Efficacy of diverse *Capsicum annuum* protease inhibitors against the adaptive plasticity of *Helicoverpa armigera* proteases

A thesis submitted to the

SAVITRIBAI PHULE PUNE UNIVERSITY

For the degree of

DOCTOR OF PHILOSOPHY

In

BIOTECHNOLOGY

By

NEHA MAHAJAN

Research Guide

Dr. A. P. GIRI

Plant Molecular Biology Unit
Division of Biochemical Sciences
CSIR-National Chemical Laboratory
Pune-411008 (India)

September 2014

This work was performed at

CSIR - NATIONAL CHEMICAL LABORATORY,

PUNE (INDIA)

Dr. Ashok P. Giri (Research supervisor)

Dr. (Mrs.) Vidya S. Gupta (Co-supervisor)

Plant Molecular Biology Unit
Division of Biochemical Sciences
CSIR-National Chemical Laboratory
Pune-411008 (India)

CERTIFICATE

Certified that the work in the Ph.D. thesis entitled **"Efficacy of diverse *Capsicum annuum* protease inhibitors against the adaptive plasticity of *Helicoverpa armigera* proteases"** submitted by **Ms. Neha Mahajan** was carried out by the candidate under my supervision. The material obtained from other sources has been duly acknowledged in the thesis.

Date:

Place:

Dr. Ashok P. Giri
(Research Guide)

Dr. Vidya Gupta
(Reseach Co-guide)

DECLARATION

I hereby declare that the thesis entitled "**Efficacy of diverse *Capsicum annuum* protease inhibitors against the adaptive plasticity of *Helicoverpa armigera* proteases"** submitted for Ph.D. degree to the **Savitribai Phule Pune University** has been carried out at CSIR- National Chemical Laboratory, Pune 411008, India. This work is original and has not been submitted by me for any degree at any other university.

Date:

Neha Mahajan
Division of Biochemical Sciences
CSIR-National Chemical Laboratory
Pune-411008 (India)

CONTENTS

Acknowledgements

Ph. D. was a dream accomplishment and I am lucky to have treasured this milestone. The course of Ph.D. has been a great experience and a lesson of perseverance for me.

My Ph.D. started after I received the CSIR-JR Fellowship and was aptly guided by Dr. G.B. Zore to join the PMB-Unit, Division of Biochemical Sciences, NCL, Pune. My sincere thanks to my supervisor, Dr. Ashok Giri, for believing in me. His support and guidance throughout has helped me tread along. His encouragement and enthusiasm always inspires everyone around. I am also grateful to my Co-supervisor, Dr. Vidya Gupta who has given her timely suggestions and honest opinions. Her devotion presents an ideal. I would also like to thank Dr. Narendra Kadoo for his help. All the three scientists shower immense love and are concerned for every member in the group.

I appreciate the constructive suggestions offered by Dr. Vaijayanti Tamhane, Dr. Manasi Mishra and Dr. Purushottam Lomate while planning my thesis and conducting experiments. I would like to thank Dr. Bhushan Dholakia, Dr. Vishal Dawkar, Rakesh J., and Sandip for their scientific inputs in my work. I also value the assistance of project trainees Pragati, Tejashree, Sonal, Yashpal, Rohan, Mehul and Veena.

I would like to mention Yojana and Priya for their joyful company. I will always remember our night-outs at PMB. A special thanks to Charu, Sheetal and Yashashree for managing the scientific requirements of our group. My heartfelt thanks to Ajit, Aditi, Amey, Amol J., Amol K., Ashish, Atul, Ashiwini, Deepika, Hemangi, Dr. Hemalata, Gouri Patil, Gauri Pathak, Gayatri, Medha, Dr. Meena, Neha K., Nidhi, Priyanka, Pranjali, Prasad, Ram, Rasika, Ramesha, Ramya, Ranjit, Reema, Rahul, Rakesh P., Dr.Rebecca, Sachin, Saleem, Smrati, Sonal, Sucheta, Tejas, Uma, Vitthal, Varsha, Yamini and all other labmates for making my experience enjoyable. Thank you all for sparing liquid nitrogen, co-autoclaving, co-deconting and numerous merry tasks that we have done together. Mr. Jagtap's bhel parties and ginger tea will always be cherished.

I am also grateful to Dr. Abhijeet Kulkarni from Department of Bioinformatics, University of Pune, for extending his help in analyzing the NGS data. I thank Dr. Mahesh Kulkarni, Sheon, Jagdisha, Yashwant and Yugendra for their help in proteomic analysis. I express my thanks to Dr. Sourav Pal, Director and Dr. S. Sivaram, ex-Director, NCL

for providing me the research facilities. I acknowledge the Council of Scientific and Industrial Research, New Delhi for the research fellowship. I also acknowledge Department of Science and technology, New Delhi for "International travel support" grant for attending "The Second International Symposium on Insect Midgut Biology" at South China Normal University, Guangzhou, China during September 24-28, 2012.

Finally, I would like to express gratitude towards all my family members. I remember my paternal and maternal grandfathers who would have been very happy to see me achieve this objective. I am indebted to my parents (Sanjyot and Sandeep Mahajan) who have made me capable of accomplishing my goal through their love and faith. My sister (Mitali), who is a delightful friend and my grannies who were keen to know when I would start writing. My husband Vivek, who has always been unconditionally supportive and cheered me all through. A very special mention of my mother-in-law and father-in-law, Vibha and Shriram Deshpande who have motivated and comforted me throughout. Apart from my family, I also owe my thanks to all my teachers and mentors all through. The blessings and good wishes of my friends and relatives whom I have mentioned or missed has helped me reach here. Above all I thank the almighty for giving me enough strength to fulfill my dreams.

-Neha

LIST OF ABBREVIATIONS

AA	Amino acids
ACN	Acetonitrile
AI	Amylase Inhibitor
Bt	*Bacillus thuringiensis*
CanPI	*Capsicum annuum* Proteinase Inhibitor
CBB	Coomassie Brilliant Blue R-250
cDNA	Complementary deoxyribonucleic acid
CI	Chymotrypsin Inhibitor
DTT	Dithiothreitol
DNA	Deoxyribonucleic acid
EDTA	Ethylenediamine tetra acetate
EMSA	Electrophoretic Mobility Shift Assay
g, mg, μg, ng	Gram, milligram, microgram, nanogram
g	Gravitational constant
GXCT	Gel- X ray film Contact-print Technique
HaTry	*Helicoverpa armigera* trypsin
HaChy	*Helicoverpa armigera* chymotrypsin
HCl	Hydrochloric acid
HGPI	*Helicoverpa armigera* Gut Proteinase Inhibitor
HGPs	*Helicoverpa armigera* Gut Proteinases
IEF	Iso-electric Focussing
IRD	Inhibitory Repeat Domain
IPG	Immobilized pH gradient
JA	Jasmonic Acid, jasmonate
kDa/kD	Kilo Dalton
L, mL, μL	Liter, milliliter, microliter
M, mM, μM	Molar, millimolar, micromolar
Fmol,mmole, μmole	Femtomole, millimole, micromole
mRNA	Messenger RNA

miRNA	MicroRNA
NaOH	Sodium hydroxide
NaProPI	Pin-II proteinase inhibitor precursor from *N. alata*
NaTPI	*N. alata* trypsin proteinase inhibitor
OD	Optical Density
PAGE	Poly-acrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
PDB	Protein data bank
PI	Protease Inhibitor
Pin-I/II	Potato Proteinase Inhibitor I/II
rCanPI	Recombinant *C. annuum* proteinase inhibitor
RNA	Ribonucleic Acid
ROS	Reactive oxygen species
rpm	Revolutions per minute
RSL	Reactive site loop
RT-PCR	Reverse Transcriptase-Polymerase Chain Reaction
smRNA	Small RNA
SA	Salicylic Acid, Salicylate
SDS	Sodium Dodecyl Sulphate
TFA	Trifluoroacetic acid
TI	Trypsin Inhibitor
TIU	Trypsin inhibitory units
TLCK	N-*p*-Tosyl-L-Lysine Chloromethyl Ketone
TPCK	N-*p*-Tosyl-L-Phenylalanine Chloromethyl Ketone
TRIS	Tris-Hydroxymethyl Aminomethane
2-D	2-dimensional gel electrophoresis

List of Figures

Chapter 1

Chapter 2A

Chapter 2B

Chapter 3

Chapter 4

List of Tables

Chapter 1

Chapter 2A

Chapter 2B

Chapter 3

Chapter 4

Thesis Summary

Stress accompanies metabolic reorganization in plants, which in turn aids in accumulation of defensive molecules. In *Capsicum annuum*, elicitation of wound response leads to the systemic induction of significant amount of protease inhibitor (PI) proteins. *Capsicum annuum* (Solanaceae) is one of the non-preferred host plants of *H.armigera*, a polyphagous pest. In order to identify the candidate molecules for effective insect control, the investigation of regulation of these PI genes as well as the molecular analysis of insect responses to these PIs needs to be performed. Carefully planned strategies keeping in view the dynamic nature of insect gut proteinases, can surely lead to robust solutions. The major objectives and the results obtained are summarized in the following points;

Characterization of induced proteins and newly identified PIs in *C. annuum*

Stress inducible proteomic changes in C. annuum leaves

Herbivore attack induces defense responses in plants, activating several signaling cascades. As a result, molecules deterrent to the herbivores are produced and accumulated in plants. Expression of defense mechanism/traits requires reorganization of the plant metabolism, redirecting the resources otherwise meant for growth. The objective of the present study was to identify/characterize the proteomic changes in the induced *C. annuum* plants. *C. annuum* plants were subjected to two types of induction treatments namely, wounding followed by application of sterile water or *H. armigera* OS (to mimic insect infestation). Uninduced/control leaves were collected from unwounded healthy plants. The proteomic profiles of systemic leaves from the treated and untreated *C. annuum* plants were analyzed by classical two-dimensional gel electrophoresis and mass spectrometric identification of proteins. We identified various proteins showing differential abundance, which were linked to different metabolic pathways/processes, in the induced leaf tissues. Majority of proteins identified as differentially accumulated, were having roles in redox metabolism and photosynthesis. For example, superoxide dismutase and NADP oxidoreductase were upregulated by 10- and 6-fold while carbonic anhydrase and fructose-1,6-bisphosphatase were downregulated by 9- and 4-fold, respectively. Selected proteins were further evaluated for their transcript accumulation by

semi-quantitative PCR at successive time points after wounding to support the mass spectrometric identification results. In general, proteins having role in defense and damage repair were upregulated while those involved in photosynthesis appeared downregulated.

Structural Features of Diverse Pin-II Proteinase Inhibitor Genes from Capsicum annuum

Plant protease inhibitors, their gene structure, expression dynamics and mechanism of action; have gained/assumed prime importance in the study of plant-insect interaction. Wound-inducible PIs like potato type I and II serine protease inhibitors are being considered as candidates for boosting plant defense in crop plants (Dunse et al., 2010). CanPI gene expression is regulated spatially, temporally as well as qualitatively and quantitatively upon different elicitation. Detailed investigation of PI genes, in *C. annuum,* revealed an array of different PI genes formed as a result of combinations of various unique inhibitory repeat domains (IRDs) (Mishra et al., 2012). These genes, possessing 1 to 4 IRD repeats, were characterized for their sequence and functional diversity. However, complete characterization of CanPI genes including 3' and 5' UTRs is necessary while considering them for potential utilization. In the current study we have characterized the gene architecture of a *C. annuum* Pin-II gene, *CanPI-7* (Beads on string- Type A). The 5'UTR, 3'UTR and intronic sequences of *CanPI-7* gene were obtained. In the genomic sequence, exon 1 (49 bp) and exon 2 (740 bp) are interrupted by a 294 bp long type I intron. We also report for the first time, the occurrence of multi-domain PIs with circularly permuted domain organization in *C. annuum*. The full CDS of three circularly permuted (Pin-II type B) genes were identified (*CanPI-69, -70, -71*) during the present investigation. *CanPI-69* was characterized for the 5'UTR, 3'UTR, and intronic sequences. *CanPI-69* possesses 18 bp and 172 bp long 5'UTR and 3'UTR, respectively. A 585 bp long intron type I was found to be located in the type B CanPI gene *CanPI-69*. The intron is flanked by exon 1 (49 bp) and exon 2 (551 bp). Both CanPI-7 and 69 genes possess several conserved features characteristic of the Pin-II type PIs. Further, the upstream sequences of *CanPI-7* (beads on string) and *CanPI-69* (clasped bracelet) were explored and various transcription factor-binding sites were predicted using database search. The binding of transcription factors to the promoter sequences was

validated by EMSA. The DNA bound proteins identified using mass spectrometry were transcription factors like MYB, MYC, zinc finger motif and MADS-box. Type A and B Pin-II protein sequences were also evaluated using structure prediction. Taking into account the gene structure of Pin-II genes, each unique coding sequence should possess unique 3'UTR sequence. Using RACE, about 24 sequences were accumulated for the 3'UTR analysis of CanPI genes. These 3'UTR sequences obtained, support the above hypothesis. This investigation of the CanPI gene architecture will enable better understanding of the genetic elements governing the CanPI gene expression.

Molecular response of *Helicoverpa armigera* upon exposure to multi-domain *Capsicum annuum* protease inhibitor-7

A multi-domain Pin-II type protease inhibitor from *C. annuum* (CanPI-7) is known to be effective against the insect pest, *H. armigera*. The present study is an attempt to investigate the optimal dose of recombinant CanPI-7 (rCanPI-7) for effective antibiosis to *H. armigera* and further to characterize the responses of digestive proteases upon rCanPI-7 ingestion. The gut protease activity was assessed biochemically and transcript accumulation pattern for selected trypsin and chymotrypsin genes was analyzed by Real-Time PCR. The growth retardation upon exposure to rCanPI-7 was more prominent in neonates as compared to third instar larvae. Influence of stage and dosage of rCanPI-7 was conspicuous on the expression and regulation of candidate trypsin and chymotrypsin genes in *H. armigera*. Monitoring the temporal response of trypsin and chymotrypsin in the third instar *H. armigera* larvae upon rCanPI-7 ingestion, depicted a transformation in the expression patterns of the transcripts; the transcripts were upregulated in the early period of feeding on rCanPI-7 and gradually stabilized through reduced upregulation and downregulation of certain genes. The transcript accumulation pattern correlated with the protease activity in rCanPI-7 exposed larvae. An overall upregulation of the energy metabolism and redox processes was shown in unbiased proteomic analysis. We conclude that early exposure and specific dose of protease inhibitor are essential for effective antibiosis despite the large diversity and plasticity in the expression of protease genes in *H. armigera*. These results highlight the requirement

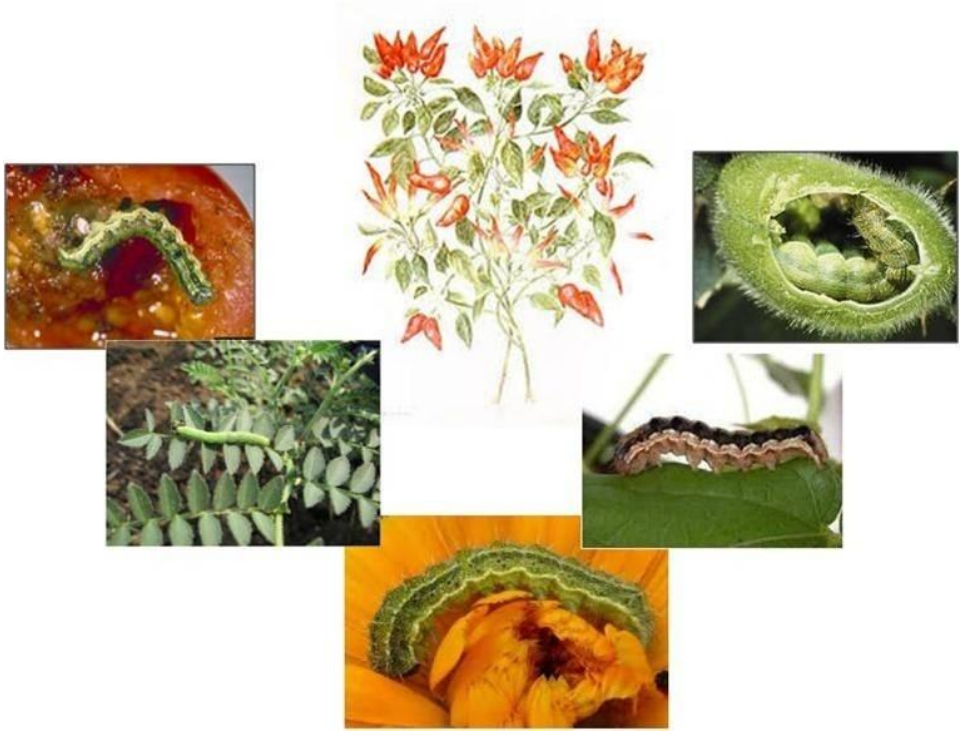
of optimal PI concentration for effective growth retardation and for inhibiting the major gut proteases of *H. armigera*.

Identification and expression profiling of *Helicoverpa armigera* microRNAs and their possible role in the regulation of digestive proteases

In *H. armigera*, proteases might be the key player in the adaptation of insects and their expression can be regulated upon exposure to PIs. Differential protease gene expression was observed in *H. armigera* upon ingestion of diet incorporated with rCanPI-7. The exact mechanism underlying the regulation of protease gene expression against PIs in insect is still unknown. Regulation of proteases might occur at transcriptional or post-transcriptional levels in polyphagous insects. In the present study, we have attempted to establish a rational correlation between the expression of microRNAs and their target proteases to gauge the role of miRNAs in protease regulation. Bioassays were conducted by feeding *H. armigera* larvae on rCanPI-7 and the whole insects were harvested at progressive time points (0.5, 2, 6, 12, 24, and 48 h) for miRNA sequencing. Using Illumina platform, small RNA sequencing was performed for the 12 libraries. Several conserved (96) and novel (90) miRNAs were identified. Bias towards occurrence of specific nucleotides at particular position is evident in the mature miRNA sequences. Most of the miRNAs were observed to be insect-specific, and more than 50 miRNAs were Lepidoptera-specific. Differential expression of several miRNAs after ingestion of CanPI-7 was observed. Negative correlation in the relative abundance of protease targeting miRNAs and their target mRNAs was apparent. The possible role of miRNAs in *H. armigera* protease gene regulation cannot be denied. However, such conclusion can be ascertained after *in vivo* functional validation.

