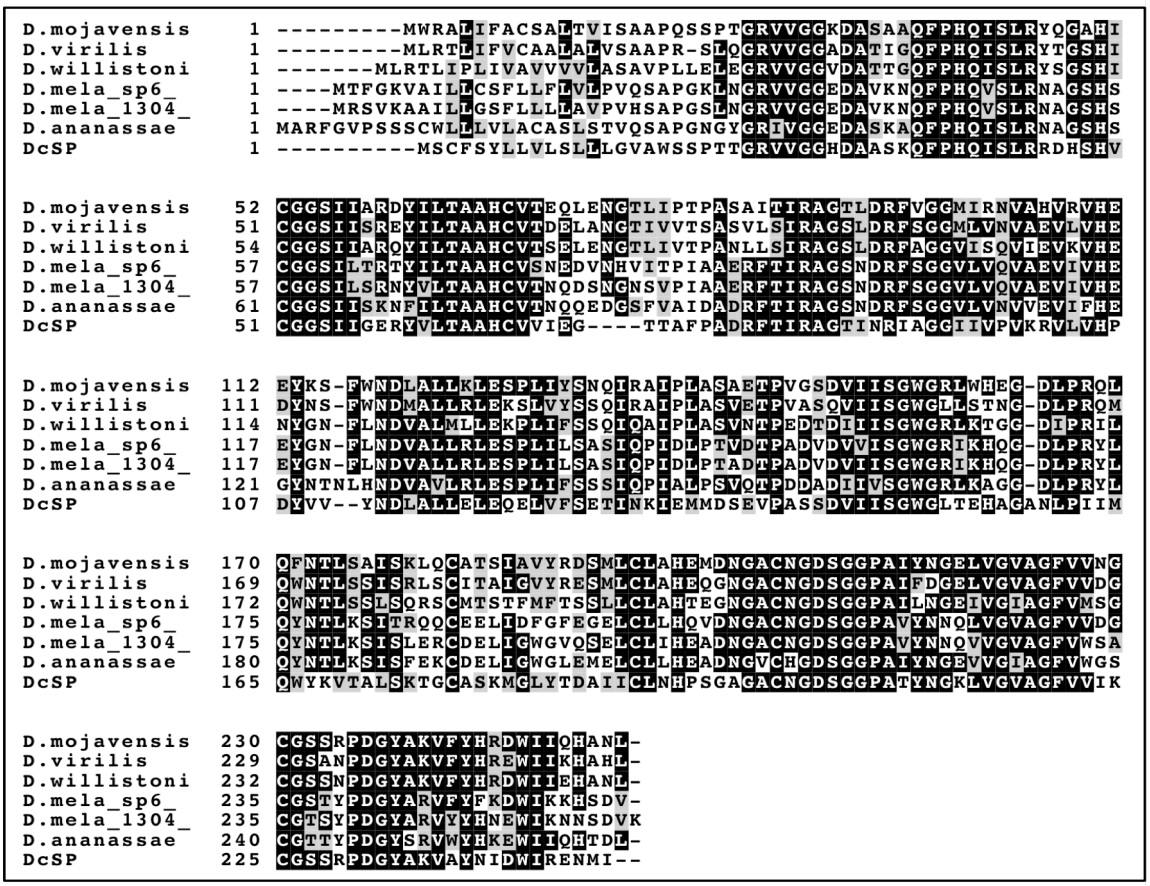


Figure 3.4 (B): CustalW-Boxshade alignment of DcSP with serine proteases from *Drosophila mojavesnsis, D. virilis, D. willistoni, D. melanogaster* (Serine Protease-6 and CG1304). Highly conserved residues are shown in black boxes and partially conserved residues in grey boxes.

3.3.2 Recombinant expression and purification of DcSP

Mature DcSP was recombinantly expressed as a TRX (Thioredoxin) fusion in *E. coli* expression vector pET32a. Thioredoxins are ubiquitous proteins of approx. 12 kD, containing catalytically active disulphide group and function in several biochemical pathways by thiol/disulphide exchange reactions (Lemaire *et al.,* 2000). Thioredoxins are structurally very stable proteins e.g. *E. coli* TRX was found to be structurally intact when analysed after hours of incubation at -80°C (Holmgren, 1985). They are also known to promote folding of fused proteins during recombinant expression. TRX in pET32a is a 327bp nucleotide long coding for 109 amino acid residues and a protein of approx.12kD that helps in high-level expression and proper folding of fused peptide. The recombinant protein produced in *E. coli,* thus was a fusion of DcSP:Trx. The expression construct for DcSP consisted of a vector derived N-terminal Trx tag

(109 amino acid residues) with 6X His tag followed by a mature coding sequence of DcSP (229 amino acid residues; SPTTGR.....RENMI) with a stop codon in the end. An enterokinase cleavage site DDDK is present between the His-tag and mature DcSP, which will release mature DcSP from Trx after treatment with Enterokinase. The details of the expression construct are shown in (Figure 3.5 A and B).

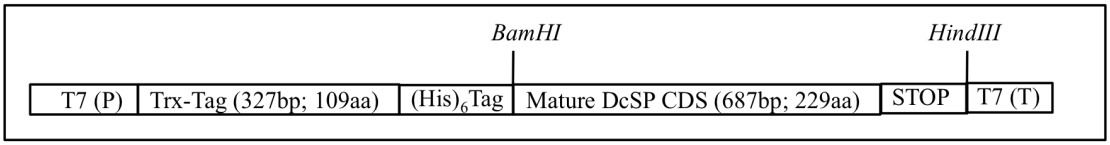


Figure 3.5 (A): Schematic representation of expression construct for DcSP in pET32a.

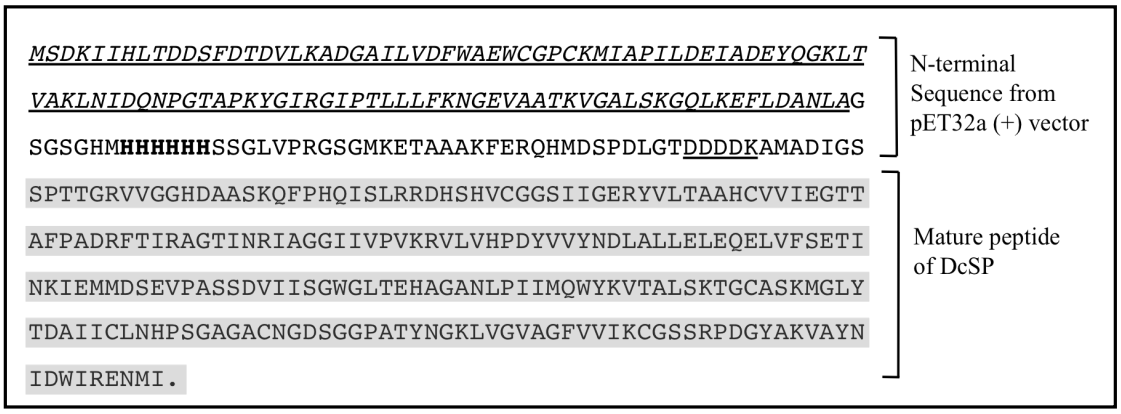


Figure 3.5 (B): Sequence details of expression construct for DcSP in pET32a. A complete coding sequence for mature DcSP containing 229 amino acid residues is inserted into MCS of an *E. coli* expression vector pET32a, between BamHI and HindIII restriction enzyme sites. Trx, the fusion partner containing 109 amino acid residues precedes DcSP with six His residues in between the fusion for detection of fusion protein and purification using metal affinity chromatography.

The predicted molecular weight of DcSP is approx. 26kD and thus the expected size of fusion DcSP:Trx is ~38kD. After induction and processing of recombinant DcSP:Trx, the samples (1) Un-induced Soluble (US), (2) Induced Soluble (IS), (3) Un-induced Insoluble (UI) and (4) Induced Insoluble (II) were analysed by SDS-PAGE followed by CBB staining. The recombinant DcSP:Trx was clearly seen in induced samples but was in insoluble fraction (figure 3.6A) migrating at approx 42kD. The presence of recombinant DcSP:Trx was confirmed by western blot using

anti $(His)_6$ antibodies and an immunoreactive band of same size was detected in induced insoluble (II) sample (figure 3.6B).

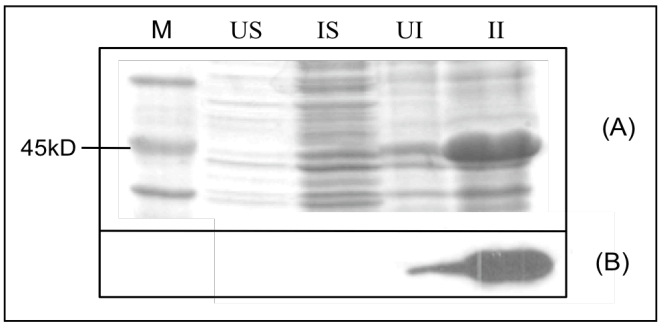


Figure 3.6 (A): SDS-PAGE followed by CBB stain and western blot analysis of DcSP:Trx fusion protein. Lane 1: molecular weight marker (SDS7), Lane 2: Un-induced Soluble fraction (US), Lane 3: Induced Soluble (IS) fraction, Lane 4: Un-induced Insoluble (UI) and (4) Induced Insoluble (II) fraction. (II) fraction shows a fusion protein of approx. 42kD which was confirmed using anti (His)6 antibodies by western blot (B)

The purification of DcSP:Trx was performed by metal affinity chromatography using a nickel-nitrilotriacetic acid (Ni-NTA) column. A standard purification profile for Ni-NTA chromatography is shown in figure 3.7. Purification samples; load (L; lysate in urea), wash (W; binding buffer wash), I_{20} (wash with binding buffer containing 20mM Imidazole) and I_{300} (binding buffer containing 300mM Imidazole) were analysed by SDS-PAGE and CBB staining (figure 3.8). Recombinant DcSP:Trx was found to be tightly bound to Ni column as no recombinant protein was seen in wash (W) sample. There was some recombinant protein seen in I_{20} that could be loosely bound to column matrix. Most of protein was eluted in almost pure form in elution buffer (I_{300}).

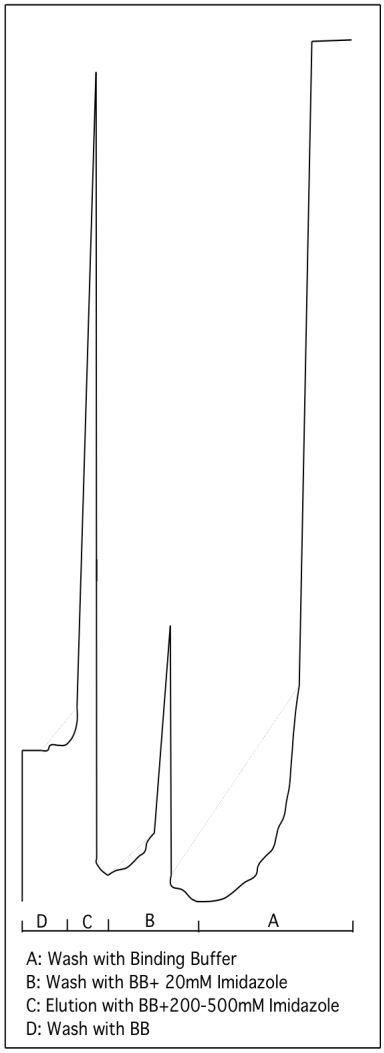


Figure 3.7: Standard Ni-NTA purification profile. Zones A to D represent different stages of elution. A: Wash with binding buffer (BB) after completion of loading for washing unbound (nonspecific) proteins, B: wash with BB containing 20mM Imidazole for washing of loosely bound proteins, C: wash with BB containing 200 to 500mM Imidazole for elution of bound recombinant proteins and D: wash with BB to equilibrate column for further loading. Peaks represent elution of protein which is measured as absorbance at 280nm with 0.5 sensitivity.

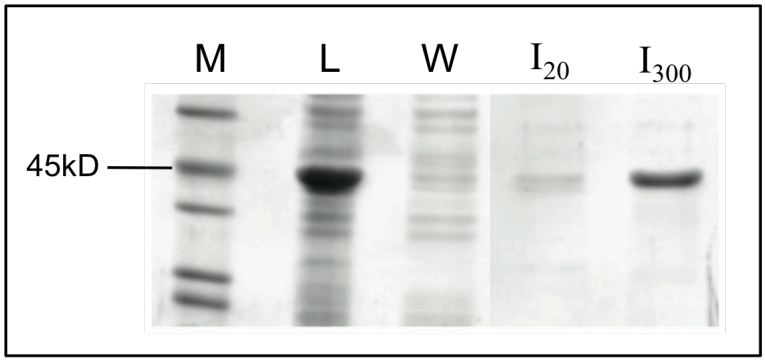


Figure 3.8: SDS-PAGE followed by CBB staining of Ni-NTA purified fractions of recombinant DcSP+Trx fusion. Lane M: Molecular weight marker (SDS7), Lane L: Load fraction; cell lysate resuspended in 6M urea, Lane W: wash fraction with binding buffer, Lane I_{20}: Wash fraction with BB containing 20mM Imidazole and Lane I_{300}: Elution fraction with B containing 300mM Imidazole. Purified recombinant DcSP:Trx fusion protein is seen at approx. 42 kD.

3.3.3 Efforts to refold DcSP in vitro:

Purified DcSP:Trx fusion was treated with Enterokinase, a serine protease that recognizes '-Asp-Asp-Asp-Asp-Lys-X' and cleaves between Lys and X very specifically. This cleavage helped release of a mature DcSP from the Trx fusion as shown in figure 3.9. The purified mature DcSP (from 500ng to 100μg) when analysed for activity using 20mM Tris buffer pH 8.0 did not show any activity with EnzChek or other substrates. Bovine trypsin, used as a positive control for the assays in a range of (50ng to 1μg) showed activity at all concentrations. The activity profile of DcSP was almost similar to an assay mixture of no enzyme control, with no hint of any activity even at the concentration of 100μg (Figure 3.10). Lack of activity in DcSP was thought to be due to misfolding of DcSP during expression or purification and efforts were taken to refold DcSP *in vitro* by dilution method. There was very low recovery of the refolded DcSP in soluble form in the refolding experiment and it did not show any activity.

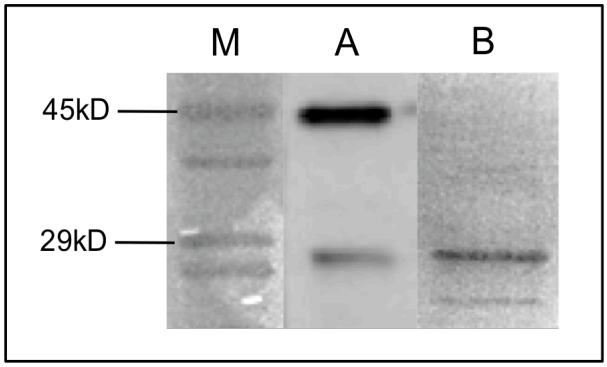


Figure 3.9: Enterokinase digestion of DcSP:Trx fusion to release mature DcSP from Trx. Lane 1: Molecular weight marker (SDS7), Lane2: partially digested DcSP:Trx fusion showing both DcSP:Trx (approx. 42kD) and released DcSP (approx 25kD) and Lane 3: completely digested mature DcSP protein.

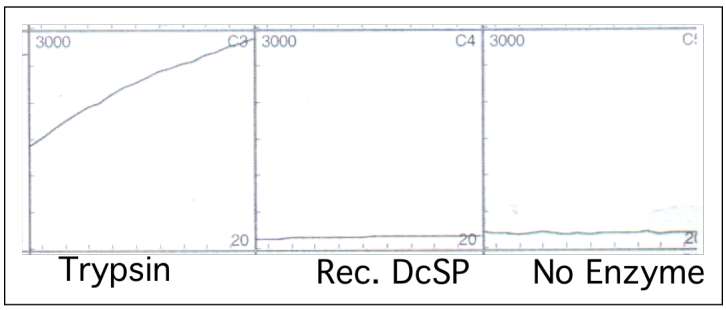


Figure 3.10 Activity assay for recombinant and purified DcSP (100ug) using EnzChek substrate. Bovine trypsin (100ng) was used as positive control and no enzyme as background control. Activity curve is shown for a part of assay, whole assay consisted of different concentrations of trypsin (50ng to 1ug) and DcSP (500ng to 100ug).

During heterologous expression of protein in *E. coli*, the formation of insoluble aggregates by inclusion body (IBs; dense aggregates of misfolded polypeptide) formation limits high yield of recombinant protein (Chrunyk *et al.,* 1993). Proteins incorporated into IBs possess elements of native structure but this is usually considered to be minimal (Middelberg, 2002). In case of recombinant enzymes, the protein needs to be released from IBs and 'refolded' or 'renatured' to give its native 3D structure to characterise it as an active enzyme. However, performing this process *in vitro* is critical and always has limited success. There are two main methods of refolding proteins *in vitro* (1) refolding by dilution and (2) column refolding with different levels of success depending of type of protein being refolded. In dilution

method protein refolding is initiated by a reduction in denaturant concentration and, in oxidative protein refolding, by altering the redox environment to enable disulphide bond formation (Middelberg, 2002). Protein samples with high denaturing conditions are delivered into large excess volume of refolding buffer. Dilution brings the unfolded sample into a rapid collapse, bypassing the intermediate denaturant concentration. But protein refolding is not a single reaction and always competes with other reactions, such as misfolding and aggregation, leading to inactive proteins (Tsumoto *et al.,* 2003). There are several factors to be considered while refolding a protein and one of them and important is to understand the structure of protein being refolded and others are composition of refolding buffer, effect of additives etc. Human trypsinogen was expressed as IBs in *E. coli* and successful refolding by dilution method to active form was obtained (Hohenblum *et al.,* 2004). As recombinant DcSP could not be obtained in active form, total protein extracted from gut tissue as a protease mixture was used for all *in vitro* assays.

3.4 Cysteine Proteinase (DcCathL) from D. coarctata gut:

3.4.1 Cloning of DcCathL:

A complete cDNA sequence encoding the *D. coarctata* cathepsin L-like cysteine proteinase; DcCathL was isolated from RNA purified from wheat bulb fly larval guts by a PCR based strategy. An initial fragment was amplified from cDNA produced by reverse transcription using degenerate primers directed towards conserved amino acid sequences in cathepsin L-like cysteine proteinases. Complete coding sequence of DcCathL was obtained using 5' and 3' RACE. An overview RACE performed for DcCathL is shown in figure 3.11 (A). A partial coding sequence of DcCathL amplified was 489bp long, and 5' and 3' RACE products were approx. 1000bp each. Actual PCR products analysed and cloned from agarose gels are shown in figure 3.11 (B). 3' RACE product had a long UTR with poly A tail (approx. 850bp) while 5' RACE product had a shorter UTR (approx. 200bp). All RACE products when aligned using Sequencher software gave a complete coding sequence of DcCathL. To confirm this, the complete coding sequence was amplified as a single PCR product using gene specific primers.

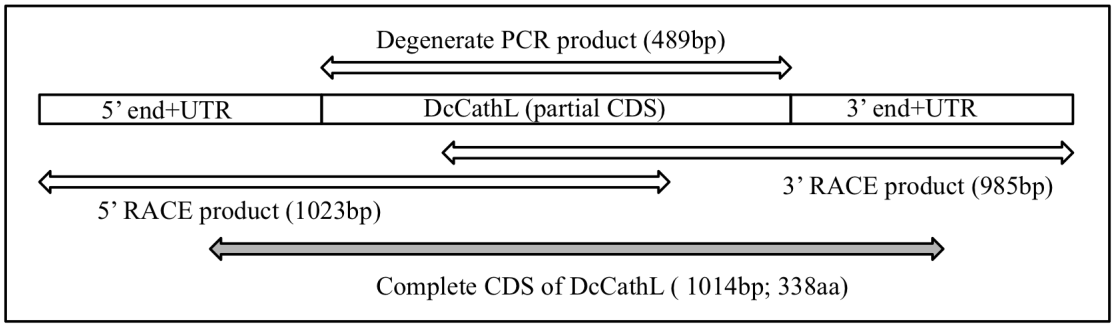


Figure 3.11 (A): Overview of RACE for *D. coarctata* cysteine proteinase (DcCathL)**.**

.

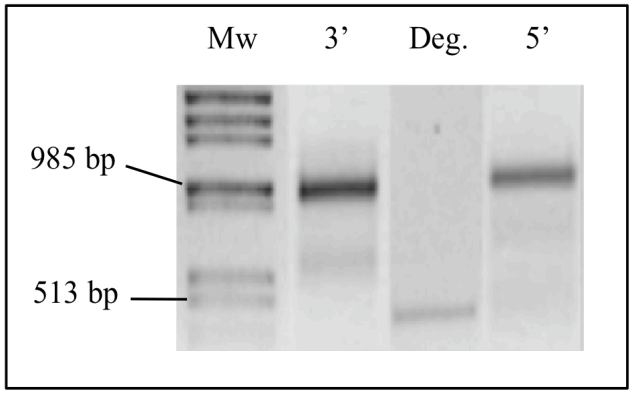


Figure 3.11 (B): Actual PCR amplification products analysed on 1% agarose gel. Lane 3': product of 3' RACE, Lane Deg.: Product of degenerate PCR and Lane 5': product of 5' RACE

3.4.2 Sequence analysis of DcCathL

The cDNA of DcCathL contains a single long open reading frame of 1014 nucleotides, which encodes a predicted polypeptide of 338 amino acids (figure 3.12). The DcCathL amino acid sequence was identical to the cathepsin L-like proteinases from *D. radicum* (cabbage root fly) and *S. peregrina* (flesh fly); these proteins have been described previously (Hegedus *et al.,* 2002; Homma *et al.,* 1994) and, in conjunction with domain and signature databases, can be used to predict features of DcCathL. SignalP prediction for signal peptide in DcCathL showed presence of a 16-residue long (MRTVLALLALVAFVQA) signal peptide. The pro-region of the protein (InterPro domain IPR013201) extends from residue 17 to 120 and the mature peptide (InterPro domain IPR000668) from residues 121 to 338. The cleavage site between the propeptide and the mature protein is predicted to occur in the sequence.....PANVQ^VPKAV...., resulting in Val^{121} being the N-terminal residue of the mature enzyme. The predicted catalytic triad residues (InterPro signature IPR000169) are Cys^{145}, His^{284} and Asn^{305}; Gln^{140}, which is also involved in catalysis

through formation of the oxyanion hole, was also conserved (Rawlings and Barrett, 1994). Ala^{332} is the predicted specificity-determining residue in the substrate-binding pocket, as is the case generally for cathepsin L-like enzymes. The propeptide region contains a unique N-glycosylation site, at Asn^{95}. Comparisons between the predicted sequence of DcCathL and the global protein database suggested that DcCathL belonged to a sub group of cathepsin L-like enzymes found in Diptera and some other orders of insects. BlastP pairwise comparisons showed that a series of groups of sequences could be defined in decreasing order of similarity to DcCathL, which included cathepsin L-like proteins as follows (most similar to least similar): in other dipteran species; in other insect orders; in other arthropods (Crustacea, mites and ticks); in the Hemipteroid Assemblage; in higher animals; other insect cysteine proteinases. Multiple sequence alignments supported the groupings identified by the pairwise analyses.

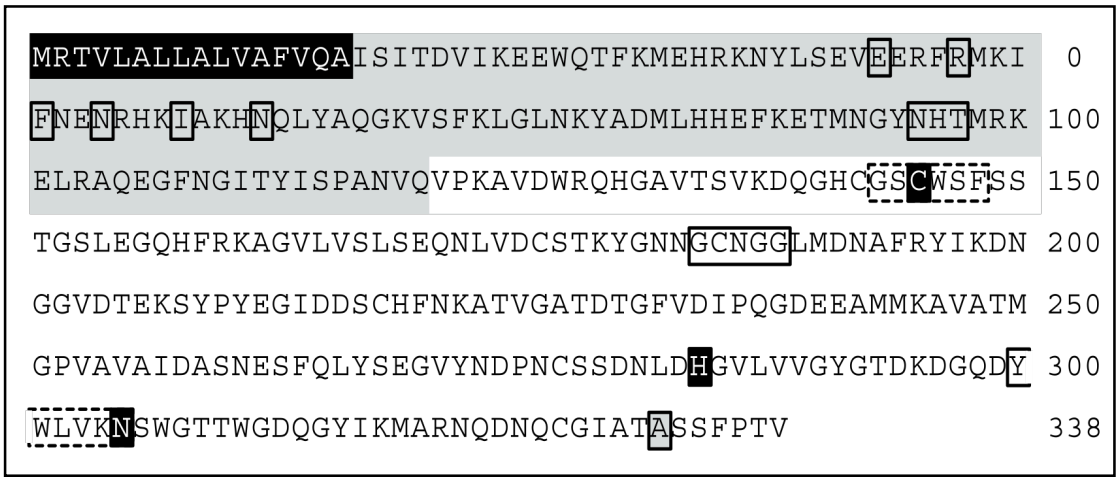


Figure 3.12: Amino acid sequence analysis of *D. coarctata* cysteine proteinase (DcCathL); Predicted signal peptide (residues 1–16) is given in white text on black background, predicted pro-region (residues 17–120) is in grey background, predicted mature polypeptide is unshaded. Catalytic triad residues (Cys145, His284 and Asn305) are indicated by inverted type. The predicted specificity-determining residue in the substrate-binding pocket (A332) is indicated by a shaded box. The pro-region ERFNIN motif (Karrer *et al.,* 1993) represented by Glu43, Arg47, Phe51, Asn54, Ile58 and Asn62 and the potential N-glycosylation site in the pro-region (–NHT–; Asn95–Thr97) are boxed. The GCNGG motif (Karrer *et al.,* 1993;Gly184–188) in the mature polypeptide is also boxed. Dotted boxes indicate regions of conserved amino acid sequence used to construct PCR primers.

A clustal analysis was used to produce a phylogenetic tree of proteins in insects, including all proteins more similar to DcCathL than the group of non-insect arthropod cathepsin L-like proteins. These proteins were compared with the complete set of cathepsin-like cysteine proteinases (Interpro Peptidase C1A signature, IPR013128) in *D. melanogaster*. The resulting tree is presented in figure 3.13, which shows that DcCathL is more closely related to a product of Cp-1 in *D. melanogaster* (Accession number **CG6692**). The other *D. melanogaster* cysteine proteinases are less related to **CG6692** than DcCathL, or any of the other proteins in a grouping including homologues of CG6692 in other *Drosophila* spp., and in other dipteran species including mosquitoes, flesh fly, and root fly, as well as similar proteins in Coleoptera, Hymenoptera and Lepidoptera. The analysis can be extended to include cathepsin L-like proteinases in non-insect arthropods, insects of the Hemipteroid assemblage and nematodes in a larger sub-group of related sequences. However, it is unclear whether this larger grouping reflects homology, since cathepsin L-like enzymes in non-insect arthropods are more similar to the **CG6692** sub-group than cathepsin L-like enzymes in other insect orders (Hemipteroid assemblage). This analysis complements the earlier work of Hegedus *et al.,* (2002) which identified the *D. radicum* cathepsin L-like cysteine proteinase and other similar proteinases as a members of the papain family. The sequence similarity of the **CG6692** sub-group of cysteine proteinases may reflect a common functional role (or roles) for these homologous proteins in different organisms. Involvement in tissue remodeling has been identified as a common theme in studies of the *S. peregrina* and *D. radicum* enzymes (Homma *et al.,* 1994; Hegedus *et al.,* 2002), where appropriate activity and patterns of expression both *in vitro* and *in vivo* (including midgut metamorphosis) have been demonstrated. Specific involvement of a lepidopteran homologue in moulting (Liu *et al.,* 2006) is consistent with this conclusion. However, **CG6692** in *D. melanogaster* is expressed throughout the insect, and in both larvae and adults (data from http://flyatlas.org), which suggests a less specific role. Further studies will be necessary before these enzymes can be categorically designated as ‘‘tissue remodelling proteinases’’.

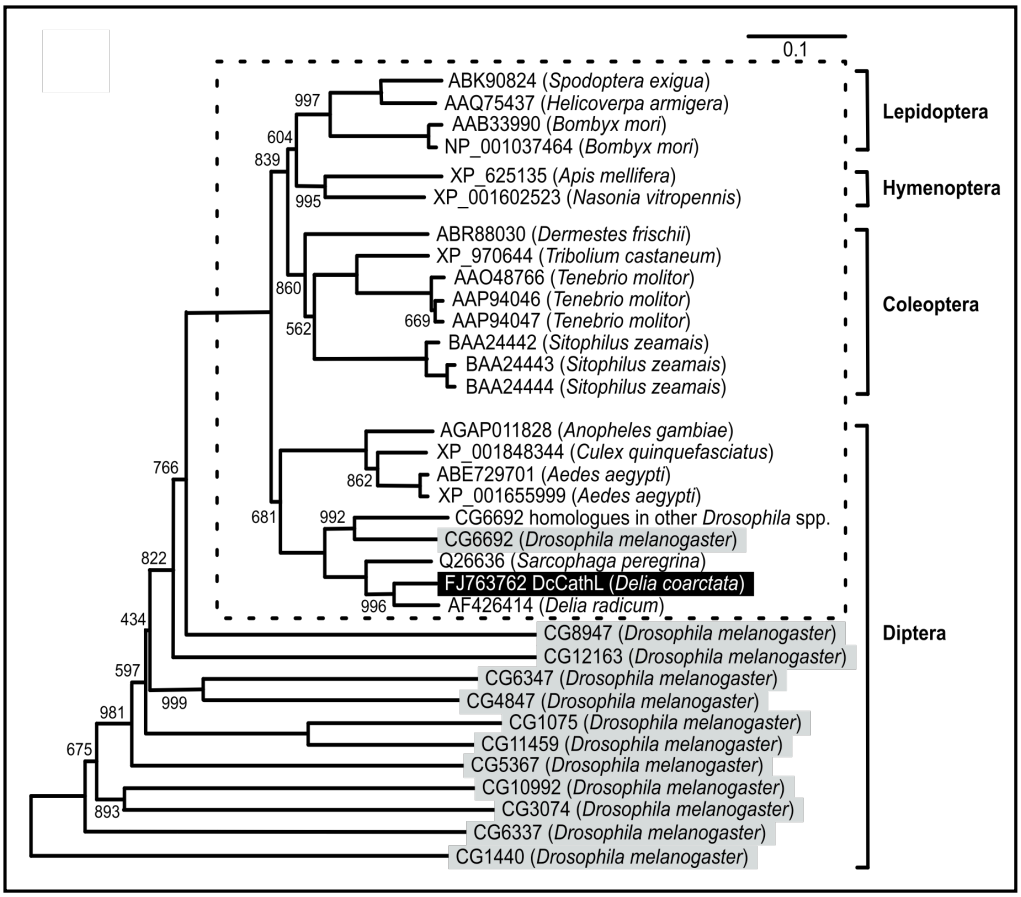


Figure 3.13: DcCathL sequence alignment and phylogeny tree. Identification of a sub-group of insect cysteine proteinases containing DcCathL. Proteins similar to DcCathL were identified by Blast pairwise sequence alignments (NCBI Blast server; http://www.ncbi.nlm.nih.gov/blast), and all insect sequences more similar to DcCathL than a group of sequences from Crustacea and other arthropods were selected. Proteins predicted by the complete genome of *Drosophila melanogaster* containing the sequence signature for cathepsin-type cysteine proteinases were obtained from Interpro (http://www.ebi.ac.uk/interpro; Peptidase C1A signature, IPR013128) and Flybase (http://www.flybase.org); alternate protein sequences from the same gene were omitted. A Clustal multiple sequence alignment was carried out, and the resulting alignment was plotted as a Neighbour-Joining tree, verified by bootstrap analysis (1000 replicates). Bootstrap values are given for all junctions where the value was <1000. Branch lengths are drawn to the same scale. *D. melanogaster* sequences are indicated by shaded boxes; wheat bulb fly DcCathL is indicated by inverse type. Homologues of *D. melanogaster* CG6692 in other Drosophila spp. are not shown individually; these comprise the following accessions: GI21205, GJ20806, GA25021, GL17172, GK19626, GG20414, GM21500, GD10995, GE12574, GF13722. All were of equal or greater similarity to CG6692 than the branch shown.

3.4.3 Recombinant expression and purification of DcCathL:

The expression construct for DcCathL contained a full coding sequence of the proenzyme inserted into pGAPZαB expression vector, in frame with the yeast a-mating factor N-terminal secretary signal. The translation product is predicted to undergo co- and post-translational proteolysis, resulting in a secreted polypeptide containing the complete pro-protein sequence, with an extra Alanine (Ala) residue at the N-terminus (from restriction site in expression vector) and a vector derived C-terminal extension of 27 residues containing a Myc antigenic determinant and a $(His)_6$ C-terminal tag (figure 3.14 A and B). *P. pastoris* transformants were screened for expression by western blotting of culture supernatants, using Anti-(Myc) antibodies. Clones expressing DcCathL showed two immunoreactive bands on the resulting blot, at approx. 50 and 32 kDa (figure 3.16, lane2). From previous results, the larger band can be assigned to un-processed pro-protein, with the smaller band representing cleaved mature DcCathL (Philip *et al.,* 2007). The calculated molecular weight of pro-DcCathL is 35.9 kDa, and mature DcCathL 23.4 kDa; the discrepancy between calculated and observed molecular weights for these cathepsin L-like proteins has been observed previously (Bown *et al.,* 2004; Philip *et al.,* 2007). Interestingly, recombinant DcCathL also reacted with antibodies raised against flesh fly cathepsin L-like enzyme (ScathL).

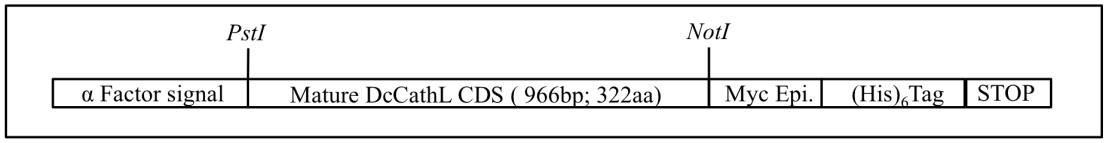


Figure 3.14 (A): Schematic representation of expression construct for DcCathL in pGAPZαB.

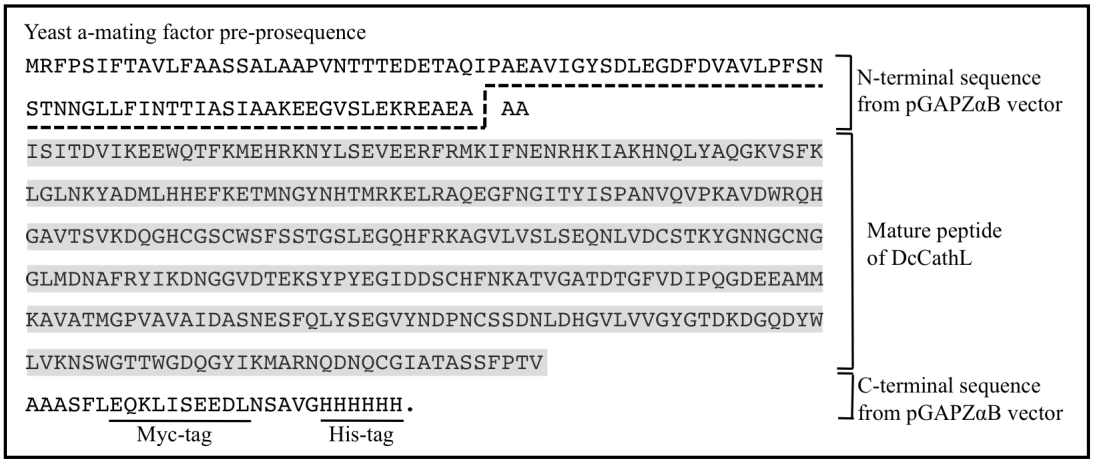


Figure 3.14 (B): Sequence details for expression construct of DcCathL in pGAPZαB. A complete coding sequence for mature DcCathL containing 322 amino acid residues is inserted into MCS of yeast expression vector pGAPZαB, between PstI and NotI restriction enzyme sites. α-factor signal sequence containing 89 amino acid residues, helps secretion of recombinant protein into the culture media. C-terminus contains a Myc epitope (for immunodetection of recombinant protein) and six His tag with residues (for immunodetection of recombinant protein and purification by metal affinity chromatography)

Recombinant protein expression was scaled up using a laboratory scale fermentor. The culture supernatant was purified by ion exchange chromatography on S-Sepharose and a yield of approx. 20 mg DcCathL per litre of culture supernatant was obtained. Elution profile of a standard S-sepharose column chromatography is shown in figure 3.15. Purified fractions when analysed on SDS-PAGE showed only a mature peptide band (approx. 32 kDa) (figure 3.16, lanes 3 and 4), suggesting that autocatalytic cleavage of the proprotein had occurred during expression or purification. Recombinant DcCathL was confirmed by western blot using anti Myc and anti $(His)_6$ antibodies (figure 3.16, lanes 5 and 6). The discrepancy in the migration of DcCathL was also observed for ScathL expression in insect cells, where complete conversion of proprotein to mature protein took place after 24 hours at 4°C (Homma *et al.,* 1994). The activation of pro-cathepsin L by limited proteolysis can occur autocatalytically (Salminen and Gottesman, 1990), and previous experiments in which insect cathepsin L-like proteinases have been expressed in *P. pastoris* have resulted in purification of the mature protein rather than proprotein (Bown *et al.,* 2004; Philip *et al.,* 2007).

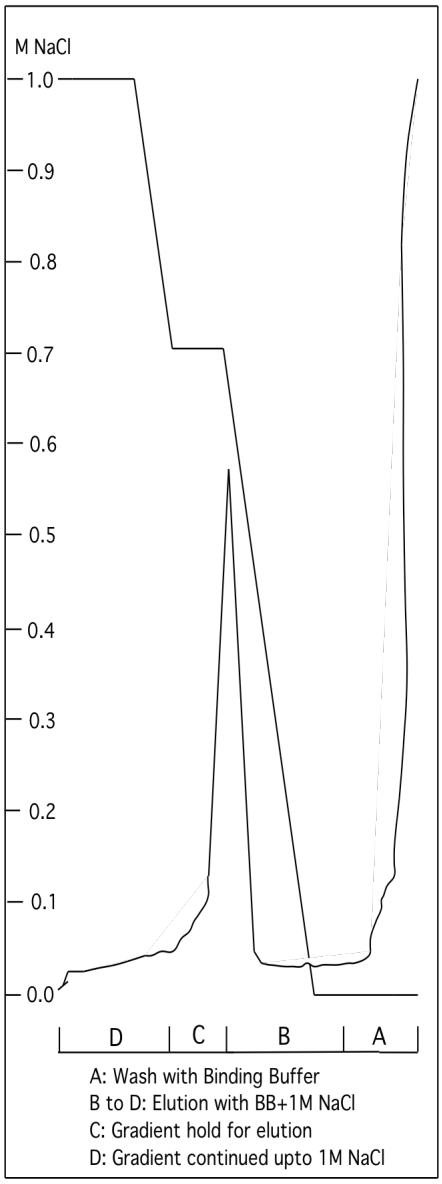


Figure 3.15: Standard S-Sepharose purification profile. Zones A to D represent different stages of elution. A: Wash with binding buffer (BB) after completion of loading for washing unbound (nonspecific) proteins, B to D: wash with BB containing 1M NaCl over a gradient from 0 to 1M NaCl. Peak represents elution of recombinant protein which is measured as absorbance at 280nm with 0.5 sensitivity.

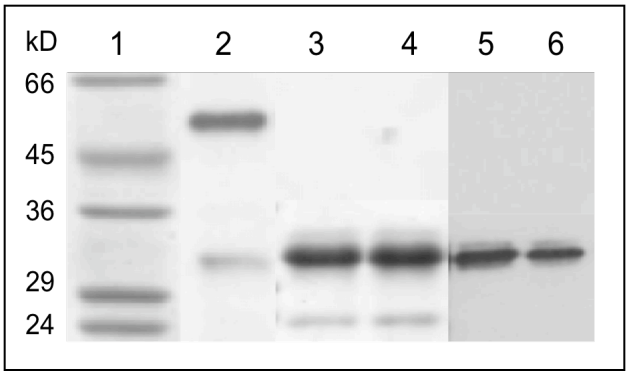


Figure 3.16: SDS-PAGE and western blot followed by immunodetection of DcCathL produced as a recombinant protein in *P. pastoris*. Lane 1: Molecular weight marker, Lane 2: western blot of culture supernatant, showing immunoreactive DcCathL, propeptide (~50 kDa) and mature peptide (~32 kDa). Lanes 3 and 4: SDS-PAGE of recombinant DcCathL after purification by ion exchange chromatography. Lanes 5 and 6 western blot of purified DcCathL using Anti-Myc antibodies and Anti-ScathL antibodies (respectively).