Chapter 1

Introduction and Review of literature



Chapter 1: Introduction and Review of literature

1.1 Plant-insect interaction: Co-evolution for survival

1.2 Losses in field crops due to insect pests

1.3 *Helicoverpa armigera*: A polyphagous insect pest

- 1.3.1 Adaptive physiology and biochemistry of *H.armigera*
- 1.3.2 Current strategies for *H.armigera* control

1.4 Plant defense mechanisms

1.5 Plant protease inhibitors

- 1.5.1 Plant protease inhibitor families
- 1.5.2 Potato inhibitor type II (Pin-II): Role in endogenous and defense functions
- 1.5.3 Structure and mechanism of inhibition
- 1.5.4 Insect resistant transgenic plants expressing PIs

1.6 *Capsicum annuum*: Occurrence and diversity in Pin-II genes

1.7 Antibiosis of CanPIs to *H. armigera*

- 1.7.1 Effect of CanPIs on insect growth and development
- 1.7.2 Merits of CanPI-7

Chapter 1: Introduction and Review of literature

1.1. Plant-insect interaction: Co-evolution for survival

Plants and insects constitute the two most pervasive taxa in the ecosystem. The study of the interactions between these counterparts of nature has unveiled the elements of co-evolution. The plants attempt to avoid being consumed while herbivorous insects try to obtain nutrition from them. Plants have evolved intricate defense mechanisms to resist insect pests at the same time insects have developed physiological and behavioral traits to counter plant defenses (Schoonhoven et al., 1998). Such interactions assume greater magnitude when applied to the plant-pest context. Investigation of the plant-insect interactions in nature has yielded cues about how plants have survived insect attacks over the period of time. In agriculture insect pests become a major threat and to control them use of heavy chemical pesticide creates severe hazards to our mother nature. Thus improvement in agricultural production requires reduction in damage by insect pests and simultaneously minimizing the use of insecticides. The prudent approach is to preferably include biopesticides while practicing the classical integrated pest management strategies. Additionally, exploring and utilizing natural plant defense molecules in cultivating pest resistant crops comprises one of the key biotechnological approaches in modern day agriculture. However, it is also important to realize that the biochemical evolution in plants has also affects the diversification in insects. Thus, reduction in crop losses due to insect pests requires ways that are compatible with sustainable production and should be derived using holistic approach involving agronomical ecosystems. The reciprocal aspects of these interactions govern the community biology in nature (Ehrlich and Raven, 1964). Addressing problems pertaining to such a complex relationship necessitates the evaluation of the communal responses in order to identify appropriate solutions.

1.2. Losses in field crops due to insect pests

In terms of monetary value, Indian agriculture suffers an annual loss of about USD 42.66 millions due to insect pests out of which USD ~18.22 million is caused by *Helicoverpa armigera* alone. In India, the losses due to *H. armigera* infestation are estimated to be 4.2 to 39.7% in pulses, chickpea alone being 29.2%; 25 to 79% in cotton, 18 to 26% in sorghum, 50% in sunflower, 15 to 46% in tomato and over 40% in okra

(Singh et al., 2014). The annual control costs and production losses due to *H. armigera* worldwide, amount to USD 5 billion (Lammers and McLeod, 2007). **Table 1.1** summarizes the estimated crop losses due to insect pests to major cash crops.

Table 1.1: Estimation of crop losses due to insect pests to major agricultural cash crops in India (adopted from Dhaliwal et al., 2010).

Crop	**Actual Production**	**Estimated loss in yield**		**Monetary value of estimated losses**
		%	**Total**	
	(million tonnes)		**(million tonnes)**	**(million INR)**
Cotton	44.03	30	18.9	339660
Rice	96.7	25	32.2	240138
Maize	19	20	4.8	29450
Sugarcane	348.2	20	87.1	70667
Groundnut	9.2	15	1.6	25165
Other oilseeds	14.7	15	2.6	35851
Pulses	14.8	15	2.6	43551
Coarse cereals	17.9	10	2	11933
Wheat	78.6	5	4.1	41368

In the five most important crops of the semi-arid tropics *viz.,* sorghum, pearl millet, pigeonpea, chickpea and groundnut, the biotic and abiotic stress factors have been estimated to cause loss of USD 15.74 billion (Sharma et al., 2001).

1.3. Helicoverpa armigera: A polyphagous insect pest

Helicoverpa armigera Hübner (Lepidoptera: Noctuidae) is a notorious insect pest worldwide. Lepidoptera is the second largest order in the class Insecta, which includes moths and butterflies. *H. armigera* infests about 300 plant species and the most important host crops include cotton, legumes (chickpea, pigeonpea), tomato, tobacco, okra, potato, sunflower, safflower, maize, groundnut (Fitt, 1989; Rajapakse and Walter, 2007). It also occupies diverse habitat in countries of Asia, Europe, Australia and Africa. The features of *H. armigera*, which make it a successful pest and threat to agriculture are, polyphagy, high reproductive rate, mobility and facultative diapause (Fitt, 1989).

H. armigera has a typical lepidopteran life-cycle spanning 25 to 35 days **(Fig. 1.1)**. The eggs hatch after 3 to 4 days of laying. The larval period is divided into six instars, which lasts for 12 to 16 days. The fourth and the fifth instar larvae are highly voracious and the most damaging ones (Tamhane et al., 2005a). A single larva eats up to 8 to 17 pods in its lifetime. The larvae are occasionally carnivorous and when the opportunity arises, cannibalistic. Pupa is a non-feeding stage, which lasts normally for 6 to 10 days. The pupa molts into a moth and remains alive for 4 to 5 days and feeds on nectar. Each female can lay several hundred eggs on lower surface of leaves, flowers and young pods, where newly hatched larvae feed preferentially (Tamhane et al., 2005a; Rajapakse and Walter, 2007). In case of unfavourable conditions, the larvae undergo facultative diapause i.e. a state of suppressed metabolism as pupae. There is a great variation in the colour of larval instars and the moth depending on the environment and the food on which it feeds upon.

Figure 1.1:

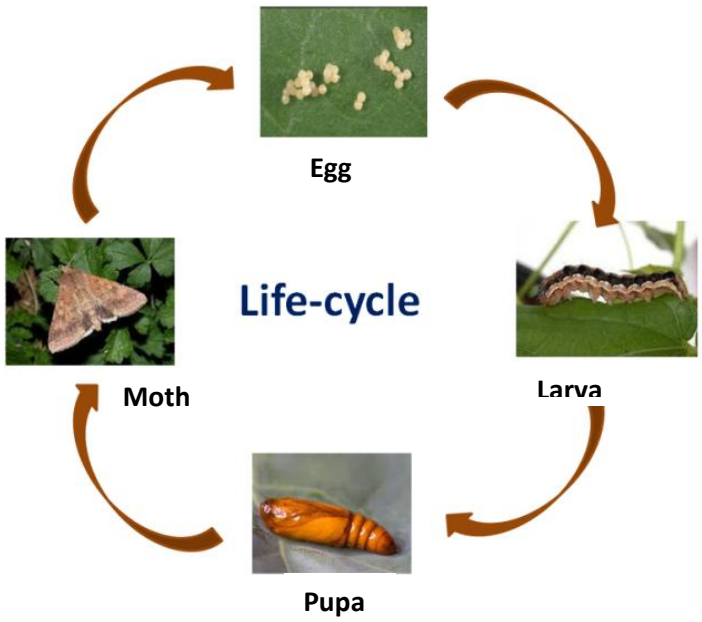


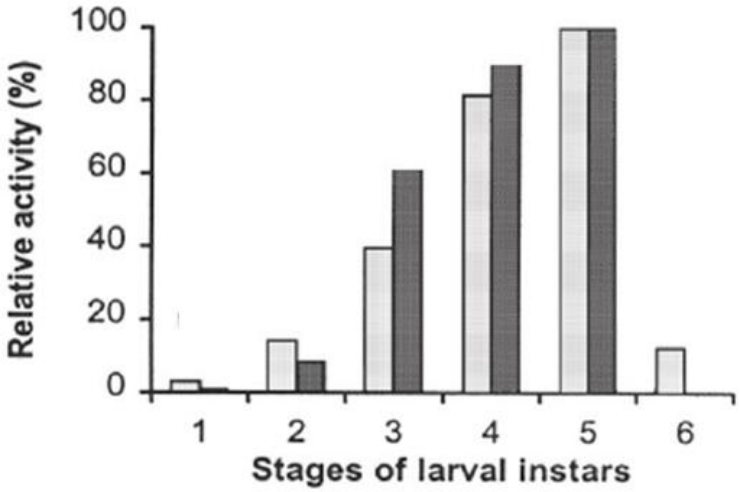
Figure 1.1: Stages in the life- cycle of *H. armigera*.

1.3.1. Adaptive physiology and biochemistry of H. armigera

The digestive system or gut of *H. armigera* is like constant flow-through system extracting nutrients and can be divided in to fore-, mid-, and hindgut. The dietary components are completely macerated by mandibles prior to its entering in the foregut. Most of the digestion and absorption of nutrients takes place in the midgut. In the hindgut water is removed and waste is filtered and is eventually excreted as frass or fecal matter. The gut is characterized by alkaline pH and presence of free Glycine (Johnston et al., 1991). The total digestive activity is constituted by trypsin (90%), chymotrypsin (5%), elastase (1%), carboxypeptidase (1%), aminopeptidase (1%), cathepsin B-like (1%) and metalloprotease (1%). Thus, serine proteases majorly contribute to the gut protease complement. Regulation of protease gene expression is dependent on the developmental stage as well as the diet composition **(Fig. 1.2A and B)** (Bown et al., 1997; and references therein).

Figure 1.2:

A



B

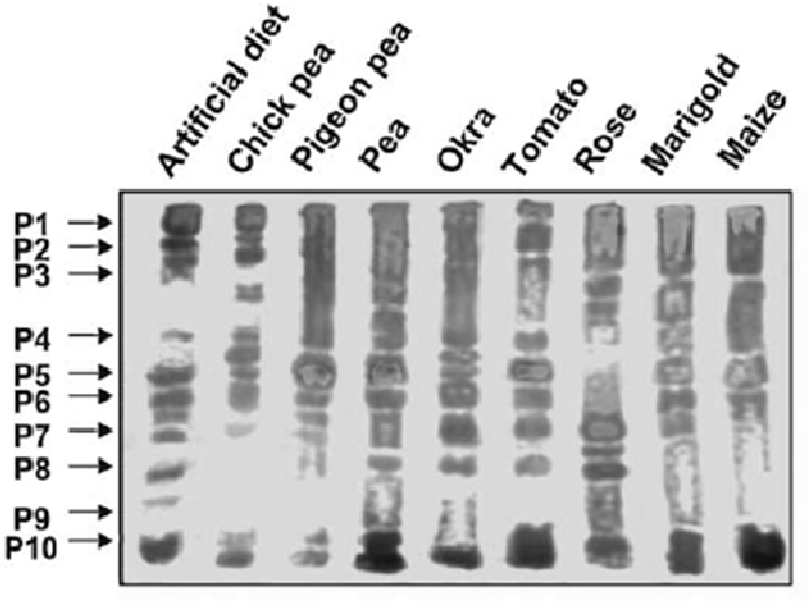


Figure 1.2: *H. armigera* gut proteinase activity profiles (A) during the stages of larval development (adopted from Patankar et al., 2001) (B) upon feeding of larvae on varied diets. P1-P10 indicate diverse isoforms (Adopted from Kotkar et al., 2009).

Polyphagy requires metabolic regulation in order to maximize the benefits from diverse diets like protein-rich plant reproductive structures or carbohydrate-rich leaves (Sarate et al., 2012). Digestive enzymes of *H. armigera* (amylases, proteases and lipases) are modulated on the basis of the composition of diet. Nutritional quality of the diet significantly affects larval and pupal mass gain, fertility and fecundity as well as life span. Broad flexibility in adapting to various host plants *viz.* legumes (chickpea and pigeon pea), vegetables (tomato and okra), flowers (rose and marigold), and cereals (sorghum and maize) has been observed in *H. armigera* (Kotkar et al., 2009; Sarate et al., 2012; Chikate et al., 2013). A dynamic system of proteinase gene expression enables the insects to utilize the plant proteins from various host plants efficiently, with minimal complement of proteinases and also avoids unnecessary expenditure of metabolic energy by reducing the expression of other protease genes (Chougule et al., 2005; Srinivasan et al., 2006).

1.3.3. Current strategies for H.armigera infestation control

Several strategies like cultural practices, chemical insecticides, biological control, biotechnological approaches and integrated pest management have been devised for controlling the *H. armigera* infestation in fields. Cultural/mechanical practices which include deep ploughing of soil, hand picking of larvae, shaking of plants, weeding, intercropping, use of pheromone traps, following a good time of sowing are laborious but environment-friendly and economical (Dahiya et al., 1999). Chemical/synthetic pesticides are being extensively applied for effectively controlling *H. armigera*. However, excessive and continuous usage has led to the development of resistance against many chemical insecticides including organochlorides, organophosphates, carbamates, pyrethroids etc. (Dawkar et al., 2013). Besides bioaccumulation and incorporation of chemical pesticides in the food chain exerts toxic effects on non-target organisms as well as human beings. Also, farmers often resort to insecticide mixtures to minimize the insect damage to crops. This not only increases the cost of pest control, but also results in insecticidal hazards and pollution of the environment (Sharma et al., 2000). Biological control methods explore the natural antagonistic interactions between organisms. Predators and parasitoids; biopesticides with live nematodes, fungi, bacteria, viruses or products derived from them and other plant products are utilized (Gurjar et al., 2010). For example, entomopathogenic fungus such as *Beauveria bassiana* and *Metarhizium anisopliae*; Nuclear Polyhedrosis Virus (NPV) and *Bacillus thuringiensis* (Bt) formulations are being employed. Neem products, vegetable oils, crude plant oils are other popular plant products. However, setting up such control systems requires proficient planning, supervision and might lack specificity.

The use of recombinant technologies allows a greater modification and a better predictability of trait expression than the conventional breeding methods. Biotechnological approaches are being employed to contain the pest damage, both in the developed and the developing countries. Attempts have been made to obtain insect resistance in plants by transferring heterologous genes from various sources (Ferry et al., 2006). Cry toxins from *B. thuringiensis* have been effectively used to control several field pests. *B. thuringiensis* insecticidal proteins are highly specific insect-gut poisons. Successful expression of *Bt* genes against the lepidopteran pests has been obtained in

cotton, maize, tomato, potato, brinjal, groundnut and chickpea. However, altered processing of protoxins or modification of the receptor site resulted in emergence of Bt resistance among the target pests (Frutos et al., 1999 and references therein). So far, resistance to Bt has not caused any major crop losses, however, emergence of resistance to Bt continues to be a threat to global food production. This has encouraged the attempts towards probing for alternative novel bio-molecules that can replace or be used in combination with Bt as part of resistance management strategies. Inhibitors of amylases and proteases of insects and lectins are few of the most studied plant defense molecules exhibiting potential while devising insect resistance strategies. Lectins disrupt the gut cell wall of insects by binding to glycoproteins of brush border cells and thus, the nutrient uptake is hampered. Alpha-amylase inhibitors (α-AIs) occur in many plants as a part of the natural defense mechanisms. They complex with the insect gut amylases and have demonstrated successful use in plant protection from insect damage (Franco et al., 2002). Initial observations about induction of protease inhibitors (PIs), upon insect damage and consequent reduction in insect fitness inspired researchers to explore the potential of PIs to limit herbivory from insect pests. In a set of elaborate laboratory and field experiments, two PIs from the structurally distinct potato type I and II families demonstrated the application of these molecules for crop protection. This accompanied with the understanding of mechanisms employed by the insects to overcome the negative effects of PIs may enable the design of PIs to avoid emergence of resistance (Dunse and Anderson, 2011). Also, proteinase inhibitors have been shown to be synergistic to Cry toxin in retarding larval growth (MacIntosh et al., 1990; Valueva and Mosolov, 2004).

Most of the PIs in plants are proteinaceous, competitive inhibitors. Following inhibition, the target proteases can no longer cleave peptide bonds, hampering dietary protein assimilation in herbivorous insects leading to significant delays in growth and development. PIs do not have a wipe out effect like synthetic pesticides and thus, do not impose strong selection pressure. Many transgenic plants have been developed using the PI genes (Schlüter et al., 2010). Recently several genomic approaches are underway in order to identify new functional genes or sites that can be effectively targeted for the control of lepidopteran insects. International Lepidopteran Genome Project (http://papilio.ab.a.u-tokyo.ac.jp/lep-genome/index.html) has been started to undertake

comparative genomics of economically and scientifically important lepidoptera like, *B. mori* and *H. armigera*.

1.4. Plant defense mechanisms

Plant defense against herbivory includes a range of adaptations evolved by plants in order to reduce the impact of herbivores and sustain survival and reproduction. To protect themselves from herbivore attack, plants use constitutive and induced defenses. Plants activate many anti-insect defense mechanisms in response to mechanical tissue damage and/or elicitors perceived upon insect attack. Following wounding, plants transiently produce reactive oxygen species (ROS), such as the superoxide anion, locally (i.e. in the damaged tissue) and H_2O_2 both locally and systemically (including parts away from damaged tissue). Responses like the upregulation of genes required for cell repair and response to osmotic stress, would likely occur as a result of either herbivory or mechanical wounding (Howe and Jander, 2008). Locally or systemically expressed wound-responsive proteins have known to play diverse roles (Ryan, 1990; Walling, 2000). Many plants produce secondary metabolites and proteins that can influence herbivore metabolism, growth and survival. Plant's indirect defenses such as green leaf volatiles, volatile organic compounds, and extra-floral nectars attract the natural enemies (such as parasitoids) of herbivores (Wu and Baldwin, 2010). One such mechanism relies on causing indigestion thereby disrupting the nutrient acquisition system in the infesting insects. Many plants produce proteins that specifically inhibit pest or pathogen digestive enzymes, for example, proteinase inhibitors, lectins, amylase inhibitors and defensins. Wound-induced PIs have been shown to enhance plant's resistance to insects by inhibiting the proteolytic enzymes of the attacking pest (Green and Ryan, 1972).

Jasmonates play a significant role in regulating defense responses to herbivores upon tissue damage. The jasmonate (JA) pathway has become known as the major signaling pathway that transforms information at the plant–insect interface into varied defense responses. Jasmonate inducible proteins, having role in direct defense include polyphenol oxidase (PPO), arginase, threonine deaminase (TD), leucine amino peptidase and a broad spectrum of PIs.