3.5 Biochemical Characterisation of DcCathL:

3.5.1 Hydrolytic activity of recombinant DcCathL:

A recombinant cathepsin L-like cysteine proteinase from *S. peregrina* (ScathL) did not show any hydrolytic activity towards synthetic proteinase substrates after purification at pH 4.0, under conditions where reducing agents were not added to buffers. However, the activity was restored by the addition of thiol compounds (either DTT or Glutathione), suggesting that the enzyme had become oxidised during culture growth or purification (Philip *et al.,* 2007). Both DTT and reduced glutathione have a potential role in reducing disulphides in a protein. DcCathL was also inactive after purification and addition of either DTT (0.1mM) or glutathione (1mM) caused immediate activation of DcCathL. Reduced glutathione was used as reducing agent for activation of DcCathL, as the activity was more stable in presence of glutathione than DTT. Optimum concentration of glutathione required for activation of DcCathL was 1.25 mM (figure 3.17), although activity decreased over time when incubated at room temperature even in presence of reducing agent (time for loss of half of activity approx. 10 min in presence of 1 mM glutathione) (figure 3.18). This loss of activity was interpreted as self-degradation of enzyme due to high activity and absence of substrate. The recombinant cathepsin L-like enzyme BmCL1 from the tick *Boophilus microplus* also required activation, which was carried out by incubating a proenzyme at 37°C for 1 hour in 25 mM sodium-acetate pH 3.5 and 5 mM DTT (Renard *et al.,*

2000). Activated DcCathL hydrolysed the synthetic peptide substrates Z-Phe-Arg-AMC and Z-Arg-Arg-AMC, with a higher maximal rate towards the former substrate. The role of reducing compounds in activation of cathepsin L-like proteases was characterised for ScathL (Philip *et al.,* 2007). Activation of ScathL by reducing agents required a significant time. DTT present at 0.1mM, activity did not reach 75% of the maximum value until approx. 1hour after addition of the reducing agent and activity was maximal after 1hour when enzyme was activated with 1.0mM glutathione (Philip *et al.,* 2007). Activity declined with time after reaching a maximum value, with less than 50% of activity present after 6hours when DTT was used for activation. The enzyme activity was more stable in the presence of 1.0mM glutathione, with activity only falling to approx. 80% of maximum after 6 hours. Enzyme from which reducing agent was removed by desalting, lost activity rapidly (Philip *et al.,* 2007). In contrast to this DcCathL was activated immediately after addition of reducing agents showing decrease in activity with increase in time of incubation.

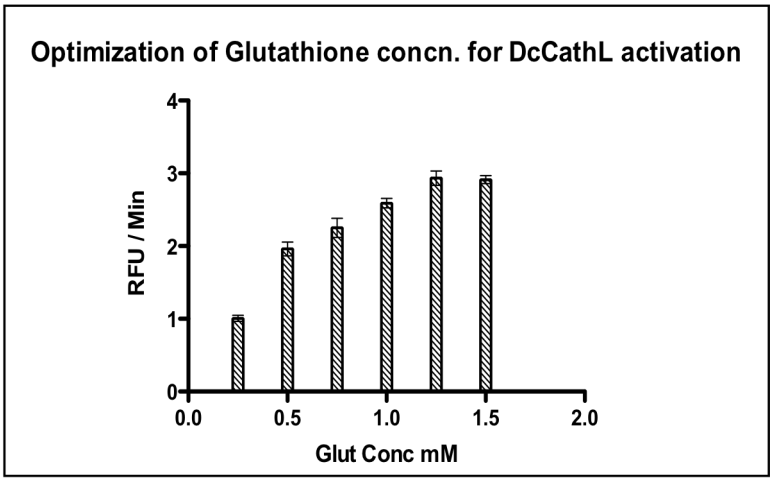


Figure 3.17: Optimization of Glutathione concentration for activation of DcCathL. Assay carried out in triplicate using 30mM sodium acetate buffer pH 5.5, glutathione (0.25 to 1.5 mM) was added for activation of DcCathL and readings were taken immediately, optimum concentration of glutathione required was 1.25mM. Error bars indicate $\pm$ SE (n=3).

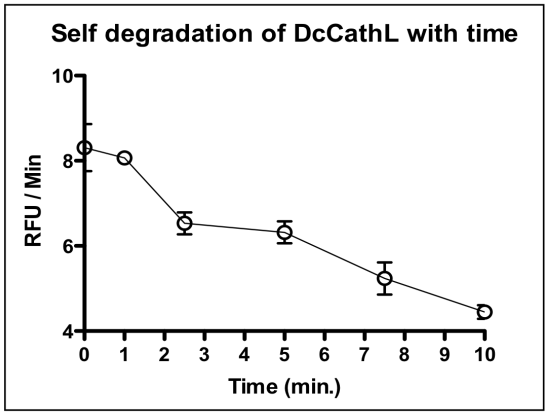


Figure 3.18: Self-degradation of DcCathL with time. DcCathL showed self-degradation in presence of 1mM glutathione and absence of any substrate. DcCathL was highly active immediately after addition of glutathione. After 10 minutes only minimal activity of DcCathL was detected. Points indicate mean values ± SE (N=3).

Hydrolysis of peptide substrates by DcCathL showed non-linear kinetics, with declining rates even at high substrate concentrations when assays were performed for long periods. This was considered to be a result of enzyme instability. However when the assays were performed for a short duration reliable rate measurements were possible. A standard Michaelis–Menten analysis for kinetics of DcCathL hydrolysis at optimal pH (see below) gave a Km value of 2.0 ±1.1 X 10^{-4} M using Z-Phe-Arg-AMC as substrate (figure 3.19), similar to values previously determined for the tick cathepsin L BmCL1. Km values for hydrolysis of Z-Phe-Arg-AMC and Z-Arg-Arg-AMC were 1.88 X 10^{-4} M and 2.36 X 10^{-4} M respectively (Renard *et al.,* 2000).

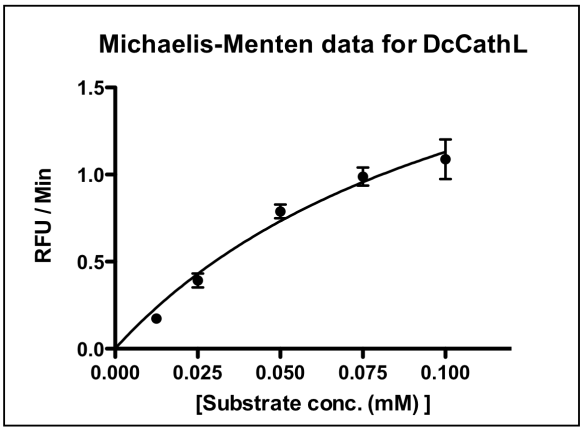


Figure 3.19: Michaelis-Menten (Km) curve for DcCathL with Z-Phe-Arg-AMC substrate. Assay was carried out using 30mM sodium acetate buffer pH 5.5, 1mM glutathione. Points indicate mean values ± SE (N=3). Km Curve was fitted using Prism Software.

3.5.2 Effect of pH and ionic strength on DcCathL activity:

Activity of recombinant cathepsin L-like proteinase from *S. peregrina* was found to be highly dependent on pH and ionic strength (Philip *et al.,* 2007). Assays carried out using the isoionic AMT buffer system over the pH range 3.0–8.0 showed a similar pH dependence for DcCathL with optimum activity at pH 4.5 and a decrease to zero activity at pH 7.5 (Figure 3.20). No activity was seen above pH 7.5, and the enzyme was irreversibly inactivated by exposure to pH 7.5 and above, which was similar to ScathL (Philip *et al.,* 2007). The cathepsin L-like recombinant enzyme from *B. microplus* showed optimum activity at pH 5.5 and no activity was seen above pH 7.0 (Renard *et al.,* 2000).

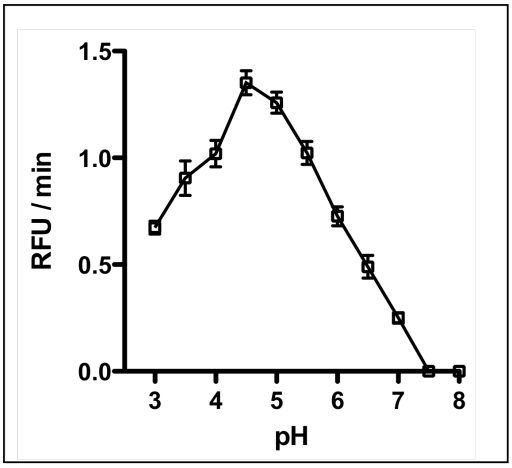


Figure 3.20: pH curve for activity of DcCathL. AMT buffer system and 1mM glutathione was used to determine the optimum pH for DcCathL activity**.** DcCathL was highly active at pH 4.5 to 5.0 with no activity above pH 7.5. Points indicate mean values ± SE (N=3).

The activity of human cathepsin L was found to be highly dependent on pH and ionic strength where activity decreased markedly with increased ionic strength (Dehrmann *et al.,* 1995). Activity of mammalian cathepsin L was found to be sensitive to pH and ionic strength (Kirschke *et al.,* 1998). Human cathepsin L is rapidly and irreversibly inactivated at pH values above 7.0 (Turk *et al.,* 1993). However our results showed that ionic strength had no significant effect of the activity of DcCathL as the activity of DcCathL was seen persistent up to 2 M NaCl (Figure 3.21). Structural analysis of cathepsin L-like cysteine proteinases suggests that stability of enzyme is regulated by a series of interactions involving charged amino acids in the substrate binding cleft region (Turk *et al.,* 1994). The substrate cleft or "zipper region" is maintained by a

series of synergistic interactions. Increase in pH or ionic content interferes with these charge-based interactions causing a progressive destabilization of this region and resulting in protein denaturation followed by loss of activity, in an "unzipping" action. DcCathL sensitivity to loss of activity at higher pH may constitute another mechanism for regulation of the activity of this proteinase in developmental processes such as those described by Homma *et al.,* (1994).

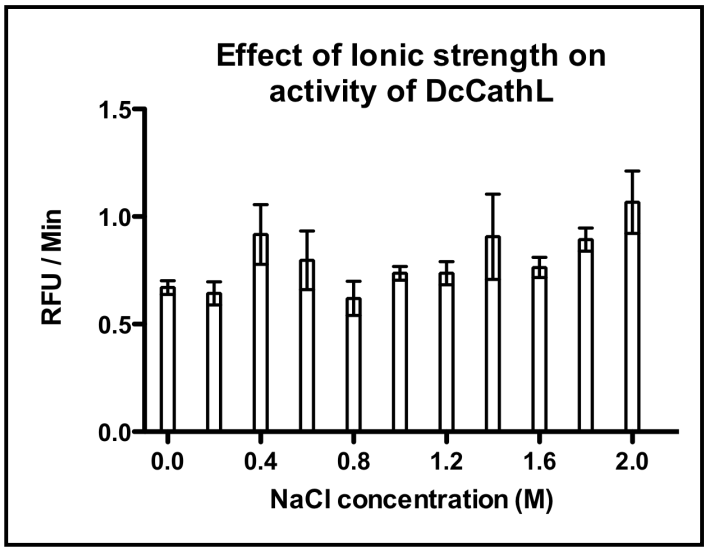


Figure 3.21: Effect of ionic strength on activity of DcCathL. 1mM glutathione and NaCl from 0 to 2M was added to standard assay buffer (30mM sodium acetate, pH 5.5) did not show any significant effect on DcCathL activity. Points indicate mean values ± SE (N=3).

3.5.3 Inhibition of DcCathL activity with E-64:

E-64 [trans-epoxysuccinyl-L-leucylamido-(4-guanidino)butane] is an irreversible, potent and highly selective cysteine protease inhibitor. It does not inhibit serine proteases (except trypsin) like other cysteine protease inhibitors, leupeptin and antipain (Barrett *et al.,* 1982). E-64 has been used as an active site titrant for many cysteine proteases (Barrett *et al.,* 1982; Buttle and Barrett 1984). The trans-epoxysuccinyl group in E-64 irreversibly binds to an active thiol group of many cysteine proteases such as papain, actinidase, and cathepsins B, H and L (Katunama and Kominami, 1995; Sreedharan *et al.,*1996) to form a thioether linkage. E-64, was an effective inhibitor of DcCathL activity, with the enzyme showing about 25% residual activity at 1 μM E-64 concentration with a further gradual decrease in activity up to 5 μM E-64 (figure 3.22). However, the reagent did not act as an active

site titrant, as was the case for other cysteine proteinases, and residual activity was present even when E-64 was present in large molar excess. Recombinant BmCL1 from *B. microplus* showed 100% inhibition with 50 μM E-64 concentration (Renard *et al.,* 2000). The residual activity of DcCathL was almost constant above 5 μM up to 50 μM suggesting saturating inhibition.

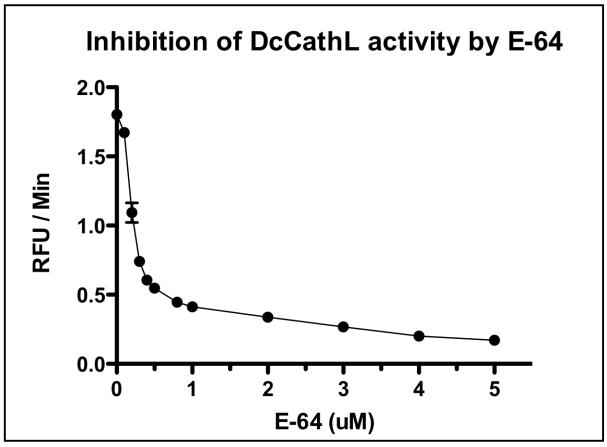


Figure 3.22: Inhibition of DcCathL activity with an irreversible inhibitor of cysteine proteases, E-64 [(2S,3S)-3-(N-{(S)-1-[N-(4-guanidinobutyl)-carbamoyl]3-methylbutyl} carbamoyl) oxirane-2-carboxylic acid)]. Points indicate mean values ± SE (N=3).

3.6 Inhibition of gut protease activity of D. coarctata

Serine proteases like DcSP and Cysteine proteases like DcCathL could play an important role in *D. coarctata* gut and are potential targets to consider for crop protection strategy. Inhibiting the activity of either of them or both using plants' serine or cysteine proteinase inhibitors (PIs) could be a potential approach to confer resistance to wheat against *D. coarctata.* To assess the role of host PIs on these proteases, two proteinase inhibitors subtilisin/chymotrypsin inhibitor (WSCI) and cysteine proteinase inhibitor (WCPI) were cloned and recombinantly expressed using *P. pastoris* expression system. Both purified PIs were active and inhibited their target proteases. The details of cloning, expression, purification and inhibition are discussed in **chapter 5**. WSCI and WCPI showed inhibition of gut proteolytic activity of *D. coarctata in vitro* (figure 3.23). Total protein extract (10ug) prepared from dissected entire guts was used for activity assays. WSCI inhibited activity of gut proteases by about 85% at the concentration of 8μg while WCPI required 8μg protein to inhibit

40% of gut protease activity. Ovalbumin used as control protein did not show any inhibition of proteolytic activity, which shows inhibitory activity of recombinant PIs.

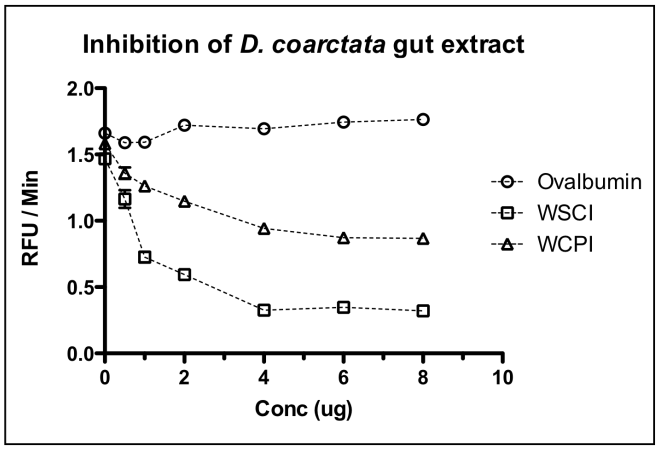


Figure 3.23 Inhibition of *D. coarctata* gut protease activity by recombinant wheat proteinase inhibitors; Subtilisin/ Chymotrypsin inhibitor (WSCI) and cysteine proteinase inhibitor (WCPI). Points indicate mean values ± SE (N=3).

Both PIs also showed inhibition of recombinant DcCathL activity *in vitro* (figure 3.24). WCPI inhibited activity of 1µg purified DcCathL by about 50% at the concentration of 2.5µg whereas WSCI showed almost complete inhibition of DcCathL at the same concentration. WSCI being an inhibitor of bacterial subtilisin and chymotrypsin inhibitor showed better inhibition of DcCathL than WCPI which is a specific inhibitor of cysteine proteases.

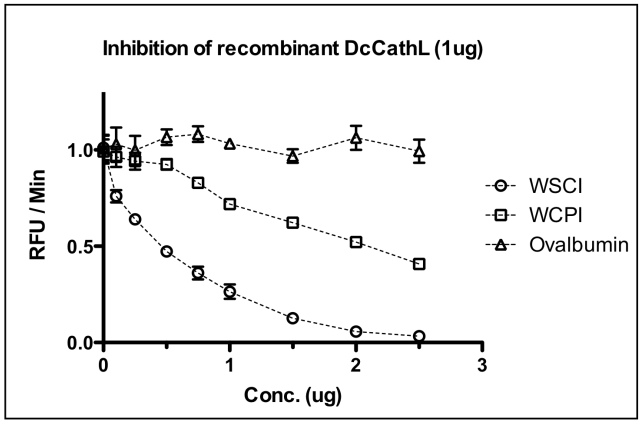


Figure 3.24 Inhibition of recombinant cysteine protease (DcCathL) activity by recombinant wheat proteinase inhibitors; Subtilisin/ Chymotrypsin inhibitor (WSCI) and cysteine proteinase inhibitor (WCPI). Points indicate mean values ± SE (N=3).

WSCI showed greater range of inhibition by inhibiting its target proteases bacterial

subtilisin, bovine chymotrypsin (as discussed in chapter 5) and DcCathL like cysteine protease. Feeding bioassay experiments with recombinant WSCI to *D. coarctata* showed interesting results. Purified recombinant WSCI at the concentration of 5µg was fed to neonates of *D. coarctata* and survival of larvae was monitored for a week. Larvae fed on recombinant WSCI showed more mortality as compared to larvae that were fed on control diet. By the seventh day of assay WSCI fed larvae showed only 30% survival where as larvae fed on control diet showed about 80% survival (figure 3.25). These results could suggest that endogenous defence proteins like WSCI may be prospective candidates to be considered for protection of wheat against *D. coarctata.*

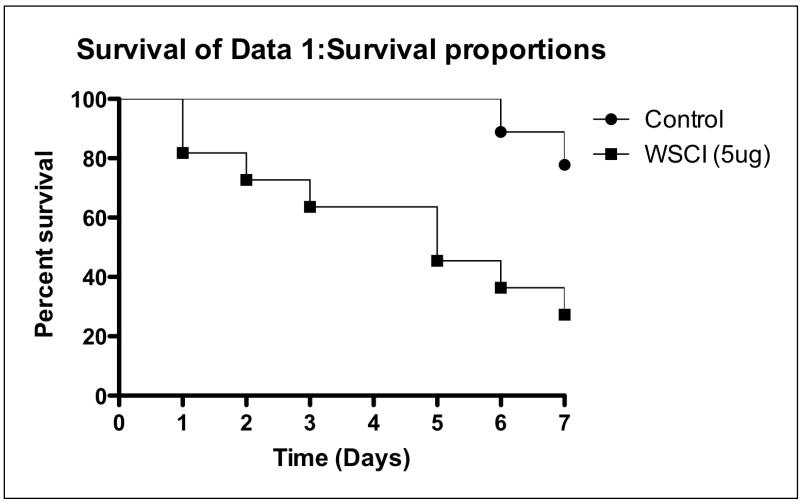


Figure 3.25 Effect of recombinant wheat Subtilisin/ Chymotrypsin inhibitor (WSCI) on survival of *D. coarctata* neonates. (N=10 per treatment). Survival data was analysed using Graphpad Prism software.

Chapter 4:

Insecticidal activity of the *Delia coarctata* cysteine proteinase (DcCathL):

Amino acid sequence comparison of DcCathL and ScathL showed about 85% sequence similarity and both showed similar biochemical properties in terms of activity. ScathL has shown insecticidal effects to lepidopteran insects when injected into hemolymph. Injecting the active form of DcCathL into *Lacanobia oleracea* and *Mamestra brassicae* larval hemolymph showed similar melanisation patterns and mortality as observed for ScathL (Phillip *et al.,* 2007). Larvae injected with higher dose of DcCathL showed excessive fluid loss and development of melanised patches spreading from the site of injection. Larvae injected with lower doses of DcCathL showed reduced feeding leading to growth inhibition. On the contrary buffer injected larvae (control) showed very small melanised patches at the site of injection. The melanisation in control larvae was localised to wound site and disappeared in 2-3 days after injecting, larvae recovered soon and fed on artificial diet normally. Systemic and extensive melanisation induced due to activity of DcCathL/ScathL like cysteine proteinases was worthy of note and efforts were undertaken to investigate the role of injected DcCathL in larval hemolymph that caused extensive and systemic melanisation. The hypothesis was that DcCathL might be targeting 'something' in the hemolymph that is related to immune response and melanisation cascade. Serine proteinase inhibitors (Serpins) in the hemolymph, the negative regulators of melanisation cascade, were considered as the possible targets of DcCathL. Due to interference in normal melanisation cascade, melanisation process loses its control in DcCathL injected larvae leading to extensive rather than localised melanisation.

4.1 Insecticidal activity of DcCathL:

Different concentrations of recombinant DcCathL (0.5 to 5μg) activated with (1mM) glutathione as reducing agent were used to inject fifth instar cabbage moth (*Mamestra brassicae* L.; Lepidoptera) larvae. Larvae showed excessive fluid loss and extensive systemic melanisation. Larvae also showed shrinkage and blackening spreading from

the site of injection leading to death. The effect of DcCathL was more lethal in the first two days after injection and was maximal in the initial 4 h after injection. The larvae surviving after this period showed poor wound healing at the site of injection and reduced feeding. The dose response curve for DcCathL injection indicated an LD_{50} of approx. 2 μg per larva, with no mortality at doses 0.5 μg, and maximal effect (> 90% mortality) at doses ≥ 5μg (figure 4.1). Surviving larvae injected with DcCathL showed a dose-dependent decrease in weight gain at doses above 1 μg, of up to approx. 50% compared to control larvae. This loss in larval weight was interpreted as due to loss of fluid to some extent (which was not measured quantitatively as larvae were very fragile and vulnerable) and reduced feeding on artificial diet (figure 4.2).

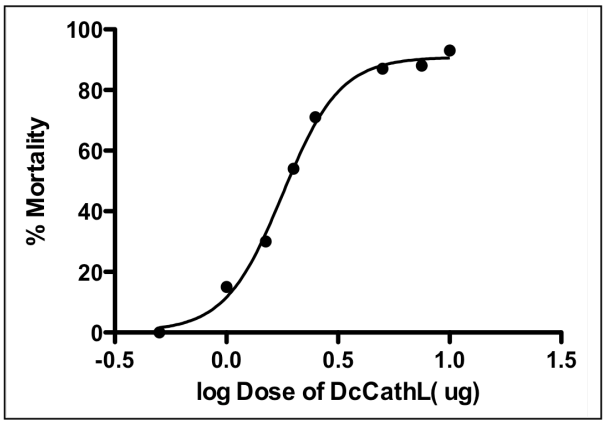


Figure 4.1: Insecticidal effect of DcCathL on *M. brassicae* larvae (Mortality). Dose–response curve for insecticidal activity of DcCathL when injected. 5th instar *M. brassicae* larvae were injected with varying amounts of DcCathL (log values of concentration). The dose response curve was fitted using non-linear regression with Prism software, and parameters were estimated from the fitted curve. Points indicate mean values ± SE (N=3).

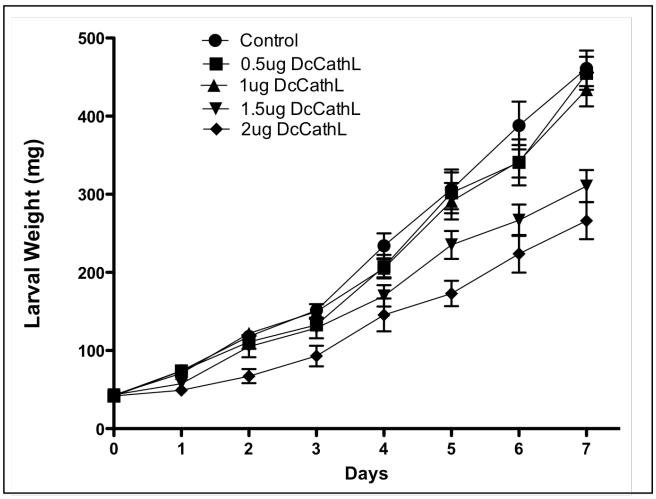


Figure 4.2: Effect of DcCathL on growth of surviving larvae of *M. brassicae*. 5th instar larvae injected with lower dose of DcCathL did not show significant mortality but their growth was reduced probably due to non-feeding. 1.5µg and 2µg DcCathL injected larvae show significant reduction in growth as compared to control and 0.5 & 1µg DcCathL. Points indicate mean values ± SE (N=20).

The physiological responses of *H. virescens* after injection of recombinant ScathL have been studied in detail (Li *et al.,* 2008). This insect was less sensitive to the cysteine proteinase than *L. oleracea*, with an estimated LD_{50} of 11 µg per larva. In *L. oleracea* and *M. brassicae*, mortality was associated with melanisation, which occurred in the hemolymph and internal organs (principally trachaea). Melanisation must result from the activity of phenoloxidase (PO), but no evidence for increased PO activity in the hemolymph of insects injected with ScathL, or infected with ScathL-expressing baculoviruses, was observed, and the prophenoloxidase pool decreased. ScathL was not able to activate pro-PO *in vitro*, and thus the increase in melanisation *in vivo* is unlikely to have resulted from direct action of ScathL on pro-PO (Li *et al.,* 2008).

4.2 Effect of DcCathL on M. brassicae hemolymph proteins:

To study the effect of DcCathL on *M. brassicae* hemolymph proteins, the cell-free haemolymph was extracted from 5th instar larvae and assayed for proteolysis *in vitro*. Analysis of a time course for incubation of the enzyme with hemolymph showed that

DcCathL was able to digest a range of hemolymph proteins over extended incubation periods when present at ratios ≥ 1:10 in relation to hemolymph proteins, although some major components were not degraded (results not shown). Previous results showed only minor effects of ScathL on hemolymph proteins in *Lacanobia oleracea* larvae *in vivo* after injection (Philip *et al.,* 2007). The extracted hemolymph had low levels of proteolytic activity, as measured by a general protease substrate; EnzChek, whereas DcCathL activity could be readily detected using this substrate. Approx. 80 μg of hemolymph protein had a similar activity to 1 μg of purified recombinant DcCathL (data not shown). DcCathL was still active in the presence of a large excess of hemolymph proteins (figure 4.3), although the total proteinase activity detected was less than the sum of activities of DcCathL plus the hemolymph, especially at ratios of DcCathL:hemolymph proteins ≥ 1:20. Inhibition of DcCathL by hemolymph components is implied by these results. On the other hand, in both the assays shown, and others carried out at varying ratios of DcCathL:hemolymph proteins, incubation of hemolymph proteins with DcCathL did not increase protease activity above the level caused by DcCathL itself, showing that no activation of hemolymph proteinases was detected.

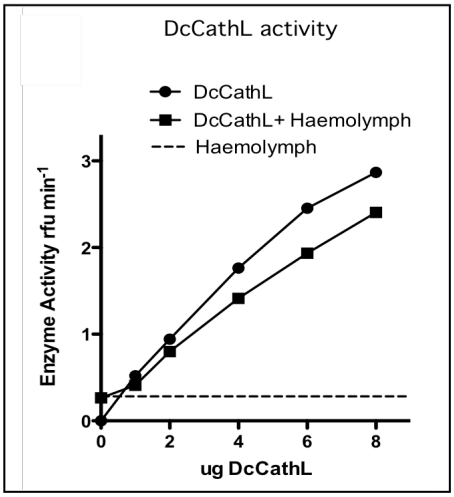


Figure 4.3: Proteolytic activity (protein substrate) of DcCathL in the presence or absence of haemolymph from *M. brassicae*. Activity of *M. brassicae* haemolymph control is represented by the horizontal dashed line. Enzyme activity is given as increase in fluorescence per minute (RFU/min). Points indicate mean values ± SE (N=3).

These results show that DcCathL and similar enzymes can be active in hemolymph if present at a sufficient concentration. The activity of these proteases in hemolymph is

seen even in presence of cysteine proteinase inhibitors as components of the hemolymph proteins. A cysteine protease inhibitor (MsCPI) has been purified from larval hemolymph of *Manduca sexta* which showed strong inhibition of the plant cysteine protease, papain. Recombinant MsCPI showed inhibition of cathepsins B, H and L (Miyaji *et al.,* 2007). The proteolytic activity of these cathepsin L-like enzymes towards a range of protein substrates is non-specific if substrates are unfolded, but specific cleavages are observed when the enzymes act on folded proteins (Philip *et al.,* 2007), implying that exposed and unfolded regions of proteins are cleaved selectively but non-specifically. Although it is possible that DcCathL and similar enzymes are able to activate proteinases involved in the pro-PO activation cascade, the failure to observe increased levels of PO activity following injection (Li *et al.,* 2008), suggests that a different mechanism is operating to cause melanisation after injection.

4.3 Effect of DcCathL on recombinant serpins of M. brassicae:

To study the effect of DcCathL on lepidopteran hemolymph serpins, we cloned two serpins (MbSpn1A and MbSpn1B/C) from *M. brassicae.* Both serpins were recombinantly expressed in *P. pastoris*, purified and characterised for their inhibitory properties towards target proteases.

4.3.1 Cloning and sequence analysis of M. brassicae serpins:

cDNA sequences are available for serpins derived from the *Mamestra configurata* (bertha armyworm) serpin1 gene by alternative splicing (Chamankhah *et al.,* 2003; Hegedus *et al.,* 2008). *M. configurata* serpin sequences were used to isolate homologues from *Mamestra brassicae* by RT-PCR, using primers based on the *M. configurata* sequences (mRNA accession nos. **AY148483** coding for protein **AAN71632**, **AY 148484** coding for protein **AAN71633** and **AY 148485** coding for protein **AAN71634**) and *M. brassicae* RNA as a template. *M. configurata* serpins were selected on the basis of differences in the Reactive Centre Loop (RCL) regions leading to differing predicted inhibition specificity; serpin-1A was predicted to inhibit chymotrypsin-like proteinases (predicted cleavage C-terminal to Tyrosine) and serpin-1B/C was predicted to inhibit trypsin-like proteinases (predicted cleavage C-terminal to Arginine).

The coding sequences of the *M. brassicae* serpins MbSpn1A and MbSpn1B/C were very similar to serpins from *M. configurata* and *Manduca sexta*. Both MbSpn1A and MbSpn1B/C showed presence of a common N-terminal region that is conserved throughout (residues 1–351). Both serpin amino acid sequences when analyzed by SignalP server showed presence of a signal sequence (residues 1–16; MKLFICLALAATAMA). There was a variable C-terminal region (residues 352–392) that comprises a hinge with conserved residues-EGAEAAAAN, Reactive Centre Loop (RCL) with variable reactive site residues and a serpin signature containing 8 amino acid residues. Deduced amino acid sequences of MbSpn1A and MbSpn1B/C showed presence of most conserved regions similar to other serpins. The conserved N-glycosylation sequence (NVTK) at amino acid positions 84 to 87 and a potential N-myristylation site (GAVLND) at amino acid residues 116 to 121 as reported by Chamankah *et al.,* (2003) were present (figure 4.4). Serpin signature sequences PFVFYLM (MbSpn1A) and PFFYALK (MBSpn1B/C), which are responsible for the specificity of the protease inhibitory activities were also present.

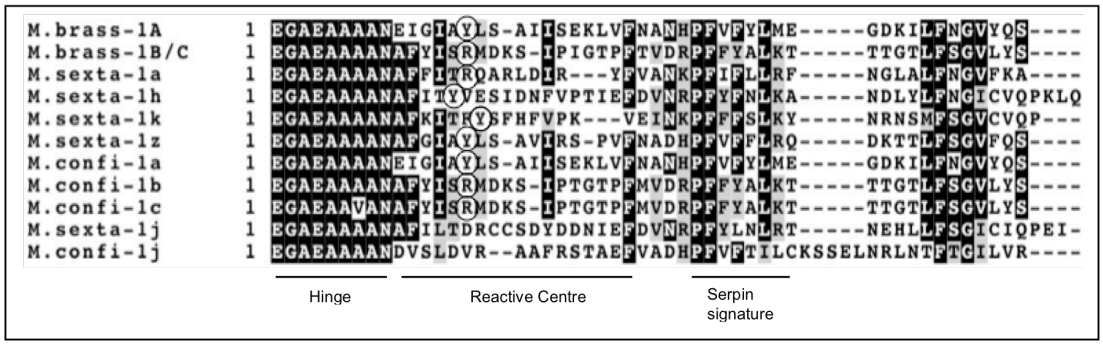
```
Mb-serpin-1A       MKLFICILALAATAMAEESNVDLLKSSNEVFTANMFQEVVKAKPGENVVLSAFSVLSPLA 60
Mb-serpin-1B/C     MKLFICILALAATAMAEESNVDLLKSSNEVFTANMFQEVVKAKPGENVVLSAFSVLSPLA 60
                   ************************************************************

Mb-serpin-1A       QLSLASVGESHDEILKAIGLPNDNVTKEVFTDVSKQLRSVKGVELRLANKVYVRQGAVLN 120
Mb-serpin-1B/C     QLSLASVGESHDEILKAIGLPNDNVTKEVFTDVSKQLRSVKGVELRLANKVYVRQGAVLN 120
                   ************************************************************

Mb-serpin-1A       DEFAAVSKDVFNSDVKNVDFTKNVEAAKEINEWVEENTNHKIKDLVSSESLDASTAAVLV 180
Mb-serpin-1B/C     DEFAAVSKDVFNSDVKNVDFTKNVEAAKEINEWVEENTNHKIKDLVSSESLDASTAAVLV 180
                   ************************************************************

Mb-serpin-1A       NAIYFKGKWKKPFDESATRDLDFFVTKDQPIKKPTMHKSGDFKYAESKELDAKLLELPYE 240
Mb-serpin-1B/C     NAIYFKGKWKKPFDESATRDLDFFVTKDQPIKKPTMHKSGDFKYAESKELDAKLLELPYE 240
                   ************************************************************

Mb-serpin-1A       GDQSSLLIVLPNEIDGIGSLVEKLKDPTALSKAVENMFYNEVNVDLPKFKIETTTDLKAV 300
Mb-serpin-1B/C     GDQSSLLIVLPNEIDGIGSLVEKLKDPTALSKAVENMFYNEVNVDLPKFKIETTTDLKAV 300
                   ************************************************************

Mb-serpin-1A       LKKMNIVKLFTGGEARLNNLIKGESDLFITDAIQKAFIDVNEEGAEAAAANEIGIAYLSA 360
Mb-serpin-1B/C     LKKMNIVKLFTGGEARLNNLIKGESDLFITDAIQKAFIDVNEEGAEAAAANAFYISRMDK 360
                   ************************************************** : *: :.

Mb-serpin-1A       IISEKLVFNANHPFVFYLMEGDKILFNGVYQS 392
Mb-serpin-1B/C     SIPIGTPFTVDRPFFYALKTTTGTLFSGVLYS 392
                    *.    *..::**.: *      **.**  *
```

Hinge, Reactive centre Loop (RCL) and serpin Signature

Figure 4.4: Amino acid sequence analysis of MbSpn-1A and MbSpn-1B/C. Amino acid sequence of *M. brassicae* MbSpn1A and 1B/C, from cDNA sequences. Predicted signal sequences (residues 1–16) are underlined, NVTK: N-glycosylation site (in bold), GAVLND: N-myristylation site (bold and underlined). Inhibitor region(residues 341–379) at carboxyl terminal: Hinge (residues 341–350), Reactive Centre Loop (residues 351–367) and serpin

signature (residues 373–379) are indicated.

The predicted inhibition specificity of the *M. brassicae* serpins, shown by a comparison of the RCLs of serpins from *M. sexta*, *M. configurata* and *M. brassicae* (figure 4.5), was similar to their *M. configurata* homologues. MbSpn1A showed presence of predicted reactive site inhibitory residue Tyrosine (Y^{357}) similar to *M. sexta* serpin 1h, 1K, 1Z and *M. configurata* serpin 1a while MbSpn1B/C contains a



predicted reactive site inhibitory residue Arginine (R^{357}) similar to *M. sexta* serpin 1a and *M. configurata* serpin 1b and 1c.

Figure 4.5: Alignment for carboxyl-terminal variable inhibitory regions of *M. brassicae* MbSpn1A and 1B/C with serpins from *M. sexta* and *M. configurata*. Predicted inhibitory reactive sites (P1) residues are encircled.

4.3.2 Recombinant expression and purification of MbSpn-1A and MbSpn-1B/C:

For recombinant expression of MbSpn1A and 1B/C in the yeast *P. pastoris,* an expression construct was assembled in pGAPZαB expression vector to encode a product containing the yeast α-mating factor prepro-sequence, an N-terminal $(His)_6$ tag for detection and purification of recombinant protein, and the mature serpin coding sequence. C-terminal tags from vector were excluded to avoid interfering with the C-terminal active site loop region of the serpins. The details of expression constructs are shown in figure 4.6 (MBSpn1-A) and figure 4.7 (MBSpn-1B/C). The recombinant proteins were produced in the culture supernatant of yeast transformants (at a yield of approx. 3 mg $litre^{-1}$ of shake flask culture) as polypeptides of approx. 43 kDa that showed immunoreactivity towards anti $(His)_6$ antibodies in western blot (Figure 4.8). The recombinant serpins were purified by metal affinity chromatography. Pemberton and Bird (2003) have suggested previously that serpins could be produced as recombinant proteins in yeast, although bacterial expression

systems have been more commonly used; serpin-1B from *M. sexta* has been characterised by recombinant expression in an *E. coli* expression system (Jiang *et al.,* 1995), as have multiple serpins from *Ctenocephalides felis* (Brandt *et al.,* 2004). Baculovirus-infected insect cells have also been used as an expression system (Brandt *et al.,* 2004). However we were successful in producing the active recombinant serpins using *P. pastoris* expression system.

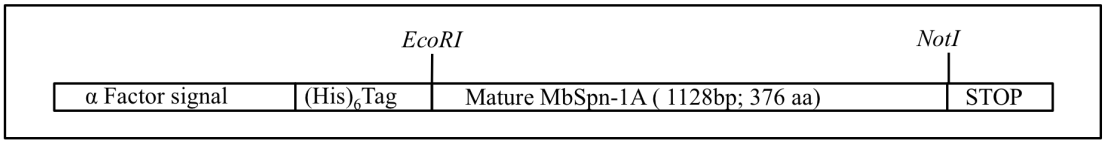


Figure 4.6 (A): Schematic representation of expression construct for MbSpn-1A in pGAPZαB.

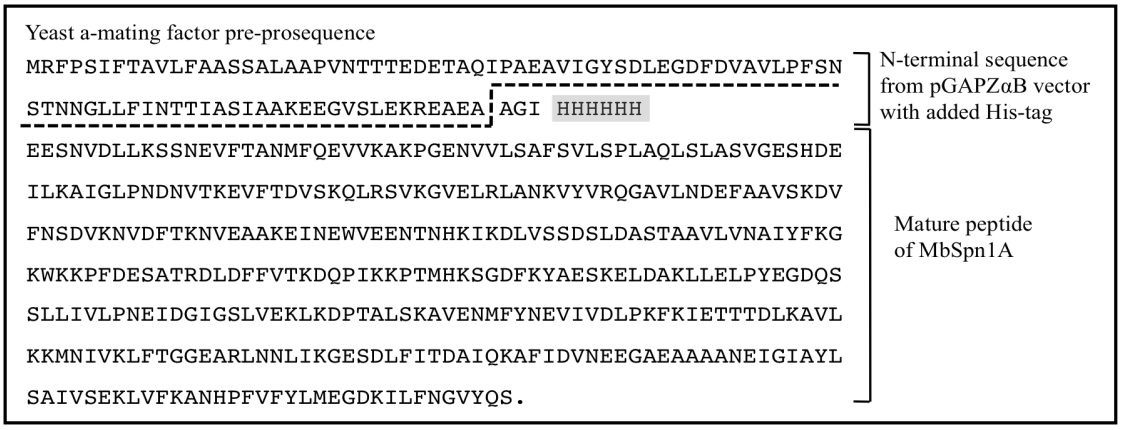


Figure 4.6 (B): Sequence details of expression construct for MbSpn-1A in pGAPZαB. A complete coding sequence for mature MbSpn-1A containing 376 amino acid residues is inserted into MCS of yeast expression vector pGAPZαB, between EcoRI and NotI restriction enzyme sites. α-factor signal sequence containing 89 amino acid residues, helps secretion of recombinant protein into the culture media. Standard C-terminal Myc epitope and (His)6 tags were excluded and a (His)6 tag was included at N-terminus of MbSpn-1A after α-factor signal sequence for immunodetection of recombinant protein and purification by metal affinity chromatography.

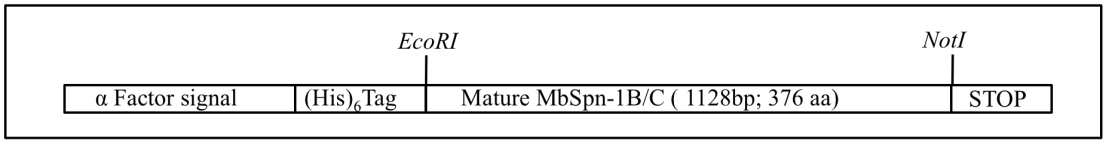


Figure 4.7 (A): Schematic representation of expression construct for MbSpn-1B/C in pGAPZαB.

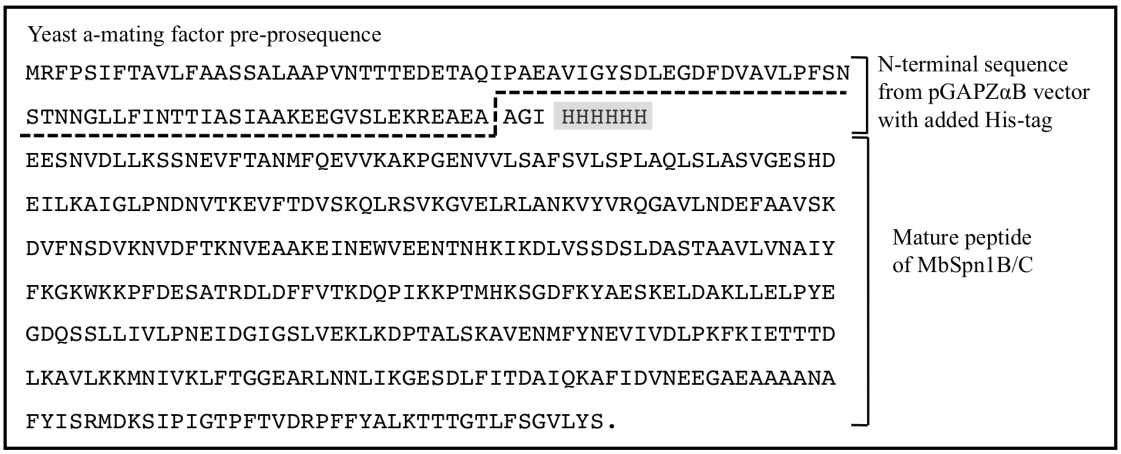


Figure 4.7 (B): Sequence details of expression construct for MbSpn-1B/C in pGAPZαB. A complete coding sequence for mature MbSpn-1B/C containing 376 amino acid residues is inserted into MCS of yeast expression vector pGAPZαB, between EcoRI and NotI restriction enzyme sites. α-factor signal sequence containing 89 amino acid residues, helps secretion of recombinant protein into the culture media. Standard C-terminal Myc epitope and (His)6 tags were excluded and a (His)6 tag was included at N-terminus of MbSpn-1A after α-factor signal sequence for immunodetection of recombinant protein and purification by metal affinity chromatography.

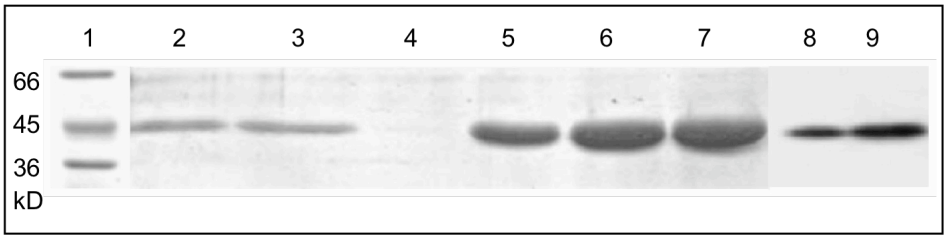


Figure 4.8: Purification of recombinant MbSpn1A (MbSpn1B/C not shown). Lane 1: Molecular weight marker, Lane2: Load (culture supernatant), Lane 3: Flow through fraction, Lane 4: Binding buffer wash fraction, Lane 5: 20 mM Imidazole wash fraction, Lanes 6 and 7: 500 mM Imidazole elution fractions. Lanes 1–7 show staining with Coomassie Blue. Lanes 8 and 9: western blot confirmation of purified MbSpn1A (MbSpn1B/C not shown) using anti-His antibodies.

4.3.3 Inhibitory properties of recombinant MbSpn-1A and MbSpn1B/C:

The purified recombinant serpins showed the expected inhibitory activities when tested against bovine enzymes. MbSpn1A showed inhibition of chymotrypsin but did not inhibit trypsin, while MbSpn1B/C showed a strong inhibition of trypsin and no inhibition of chymotrypsin (Figure 4.9). Saturating levels of inhibition under the assay conditions used were >90% for MbSpn1B/C and approx. 50% for MbSpn1A. Neither serpins inhibited recombinant DcCathL even when present in large excess. This is expected, as the reactive site loop on these inhibitors is specific for serine proteinases. Although there have been suggestions that some serpins are able to inhibit cysteine proteinases (Jiang and Kanost, 1997).

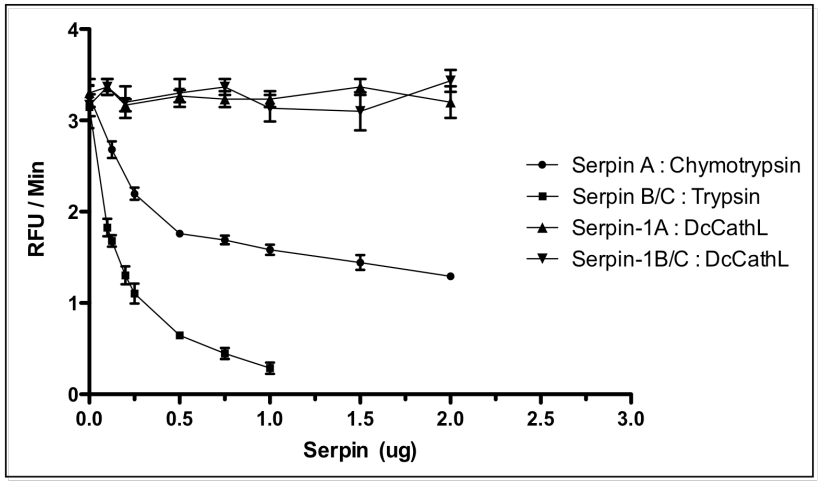


Figure 4.9: Inhibition of proteinases by recombinant MbSpn-1A and MbSpn-1B/C. Points show mean values ± SE (n =3) of residual activity of trypsin, chymotrypsin and DcCathL after incubation with *M. brassicae* MbSpn1A (serpin A) and MbSpn1B/C (serpin B/C). Enzyme activity is given as increase in fluorescence per minute (RFU/min). Points indicate mean values ± SE (N=3).

4.3.4 Degradation of recombinant serpins by DcCathL in vitro:

Since the *M. brassicae* serpins do not inhibit DcCathL, the possibility that the proteinase could rapidly degrade these negative regulators of PPO activation was explored by carrying out *in vitro* digestion assays. Degradation of *M. brassicae* serpins was compared to a standard protein, myoglobin (from horse skeletal muscle), in a time course experiment where both serpin and myoglobin were simultaneously present. In a typical experiment (enzyme:substrate ratio 1:10), analysis by SDS-PAGE

showed that more than 50% of intact serpin was digested within first 15 min of treatment with DcCathL, whereas digestion of myoglobin was much slower, with no significant decrease in the amount of intact protein after 4 h incubation with DcCathL (figure 4.10). The experiment shows that *M. brassicae* hemolymph serpins are susceptible to degradation by DcCathL *in vitro*.

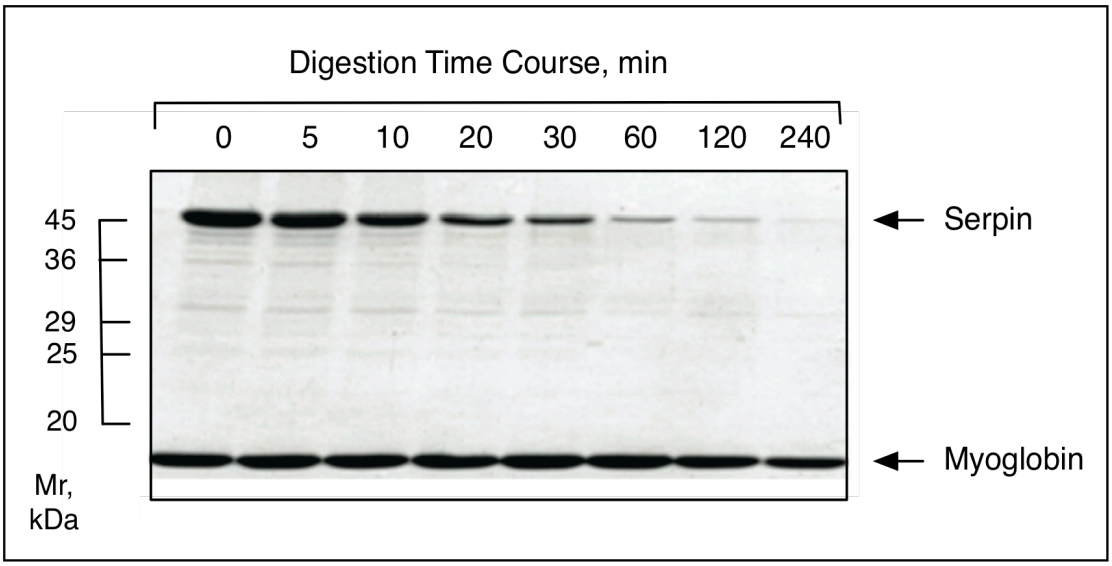


Figure 4.10: Serpins are susceptible to degradation by DcCathL. Comparison of degradation of recombinant serpin and myoglobin by DcCathL. 50 μg each of recombinant MbSpn1A and horse skeletal myoglobin were digested with 5 μg of recombinant DcCathL. Reaction aliquots at varying times of digestion (as shown) were analysed by SDS-PAGE followed by Coomassie Blue staining. Bands of recombinant serpin (43kDa) and myoglobin (17kDa) are indicated.

4.4 Degradation of recombinant serpins in hemolymph:

To confirm that DcCathL can selectively degrade serpins as components of insect hemolymph, recombinant serpins were added to cell-free hemolymph extracted from *M. brassicae* larvae. Time course analyses were carried out with and without DcCathL addition. Analysis by SDS-PAGE followed by staining showed that DcCathL had no major effect on hemolymph protein profiles under the conditions used (figure 4.11). When western blotting to detect recombinant serpin was carried out, the serpin was observed to be stable over a 26 h incubation in hemolymph in the absence of added DcCathL, but was rapidly degraded in hemolymph to which DcCathL had been added, with a half life of approx. 2 to 4 hours under the conditions

used (Figure 4.11). The stained gels show that hemolymph proteins are present in large excess over both the recombinant serpin and the recombinant proteinase, but a specific degradation of the serpin by DcCathL has still taken place. If a similar degradation of serpins by DcCathL occurs *in vivo* when the proteinase is injected into hemolymph, then this would prevent serpins acting as negative regulators of phenoloxidase activation.

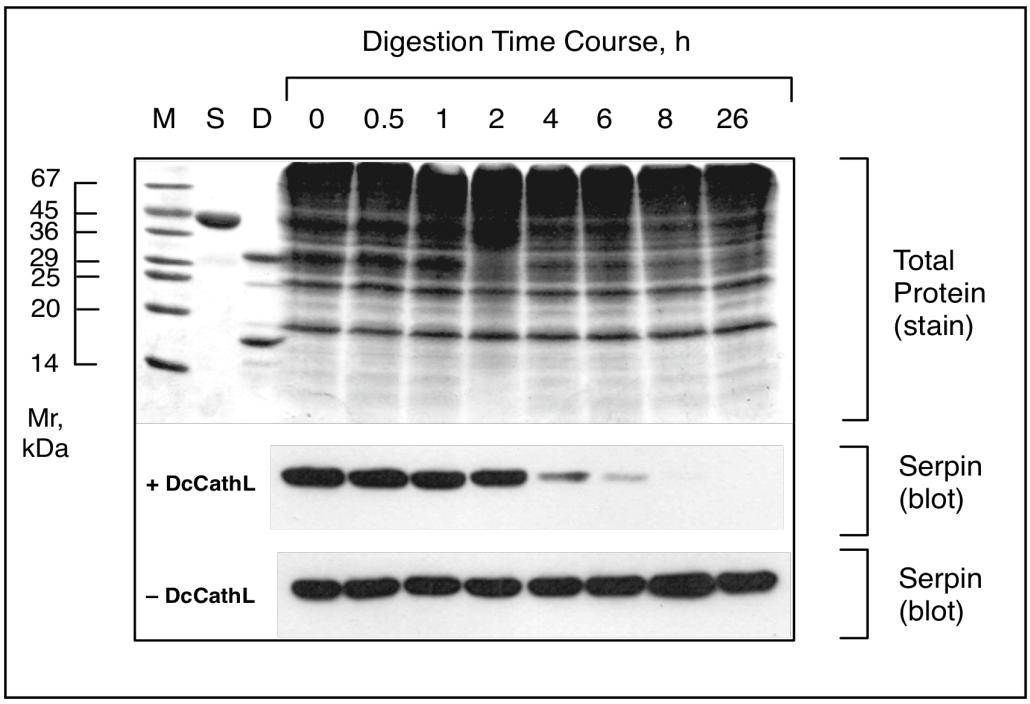


Figure 4.11: DcCathL selectively degrades Serpins in the presence of excess haemolymph proteins. Approx. 160 μg *M. brassicae* haemolymph proteins were incubated with 16 μg recombinant MbSpn1A in the presence or absence of 16 μg recombinant DcCathL. Reaction aliquots were sampled were collected after varying times of incubation (as shown) and analysed by SDS-PAGE. Gels were stained for total protein (Coomassie blue) or subjected to western blotting and probed using anti-(His)6 antibodies. Gel stained for total protein shows samples incubated with DcCathL; M: molecular weight markers, S: recombinant serpin standard (2 μg), D: recombinant DcCathL standard (2 μg).

4.5 M. brassicae serpins block insecticidal activity of DcCathL

It is assumed that injected DcCathL interferes with negative regulation of the phenoloxidase activation cascade in larval hemolymph by degrading endogenous serpins. However, maintaining the pool of serpins in the hemolymph exogenously should prevent this effect, even though the serpins have no direct effect on DcCathL activity. To confirm this hypothesis, *M. brassicae* larvae were co-injected with a

lethal dose of recombinant DcCathL and serpins (MbSpn1A and MbSpn1B/C). Buffer-injected control larvae showed a highly localised melanisation at the site of the injection wound, which healed with minimal fluid loss (Figure 4.12, panel A); they remained healthy and fed normally. The fluid loss was not measured quantitatively as the larvae were too fragile to handle and was kept to visual observation when the bleeding hemolymph was seen on the tissue paper the larvae were kept on. The larvae were of about the same size when they were injected, but different levels of fluid losses altered the size of larvae. Similar minimal fluid loss, healthy behaviour, and highly localised melanisation were observed for larvae injected with MbSpn1A or MbSpn1B/C (Figure 4.12, panel B), demonstrating that exogenous serpins do not prevent melanisation involved in wound healing. Larvae injected with DcCathL at the dose employed showed extensive fluid loss, cessation of movement and feeding, and systemic melanisation to a greater or lesser degree (Figure 4.12, panel E–G). Approx. 80% of the DcCathL-injected larvae died within 4 hours of injection, and all died within 24 hours. In contrast, larvae co-injected with either MbSpn1A or MbSpn1B/C and DcCathL did not show excessive fluid loss or cessation of feeding. Larvae showed more melanisation than observed in control larvae although this remained localised to the site of injection (Figure 4.12, panel C, D). After 24 hours larvae co-injected with serpins and DcCathL still showed a bigger melanisation spot than controls, but remained healthy, active and feeding normally. Scoring of mortality and melanisation showed a statistically significant difference between DcCathL-only injected insects and all other treatments (Table 4.1).

Colour Plate-IV

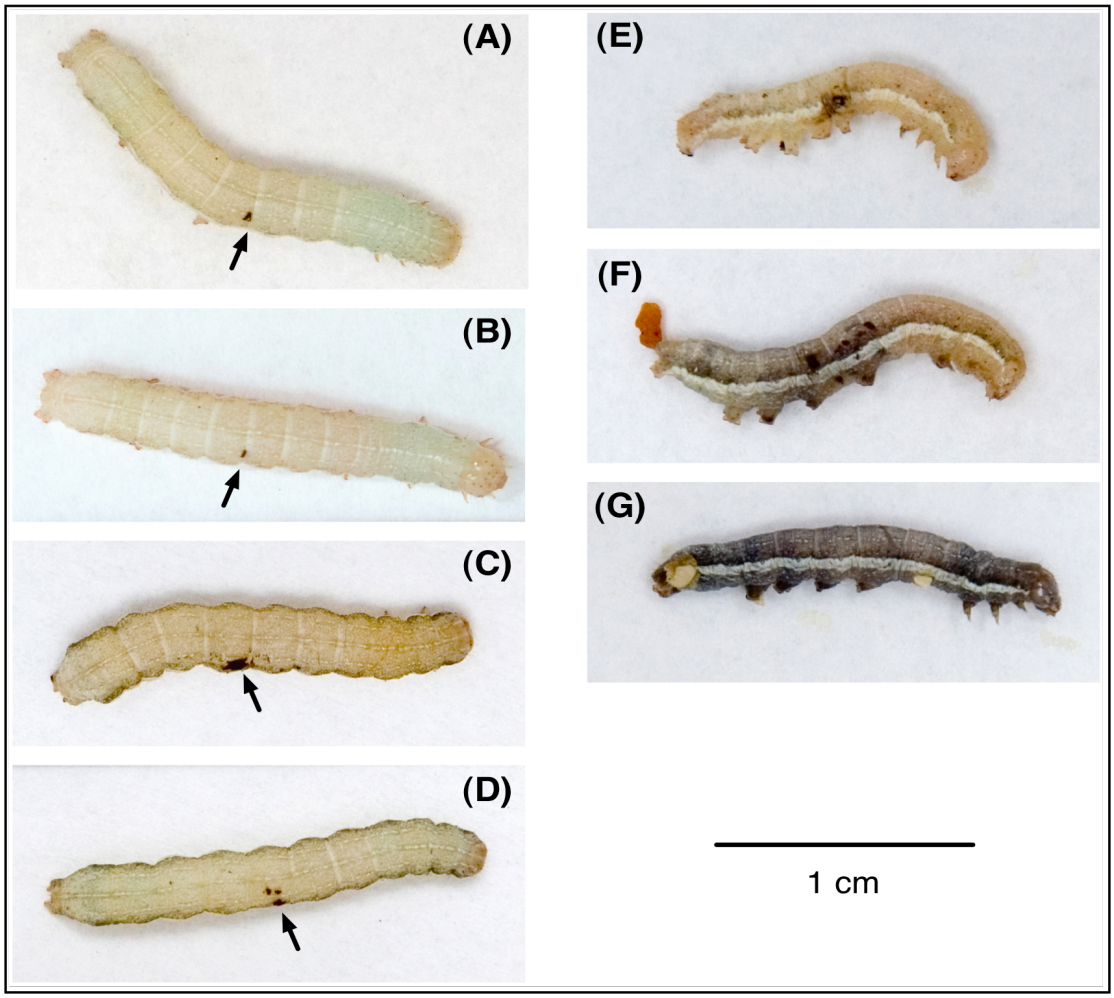


Figure 4.12: Co-injections (DcCathL+ rec. serpin); Serpins can block insecticidal effects of DcCathL. Photographs show representative effects of injecting recombinant DcCathL and *M. brassicae* serpins into 5th instar *M. brassicae* larvae. All larvae were injected with a total volume of 5µl of solution. (A) Control (buffer injected); (B) 5µg serpin (MbSpn1A and MbSpnB/C gave identical results); (C) 5µg + MbSpn1A + 5mg DcCathL; (D) 5µg MbSpnB/C + 5µg DcCathL. Larvae (A)–(D) all survived >24 h after injection; arrows denote site of injection. Larvae (E)–(G) were injected with 5mg DcCathL, and show different extents of systemic melanisation. (E) Spreading from injection site; (F) partially melanised; (G) fully melanised. Larvae (E)–(G) all died <4 h after injection.

Treatment	Mortality (%, $n = 12$)		Melanisation (mean score)	
	4 h PI	24 h PI	4 h PI	24 h PI
Control	0	0	1.0	1.0
MbSpn1A	0	0	1.1	1.0
MbSpn1B/C	0	0	1.2	1.0
DcCathL ($n = 16$)	75	100	3.2	3.4
DcCathL + MbSpn1A	0	0	1.0	1.3
DcCathL + MbSpn1B/C	8	8	1.2	1.4

Table 4.1: Mortality and melanisation of 5th instar M. brassicae larvae injected with DcCathL (5 μg) and/or serpins MbSpn1A and MbSpn1B/C (5 μg each). (PI) post injection. Mortality was analysed by survival analysis; DcCathL treatment was different to all other treatments ($p < 0.01$). Melanisation score was analysed by 1-way ANOVA, using a non-parametric Kruskal–Wallis comparison and Dunn's multiple comparison test; DcCathL treatment was different to all other treatments ($p < 0.05$).

These observations suggest that MbSpn1A and MbSpn1B/C can act as negative regulators of the phenoloxidase activation cascade in *M. brassicae*. Although they do not inhibit DcCathL, they were able to prevent the systemic melanisation caused by this enzyme when injected into the insect. Although MbSpn1A and MbSpn1B/C would be subject to degradation by DcCathL in the hemolymph, sufficient exogenously supplied inhibitors were present to inhibit prophenoloxidase activation, and thus systemic melanisation. This would depend on the rate of serpin degradation by DcCathL, the stability of DcCathL in the hemolymph, and the levels of endogenous serpins. However, the results do not establish that MbSpn1A and MbSpn1B/C are directly involved in regulating conversion of prophenoloxidase to phenoloxidase by prophenoloxidase activating enzyme (PPAE). The differing specificities of inhibition of the two serpins make it highly unlikely that both could inhibit PPAE, which cleaves prophenoloxidase with a trypsin-like specificity in *M. sexta.* Serpin inhibition of pro-PO activation can occur at any of the several proteolytic steps involved in the activation cascade in *M. sexta* (Tong *et al.,* 2005), and the observation that both MbSpn1A and MbSpn1B/C are effective as inhibitors of systemic melanisation suggests that different proteinases in the cascade are being targeted.

Although the inhibitory action of serpins is thought to restrict melanisation in insects

to the site of injury or microbial infection, the mechanism of this negative regulation is not fully understood, particularly with respect to how phenoloxidase activity is restricted to the site(s) where it is required. In particular, it is not clear why insects injected with the cysteine proteinase suffer massive fluid loss, unless serpin regulation of phenoloxidase activity is necessary to allow the ‘‘plugging’’ of the injection hole to occur properly.

4.6 Potential of DcCathL as insecticidal protein:

Proteinases like DcCathL could be potential candidates for crop protection. An extensive study by Li *et al.,* (2008) showed that baculovirus expressed ScathL was insecticidal to two insect pests *H. armigera* and *A. pisum,* causing fragmented fat body, ruptured gut and malpighian tubules and melanised tracheae. The role of cathepsins in insect tissue remodelling and potential as insecticidal proteins was discussed. The basement membrane (BM) is the ultimate target for the cysteine proteases like DcCathL and ScathL and has been considered as a potential target for insect pest management (Harrison and Bonning, 2001; Liu *et al.,* 2006) as the degradation of BM may lead to tissue degradation and cause death. BMs are extracellular protein sheets that surround all tissues. BMs play and an important role in cell adhesion, cell signaling, and maintenance of tissue structure (Yurchenco and O’rear, 1993). Basement membranes in insects need to be remodelled during embryonic development, tissue and cell differentiation and metamorphosis (Page-McCaw *et al.,* 2003) which involve many enzymes and mainly the basement degrading enzymes; cathepsins (Homma and Natori, 1996; Homma *et al.,* 1994). It is necessary to control tightly the expression and activation of these BM-degrading enzymes as it may lead to uncontrolled and potential damage to other tissues (Llano *et al.,* 2002). This could ultimately lead to interference with insect physiological processes resulting in insect death (Liu *et al.,* 2006; Tang *et al.,* 2007). The specificity and intra hemocoelic toxic effect makes cathepsins the ideal candidates for crop protection strategy.

However there is a major drawback that these enzymes must be introduced into the hemolymph to be effective as both the effects on basement membrane and on the melanisation cascade depend on degradation of proteins only accessible in hemolymph. ScathL a cysteine proteinase similar to DcCathL was fed to lepidopteran

larvae by incorporating into artificial diet but did not show insecticidal activity. This was considered to be due to degradation or inactivation in the insect gut (Li *et al.,* 2008). Some proteins like plant lectins (GNA; a lectin-agglutinin from Snowdrop, *Galanthus nivalis*) are able to cross insect gut and enter hemolymph when fed in artificial diet. GNA is a mannose specific lectin and when fed to insects, it binds to gut epithelium and passes into the hemolymph. (Fiches *et al.,* 2004b). GNA when fed to lepidopteran larvae shows limited insecticidal activity by reducing larval weight gain and to slow the developmental rate of first stadium *Lacanobia oleracea* larvae (Fitches *et al.,* 1997; Gatehouse *et al.,* 1997). Though the mechanism of antifeedant behaviour of GNA is not clear it might bind to glycoproteins of the gut epithelium (Fitches *et al.,* 1999). GNA is a hardy protein that can withstand highly proteolytic conditions of lepidopteran guts and resistant to gut proteolysis (Fitches *et al.,* 2001). There is a possibility of formation of 'leaky' junctions at the binding sites of GNA to gut epithelium that might be responsible for passive transport of insecticidal toxins across the gut to hemolymph (Fitches *et al.,* 2004b).

Insecticidal proteins could be recombinantly co-expressed with such proteins acting as 'carrier' proteins and can potentially carry the insecticidal protein to the insect hemolymph. GNA was found to deliver an insect neuropeptide (*Manduca sexta* allatostatin; Manse-AS) to the hemolymph of lepidopteran larvae when fed in artificial diet and showed significant reduction in growth of fed larvae (Fitches *et al.,* 2002). An insecticidal protein, chitinase was recombinantly produced as a fusion protein with GNA, however its effectiveness was not enhanced due to carrier protein (Fitches *et al.,* 2004a). In another attempt, insecticidal spider venom neurotoxin (*Segestria florentina* toxin SFI1) fusion with GNA was highly insecticidal to tomato moth, *L. oleracea.* Delivery of intact SFI1/GNA fusion protein to the haemolymph in these insects was confirmed, where haemolymph samples from fusion-fed larvae contained a GNA-immunoreactive protein of the same molecular weight as the SFI1/GNA fusion (Fitches *et al.,* 2004b). DcCathL fusion with GNA like carrier proteins could be a potential candidate as an insecticidal protein against lepidopteran insect pests. Crop plants expressing such fusion proteins can deliver resistance to plant against wide range of insect pests. However, considerable technical problems would need to be overcome to produce functionally active DcCathL / GNA fusion proteins either in recombinant proteins expression systems, or *in planta.*

Chapter 5

Cereal aphid (*Sitobion avenae*) gut proteolysis and antimetabolic effects of recombinant wheat PIs on survival and growth of *S. avenae*

5.1 Proteolytic activity and types of proteases in S. avenae guts

5.1.1 Proteolytic activity in S. avenae gut:

Extracts of total soluble proteins from whole cereal aphids and dissected guts showed similar levels of proteolytic activity towards a protein substrate (labelled casein), when compared on a per insect basis, suggesting that guts contained the majority of the proteolytic activity in the insect (result not shown). In agreement with this conclusion, no proteolytic activity above background could be detected in aphid haemolymph. All subsequent results were obtained with protein extracts prepared from dissected guts, and include enzyme activities present in gut tissues, and gut contents.

The proteolytic activity towards a protein substrate in aphid guts was maximal at pH 7.0, with negligible activity at pH $\leq$ 4.5 or $\geq$ 9.0 (figure 5.1). Assays were carried out in an isoionic buffer system but similar results were obtained with other buffers (not shown). Addition of a reducing agent, glutathione, stimulated the proteolytic activity by approx. 10% at the optimum pH. Proteolytic activity was decreased by chemical inhibitors with specificity for both cysteine and serine proteinases (figure 5.2). Both chymostatin (specific for chymotrypsin-like enzymes) and E-64 (specific for cathepsin-like cysteine proteinases) decreased activity by approx. 75% when used at concentrations of 10μM and 1μM respectively. The specific serine protease inhibitor AEBSF (100μM) decreased activity by approx. 70%, whereas chloroketone inhibitors specific for trypsin and chymotrypsin (TLCK and TPCK; 10μM each) caused approx. 30% decrease in activity. Specific synthetic substrates were used to further characterise the types of proteolytic activities present in aphid guts (figure 5.3). Similar levels of hydrolytic activity towards Z-Arg-Arg-AMC (substrate for cathepsin B) and Suc-Ala-Ala-Pro-Phe-AMC (substrate for chymotrypsin) were observed at

substrate concentrations giving high activity, with approx. 60% activity shown towards Z-Phe-Arg-AMC (substrate for cathepsin L).

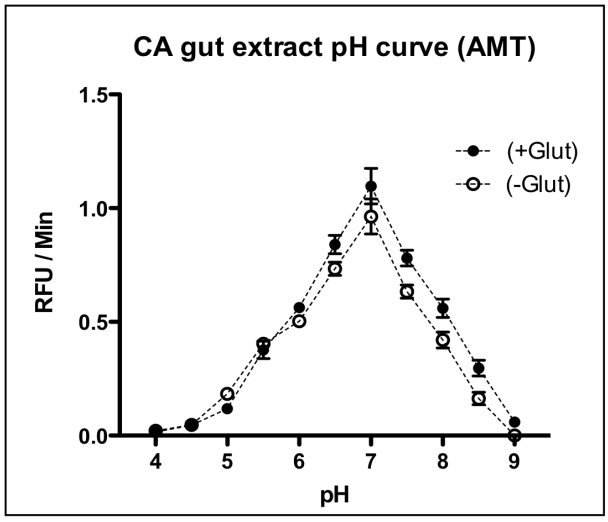


Figure 5.1 pH dependent activity of Cereal aphid (CA); *S. avenae* gut proteases: Gut total protein extract activity curve at pH 4 to 9 using EnzChek substrate and AMT buffer system. Points show mean values ± SE (n=3). Gut protease activity is shown as relative fluorescence unit per min (RFU/Min). Activity of gut proteases in absence and presence of 1mM Glutathione as reducing agent is shown.

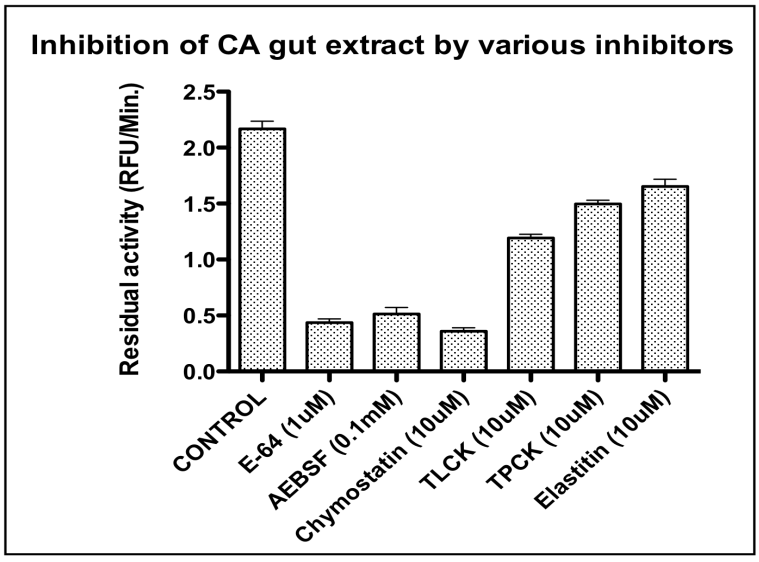


Figure 5.2 Inhibition of Cereal aphid (CA); *S. avenae* gut extract by different protease inhibitors: The following synthetic protease inhibitors were used for inhibition of CA gut protease activity to determine type of protease activity using EnzChek as assay substrate. E-64 (Papain type cysteine proteases), AEBSF (Serine proteases), Chymostatin (chymotrypsin and cysteine proteases), TLCK and TPCK (both trypsin protease inhibitors), all inhibitors

showed at least 10-15% inhibition of protease activity. Chymostatin and E-64 showed highest inhibition of protease activity followed by AEBSF, TLCK and TPCK. Error bars indicate mean values ± SE (n=3).

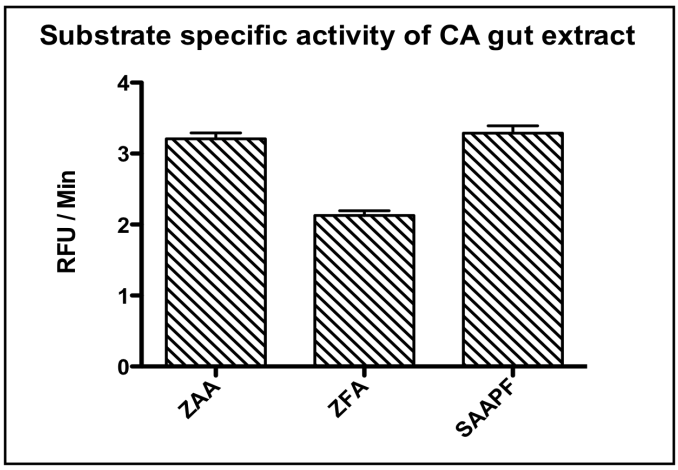


Figure 5.3 Activity of *S. avenae* gut extract with protease specific substrates: Protease specific synthetic flourimetric substrates were used to detect type of protease activity in the gut extracts. Highest relative activity was seen with Z-Arg-Arg-AMC (ZAA; cathepsin like cysteine proteases) and Suc-Ala-Ala-Pro-Phe-AMC (SAAPF; chymotrypsin like serine protease activity. Gut extract also showed activity with Z-Phe-Arg-AMC (ZFA), which is a common substrate for trypsin-like serine protease and cathepsin-like cysteine proteases. Error bars indicate mean values ± SE (n=3).

The biochemical characterisation of proteinase activity in the aphid gut reported here is consistent in part with identification of genes likely to encode gut proteinases. The bioinformatic analysis identifies cysteine proteinases with sequence similarity to cathepsin B as the most likely enzymes responsible for the observed proteolytic activity (Rispe *et al.,* 2007). The neutral pH optimum for proteolysis and inhibition by E-64 are consistent with this prediction, as is the hydrolysis of a typical cathepsin B substrate, Z-Arg-Arg-AMC. Cathepsin B-like enzymes will also hydrolyse the cathepsin L substrate, Z-Phe-Arg-AMC (Bown *et al.,* 2004). Cathepsin L-like enzymes may also be present; a gut proteinase cloned as cDNA from cotton aphid, *Aphis gossypii,* was shown to be cathepsin L-like on the basis of sequence similarity (Deraison *et al.,* 2004), although localisation of this enzyme suggested it was intracellular rather than secreted. The bioinformatic analysis did not support the presence of digestive serine proteinases in aphids, although the biochemical data did

suggest some evidence for serine proteinase activity (Deraison *et al.,* 2004), shown by partial inhibition of proteolysis by high levels of AEBSF and effective inhibition of proteolysis by chymostatin, albeit at a higher concentration (10μM) than optimal for specific inhibition of chymotrypsin (1μM). Chymostatin can inhibit cysteine proteinases (Kirschke, 1998), so the inhibition observed for cereal aphid gut extracts, as well as those from *A. pisum* and *A. gossypii* (Christofoletti *et al.,* 2004; Deraison *et al.,* 2004) is not proof of the presence of a chymotrypsin-like serine proteinase. Further evidence for a chymotrypsin-like serine proteinase was shown by hydrolysis of a typical chymotrypsin substrate, Suc-Ala-Ala-Pro-Phe-AMC, at a rate comparable to the cathepsin B substrate. In comparison, a detailed biochemical analysis of proteinase activity in the gut of pea aphid, *A. pisum,* concluded that the major proteinase activity in the gut was due to cathepsin cysteine proteinase(s) (Cristofoletti *et al.*, 2004), with no hydrolysis of a synthetic chymotrypsin substrate (Suc-Ala-Ala-Phe-AMC) observed. The presence of a chymotrypsin-like serine proteinase in the cereal aphid gut cannot be ruled out, and further characterisation of individual enzymes will be necessary to determine whether the chymotrypsin-like activity observed can be accounted for by cysteine proteinases.

5.1.2 Cathepsin B like cysteine proteases from S. avenae gut

Gut extract assays of *S. avenae* showed presence of cysteine type protease activity in the gut and it is believed that Hemipteran insects rely mainly on cysteine type proteases for their digestion (Terra, 1990). Cysteine proteinase genes of the cathepsin B family were found to be present in multiple copies, with 28 cathepsin B-like genes identified in the *A. pisum,* 5 of which were preferentially expressed in gut (Rispe *et al.,* 2007). Homologues of *A. pisum* cathepsin B-like cysteine proteases CathB-16A, CathB-2744 and CathB-84 were reported to be present in *S. avenae* (Rispe *et al.,* 2007). Complete coding sequences for the *S. avenae* genes CathB-16D, CathB-2744 and CathB-84 were isolated by RT-PCR from RNA purified from dissected cereal aphid guts, using gene-specific primers derived from *A. pisum* sequences. The predicted protein sequences were very similar (>95% identity) to the *A. pisum* orthologues (Figure 5.4 A, B and C). These result show that these genes are expressed in cereal aphid gut, and thus could be potential digestive enzymes.

Figure 5.4 (A) ClustalW-Boxshade alignment of *S. avenae* SitCathB-16A with *A.pisum* CathB-16A (Accession No. **ABV 03124**). Residues with black background are highly conserved and residues with light grey background are partially conserved.

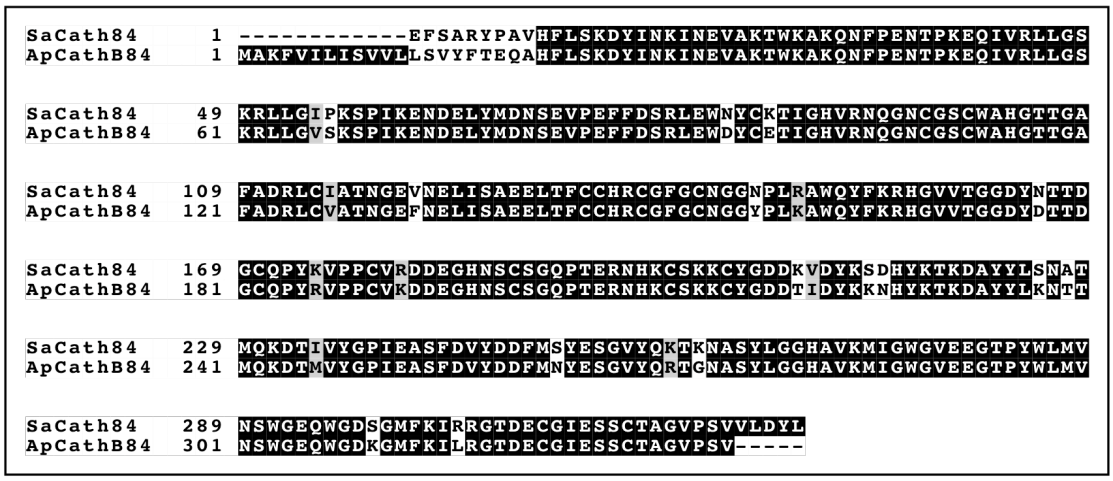


Figure 5.4 (B) ClustalW-Boxshade Alignment of *S. avenae* SitCathB-84 with *A.pisum* CathB-84 (Accession No. **ABV 03134**). Residues with black background are highly conserved and residues with light grey background are partially conserved.