Figure 1.3:

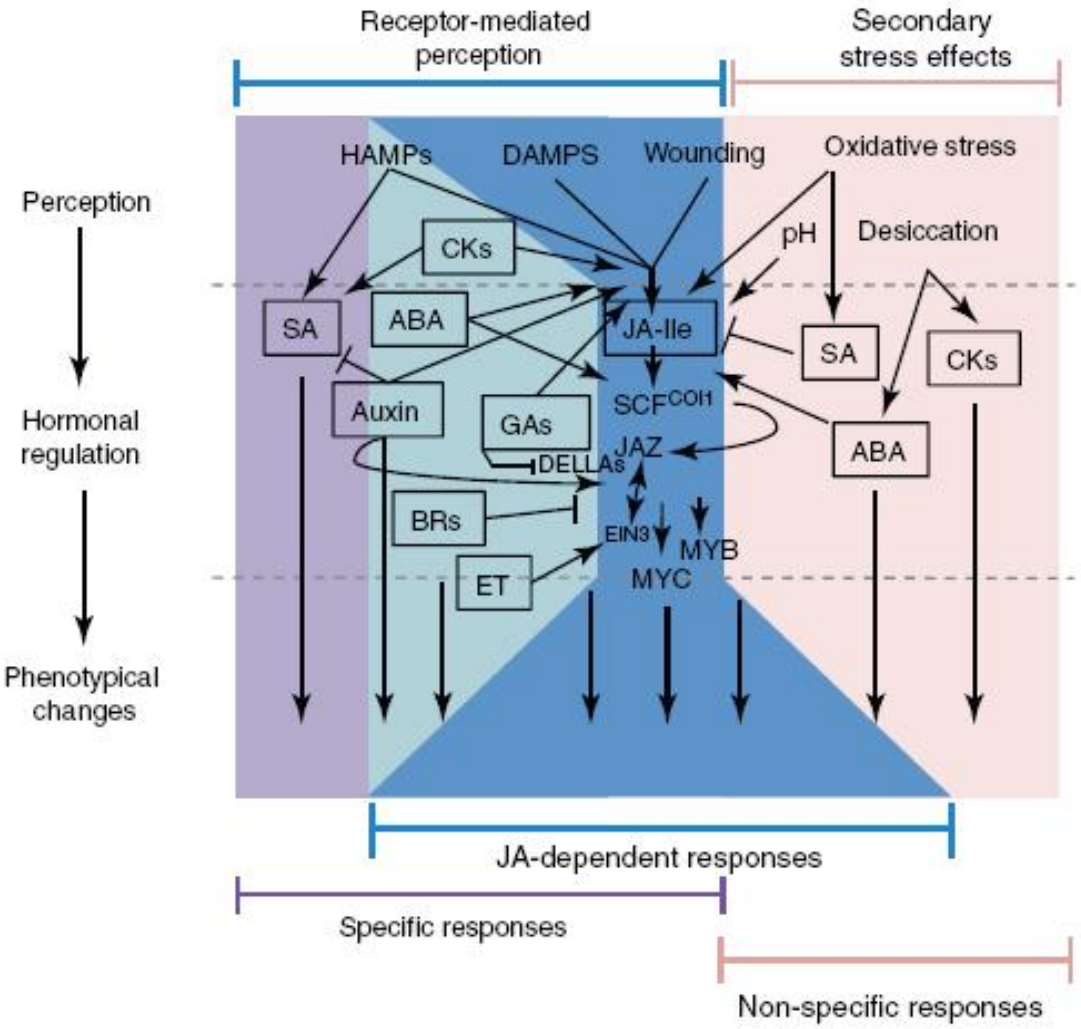


Figure 1.3: The jasmonate (JA) core pathway and its modulating factors. Herbivore-associated molecular patterns (HAMPs), damage-associated molecular patterns (DAMPs) and wounding, activate the JA pathway (blue area). Increased accumulation of jasmonoyl-L-isoleucine (JA-Ile) promotes the interaction of JAZ proteins with the SCF ubiquitin ligase SCFCOI1. Ubiquitin-dependent degradation of JAZs releases transcription factors from their JAZ-bound repressed state. This activates the expression of transcriptional regulons that promote defense and inhibit vegetative growth. JA-independent hormonal pathways are also induced (purple area), and several hormones, including salicylic acid (SA), ethylene (ET), auxin, gibberellins (GA), cytokinins (CK) and brassinosteroids (BR) modulate JA metabolism and signaling (light-blue area). Herbivory also leads to oxidative stress, changes in intracellular pH and desiccation, which modulate the JA pathway either directly or indirectly through other hormones. Together, this leads to complex phenotypic changes that comprise both specific and general responses. (Adopted from Erb et al., 2012)

Jasmonates are also involved in the regulation of tritrophic interactions; host plant resistance to phloem-feeding insects, trichome-based defenses, pathogen resistance and systemic transmission of defense signals. In addition, jasmonates promote defensive and reproductive processes inhibiting the growth and photosynthetic output. Although chewing herbivores are believed to activate JA responsive genes, the expression of defense genes is modulated by the crosstalk of three signal pathways, namely JA, ethylene, and salicylic acid (SA) **(Fig. 1.3)**. The elicitation of multiple pathways is likely to tailor plant responses adaptively to the diverse herbivore species. By eliciting signal molecules from the three pathways to different degrees, a plant tailors its defensive responses to a specific attacker (Reymond and Farmer, 1998).

1.5. Plant protease inhibitors

Protease inhibitor is one of the most studied classes of plant defense proteins. PIs are abundantly present in leaves, flowers, seeds and tubers and are induced/ upregulated in response to herbivore attack (Green and Ryan, 1972). PIs function by blocking the digestive proteases in the larval gut thereby limiting the release of amino acids from dietary proteins (Hilder and Boulter, 1999; Tamhane et al., 2005a), reducing overall fitness of the insect and thus reduce its impact on the plant. They also interfere with proteolytic activation of enzymes as well as molting and water-balance of the insects (Boulter, 1993). The digestive proteolytic machinery is also linked with adaptation to newer hosts and resistance to antagonistic agents. PIs from various families of plants have been found to have anti-nutritional effects on insects. PIs do not exert strong selection pressure on insects since they do not cause mass death of insect populations (Srinivasan et al., 2006).

1.5.1. Plant protease inhibitor families

Several families of PIs have been classified depending on specificity towards target proteases, molecular mass and structure. A database for plant PIs was developed by De Leo et al., (2002) correlating information about primary sequence and functional analysis of 351 PIs (http://bighost.area.ba.cnr.it/PLANT-PIs). Serpin (Serine PI) family, Bowman Birk inhibitors (BBIs) family, Kunitz family, Squash inhibitors, Cereal

trypsin/α-amylase inhibitors, Mustard (Sinapis) trypsin inhibitor (MSI), Potato type I PIs (PI 1), Potato type II PIs (PI 2), Cysteine PIs (CYS), Metallocarboxypeptidase inhibitors are the families described. Among these, inhibitors from four families belong to inducible serine PIs- Potato type I and II, Bowman Birk (BBIs) and Kunitz families (De Leo et al., 2002). Perpetual addition to the PI sequence databases has led to revising the approaches used to classify and organize them. Based on sequence homologies of inhibitor domains, PIs have been classified into 48 families in the MEROPS database (http://merops.sanger.ac.uk) (Rawlings et al., 2004). Rawlings et al., (2004) pointed a few fundamental features of PIs- (1) an effective inhibitor unit may contain as few as 14 amino acid residues, (2) reactive-site residues in inhibitor units are often not conserved in the way that active-site residues are in peptidases and (3) inhibitor sequence units in several families have duplicated during evolution giving rise to multiple divergent copies. Present work is specifically focused on Potato inhibitor type II (Pin-II) from *Capsicum annuum* L.

1.5.2. Potato inhibitor type II (Pin-II): Role in endogenous and defense functions

Pin-II or Pot-II family of serine protease inhibitors has been explored at gene, protein and functional level. The Pin-II PIs, predominant in Solanaceae, have a unique single or multiple inhibitory repeat domains (IRDs) with sequence and structure variations. They show wound induced up-regulation and expression, post-translational interactions with proteases leading to modification in PI protein structure, activity and function. Together all these features make them an intriguing subject area for plant and insect biologists. The Pin-II PIs have been classified under I20 ('I' stands for inhibitor and '20' denotes the serial number) in the MEROPS database (Rawlings et al., 2004).

Plant organs that express Pin-II protein include leaves, flowers, fruits, stem, tubers and roots (Tamhane et al., 2012). Wound induction of Pin-II PIs and their role in herbivore defense was correlated by the pioneering work of Green and Ryan, (1972), which later led to the discovery of several types of PIs and their activities to retard growth and development of insects. Leaves from wounded tomato plants have been shown to accumulate >200 μg of potato inhibitors I and II per g of leaf tissue and to

severely reduce the growth of larvae of *Spodoptera exigua*, the beet armyworm (Jongsma et al., 1995).

There are some evidences that Pin-II PIs can regulate cell proteolysis by their action on endogenous proteinases, thereby controlling protein turnover and metabolism (Horn et al., 2005; Sin and Chye, 2004). They are believed to play role in several endogenous processes in cells. *Solanum americanum* has two well characterized PIs SaPin-IIa and SaPin-IIb. Abundant expression of SaPin-IIa in stems especially in companion cells (CC) and sieve elements (SE) of phloem suggested their role in regulating proteolysis in SE as well as in phloem development (Xu et al., 2001; 2004). SaPin-II a and b are strongly expressed in the floral buds that are destined to undergo developmental programmed cell death (PCD) (Sin and Chye, 2004). Expression profiles of SaPin-II a and b, suggest their differential regulation and probably overlapping and complementary roles in floral development. They probably function by confining the PCD to the specific tissues, thereby protecting the adjacent tissues (Peña Cortés et al., 1991; Sin and Chye, 2004). The PIs in developing seeds of *S. americanum* play a role in protection of the endosperm and embryo by regulating proteinases generated within the seed (Sin et al., 2006). SaPin-IIb was also found to play a role in trichome-based defense by functioning as a constitutive component of trichome chemical defense and/or by regulating the development of glandular trichomes (Liu et al., 2006; Luo et al., 2009).

Sin et al., (2006) have reported an increase in flower size and an 80% seed abortion after silencing homologs of *SnSPI2a* and *SnSPI2b* in *S. americanum*. However, in case of *S. nigrum*, no effect on flower size was detected and only 0.7 to 2.8% of the seeds were aborted or defective upon silencing SPI2a and SPI2b (Hartl et al., 2010; 2011). Four Pin-II PIs corresponding to their respective activity isoforms namely *SnSPI1*, *SnSPI2a*, *SnSPI2b and SnSPI2c* were identified in *S. nigrum*. The *SnSPI2a*, *SnSPI2b* and *SnSPI1* were found to be strong subtilisin-inhibitors, where as *SnSPI2c* was identified as a strong inhibitor of trypsin and chymotrypsin (Hartl et al., 2010). Hartl et al., (2010) have shown that the PIs from *S. nigrum* exhibit a certain degree of functional differentiation but also considerable functional overlap. The highly abundant PI (*SnSPI2c)* displays typical characteristics of a defense related gene where as the other two, *SnSPI2a* and *SnSPI2b*, show an overlap of defensive and developmental roles.

In *Nicotiana attenuata*, silencing the PI gene abolishes the plant's capacity to produce PIs and allows it to grow faster, flower earlier, and produce more seed capsules compared with PI-producing genotypes (Zavala et al., 2004). Bezzi et al., (2010) provide evidence for a role of PIs in the processing and secretion of nectar proteins, which, in turn results in higher levels of nectar H_2O_2. Thus, Pin-II PIs assume a significant role in endogenous functions, namely regulation of proteolysis, macromolecular trafficking, PCD and consequently aid the plant growth and development in respective tissue.

A qualitative as well as quantitative analysis of endogenous PI activity in different tissues of field grown tomato and *Capsicum annuum* plants shows substantially high activity in flowers. This observation suggests a survival strategy by the plant, diverting defense molecules to the reproductive organ i.e. the flower (Damle et al., 2005; Tamhane et al., 2005b). In order to investigate the *in vitro* activity of tomato PIs, Damle et al., (2005) tested the inhibitory potential of these PIs against gut proteases from chickpea-fed and tomato-fed *H. armigera* insects. In both the cases, the gut protease activity was inhibited to 60 and 90%, respectively and in feeding experiments, these PIs showed adverse effects on *H. armigera* development in a dose dependent manner. The feeding of *C. annuum* leaf extracts to *H. armigera* larvae led to retarded larval growth and development, delay in pupation period and reduction in fecundity and fertility (Tamhane et al., 2005b). Significance and utilization of the defense related property of Pin-II PIs has been further discussed in section *1.5.4.*

1.5.3. Structure and mechanism of inhibition

The gene structure within Pin-II family is conserved. It consists of an exon encoding the N-terminus of the signal peptide followed by second main exon encoding the C-terminus of the signal peptide and variable number of IRDs. The exons are always separated by an intron of 100-200 bp **(Fig. 1.4)** (Barta et al., 2002).

Figure 1.4:

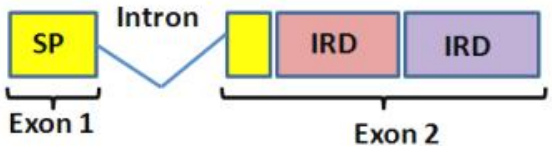


Figure 1.4: The consensus gene structure with organization of the Pin-II PI protein domains (SP: Signal peptide, IRD: Inhibitory Repeat Domain).

The conserved Pin-II PI protein consists of an endoplasmic reticulum signal peptide of 25 amino acid (aa) followed by variable number of IRDs of ~55 aa. The IRDs are separated by 5 aa linker regions. In some Pin-II PIs, a vacuolar sorting signal is present at the extreme 3′ region. The partial IRDs at the N- and C- terminal of the Pin-II PIs form covalent bonds to generate a functional IRD as one. The striking feature of Pin-II PIs is the presence of variable number of IRDs, structurally forming multi-domain proteins. The number of IRDs varies from 1- to 8-IRDs amongst different members of Solanaceae. Unequal crossing over and duplication of the inhibitory domain sequence is presumed to be responsible for the expansion of the repeated domains. The presence of multiple IRDs in many Pin-II PIs has suggested the functional significance of the combination of IRDs within a single PI (Barta et al., 2002; Christeller, 2005). The sequence of linker regions of 5 aa residues between IRDs is almost conserved in *Nicotiana* sp. (EEKKN), whereas linker regions in Pin-II PIs of other genera it is different though it functions similarly. The cysteines are involved in formation of four disulphide bonds, which stabilize the repeat structure **(Fig.1.5A)**.

The sequence of IRDs is highly variable; however, presence of eight cysteines, a single proline residue and an active site either for trypsin or chymotrypsin inhibition is conserved throughout IRDs. Single domain PIs, formed by proteolytic cleavage at the linker regions of the multi-domain precursor, have either trypsin inhibitory (TI) or chymotrypsin inhibitory (CI) sites (Heath et al., 1995; Lee et al., 1999).

Figure 1.5:

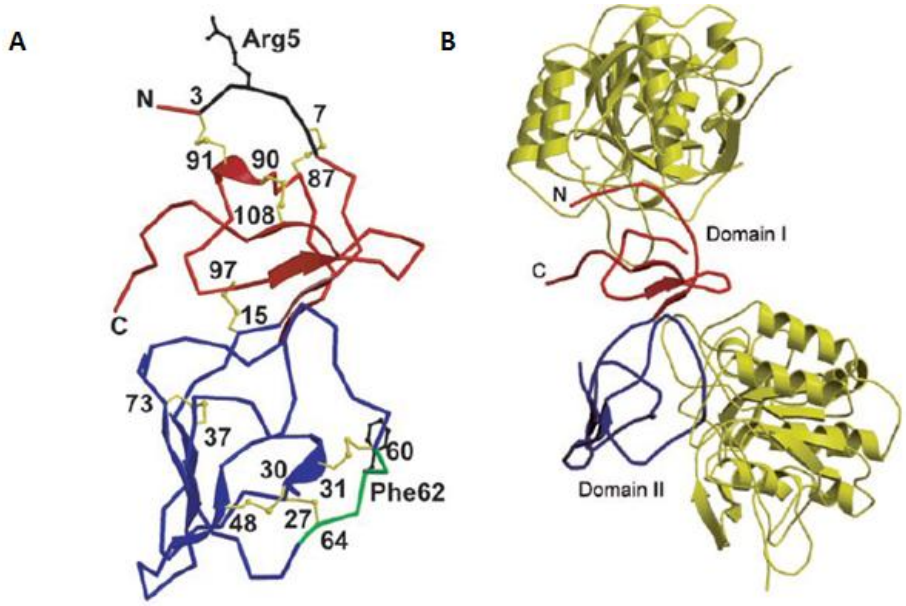


Figure 1.5: Crystallographic structure (A) ribbon diagram of unbound TI-II. IRD1 is shown in red, whereas IRD 2 is shown in blue. Reactive site loops 1 and 2 are shown in black and green, respectively. The side chain of the P1 residue in each reactive site loop is drawn in black. Disulfide bonds are drawn in yellow and residue numbers are given in black. (B) Structure of the TIII (subtilisin)2 complex. Subtilisin molecules are drawn in yellow, and the two IRDs of TI-II are drawn in red and blue. Simultaneous binding of two protease molecules to the 2 IRD PI is evident. (Adopted from Barette-Ng et al., 2003a; 2003b)

The conserved presence of two disulfide bonds anchoring the reactive site loop, among all known Pin-II PIs indicates their significance in stabilizing the loop. Other than these, there are few more highly conserved residues, which are structurally important, such as Pro-18, Gly-38 and Gly-46 as they belong to the three β-turns, respectively (Kong and Ranganathan, 2008).

Single IRD of Pin-II PIs can bind to a single protease, while two domain PIs of tomato and potato can simultaneously inhibit two protease molecules. The P1 residue of the reactive site, which reacts with protease active site, determines its specificity. Presence of lysine ‘K’, arginine ‘R’ or Leucine ‘L’ in the P1 position confers the inhibitor with either trypsin (K or R) or chymotrypsin (L) inhibitory potential. P3 to P2’ a stretch of 5 aa close to the reactive site is important in determining enzyme specificity of the inhibitor. The reactive site loop (RSL) P4 to P3’ of the inhibitor domain interacts with S6

to S2' of protease pocket to bring about its inhibition by mimicking a substrate (Barrette-Ng et al., 2003a). The structural basis of inhibition of a multi-domain Pin-II inhibitor has been shown by its ternary complex with two subtilisin Carlsberg molecules which revealed how it can bind to and simultaneously inhibit two enzyme molecules within a single ternary complex (Barette-Ng et al., 2003b). The inhibitory reactive site loop in each IRD is positioned at opposite ends of the elongated molecule facilitating inhibition of two protease molecules **(Fig.1.5B)**.

For multi-IRD Pin-II protein, there are two possible domain organizations: (1) tandem repeat domain organization where domains are arranged in beads-on-a-string way; or (2) circularly permuted domain organization which is formed by the association between two terminal half-repeats to form a PI domain (Lee et al., 1999; Schirra and Craik, 2005). The three-dimensional structure of several Pin-II PIs, single- as well as multi-domain, have been determined either by X-ray crystallography or NMR provide good outline of the structure and dynamics of this class. The 43-kDa precursor PI of *Nicotiana alata* NaProPI forms a circular 'clasped bracelet' conformation as a result of formation of disulfide bridges between the partial repeat regions at the N and C terminal of the precursor (Scanlon et al., 1999). It has been demonstrated that the aa sequence repeat in NaProPI does not correlate with the structural repeat in this 43-kDa inhibitor. An active two-chain domain (C2) is formed by joining two partial domains in addition to five single chain domains (TI1-TI4 and CI1) **(Fig.1.6)**.