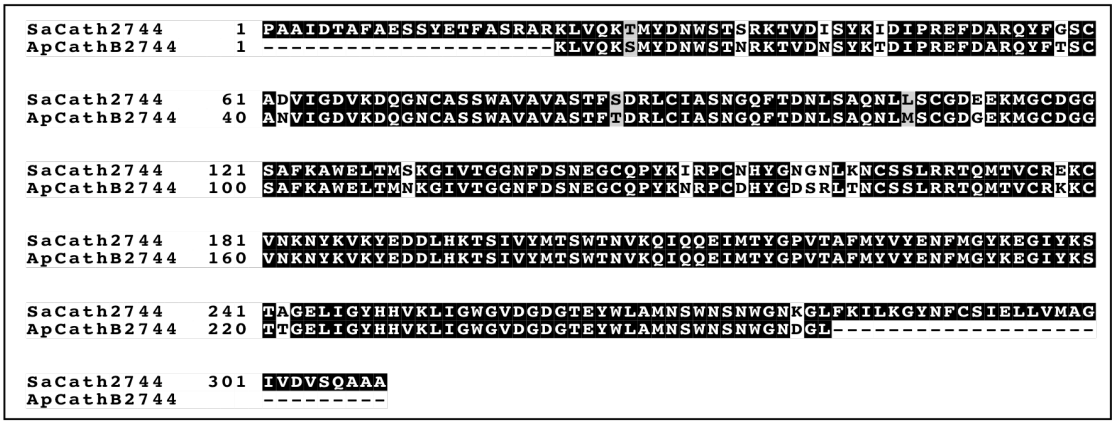


Figure 5.4 (C) ClustalW-Boxshade Alignment of *S. avenae* SitCathB-2744 with *A.pisum* CathB-2744 (Accession No.**ABV 03127)**. Residues with black background are highly conserved and residues with light grey background are partially conserved.

While cysteine proteinases are clearly identified as potential digestive enzymes in aphids on the basis of genomic analysis, similar evidence for serine proteinases is lacking. The sequence of a putative digestive serine proteinase isolated from gut tissue of the hemipteran sap-sucking herbivore *N. lugens* (rice brown planthopper; accession no. **CAC87119;** Foissac *et al*., 2002) was used to search for similar predicted proteins in *A. pisum*, but identified proteins with only limited similarity. The most similar predicted protein (XP_001952684) had 38% identity, 54% similarity (Blast score 169; BlastP software) and was similar in length to the *N. lugens* serine proteinase. However, most predicted serine proteinases in *A. pisum* with less similarity to **CAC87119** have additional domains diagnostic of non-digestive roles, or are similar to proteins with known non-digestive functions. In comparison, the *A. pisum* genome contains many predicted proteins with similarity to a *N. lugens* cathepsin B-like proteinase (**CAC87118**) isolated from gut tissue (28 predicted proteins have Blast scores >200; most similar *A. pisum* protein, NP_001119608, Blast score 303), indicating that similar digestive cathepsin B-like cysteine proteinases are widely distributed in hemipteran insects. Recombinant expression and characterisation of these proteases and inhibition with cysteine proteinase inhibitors could throw more light on functions of these proteases. I tried to express CathB-16D recombinantly in *P. pastoris* by following the same methods as DcCathL however

was not successful, therefore entire gut extracts were used as representative of gut proteolytic activity for *in vitro* assays.

5.2 Utilisation of ingested protein by cereal aphids

As we found both serine and cysteine type protease activity in the *S. avenae* gut, it was interesting to find out whether these gut proteases can digest an ingested digestible protein. Degradation of the ingested digestible protein such as ovalbumin could be easily detected by analysing the honeydew, which could prove the proteolysis in aphid gut. Aphids were fed on liquid diets with normal levels of free amino acids, and from which free amino acids were omitted ("control" and "-AA" diets) and a diet lacking free amino acids but supplemented with ovalbumin at 1mg/ml as a protein source ("-AA + ovalbumin" diet)**.** As expected, aphid survival on the diet lacking free amino acids was poor compared to control diet, with all aphids dead by day 21 of the assay compared to 90% survival for controls (Figure 5.5 A). Growth on "-AA" diet was strongly retarded, with aphids failing to reach maturity; at day 15 length and width were both decreased by 55-60% compared to controls (Figure 5.5 B). Supplementation of the "-AA" diet with ovalbumin improved survival, and increased growth. Although all aphids on "-AA + ovalbumin" diet died by day 21 of the assay, survival curves for treatments with and without ovalbumin were significantly different ($p < 0.001$, log-rank test), with no mortality until day 13, whereas mortality for aphids fed "-AA" diet commenced at day 5. At day 15, surviving aphids fed "-AA + ovalbumin" diet were approx. 25% longer and 20% wider than aphids fed "-AA" diet ($p < 0.001$ for length, 2-way ANOVA). Similar results were obtained when ovalbumin was added to the diet at 5mg/ml and 10mg/ml, with the 10mg/ml diet giving better growth than 1 mg/ml, although still reduced by approx. 35% (length and width) when compared to control "+AA" diet (Figure 5.6). Protein supplementation of the "-AA" aphid diet with ovalbumin, at any level tested, was not sufficient to allow normal development, and aphids failed to reach maturity.

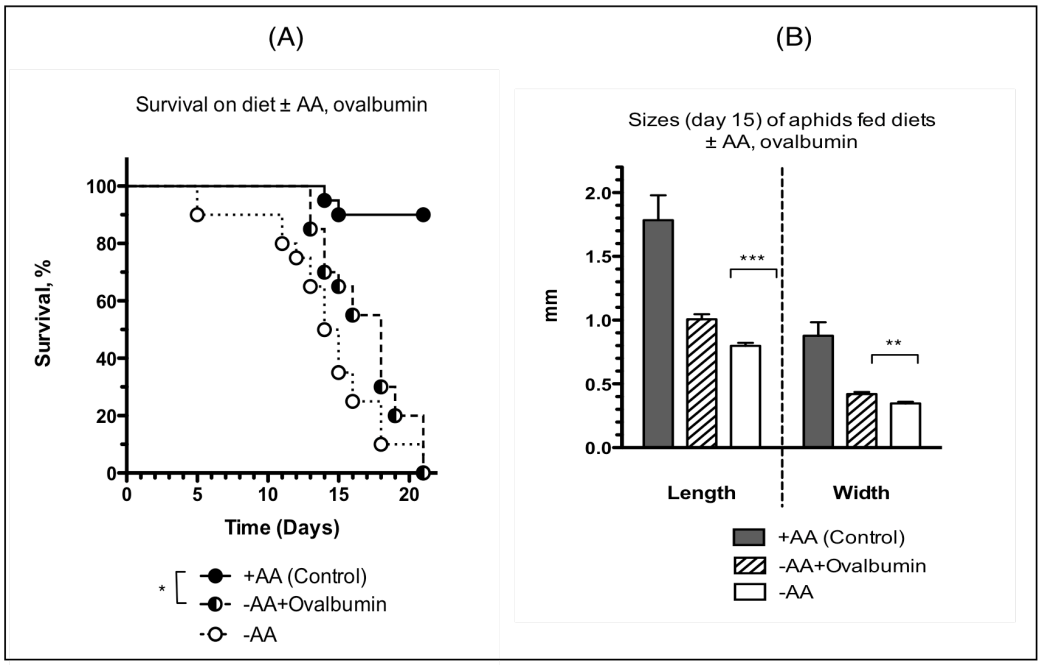


Figure 5.5 Putative digestion of ovalbumin in *S. avenae* gut; Effect of supply of free amino acids (+AA), no free amino acids (-AA) and ovalbumin as a digestible protein on (A) survival and (B) growth of *S. avenae*: Control diet (+AA) with a full supplement of free amino acids shows lower mortality and better growth, whereas aphids fed on no free amino acids diet (-AA) show less survival than control and the growth is highly affected. Ovalbumin, a digestible protein when supplied (1mg/ml) with no amino acid diet help aphids to survive better and growth is better as compared to (-AA) diet. Points show mean values ± SE (n=20).

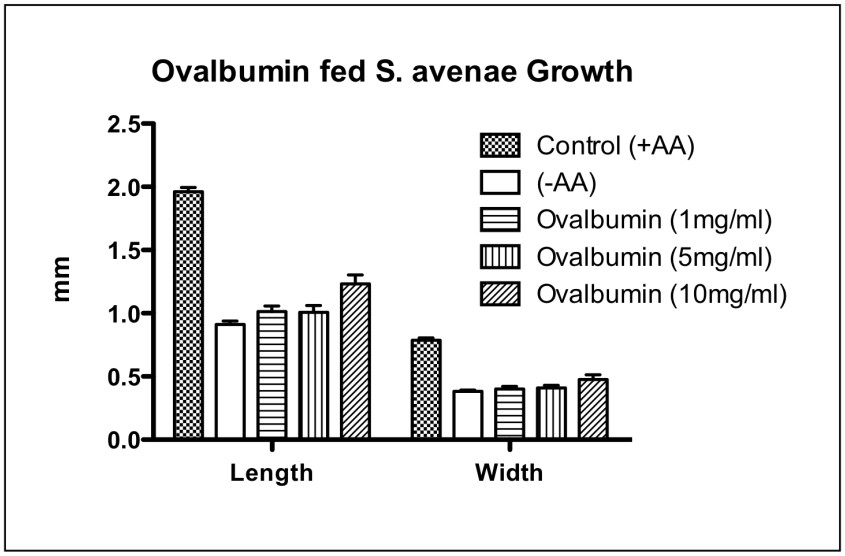


Figure 5.6 Effect of ovalbumin concentration on growth of *S. avenae*. Ovalbumin was added to diet containing no free amino acids (-AA) at 1, 5, 10 mg/ml concentration and

growth of aphids was measured after 6 days of feeding. Control aphids were fed with full supply of free amino acids (+AA). Error bars indicate mean values ± SE (n=20)

Analysis of honeydew collected over 72h from aphids fed "-AA + ovalbumin" diet by SDS-PAGE, followed by silver staining (figure 5.7), showed only small amounts of intact ovalbumin, estimated as less than 1% of the protein present in the ingested diet. The honeydew did not contain any prominent protein bands at all, suggesting that ovalbumin fragments resulting from partial digestion were not being excreted in detectable amounts, and that ovalbumin was being digested to peptides too small to be detected on SDS-PAGE, or to free amino acids. In contrast, ovalbumin was stable in diet over the period of the assay. Similar results were seen in Silverleaf whitefly where proteins were not only ingested but also digested using gut proteases producing free amino acids that were either excreted via honeydew or were used for *de novo* protein synthesis (Salvucci *et al.,* 1998).

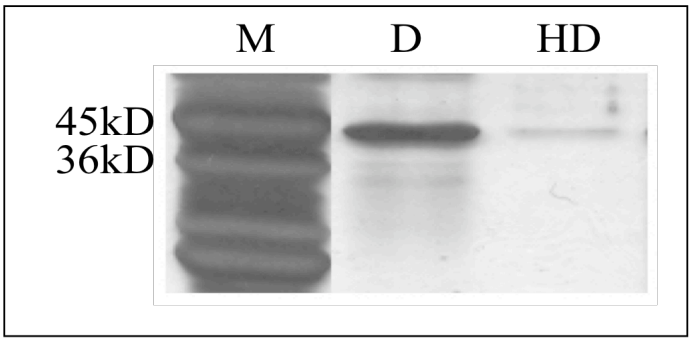


Figure 5.7 Putative digestion of ovalbumin in *S. avenae* gut. Honeydew analysed by SDS-PAGE and silver staining. Lane M: Molecular Weight marker, Lane D: Diet Control and Lane HD: Honeydew, showing ovalbumin band at ~45kD

Comparatively better growth in aphids fed on ovalbumin containing diet could be due to utilisation of ovalbumin as protein source to compensate the amino acid starvation. This finding is in agreement with findings of Foissac *et al.,* (2002), that digestive proteolysis is common in aphids and can make a significant contribution to nutrition. However, in *A. pisum* no apparent modification or degradation was observed in the ingested lectins in the honeydew samples collected (Rahbe *et al.,* 1995) where absence of lectins in honeydew are related to peculiar interaction with the aphid digestive tract. In another experiment by Habibi *et al.,* (2002) in *Lygus hesperus*, labelled casein was found to be extensively degraded in the gut to lower molecular

weight proteins in initial 2 hours and amount of lower molecular weight protein increased with increase in retention time in gut. After 12 hours no degraded products were detected suggesting all the ingested casein was digested. On the contrary, no detectable degradation was observed for the another protein fed, GFP, which is a non-digestible protein, it was transferred to hemolymph as a holoprotein.

5.3 Plant proteinase inhibitors against aphids:

There are many reports of recombinant expression of plant proteinase inhibitors using *E.coli* or *P. pastoris* expression system and these purified recombinant PIs have shown inhibition of insect gut protease activity *in vitro* and *in vivo* in feeding assays. Wheat Subtilisin/chymotrypsin inhibitor (WSCI) belongs to potato inhibitor I family and shares 87% sequence identity to barley chymotrypsin inhibitor-2A (CI-2A) (Facchiano *et al.,* 2006). Facchiano *et al.,* (2006) is the first report of wheat PI that is active against animal chymotrypsins and bacterial subtilisins. WSCI has been cloned and has been recombinantly expressed as a fusion protein with GST using *E.coli* (Di Gennaro *et al.,* 2005). 3D structure of WSCI revealing it reactive sites using molecular dynamics simulations has been reported earlier (Facchiano *et al.,* 2006). Four cDNAs encoding wheat Cystatins (WC1, WC2, WC3 and WC4) have also been cloned, characterized and recombinantly expressed as GST fusion using pGEX3 expression system (Kuroda *et al.,* 2001). A cDNA coding for phytocystatin (WC5) was isolated and its expression during caryopsis development has been characterised and a recombinant protein was produced in *E.coli* using pQEX-30 expression system (Corre-Menguy *et al.,* 2002).

The crop protection strategies developed against insect pests so far have been limited mostly to herbivorous pests belonging especially to Lepidoptera and Coleoptera. Due to assumption that aphids feed on plant phloem and need not rely on gut protein digestion, no special approach related to aphid digestion has been used. However with the establishment of presence of digestive proteases in aphid guts and putative digestion of ingested proteins, several proteinase inhibitors have been tested against aphid gut proteases. A cysteine proteinase inhibitor from rice (Oryzacystatin-OC-I) has been shown to induce moderate but significant growth inhibition on pea aphid (*A. pisum*), cotton aphid (*Aphis gossipii*) and peach potato aphid (*Myzus persicae*). OC-I was recombinantly expressed in leaves and phloem sap of transgenic oil seed rape

showed growth retardation, reduced fecundity and aphid biomass in *M. persicae* and the effect was correlated to inhibition of a major cathepsin L/H-type cysteine protease activity which was detected in whole insect extracts. The results suggested that OC-I affects *M. persicae* through digestive tract targets as well as by reaching hemolymph, and inhibiting extra-digestive proteolytic activities (Rahbe *et al.,* 2003a).

A Bowman-Birk type chymotrypsin inhibitor from pea seeds showed antimetabolic effect to pea aphid *A. pisum*. However, chymotrypsin like activity was not detected in the guts of *A. pisum* by using chromogenic substrates. This suggests that though chymotrypsin inhibitor shows antimetabolic effects on aphid, its physiological target is not present in the gut (Rahbe *et al.,* 2003b). PIs like these could also be involved in the inhibition of extra-digestive proteases that are involved in other metabolic activities. Five recombinant PIs from potato were fed to and tested for their ability to control 3 species of cereal aphids *Diuraphis noxia, Schizaphis graminum* and *Rhoalosiphum padi.* Of the five, two PIs from potato, I and II showed great antimetabolic effect on aphids by increasing mortality and reducing production of nymphs. This work suggests that PIs from potato are potential candidates for the control of cereal aphid species (Tran *et al.,* 1997). WSCI a similar protein to potato inhibitors could show similar effects against aphids. Genetically modified eggplant expressing oryzacystatin showed reduced net reproductive rate, the instantaneous rate of population increase and the finite rate of population increase of two aphids *M. persicae* and *Macrosiphum euphorbiae* suggesting that oryzacystatin has a negative impact on aphid population growth and mortality rates of tested aphids (Ribeiro *et al.,* 2006).

5.4 Wheat subtilisin/chymotrypsin inhibitor (WSCI) and cysteine proteinase inhibitor (WCPI):

5.4.1 Cloning of WSCI and WCPI

Two wheat PIs, cystatin (WCPI) an inhibitor of cysteine proteinases and WSCI, an inhibitor of chymotrypsin and subtilisin serine proteinases, were selected for assay of antimetabolic effects on *S. avenae* and effects on aphid gut proteolytic activity. cDNAs encoding both inhibitors were isolated from total RNA purified from wheat leaves by RT-PCR. The coding sequences of tha isolated cDNAs were identical to

sequences in the global database (WSCI: accession no. **AAY45744** and WCPI: **BAB18766**).

5.4.2 Sequence analysis of WSCI

Complete coding sequence of WSCI cDNA consists of 252 nucleotides coding for 84 amino acid residues. The complete amino acid sequence characterisation of WSCI has been reported earlier by Poerio *et al.,* (2003). Deduced amino acid sequence of WSCI was analysed for signal peptide (http://www.cbs.dtu.dk/services/SignalP/) showed presence of signal peptide with 12 amino acid residues MSSVVKKPEGGN. The mature peptide of WSCI contains 72 amino acid residues (TDT.....RVG). WSCI is a similar protein like potato inhibitor type-I (Poerio *et al.,* 2003) and all conserved amino acid residues similar to Potato Inhibitor I family members, K^{22}, T^{23}, $W^{25,}$ P^{26}, E^{27}, G^{30}, A^{36}, I^{40}, V^{57}, R^{66} and V^{67} (WSCI numbering) were present in WSCI. Absence of cysteinyl residues and presence of only one methioninyl and tryptophanyl residue along with presence of high content of essential amino acids (41 residues out of 71) are the peculiar characteristic of this inhibitor (Poerio *et al.,* 2003). The reactive site identified in PIN-I family member was Met^{60}-Glu^{61}. All the amino acid sequence features of WSCI are shown in figure 5.8.

```
1    ATGAGTTCTGTGGTGAAGAAGCCGGAGGGAGGGAACACCGATACTGGTGACCATCACAACCAGAAGACG
1     M  S  S  V  V  K  K  P  E  G  G  N  T  D  T  G  D  H  H  N  Q [K  T]
70   GAGTGGCCAGAGTTGGTGGGGAAGTCGGTGGAGGAGGCCAAGAAGGTGATTATGCAGGACAAGTCAGAG
24    E [W  P  E] L  V [G] K  S  V  E  E [A] K  K  V [I] M  Q  D  K  S  E
139  GCACAGATCGTAGTTCTACCGGTGGGGACAATTGTGACCATGGAATATCGAATCGACCGTGTCCGCCTC
47    A  Q  I  V  V  L  P  V  G  T  I [V] T  M ●E  Y  R  I  D [R  V] R  L
208  TTTGTTGACAGTCTCGACAAAATTGCcCAGGTCCCCAGGGTCGGC
70    F  V  D  S  L  D  K  I  A  Q  V  P  R  V  G
```

Figure 5.8 Nucleotide and deduced amino acid sequence of WSCI. Nucleotide bases and amino acid residues are numbered on left. Nucleotide sequences shown in italics were primers used for cloning WSCI.Putative signal peptide is shadowed. Highly conserved residues for serine proteases similar to Potato Inhibitor I family are boxed. (●) represents a reactive site either directly or indirectly identified (Poerio *et al,* 2003)

5.4.3 Sequence analysis of WCPI:

Complete coding sequence of WCPI cDNA consists of 426 nucleotides coding for 142 amino acid residues. The complete sequence characterisation has been reported earlier by Corre-Menguy *et al.,* (2002) and Kuroda *et al.,* (2001). Deduced amino acid sequence of WCPI was analysed for signal peptide (http://www.cbs.dtu.dk/services/SignalP/) showed presence of signal peptide with 28 amino acid residues MEMWKYRVVGSVAALLLLLAIVVPFTQT. The mature peptide of WCPI contains 114 amino acids (QTQ.....PSA). Amino acid sequence showed all the conserved motifs and residues typical to plant cystatins. Two highly conserved cystatin signature sequences were presence of QTVAG (residues 93 to 97; WCPI numbering) and W^{123} (Corre-Menguy *et al.,* 2002) and partially conserved residues are T^{94} and A^{96}. Residue Ala^{40} shows partial conservation, as it is replaced by Glycine in other plant cystatins. All the amino acid sequence features of WCPI are shown in figure 5.9.

```
1    ATGGAGATGTGGAAATATCGGGTCGTGGGATCGGTTGCTGCCCTCCTCTTGCTACTCGCCATCGTCGTG
1     M  E  M  W  K  Y  R  V  V  G  S  V  A  A  L  L  L  L  L  A  I  V  V
70   CCGTTTACTCAGACCCAGACGCAGAGCGCACGGGACAAAGCTGCCATGGCGGAAGACGCGGGGCCGCTG
24    P  F  T  Q  T  Q  T  Q  S  A  R  D  K  A  A  M [A] E  D  A  G  P  L
139  GTGGGAGGCATCAGTGACTCGCCGATGGGGCAAGAAAACGACCTCGACGTCATCGCGCTCGCCCGCTTC
47    V  G  G  I  S  D  S  P  M  G  Q  E  N  D  L  D  V  I  A  L  A  R  F
208  GCCGTCTCCGAGCACAACAACAAGGCCAATGCCCTGCTGGAGTTCGAGAATGTGGTGAAGGTGAAGAAG
70    A  V  S  E  H  N  N  K  A  N  A  L  L  E  F  E  N  V  V  K  V  K  K
277  CAAACTGTTGCTGGCACGATGCACTACATTACAATCCGGGTCACTGAAGGTGGGGCCAAGAAGCTCTAT
93   [Q  T  V  A  G] T  M  H  Y  I  T  I  R  V  T  E  G  G  A  K  K  L  Y
346  GAAGCTAAGGTGTGGGAGAAACCATGGGAGAACTTTAAGAAGCTCGAGGAGTTCAAGCTGGTGGAGGAC
116   E  A  K  V  W  E  K  P [W] E  N  F  K  K  L  E  E  F  K  L  V  E  D
415  GTTCCAAGCGCA
139   V  P  S  A
```

Figure 5.9 Nucleotide and deduced amino acid sequence of WCPI. Nucleotide sequences shown in bold are the primer sequences for cloning WCPI. Predicted signal peptide is shadowed.Boxed residues depict two highly conserved cystatin signatures. Of conserved residues QTVAG, Q93, V95 and G97 are highly conserved and T94 and A96 are partially conserved in other plant cystatins. Boxed and shadowed residue A40 shows partial conservation, as it is G (Glycine) in other plant cystatins (Corre-Menguy *et al*, 2002)

5.5 Recombinant expression and purification of WSCI and WCPI

The inhibitors were produced as recombinant proteins by expression in the yeast *Pichia pastoris*. Only mature peptides of WSCI and WCPI excluding the signal peptide [WSCI (TDT.....RVG.) and WCPI (QTQ.....PSA.)] were used for the preparation of expression constructs using pGAPZαB yeast expression vector. Coding

sequences for mature peptide were inserted in frame with the yeast α-mating factor N-terminal secretary signal. The expression vector pGAPZαB contained a C-terminal extension of 27 residues containing a Myc antigenic determinant and a $(His)_6$ C-terminal tag for detection and purification of recombinant protein. The details of expression constructs for WSCI and WCPI are shown in figure 5.10 and figure 5.11 respectively. *P. pastoris* clones were screened for recombinant protein expression by growing small YPG cultures for 3 days and the supernatants were subjected to SDS-PAGE (17.5%) followed by western blot. The clones expressing recombinant inhibitors showed immunoreactive bands with both anti $(His)_6$ and anti-Myc antibodies. The band for WSCI was at approx. ~15 kD and WCPI at approx. ~18 kD. The predicted molecular weights of mature WSCI and WCPI are 8.14 kD and 12.6 kD respectively (Figure 5.12).

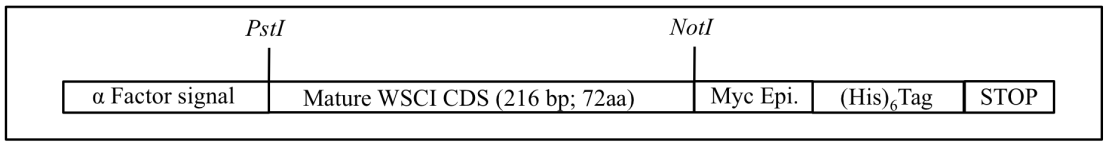


Figure 5.10 (A): Schematic representation of expression construct for WSCI in pGAPZαB.

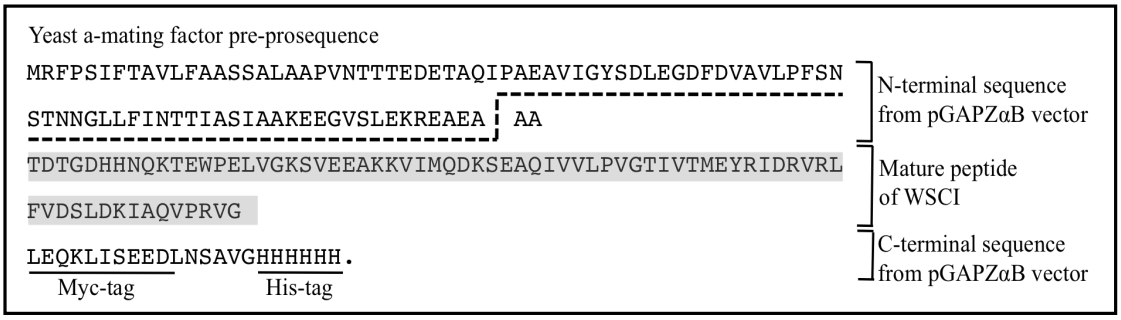


Figure 5.10 (B): Sequence details of Expression construct for WSCI in pGAPZαB. A complete coding sequence for mature WSCI containing 72 amino acid residues is inserted into MCS of yeast expression vector pGAPZαB, between PstI and NotI restriction enzyme sites. α-factor signal sequence containing 89 amino acid residues, helps secretion of recombinant protein into the culture media. C-terminus contains a Myc epitope (for immunodetection of recombinant protein) and His tag with six residues (for immunodetection of recombinant protein and purification by metal affinity chromatography)

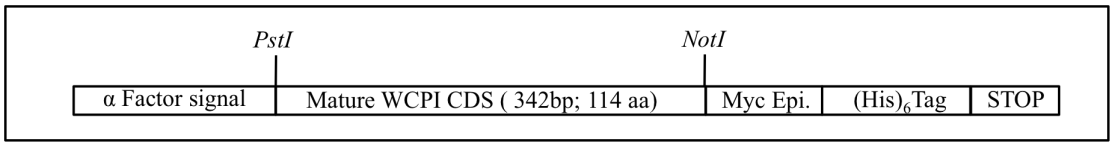


Figure 5.11 (A): Schematic representation of expression construct for WCPI in pGAPZαB.

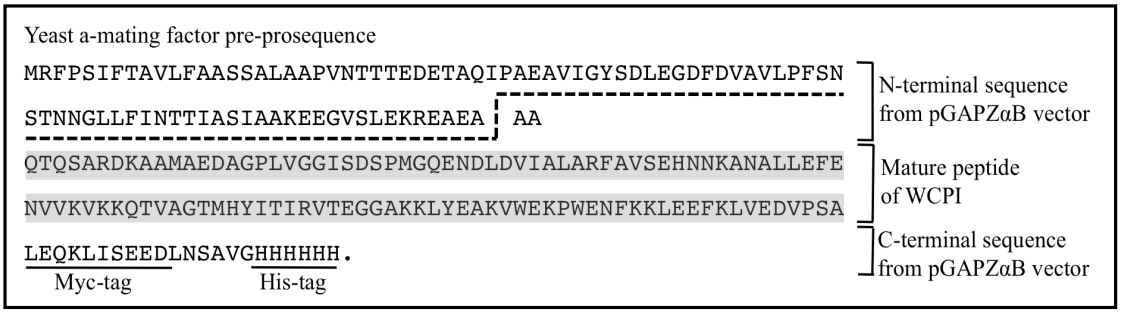


Figure 5.11 (B): Sequence details of expression construct for WCPI in pGAPZαB. A complete coding sequence for mature WCPI containing 114 amino acid residues is inserted into MCS of yeast expression vector pGAPZαB, between PstI and NotI restriction enzyme sites. α-factor signal sequence containing 89 amino acid residues, helps secretion of recombinant protein into the culture media. C-terminus contains a Myc epitope (for immunodetection of recombinant protein) and His tag with six residues (for immunodetection of recombinant protein and purification by metal affinity chromatography)

Selected positive expressing clones were used for production of inhibitor on large scale by bench-top fermentation. Recombinant WSCI and WCPI were purified from culture supernatant by ion-exchange chromatography at pH 4.0 on S-Sepharose, with elution by salt gradient. Both recombinant proteins were eluted at 0.65M NaCl concentrations in elution buffer. Purified protein fractions were analysed and confirmed by SDS-PAGE followed by CBB staining. Both inhibitors showed purified bands of respective sizes in stained gel, which also were immunoreactive when analysed by western blot using anti Myc and anti His antibodies (figure 5.12). Yields of both inhibitors were approx. 30 mg inhibitor / litre fermented media.

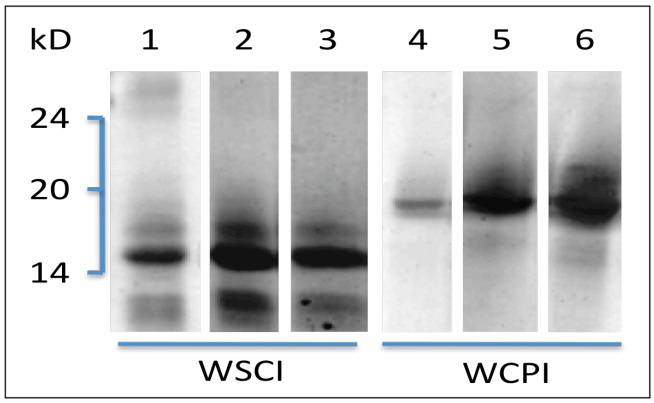


Figure 5.12 SDS-PAGE and western blot analysis of recombinant WSCI and WCPI expressed in *P. pastoris*. Lane 1: Fermentor supernatant purified by S-sepharose column chromatography, separated on 17.5% acrylamide-SDS-PAGE followed by CBB staining shows a 15kD WSCI band. Lane 2: Purified WSCI confirmation with Anti-Myc and Lane 3: with Anti-His6 antibodies by SDS-PAGE followed by western blot and detected using chemiluminescent detection system. Lane 4: Fermentor supernatant purified by S-sepharose column chromatography, separated on 17.5% acrylamide-SDS-PAGE followed by CBB staining shows a 18kD WCPI band. Lane 5: Purified WCPI confirmation with Anti-Myc and Lane 6: with Anti-(His) 6 antibodies by SDS-PAGE followed by western blot and detected using chemiluminescent detection system.

Previously repeated recombinant expression of WSCI has utilised bacterial expression systems like *E. coli.* A cDNA coding WSCI was cloned and recombinantly expressed using pGEX-2T as GST expression vector in *E.coli*. Recombinant WSCI when released from GST migrated at about 8.27 kD as per predicted molecular weight of the protein on 16.5 % acrylamide gel (Di Gennaro *et al.,* 2005). There are several reports of recombinant expression and purification of phytocystatins. Three cystatins from rice Oryzacystatin-I, II and III were expressed as GST fusion proteins (Ohtsubo *et al.,* 2005). Two cystatins cysM and cysP from *Coix lacryma-jobi* were expressed recombinantly in *E.coli* using pQE30 expression vector, cysM was obtained as soluble protein while cysP was in the insoluble fraction. Purified inhibitors migrated at 14 kD (cysM) and 15 kD(cysP) on 15% acrylamide gel (Yoza *et al.,* 2002). Corn cystatin I was expressed as mature protein in *E. coli* and showed a molecular mass of approx. 13kD as detected by western blot. It showed strong inhibitory activities against papain and cathepsin H but showed less inhibition of Cathepsin B (Abe *et al.,*

1992). Wheat cysteine proteinase inhibitors are also expressed recombinantly using *E. coli* expression systems (Kuroda *et al.,* 2001; Corre-Menguy *et al.,* 2002). Our results show the first report expression of wheat cystatin using the yeast (*P. pastoris*) expression system.

5.6 Inhibitory properties of WSCI and WCPI

Recombinant WSCI and WCPI were confirmed to be fully functional by assays against standard proteases. Using a protein substrate, WCPI inhibited papain at approx. 1:1 molar stoichiometry (figure 5.13) to >80% inhibition, with an extrapolated equivalence of 0.3μg WCPI to 1.0μg papain. WSCI strongly inhibited both subtilisin and chymotrypsin digestion of a protein substrate, although in neither case was inhibition stoichiometric. Approx. 0.7μg and 0.8μg of WSCI was required to inhibit the activity 1μg of subtilisin and chymotrypsin, respectively, by 50%, and complete inhibition was not observed even at 10-fold molar excess of inhibitor (Figure 5.14 and 5.15). WCPI was not active against chymotrypsin and subtilisin, as expected (data not shown), but, surprisingly, WSCI was an effective inhibitor of protein digestion by papain (Figure 5.13). Amounts of WSCI required to give 40-90% inhibition of papain activity were comparable to WCPI, although WSCI was less effective both at low levels of inhibitor, and when inhibitor was present in excess. WSCI activity against cysteine proteinases has not been observed previously.

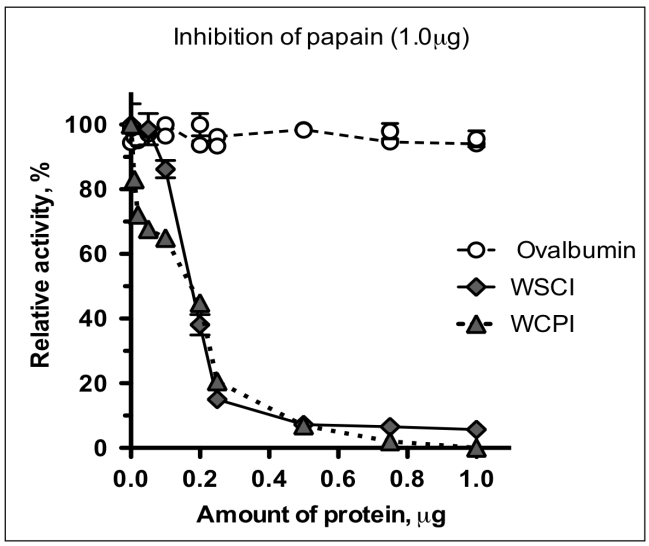


Figure 5.13 Inhibition of cysteine protease (Papain) by recombinant WSCI and WCPI. Points show mean values ± SE (n=3)

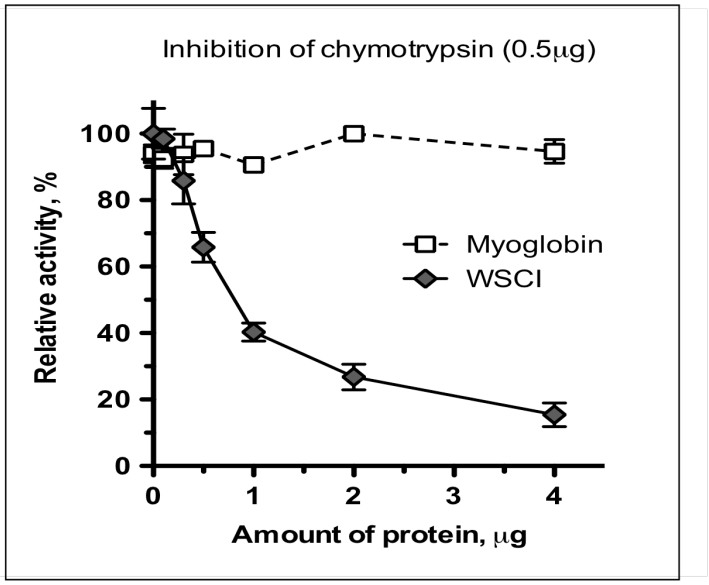


Figure 5.14 Inhibition of Bovine chymotrypsin activity by recombinant WSCI

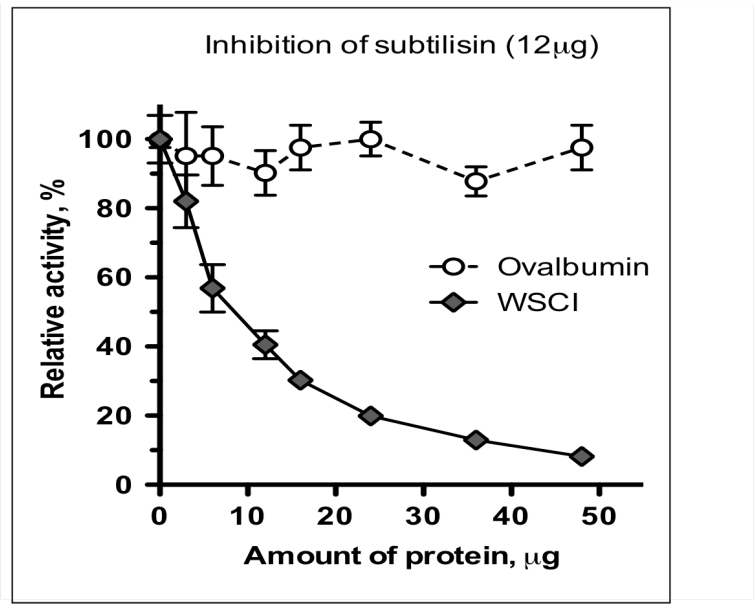


Figure 5.15 Inhibition of bacterial subtilisin activity by recombinant WSCI. Points show mean values ± SE (n=3)

WSCI also showed inhibition of recombinant cathepsin L like cysteine proteinase (DcCathL) from *D. coarctata* (as discussed earlier in **chapter 3**). WCPI being a specific cysteine proteinase inhibitor showed weaker inhibition of DcCathL than WSCI. Recombinant WSCI produced in *E.coli* showed inhibition of bacterial subtilisin and pancreatic chymotrypsin (Di Gennaro *et al.,* 2005). 1μg recombinant WCPI inhibited cathepsin like activity of 1μg Papain by almost 100%. Recombinant

WCs (WC- 1 to 4) produced in *E.coli* showed inhibition of rat liver Cathepsins and purified wheat cysteine proteinase (WCP-3) (Kuroda *et al.,* 2001).

5.7 Effect of WSCI and WCPI on S. avenae proteases in vitro

The proteolytic activity of cereal aphid gut extract against a labelled protein substrate *in vitro* was strongly inhibited by both WCPI and WSCI, with maximal inhibition of activity approx. 90% for WCPI and 95% for WSCI (excess inhibitor over gut extract protein) (Figure 5.16). A control protein (ovalbumin) gave no inhibition when present at similar amounts. The high level of inhibition given by excess WCPI, which is specific for cysteine proteinases, confirms that most of the gut proteolytic activity detected is due to these enzymes. WSCI was more effective inhibitor of the two over the whole range tested (0 – 4μg inhibitor per 3μg gut extract protein), inhibiting by >80% when present at 0.5μg per assay compared to approx. 60% for WCPI, and showing no increase in inhibition after 2μg per assay. However, both inhibitors gave curved plots for activity vs. amount of inhibitor.

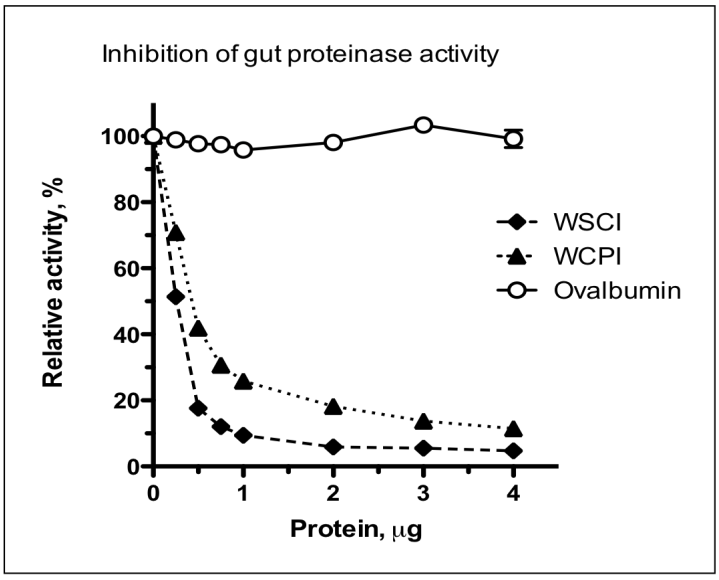


Figure 5.16 Inhibition of *S. avenae* gut proteolytic activity by recombinant WSCI and WCPI. Points show mean values ± SE (n=3)

5.8 Effect of WSCI and WCPI on S. avenae gut proteolytic activity in vivo (feeding bioassays)

To study the effects of recombinant protease inhibitors WSCI and WCPI on *S. avenae* gut proteases *in vivo*, these PIs were added to the (-AA) diet containing a digestible protein ovalbumin (1mg/ml). Aphids were fed on this diet for 72 hours and equal volumes of diet and honeydew were analysed by SDS-PAGE followed by silver staining. Bands corresponding to inhibitors were present in honeydew collected from aphids feeding on WCPI- and WSCI-containing diets, showing that the inhibitors themselves were not digested by aphid gut proteolytic activity. The amount of ovalbumin detected in honeydew was very much lower than in the diet, suggesting almost complete proteolytic degradation in gut by proteases (figure 5.17; lane 6). However, diets to which inhibitors were added showed presence of comparatively more ovalbumin in it, than control honeydew. This presumably due to inhibition of gut digestive proteases by recombinant inhibitors due to which proteases could not act on ingested ovalbumin. The relative band intensities suggested that WSCI was more effective inhibitor of gut proteases *in vivo* than WCPI, giving higher amounts of intact ovalbumin in honeydew (figure 5.17; lane 8 and 10). Ovalbumin was seen to be stable in diet for a period of 72 hours without any self-degradation and was seen in the diet analysed along with honeydew lane. These results support the hypothesis that recombinant PIs could protect ingested ovalbumin from the proteolysis by gut proteases and gut proteases of *S. avenae* play an important role in gut proteolytic digestion as seen in other aphids. The results reveal that *S. avenae* if forced to feed on diet containing no free amino acids and provided with additional digestible protein, could utilise the gut proteases to digest the ingested protein. Though the digestion of ingested protein cannot completely compensate the dietary requirement of aphids. Activity of recombinant WSCI and WCPI in the insect gut and inhibition as seen by protection of ovalbumin in highly proteolytic gut environment suggests that *S. avenae* gut proteases can be utilised for breaking down the ingested protein.

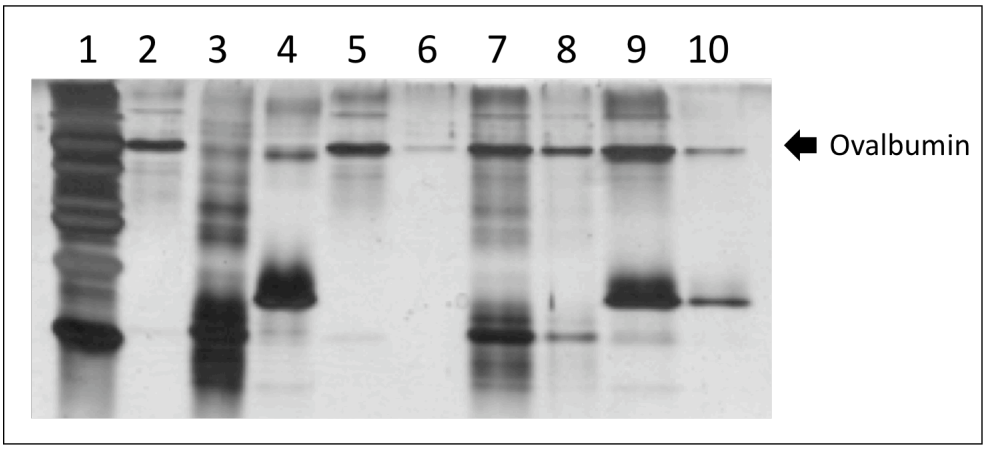


Figure 5.17 Analysis of honeydew by SDS-PAGE followed by silver staining to determine fate of ingested ovalbumin. Aphids were fed for 72 hours on diet without free amino acids (-AA) and supplemented with 1mg/ml Ovalbumin and 1mg/ml concentration of WSCI and WCPI. Honeydew was collected after 72 hours and samples were run on SDS-PAGE. Lane1: Molecular weight marker, Lane2: Ovalbumin (2.5 µg), Lane3: WSCI (2.5 µg), Lane4: WCPI (2.5 µg), Lane5: Control Diet, Lane6: Control honeydew (honeydew collected from aphids that did not contain any inhibitor), Lane7: Diet (-AA+Ovalbumin +WSCI), Lane8: honeydew (-AA+Ovalbumin+WSCI), Lane9: Diet (-AA+Ovalbumin+WCPI), Lane10: honeydew (-AA+Ovalbumin+WCPI). Respective diets were loaded to compare the degradation level of ovalbumin (~45kD band) in honeydew. Control honeydew shows almost complete degradation ingested ovalbumin whereas in WSCI honeydew, WSCI protects degradation of ingested ovalbumin from the activity of gut proteases and WCPI shows a protection of ovalbumin to some extent.

5.9 Effects of recombinant PIs on S. avenae growth and survival:

As the recombinant inhibitors fed to cereal aphids did not show any proteolytic degradation in gut, they were used to replace ovalbumin as proteins substrates in aphid diets lacking free amino acids ("-AA" diet), aphid mortality was significantly increased (from approx. 30% to approx. 80%; $p < 0.01$, log rank test, survival analysis) over an 8-day assay (Figure 5.18). However, aphid performance was generally poor on this diet, and the effects of inhibitors may thus be over-estimated.

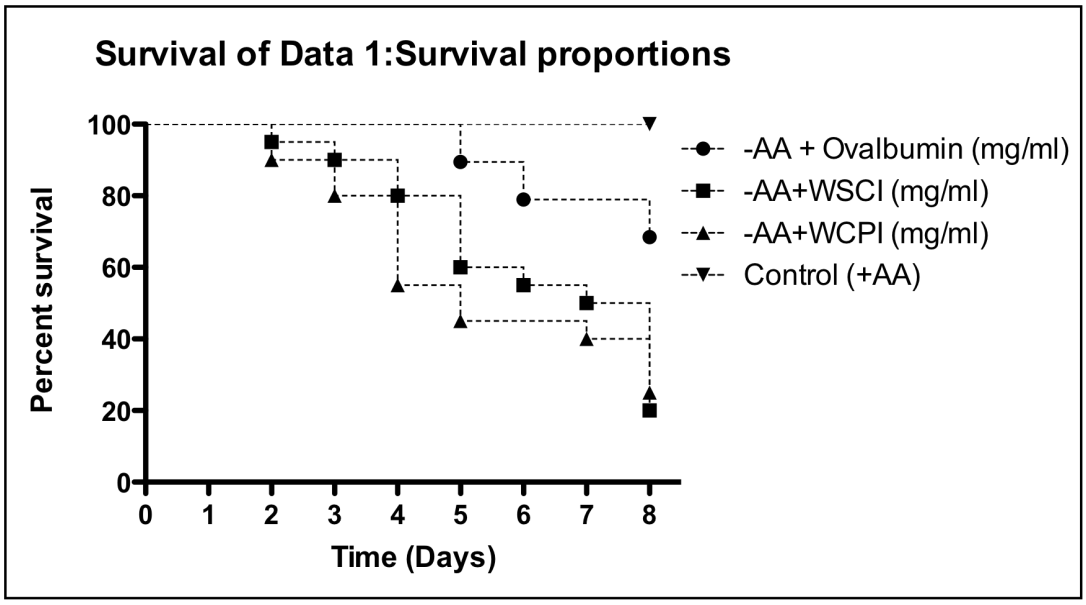


Figure 5.18 Antimetabolic effects of recombinant WSCI and WCPI on survival of *S. avenae* when added to diet without free amino acids (-AA). Aphids fed on control diet (+AA) with free amino acids did not show any mortality. Ovalbumin (1mg/ml) was used as protein control and did not show significant mortality. Recombinant PIs when added (1mg/ml) to (-AA) diet, aphids showed significant mortality upto 80% by day 8 of feeding assay (N=20 per assay). Survival data was analysed using Graphpad Prism software.

The presence of a large excess of free amino acids in the normal aphid diet ("+AA" diet) would be expected to prevent dietary inhibitors having any effect on cereal aphids, since digestive proteolysis would not be required. However, both WSCI and WCPI showed significant effects on aphid growth and survival even when fed as components of this diet. After 10 days, aphids fed control diet or diet + ovalbumin showed 100% survival, whereas aphids fed +AA diet containing WCPI or WSCI showed 30% or 65% mortality respectively. The inhibitors had a negative effect on

growth of surviving aphids, reducing length and width by 30-35%, with WSCI marginally more effective than WCPI (figure 5.19 A and B).

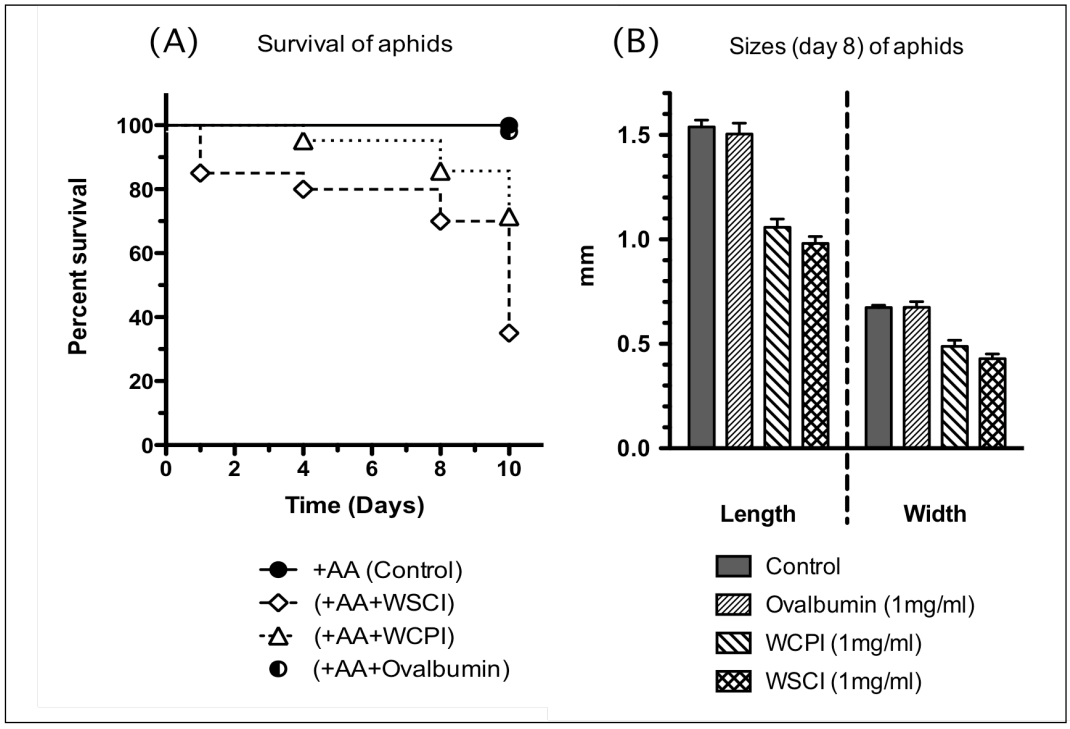


Figure 5.19 Antimetabolic effects of recombinant WSCI and WCPI on *S. avenae* when fed with full supplement of free amino acids (+AA diet) in presence or absence of digestible protein ovalbumin (1mg/ml). **(A) Survival:** Aphids showed better survival on the full free amino acid supplied diet (+AA) as compared to no free amino acid (-AA). Addition of ovalbumin to the (+AA) diet did not show any significant change in the survival. Aphids fed on diet containing WSCI and WCPI showed mortality with WSCI being more antimetabolic than WCPI. Survival data was analysed using Graphpad Prism software. **(B) Growth:** Addition of ovalbumin to (+AA) diet did not improve the growth of aphids and was similar to (+AA) diet. Addition of WSCI and WCPI to (+AA) diets caused retardation in the growth of the aphids. WSCI was more antimetabolic than WCPI (N=20 per assay).

To approximate the potential effects of WSCI and WCPI in a diet closer to phloem sap, containing free amino acids and protein, a diet with half the normal concentration of free amino acids, which could be supplemented with ovalbumin at 1 mg/ml, was used ("½ AA±Ovalbumin" diet). Aphids feeding on (½ AA) diet and (½ AA+ Ovalbumin) diet did not show significant mortality or growth inhibition compared to control aphids fed on +AA diet, with 100% survival up to 12 days, by which time

aphids were mature. Diet containing (½ AA+ Ovalbumin) and 1 mg/ml WSCI decreased the survival of aphids significantly, with 100% mortality by day 10 of assay. WCPI added at 1 mg/ml to the to (½ AA+ Ovalbumin) diet also decreased survival, with approx. 60% mortality at day 10, increasing to about 70% mortality by day 14 (Figure 5.20 A). The inhibitors also had a negative effect on growth of surviving aphids. Protein supplementation had a small positive effect on aphid growth on the (½ AA) diet, with aphids feeding on (½ AA+ Ovalbumin) diet showing slightly better growth than aphids feeding on (½ AA) diet (approx. 10% increase in length and width at day 4). However, growth was reduced by both inhibitors, with WSCI showing greater reductions than WCPI (approx. 40% and 20% reduction in length and width for WSCI and WCPI, respectively, compared to (½ AA+ Ovalbumin diet) (Figure 5.20 B).

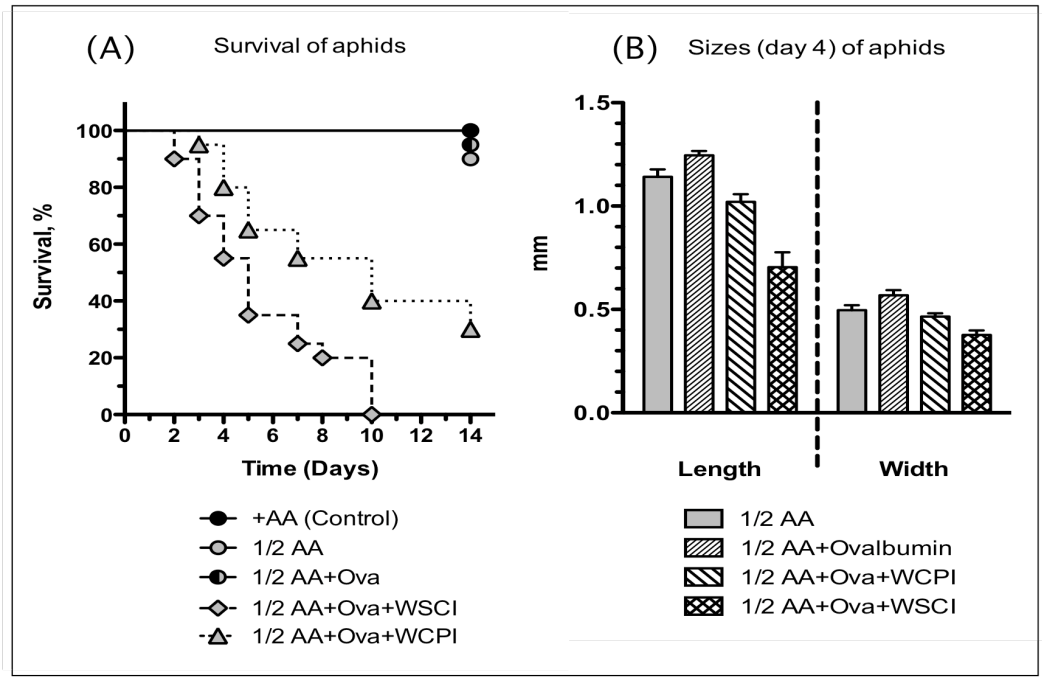


Figure 5.20 Antimetabolic effects of recombinant WSCI and WCPI on *S. avenae* fed with half supplement of free amino acids (1/2 AA) in presence or absence of digestible protein ovalbumin (1 mg/ml). (A) Survival: Aphids feeding on control diet (+AA), half amino acids diet (1/2 AA) and (1/2 AA + Ovalbumin) showed almost no mortality in comparison to diets with inhibitors. WSCI shows a greater mortality as compared to WCPI. Survival data was analysed using Graphpad Prism software. **(B) Growth**: Aphid feeding on (1/2 AA) diet supplemented with ovalbumin showed slightly better growth that (1/2 AA) fed aphids. Growth retardation was observed in inhibitor containing diets. WSCI had a greater effect on aphid growth than WCPI. (N=20 per assay).

The data presented here establishes that cereal aphids are able to digest ingested protein, and can use it to partially compensate for a lack of free amino acids in their diet. However, dietary protein is not an adequate substitute for free amino acids, suggesting that these insects do not have sufficient protein digestive capacity to supply their nutritional requirements. A recent analysis of sieve tube exudate from wheat by aphid stylectomy has shown that mean total amino acid concentration in phloem sap is approx. 460mM, varying from approx. 130mM – 1050mM in different samples (Gattolin *et al.,* 2008). The predominant amino acid in phloem was glutamate (approx. 170mM), with serine (approx. 60mM) and aspartate (approx. 50mM) as the other significant components. Concentrations of other amino acids were <25mM. These estimates are higher than previous values for wheat (total amino acid concentration approx. 260mM; Hayashi and Chino, 1986), but the proportions of different amino acids are similar. The phloem sap of wheat contains at least 100 soluble proteins at an estimated total concentration of 0.1mg/ml (Fisher *et al.,* 1992), although a more recent study in barley identified over 250 phloem proteins, at a total concentration of approx. 0.4 mg/ml. The maximum level of amino acids in the form of protein in phloem of host plants for cereal aphids is thus approx. 3mM (Gaupels *et al.,* 2008), which is insignificant compared to the level of free amino acids, reinforcing the commonly-held view that aphids do not need to carry out protein digestion for nutritional requirements. So why is proteinase activity present? One possibility is that the digestive proteinase activity allows a limited supply of amino acids to be generated from ingested proteins if required. This might give the aphid more flexibility in responding to changing phloem composition (Gattolin *et al.,* 2008), or to deficiencies of essential amino acids (Douglas *et al*., 2009).

Both recombinant PIs showed antimetabolic effect on *S. avenae* affecting survial and growth. Use of such PIs in providing wheat resistance against aphids could be exploited to further. Some aphids are shown to be insensitive to serine protease inhibitors from plants, however some members of Bowman-Birk family (BBI) of protease inhibitors have shown significant toxicity to *A. pisum* (Rahbe *et al.,* 2003b). WSCI and WCPI are the potential candidates in wheat those can be utilised to provide wheat resistance against *S. avenae* and possibly other phloem feeding insects. Transgenic wheat over-expressing or constitutively expressing either of these PIs could pertain resistance to wheat against *S. avenae.*

Chapter 6

Other approaches to control *D. coarctata*:

6.1 Introduction: Lectin-like endogenous defence proteins in wheat:

Lectins, or carbohydrate-binding proteins, are accumulated in plant storage and reproductive tissues, and are considered to act as defensive compounds against herbivores and pathogens (Murdock and Shade, 2002). These proteins are widely distributed in plants, and, like proteinase inhibitors, may play multiple roles in physiology and interactions with foreign organisms. Wheat Germ Agglutinin (WGA) is the major lectin from wheat; it is accumulated in grains, and has been extensively studied and characterised (Levine *et al.,* 1972). Three isoforms of WGA, isolectins A, D and B have been identified, and cDNAs encoding them have been cloned and characterised (Smith and Raikhel, 1989; Raikhel and Wilkins, 1987). Although WGA has not been directly shown to be causative in resistance of wheat to insect pests, the purified protein has insecticidal activity, and several reports of its antimetabolic effects on insects have been published. WGA showed antimetabolic effects on nymphs of brown leafhopper (*Orosius orientalis*), which is an important vector of many viruses and phytoplasmas worldwide (Trebicki *et al.,* 2009). Artificial feeding of WGA to cowpea weevil larvae (*Callosobrochus maculates* L.) delayed larval development (Murdcok *et al.,* 1990). A dipteran insect Blowfly (*Lucilia cuprina* Weiderman) showed inhibited larval growth and increased larval mortality when fed with WGA (Eisemann *et al.,* 1994). WGA-fed European corn borer (*Ostrinia nubilalis* Hubner) showed abnormalities in the peritrophic matrix (PM) structure including large holes in the chitin meshwork of the envelope and disruption of the microvillar organization in the midgut (Harper *et al.,* 1998). These toxic properties of chitin binding WGA-like lectins may result in them being perceived by the insect as feeding deterrent compounds after detection by gustatory receptors. The resulting reduced nutrient consumption could then lead to death by starvation (Giovanini *et al.,* 2007).

Lectin-like proteins with a possible defensive role have also been reported as specifically expressed in wheat in response to herbivory by Hessian fly (*M. destructor*). Interactions between *M. destructor* and wheat operate on a gene-for-gene basis and result in up-regulation of expression of a series of genes designated *Hessian fly responsive* (*Hfr*) in wheat, including three extensively studied genes designated *Hfr-1*, *Hfr-2* and *Hfr-3*. The protein products Hfr-1, Hfr-2 and Hfr-3 all contain at least one lectin-like domain, which may be important in defence against *M. destructor. Hfr-1* is constitutively expressed in wheat, however, mRNA levels were found to increase with the infestation of a virulent strain of *M. destructor*. The Hfr-1 protein was found to be similar to a high mannose N-glycan-specific lectin-like protein, maize beta-glucosidase aggregating factor (BGAF) (Williams *et al.,* 2002), which plays an important role in defence against insects (Blanchard *et al.,* 2001). Hfr-1 has been recombinantly expressed in soluble and active form in *E. coli* and has been characterised (Subramanyam *et al.,* 2008). It was found to show antinutritive effects on *D. melanogaster* (Diptera) larvae (Subramanyam *et al.,* 2008). Hfr-2 was found to have a region similar to genes encoding seed-specific agglutinin proteins from *Amaranthus spp.* and additionally it contained a region similar to haemolytic proteins from both mushroom and bacteria that are able to form pores in cell membranes of mammalian red blood cells (Puthoff *et al.,* 2005). Hfr-3 is a lectin-like protein similar to wheat germ agglutinin with 68-70% amino acid identity (Giovanini *et al.,* 2007). As Hfrs are specifically expressed in response to herbivory by Hessian fly, they could equally be effective against other dipteran insects like *D. coarctata.*

6.2 Hessian fly responsive proteins (Hfrs) from wheat:

6.2.1 Cloning and sequence characterisation of Hfr-1, Hfr-2 and Hfr-3:

Complete coding sequences of Hfr-1, 2 and 3 were cloned by RT-PCR from wheat leaf. A complete sequence characterisation of cDNA coding for Hfr-1 (Genbank accession number **AF483596**) has been reported earlier by Williams *et al.,* (2002). The coding sequence of Hfr-1 consists of a single open reading frame of 1035 nucleotides coding for a 37.7 kDa protein containing 345 amino acid residues. It was analysed for the presence of signal peptide and was found to be a non-secreted protein. Hfr-1 contains a 'dirigent' domain as represented by amino acids L^{44} to C^{179} and Jacalin like mannose-binding lectin domain represented by D^{191} to V^{337} (figure 6.1). The protein has 44% amino acid sequence similarity to the Maize β–glucosidase-

aggregating factor (BGAF; see above). Sequence comparisons with other similar proteins from rice, crocus, hedge bindweed, Jerusalem artichoke have suggested mannose-binding jacalin-like-related lectin activity (Bourne *et al.,* 1999).

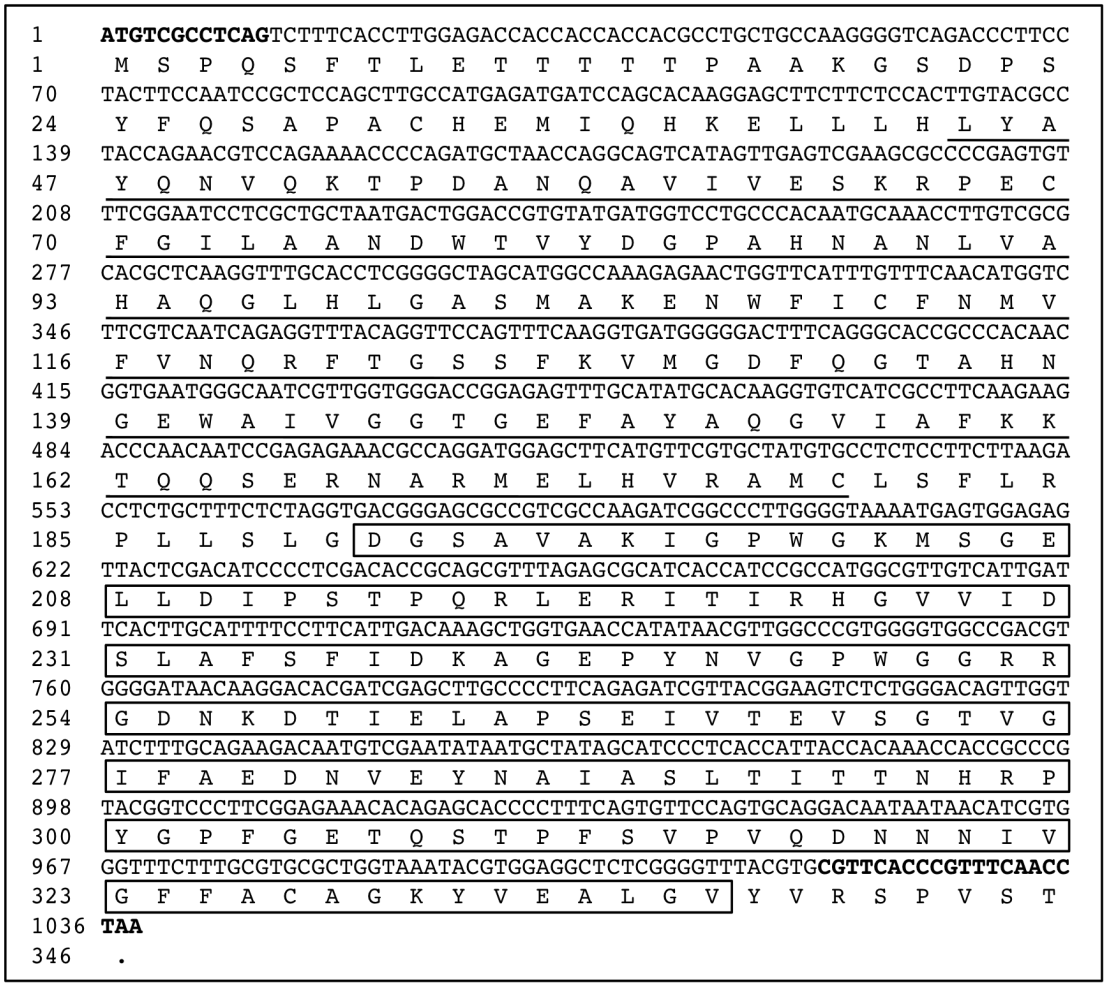


Figure 6.1 Nucleotide and deduced amino acid sequence of Hfr-1 (Williams *et al.,* 2002). Nucleotide sequence shown in bold was used for PCR amplification of coding sequence. "dirigent" domain is underlined (residues L44 to C179) Jacalin-like mannose-binding lectin domain is boxed (residues D191 to V337).

Hfr-2 (Accession number **AY587018**) consists of a single open reading frame of 1479 nucleotides (493 amino acid residues) with two different conserved domains, a lectin like domain and a pore-forming toxin like domain. It was analysed for the presence of signal peptide and was found to be non-secreted protein. Residues Y^{189} to Q^{282} code for an agglutination domain and residues F^{331} to F^{452} code for a pore-forming domain (Puthoff *et al.,* 2005) (figure 6.2).

```
1     ATGTTTCCAGTGACGGGCTCGCCGAAATCCGTTGCTTTCCGATCAAAGCATACCCGCAAGTACCTAGGT
1      M  F  P  V  T  G  S  P  K  S  V  A  F  R  S  K  H  T  R  K  Y  L  G
70    AGCGTGCAAGCAGGGAGCGAGGAGAGCGCCGGCGGAGGCAAGTTCTTCGAGGAGCTGAGCCACGGCGCC
24     S  V  Q  A  G  S  E  E  S  A  G  G  G  K  F  F  E  E  L  S  H  G  A
139   GACGACGTCGATGTCCTTGCAAGCCCGTACACTAGATTCTACCTGGAGCCATCCAAGGAGCACGACGGG
47     D  D  V  D  V  L  A  S  P  Y  T  R  F  Y  L  E  P  S  K  E  H  D  G
208   CTCCTGCACGTCAGGTGCTGCCACAACAACAAGTACTGGGTGGCCAAACATGTCGGTGAAGGCAGCGGC
70     L  L  H  V  R  C  C  H  N  N  K  Y  W  V  A  K  H  V  G  E  G  S  G
277   CACTGGATCATTGGCATCGTCAATGAACCGGAGGACGACCTGTCCAAGCCGTCATGCACGCTTTTTGAG
93     H  W  I  I  G  I  V  N  E  P  E  D  D  L  S  K  P  S  C  T  L  F  E
346   CCCATCCCTCTCGCGGACACGGACAATAATCTATCCATCAGGTTCTTCCGTCCTCAGCAGACAACAAGC
116    P  I  P  L  A  D  T  D  N  N  L  S  I  R  F  F  R  P  Q  Q  T  T  S
415   TCTGAATCTGATATGACCAAGGAAAAGGGGACAACTGAAGAAGCCTACCTGTTTCTGGGAACCGGAGGG
139    S  E  S  D  M  T  K  E  K  G  T  T  E  E  A  Y  L  F  L  G  T  G  G
484   CATGAAAAAGCAGTTGATCAAGTTAAATCTCTTCATGACTTCTCCGCCATTGATCTGTCGAAGCAATTG
162    H  E  K  A  V  D  Q  V  K  S  L  H  D  F  S  A  I  D  L  S  K  Q  L
553   GTACTACCTAAATATGTTGCTTTTAAAGGCGACAATGACATGTACCTCCGGGCAAGGATCATCCAGAAA
185    V  L  P  K  Y  V  A  F  K  G  D  N  D  M  Y  L  R  A  R  I  I  Q  K
622   CGCAATTACCTGGAATTCTCATCATCTGATATTGCAGATTCAACTGTGGTCAACACTATTTTCCCCAAC
208    R  N  Y  L  E  F  S  S  S  D  I  A  D  S  T  V  V  N  T  I  F  P  N
691   TATGCTAATGGGAACGTGCGCATAAAATCTAACCACTTCAATAGGTTTTGGAGGCTCAGCCCCAACTGG
231    Y  A  N  G  N  V  R  I  K  S  N  H  F  N  R  F  W  R  L  S  P  N  W
760   ATCTGGGCTGACTCGGCCGACACCAGTAGCAGAGACCGTGACACGCTCTTCAGGGTGGTCATGTTGCCC
254    I  W  A  D  S  A  D  T  S  S  R  D  R  D  T  L  F  R  V  V  M  L  P
829   GACTACATCGGTCTCCAGAACCTAGGAAACTCCAGGTACTGCAAGAGGCTAACTGCCGACAAGAAGACA
277    D  Y  I  G  L  Q  N  L  G  N  S  R  Y  C  K  R  L  T  A  D  K  K  T
898   AGCTGCCTTAACGCCGCCGTCGATACCATCACCCTTGAAGCAAGGTTACGGGTGGAAGAAGCCGTACTT
300    S  C  L  N  A  A  V  D  T  I  T  L  E  A  R  L  R  V  E  E  A  V  L
967   TCGCGAGAGGTCTATGGCGTGGAGTTCAAGCTCTCCGAAGCTAGGATCTACGGCGAGAAGCCTCTTACC
323    S  R  E  V  Y  G  V  E  F  K  L  S  E  A  R  I  Y  G  E  K  P  L  T
1036  TTTCCCAGCATGACTTCAACCAATGACACCAACGAAACACATGCCAAGACCCTGACCCTGAAATACGAG
346    F  P  S  M  T  S  T  N  D  T  N  E  T  H  A  K  T  L  T  L  K  Y  E
1105  GAGACGCAGGCCAAGACCTGGAGCTCGACTGTTAGCCTCAAGATCGGTGTCACGGCCAAGTTAAGAGCC
369    E  T  Q  A  K  T  W  S  S  T  V  S  L  K  I  G  V  T  A  K  L  R  A
1174  GGGATCCCGGTCATAGCAGAAGGGAAGGTTGAGGTATCCACTGAATTCAACTCAGAGTACGAGTGGGGG
392    G  I  P  V  I  A  E  G  K  V  E  V  S  T  E  F  N  S  E  Y  E  W  G
1243  TCATCCATTCAGACAAGGACATCACAAGAGGCCTCCTACCAGGCTGTGGTGCCTCCGATGACCAAGGTG
415    S  S  I  Q  T  R  T  S  Q  E  A  S  Y  Q  A  V  V  P  P  M  T  K  V
1312  ACAATCAGAGCGGCTGCGACTCAGGGTTCTATTGACGTCCCGTTCTCCTACACTCAGAGGGACATTCTG
438    T  I  R  A  A  A  T  Q  G  S  I  D  V  P  F  S  Y  T  Q  R  D  I  L
1381  ACCACAGGAGAGGTTGTTACCTACAAGATGGACGATGGCCTGTTCACTGGTATGAACAACTATAATTTC
461    T  T  G  E  V  V  T  Y  K  M  D  D  G  L  F  T  G  M  N  N  Y  N  F
1450  CAATTTGAAGCCACACAAGAGCCCATCTGA
484    Q  F  E  A  T  Q  E  P  I  .
```

Figure 6.2 Nucleotide and deduced amino acid sequence of Hfr-2 (Puthoff *et al.*, 2005). Nucleotide sequences shown in bold were used for primers to clone Hfr-2. Shadowed residues (Y189 to Q282) represent agglutination domain. Shadowed and underlined residues (F331 to F452) represent Pore forming domain.

Amino acid sequence analysis revealed that Hfr-3 (accession number **DQ462308**) consists of a single open reading frame of 594 nucleotides coding for 198 amino acid residues. Giovanini *et al.,* (2007) have reported a structural analysis of *Hfr-3* gene and no introns were observed. Hfr3 was analysed for presence of signal peptide and found

to be a secreted protein containing a signal peptide of 20 amino acid residues (MKGLLLCALAFAAVTTHA). Hfr-3 amino acid sequence shows presence of four hevein domains, (1) residues T^{27} to N^{72}, (2) K^{73} to D^{114}, (3) L^{115} to D^{157} and (4) K^{158} to N^{198} with saccharide binding residues in them (figure 6.3A). The deduced amino acid sequence of Hfr-3 was found to be 68% identical to *Triticum aestivum* WGA-A (e-value 6e-67; Accession no. **P10968**), 70% identical to WGA-B (e-value 2e-69; Accession no. **P10969**) and 69% identical to WGA-D (e-value 5e-68; Accession no. **P02876**). However ClustalW-Boxshade alignment of Hfr-3 sequence with wheat germ agglutinin from *Triticum turgidum* (accession no. **AAA34257**), showed almost 80% similarity to Hfr-3 (figure 6.3B).

```
1    ATGAAGGGCCTCTTGCTGTGCGCGCTCGCACTTGCCTTTGCTGCGGTTACCACCCACGCCCAGCTGCAG
1     M  K  G  L  L  L  C  A  L  A  L  A  F  A  A  V  T  T  H  A  Q  L  Q
70   TCCTGCCCGACACGCTGCGGCAAGCAGGCCGACGGCATGGAGTGCCCCAACAACCTCTGCTGCAGCAAG
24    S  C  P  T  R  C  G  K  Q  A  D  G  M  E  C  P  N  N  L  C  C  S  K
139  GATGGGTACTGCGGCCTGGGCGTCGACTACTGCAGCGCTGGCGCCGGCTGCCAGAGCGGCGCCTGCTAT
47    D  G  Y  C  G  L  G  V  D  Y  C  S  A  G  A  G  C  Q  S  G  A  C  Y
208  GACAACAAGATCTGCGGCGCGCAGGCCAATGGTACATTGTGCCGTAACAACCACTGTTGTAGCTCGGGT
70    D  N  K  I  C  G  A  Q  A  N  G  T  L  C  R  N  N  H  C  C  S  S  G
277  GGTCGCTGCGGCTACGGCAGAGAATACTGCAGCAACGGCTGCCAGGGCGGTCCTTGCTGGGCTGATCTC
93    G  R  C  G  Y  G  R  E  Y  C  S  N  G  C  Q  G  G  P  C  W  A  D  L
346  AAGTGTGGCCACCTGGACAATGGTAAGCTATGCCCCAACAACCTCTGCTGCAGCCAGTACGGTTACTGT
116   K  C  G  H  L  D  N  G  K  L  C  P  N  N  L  C  C  S  Q  Y  G  Y  C
415  GGCCTAGGACCGGAGTTCTGTGGTACTGGCTGCCAAAATGGCGCTTGCAGCACAGACAAGCCGTGTGGC
139   G  L  G  P  E  F  C  G  T  G  C  Q  N  G  A  C  S  T  D  K  P  C  G
484  AACAAGGCTAATGGTGCACCATGCACCAACAACTATTGTTGCAGCCAGTATGGGTCTTGTGGTCTTGGC
162   N  K  A  N  G  A  P  C  T  N  N  Y  C  C  S  Q  Y  G  S  C  G  L  G
553  AAGGATTACTGTGGTACCGGCTGCCAGAATGGTGCATGCAACTAG
185   K  D  Y  C  G  T  G  C  Q  N  G  A  C  N  .
```

Figure 6.3 (A) Nucleotide and deduced amino acid sequence of Hfr-3. Predicted signal peptide is shadowed. Putative four hevein domains are shown domain-1: residues in italics, domain-2 is italics and underlined, domain-3 is in bold and domain-4 is bold and underlined. Saccharide-binding residues conserved in hevein domain are boxed. Nucleotide sequences shown in italics were used for primer design to clone Hfr-3

Figure 6.3 (B) ClustalW-Boxshade alignment of amino acid sequence of Hfr3 with Wheat germ agglutinin (WGA). Deduced amino acid sequence of Hfr3 was aligned with amino acid sequence of *Triticum turgidum* available at NCBI (Accession No.**AAA34257**). Residues with black background are highly conserved, residues with grey background are partially conserved.

6.2.2 Recombinant expression and purification of Hfr-3

Many lectins and lectin-like proteins have been expressed as recombinant proteins, but expression in *E. coli* has generally yielded insoluble proteins which have to be refolded by denaturation-renaturation techniques. The use of *P. pastoris* as an expression system has been successful in producing a number of monocot lectins as soluble, functional proteins (Fitches *et al.,* 1997). Garlic lectin-like agglutinins (*Allium sativum* agglutinin; ASA-I and ASA-II) have been expressed recombinantly in *P. pastoris* (Fitches *et al.,* 2008). Based on these reports, *P. pastoris* was used as a recombinant expression system for Hfr-1 and Hfr-3. Hfr-1 has been recombinantly expressed in soluble and active form in *E. coli* and has been characterised (Subramanyam *et al.,* 2008). However, my efforts to express Hfr-1 using *P. pastoris* expression system were not successful.