In various members of Pin-II precursors studied there is a combination of trypsin/ chymotrypsin inhibitory domains. For example the six domain *N. alata* PI (NaProPI) possesses four TI domains and two CI domains. However, the NaProPI with 6 IRDs cannot bind to six proteases simultaneously because of stearic interference. This six IRD PI molecule of *N. alata* could inhibit maximum of four chymotrypsin or 2.6 trypsin molecules. In order to realize total inhibition potential, individual IRDs must be released from the precursor. It is thus important to have a proteolytically processed precursor for maximum protease inhibition (Heath et al., 1995).

Figure 1.6:

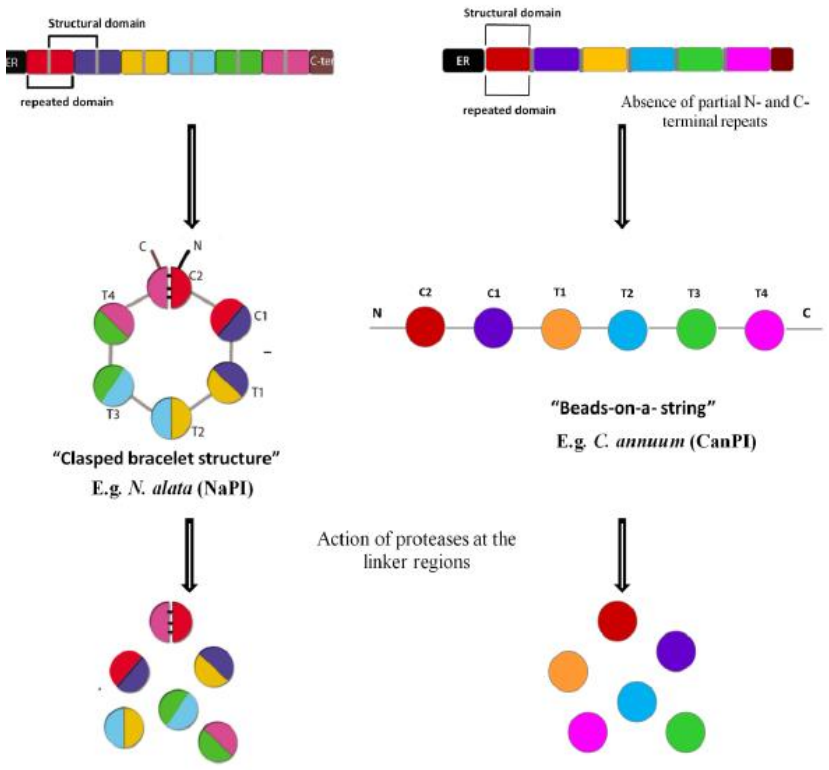


Figure 1.6: Models of Pin-II PI precursors. Structure and processing of the 6 domainNa-proPI. The 40.3 kDa PI precursor consists of 6 IRDs preceded by a signal sequence and terminated by a vacuolar targeting signal. C2 is formed by joining of N- and C- terminal partial ends. In case of lack of these partial ends, beads-on-a-string like structure is expected to be assumed by the precursor. 'T' and 'C' indicate trypsin and chymotrypsin IRDs respectively (Modified from Dunse and Anderson, ISB news report, June 2011).

1.5.4. Insect resistant transgenic plants expressing PIs

Elaborate experimental approaches were used by Heath et al., (1997) to explore the inhibitory potential of a 5-domain (4 TIs and 1 CI) *N. alata* PI (NaPI) towards *Helicoverpa punctigera* and *Teleogryllus commodus*. Firstly, the *in vitro* protease inhibitory activity of the individual and pooled TI and CI domains was determined, which ranged from 37 to 79%. Secondly, on feeding *H. punctigera* larvae on artificial diet incorporated with 0.26% (w/w) NaPI, delayed development and reduced larval mass. Thirdly, transgenic *Nicotiana tabacum* plants expressing NaPI (0.2% of soluble protein) were developed. However, the effect on growth in *H. punctigera* larvae fed on such

leaves varied among the colonies of the larvae. In another study, *Manduca sexta* larvae fed on transgenic tobacco plants expressing inhibitor II proteins from the tomato and potato showed severe inhibition at 50 µg inhibitor/g of tissue while more inhibition and mortality was noted at 100 µg/g tissue (Johnson et al., 1989). Interestingly, greater insecticidal effect was observed in tobacco plants transformed with the genomic sequence of the tomato PI-II than in those transformed with the cDNA sequence. This indicated that the presence of intron responsible for its enhanced expression and appropriate splicing of exogenous sequences in the transgenic plants to obtain the active protein (Zhang et al., 2004). The defensive effect of endogenous trypsin proteinase inhibitors (NaTPIs) on the herbivore *M. sexta* was demonstrated by genetically altering NaTPI production in host plant, *N. attenuata* (Zavala et al., 2008). The effects of NaTPI on *M. sexta* gut proteinase activity in different larval instars were measured by feeding larvae freely on *N. attenuata* plants where NaTPI expression was manipulated. Second and third instars larvae that fed on NaTPI-producing (wild type) genotypes were lighter and had less gut proteinase activity compared to those that fed on genotypes with either little or no NaTPI activity. It was found that the larvae can minimize the effects of high NaTPI levels by feeding on leaves with high protein and low NaTPI activity. However, the host plant's endogenous NaTPIs remain an effective defense against *M. sexta*, inhibiting gut proteinase and affecting larval performance (Zavala et al., 2008)

Dunse et al., (2010) explored the consequences of feeding *H. Punctigera* and *H. armigera* with Pin-I and -II inhibitor proteins. *H. punctigera* larvae were fed with a cotton leaf-based artificial diet, containing *N. alata* proteinase inhibitor (NaPI) which is a Pin-II type inhibitor. They detected a higher mortality (80%) as well as lower larval mass gain in NaPI-fed larvae as compared to those fed with the control diet (i.e. 40% mortality). Interestingly, the consumption of NaPI by the larvae, led to the induction of a chymotrypsin, which was found to be resistant to inhibition by NaPI. However, the activity of this chymotrypsin was found to be inhibited by a *S. tuberosum* Pin-I inhibitor (StPin1A) isolated from wound-induced leaves. *H. armigera* larvae fed with diets containing NaPI and StPin1A were reported to weigh less than the larvae fed with control diet by 50% and 40%, respectively; while 90% smaller larvae were observed when fed with an artificial diet composed of both the types of inhibitor proteins. Further, they

conducted field trials with transgenic cotton plants, expressing the individual inhibitors as well as both types of inhibitors (NaPI-StPin1A). These plants were artificially subjected to *H. armigera* infestation along with the natural prevalence of *H. punctigera* at the field site. They recorded a statistically significant increase in number of cotton bolls in the transgenic line expressing both StPin1A and NaPI than the parental line (untransformed) and also a boost in lint weight per plant for these transgenic lines, when evaluated against the control line.

Pin-II PIs of potato have also been used in transgenic rice and wheat plants to control biotic infestations by *Sesmia inferens* (rice) (Duan et al., 1996) and *Heterodera avenae* a nematode (wheat) (Vishnudasan et al., 2005).

1.6. Capsicum annuum: Occurrence and diversity in Pin-II genes

Capsicum annuum ("Chili pepper") is a domesticated species of genus *Capsicum* from Solanaceae family. *C. annuum* has the lowest number of chromosomes amongst different members of Solanaceae i.e. (2X = 24). It is one of the non-preferred hosts of *H. armigera*. Sixty seven *Capsicum annuum* PI (*CanPI)* genes were identified which showed homology to Pin-II family and constituted 1, 2, 3, or 4-IRD PIs (Shin et al., 2001; Kim et al., 2001; Tamhane et al., 2009; Mishra, 2013 doctoral dissertation). Among these 67 *CanPI* genes, 11 contained 4-IRDs, 37 contained 3-IRDs, 16 contained 2-IRDs and 3 contained a single IRD thus contributing to the PI diversity **(Fig. 1.7)**. Sequence analysis revealed highly homologous *CanPIs* with an average variance of 4%. The diversity in CanPIs can be attributed to individual IRDs, which displayed a sequence variation ranging from 2 to 25% within the vicinity of the reactive site loops and C-terminal region. Fifty-five unique IRDs, constituting 11 CIs and 44 TIs, followed the H-L type topology, where the sequence repeat was identical to the structural repeat (Kong and Ranganthan, 2008). CanPIs characterized so far do not possess the N- and C- terminal partial IRDs, hence are expected to assume the beads-on-a-string type structure (Tamhane et al., 2009). Active site variants of TIs 'CPRNC', 'CPKNC', 'CPRYC' and 'CPRDC' and two types of CI sites 'CTLNC' and 'CTPNC' were present among all the identified 55 IRDs.

Figure 1.7:

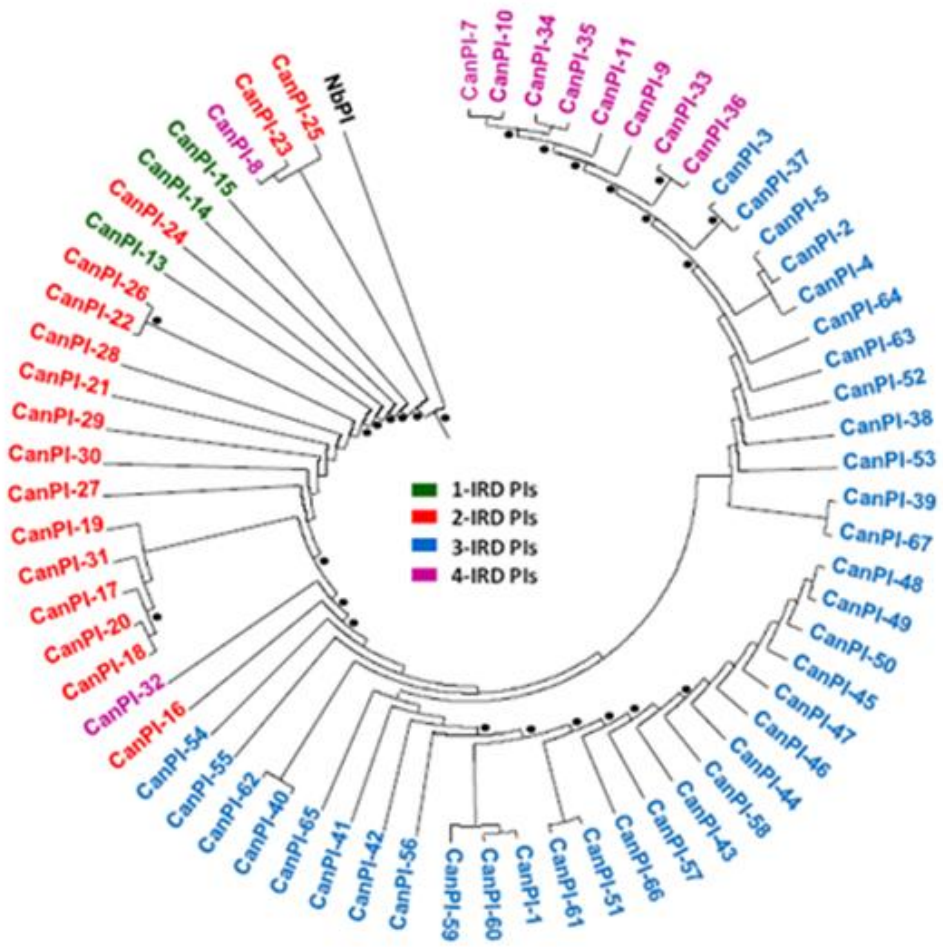


Figure 1.7: Dendrogram of CanPIs based on deduced amino acid sequences of full length genes isolated from *C. annuum*. Pin-II type PI from *Nicotiana benthamiana* (NbPI, NCBI: ABA42892) is used as an out-group. "•"denotes bootstrap value ≥80 as calculated by Lasergene software. (Adopted from Mishra, 2013, doctoral dissertation)

Expression patterns of CanPIs in the different tissues vary qualitatively, quantitatively as well as spatially and temporally. The flower tissue has significantly higher level of PI activity compared to the leaf, stem and fruit tissues (Tamhane et al., 2009). Several novel and diverse Pin-II PIs having 1- to 4-IRDs were isolated from developing fruit and stem tissues of *C. annuum.* Though all the four IRD forms were represented in both the tissues, stem tissue showed higher proportion of expression of 1- and 2-IRD CanPIs while fruit tissue showed higher expression of 3- and 4-IRD *CanPIs* (Tamhane et al., 2009). Significantly high expression levels of CanPIs observed upon natural infestation by aphids and lepidopteran insects, indicated the involvement of

CanPIs in plant defense (Tamhane et al., 2009). Wounding and biotic (virus, aphid and lepidopteran insect) stress to the plant tissues induced variable CanPI profiles (Mishra et al., 2012). Recent reports on the endogenous and/or defensive roles of PIs from various solanaceous species and simultaneous expression of multiple CanPIs highlights their prospective involvement in many of the plant's complex processes (Sin and Chye, 2004; Xu et al., 2004; Wu et al., 2006; Johnson et al., 2007; Tamhane et al., 2009; Hartl et al., 2011).

1.6. Antibiosis of CanPIs towards H. armigera

The efficacy of CanPIs against *H. armigera* gut proteases as well as larval growth and development was demonstrated by Tamhane et al., (2005b) **(Fig. 1.8A)**. *In vitro* assays showed that most of the trypsin-like activity of the *H. armigera* gut protease isoforms was sensitive to inhibition (upto 68 to 91%) by CanPIs. CanPIs inhibited more than 60% total proteolytic activity of larvae fed on cotton and chickpea plants. The feeding of *C. annuum* leaf extracts and purified PIs in varied doses to *H. armigera* larvae for two successive generations demonstrated retarded larval growth and development. Delayed pupation and reduction in fecundity and fertility were also observed. Further exploiting the interaction(s) of recombinant CanPIs with *H. armigera* gut proteases by Intensity Fading Matrix Assisted Lased Desorption/ Ionization Time of Flight (IF-MALDI-TOF) analysis, Mishra et al., (2010) revealed PI processing patterns and the stability of recombinant CanPIs in presence of gut proteases of *H. armigera*. *In silico*, *in vitro* and *in vivo* studies with individual IRDs (IRD-7, -9 and -12) of *C. annuum* not only demonstrated their efficacy in inhibition of insect proteases but also the effect of sequence variation on inhibition potential was evident (Joshi et al., 2014a; 2014b).

Figure 1.8:

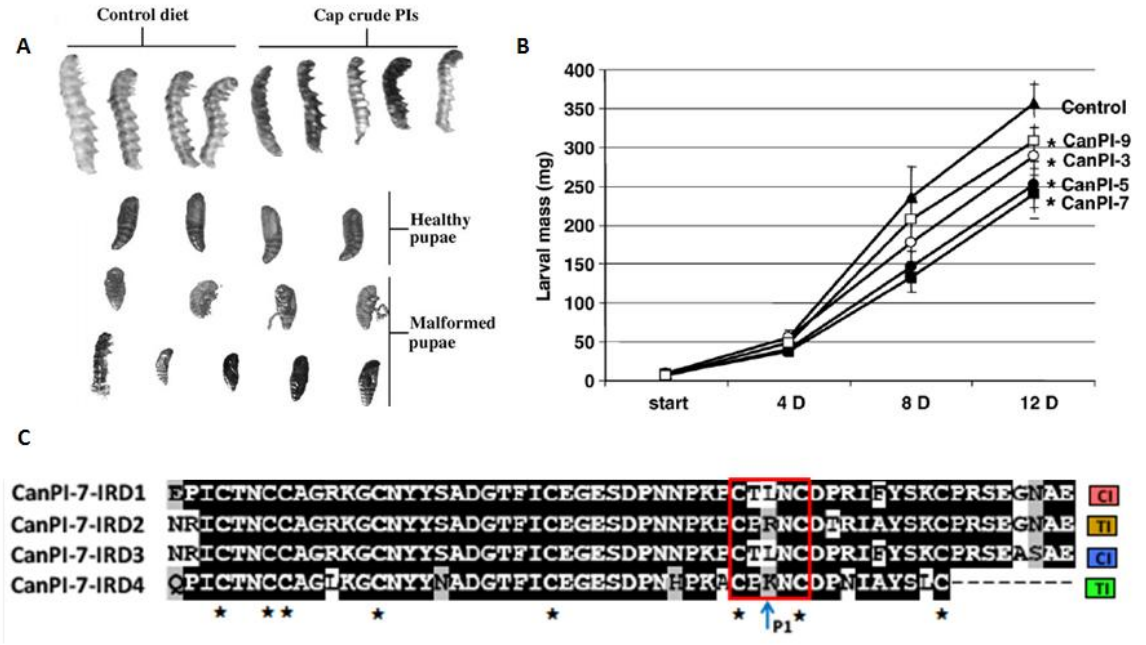


Figure 1.8: (A) The effect of *C. annuum* PIs on the growth and development of *H. armigera* larvae and pupae (Tamhane et al., 2005). (B) Reduced larval weight gain in the larvae fed on recombinant CanPIs (Tamhane et al., 2007). (C) The multiple sequence alignment of deduced aa sequences of CanPI-7 IRDs. Active site is marked by a red box and the P1 residue is marked with a blue arrow.

1.6.1 Merits of rCanPI-7

Diverse CanPIs with 1- to 4-IRDs were expressed heterologously in *Pichia pastoris* and the recombinant proteins were characterized for their insect inhibitory potential by Tamhane et al., (2007). *H. armigera* larvae fed on rCanPI diet showed 30% mortality and 40% lower weight among the survivors, in the early instars. Pupal mass reduction of 12 to 25% was recorded, leading to decreased fecundity. The 4-IRD rCanPI-7 with two chymotrypsin inhibitory sites and two trypsin inhibitory sites showed the strongest anti-metabolic effect on *H. armigera* among the selected rCanPIs **(Fig. 1.8B and C)**. The stoichiometry of rCanPI-7 with protease molecules was analyzed by molecular docking. The static complex of CanPI-7 representing at least three protease molecules in close proximity to the RSLs also indicates its higher potency (Mishra et al., 2013).

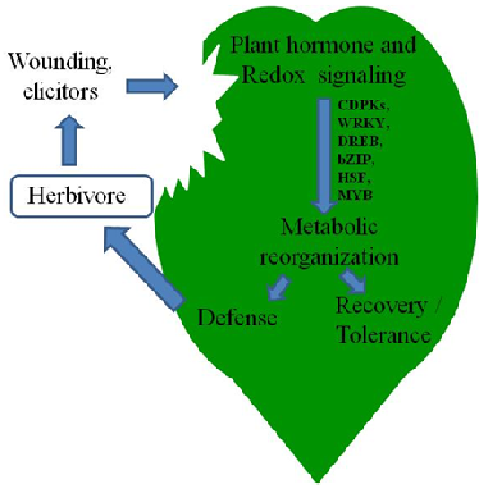
Genesis of thesis

Lepidopteran insect pests are responsible for causing severe losses in several crops. Pin-II PIs are one of the natural plant defense proteins that are being explored for their inhibitory potential towards these pests. Successful development of transgenic plants fortified with Pin-II PIs has encouraged assessment of plant and pest interaction on a case-by-case basis. The plant and insect systems considered here are *Capsicum annuum* and *Helicoverpa armigera*, respectively.

C. annuum, a solanaceous plant, is non-preferred host for *H. armigera*. Early observations showing reduced fitness of *H. armigera* larvae upon ingestion of *C. annuum* PIs (CanPIs) inspired us to further investigate this interaction. Study of the induced PI diversity, in *C. annuum*, revealed an array of different PI genes formed as a result of combinations of various unique inhibitory repeat domains (IRDs). Characterization of induced proteins and newly identified PIs in *C. annuum* was undertaken. Natural occurrence of these diverse PI sequences was intriguing; we therefore initiated the investigation of genetic elements like promoter, intron and UTRs of CanPIs.

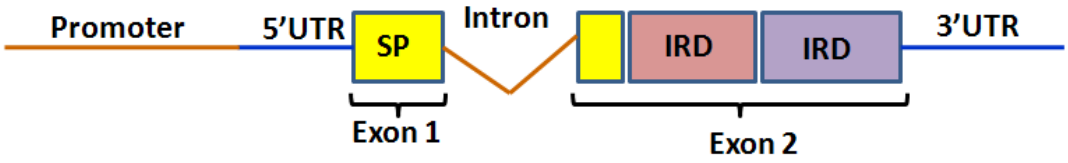
On the other hand, the molecular response of the target insect should be thoroughly evaluated for efficient design of PI strategy. Carefully planned strategies keeping in view the dynamic nature of insect gut proteases, can surely lead to robust solutions. Hence, the analysis of molecular response of *H. armigera* larvae upon ingestion of rCanPI-7 was undertaken.

To sum-up, mystifying structural and functional diversity amongst these CanPIs and the regulation of gut protease complement by *H. armigera* in response to these PIs are the issues to be addressed. Understanding these aspects will provide valuable knowledge for the development of PIs as a viable transgenic plant protection technology.



Chapter 2

Characterization of induced proteins and newly identified PIs in *Capsicum annuum*



The research work described in Chapter 2A is part of full-length paper, which has been published in Plant Physiology and Biochemistry, 2014, 74, 212-217.

The research work described in Chapter 2B is part of full-length paper, which has been accepted for publication in Planta, 2014, (In Press).

Chapter 2: Characterization of induced proteins and newly identified PIs in *Capsicum annuum*

Chapter 2A

Stress inducible proteomic changes in *Capsicum annuum* leaves

2A.1 Introduction

2A.2 Materials and Methods

2A.3 Results

- 2A.3.1 Proteomic changes in induced *C. annuum* leaves
- 2A.3.2 Semi-quantitative expression analysis of candidate genes

2A.4 Discussion

2A.5 Conclusion

Chapter 2B

Structural features of diverse Pin-II proteinase inhibitor genes from *Capsicum annuum*

2B.1 Introduction

2B.2 Materials and Methods

2B.3 Results

- 2B.3.1 Occurrence of Clasped bracelet type structural organization in *CanPI*s
- 2B.3.2 Gene architecture revealed several conserved characteristics of *CanPI*s
- 2B.3.3 *CanPI* genes represent unique 3'UTR sequences
- 2B.3.4 *CanPI* promoters are conserved and responsive to various stimuli
- 2B.3.5 Activity of *CanPI* promoters
- 2B.3.6 Structural attributes of clasped bracelet type CanPIs

2B.4 Discussion

2B.5 Conclusion

Chapter 2A: Stress inducible proteomic changes in *Capsicum annuum* leaves

Abstract

Herbivore attack induces defense responses in plants, activating several signaling cascades. As a result, molecules deterrent to the herbivores are produced and accumulated in plants. Expression of defense mechanism/traits requires reorganization of the plant metabolism, redirecting the resources otherwise meant for growth. In the present work, protein profile of *Capsicum annuum* leaves was examined after herbivore attack/induction. Majority of proteins identified as differentially accumulated, were having roles in redox metabolism and photosynthesis. For example, superoxide dismutase and NADP oxidoreductase were upregulated by 10- and 6-fold while carbonic anhydrase and fructose-1,6-bisphosphatase were downregulated by 9- and 4-fold, respectively. Also, superoxide dismutase, NADPH quinone oxidoreductase and NADP dependent isocitrate dehydrogenase transcripts showed a higher accumulation in induced leaf tissues at early time points. In general, proteins having role in defense and damage repair were upregulated while those involved in photosynthesis appeared downregulated. Thus metabolic reconfiguration to balance defense and tolerance was evident in the stress-induced leaves.

2A.1. Introduction

To sustain and survive in nature, plants adopt strategies to combat varied biotic and abiotic stresses. Pathogenic microorganisms and herbivorous insects are primarily the main types of biotic stress encountered by plants. Various microbial/pathogen associated molecular patterns (MAMPs or PAMPs) or herbivore attack associated molecular patterns (HAMPs) enable the perception of the challenge posed by these attacks (Wu and Baldwin, 2010). HAMPs such as glucose oxidase, β-glucosidase, inceptin, fatty acid-amino acid conjugates (FACs), caeliferins and oviposition fluid are known to elicit plant defenses (Reymond and Farmer, 1998). Plant hormones namely, salicylic acid (SA), jasmonic acid (JA) and ethylene (ET) play significant role in regulating the signaling cascades during stress. SA is involved in plant defense against pathogens and induction of pathogenesis related (PR) genes. JA and ET pathways are fundamental to wounding and anti-herbivore defense in the plants (Walling, 2000). Also, it has been observed that a cross talk between these pathways exists, which actually governs the nature of plant response (Erb et al., 2012).

Plants have evolved a plethora of chemical strategies to deal with the phytopathogens and herbivores. Investigation of antiherbivore defenses in plants has highlighted occurrence of several such molecules and signaling cascades, which ultimately restrict herbivore attack (Walling, 2000; Felton, 2005). Also systemic induction of defense traits indicates the existence of "complex regulatory networks". Mobile signaling molecules travel to the undamaged regions from the damaged parts of the plant to initiate defense reactions and herbivore attack alert signals in systemic tissues (Schilmiller and Howe, 2005). Plants sense the presence of herbivores and the changes in several signaling pathways are initiated, by means of Ca^{2+} flux, membrane depolarization, kinase activation, etc. These pathways form sophisticated intertwined regulatory networks that orchestrate specific defense responses according to the attacking species of the herbivore (Wu and Baldwin, 2010). Redox associated or reactive oxygen species (ROS) mediated signaling has been found to play a major role in amplification of defense responses in plants (Mittler et al., 2004). Wounding/injury alone can activate multiple signal transduction pathways to initiate damage repair and prevent further damage. Wounding essentially accompanies herbivore attack, therefore, events like

oxidative burst and release of oligosaccharides from damaged cell walls, take place in either of the conditions. However, plants perceive wounding and herbivore attack differently due to the presence of elicitors in insect oral secretion. Hence, despite similarities in plant responses to wounding and herbivore attack, they differ qualitatively as well as in magnitude (Walling, 2000). Treatments like wounding and wounding followed by application of *Helicoverpa armigera* oral secretion (OS), led to the systemic induction of significant amounts of diverse protease inhibitors (PI) in *Capsicum annuum*. These PIs also show differential accumulation in response to the respective treatments, indicating differential perception of elicitors in the OS (Mishra et al., 2012).

The objective of the present study was to identify/characterize the proteomic changes in the induced *C. annuum* plants. *C. annuum* plants were subjected to two types of induction treatments namely, wounding followed by application of sterile water or *H. armigera* OS (to mimic insect infestation). Uninduced/control leaves were collected from unwounded healthy plants. The proteomic profiles of systemic leaves from the treated and untreated *C. annuum* plants, were analyzed by classical two-dimensional gel electrophoresis and mass spectrometric (nano-Liquid Chromatorgraphy-Mass Spectrometry^{E}) identification of proteins. We identified various proteins showing differential abundance, which were linked to different metabolic pathways/processes, in the induced leaf tissues. Selected proteins were further evaluated for their transcript accumulation by semi-quantitative PCR at successive time points after wounding to support the mass spectrometric identification results. Overall, proteins having role in defense and damage repair were upregulated, while those involved in photosynthesis were downregulated.

2A.2. Materials and Methods

2A.2.1. Plant material and induction treatments

C. annuum seeds, variety Phule Jyoti (from Mahatama Phule Agricultural University, Rahuri, India) were grown in pots with soilrite and were watered daily with Hoagland solution. Plants were cultivated in greenhouse at 23° C (+/-2°C) and a photoperiod of 14 h of light. The 30-day-old seedlings were transferred to individual pots. Induction treatments of wounding and insect infestation were made on 3 month old

plants. Wounding was induced in greenhouse with vertical and horizontal cuts by pattern wheel on the upper surface of the leaves. This was followed by the application of sterile water or *H. armigera* OS (to mimic natural insect infestation) on the punctured/wounded surface. These were considered as local tissue whereas the non-wounded leaves one node above or below were collected as systemic tissue. Unwounded healthy plants were used as the source of uninduced/control tissue. Systemic tissue was collected after 30 h of the above mentioned treatments for proteomic analysis. For gene expression analysis, systemic leaves were harvested 3, 6, 9 and 12 h after wounding. Tissues were flash-frozen in liquid nitrogen and stored at -80° C till further use.

2A.2.2. Preparation of protein extracts

Total soluble protein was extracted from water (5% polyvinylpolypyrrolidone) extract by acetone precipitation and resuspended in rehydration buffer (6 M urea, 2 M thiourea, 4% CHAPS, 50 mM dithiothreitol) (Görg et al., 2004). Protein content was estimated using the Bradford reagent (Bio-Rad protein assay dye reagent concentrate, Bio-Rad Laboratories Inc., Hercules, CA, USA) (Bradford, 1976).

2A.2.3. Separation by two-dimensional gel electrophoresis (2-DE) and image analysis

2-DE separation of proteins was done for total protein profiling. For first dimension, 11-cm IPG strips pH 3-10 NL (Bio-Rad Laboratories Inc.) were rehydrated with 200 µg of protein (in 200 ml rehydration buffer) for 14 h at 25 °C. The proteins were focused on a Protean IEF Cell (Bio-Rad Laboratories Inc.) at 20 °C, rapid voltage ramping, 20,000 V h, and 50 mA current per IPG strip. Preceding second-dimension separation, the IPG strips were equilibrated with a DTT buffer (I) (6 M urea, 0.375 M Tris HCl pH 8.8, 2% sodium dodecyl sulfate, 20% glycerol and 2% DTT) followed by an iodoacetamide buffer (II) (DTT replaced by 2.5% iodoacetamide), each for 10 min at RT on a shaker. Second dimension separation was done on 12% SDS-PAGE gel using Hoefer electrophoresis unit (GE Healthcare Bio-sciences AB, Buckinghamshire, UK) at 25 °C and 200 V constant. The gels were stained with Coomassie R-250. Gel images were taken using a GS-800 calibrated densitometer (Bio-Rad Laboratories Inc.). The software PDQuest version 7.3.1 (Bio-Rad Laboratories Inc.) was used for matching,

normalization, spot detection and quantification. The ratios of the relative spot quantities were calculated. All the analysis was performed using at least three 2-DE images. Protein spots of interest were excised for tryptic digestion.

2A.2.4. In-gel digestion and peptide extraction

The gel pieces with protein spots were excised and destained by washing in 50% acetonitrile (ACN)/50% 50 mM ammonium bicarbonate followed by dehydration with 100% ACN. Gel pieces were reduced in 10 mM dithiothreitol (DTT)/100 mM ammonium bicarbonate for 60 min at 56° C. Alkylation was done by incubation with 55m M iodoacetamide/ 100mM ammonium bicarbonate for 45 min in dark at 25° C. Gel pieces were again washed and dehydrated with ammonium bicarbonate and ACN as earlier. Trypsin digestion was carried out overnight with 50 ng of trypsin (Porcine trypsin; Sigma-Aldrich Chemical Co., St. Louis, MO, USA) in 20 µL of 50 mM ammonium bicarbonate at 37° C. The digest solution was collected in separate tubes. Peptides were extracted with 30 µL extraction buffer (50% ACN/2% formic acid) by vortexing and sonication. The extraction step was repeated thrice. The solution containing the peptide mixture was then vacuum dried.

2A.2.5. LC-MS^E Analysis and protein identification by database searches

All the samples were analyzed by nanoscale capillary LC-MS^E using a NanoAcquity ultra performance liquid chromatography (UPLC) system (Waters Corporation, Milford, CA, USA) coupled to a Q-TOF MALDI-SYNAPT High Definition Mass spectrometer (Waters Corporation). Dry peptides were reconstituted in 10 µL of aqueous 3%ACN/ 0.1% formic acid. The nano-LC separation was performed using a bridged-ethyl hybrid (BEH) C18 reversed phase column (1.7 µm particle size) with an internal diameter of 75 µm and length of 150 mm (Waters Corporation). The binary solvent system used comprised 99.9% water and 0.1% formic acid (mobile phase A) and 99.9% acetonitrile and 0.1% formic acid (mobile phase B). The samples were initially pre-concentrated and desalted online at a flow rate of 5 µL/min using a Symmetry C18 trapping column (internal diameter 180 µm, length 20 mm) (Waters Corporation) with a 0.1% B mobile phase. Each sample (total digested protein) was applied to the trapping

column and flushed with 0.1% solvent B for 3 min at a flow rate of 15 µL/min. After each injection, peptides were eluted into the ion source at a flow rate of 300 nL/min using a gradient of 2-40% B over 30 min. The lockmass calibrant peptide standard, 600 fmol/µL glu-fibrinopeptide B, was infused into the ion source at a flow rate of 300 nL/min and was sampled during the acquisition at 30 s intervals. The mass spectrometer was operated in V-mode at a resolution of at least 9000 full width at half height (fwhh). For LC-MS^E, full scan (m/z 50-2000) LC-MS^E data were collected using the "expression" mode of acquisition, which acquires alternating 1 s scans of normal and elevated collision energy. Data were collected at a constant collision energy setting of 4 V during low-energy MS mode scans, whereas a step from 20 to 40 V of collision energy was used during the high-energy MS^E mode scans.

Protein Lynx Global server (PLGS) browser version 2.1.5 software (both from Waters Corporation) were used for database searches. The data was searched against the Solanaceae database (www.uniprot.org) with the following parameters: peptide tolerance of 20 ppm, fragment tolerance of 0.05D, one missed cleavage, carbamido methylation of cysteines and possible oxidation of methionine. The Ion Accounting search parameters that were used to search the data independent analysis (DIA) data included precursor and product ion tolerance (automatic setting), minimum number of peptide matches (1), minimum number of product ion matches per peptide (5), minimum number of product ion matches per protein (7). The false positive rate was 4%. Search results of the proteins and the individual MS/MS spectra with confidence level at or above >95% were accepted. Triplicate replications were performed for each sample. Gene ontology annotation of the identified proteins was obtained from UniProt Knowledgebase (www.uniprot.org).

2A.2.6. Transcriptomic analysis

Total RNAwas isolated from the leaf tissues using Trizol reagent (Invitrogen, Carlsbad, CA, USA). Integrity of RNA was determined by agarose gel electrophoresis and quantified using Nanodrop (Thermo Scientific, Waltham, MA, USA). cDNA synthesiswas done using a reverse transcriptase (RT) kit (Promega, Madison, WI, USA) with 2 µg total RNA. Normalization of the template was done using 18S rRNA. PCR

conditions for gene specific amplification were as follows: 95° C for 5 min for initial denaturation, followed by 34 cycles of 95° C for 30 s, 55° C/60° C for 45 s and 72° C for 90 s, with final extension of 72° C for 10 min. Sequences of gene-specific primers used were as mentioned in **Appendix I Table 1**.

2A.3. Results

2B.3.1. Proteomic changes in induced C. annuum leaves

Biotic stress entails restrain on normal plant growth and development. Plants try to evade the damage by activating defenses, which in turn requires metabolic reorganization (Walling, 2000). Induction treatments to *C. annuum* leaves, showed differential accumulation of proteins with respect to uninduced leaves. These proteins were involved in varied functions including cellular processes, metabolism and stress response **(Fig. 2A.1 and Table 2A.1)**. The detailed information of the identified proteins and the fold differences in abundance of proteins under different induction treatments is listed in **Table 2A.1**.

Figure 2A.1:

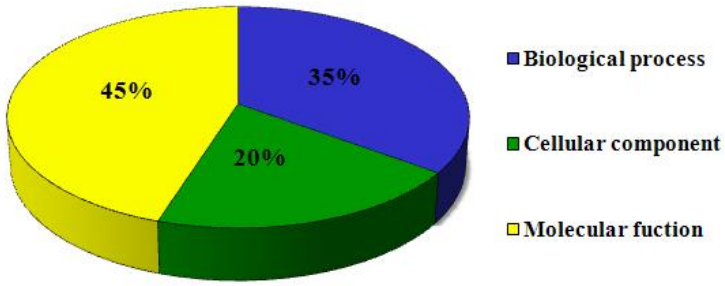


Fig. 2A.1: Illustration of classification of proteins identified in the present study on the basis of gene ontology.

Twenty-two protein spots showing differential protein accumulation were identified. The proteins were identified with sequence coverage ranging from 12.3% to 65.5% and the number of matching peptides ranged from 9 to 101. In the present study, both the induction treatments showed similar trends of protein accumulation **(Fig. 2A.2)**. The upregulated proteins were NADPH quinone oxidoreductase (spot 1, 2, 10, 22);

glycolate oxidase (spot 12); pectinesterase (spot 13); chloroplast ferredoxin NADP oxidoreductase (spot 14); superoxide dismutase Cu/Zn (spot 15); 23 kDa polypeptide of photosystem II (spot no. 16), while chloroplast sedoheptulose-1,7-bisphosphatase (spot 4, 5, 6, 11), fructose-1,6-bisphosphatase (spot 3) and carbonic anhydrase (spot 7, 19, 20) showed lower fold accumulation as compared to uninduced leaves **(Fig. 2A.2)**. Superoxide dismutase showed the highest upregulation (10-fold) in either of the induced tissues while, carbonic anhydrasewas downregulated by 9-fold in wounded leaves treated with *H. armigera* OS. Spots showing marked visual difference in protein accumulation are represented as magnified images in **Fig. 2A.3**.

Table 2A.1: Protein spots identified by mass spectrometry. Mass spectrometric identification of protein spots using Solanaceae database. The upregulation (↑) or downregulation (↓) of proteins in *Capsicum annuum* leaves after induction treatments with respect to uninduced leaves, listed along with their fold difference. Induction treatments: Wounding +water (W+ W) and wounding + *H. armigera* oral secretion (W+S). GO: Gene ontology; MW: Molecular weight; BP: Biological process; MF: Molecular function; CC: Cellular component; pI: Isoelectric point.