For recombinant expression of Hfr-3, only mature coding sequence (QLQS.....GACN) excluding a signal peptide was used for the expression construct in pGAPZαB, which was assembled by standard recombinant DNA techniques. A complete coding sequence for mature Hfr-3 containing (178 amino acids) was inserted into MCS of yeast expression vector pGAPZαB, between PstI and NotI restriction enzyme sites in frame with yeast a- mating factor N-terminal secretary signal. The translation product is predicted to undergo co- and post-translational

proteolysis, resulting in a secreted polypeptide containing the complete proprotein sequence, with an extra Ala residue at the N-terminus (from restriction site in expression vector) and a vector derived C-terminal extension of 27 residues containing a Myc antigenic determinant and a $(His)_6$ C-terminal tag (Figure 6.4). Expression construct prepared as above were mobilised into *E. coli* using standard methods. Clones expressing Hfr3 were selected by screening culture supernatants. Probing a western blot with anti $(His)_6$ antibodies showed two immunoreactive bands of approx 40kD and 23kD (figure 6.5 lanes 1 and 2). The predicted molecular weight of recombinant Hfr-3 including the C-terminal Myc and $(His)_6$ is approx. 20kD. Two forms of Hfr-3 were observed in the immunodetection of total protein extract from wheat with Hfr-3 specific antibodies (Giovanini *et al.,* 2007). The molecular mass of monomeric form of Hfr-3 was 18 kD and of the dimer was 32 kD. The immunoreactive band observed at approx. 40kD in our experiments could therefore be a dimeric form of Hfr-3. Similar hetorodimeric or homodimeric forms of lectins were observed for a recombinant garlic lectin (ASA-I) (Fitches *et al.,* 2008). ASA-I was recombinantly expressed in *P. pastoris* and purified ASA-I showed two polypeptides of 13.4 and 14 kDa and some preparations also showed a presence of a protein of 32 kDa and a few high molecular weight bands. All these bands showed immunoreactivity with anti-ASA antibodies and it was concluded that 32 kDa and higher molecular weight bands are the heterodimers of ASA-I, which represent the precursor forms of 13.5 and 14kDa protein with varying degrees of glycosylation. Deglycosylation of these proteins with N-glycosidase improved their migration of gel to some extent (Fitches *et al.,* 2008).

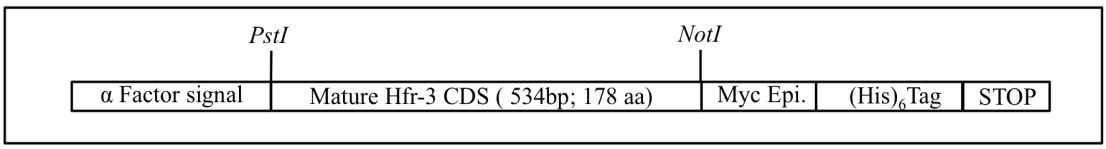


Figure 6.4 (A): Schematic representation of expression construct for Hfr-3 in pGAPZαB.

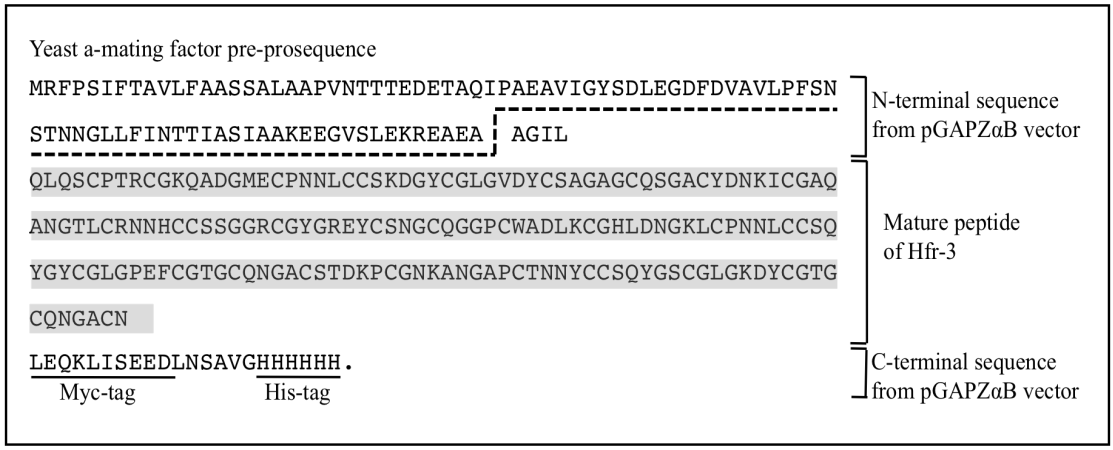


Figure 6.4 (B): Sequence details of expression construct for Hfr-3 in pGAPZαB. A complete coding sequence for mature Hfr-3 containing 178 amino acid residues is inserted into MCS of yeast expression vector pGAPZαB, between PstI and NotI restriction enzyme sites. α-factor signal sequence containing 89 amino acid residues, helps secretion of recombinant protein into the culture media. C-terminus contains a Myc epitope (for immunodetection of recombinant protein) and His tag with six residues (for immunodetection of recombinant protein and purification by metal affinity chromatography)

Recombinant Hfr-3 was purified from culture supernatant by Ni-NTA metal affinity chromatography. Fractions were analysed by SDS-PAGE; Hfr-3 containing fractions showed both 20 kD and 40 kD bands (figure 6.5 lane 3). These bands also showed immunoreactivity with anti-$(His)_6$ antibodies in western blot (figure 6.5 lane 4). A yield of approx. 5-7 mg recombinant Hfr-3 per litre of fermented media was obtained, which is less than other lectins produced in *P. pastoris.* A lectin from *Phaseolus vulgaris* (PHA-E) has been produced in *P. pastoris* with yield of 100 mg per litre of fermented media and a lectin from snowdrop at the rate of 80 mg per litre (Baumgartner 2002 and 2003). Our yield of recombinant Hfr-3 was comparable to yield of 10-20 mg per litre for ASA from garlic (Fitches *et al.,* 2008), although it is not clear whether expression could be improved to obtain greater yields.

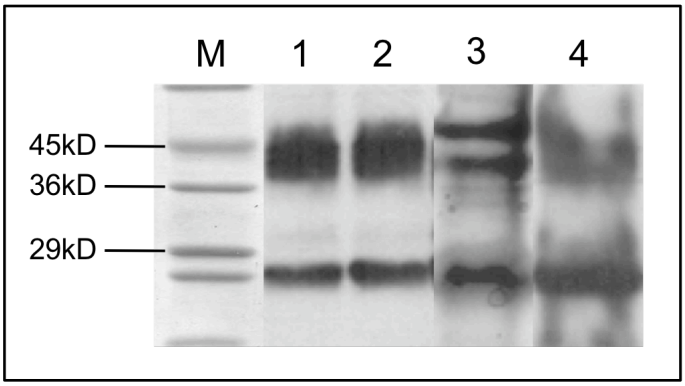


Figure 6.5 SDS-PAGE and western blot analysis of recombinant Hfr-3. Lane M: Molecular weight marker, Lanes 1and 2: Screening western blot for two clones expressing Hfr-3 with Anti His(6) antibodies, Lane 3: Ni-NTA purified Hfr-3 and Lane 4: purified recombinant Hfr-3 and confirmation of purified recombinant Hfr-3 by western blot using anti His(6) antibodies

6.2.3 Antimetabolic effects of Hfr-3 to S. avenae

To study the antimetabolic effects of Hfr-3 on insects, purified Hfr-3 was fed to *S. avenae* by incorporating into artificial diet containing full supply of amino acids (+AA). Due to sequence similarity with Hfr-3 and antimetabolic effects of wheat germ agglutinin on the insects (WGA), it was used as the positive control for the feeding assays. The effects of WGA and Hfr-3 on survival and growth of *S. avenae* were monitored. Aphids feeding on Hfr-3 showed 100% mortality in 8 days of feeding, whereas aphids feeding on WGA containing diet showed only limited mortality (up to 15%; figure 6.6). Growth of aphids feeding on either WGA or Hfr-3 was significantly affected when measured at day 5. Overall growth in aphids feeding on Hfr-3 was reduced by about 50% whereas of those feeding on WGA was reduced by 25% compared to control aphids (Figure 6.7). WGA fed aphids when measured on day 8, were almost similar to control aphids, suggesting that they overcame the effect of WGA. However the aphids feeding on Hfr-3 did not show reversal of the growth inhibition effect. Feeding Hfr-3 at different doses from 0.125mg/ml to 1mg/ml showed dose dependent effect on *S. avenae* survival (figure 6.8). After 8 days of feeding, 1mg/ml Hfr-3 caused 100% mortality whereas at the lowest concentration 0.125mg/ml of Hfr-3 showed about 40% mortality.

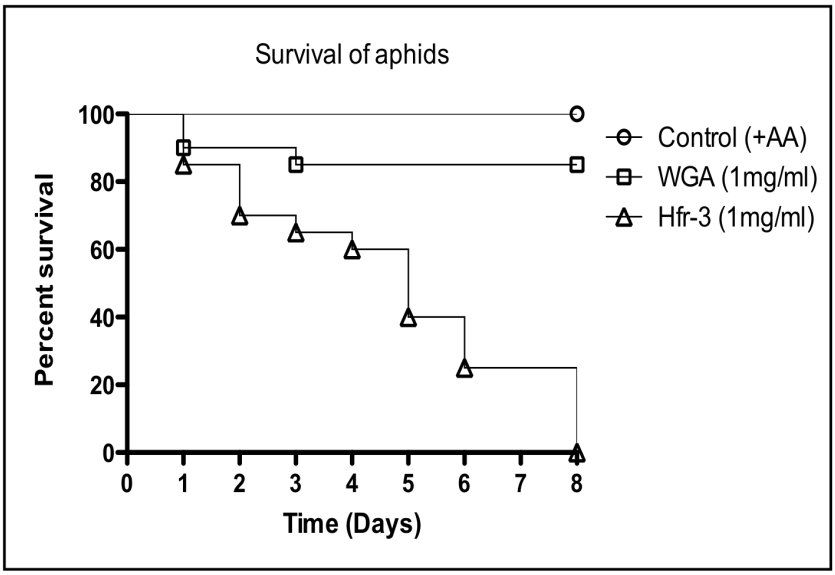


Figure 6.6 Antimetabolic effects of wheat germ agglutinin (WGA) and recombinant Hfr-3 on survival of *S. avenae*. (N=20 per assay). Survival data was analysed using Graphpad Prism software.

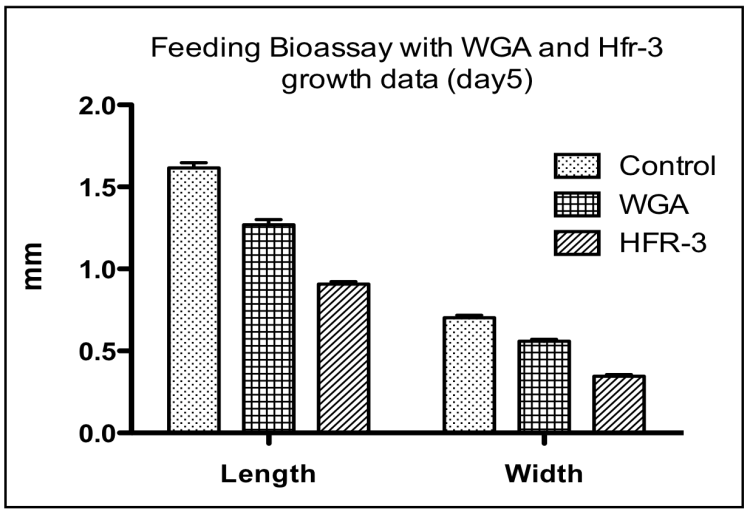


Figure 6.7 Effect of wheat germ agglutinin (WGA) and recombinant Hfr-3 on growth of *S. avenae.* Error bars indicate mean values $\pm$ SE (n=10)

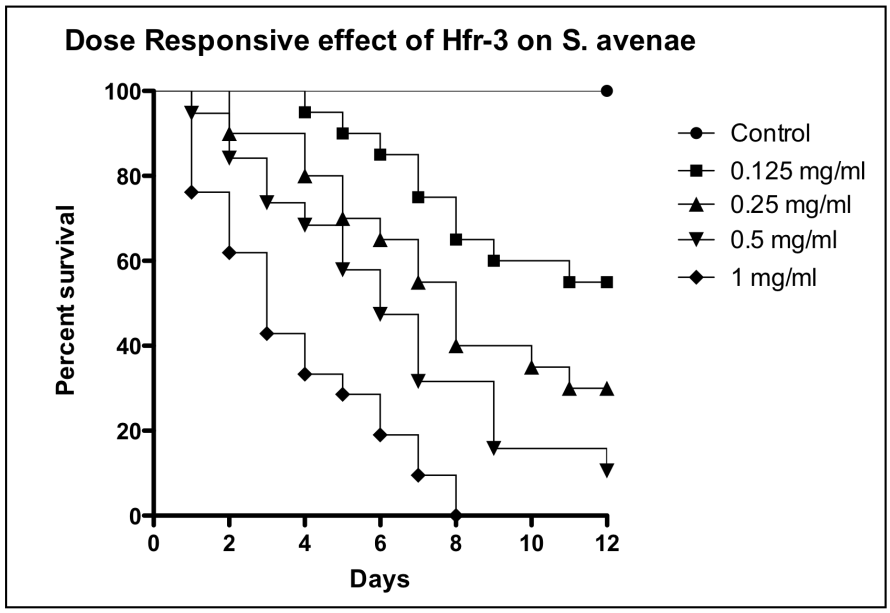


Figure 6.8 Dose dependent effect of recombinant Hfr3 on survival of *S. avenae.* (N=20 per assay). Survival data was analysed using Graphpad Prism software.

Recombinant plant lectins have been shown to have toxic or chronic insecticidal effects on Hemipteran insects. Two recombinant agglutinin-like lectins from Garlic bulbs (ASA-I and ASA-II) showed significant effect on the survival of pea aphids (*Acyrthosiphon pisum*), which was dose dependent (Fitches *et al.,* 2008). The effect of ASA on the aphids was divided into two phases (1) acute toxicity leading to mortality in the initial 48 hours of feeding and (2) chronic toxicity causing mortality over the extended time (upto 28 days). In all feeding assays, aphids feeding on ASA showed less survival that control aphids which were fed on non-ASA diet. These proteins also showed dose dependent toxicity to two other hemipteran insect pests, peach-potato aphid *Myzus persicae* and rice brown plant hopper *Nilaparvata lugens*. The toxicity of Hfr-3 to *S. avenae* is broadly similar to other lectins on hemiptera (Rahbe *et al.,* 1995, Powell *et al.,* 2001), but the high level of mortality shown by Hfr-3 is not observed in many other cases. Transgenic plants expressing mannose-binding lectins have shown deleterious effects on aphid survival and development (Hilder *et al.,* 1995, Down *et al.,* 1996; Rao *et al.,* 1998; Stoger *et al.,* 1999). All these results suggest that lectin like proteins are not only insecticidal against Lepidopteran pests (Fitches *et al.,* 1997) but also to phloem feeding aphids. Hfr proteins could be the potential candidates to be considered for developing insect resistance in wheat against Hemipteran and other

pests. Transgenic wheat plant over-expressing or constitutively expressing such proteins could be a potential approach.

6.2.4 Antimetabolic effects of Hfr-3 to D. coarctata

Purified recombinant Hfr-3 and WGA were fed to neonates of *D. coarctata* at the concentration of 3µg and survival of larvae was monitored for a week. Larvae fed with Hfr-3 showed 50% mortality while the same concentration of WGA showed only 30% mortality (figure 6.9). The effects of Hfr-3 on *D. coarctata* show similar results as for *S. avenae* where Hfr-3 has stronger antimetabolic effects than WGA. These results suggest that endogenous resistance proteins like Hfr-3 could be potential candidates to be considered for protection of wheat against *D. coarctata.*

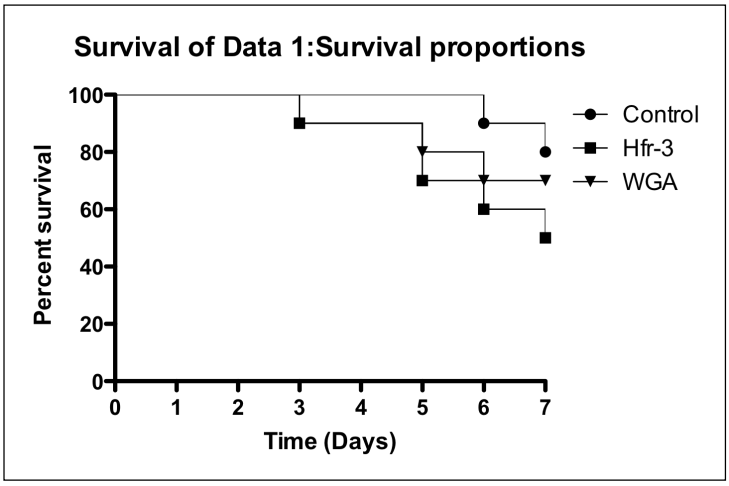


Figure 6.9 Effect of wheat germ agglutinin (WGA) and recombinant Hfr-3 from wheat on survival of *D. coarctata.* Survival was calculated using Graphpad Prism software (N=10).

6.3 Gene knock down by RNA interference in D. coarctata:

Injecting an exogenous dsRNA triggers sequence-specific degradation of target endogenous mRNA in target cell/organism (Jaubert-Possamai et al, 2007) and the functionality of a specific gene in the organism can be silenced. RNAi technology has been useful for systems where stable transgenesis involving other resistance related gene is not available (Jaubert-Possamai *et al.,* 2007). Use of RNAi mediated crop protection by down-regulation of the expression of potentially crucial and specific genes by injecting or feeding dsRNA (double stranded RNA) could be another potential strategy for controlling *D. coarctata*. Several examples of gene knockout

using RNAi in insects are available. *D. melanogaster* (Diptera) is an ideal insect system studies extensively so far for RNAi along with a Coleopteran insect *Tribolium castaneum* in which RNAi system works very well. Two crucial genes, calreticulin (Ap-CRT-I) and cathepsin-L (Ap-Cath-L), have been knocked down by injecting dsRNAs into *Acyrthosiphon pisum* (pea aphid) which transiently silenced the expression of these two genes (Jaubert-Possamai *et al.,* 2007). Three important genes expressed uniquely in the midgut and salivary gland, BtCG5885, BtGATAd and BtSnap were silenced by injecting long dsRNAs in White fly (*Bemisia tabaci* Gennadius; Homoptera), which is an important agricultural pest worldwide. The expression of targeted mRNA was reduced up to 70% compared to the control flies injected with buffer or green fluorescent protein (GFP)-specific dsRNA (Ghanim *et al.,* 2007).

For insect pests, feeding or ingestion of dsRNA is another option to induce RNAi, as injecting is not feasibly and practically possible at field level (Price and Gatehouse, 2008). Feeding dsRNA has been used in some insect system have shown success not as efficient as injecting dsRNA. Silencing of a salivary nitrophorin (NP2; hemeprotein present in salivary glands with multifunctional activities) in *Rhodnius prolixus*, Stal; (Hemiptera) was carried out using both injection and ingestion of dsRNA. RT-PCR of salivary glands revealed that 15μg of NP2 dsRNA (in fourth instar nymphs) reduced gene expression by 75± 14% and feeding 1μg /μl of NP2 dsRNA (in second instar nymphs) reduced gene expression by 42± 10% (Araujo *et al.,* 2006). Feeding of dsRNA at field level would require very large quantities of dsRNA. Host delivered RNAi (HD-RNAi) is a promising solution for feeding dsRNA to insects. Host plants are transformed with RNAi construct and plants express double stranded RNA constitutively. dsRNA is thus delivered to the insects via food while insects' feeding. It is necessary to select the gene for RNAi carefully by taking into consideration many factors such as effect on non-target organisms in the field, this is only possible in RNAi where the gene being used is common to other insect species and shares a good degree of sequence homology. For control of insects by RNAi, the genes related to insects' metabolism that will affect its growth and development are important considerations.

RNAi can also be used to take advantage of plants' innate defence mechanism in the form of production of secondary metabolites. Insects use their detoxifying enzymes to overcome the plant secondary metabolites. These detoxifying enzymes can be silenced to make insects vulnerable to plant secondary metabolites. A cytochrome P450 monoxygenase (CYP6AE14) gene in cotton bollworm is important for detoxification of gossypol present in cotton leaves. Feeding cotton bollworms on transgenic tobacco and *Arabidopsis* producing CYP6AE14 dsRNA showed retardation in growth and sensitivity to gossypol (Mao *et al.,* 2007). However the efficacy of these under field condition is not validated yet, if done so RNAi could be the best alternative to chemical insecticides and traditional insecticidal proteins like Bt endotoxin. Amino acid transporters located in the insect guts play an important role in amino acid turnover and thus are potential targets for the RNAi.

6.4 Amino acid transporters as potential target for RNAi

Biosynthesis of structural and signaling molecules in animals highly depends on intracellular concentrations of essential amino acids. This concentration is maintained by a specific system of plasma membrane transporters. Most insect cell membranes contain uniporters that facilitate the diffusion of amino acids into and out of the cells (Wolfersberger, 2000) and in addition to these there is a passive diffusion system. Insects contain amino acid transporters (AATs); the membrane bound transport proteins belonging to solute carrier family. Many amino acid transport proteins have been investigated and reported so far and categorized in to many gene families (Castagna *et al.,* 1997). The amino acid transporters that transport cationic amino acids across the membrane are called cationic amino acid transporters (CAAT or CAT or KAAT). These AATs help their cells to accumulate certain amino acids from the extracellular medium. CAT proteins are integral membrane proteins containing 12–14 putative transmembrane domains (Kim *et al.,* 1991; Wang *et al.,* 1991; Closs *et al.,* 1993a and 1993b). A potassium coupled cationic amino acid transporter KAAT1 from *Manduca sexta* larval midgut has been cloned and expressed in *Xenopus* oocytes. Its role in coupling with K+ rather than NA+ or H+ to amino acid absorption has been discussed (Castagna 1998). The cationic amino acid transport system y+ mediates the entry of positively charged amino acids down their electrochemical gradient. Cationic amino acids are taken up by cells via y+ system and the system also accepts some neutral amino acids in the presence of Na+, resulting in electrogenic transport. CAT

proteins catalyse Na+ dependent uptake of essential amino acids Arginine, Lysine and Ornithine required which play crucial role in metabolism.

6.5 RNAi of Cationic amino acid transporter from D. coarctata (DcCAAT)

6.5.1 Cloning of DcCAAT, sequence analysis and comparison with dipteran amino acid transporters

Complete coding sequence of DcCAAT was obtained from *D. coarctata* gut tissue by degenerate PCR followed by 3' and 5' RACE. An overview of RACE for DcCAAT is shown in figure 6.10 (A). Partial coding sequence of DcCAAT amplified was 603bp long, and 5' and 3' RACE products were approx. 1000bp and 1300bp respectively. Actual PCR products analysed and cloned from agarose gels are shown in figure 6.10 (B). 3' RACE product had UTR with polyA tail (approx. 250bp) while 5' RACE product had a UTR of approx. 200bp. All RACE products when aligned in Sequencher and analysed gave a complete coding sequence DcCAAT with a single open reading frame of 1624bp with 541 amino acid residues. Complete coding sequence of DcCAAT with deduced amino acid sequence is shown in figure 6.11. Amino acid sequence of DcCAAT was analysed for the presence of transmembrane domains with TMHMM server (http://www.cbs.dtu.dk/services/TMHMM/) showed presence of 11 transmembrane domains as shown in figure 6.12. The details of amino acid residues with transmembrane domains are shown in figure 6.11. Amino acid sequence of DcCAAT was aligned with Dipteran cationic amino acid sequences showed about 90% identical amino acid residues (figure 6.13)

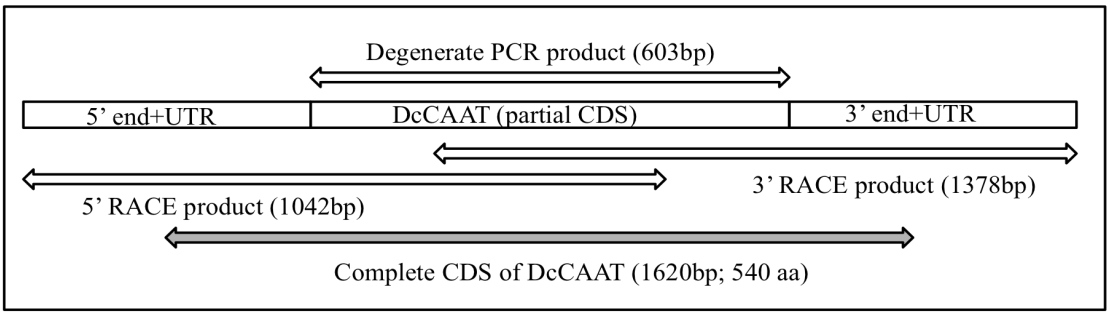


Figure 6.10 (A) Overview of RACE for DcCAAT

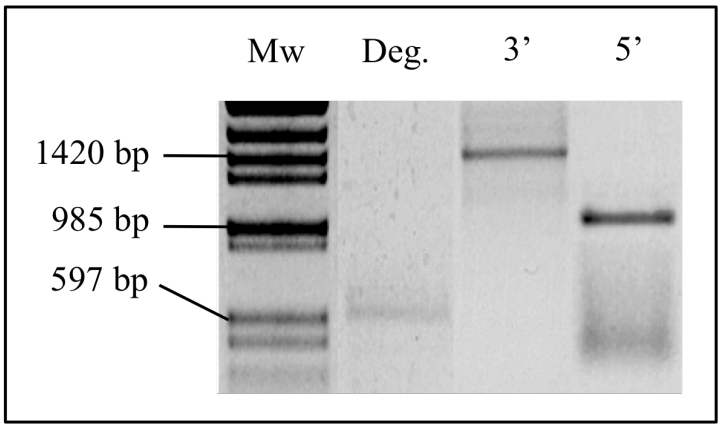


Figure 6.10 (B) Actual PCR amplification products analysed on 1% agarose gel. Lane Deg.: Product of Degenerate PCR, lane3': product of 3' RACE and Lane 5': Product of 5' RACE.

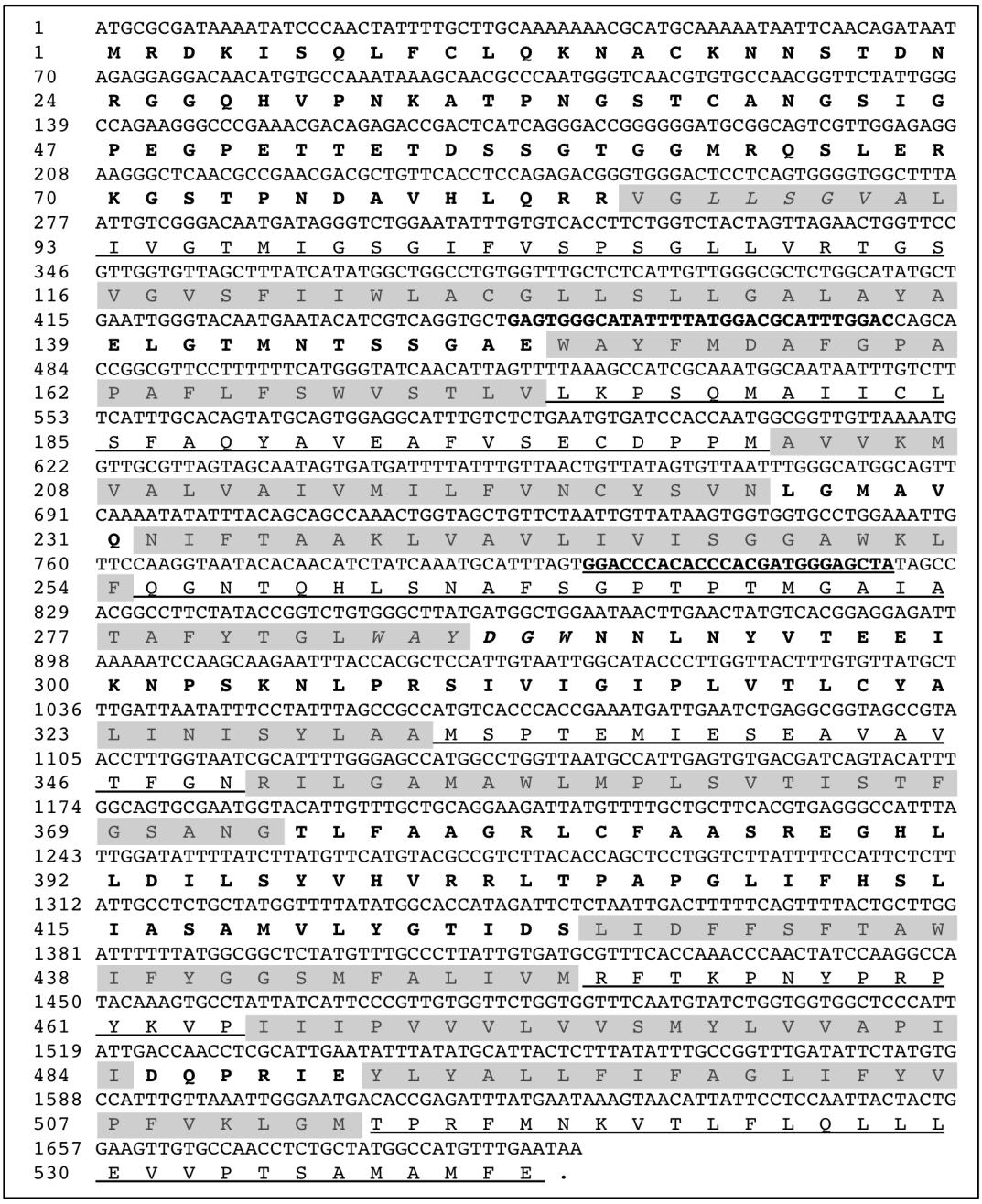


Figure 6.11 Nucleotide and deduced amino acid sequence for complete coding sequence of DcCAAT. Trans-membrane domains are shadowed, IN residues are shown in bold and OUT residues with underline. Nucleotide sequence in italics used for degenerate primers. Nucleotide sequence in bold used as GSP for 3' RACE and bold with underline used as GSP for 5' RACE. Conserved residues shown in italics *LLSGVA* and *WAYDGW* were used to design degenerate primers

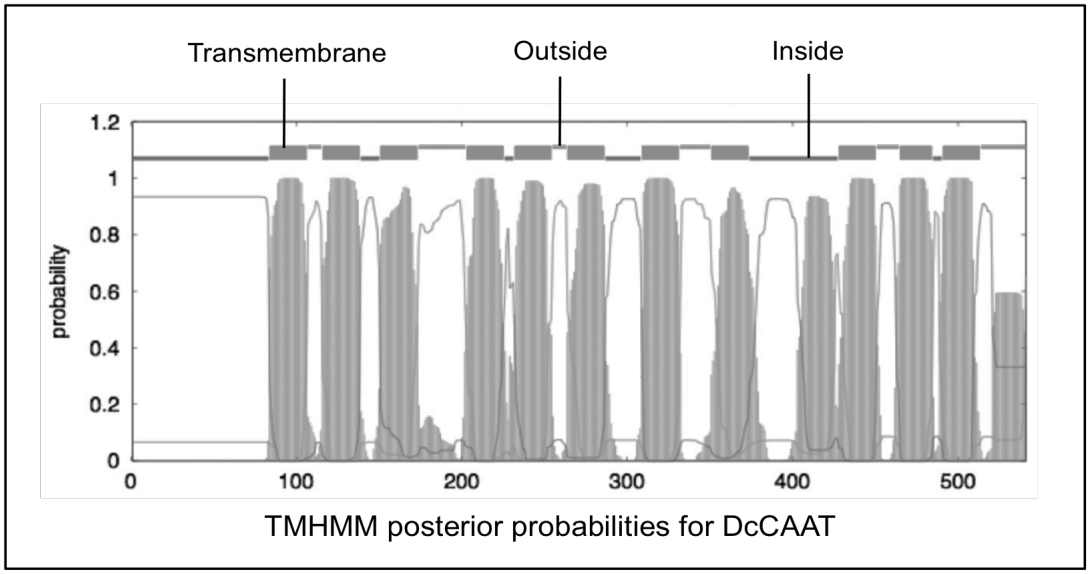


Figure 6.12 TMHMM prediction of transmembrane helices for DcCAAT

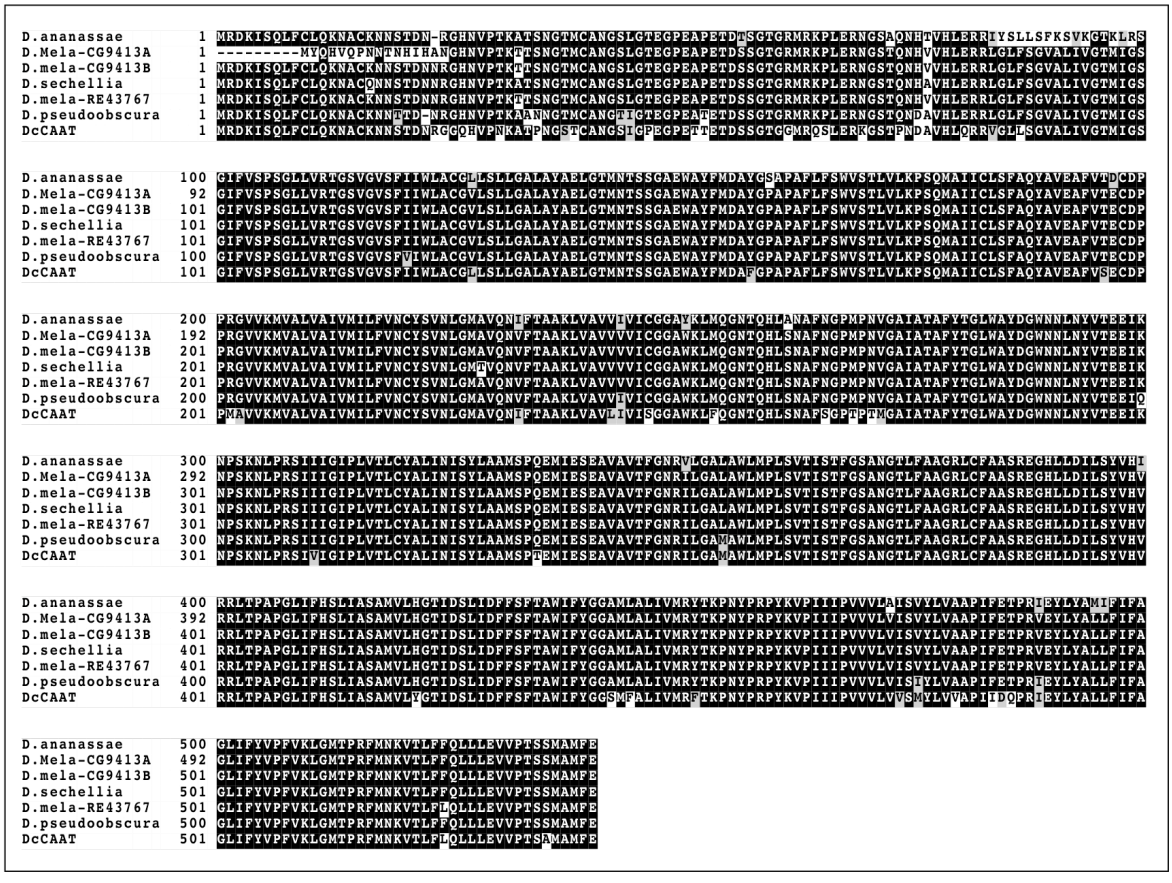


Figure 6.13 ClustalW-Boxshade alignment of DcCAAT with Dipteran cationic amino acid transporters. Residues with black background are highly conserved and residues with grey background are partially conserved

6.5.2 RNAi for DcCAAT in D. coarctata:

Partial coding sequence of DcCAAT (831bp) was used for preparation of RNAi construct in pJET1.2 cloning vector as shown in figure 6.14A. For sense strand vector containing DcCAAT partial sequence was linearised with XbaI and for antisense strand with XhoI (Figure 6.14B). In either case T7 promoter was present for *in vitro* transcription. The details of nucleotide sequence used RNAi for is shown in figure 6.14 C. Preparation of double stranded RNA by *in vitro* transcription, injection or feeding dsRNA to *D. coarctata* are in progress. It would be interesting to see the knock out effects of DcCAAT in *D. coarctata*.

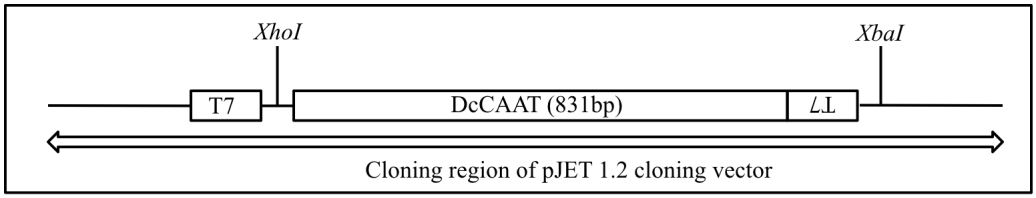


Figure 6.14 (A) Overview of RNAi construct for DcCAAT in pJET1.2 vector

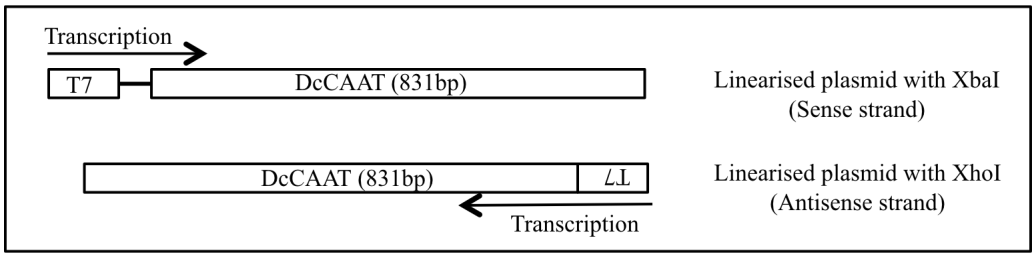


Figure 6.14 (B) Overview of plasmid linearisation and *in vitro* transcription

```
..........GGCGTAATACGACTCACTATAGGGAGAGCGGCCAGATCTTCCGGATGGCTCGAGTTTTTCAGC
AAGATATGAATTCATGATTTTATTTGTTAACTGTTATAGTGTTAATTTGGGCATGGCAGTTCAAAATATATTTA
CAGCAGCCAAACTGGTAGCTGTTCTAATTGTTATAAGTGGTGGTGCCTGGAAATTGTTCCAAGGTAATACACAA
CATCTATCAAATGCATTTAGTGGACCCACACCCACGATGGGAGCTATAGCCACGGCCTTCTATACCGGTCTGTG
GGCTTATGATGGCTGGAATAACTTGAACTATGTCACGGAGGAGATTAAAAATCCAAGCAAGAATTTACCACGCT
CCATTGTAATTGGCATACCCTTGGTTACTTTGTGTTATGCTTTGATTAATATTTCCTATTTAGCCGCCATGTCA
CCCACCGAAATGATTGAATCTGAGGCGGTAGCCGTAACCTTTGGTAATCGCATTTTGGGAGCCATGGCCTGGTT
AATGCCATTGAGTGTGACGATCAGTACATTTGGCAGTGCGAATGGTACATTGTTTGCTGCAGGAAGATTATGTT
TTGCTGCTTCACGTGAGGGCCATTTATTGGATATTTTATCTTATGTTCATGTACGCCGTCTTACACCAGCTCCT
GGTCTTATTTTCCATTCTCTTATTGCCTCTGCTATGGTTTTATATGGCACCATAGATTCTCTAATTGACTTTTT
CAGTTTTACTGCTTGGATTTTTTATGGCGGCTCTATGTTTGCCCTTATTGTGATGCGTTTCACCAAACCCAACT
ATCCAAGGCCATACAAAGTGCCTATTATCATTCCCGTTGTGGTTCTGGTGGTTTCAATGTATCTGGTGGTGGCT
CCCATTATTGACCAACCTCGCATTGAATATCTATAGTGAGTCGTATTATCTTTCTAGAAGATC..........
.
```

Figure 6.14 (C) Partial coding sequence of DcCAAT used for RNAi

Chapter 7

General discussion and conclusions:

Wheat is an important cereal crop worldwide. Two insect pests, Cereal aphid (*Sitobion avenae*) and Wheat bulb fly (*Delia coarctata*) are a major threat to cultivation of wheat in UK, that significantly reduce the yield under heavy infestation as discussed earlier in **Chapter 1**. Various synthetic insecticides used currently have been suggested for the control of these insect pests. However excessive use of insecticides is not a suitable approach from the environmental point of view. It has become immensely important to think about environmentally safe methods to control insect pests. Improving the genetic quality of wheat for resistant factors against these insect pests is a prospective solution. Traditional breeding of wheat has given rise to many high yielding varieties, resistant to many biotic and abiotic stresses, which are much superior than their ancestral wild varieties. However more efforts are required in the direction of development of insect resistant wheat varieties. Genetic improvement of wheat using modern biotechnological tools alone, or in combination with traditional breeding is one of the finest approaches. The genetic modifications are very specific and precise as compared to breeding methods. On the other hand, genetic modifications have to be environmentally safe not to affect the non-target organisms in the field by interfering with other trophic interactions. For a successful genetic improvement in crop plants, it is very important to understand the biochemistry of insect pest and host plant, and their interactions at molecular level. Interaction studies at protein level between crop plant and insect pests have revealed several endogenous defence compounds and their crucial role in plants' protection against pests. Studies on insect biochemistry have also revealed the potential targets such as insect guts and proteases, which can be further exploited to develop insect resistant crops.