Spot No.	Accession No.	Identified protein	GO	Protein identification data				Fold change	
				No. of peptides/ coverage	MW (kD)	pI	PLGS Score	W+W	W+S
4, 5, 6, 11	C5IU71	Chloroplast sedoheptulose-1,7-bisphosphatase	BP, MF	41/ 43.6	42.56	6.01	7097	↓1.2	↓1.2
1, 2, 9, 10, 22	B9VXZ6	NAD P H quinone oxidoreductase	BP, MF, CC	19/ 50.7	21.99	7.09	5904	↑1.2	↑3.5
3	Q9XF82	Fructose-1, 6-bisphosphatase	BP, MF	25/ 34.3	44.43	5.07	2549	↓4.0	↓1.0
16	Q40458	23 kDa polypeptide of water oxidizing complex of photosystem II Fragment	BP, MF, CC	12/ 17.5	21.96	5.09	1711	↑4.4	↑5.5
7,8,17,20,21	Q5NE20	Carbonic anhydrase	BP, MF	8/ 27.1	34.45	6.73	4460	↓4.4	↓6.0
19	Q8W183	Carbonic anhydrase		25/ 26.7	34.43	6.19	6491	↓4.4	↓6.0
12	E1AXT8	Glycolate oxidase	MF	101/ 47.7	40.57	9.37	11768	↑5.0	↑3.5
13	Q9SEE7	Pectinesterase Fragment	BP, MF, CC	9/ 12.26	58.03	6.49	8452	↑5.4	↑7.0
14, 18	Q9M4D2	Chloroplast ferredoxin NADP oxidoreductase	MF, CC	41/ 65.47	40.38	8.54	3601	↑6.0	↑6.0
15	Q7XAV2	Superoxide dismutase Cu Zn	BP, MF	21/ 25.81	22.26	6.03	4630	↑10.0	↑10.0

Figure 2A.2:

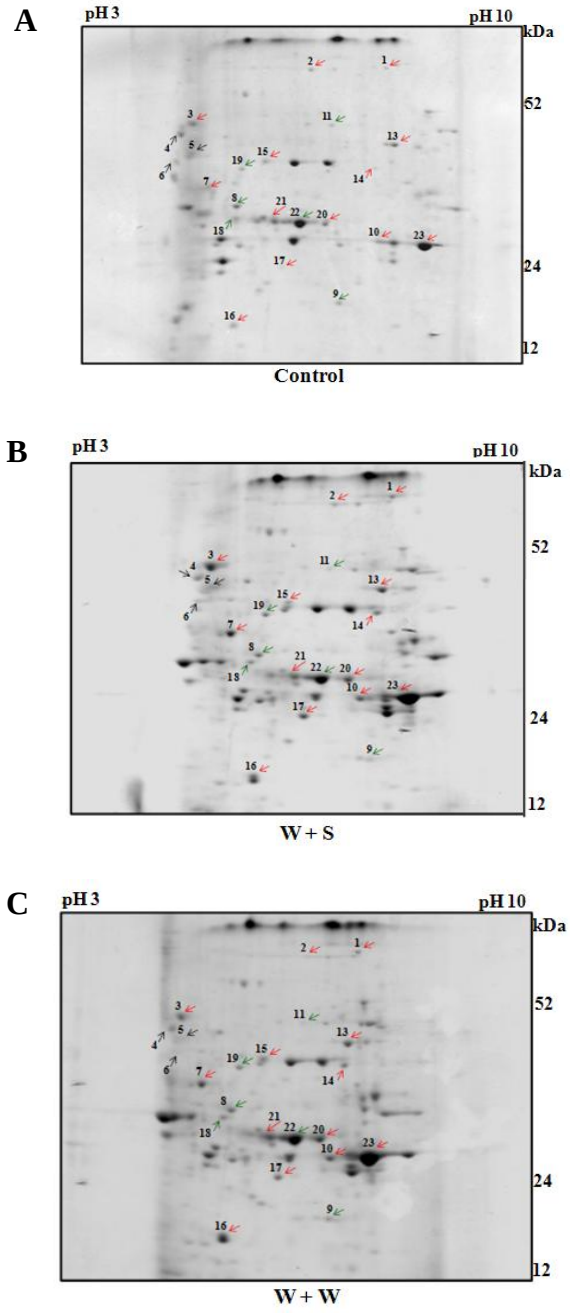


Figure 2A.2: Two-dimensional gel electrophoresis of uninduced (control) and induced leaf proteins. The 2D gel electrophoresis of proteins isolated from uninduced and induced leaves. The spot numbers correspond to the identified proteins listed in Table 2A.1. (A) Control (B) Wounding + oral secretion (W+S) (C) Wounding + water (W+W). Red or green arrows indicate upregulation or downregulation respectively. Increased or decreased intensity indicate upregulation or downregulation respectively. Red or green arrows indicate upregulation or downregulation respectively.

Figure 2A.3:

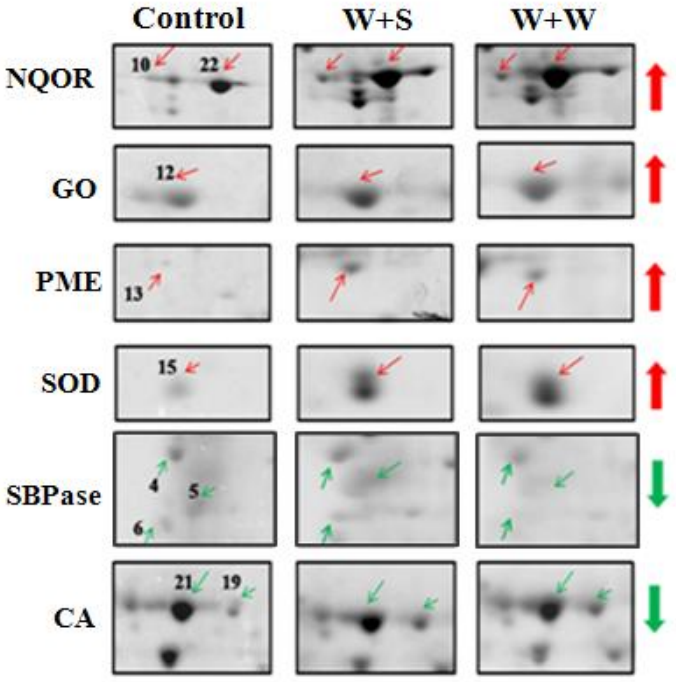


Figure 2A.3: Magnified spots from 2D gels. Increased or decreased intensity indicate upregulation or downregulation respectively. Red or green arrows indicate upregulation or downregulation respectively. NQOR (NADPH quinone oxidoreductase), GO (glycolate oxidase), PME (pectinesterase), SOD (superoxide dismutase), SBPase (Chloroplast sedoheptulose-1,7-bisphosphatase), CA (carbonic anhydrase).

2B.3.2. Semi-quantitative expression analysis of candidate genes

To validate the accumulation patterns of proteins, the transcript accumulation was examined for few candidate genes, at 3, 6, 9 and 12 h after induction **(Fig. 2A.4)**. Since the proteomic responses in both types of inductions were similar, only wound induced tissue was used for gene expression analysis. Except for carbonic anhydrase, the transcript accumulation of selected genes was in accordance with the proteomic data.

Figure 2A.4:

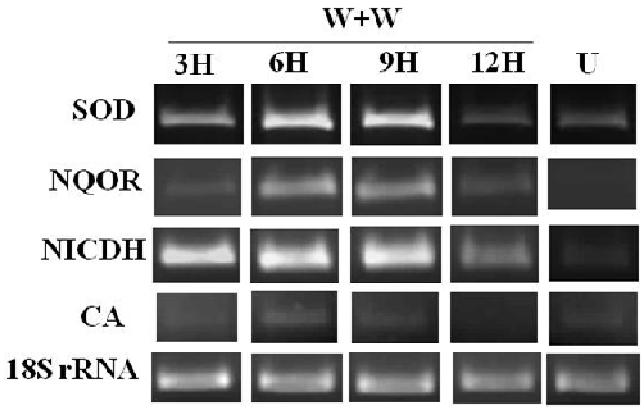


Figure 2A.4: Gene expression analysis. Candidate genes showing differential expression in C. annuum leaves at progressive time points after wounding. Upregulation of genes in response to wounding is evident from the transcript accumulation pattern. SOD (superoxide dismutase), NQOR (NADPH quinone oxidoreductase), NICDH (NADP dependent isocitrate dehydrogenase), CA (carbonic anhydrase), U (uninduced).

2A.4. Discussion

Activation of such defenses in plants is associated with fitness costs due to alteration/shift in allocation of resources (Zavala and Baldwin, 2004). Allocating resources from growth to defense leading to a reduction of photosynthetic capacity in leaf tissues characterize the cost for defense (Giri et al., 2006; Hermsmeier et al., 2001). The roles of various proteins found to be differentially expressed during the current study are discussed below.

Glycolate oxidase (GOX) converts glycolate into glyoxylate, during photorespiration, producing H_2O_2 as a by-product. Upregulation of GOX and consequent generation of H_2O_2 is also associated with the hypersensitive response in plants. H_2O_2 was also found to induce SA signaling cascades and regulate JA and ET pathways as well (Rojas et al., 2012). Redox signaling associated with photorespiration aids in maintaining cellular homeostasis (Foyer et al., 2009). H_2O_2, which is a small diffusible molecule, has been widely accepted as a signaling molecule when present in low concentrations. At high concentrations it induces programmed cell death leading to hypersensitive response. It also plays a role in regulating the cascades involving physiological processes like nonhost resistance, phytoalexin production, strengthening of cell wall, senescence, photosynthesis, and the cell cycle. Plant growth and defense responses are also regulated

by redox signaling cascades. Transcription factors like CDPKs, WRKY, DREB, bZIP, HSF and MYB have been found to be associated with H_2O_2 signaling/burst during biotic or abiotic stress (Petrov and Breusegem, 2012). Apart from signaling, the accumulated H_2O_2 also plays a direct defensive role against herbivores and pathogens, which in turn might subject the plant to oxidative stress (Ruuhola and Yang, 2006).

NADPH quinone oxidoreductase (NQOR) was upregulated by 2.3 fold (spot no. 9) in case of leaves treated with wounding followed by OS than that in leaves treated with wounding. NQOR plays a role in detoxification of plant quinones (Sparla et al., 1996). Lipid peroxides generated in plants during biotic stress can be degraded to form toxic aldehydes like α,β-unsaturated aldehydes. P1-ζ-crystallin, an NADPH quinone oxidoreductase in Arabidopsis thaliana was found to be induced under oxidative stress to detoxify the lipidperoxide-derived α,β-unsaturated aldehydes (Mano et al., 2002). Superoxide dismutase (SOD) Cu/Zn is a scavenger of superoxide radicals and leads to the generation of H_2O_2. Thus, it limits oxidative damage in the plants during stress leading to increased tolerance (Gupta et al., 1993). The upregulation of NQOR and SOD transcripts was also observed at early time points after induction i.e. from 3 to 6 h which was maintained up to 9 h **(Fig. 2A.4)**. At 12 h, the level of transcripts in case of SOD appeared comparable to uninduced leaves, whereas that for NQOR was slightly higher. Chloroplast ferredoxin NADP oxidoreductase, another redox related protein identified, is involved in electron transfer reactions of photosynthesis. It was also implicated in quenching of ROS and preserving the NADP+/NADPH balance in chloroplast, under oxidative stress (Mulo, 2011). Upregulation of proteins like SOD, NQOR and chloroplast ferredoxin NADP oxidoreductase contributes in preventing the excessive oxidative damage in plants under stress. Upregulation of proteins associated with the ROS metabolism/- H_2O_2 production or scavenging highlights the important role of ROS as a signaling molecule in stress management. Formation of ROS is usual even under stress free conditions and is kept under control by antioxidant enzymes like superoxide dismutase, catalase, glutathione reductase etc. At high concentrations, H_2O_2 can oxidize proteins, lipids and nucleic acids leading to inactivation of enzymes and the photosystems I and II (Mano et al., 2002; Sharma et al., 2012). The plant has to cope with this oxidative

burst of ROS in response to wounding/herbivore attack or pathogen attack so as to prevent excessive cellular damage.

Photosynthesis related proteins identified, in the present study, appeared downregulated in the induced leaves with respect to the uninduced leaves. The carbonic anhydrase (CA) accumulation in the leaves treated with wounding followed by insect OS was 7-fold (spot no. 19) lower than that in leaves treated with wounding. CA, which primarily catalyzes hydration of CO_2 to bicarbonate, is one of the abundant (~2% of total soluble leaf protein) proteins with multiple isoforms and varied roles in C3 plants. CA enables efficient CO_2 fixation by Rubisco (ribulose-1,5-bisphosphate carboxylase/oxygenase) and also aids the diffusion of CO_2 and HCO_3^- across the chloroplast (Badger and Price, 1994). Downregulation of CA, might lead to reduced supply of CO_2 to Rubisco thus affecting photosynthesis. However, the CA transcripts showed modest upregulation in induced tissue with higher levels at 6 h. Hara et al., (2000) detected an increase in transcripts of chloroplastic carbonic anhydrase in mechanically wounded leaves. This increase was assumed to be aiding the maintenance of plant cellular homeostasis by buffering the stroma or regulating the proton gradient across the thylakoid membranes (Hara et al., 2000; Lazova et al., 2009).

The other two photosynthesis related proteins identified were sedoheptulose-1,7-bisphosphatase and fructose-1,6-bisphosphatase. The enzyme sedoheptulose-1,7-bisphosphatase (SBPase) catalyzes the dephosphorylation of sedoheptulose-1,7-bisphosphate (SBP) in the regenerative phase of the Calvin cycle. Reduced levels of SBPase activity can further reduce the photosynthetic capacity as hindered regeneration of ribulose-1,5-bisphosphate (RuBP) limits carbon fixation (Harrison et al., 1998; Ölçer et al., 2001). Fructose-1,6-bisphosphatase (F16BPase) was downregulated by 3 fold (spot no. 3) in case of leaves treated with wounding followed by application of OS than that in leaves treated with wounding. In plants, two isoenzymes of F16BPase occur. One is the chloroplastic isoform, which participates in the regeneration of ribulose bisphosphate and another is localized in the cytosol and is involved in synthesis of sucrose from triose phosphates (Raines et al., 1999). Inhibition of chloroplastic F16BPase negatively affects photosynthesis (Koßmann et al., 1994). The activation of plant defenses requires channeling of resources towards protection rather than towards growth and development

(Walling, 2000). Hence the costs of launching defenses are compensated by downregulation of photosynthesis and growth (Giri et al., 2006; Bilgin et al., 2010; Ashraf and Harris, 2013). The 23 kDa polypeptide of photosystem II (PSII), which is found to be upregulated in the present study, is associated with a supercomplex of 20 subunit proteins and three inorganic ions, manganese (Mn), calcium (Ca), and chloride (Cl), located in thylakoid membranes in the chloroplasts of photosynthetic organisms. This machinery splits water to molecular oxygen using light energy. Three extrinsic proteins PsbO (33 kD), PsbP (23 kD), and PsbQ (17 kD) are present in higher plants, which are vital for assembly and functioning of PSII. The 23 kD protein (PsbP), is expected to guard the Mn cluster from exogenous reductants (like H_2O_2) or preserve the reduced Mn (Mn++) (Ifuku et al., 2005). PsbP plays an important role in grana stacking and maintaining the structural integrity of PSII supercomplex, especially during photoactivation of the PSII complex through the repair cycle. The PsbP and PsbQ proteins collectively maintain the inorganic environment for water oxidation (Suorsa and Aro, 2007). Hence upregulation of PsbP protein indicated the plant's attempts to retain/revive the vital functions under stress. Pectin methylesterase (PME), the cell wall hydrolyzing enzyme, was upregulated in induced leaves. PME catalyses the demethylation of pectin, the cell wall polymer. The OS of larvae due to its alkaline pH was found to elicit the PME transcripts and activity, which were associated with the methanol emissions from the wound site (Giri et al., 2006). Pectin methylesterase (PME) is also known to be potential pathogen response gene and has been found to be expressed at elevated levels upon wounding (Schafleitner and Wilhelm, 2001; Delessert et al., 2004). Transcript accumulation pattern in induced tissues was examined for NADP dependent isocitrate dehydrogenase (NICDH), known for its role in wound induced response. NICDH showed comparable upregulation at 3, 6 and 9 h. NICDH catalyses the conversion of isocitrate to 2-oxoglutarate, generating NADPH during the reaction. Thus, NADPH (a reducing equivalent) is supplied for protection against oxidative damage in mitochondria, cytosol, chloroplast, glyoxysomes and peroxisomes (Corpas et al., 1999). Also the oxidative burst and subsequent accumulation of H_2O_2 is intervened by the action of NICDH and NADPH (Leterrier et al., 2007). Upon biotic stress, the expression of defense genes is regulated by the cross talk of signaling pathways involving jasmonic

acid, salicylic acid and ethylene (Reymond and Farmer, 1998; Gulati et al., 2013). However, redox related mechanisms have also been reported to have significant influence on cell biochemistry. Also "redox dependent sensing" might play a role in monitoring environmental cues and triggering changes in gene expression (Noctor, 2006). Mejía-Teniente et al., (2013) observed that application of elicitors like SA, H_2O_2, and chitosan exogenously in *C. annuum* led to increase in endogenous H_2O_2, accompanied by increased gene expression and enzymatic activities of phenylalanine ammonia lyase and catalase. H_2O_2 accumulation was assumed to induce a "stress alert" condition in *C. annuum*.

2A.5. Conclusion

Wounding and herbivore attack altered the metabolic processes in *C. annuum* plants. Redox related proteins were upregulated, implying the role of ROS production and scavenging in stress response and signaling. Photosynthesis related proteins were downregulated, perhaps to counterbalance the upregulation of defense traits. Materialization of such counter responses requires a coordinated expression of several signaling cascades. The results presented here suggest that the interplay of defense and tolerance/ recovery enables the plants to sustain biotic stress.

Chapter 2B: Structural Features of Diverse Pin-II Proteinase Inhibitor Genes from *Capsicum annuum*

Abstract

Several potato inhibitor II (Pin-II) type proteinase inhibitor (PI) genes have been analyzed from *Capsicum annuum* (L.) with respect to their differential expression, during plant defense response. However, complete gene characterization of any of these *C. annuum* PIs (*CanPI*s) has not been carried out so far. Complete gene architectures of a previously identified *CanPI-7* (Beads-on-string, Type A) and a member of newly isolated Bracelet like (Type B), *CanPI-69* are reported in this study. The 5'UTR (untranslated region), 3'UTR, and intronic sequences of both the *CanPI* genes were obtained. The genomic sequence of *CanPI-7* exhibited, exon 1 (49 base pair, bp) and exon 2 (740 bp) interrupted by a 294 bp long type I intron. We noted the occurrence of three multi-domain PIs (*CanPI-69, 70, 71*) with circularly permuted domain organization. *CanPI-69* was found to possess exon 1 (49 bp), exon 2 (551 bp) and a 584 bp long type I intron. The upstream sequence analysis of *CanPI-7* and *CanPI-69* predicted various transcription factor-binding sites including TATA and CAAT boxes, hormone responsive elements (ABRELATERD1, DOFCOREZM, ERELEE4), and a defense responsive element (WRKY71OS). Binding of transcription factors such as zinc finger motif, MADS-box and MYB to the promoter regions was confirmed using electrophoretic mobility shift assay followed by mass spectrometric identification of proteins bound to promoter region. The 3'UTR analysis for 25 *CanPI* genes revealed unique/distinct 3'UTR sequence for each gene. Structures of 3 repeat CanPIs of type A and B were predicted and further analyzed for their attributes. This investigation of *CanPI* gene architecture will enable the better understanding of the genetic elements present in CanPIs.