The objectives of this work were to explore molecular approaches to increase the resistance of wheat towards *S. avenae* and *D. coarctata*. Both insects were studied to understand their digestive biochemistry and suitability of their guts as potential target

for development of wheat resistant to these pests. Wheat was explored for presence of endogenous resistance factors or defence proteins such as proteinase inhibitor and lectins, which are produced either systemically or in response to insect herbivory. These defence proteins could be utilised to pertain resistance to wheat against insect pests.

Biochemical assays revealed that *D. coarctata* possesses serine as well as cysteine proteases in gut which play an important role in digestion of a protein diet. The pH of gut was found to be neutral which could be suitable for the activity of both types of proteases. For characterising gut the proteases, a complete coding sequence of serine proteinase from *D. coarctata* gut tissue (DcSP) was cloned by degenerate PCR and rapid amplification of cDNA ends. Deduced amino acid sequence of DcSP showed all necessary active site conserved residues and similarity to dipteran serine proteases. DcSP was recombinantly expressed in *E. Coli* as a fusion protein with thioredoxin. Recombinant DcSP was obtained in insoluble aggregate fraction and was purified by metal affinity chromatography. However, efforts to refold recombinant DcSP by dilution method *in vitro* were not successful. Released mature DcSP from thioredoxin did not show any proteolytic activity.

A cysteine proteinase from *D. coarctata* (DcCathL) was cloned from the gut tissue by using the same method as described for DcSP. Deduced amino acid sequence of DcCathL showed necessary active site conserved residues and sequence similarities to Dipteran cathepsin L-like cysteine proteases. DcCathL was recombinantly expressed in *P. pastoris* and was produced on large scale using a bench top fermentor. DcCathL was purified using ion exchange chromatography and showed hydrolytic activity towards synthetic substrates. A complete biochemical characterisation of DcCathL was done. DcCathL could be instantly activated by addition of reducing agents and the activity was seen to reduce over time in absence of substrate due to self-degradation. Optimum pH for activity of DcCathL was 4.5 and no activity was seen at pH 7.5, thus confirming it to be a lysosomal cysteine proteinase. DcCathL activity did not show any dependence on ionic strength, however the activity could be inhibited using E-64, an irreversible inhibitor of cysteine proteinases.

Identification of types of proteases present in *D. coarctata* gut would help in use of endogenous resistance factors, proteinase inhibitors (PIs) from wheat to provide resistance against this pest. Two PIs from wheat, subtilisin/ chymotrypsin inhibitor (WSCI) and cysteine proteinase inhibitor (WCPI) were cloned by RT-PCR from wheat leaf. Both WSCI and WCPI were recombinantly expressed in *P. pastoris* and expression was scaled up using bench top fermentation. Purification of recombinant inhibitors was carried out using ion exchange chromatography. Both inhibitors showed inhibition of their target proteases available commercially. WCPI inhibited cysteine protease activity of papain and recombinant DcCathL from *D. coarctata*. WSCI inhibited bovine chymotrypsin and bacterial subtilisin. WSCI also showed inhibition of cysteine proteinases including recombinant DcCathL. Both recombinant PIs showed inhibition of gut protease activity of *D. coarctata* in *in vitro* assays. In feeding experiments, it was found that WSCI was lethal to *D. coarctata* by causing mortality. As the wheat plant fails to show resistance against *D. coarctata*, it is possible that it does not have enough levels of these PIs systemically to be resistant to these pests. Constitutive and / or over-expression of such PIs in wheat could provide resistance against *D. coarctata.*

There is long time debate on the gut protein digestion in phloem feeding insects. There are two different points of views about aphid gut digestion, some believe that they could use gut proteases to digest ingested proteins while others believe that they only have limited capacity to do it (Rahbe *et al.,* 1995, Habibi *et al.,* 2002, Foissac *et al.,* 2002 and Salvucci *et al.,* 1998). For *S. avenae*, it was necessary to establish that they could use their gut proteases to digest ingested proteins. It is believed that Hemipteran insects rely mainly on cysteine type proteases for their gut proteolytic digestion (Terra *et al.,* 1996). I cloned three cathepsin B-like cysteine proteases from *S. avenae* gut tissue. All three cysteine proteases (CathB 16, CathB 2744 and CathB 84) were identical to cathepsin B genes from Pea aphid (*Acyrthosiphon pisum*). My feeding experiments with digestible protein suggested that *S. avenae* could use their gut proteases and metabolise the ingested protein. In this simple experiment, *S. avenae* was fed on a diet without free amino acids and a diet that was added with ovalbumin as a digestible protein. Overall the aphids fed on ovalbumin showed better survival and growth than aphids that were fed on diet without free amino acids. When the honeydew was analysed from the above aphids, ovalbumin was detectable in the

diet while it did not show up in the honeydew collected from the fed aphids. Also when the recombinant wheat PIs (WSCI and WCPI) were included in the diet showed protection of ovalbumin from the gut proteases that could clearly depict the activity of proteases in the *S. avenae* gut. This was an important finding as it opens up a new arena of research on use of host plant proteinase inhibitors against phloem feeding insects. Due to the specialist feeding habits of aphids and considering that they need not do gut proteolytic digestion, proteinase inhibitors were not in consideration against aphids before. Presence of proteinase inhibitors in plant phloem sap suggests that they might act against these phloem-feeding insects, however sometimes they are not effective to deter insect from feeding.

Insect gut proteases, as pivotal targets in crop protection strategy have been suggested by many researchers and have been used in development of insect resistant crop plants. Results on antimetabolic effects of plant proteinase inhibitors inhibiting insect gut proteases thereby reducing their survival and growth suggest the importance of insect gut proteases. Incorporation of cystatin like proteases in transgenic plants has shown resistance towards nematodes without any toxic risks to humans (Atkinson *et al.,* 2003). This is an interesting finding that supports feasibility of using proteinase inhibitors for developing insect resistant food crops.

Use of insecticidal proteins in crop protection is another promising approach, though it has important concerns of effect on non-target insects in the field. ScathL, a cysteine proteinase from flesh fly (*Sarcophaga peregrina*) has been reported to be a potential insecticidal protein if delivered to insect hemolymph where it is most active (Philip *et al*., 2007). Recombinant DcCathL showed similar insecticidal effect on lepidopteran insects, *Lacanobia oleracea* and *Mamestra brassicae* when injected into hemolymph by causing extensive melanisation leading to death. Efforts were undertaken to explore the relation between systemic melanisation and target of DcCathL in hemolymph. Literature review revealed important role of hemolymph serpins in the negative regulation of melanisation cascade. Therefore, two serine proteinase inhibitors (Serpins; MbSpn-1A and MbSpn-1B/C) from *Mamestra brassicae* were cloned and recombinantly expressed using the *P. pastoris* expression system. Both recombinant serpins were purified as active PIs and showed inhibition of their target proteases. Co-injecting serpins with DcCathL showed controlled

melanisation rather than extensive melanisation. This helped me to prove the hypothesis that serpins in lepidopteran insects are the targets for the cysteine proteases like cathepsin. Proteases like DcCathL could be likely candidates for crop protection strategy if they are delivered to insect hemolymph. Insecticidal activity of DcCathL, is based on interference with the normal negative regulatory mechanism which prevents systemic activation of the melanisation cascade. However use of DcCathL like proteases in crop protection suffer from the drawback that they must be introduced into the haemolymph to be effective. Similar cysteine proteinase ScathL did not show insecticidal effects when orally ingested by lepidopteran larvae, which was considered to be due to degradation or inactivation in the insect gut (Li *et al.,* 2008). However, a fusion protein approach, in which the enzyme is linked to a carrier protein, able to cross from the gut into hemocoel (Fitches *et al.,* 2004b), might be able to deliver these insecticidal proteins to the insect hemolymph.

Wheat produces insect specific defence proteins in response to herbivory by insects. Hessian fly responsive genes (Hfrs) are up-regulated in response to herbivory by Hessian fly (*Mayetiola destructor*). Three Hfr genes from wheat leaf (Hfr-1, Hf-2 and Hfr-3) were successfully cloned by PCR based methods. As Hfrs are over expressed in response to herbivory by Hessian fly, the recombinant proteins could be equally potential against other Dipteran insects like *D. coarctata.* The efforts to produce Hfr-1 recombinantly in *P. pastoris* were not successful, however Hfr-3 was recombinantly expressed using the same expression system. Hfr-3, a wheat germ agglutinin (WGA) lectin like protein showed antimetabolic effects towards *S. avenae* by causing death and also showed retardation in growth. Lectin like proteins from wheat and especially wheat germ agglutinin (WGA) has proven to be insecticidal to many insect pests. Similarity of Hfr-3 to WGA and our results on feeding Hfr-3 to *S. avenae* makes it a suitable candidate for the development of wheat resistant to similar pests. Higher levels of expression and / or systemic expression of genes like Hfr-3 could be another possible approach to pertain wheat resistance against Dipteran insects like *D. coarctata.*

RNA interference (RNAi) mediated crop protection is the current hotspot among strategies for crop protection against insect pests. Knocking down a crucial gene from insects by RNAi has shown to be a specific method to control the insect pests. Double

stranded RNA (dsRNA) that is used in RNAi technology is sturdy molecule and can be either injected or even fed to insects to exhibit knock down effects. As the RNA is a specific molecule for a particular insect species, there would be no harm to other insects in the field, unless there is a strong sequence homology between the genes being used. This could take care of another major concern of non-target insects in the field being killed. To carry out the RNAi in *D. coarctata,* a cationic amino acid transporter (DcCAAT) from the gut tissue was cloned. DcCAAT is an important gene involved in transport of essential amino acids across the gut membrane. Silencing this gene would be expected to drastically affect the insect metabolism and amino acid transport across the gut. However application of dsRNA to a field level has two major concerns, as it requires a dsRNA to be produced on large scale and cost effectiveness of production. Recent studies in efforts to produce dsRNA on a large scale using ϕ6 bacteriophage RNA-dependent RNA polymerase, have shown that dsRNA could be produced at the rate of 1.6 mg /g of wet cells (~10% of the total RNA synthesized in the cell) and the process can be optimised to produce kilogram quantities of dsRNA using industrial scale bioreactors (Aalto *et al.,* 2007).

Recently a demonstration of host delivered RNA interference (HD-RNAi) has given new insights into engineering plants to produce small RNA molecules, which will be ingested by pests (pathogens, nematodes and insects) and degrade the targeted mRNAs and thereby silencing essential genes in the pest. Transgenic maize plants expressing dsRNA for Western corn rootworm (*Diabrotica virgifera*) V-type ATPaseA silenced this vacuolar gene in midgut cells of WCR and caused reduced feeding from WCR (Baum *et al.,* 2007). Thus one of the major advantages of RNAi technology is that, the crop plants can be modified for resistance against insects without introducing new proteins into food and feed products, which always is a major concern in GM crops.

The results presented in this thesis could be useful while developing wheat resistant to *S. avenae, D. coarctata* and similar insect pests. Wheat breeding can be carried out to enhance the effects of endogenous resistance factors discussed here. The expression levels of these endogenous resistance factors can also be increased using biotechnological tools such as transformation. Wheat has been genetically modified

for the resistance against *Fusarium* head blight (FHB) using genetic modification methods. The transgenic wheat showed resistance to *Fusarium* under greenhouse and field conditions (Mackintosh *et al.,* 2007). Incorporating constitutive promoters, which will allow continual transcription of its associated defence related endogenous genes is one of the possibilities. Insecticidal proteins such as DcCathL fused with 'carrier' proteins could be transformed in crop plants for *in planta* expression. However, these methods require an appropriate gene expression level which needs to be tested under laboratory / green house conditions initially before testing their efficacy under field conditions. Production of such fusion proteins using commercial or large-scale fermentation methods would give us the bio-pesticides, which could benefit field applications.

References:

Aalto, A.P., Sarin, L.P., Van Dijk, A.A., Saarma, M., Poranen, M.M., Arumae, U. and Bamford, D.H., (2007). Large-scale production of dsRNA and siRNA pools for RNA interference utilizing bacteriophage f6 RNA-dependent RNA polymers. RNA 13: 422–429.

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.

Abe, M., Abe, K., Kuroda, M. and Arai, S., (1992). Corn kernel cysteine proteinase inhibitor as a novel cystatin superfamily member of plant origin. Molecular cloning and statement studies. Eur. J. Biochem. 209: 933-937.

Allsopp, P.G. and McGhie, T.K., (1996). Snowdrop and wheatgerm lectins and avidin as antimetabolities for the control of sugarcane whitegrubs. Entomol. Exp. Appl. 80: 409-414.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J., (1990). Basic local alignment search tool. J. Mol. Biol. 215: 403-410.

Araujo, R.N., Santos, A., Pinto, F.S., Gontijo, N.F., Lehane, M.J. and Pereira, M.H., (2006). RNA interference of the salivary gland nitrophorin 2 in the triatomine bug *Rhodnius prolixus* (Hemiptera: Reduviidae) by dsRNA ingestion or injection. Insect Biochem. Mol. Biol. 36: 683–693.

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

Ashida, M. and Yoshida, H., (1988). Limited proteolysis of prophenoloxidase during activation by microbial products in insect plasma and effect of phenoloxidase on electrophoretic mobilities of plasma proteins. Insect Biochem. 18: 11–19.

Auer, C. and Frederick, R., (2009). Crop improvement using small RNAs: application and predictive ecological risk assessments. Trends Biotechnol. 27: 644-651.

Barnes, A., Bale, J., Constantinidou, C., Ashton, P., Jones, A. and Pritchard, J., (2004). Determining protein identity from sieve element sap in *Ricinus communis* L. by quadrupole time of flight (Q-TOF) mass spectrometry. J. Exp. Botany 55: 1473–1481.

Barrett, A.J., Kembhavi, A.A., Brown, M.A., Kirshcke, H., Knight, C.G., Tamai, M. and Hanada, K., (1982). L-trans-Epoxysuccinyl-leucylamido(4-guanido)butane (E-64) and its analogues as inhibitors of cysteine proteinases including cathepsins B, H and L. Biochem. J. 201: 189-198.

Barrett, A.J., Rawlings, N.D. and Woessner, J.F., (2003). The Handbook of Proteolytic Enzymes, 2nd ed. Academic Press.

Barrett, A.J., Rawlings, N.D. and Brien, E.A., (2001). The MEROPS database as a protease information system. J. Struct. Biol. 14: 95-102.

Baum, J.A., Bogaert, T., Clinton, W., Heck, G.R., Feldmann, P., Ilagan, O., Johnson, S., Plaetinck, G., Munyikwa, T., Pleu, M., Vaughn, T., and Roberts, J., (2007). Control

of coleopteran insect pests through RNA interference. Nat. Biotechnol. 25: 1322–1326.

Baumgartner, P., Harper, K., Raemaekers, R.J., Durieux, A., Gatehouse, A.M.R., Davies, H.V. and Taylor, M.A., (2003). Large-scale production and purification of recombinant *Galanthus nivalis* agglutinin (GNA) expressed in the methylotrophic yeast *Pichia pastoris*. Biotechnol. Lett. 25: 1281–1285.

Baumgartner, P., Raemaekers, R.J., Durieux, A., Gatehouse, A., Davies, H. and Taylor, M., (2002). Large-scale production, purification, and characterisation of recombinant *Phaseolus vulgaris* phytohemagglutinin E-form expressed in the methylotrophic yeast *Pichia pastoris*. Protein Exp. Purif. 26: 394–405.

Bennett, R.N., and Wallsgrove, R.M., (1994) Secondary metabolites in plant defense mechanisms. New Phytologist, 127, 617-633.

Blanchard, D.J., Cicek, M., Chen, J. and Esen, A., (2001). Identification of glucosidase aggregating factor (BGAF) and mapping of BGAF binding regions on maize glucosidase. J. Biol. Chem. 276:11895–11901.

Blandin, S., Moita, L.F., Kocher, T., Wilm, M., Kafatos, F.C. and Levashina, E.A., (2002). Reverse genetics in the mosquito *Anopheles gambiae*: targeted disruption of the Defensin gene. EMBO Rep 3:852-856.

Blum, H., Beier, H. and Gross, H.J., (1987) Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels. Electrophoresis 8: 93-99.

Bode, W., Engh, R., Musil, D., Thiele, U., Huber, R., Karshikov, A., Brzin, J., Kos, J. and Turk, V., (1988). The 2.0 Å X-ray crystal structure of chicken egg white cystatin and its possible mode of interaction with cysteine proteinases. EMBO J. 7: 2593–2599.

Botella, M.A., Xu, Y., Prabha, T.N., Zhao, Y., Narasimhan, M.L., Wilson, K.A., Nielsen, S.S., Bressan, R.A. and Hasegawa, P.M., (1996). Differential expression of soybean proteinase inhibitor genes during development and in response to wounding and methyl jasmonate. Plant Physiol. 112: 1201–1210.

Bourne, Y., Zamboni, V., Barre, A., Peumans, W.J., Van Damme, E.J.M., and Rouge, P., (1999). *Helianthus tuberosus* lectin reveals a widespread scaffold for mannose-binding lectins. Structure 7:1473–1482.

Bown, D.P., Wilkinson, H.S. and Gatehouse, J.A., (1997). Differentially regulated inhibitor-sensitive and insensitive protease genes from the phytophagous insect pest, *Helicoverpa armigera*, are members of complex multigene families. Insect Biochem. Mol. Biol. 27: 625–638.

Bown, D.P., Wilkinson, H.S., Jongsma, M.A. and Gatehouse, J.A., (2004). Characterisation of cysteine proteinases responsible for digestive proteolysis in guts of larval western corn rootworm (*Diabrotica virgifera*) by expression in the yeast Pichia pastoris. Insect Biochem. Mol. Biol. 34: 305–320.

Boyd, W.C. and Shapleigh, E., (1954). Antigenic relations of blood group antigens as suggested by tests with lectins. J. Immunol. 73: 226–231.

Brandt, K.S., Silver, G.M., Becher, A.M., Gaines, P.J., Maddux, J.D., Jarvis, E.E. and Wisnewski, N., (2004). Isolation, characterization, and recombinant expression of multiple serpins from the cat flea, *Ctenocephalides felis*. Arch. Insect. Biochem. Physiol. 55: 200–214.

Broadway, R.M., (1995). Are insects resistant to plant proteinase inhibitors? Journal of Insect Physiology 41: 07-116.

Broadway, R.M., (1997). Dietary regulation of serine proteinases that are resistant to serine proteinase inhibitors. Insect Biochemistry and Molecular Biology 43: 855-874.

Brotman, Y., Silberstein, L., Kovalski, I., Perin, C., Dogimont, C., Pitrat, M., Klingler, J., Thompson, A. and Perl-Treves, R., (2002). Resistance gene homologues in melon are linked to genetic loci conferring disease and pest resistance. Theor. Appl. Genet. 104: 1055–1063.

Buchner, P., (1965). Endosymbiosis of Animals with Plant Microorganisms. Interscience, New York.

Buttle, D.J. and Barrett, A.J., (1984). Chymopapain. Chromatographic purification and immunological characterization. Biochem J. 223: 81–88.

Castagna, M., Shayakul, C., Trotti, D., Sacchi, V.F., Harvey, W.R. and Hediger, M.A., (1998). Cloning and characterization of a potassium-coupled amino acid transporter. Proc. Natl. Acad. Sci. USA 95: 5395-5400.

Castagna, M., Shayakul, C., Trotti, D., Sacchi, V.F., Harvey, W.R. and Hediger, M.A., (1997). Molecular characteristics of mammalian and insect amino acid transporters: Implications for amino acid homeostasis. J. Exp. Biol. 200: 269–286.

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

Charity, J.A., Anderson, M.A., Bittisnich, D.J., Whitecross, M. and Higgins, T.J.V., (1999). Transgenic tobacco and peas expressing a proteinase inhibitor from *Nicotiana alata* have increased insect resistance. Molecular Breeding 5: 357-365.

Chen, M., (2008). Inducible direct plant defense against insect herbivores: A review. Insect Science 15: 101-114.

Cho, H., Choi, K., Yamashita, M., Morikawa, H. and Murooka, Y., (1995). Introduction and expression of the *Streptomyces* cholesterol oxidase gene (choA), a potent insecticidal protein active against boll-weevil larvae, into tobacco cells. Appl. Microbiology and Biotech. 44:133-138.

Chougule, N.P., Doyle, E., Fitches, E. and Gatehouse, J.A., (2008). Biochemical characterisation of midgut digestive proteases from *Mamestra brassicae* (cabbage moth; Lepidoptera: Noctuidae) and effect of soybean Kunitz inhibitor (SKTI) in feeding assays. J. Insect Physiol. 54: 563-572.

Chougule, N.P., Giri, A.P., Sainani, M.N. and Gupta, V.S., (2005). Gene expression patterns of *Helicoverpa armigera* gut proteases. Insect Biochem. Mol. Biol. 35: 355-367.

Christeller, J.T., (2005). Evolutionary mechanisms acting on proteinase inhibitor variability. FEBS Journal, 272: 5710–5722.

Christeller, J.T., Farley, P.C., Ramsay, R.J., Sullivan, P.A. and Laing, W.A., (1998). Purification, characterization and cloning of an aspartic proteinase inhibitor from squash phloem exudate. Eur. J. Biochem. 254: 160–167.

Christensen, B.M., Li, J., Chen, C.C. and Nappi, A.J., (2005). Melanization immune responses in mosquito vectors. Trends Parasitol 21:192–199.

Christophides, G.K., Zdobnov, E., Barillas-Mury, C., Birney, E., Blandin, S., Blass, C., Brey, P.T., Collins, F.H., Danielli, A., Dimopoulos, G., Hetru, C., Hoa, N.T., Hoffmann, J.A., Kanzok, S.M., Letunic, I., Levashina, E.A., Loukeris, T.G., Lycett, G., Meister, S., Michel, K., Moita, L.F., Müller, H., Osta, M.A., Paskewitz, S.M., Reichhart, J., Rzhetsky, A., Troxler, L., Vernick, K.D., Vlachou, D., Volz, J., von Mering, C., Xu, J., Zheng, L., Bork, P., and Kafatos, F.C., (2002). Immunity-related genes and gene families in *Anopheles gambiae*. Science 298:159 – 165.

Chrunyk, B., Evans, J., Lillquist, J., Young, P. and Wetzel, R., (1993). Inclusion body formation and protein stability in sequence variants of Interleukin-1β. J. Biol. Chem. 268: 18053-18061.

Cino, J., (1999). High yield protein production from *Pichia pastoris*: a protocol for benchtop fermentation. New Brunswick Co. Inc. (http://www.nbsc.com/papers/Default.asp.).

Clark, J.M. and Yamaguchi, I., In: Clark J. and Yamaguchi I, editors, (2002). Scope and status of pesticide resistance. Agrochemical resistance extent, mechanism, and detection. American Chemical Society, Washington, DC.

Closs, E.I., Albritton, L.M., Kim, J.W. and Cunningham, J.M., (1993a). Identification of a low affinity, high capacity transporter of cationic amino acids in mouse liver. J. Biol. Chem. 268: 7538–7544.

Closs, E.I., Lyons, C., Kelly, C. and Cunningham, J.M., (1993b). Characterization of the third member of the MCAT family of cationic amino acid transporters. J. Biol. Chem. 268: 20796–20800.

Corre-Menguy, F., Cejude, F.J., Mazubert, C., Vidal, J., Lelandais-Briere, C., Torres, G., Rode, A. and Hartmann, C., (2002). Characterization of the expression of wheat cystatin gene during caryopsis development. Plant Molecular Biology 50: 687-698.

Cristofoletti, P.T., Ribeiro, A.F., Deraison, C., Rahbe, Y. and Terra, W.R., (2003). Midgut adaptation and digestive enzyme distribution in a phloem feeding insect, the pea aphid *Acyrthosiphon pisum*. J Insect Physiol. 49:11–24.

Cronshaw, J. and Sabnis, D.D., (1990). Phloem proteins. In Behnke H-D, Sjolund RD (eds): Sieve Elements. Comparative Structure, Induction and Development. Berlin: Springer-Verlag, pp 257–283.

Dadd, R.H., (1985). Nutrition: Organisms. In: G. A. Kerkut & L. I. Gilbert (eds), Comprehensive Insect Physiology Biochemistry and Pharmacology. Vol. 4, Regulation: Digestion, Nutrition, Excretion. Pergamon Press, Oxford, pp. 313–390.

Dale, C. and Moran, N.A., (2006). Molecular interactions between bacterial symbionts and their hosts. Cell 126: 453 – 465.

De Gregorio, E., Han, S.J., Lee, W.J., Baek, M.J., Osaki, T., Kawabata, S.I., Lee, B.L., Iwanaga, S., Lemaitre, B. and Brey, P.T., (2002). An Immune-Responsive Serpin Regulates the Melanization Cascade in Drosophila. Developmental Cell, Vol. 3, 581–592.

De Leo, F., Bonadé-Bottino, M.A., Ceci, L.R., 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.

De Leo, F., Ceci, L.R., Jouanin, L., Gallerani, R., (2001). Analysis of mustard trypsin inhibitor-2 gene expression in response to developmental or environmental induction. Planta 212: 710-717.

De Luca, V. and St Pierre, B., (2000). The cell and developmental biology of alkaloid biosynthesis. Trends Plant Sci. 5, 168.

Dehrmann, F.M., Coetzer, T.H., Pike, R.N., Dennison, C., (1995). Mature cathepsin L is substantially active in the ionic milieu of the extracellular medium. Arch. Biochem. Biophys. 324: 93-8.

Delibes, A., Delmoral, J., Martinsanchez, J.A, Mejias, A., Gallego, M., Casado, D., Sin, E., and Lopezbrana, I., (1997). Hessian fly-resistance gene transferred from chromosome 4M (V) of *Aegilops ventricosa* to *Triticum aestivum*. Theor. Appl. Genet. 94:858–864.

Deraison, C., Darboux, I., Duportets, L., Gorojankina, T., Rahbe, Y. and Jouanin, L., (2004). Cloning and characterization of a gut-specific cathepsin L from the aphid *Aphis gossypii*. Insect Mol. Biol 13: 165-177.

Di Gennaro, S.D., Ficca, A.G., Panichi, D. and Poerio, E., (2005). cDNA cloning and heterologous expression of a wheat proteinase inhibitor of subtilisin and chymotrypsin (WSCI) that interferes with digestive enzymes of insect pests. Biol. Chem. 386: 383-389.

Ding, X., Gopalakrishnan, B., Johnson, L.B., White, F.F., Wang, X., Morgan, T.D., Kramer, K.J. and Muthukrishnan, S., (1998). Insect resistance of transgenic tobacco expressing an insect chitinase gene. Transgenic Research 7: 77-84.

Douglas, A.E., (2006a). Phloem-sap feeding by animals: problems and solutions. J. Exp. Bot. 57: 747–754.

Douglas, A.E., (2005). Bugs within bugs: symbiotic bacteria in garden insects. Microbiology Today 32: 76-79.

Douglas, A.E., (2009). The microbial dimension in insect nutritional ecology. Functional Ecology 23: 38-47.

Douglas, A.E., (1988). Sulphate utilization in an aphid symbiosis. Insect Biochem. 18: 599–605.

Douglas, A.E., (1993). The nutritional quality of phloem sap utilized by natural aphid populations. Ecol. Entomology 18: 31–38.

Douglas, A.E., (2003). Nutritional physiology of aphids. Adv. Insect Physiol 31: 73–140.

Douglas, A.E., Price, D.R.G., Minto, L.B., Jones, E., Pescod, K.V., Francois, C.L.M.J., Pritchard, J. and Boonham, N., (2006b). Sweet problems: insect traits defining the limits to dietary sugar utilisation by the pea aphid, *Acyrthosiphon pisum.* J. Exp. Biol. 209: 1395–1403.

Down, R.E., Gatehouse, A.M.R., Hamilton, W.D.O. and Gatehouse, J.A., (1996). Snowdrop lectin inhibits development and decreases fecundity of the glasshouse potato aphid (*Aulacorthum solani*) when administered in vitro and via transgenic plants both in laboratory and glasshouse trials. J. Insect Physiol. 42: 1035-1045.

Eamens, A., Wang, M., Smith, N.A. and Waterhouse, P.M., (2008). RNA silencing in plants: yesterday, today, and tomorrow. Plant Physiol. 147: 456–468.

Ehlrich, P.R. and Raven, P.H., (1964). Butterflies and plants: a study in co-evolution. Evolution 18: 586-608.

Eisemann, C.H., Donaldson, R.A., Pearson, R.D., Cadogan, L.C., Vuocolo, T. and Tellam, R.L., (1994). Larvicidal activity of lectins on *Lucilia cuprina*: Mechanism of action. Entomol. Exp. Appl. 72: 1-11.

Ellis, K.J. and Morrison, J.F., (1982). Buffers of constant ionic strength for studying pH-dependent processes. Methods Enzymol. 87: 405–426.

Elton Robinson, (2008). World wheat production-expected higher for 2008-2009. Farm Press Editorial staff. July 25, 2008 (http://westernfarmpress.com/news/wheat-production-0625/).

Espinoza-Fuentes, F.P. and Terra, W.R., (1987). Physiological adaptations for digesting bacteria. Water fluxes and distribution of digestive enzymes in *Musca domestica* larval midgut. Insect Biochem. 17: 809-817.

Facchiano, A.M., Costantini, S, Maro, A.D., Panichi, D., Chambery, A., Parente, A., Gennaro, S.D. and Poerio, E., (2006). Modeling the 3D structure of wheat subtilisin/chymotrypsin inhibitor (WSCI). Probing the reactive site with two susceptible proteinases by time-course analysis and molecular dynamics simulations, Biol. Chem. 387: 931–940.

Febvay, G., Rahbe, Y., Rynkiewicz, M., Guillaud, J. and Bonnot, G., (1999). Fate of dietary sucrose and neosynthesis of amino acids in the pea aphid, *Acyrthosiphon pisum*, reared on different diets. J. Exp. Biol. 202: 2639–2652.

Ferry, N., Edwards, M.G., Gatehouse, J.A. and Gatehouse, A.M.R., (2004). Plant-insect interactions: Molecular approaches to insect resistance. Current Opinion in Biotechnol. 15: 155-161.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E. and Mello, C.C., (1998). Potent and specific genetic interference by double- stranded RNA in *Caenorhabditis elegans*. Nature 39: 806–811.

Fisher, D.B., Wu, Y. and Ku, M.S.B., (1992). Turnover of soluble proteins in the wheat sieve tube. Plant Physiol. 100: 1433–1441.

Fitches, E., Wiles, D., Douglas, A.E., Hinchliffe, G., Audsley, N., Gatehouse, J.A., (2008). The insecticidal activity of recombinant garlic lectins towards aphids. Insect Biochem. Mol. Biol. 38: 905–915.

Fitches, E., (1999). The mechanisms of action of insecticidal lectins from snowdrop (GNA) and jackbean (Concanavalin A) on tomato moth larvae. Ph.D. thesis, University of Durham, UK.

Fitches, E. and Gatehouse, J.A., (1998). A comparison of the short and long term effects of insecticidal lectins on the activities of soluble and brush border enzymes of tomato moth larvae (*Lacanobia oleracea*). J. Insect Physiol. 44: 1213-1224.

Fitches, E., Audsley, N., Gatehouse, J.A., Edwards, J.P., (2002). Fusion proteins containing neuropeptides as novel insect control agents: snowdrop lectin delivers fused allatostatin to insect haemolymph following oral ingestion. Insect Biochem. Mol. Biol. 32: 1653–1661.

Fitches, E., Edwards, M.G., Mee, C., Grishin, E., Gatehouse, A.M.R., Edwards, J., Gatehouse, J.A., (2004b). Fusion proteins containing insect-specific toxins as pest control agents: snowdrop lectin delivers fused insecticidal spider venom toxin to insect haemolymph following oral ingestion. J. Insect Physiol. 50: 61–71.

Fitches, E., Gatehouse, A.M.R., Gatehouse, J.A., (1997). Snowdrop lectin (GNA) delivered via artificial diet and transgenic plants on the development of the tomato moth (*Lacanobia oleracea*) larvae in the laboratory and glasshouse trials. J. Insect Physiol. 43: 727–739.

Fitches, E., Wilkinson, H., Bell, H., Bown, D.P., Gatehouse, J.A., Edwards, J.P., (2004a). Cloning, expression and functional characterisation of chitinase from larvae of tomato moth (*Lacanobia oleracea*): a demonstration of the insecticidal activity of insect chitinase. Insect Biochem. Mol. Biol. 34: 1037–1050.

Fitches, E., Woodhouse, S., Edwards, J.P. and Gatehouse, J.A., (2001). *In vitro* and *in vivo* binding of snowdrop (*Galanthus nivalis* agglutinin; GNA) and jackbean (*Canavalia ensiformis*; Con A) lectins within tomato moth (*Lacanobia oleracea*) larvae; mechanisms of insecticidal action. J. Insect Physiol. 47: 777–787.

Flor, H.H., (1955). Host–parasite interaction in flax rust; its genetics and other implications. Phytopathology, 45: 680–685.

Foissac, X., Edwards, M.G., Du, J.P., Gatehouse, A.M.R. and Gatehouse, J.A., (2002). Putative protein digestion in a sap-sucking homopteran plant pest (rice brown plant hopper; *Nilaparvata lugens*: Delphacidae)-identification of trypsin-like and cathepsin B-like proteases. Insect Biochem. Mol. Biol. 32: 967–978.

Fourney, R.M., Miyakoshi, S., Day, R.S. and Peterson, M.C., (1988). Northern blotting: efficient RNA staining and transfer. Focus 10: 5–6.

Gallun, R.L., (1978). Genetics of biotypes B and C of the Hessian fly. Ann. Entomol. Soc. Am. 71: 481–486.

Garcia-Olmedo, F., Salcedo, G., Sanchez-Monge, R., Gomez, L., Roys, J. and Carbonero, P., (1987). Plant proteinaceous inhibitors of proteases and amylases. Oxford Survey Plant Mol. Cell. Biol. 4: 275-334.

Gatehouse, A.M.R., Powell, K.S., Peumans, W.J., Van Damme, E.J.M. and Gatehouse, J.A., (1994). Insecticidal properties of plant lectins: their potential role in plant protection. In: Pusztai A.J., Barcods S., eds. Lectins biochemical perspectives. Basingstoke: Francis and Taylor, 35-57.

Gatehouse, J.A., (2002). Plant resistance towards insect herbivores: a dynamic interaction. New Phytologist 156: 145-169.

Gatehouse, A.M.R., Down, R.E., Powell, K.S., Sauvion, N., Rahbé, Y., Newell, C.A., Merryweather, A., Hamilton, W.D.O. and Gatehouse, J.A., (1996). Transgenic potato plants with enhanced resistance to the peach-potato aphid *Myzus persicae*. Entomologia Experimentalis et Applicata 79: 295–307.

Gatehouse, A.M.R., Davidson, G.M., Newell, C.A., Merryweather, A., Hamilton, W.D.O., Burgess, E.P.J., Gilbert, R.J.C. and Gatehouse, J.A., (1997). Transgenic potato plants with enhanced resistance to the tomato moth *Lacanobia oleracea*: growth room trials. Mol. Breeding, 3: 49-63.

Gattolin, S., Newbury, H.J., Bale, J.S., Tseng, H.M., Barrett, D.A. and Pritchard, J., (2008). A diurnal component to the variation in sieve tube amino acid content in wheat (*Triticum aestivum* L.). Plant Physiol. 147: 912– 921.

Gaupels, F., Knauer, T. and van Bel, A.J.E., (2008). A combinatory approach for analysis of protein sets in barley sieve-tube samples using EDTA-facilitated exudation and aphid stylectomy, J. Plant. Physiol. 165: 95–103.

Gettins, P.G.W., (2002). Helix F plays an active, critical role in the serpin mechanism. FEBS Lett. 523: 2-6.

Ghanim, M., Kontsedalova, S. and Czobsnek, H, (2007). Tissue-specific gene silencing by RNA interference in the whitefly *Bemisia tabaci* (Gennadius). Insect Biochem. Mol. Biol. 37: 732–738.

Gianoli, E. and Nieneyer, H.M., (1998). DIBOA in wild Poaceae: Sources of resistance to the Russian wheat aphid (*Diurahis noxia*) and the greenbug (*Schizaphis graminum*). Euphytica 102:317–321.

Gill, K.S. and Gill, B.S., (1994). Mapping in the realm of polyploidy: The wheat model. BioEssays 16:841-846.

Gillespie, J.P. and Kanost, M.R., (1997). Biological Mediators Of Insect Immunity. Annu. Rev. Entomol. 42:611–43.

Giovanini, M.P., Saltzmann, K.D., Puthoff, D.P., Gonzalo, M., Ohm, H.W., and Williams, C.E., (2007). A novel wheat gene encoding a putative chitin-binding lectin is associated with resistance against Hessian fly. Mol. Plant Pathol. 8: 69–82.

Givovich, A., Morse, S., Cerda, H., Niemeyer, H.M., Wratten, S.D. and Edwards, P.J., (1992). Hydroxamic acid glucosides in the honeydew of aphids feeding on wheat. J. Chem. Ecol. 18: 841–846.

Greaney, M.G. and Rosteck, P.R., (1994). Expression vectors for the bovine trypsin and trypsinogen and host cells transformed therewith. European Patent (EP597681).

Gunduz, A.E. and Douglas A.E., (2009). Symbiotic bacteria enable insect to use a nutritionally inadequate diet. Proc. R. Soc. B. 276: 987-991.

Habib, H. and Fazili, K.M. (2007). Plant protease inhibitors: a defense strategy in plants. Biotechnol. Mol. Biol. Review 2: 68-85.

Habibi, J., Brandt, S.L., Coudron, T.A., Wagner, R.M., Wright, M.K., Backus, E.A., and Huesing, J.E., (2002). Uptake, flow, and digestion of casein and green fluorescent protein in the digestive system of *Lygus hesperus* Knigh. Archives of Insect Biochem. Physiol. 50: 62–74.

Haebel, S. and Kehr, J., (2001). Matrix-assisted laser desorption/ionization time of flight mass spectrometry peptide mass fingerprints and post source decay: a tool for the identification and analysis of phloem proteins from *Cucurbita maxima* Duch. separated by two dimensional polyacrylamide gel electrophoresis. Planta 213: 586–593.

Haq, S.K., Atif, S.M. and Khan, R.H. (2004). Protein proteinase inhibitor genes in combat against insects, pests, and pathogens: natural and engineered phytoprotection. Arch. Biochem. Biophys. 43: 145-159.

Harborne, J.B., (1988). The flavonoids: recent advances. In Goodwin, T.W., (ed), 'Plant Pigments', pp. 299-344. Academic press, London.

Harper, M.S., Hopkins, T.L. and Czapla, T.H., (1998). Effect of wheat germ agglutinin on formation and structure of the peritrophic membrane in European corn borer (*Ostrinia nubilalis*) larvae. Tissue Cell 30: 166-176.

Harper, S.M., Crenshaw, R.W., Mullins, M.A., Privalle, L.S.J., (1995). Lectin-binding to insect brush-border membranes. Econ. Entomol. 88, 1197-1202.

Harrison, R.L. and Bonning, B.C., (2001). Use of proteases to improve the insecticidal activity of baculoviruses. Biol. Control 20: 199–209.

Harsulkar, A.M., Giri, A.P., Gupta, V.S., Sainani, M.N., Ranjekar, P.K. and Deshpande, V.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.

Hashmi, S., Britton, C., Liu, J., Guiliano, D.B., Oksov, Y., Lustigman, S., (2002). Cathepsin L is essential for embryogenesis and development of *Caenorhabditis elegans*. J. Biol. Chem. 277: 3477–3486.

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

Hayashi, H. and Chino, M., (1986). Collection of pure phloem sap from wheat and its chemical composition. Plant Cell Physiol. 27: 1387-1393.

Hebert, C.G., Valdes, J.J. and Bentley, W.E., (2008). Beyond silencing–engineering applications of RNA interference and antisense technology for altering cellular phenotypes. Curr. Opin. Biotechnol. 19: 500–505.

Hegedus, D. D., O'Grady, M., Chamankhah, M., Baldwin, D., Gleddie, S., Braun, L., Erlandson, M., (2002). Changes in cysteine protease activity and localization during midgut metamorphosis in the crucifer root maggot (*Delia radicum*). Insect. Biochem. Mol. Biol. 32: 1585–1596.

Hegedus, D.D., Erlandson, M., Baldwin, D., Hou, X., Chamankhah, M., (2008). Differential expansion and evolution of the exon family encoding the serpin-1 reactive centre loop has resulted in divergent serpin repertoires among the Lepidoptera. Gene 418: 15–21.

Herbert, R.B., (1989). The Biosynthesis of Secondary Metabolites. 2nd edn., Chapman and Hall, London. 232 pp.

Higgins, D.R. and Cregg, J.M., (1998). Introduction to *Pichia pastoris*. In: D.R. Higgins and J.M. Cregg, Editors, Methods in Molecular Biology: Pichia Protocols, Humana Press, Totowa, NJ, pp. 1–15.

Hilder, V.A., Gatehouse, A.M.R. and Boulter, D., (1993). Transgenic plants conferring insect tolerance: protease inhibitor approach. In: Kung, S.D., Wu, R. (Eds.), Transgenic plants. Academic Press, London (GBR), pp. 317-338

Hilder, V.A., Gatehouse, A.M.R., Sheerman, S.S., Barker, R.S., Boulter, D., (1987). A novel mechanism of insect resistance engineered into tobacco. Nature 330: 160-163.

Hofman, J. and Hofmanova, O., (1969). 1,4-Benzoxazin derivatives in plants. Sephadex fractionation and identification of a new glucoside. Eur. J. Biochem. 8: 109–112.

Hohenblum, H., Vorauer-Uhl, K., Katinger and Mattanovich, D., (2004). Bacterial expression and refolding of human trypsinogen. J. Biotechnol. 109:3–11.

Holmgren, A., (1985) Thioredoxin. Ann. Rev. Biochem. 54: 237-271.

Holt, R.A., Subramanian, G.M., Halpern, A., Sutton, G.G., Charlab, R., Nusskern, D.R., Wincker, P., Clark, A.G., Ribeiro, J.M., Wides, R., *et al.,* (2002). The genome sequence of the malaria mosquito *Anopheles gambiae*. Science 298: 129-149.

Homma, K. and Natori, S., (1996). Identification of substrate proteins for cathepsin L that are selectively hydrolyzed during the differentiation of imaginal discs of *Sarcophaga peregrina*. Eur. J. Biochem. 240: 443-447.

Homma, K., Kurata, S. and Natori, S., (1994). Purification, characterization, and cDNA cloning of procathepsin L from the culture medium of NIH-Sape-4, an embryonic cell line of *Sarcophaga peregrina* (flesh fly), and its involvement in the differentiation of imaginal discs. J. Biol. Chem. 269: 15258–15264.

Hopkins, T.L. and Kramer, K.J., (1992). Insect cuticle sclerotization. Annu. Rev.

Entomol. 32:71–93.

Houseman, J.G. and Downe, A.E.R., (1980). Endoproteinase activity in the posterior midgut of the *Rhodnius prolixus* Stal (Hemiptera: Reduviidae). Insect Biochem. 10: 363–366.

Hughes, C.L. and Kaufman, T.C., (2000). RNAi analysis of deformed, proboscipedia and sex combs reduced in the milkweed bug *Oncopeltus fasciatus*: novel roles for Hox genes in the hemipteran head. Development 127: 3683–3694.

Ivanov, D., Emonet, C., Foata, F., Affolter, M., Delly, M., Fisseha, M., Blum-Sperisen, S., Kochhar, S. and Arigoni, F., (2006). A serpin from the gut bacterium *Bifidobacterium longum* inhibits eukaryotic elastase like serine proteases. J. Biol. Chem. 281: 17246-52.

Janciauskiene, S., (2001). Conformational properties of serine proteinase inhibitors (serpins) confer multiple pathophysiological roles. Biochim. Biophys. Acta 1535: 221–235.

Jaubert-Possamai, S., Trionnaire, G., JoëlBonhomme, Christophides, G.K., Rispe, C. and Tagu, D., (2007). Gene knockdown by RNAi in the pea aphid *Acyrthosiphon pisum*. BMC Biotechnology, 7:63.

Ji, C., Wang, Y., Guo, X, Hartson, S. and Jiang, H., (2004). A pattern recognition serine proteinase triggers the prophenoloxidase activation cascade in the tobacco hornworm, *Manduca sexta*. J. Biol chem. 279: 34101-6.

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

Jiang, H., Wang, Y., Yu, X.Q., Zhu, Y. and Kanost, M. R., (2003). Prophenoloxidase-activating proteinase-3 (PAP-3) from *Manduca sexta* hemolymph: a clip-domain serine proteinase regulated by serpin-1J and serine proteinase homologs. Insect Biochem. Mol. Biol. 33: 1049-60.

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

Jiang, H., Mulnix, A.B. and Kanost, M.R., (1995). Expression and characterization of recombinant *Manduca sexta* serpin-1B and site-directed mutants that change its inhibitory selectivity. Insect Biochem. Mol. Biol. 25: 1093–1100.

Joanitti, G.A., Freitas, S.M., Silva, L.P., (2006). Proteinaceous Protease Inhibitors: Structural features and multiple functional faces. Current Enz. Inhib. 2: 199-217.

Johnson, R., Narvaez, J., An, G. and 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. U.S.A. 86: 9871-9875

Jongsma, M.A., Bakker, P.L., Peters, J., Bosch, D. and Stiekema, W.J., (1995). Adaptation of *Spodoptera exigua* larvae to plant proteinase inhibitors by induction of gut proteinase activity insensitive to inhibition. Proc. Natl. Acad. Sci., 92: 8041-8045.

Jouanin, L., Bonadé-Bottino, M., Girard, C., Morrot, G. and Giband, M., (1998). Transgenic plants for insect resistance. Plant Sci. 131: 1–11.

Kanost, M.R, Jiang, H., and Yu, X.Q. (2004) Innate immune responses of a Lepidopteran insect, *Manduca sexta*. Immun. Reviews. 198: 97-105.

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

Karrer, K.M., Pfeiffer, S.L. and Ditomas, M.E., (1993). Two distinct gene subfamilies within the family of cysteine protease genes. Proc. Natl. Acad. Sci. USA 90: 3063–3067.

Kehr, J., (2006). Phloem sap proteins: their identities and potential roles in the interaction between plants and phloem-feeding insects. J. Exp. Bot. 57: 767-774.

Kennerdell, J. R. and Carthew, R. W. (2000) Heritable gene silencing in *Drosophila* using double-stranded RNA. Nature Biotechnol. 18: 896–898.

Kennerdell, J.R. and Carthew, R.W. (1998) Use of dsRNA-mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway. Cell 95: 1017–1026.

Kent, N.L. and Evers, A.D. (1994). Kent's Technology of cereals: An introduction for students of food science and agriculture. Pp17-18, 50, 96-97, New York, USA: Elsevier Science Ltd.

Khush, G.S, and Brar, D.S, (1991) Genetics of resistance to insects in crop plants, Adv. Agron. 45: 223-274.

Kim, J.W., Closs, E.I, Aalbritton, L.M. and Cunningham, J.M., (1991). Transport of cationic amino acids by the ecotropic virus receptor. Nature, 352: 725–728.

Kirschke, H. (1998). Cathepsin L. In Handbook of Proteolytic Enzymes. Barrett, A.J., Rawlings, N.D., Woessner, J.F. eds. London, Academic Press: pp. 617-621.

Koliatakudy, P.E. and Koller, W., (1983). Fungal penetration of the first line defensive barriers of plants, ln : Callow J A. ed. Biochemical plant pathologv. Chichester: John Wiley.

Kopetzki, E., Hopfner, K.P., Bode, W. and Huber, R., (1999). Zymogenic protease precursors that can be autocatalytically activated and their use. Int. Patent WO 99/10503.

Kramer, K.J., Morgan, T.D., Throne, J.E., Dowell, F.E., Bailey, M. and Howard, J.A., (2000). Transgenic avidin maize is resistant to storage insect pests. Nature Biotech. 18: 670–764.

Kunkel, H., (1976). Membrane feeding systems in aphid research. In: K.F. Harris and K. Maramorosch, Editors, Aphids as Virus Vectors, Academic Press, New York, pp. 311–337.

Kuroda, M., Kiyosaki, T., Matsumoto, I., Misaka, T., Arai, S. and Abe, K. (2001). Molecular cloning, characterization and expression of wheat cystatins. Biosci. Biotechnol. Biochem., 65: 22-28.

Kutsukake, M., Shibao, H., Nikoh, N., Morioka, M., Tamura, T., Hoshino, T., Ohgiya, S. and Fukatsu, T., (2004). Venomous protease of aphid soldier for colony defense. Proc Natl Acad Sci USA. 101:11338–11343.

Laemlli, U.K., (1970). Clevage of structural proteins during the assembly of head of bacteriophage T4. Nature, 227:680–685.

Lawrence, P.K. and Koundal, K.R., (2002). Plant protease inhibitors in control of phytophagous insects. Electron J Biotechnol. [online]; 5(1). Available from: http://www.ejbiotechnology. info/content/vol5/issue1/full/3/.

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

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

Lemaire, S.D., Richardson, J.M., Goyer, A., Keryer, E., Lancelin, J.M., Makhatadze, G.I., and Jacquot, J.P., (2000) Primary structure determinants of the pH and temperature dependent aggregation of thioredoxin *h*. Biochem. Biophys. Acta 1476: 311-323.

Lemos, F.J.A. and Terra, W. R., (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.

Leple, J.C., Bonade, M., Augustin, S., Pilate, G., Dumanois, V., Delplanque, A., Cornu, D. and Jouanin, L., (1995). Toxicity to *Chrysomela tremulae* (Coleoptera: Chrysomelidae) of transgenic poplars expressing a cysteine proteinase inhibitor. Mol Breed. 1: 319–328.

Leszczynski, B., Matok, H. and Dixon, A.F.G., (1992). Resistance of cereals to aphids: the interaction between hydroxamic acids and UDP-glucose transferases in the aphid *Sitobion avenae* (Homoptera:Aphididae). J. Chem. Ecol. 18: 1189-1200.

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

Levine, D., Kaplan, M.J. and Greenway, P.J., (1972). The Purification and Characterization of Wheat-Germ Agglutinin. Biochem. J. 129: 847-856.

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

Li, H., Tang, H., Sivakmar, S., Philip, J., Harrison, R.L., Gatehouse, J.A., Bonning, B.C., (2008). Insecticidal activity of a basement membrane-degrading protease against *Heliothis virescens* (Fabricius) and *Acyrthosiphon pisum* (Harris). J. Insect Physiol. 54: 777–789

Lindbo, J.A. and Dougherty, W.D. (2005) Plant pathology and RNAi: A brief history. Annu. Rev. Phytopathol. 43: 191–204.

Lipke, H., Liener, I.E. and Fraenkel, G., (1954). Effects of soybean inhibitors on growth of *Tribolium confusum*. J. Agric. Food Chem., 2: 410-414.

Liu, J., Shi, G.P., Zhang, W.Q., Zhang, G.R. and Xu, W.H., (2006). Cathepsin L function in insect moulting: molecular cloning and functional analysis in cotton bollworm, *Helicoverpa armigera*. Insect Mol. Biol. 15: 823–834.

Llano, E., Adam, G., Pendas, A. M., Quesada, V., Sanchez, L. M., Santamaria, I., Noselli, S. and Lopez-Otin, C. (2002). Structural and enzymatic characterization of *Drosophila* Dm2-MMP, a membrane-bound matrix metalloproteinase with tissue-specific expression. J. Biol. Chem. 277: 23321 -23329.

Ma, C. and Kanost, M.R., (2000). A beta-1,3-glucan recognition protein from an insect, *Manduca sexta*, agglutinates microorganisms and activates the phenoloxidase cascade. J. Biol. Chem. 275: 7505–7514.

Machleidt, W., Thiele, U., Laber, B., Assfalg-Machleidt, I., Esterl, A., Wiegand, G., Kos, J., Turk, V. and Bode, W. (1989). Mechanism of inhibition of papain by chicken egg white cystatin. FEBS Lett. 243: 234–238.

Mackintosh, C.A., Lew, J.L., Radmer, L.E., Shin, S., Heiinen, S.J., Smith, L.A., Wyckoff, M.N., Dill-Macky, R., Evans, C.K., Kravchenko, S., Baldridge, G.D., Zeyen, R.J. and Muehlbauer, G.J., (2007). Overexpression of defense response genes in

transgenic wheat enhances resistance to *Fusarium* head blight. Plant Cell Rep., 26: 479-488.

Mangelsdorf, P.C (1953). "Wheat." *Scientific American*: 2–11.

Mao, Y.B., Cai, W.J., Wang, J.W., Hong, G.J., Tao, X.Y., Wang, L.J., Huang, Y.P. and Chen, X.Y., (2007). Silencing a cotton bollworm P450 monooxygenase gene by plant-mediated RNAi impairs larval tolerance of gossypol. Nat. Biotechnol. 25: 1307–1313.

Mao, J., Burt, A.J., Ramputh, A.L., Simmons, J., Cass, L., Hubbard, K., Miller, S., Altosaar, I. and Arnason, J.T. (2007) Diverted secondary metabolism and improved resistance to European corn borer (*Ostrinia nubilalis*) in maize (*Zea mays* L.) transformed with wheat oxalate oxidase. J. Agric. Food Chem. 55: 2582-9

Marie, B., Bacon, J.P. and Blagburn, J.M., (2000). Double-stranded RNA interference shows that Engrailed controls the synaptic specificity of identified sensory neurons. Curr Biol. 10:289-92.

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. 121: 464–476.

Matsumoto, I., Watanabe, H., Abe, K., Arai, S. and Emori, Y., (1995). A putative digestive cysteine proteinase from *Drosophila melanogaster* is predominantly expressed in the embryonic and larval midgut. Eur. J. Biochem. 227: 582–587.

Mazumdar-Leighton, S. and Broadway, R.M., (2001). Transcriptional induction of diverse midgut trypsins in larval *Agrotis ipsilon* and *Helicoverpa zea* feeding on the soybean trypsin inhibitor. Insect Biochem. Mol. Biol. 31: 645–657.

Middelberg, A.P.J., (2002). Preparative protein refolding, Trends Biotechnol. 20: 437–443. U.S. Patent, 4977248.

Misquitta, L. and Paterson, B.M., (1999) Targeted disruption of gene function in *Drosophila* by RNA interference (RNAi): a role for nautilus in embryonic somatic muscle formation. Proc. Natl. Acad. Sci. U. S. A. 96: 1451–1456

Mittapalli, O., Stuart, J.J. and Shukle, R.H. (2005). Molecular cloning and characterization of two digestive serine proteases from the Hessian fly, *Mayetiola destructor*. Insect Mol. Biol. 14: 309-318.

Miyaji, T., Kouzuma, Y., Yaguchi, J., Matsumoto, R., Kanost, M.R., Kramer, K.J., Yonekura, M., (2007). Purification of a cysteine protease inhibitor from larval hemolymph of the tobacco hornworm (*Manduca sexta*) and functional expression of the recombinant protein. Insect Biochem. Mol. Biol. 37: 960–968.

Mukanganyama S., Figueroa, C.C., Hasler, J.A. and Niemeyer, H.M., (2003). Effects of DIMBOA on detoxification enzymes of the aphid *Rhopalosiphum padi* (Homoptera: aphididae) J. Insect Physiol. 49: 223–229.

Murdock, L.L. and Shade, R.E., (2002). Lectins and protease inhibitors as plant defenses against insects. J. Agric. Food Chem., 50: 6605-6611.

Murdock, L.L., Huesing, J.E., Nielsen, S.S., Pratt, R.C. and Shade, R.E., (1990). Biological effects of plant lectins on the cowpea weevil. Phytochemistry 29: 85–89.

Murray, C. and Christeller, J.T., (1995). Purification of a trypsin-inhibitor (Pfti) from pumpkin fruit phloem exudate and isolation of putative trypsin and chymotrypsin inhibitor cDNA clones. Biological Chemistry Hoppe-Seyler 376: 281–287.

Nandi, A.K., Basu, D., Das, S. and Sen, S.K. (1999). High level expression of soybean trypsin inhibitor gene in transgenic tobacco plants failed to confer resistance against damage caused by *Helicoverpa armigera*. J. Biosci.24: 445-452.

Nappi, A.J., Frey, F. and Carton, Y., (2005). *Drosophila* serpin 27A is a likely target for immune suppression of the blood cell-mediated melanotic encapsulation response. J. Insect. Physiol. 51: 197–205.

Niemeyer, H.M. and Jerez, J.M., (1997). Chromosomal location of genes for hydroxamic acid accumulation in *Triticum aestivum* L. (wheat) using wheat aneuploids and wheat substitution lines. Heredity 79:10–14.

Niemeyer, H.M. and Perez, F.J., (1995). Potential of hydroxamic acids in the control of cereal pests, diseases, and weeds. In: Inderjit, Dakshini, K.M.M., Einhellig, F.A. (Eds.), Allelopathy: Organisms, Processes, and Application. ACS Symposium Series Chemical Society, Washington, DC, pp. 260–270.

Niemeyer, H.M., Pesel, E., Franke, S. and Francke, W., (1989). Ingestion of the benzoxazinone DIMBOA from wheat plants by aphids. Phytochemistry 28: 2397–2310.

Oerke, E.C., Dehne, H.W., Schonbeck, F. and Weber, A. (1994). Crop production and crop protection: Estimated losses in major food and cash crops. Elsevier, Amsterdam, The Netherlands.

Ohtsubo S., Kobayashi H., Noro W., Taniguchi M. and Saitoh E., (2005). Molecular cloning and characterisation of Oryzacystatin-III, a Novel member of Phytocystatin in Rice (*Oryza sativa* L. *japonica*). J. Agric. Food Chem. 53: 5218-5224

Ojima, A., Shiota, H., Higashi, K., Kamada, H., Shimma, Y.I., Wada, M. and Satoh, S., (1997). An extracellular insoluble inhibitor of cysteine proteinases in cell cultures and seeds of carrot. Plant Mol. Biol. 34: 99–109.

Page-McCaw, A., Serano, J., Sante, J.M. and Rubin, G.M., (2003). Drosophila matrix metalloproteinases are required for tissue remodeling, but not embryonic development. Develop. Cell 4: 95–106.

Patankar, A.G., Giri, A.P., Harsulkar, A.M., Sainani, M.N., Deshpande, V.V., Ranjekar, P.K. and Gupta, V.S. (2001) Complexity in specificities and expression of *Helicoverpa armigera* gut proteinases explains polyphagous nature of the insect pest. Insect Biochem. Mol. Biol. 31: 453–464.

Patnaik, D. and Khurana, P., (2001), Wheat biotechnology: A minireview. Electronic J. Biotechnol.4: 74-102

Pemberton, P.A. and Bird, P.I., (2003). Production of serpins using yeast expression systems. Methods 32: 185–190.

Peterson, F.C., Gordon, N.C. and Gettins, P.G.W., (2001). High-level bacterial expression and ^{15}N-alanine-labeling of bovine trypsin. Application to the study of trypsin inhibitor complexes and trypsinogen activation by NMR spectroscopy. Biochemistry 40: 6275–6283.

Peumans, W. J. and Van Damme, E. J. M., (1998). Plant lectins: A composite of several distinct families of structurally and evolutionary related proteins with diverse biological roles. Crit. Rev. Plant Sci. 17: 575-692.

Peumans, W. J., Smeets, K., Van Nerum, K., Van Leuven, F. and Van Damme, E.J. M., (1997). Lectin and allinase are the predominant proteins in nectar from leek (*Allium porrum* L.) flowers. Planta 201: 298-302.

Philip, J.M., Fitches, E., Harrison, R.L., Bonning, B. and Gatehouse, J.A., (2007). Characterisation of functional and insecticidal properties of a recombinant cathepsin L-like proteinase from flesh fly (*Sarcophaga peregrina*), which plays a role in differentiation of imaginal discs. Insect. Biochem. Mol. Biol. 37: 589–600.

Pitrat, M. and Lecoq, H., (1982). Relations génétiques entre les résistances par non-acceptation et par antibiose du melon à *Aphis gossypii*. Recherche de liaisons aved'autres gènes. Agronomie 2: 503-508.

Poerio, E., Di Gennaro, S., Di Maro, A., Farisei, F., Ferranti, P.. and Parente, A., (2003). Primary Structure and Reactive Site of a Novel Wheat Proteinase Inhibitor of Subtilisin and Chymotrypsin Biol. Chem. 384: 295–304.

Porter, D.R., Burd, J.D., Shufran, K.A. and Webster, J.A., (2000). Efficacy of pyramiding greenbug (Homoptera: Aphididae) resistance genes in wheat. Environ Entomol 29: 1315-1318.

Powell, K.S., Gatehouse, A.M.R., Hilder, V.A. and Gatehouse, J.A., (1995). Antifeedant effects of plant lectins and an enzyme on the adult stage of the rice brown planthopper *Nilaparvata lugens*. Entomol. Exp. Appl. 75: 51-59.

Powell, K.S., (2001). Antimetabolic effects of plant lectins towards nymphal stages of the planthoppers *Tarophagous proserpina* and *Nilaparvata lugens*. Entomologia Experimentalis et Applicata, 99: 71-77.

Powell, K.S., Spence, J., Bharathi, M., Gatehouse, J.A. and Gatehouse, A.M.R., (1998). Immunohistochemical and developmental studies to elucidate the mechanism of action of the snowdrop lectin on the rice brown planthopper, *Nilaparvata lugens*. J. Insect Physiol. 44: 529-539.

Price, D.R.G (2004). Receptors of peptide hormones in the gut of rice brown planthopper (*Nilaparvata lugens*) as targets for novel insecticides. (PhD. Thesis, Durham University, UK)

Price, D.R.G. and Gatehouse, J.A., (2008). RNAi mediated crop protection against insects. Trends in Bitoechnol. 26: 393-400

Prosser, W.A. and Douglas, A.E., (1992). A test of the hypotheses that nitrogen is upgraded and recycled in an aphid (*Acyrthosiphon pisum*) symbiosis. J. Insect Physiol. 38: 93–99.

Purcell, J.P., Greenplate, J.T., Jennings, M.G., Ryers, J.S., Pershing, J.C., Sims, S.R., Prinsen, M.J., Corbin, D.R., Tran, M., Sammons, R.D. and Stonard, R.J., (1993). Cholesterol oxidase: A potent insecticidal protein active against bollweevil larvae. Biochem. Biophys. Res. Comm. 196:1406-1413.

Puthoff, D.P., Sardesai, N., Subramanyam, S., Nemacheck, J.A. and Williams, C.E., (2005). Hfr-2, a wheat cytolytic toxin-like gene, is up- regulated by virulent Hessian fly larval feeding. Mol. Plant Pathol. 6: 411–423.

Rahbe, Y., Deraison C., Bonade-Bottino, M., Girard, C., Nardon, C. and Jouanin, L., (2003a). Effects of the cysteine protease inhibitor oryzacystatin (OC-I) on different aphids and reduced performance of *Myzus persicae* on OC-I expressing transgenic oilseed rape. Plant Sci. 164: 441–450.

Rahbe, Y., Ferrasson, E., Rabesona, H. and Quillies, L., (2003b). Toxicity to the pea aphid *Acyrthosiphon pisum* of anti-chymotrypsin isoforms and fragments of Bowman-Birk protease inhibitors from pea seeds. Insect Biochem. Mol. Biol 33: 299-306.

Rahbe, Y., Sauvion, N., Febvay, G., Peumans, W.J. and Gatehouse, A.M.R., (1995). Toxicity of lectins and processing of ingested proteins in the pea aphid *Acyrthosiphon pisum*. Entomologia Experimentalis et Applicata 76: 143-155.

Raikhel, N.V. and Wilkins, T.A., (1987). Isolation and characterization of a cDNA encoding wheat germ agglutinin. Proc Natl Acad Sci USA 84: 6745–6749.

Rao, K.V., Rathore, K.S., Hodges, T.K., Fu, X., Stoger, E., Sudhakar, D., Williams, S., Christou, P., Bharathi, M., Bown, D.P., Powell, K.S., Spence, J.,Gatehouse, A. M. R. and Gatehouse, J.A., (1998). Expression of snowdrop lectin (GNA) in transgenic rice plants confers resistance to rice brown planthopper. Plant J. 15: 469-477.

Ratcliffe, R.H. and Hatchett, J.H., (1997). Biology and genetics of the Hessian fly and resistance in wheat. pp. 47–56 in K. Bonardi (ed.). New Developments in Entomology. Research Signpost, Trivandrum, India

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

Rawlings, N.D. and Barrett, A.J., (1994). Families of cysteine peptidases. Methods Enzymol. 244: 461–500.

Rawlings, N.D., Tolle, D.P. and Barrett, A.J., (2004). Evolutionary families of peptidase inhibitors, J. Biochem., 378: 705-716.

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. Mol. Biol. 30: 1017–1026.

Ribeiro, A.P.O., Pereira, E.J.G., Galvan, T.L., Picanc, M.C., Picoli, E.A.T., da Silva, D.J.H., Fa´ri, M.G and Otoni, W.C., (2006). Effect of eggplant transformed with oryzacystatin gene on *Myzus persicae* and *Macrosiphum euphorbiae*. J. Appl. Entomol. 130: 84-90.

Richer, J.K., Hunt, W.G., Sakanari, J.A. and Grieve, R.B., (1993). *Dirofilaria immitis*: effect of fluoromethyl ketone cysteine protease inhibitors on the third- to fourth-stage molt. Exp. Parasitol. 76: 221–231.

Rigas, D.A., Osgood, E.E. and Duerst, M., (1955). Purification and properties of the phytohemagglutinin of *Phaseolus vulgaris*. J. Biol. Chem. 212: 607–615.

Rispe, C., Kutsukake, M., doublet, V., Hudaverdian, S., Legeai, F., Simon, J.C, Tagu, D. and Fakatsu, T., (2007). Large gene family expansion and variable selective pressures for Cathepsin B in aphids. Mol. Biol. Evol. 25: 5-17.

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

Rossi, M., Goggin, F.L., Milligan, S.B., Kaloshian, I., Ullman, D.E. and Williamson, V.M., (1998). The nematode resistance gene Mi of tomato confers resistance against the potato aphid. Proc. Natl. Acad. Sci. USA 95: 9750–9754.

Royle, D.J. (1976). Structural feature.s of resistance to plant diseases. In: Friend J, Threifall DR, eds. Biochemical aspects of plant-pathogen relationships. London; Academic Press, 161-193.

Ryan, C.A., (1978). Proteinase inhibitors in plant leaves: a biochemical model for pest-induced natural plant protection. Trends Biol. Sci. 3: 148-1 50.

Ryan, C.A., (1990). Proteinase inhibitors in plants: genes for improving defenses against insects and pathogens. Annu. Rev. Phytopathol, 28: 425-449.

Ryan, S.N., Liang W.A. and McManus, M.T., (1998). A cysteine proteinase inhibitor purified from apple fruit. Phytochemistry. 49: 957-63.

Salminen, A. and Gottesman, M.M., (1990). Inhibitor studies indicate that active cathepsin L is probably essential to its own processing in cultured fibroblasts. Biochem. J. 272: 39–44.

Salvucci, M.E., Rosell, R.C. and Brown, J.K., (1998). Uptake and metabolism of leaf proteins by the silverleaf whitefly. Arch. Insect Biochem. Physiol. 39: 155–165.

Sambrook, J. and Russell, D.W., (2001). Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press.

Sandstrom, J. and Moran, N., (1999). How nutritionally imbalanced is phloem sap for aphids? Entomol. Exp. Appl. 91: 203-210.

Sandstrom, J., Telang, A. and Moran, N. A., (2000). Nutritional enhancement of host plants by aphids: a comparison of three aphid species on grasses. J. Insect Physiol. 46: 33-40.

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

Schobert, C., Baker, L., Szederkenyi, J., Großmann, P., Komor, E., Hayashi, H., (1998). Identification of immunologically related proteins in sieve-tube exudate collected from monocotyledonous and dicotyledonous plants. Planta 206: 245–252.

Shah, D.M., (1997). Genetic engineering for fungal and bacterial diseases. Curr. Opin. Biotechnol. 8: 208-214.

Sharma, H.C., Sharma, K.K., Seetharaman, N. and Ortiz, R. (2000). Prospects for using transgenic resistance to insects in crop improvement EJB Electronic Journal of Biotechnology [online]. 15 August 2000, vol.2n.3, Available from /http://www.ejb.org/content/vol3/issue2/full/3. ISSN 0717-3458.

Shelton, A.M., Zhao, J.Z. and Roush, R.T., (2002). Economic, ecological, food safety, and social consequences of the deployment of Bt transgenic plants. Ann. Rev. Entomol. 47: 845–881.

Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y., Ishikawa, H., (2000). Genome sequence of the endocellular bacterial symbionts of aphids *Buchnera sp*. APS. Nature 407: 81–86.

Shukle, R.H., Murdock, L.L. and Gallun, R.L., (1985). Identification and partial characterisation of a major gut proteinase from larvae of the Hessian fly, *Mayetiola Destructor* (say) (Diptera: Cecidomyiidae), Insect. Biochem. 15: 93-101.

Silva, C.P. and Terra, W.R., (1994). Digestive and absorptive sites along the midgut of the cotton seed sucker bug *Dysdercus peruvianus* (Hemiptera: Pyrrhocoridae). Insect Biochem. Mol. Biol. 24: 493–505.

Smith, J.J., and Raikhel, N.V. (1989). Production of an antibody specific for the propeptide of wheat germ agglutinin. Plant Physiol. 91: 473-476.

Sreedharan, S.K, Verma, C., Caves, L.S., Brocklehurst, S.M., Gharbia, S.E., Shah, H.N. and Brocklehurst, K., (1996). Demonstration that 1-trans-epoxysuccinyl-L-leucylamido-(4-guanidino) butane (E-64) is one of the most effective low Mr inhibitors of trypsin-catalysed hydrolysis. Characterization by kinetic analysis and

by energy minimization and molecular dynamics simulation of the E-64-beta-trypsin complex. Biochem J. 316: 777–786.

St. Johnston, D., (2002) The art and design of genetic screens: *Drosophila melanogaster*. Nat. Rev. Genet. 3: 176–188.

Stoger, E., Williams, S., Christou, P., Down, R. E. and Gatehouse, J.A., (1999). Expression of the insecticidal lectin from snowdrop (Galanthus niValis agglutinin, GNA) in transgenic wheat plants: effects on predation by the grain aphid *Sitobion avenae*. Mol. Breed. 5: 65-73.

Subramanyam, S., David, F., Smith, J.C., Clemens, M.A. Webb, Sardesai, N. and Williams, C.E., (2008). Functional Characterization of HFR1, a High-Mannose N-Glycan-Specific Wheat Lectin Induced by Hessian Fly Larvae. Plant Physiol. 147: 1412–1426.

Sunilkumar, G., Campbell L. M., Puckhaber L., Stipanovic R.D. and Rathore K.S., (2006). Engineering cottonseed for use in human nutrition by tissue-specific reduction of toxic gossypol. Proc. Natl. Acad. Sci. U. S. A. 103: 18054–18059.

Supuran, C.T., Scozzafava, A. and Clare, B.W., (2002). Bacterial protease inhibitors. Med. Res. Rev. 22: 329-72.

Tamhane, V.A., Chougule, N.P., Giri, A.P., Dixit, A.R., Sainani, M.N. and Gupta, V.S., (2005). In vitro and in vivo effects of Capsicum annum proteinase inhibitors on *Helicoverpa armigera* gut proteinases. Biochimica et Biophysica Acta, 1722: 156–167.

Tang, H., Li, H., Lei, S.M., Harrison, R.L. and Bonning, B.C., (2007). Tissue specificity of a baculovirus-expressed, basement membrane-degrading protease in larvae of *Heliothis virescens*. Tissue and Cell 39: 431–443.

Tao, K., Stearns, N.A., Cond, J., Wu, Q.L., Sahagian, G.G., (1994). The proregion of cathepsin L is required for proper folding, stability, and ER exit. Arch. Biochem. Biophys. 311: 19–27.

Telang, M.A, Giri, A.P., Pyati, P.S., Gupta, V.S., Tegeder, M., Franceschi, V.R., (2009a). Winged bean chymotrypsin inhibitors retard growth of Helicoverpa armigera. Gene 431: 80–85.

Telang, M.A, Pyati, P.S., Sainani, M., Gupta, V.S., and Giri, A.P., (2009b). *Momordica charantia* trypsin inhibitor II inhibits growth and development of *Helicoverpa armigera*. Insect Science :16: 371–379.

Telang, M.A., Srinivasan, A., Patankar, A., Harsulkar, A., Joshi, V., Damle, A., Deshpande, V., Sainani, M., Ranjekar, P., Gupta, G., Birah, A., Rani, S., Kachole, M., Giri, A. and Gupta, V., (2003). Bitter gourd proteinase inhibitors: potential growth inhibitors of *Helicoverpa armigera* and *Spodoptera litura*. Phytochemistry 63: 643–652.

Terra, W.R, (1990), Evolution of digestive systems of insects. Ann. Rev. Entomol., 35: 181-200.

Terra, W.R., Ferreira, C., Jordao, B.P. and Dillon, R.J., (1996). Digestive enzymes. In: Lehane, M.J., Billingsley, P.F. (Eds.), Biology of the Insect Midgut. Chapman and Hall, London, pp. 153–194

Theis, N. and Lerdau, M., (2003). The evolution of function in plant secondary metabolites. International Journal of Plant Science, 164: S93-S102.

Thompson, G.A. and Schulz, A., (1999). Macromolecular trafficking in the phloem. Trends in Plant Sci. 4: 354-360.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G., (1997).The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25: 4876–4882.

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

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

Tran, P., Cheesbrough, T.M. and Keickhefer, R.W., (1997). Plant proteinase inhibitors are potential anticereal aphid compounds. J. Econ. Entomol. 90: 1672-1677.

Trebicki P, Harding, R.M. and Powell, K.S., (2009). Anti-metabolic effects of *Galanthus nivalis* agglutinin and wheat germ agglutinin on nymphal stages of the common brown leafhopper using a novel artificial diet system Entomologia Experimentalis et Applicata 131: 99–105.

Tsumoto, K., Ejima, D., Kumagai, I. and Arakawa, T., (2003). Practical considerations in refolding proteins from inclusion bodies (Mini Review). Protein Expr. Purific. 28: 1–8.

Turk, B., Dolenc, I., Turk, V. and Bieth, J.G. (1993). Kinetics of the pH-induced inactivation of human cathepsin L. Biochemistry 32: 375-380.

Turk, B., Dolenc, I., Zerovnik, E., Turk, D., Gubensek, F. and Turk, V. (1994). Human cathepsin B is a metastable enzyme stabilized by specific ionic interactions associated with the active site. Biochemistry 33: 14800-14806.

Turk, V. & Bode, W. (1991). The cystatins: protein inhibitors of cysteine proteinases, FEBS Lett. 285: 213 -219.

Valueva, T.A. and Mosolov, V.V., (2004). Role of inhibitors of proteolytic enzymes in plant defense against phytopathogenic microorganisms. Biochemistry, 69: 1305-1309.

Van Damme, E.J.M., Allen, A.K. and Peumans, W.J., (1987). Isolation and characterization of a lectin with exclusive pecificity towards mannose from snowdrop (*Galanthus nivalis*) bulbs. FEBS Letters 215: 140–144.

Van Damme, E.J.M., Smeets, K., Torrekens, S., Van Leuven, F., Goldstein, I. J. and Peumans, W.J., (1992). The closely related homomeric and heterodimeric mannose-binding lectins from garlic are encoded by one domain and two-domain lectin genes, respectively. Eur. J. Biochem. 206: 413–420.

Van Damme, E.J.M., Peumans, W.J., Barre, A. and Rougé, P., (1998). Plant lectins: a composite of several distinct families of structurally and evolutionary related proteins with diverse biological roles. Crit. Rev. Plant Sci. 17: 575–692.

Van Damme, N. M., Horn, Mares, M. and Baldwin, I.T., (2001). Ontogeny constrains systemic protease inhibitor response in *Nicotiana attenuata*. J. Chem. Ecol. 27: 547–568.

Vasquez, J.R., Evnin, L.B., Higaki, J.N. and Craik, C.S., (1989). An expression system for trypsin. J. Cell. Biochem. 39: 265–276.

Vermehren, A., Qazi, S. and Trimmer, B.A., (2001). The nicotinic alpha subunit MARA1 is necessary for cholinergic evoked calcium transients in Manduca neurons. Neurosci. Lett. 313: 113–116.

Walz, C., Giavalisco, P, Schad, M, Juenger, M, Klose, J. and Kehr, J., (2004). Proteomics of curcurbit phloem exudate reveals a network of defence proteins. Phytochemistry 65: 1795–1804.

Wang, H., Kavanaugh, M.P., North, R.A., Kabat, D., (1991). Cell-surface receptor for ecotropic murine retroviruses is a basic amino-acid transporter. Nature 352: 729–731.

Wang, Z. B. and Guo, S. D., (1999). Expression of two insect-resistant genes cryIA (b&c) GNA in transgenic tobacco plants results in added protection against both cotton bollworm and aphids. Chinese Sci. Bull., 44: 2051-2058.

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

Wink, M., (1999). Functions of plant secondary metabolites and their exploitation in biotechnology. Annual Plant Reviews, Vol. 3 (ed. M. Wink), Sheffield Academic Press, Sheffield, UK. 304 pp.

Wolfersberger, M. G., (2000) Amino acid transport in insects. Annu. Rev. Entomol. 2000. 45: 111–120.

Wu, Y., Llewellyn, D., Mathews, A., and Dennis, E.S. (1997). Adaptation of *Helicoverpa armigera* (Lepidoptera, Noctuidae) to a proteinase inhibitor expressed in transgenic tobacco. Mol. Breed. 3: 371–380.

Yamamoto, Y. and Takahashi, S.Y., (1993). Cysteine proteinase from *Bombyx* eggs: role in programmed degradation of yolk proteins during embryogenesis. Comp. Biochem. Physiol. 106: 35–45.

Yan, J., Cheng, Q., Li, C.B. and Aksoy, S., (2002). Molecular characterization of three gut genes from *Glossina morsitans morsitans*: cathepsin B, zinc-metalloprotease and zinc-carboxypeptidase. Insect Mol. Biol. 11: 57–65.

Yee, L. and Blanch, H.W., (1993). Recombinant trypsin production in high cell density fed-batch cultures in Escherichia coli. Biotechnol. Bioeng. 41: 781–790.

Yoo, B.C, Aoki. K., Xiang, Y., Campbell, L.R., Hull, R.J., Xoconostle-Cazares, B., Monzer, J., Lee, J.Y., Ullman, D.E. and Lucas, W.J., (2000). Characterization of *Cucurbita maxima* phloem serpin-1 (CmPS-1). J. Biol. Chem. 275: 35122–35128.

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

Yoza, K., Nakamura, S., Yaguchi, M., Haraguchi, K. and Ohtsubo, K., (2002). Molecular cloning and functional expression of cDNA encoding a Cysteine Proteinase inhibitor, cystatin, from Job's Tears (*Coix Lacryma-jobi* L. var. *Ma-yuen* Stapf). Biosc. Biotechnol. Biochem, 66: 2287-2291.

Yurchenco, P.D. and O'Rear, J., (1993). Supramolecular organization of basement membranes. In: Rohrbach, D.H., Timpl, R. (Eds.), Molecular and Cellular Aspects of Basement Membranes. Academic Press, New York, pp. 19–47.

Zdobnov, E.M., von Mering, C., Letunic, I., Torrents, D., Suyama, M., Copley, R.R., Christophides, G.K., Thomasova, D., Holt, R.A., Subramanian, G.M., et al., (2002).

Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*. Science 298: 149–159.

Zeigler, H., (1975). Nature of transported. In Zimmermann H, Milburn JA (ed): Encyclopedia of Plant Physiology, *New Series*. Berlin: Springer-Verlag, vol 1, pp 59-100.

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

Zhu-Salzman, K., Shade, R.E., Koiwa, H., Salzman, R.A., Narasimhan, M., Bressan, R.A., Hasegawa, P.M., Murdock, L.L., (1998). Carbohydrate binding and resistance to proteolysis control insecticidal activity of *Griffonia simplicifolia* lectin II. Proc. Natl. Acad. Sci. U.S.A., 95: 15123-15128.

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

Zou, Z., Wang, Y. and Jiang, H. (2005). *Manduca sexta* prophenoloxidase activating proteinase-1 (PAP-1) gene: organization, expression, and regulation by immune and hormonal signals. Insect Biochem. Mol. Biol. 35: 627-636.