2B.1. Introduction

Wound-inducible proteinase inhibitors (PIs) are one of the chemical strategies evolved by plants to deal with the biotic attackers. Potato inhibitor II type (Pin-II) PI proteins are known to occur in monocots as well as dicots, however, they have been extensively studied and characterized from Solanaceous plant species (Barta et al. 2002).

A distinct feature of Pin-II PIs is the presence of tandem repeats of around 50 amino acid termed as inhibitory repeat domain (IRD) (1 to 8) interconnected by linker peptides. Gene duplication, unequal crossing over and/or domain duplication events have resulted in the evolution of the multi-domain Pin-II PIs with structurally and functionally divergent IRDs (Hartl et al., 2010).The gene structure of Pin-II PIs is highly conserved throughout the Solanaceous plants. A typical Pin-II PI gene essentially contains a type I intron (100 to 200 bp), exon 1, and exon 2. Generally, exon 1 codes for the part of N-terminus of the signal peptide while exon 2 codes for the C-terminus of the signal peptide and the inhibitory repeat domain/s of the PI protein (Barta et al., 2002). The Pin-II PI protein consists of an endoplasmic reticulum targeting signal peptide or a vacuolar sorting signal at the C-terminal followed by a chain of one or multiple number of IRDs. The IRD forms the functional unit of the PIs consisting of an active site either for inhibition of trypsin or chymotrypsin. It is characterized by a variable amino acid composition but contains eight conserved cysteines and a proline residue (Kong and Ranganathan, 2008). Classically, a multi-IRD Pin-II protein can attain two possible domain organizations; (i) circularly permuted domain organization (clasped bracelet like conformation), in which the two partial repeats at both ends (N and C terminal of precursor proteins) of the molecule unite to form a complete functional domain, or (ii) tandem repeat domain organization (beads on a string like conformation) (Schirra and Craik 2005).

Pin-II PIs form *Capsicum annuum* (L.) have been identified and characterized for their biochemical and inducible properties (Kim et al., 2001; Shin et al., 2001; Moura and Ryan, 2001; Tamhane et al., 2009, Mishra et al., 2012). Earlier, we have reported a multi domain *C. annuum* PI namely CanPI-7 and evaluated its antagonistic effect on growth and development of *H. armigera* (Tamhane et al., 2005b; 2007). Tamhane et al., (2009)

and Mishra et al., (2012) reported the constitutive and induced expression of CanPIs when the plants were challenged with wounding and insect infestation. Although these studies showed the characteristics of CanPI proteins, the complete gene structure of any of the *CanPIs* is still not explored. In the current study, we report the genomic sequence of *CanPI-7* (beads on string) and *CanPI-69* (clasped bracelet) along with the characterization of 3'UTRs of various *CanPIs*. Furthermore, the occurrence of multi-domain PIs with circularly permuted domain organization in *C. annuum* is reported for the first time. The upstream sequences of *CanPI-7* and *CanPI-69* were also explored and various transcription factor-binding sites were predicted. As a proof of principle, the binding of transcription factors to the promoters was demonstrated using mobility shift assay followed by mass spectrometric (MS) analysis. The binding conformation of CanPI-69 and CanPI-1 (beads on string), both having 3-IRDs was evaluated using structure prediction.

2B.2. Materials and Methods

2B.2.1. Plant material and isolation of DNA

Capsicum annuum (variety Phule Jyoti, obtained from Mahatma Phule Agricultural University, Rahuri, MS, India) plants were grown in individual pots and maintained in a growth chamber. Plant growth conditions, induction treatments (wounding and wounding + *H. armigera* oral secretion) and tissue harvesting were as described in (Section 2A.2.1). Genomic DNA isolation was done for *C. annuum* uninduced (without any treatment) leaves using DNeasy kit (Qiagen, Valencia, CA, USA).

2B.2.2. Functional gene isolation

Total RNA was isolated form induced leaves of *C. annuum* using Plant Spectrum RNA kit (Sigma–Aldrich Chemical Co.). RNA was quantified using spectrophotometry and analyzed by agarose gel electrophoresis. cDNA for 5' and 3' untranslated region (UTR) analysis was prepared using Smarter RACE kit and Advantage polymerase mix (Clontech, Palo Alto, CA, USA) as per manufacturer's instructions. Gene specific primers used for the amplification of UTR sequences are listed in **Appendix I Table 2**. For the isolation of 3'UTR sequences of *CanPI* genes, CanPin1F and UPM were used in

the primary PCR reactions whereas primers N8, N1m, N5m, N12, N17, N18, N62 and V63 with NUP **(Appendix I Table 2)** were used for nested PCR reactions. Isolation of 5'UTR sequences of *CanPI-7* and *-69* genes was performed by using the primers N27 and N32 in combination with UPM. The amplicons were cloned into pGEM-T Easy vector (Promega).

2B.2.3. Gene expression analysis

A semi-quantitative estimation of gene expression in uninduced and induced leaf tissues was carried out for Pin II Type A (CanPI-7) and Pin II type B (All Type B CanPIs) using primer pair (80, N29 **Appendix I Table 2**). Gene specific primers for CanPI-7 and 18S rRNA are mentioned **(Appendix I Table 1 and 2)**. PCR conditions for amplification were as follows: 95 °C for 2 min for initial denaturation, followed by 33 cycles of 95 °C for 30 s, 55 °C/60 °C for45 s and 72 °C for 90 s, with final extension of 72 °C for 10 min.

2B.2.4. CanPI promoter amplification by genome walking

The promoter fragments of *CanPI-7* and *-69* genes were isolated using the Universal Genome Walker Kit (Clontech). The genomic DNA was digested with four alternative restriction enzymes (DraI, EcoRV, StuI and PvuII) and separately ligated with the adaptors provided in the kit to obtain four restriction fragment libraries. PCRs were carried out by using gene-specific primers (N27, N32, N51 and N54; **Appendix I Table 2**) and adapter primers (AP1 and AP2) provided with the kit. Jumpstart^{TM}AccuTaq^{TM} LA DNA polymerase mix (Sigma–Aldrich Chemical Co.) was used to perform PCR. The intronic sequences were isolated using primer pairs CanPin1F and N27 for *CanPI-7* and N61 and N29 for *CanPI-69*. The obtained amplicons were cloned into the pGEM-T Easy vector (Promega).

2B.2.5. Sequence analysis

PI sequence of *Solanum lycopersicum* was obtained from Sol Genomics (SGN; http://solgenomics.net/; Mueller et al., 2005). Sequence similarity search was carried out with NCBI BLAST (www.ncbi.nlm.nih.gov). *Cis* acting elements in the promoter were detected by using PLACE database (http://www.dna.affrc.go.jp/htdocs/PLACE/; Higo et

al., 1999). ClustalW (http://www.ebi.ac.uk/clustalw; Thompson et al., 1994) was used for making sequence alignments.

2B.2.6. Electrophoretic mobility shift assay (EMSA)

The binding of regulatory proteins to the predicted promoter regions was validated using EMSA. Plant nuclear protein extracts (NPE) were prepared from leaves of uninduced and wounding induced *C. annuum* plants using CelLytica^{TM} PN (Sigma–Aldrich Chemical Co.). The promoter fragments of *CanPI-7* and *CanPI-69* were PCR amplified from plasmid constructs and purified using GenElute™ PCR Clean-Up Kit (Sigma–Aldrich Chemical Co.). The EMSA was performed by incubating 9 µg NPE with 100 ng of promoter fragment for 45 min. The binding buffer, loading dye and SYBR® Green EMSA nucleic acid gel stain provided in Electrophoretic Mobility Shift Assay (EMSA) Kit (Invitrogen) were used. DNA proteins complexes were separated on 5% polyacrylamide or 1.8% agarose gel in Tris-acetate-EDTA (pH 7.8) buffer system as described in (Hellman and Fried, 2007). Mobility shift was not affected by the change in gel system. The polyacrylamide gel was used for preparing the samples for mass spectrometry while the agarose gel image has been used for image representation. The cytoplasmic proteins from uninduced and induced leaves were precipitated using 4 volumes of acetone and resuspended in Tris buffer (0.1M, pH 7.8). Protein content was estimated using Bradford reagent (Bio-Rad protein assay dye reagent concentrate, Bio-Rad Laboratories Inc.). The cytoplasmic proteins were used as negative controls for mobility shift assay.

2B.2.7. Mass spectrometric identification of promoter binding proteins

The gel pieces with retarded DNA (DNA protein complex) band were excised. These bands were further subjected to tryptic digestion and nano-LC-MS^{E} as described in section 2A.2.5. Prior to MS analysis, the extracted peptides were purified and desalted using ZipTip® pipette tip (Millipore, Billerica, MA, USA) as per manufacturer's protocol. Protein Lynx Global server (PLGS) browser version 2.5.3 software (Waters Corporation) were used for database searches. The plant transcription factor database, PlantTFDB 3.0 (http://planttfdb.cbi.pku.edu.cn/; Jin et al., 2014) was used for protein identification. LC-MS^{E} data were searched with a fixed carbamidomethyl modification for cysteine

residues, and a variable modification for oxidation of methionine. The Ion accounting search parameters used to search data included precursor and product ion tolerance (automatic setting), minimum number of peptide matches (1), minimum number of product ion matches per peptide (3), minimum number of production matches per protein (7) and maximum number of missed tryptic cleavage sites (2). The false positive rate was 4%. Triplicate replications were performed for each sample.

2B.2.8. Structure prediction

3D structure of CanPI-1 and CanPI-69 (both 3-IRD PIs) was predicted by *ab initio* modelling using the automated I-TASSER server (http://zhang-lab.ccmb.med.umich.edu/I-TASSER/) (Zhang, 2008; Roy et al., 2010). To deal with disulfide bonds during structure prediction, we have specified the restrain contact between all the cysteines residues according to standard Pin-II PIs disulfide bond pattern. So, when amino acid contact restrain is specified, I-TASSER will try to move the two C-alpha and their side chain center in close proximity. Thus formation of disulfide bond with the side chains and correct orientation was ensured. The best model was selected on basis of C-score and Ramchandran plot analysis. This model was subjected to an energy minimization procedure with GROMOS96 (The PyMol Molecular Graphics System, Version 1.2r3 pre, Schrödinger LLC) to optimize stereochemistry and reduce poor van der Waals contacts. The quality of final model produced was assessed by checking protein stereology using PROCHECK and the energy was checked by ProSA. The models were visualized and represented by using PyMol visualize (The PyMol Molecular Graphics System, Version 1.2r3 pre, Schrödinger LLC).

2B.3. Results

2B.3.1. Occurrence of Clasped bracelet type structural organization in CanPIs

Previously characterized CanPIs are found to have a "beads on linear string" like structure (referred to as Pin-II type A). However, the present investigation interestingly revealed the presence of "clasped bracelet" (referred to as Pin-II type B) type PIs in *C. annuum*.

Figure 2B.1:

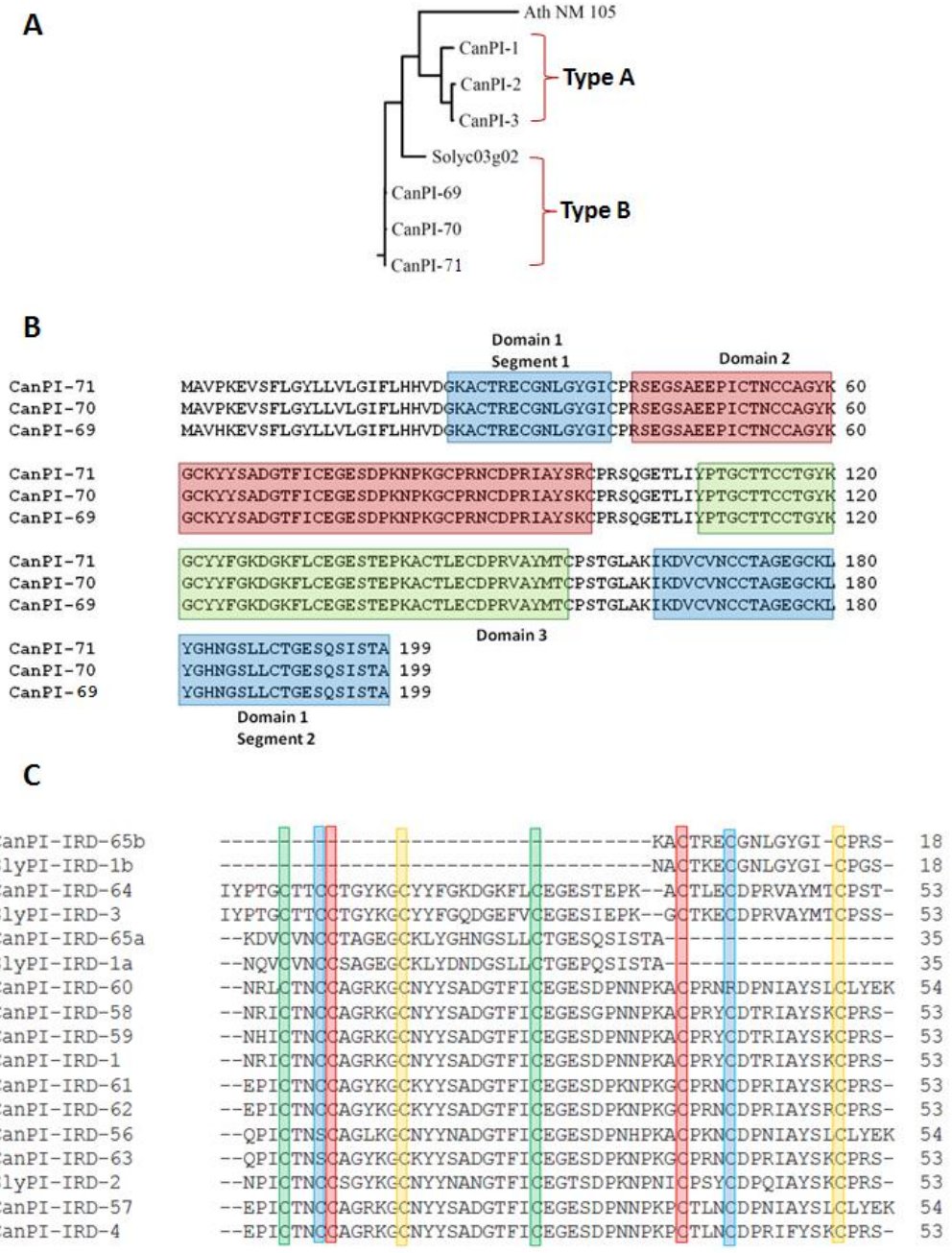


Figure 2B.1: (A) Phylogenetic tree of representative protein sequences of CanPIs of type A and B along with *Solanum lycopersicum* Pin-II PI (Solyc03g020050) and *Arabidopsis thaliana* kunitz type PI sequence as an out group. The beads-on-string type and clasped bracelet like PIs group distinctly. (B) Alignment of the Type B CanPI protein sequences obtained during the present study. Each inhibitory domain is indicated by colored boxes. Segment 1 and 2 (Partial IRDs) of domain 1 are marked in the figure. (C) Alignment of novel CanPI IRDs along with IRD4 (reported earlier) and *Solanum lycoperscium* Pin-II PI IRDs (Solyc03g020050). The cysteine residues forming disulphide bridges have been highlighted with like colors.

Identified type B PIs bear 75 to 80% similarity to *S. lycopersium* PI (Solyc03g020050). Moreover, type A and type B CanPI proteins showed 65% sequence similarity and found clustered separately **(Fig. 2B.1A)**. Protein sequence alignments of type B PIs revealed the presence of IRDs with partial segments at N- and C- terminal **(Fig. 2B.1B)**. The 5 aa linker sequences in type B CanPIs (GLAKI, QGETL and EGSAE) were different from those identified in type A (QRNAK, EENAE, EASAE, EGNAE and EETQK). IRD alignments demonstrated that CanPI IRD 64 and 65 closely aligned with S. *lycopersium* IRDs **(Fig. 2B.1C)**.

2B.3.2. Gene architecture revealed several conserved characteristics of CanPIs

While cloning the full length*CanPI-7*gene, we identified some novel *CanPI* genes (*CanPI-68* to *-79*; accession nos. KM576781-KM576792) **(Table 2B.1; Appendix II)**.

Table 2B.1: IRD composition of novel *C. annuum* Pin-II PI genes

Serial No.	***CanPI***	**PinII Type**	**Composition**				
			Signal peptide	**1st IRD**	**2nd IRD**	**3rd IRD**	**4th IRD**
1	*CanPI*-68	A	SP 2	1	25	17	-
2	*CanPI*-69	B	SP 14	65	61	64	-
3	*CanPI*-70	B	SP 15	65	62	64	-
4	*CanPI*-71	B	Partial	65	61	64	-
5	*CanPI*-72	A-2	Partial	59	25	60	-
6	*CanPI*-73	A-1	Partial	57	-	-	-
7	*CanPI*-74	A-2	SP 1	18	-	-	-
8	*CanPI*-75	A-2	Partial	1	1	17	-
9	*CanPI*-76	A-2	SP 5 like	58	1	17	-
10	*CanPI*-77	A-2	SP 5 like	1	25	17	-
11	*CanPI*-78	A-2	SP 5 like	1	25	17	-
12	*CanPI*-79	B	SP 15	65	61	64	-

We selected a previously reported *CanPI-7* (type A) and a newly identified *CanPI-69* (type B) for characterization of the complete gene architecture.

Figure 2B.2:

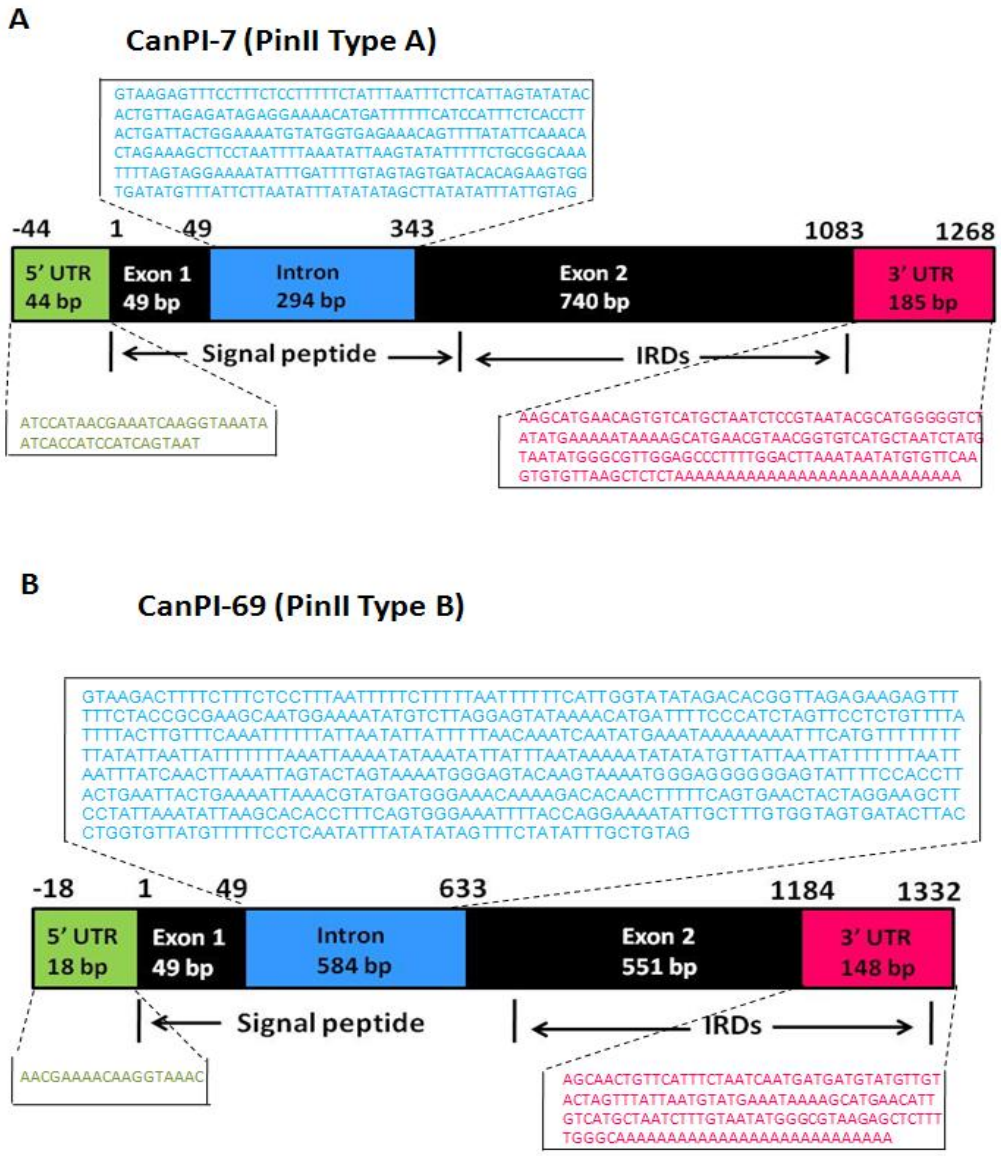


Figure 2B.2: Schematic representation of the Type A and B *CanPI* genes characterized, indicating their 5'UTR, 3'UTR and intron lengths. (A) CanPI-7 (B) CanPI-69 +1 was assigned to the "A" of start codon.

CanPI-7 (accession number DQ005913) is a four IRD PI with 789 base pair (bp) coding DNA sequence (Tamhane 2007). *CanPI-7* possessed a 44 bp long 5'UTR while 3'UTR was found to be 157 bp long. Elucidation of the genomic region of *CanPI-7* confirmed the presence of a single type I intron. Thus, exon 1 (49 bp) and exon 2 (740 bp) were interrupted by a 294 bp long intron **(Fig. 2B.2A)**. *CanPI-69* gene possessed a 121bp long 3'UTR and 18 bp long 5'UTR **(Fig. 2B.2B)**. A 584 bp long intron type I was found to be located in *CanPI-69* and the intron was flanked by exon 1 (49 bp) and exon 2 (551 bp) **(Fig. 2B.2B)**. The alignment of introns from *CanPI-7* and *-69* is shown in **Fig. 2B.3A**. The signal peptides signal peptides from type B CanPIs are aligned to signal peptides from tomato and *CanPI-7* in **Fig. 2B.3B**.

Figure 2B.3:

A

```
CanPI-7          GTAAGAGTTTCCTTTCTCCTT-----TTTCTATTTAATTTCTTCATTAGTATATA--CAC  53
CanPI-69         GTAAGACTTTTCTTTCTCCTTTAATTTTTCTTTTTAATTTTTTCATTGGTATATAGACAC  60
SolyC03G020050   GTAAGATTTTAATTT----TT-----TTTCTCCTTGGTACAATCAAATTTCTATTATAAC  51

CanPI-7          TGTTAGAGAT-------------------------------------------AGAGGAAAA  72
CanPI-69         GGTTAGAGAAGAGTTTTTTCTACCGCGAAGCAATGGAAAATATGTCTTAGGAGTATAAAA  120
SolyC03G020050   GGTGT-------------------------------------------------------  56

CanPI-7          CATGATTTTTTCATCCATTTCT-------------------------------------  94
CanPI-69         CATGATTTTCCCATCTAGTTCCTCTGTTTTATTTTACTTGTTTCAAATTTTTTATTAATA  180
SolyC03G020050   CAT----TTTTCAT-------------------------------------------------  66

CanPI-7          -------------------------------------------------------------
CanPI-69         TTATTTTTAACAAATCAATATGAAATAAAAAAAATTTCATGTTTTTTTTTTATATTAATT  240
SolyC03G020050   -------------------------------------------------------------

CanPI-7          -------------------------------------------------------------
CanPI-69         ATTTTTTTAAATTAAAATATAAATATTATTTAATAAAAATATATATGTTATTAATTATTT  300
SolyC03G020050   -------------------------------------------------------------

CanPI-7          -------------------------------------------------------------
CanPI-69         TTTTAATTAATTTATCAACTTAAATTAGTACTAGTAAAATGGGAGTACAAGTAAAATGGG  360
SolyC03G020050   -------------------------------------------------------------

CanPI-7          ----------------CACCTTACTG-ATTACTGGAAAAT----GTATGGTGAGAAACAG  133
CanPI-69         AGGGGGGAGTATTTTCCACCTTACTGAATTACTGAAAATTAAACGTATGATGGGAAACAA  420
SolyC03G020050   ------------------GAATAAT-----------------------------------  73

CanPI-7          ----------TTTTATATTCAAACACTAGAAAGCTTCCTAATTTTAAATATTAAGTATAT  183
CanPI-69         AAGACACAACTTTTTCAGTGAACTACTAGGAAGCTTCCTA---TTAAATATTAAGCACAC  477
SolyC03G020050   ----------------------ATTACTTGAAA-CT----------------------AA  88

CanPI-7          TTTTCTGCGGCAAATTTTAGTAGGAAAATATTTGATTTTGTAGTAGTGATACACAGAAGT  243
CanPI-69         CTTTCAGTGGGAAATTTTACCAGGAAAATATTG--CTTTGTGGTAGTGATACTTA--CCT  533
SolyC03G020050   TTGTCAGAG------------AG--AAATATCG----TTATAG--ATGAT---------- 118

CanPI-7          GGTGATATGTTTATTCTTAATATTTATATATAGCTTATATAT-TTATTGTAG  294
CanPI-69         GGTGTTATGTTTTTCCTCAATATTTATATATAGTTTCTATAT-TTGCTGTAG  584
SolyC03G020050   ----TATTGTTATAACGTAACG---CTATATAACTTATATTTGTAATTGTAG  163
```

B

```
SlyPI-SP        MAVYKVSFLAHLLVLGMYLLVSTVEHA--
CanPI-SP-5      MAVPKEVSFLASLLVLGILLLHVDAKACS
CanPI-SP-14     MAVHKEVSFLGYLLVLGIFLHHVDG----
CanPI-SP-15     MAVPKEVSFLGYLLVLGIFLHHVDG----
```

Figure 2B.3: (A) Alignment of introns of *CanPI-7*, *69* and *Solyc03g020050*. (B) Alignment of novel signal peptides obtained in the current study.

2B.3.3. CanPI genes represent unique 3'UTR sequences

CanPI genes were subjected to 3'UTR analysis that resulted into the accumulation of 25 distinct sequences **(Appendix III)**. Considering the gene structure of Pin-II PIs, each coding gene sequence should represent unique 3'UTR sequence. The results obtained in the present study are consistent with this hypothesis. Further, we noticed that the 3'UTR lengths were variable ranging from 90 to 189 bp (excluding polyA tail) with a GC content of ~38%. Sequence alignment analysis resulted into the formation of three distinct groups of 3'UTRs. The 3'UTRs of Pin-II type A and type B *CanPIs* clustered separately, further type A 3'UTR sequences were divided into two different groups, A1 and A2 **(Fig. 2B.4)**. The average sequence lengths of the UTRs (excluding polyA) were 156, 134 and 148bp for types A1, A2 and B, respectively. 3'UTR sequences of type A1 and A2 showed 88% similarity to each other and 74% similarity with type B sequences. The sequence following the stop codon was found highly conserved while the sequence between 84 to 120 bp downstream the stop codon revealed maximum diversity conferring uniqueness to each 3'UTR sequence. A probable polyA signal "AATAAA" can be predicted from the sequence alignments. PolyA signal was observed to be located at 56 and 60 bp downstream of the stop codon for type A and B PIs, respectively. However, multiple and diverse polyA signals are known to exist in plant 3'UTRs (Joshi, 1987; Loke et al., 2005). Plant 3'UTR sequences also exhibited conserved elements i.e. presence of a purine (A) at the first downstream position with respect to the stop codon (Joshi, 1987) and a T/U region preceding the cleavage site (Graber, 1999).

Figure 2B.4:

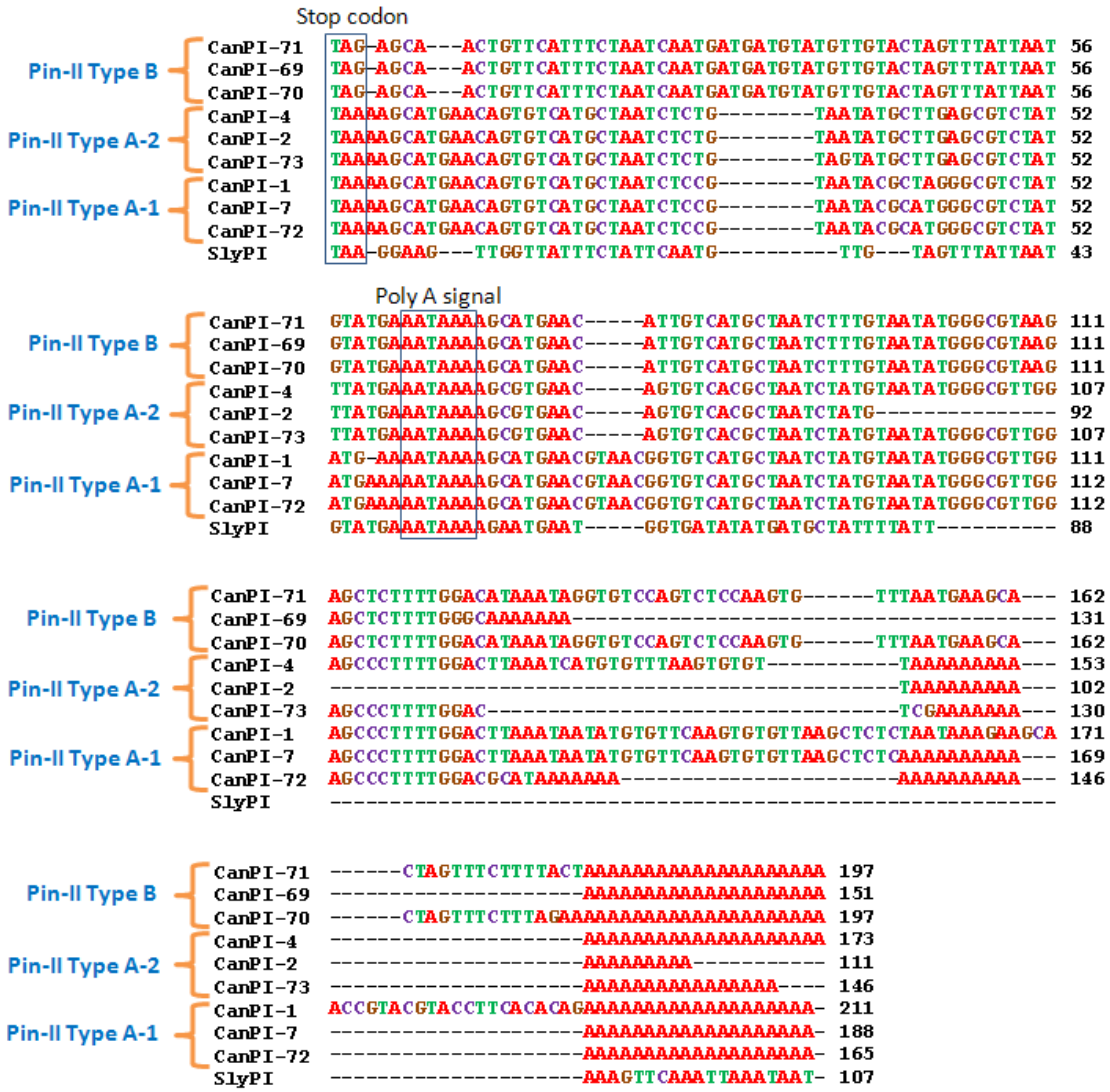


Figure 2B.4: Alignment of representative 3'UTR sequences of CanPI genes to display the diversity and groups formed by these sequences. 3'UTRs of Pin-II type A and type B CanPIs aligned separately; while type A 3'UTR sequences too were divided in two different groups (type A1 and A2).

2B.3.4. CanPI promoters are conserved and responsive to various stimuli

The 539 and 681 bp sequences upstream to the start codon of *CanPI-7* and *69* were obtained in the EcoRV fragment library and the putative *cis*- regulatory elements were predicted on these regions. As expected, elements responsive to hormones, stress, development, physiology, organ specific gene expression and defense were detected. The putative TATA box was located at -75 and -55 bp for *CanPI-7* and *CanPI-69*, respectively. The summary of elements predicted on the sense strand is provided in **Table 2B.2**. The elements predicted on both the strands are listed in **Appendix I Tables 3 and 4**. Further, these elements were classified into four categories namely defense, hormone responsive, environmental/stress and developmental/physiological for convenient representation **(Fig. 2B.5A and B)**. However, these categories are not absolute as the elements or transcription factor binding sites might operate in an overlapping fashion under various conditions/stimuli. Though most of the elements identified in the promoters of Type A and B genes were common, few elements were found to be unique for either of the genes. Elements such as, CPBCSPOR (cytokinin enhanced protein binding), PREATPRODH (abiotic stress), SEF4MOTIFGM7S (seed specific) and TAAAGSTKST1 (anther specific) were unique for *CanPI-7* whereas ACGTABOX (diverse stimuli), DPBFCOREDCDC3 (embryo specific), EECCRCAH1 (binding site for MYB), ERELEE4 (ethylene specific), MYBCOREATCYCB1 (biotic and abiotic stress), SP8BFIBSP8BIB (root specific), WBOXNTCHN48 and WBOXNTERF3 (binding site for WRKY) were exclusively found in *CanPI-69* gene. Several light responsive elements were also detected in the upstream regions of both of these genes but their significance in PI expression remains unexplored.

Table 2B.2: Putative *cis*-acting elements as predicted by PLACE. (IUPAC letter codes in DNA sequences: A = adenine, C = cytosine, G = guanine, T = thymine, R = (G/A), Y = (T/C), K = (G/T), M = (A/C), S = (G/C), W = (A/T), B = (G/T/C), D = (G/A/T), H = (A/C/T), V = (G/C/A), N = (A/G/C/T))

S. no.	Factor or Site Name	Signal sequence	PLACE reference no.	Location (-)		Function
				CanPI-7	CanPI-69	
1	ABRELATERD1	ACGTG	S000414	289	641	ABA-responsive element; Dehydration stress responses
2	ACGTABOX	TACGTA	S000130	-	80	Mediating signals from diverse environmental, developmental, and physiological stimuli
3	ACGTATERD1	ACGT	S000415	96, 289	79, 579, 634, 641	Dehydration stress and etioloation
4	ACGTTBOX	AACGTT	S000132	-	635	Similar to ACGTABOX
5	ARR1AT	NGATT	S000454	363	338, 363, 217, 368	Cytokinin primary response
6	BIHD1OS	TGTCA	S000498		263, 567	Resistance response
7	CAATBOX1	CAAT	S000028	232	159, 250, 376, 469, 604	CAAT promoter consensus sequence
8	CARGATCONSENSUS	CCWWWWWWGG	S000404	-	56, 160, 376	Embryo development
9	CARGCW8GAT	CWWWWWWWWG	S000431	76, 90	-	
10	CPBCSPOR	TATTAG	S000491	213	-	Critical for Cytokinin-enhanced Protein Binding
11	DOFCOREZM	AAAG	S000265	112, 182, 204, 406	95, 221, 307, 322, 347, 414	light, phytohormone and defense responses, seed development and germination
12	DPBFCOREDCDC3	ACACNNG	S000292	-	559	Embryo-specific expression; ABA response
13	EECCRCAH1	GANTTNC	S000494	-	216, 337	Binding site of MYB transcription factor

14	ERELEE4	AWTTCAAA	S000037	-	408	Ethylene responsive element; physiological and environmental signals
15	GT1CONSENSUS	GRWAAW	S000198	27, 185, 186, 281, 282, 526	7, 99, 242, 325	Tissue-specific (pollen), defense-related, light-repressed, and circadian clock-controlled
16	GT1GMSCAM4	GAAAAA	S000453	185	325	Pathogen- and salt-induced gene expression
17	GTGANTG10	GTGA	S000378	63, 120, 376, 497	350, 426	Pollen specific
18	MYBCORE	CNGTTR	S000176	-	511	Plant development, secondary metabolism, hormone signal transduction, biotic and abiotic stress tolerance
19	MYBCOREATCYCB1	AACGG	S000502	-	679	
20	MYBST1	GGATA	S000180	-	38, 100, 647	
21	MYCCONSENSUSAT	CANNTG	S000407	-	558	JA- and ABA-regulated signaling
22	NODCON1GM	AAAGAT	S000461	406	-	Root nodule-specific expression
23	NODCON2GM	CTCTT	S000462	294	178	
24	OSE1ROOTNODULE	AAAGAT	S000467	406	-	
25	OSE2ROOTNODULE	CTCTT	S000468	294	178	
26	POLLEN1LELAT52	AGAAA	S000245	206	412	Pollen specific expression
27	PREATPRODH	ACTCAT	S000450	389	-	Proline metabolism; abiotic stresses; hydroxyproline-rich extensin glycoproteins upregulated upon wounding
28	PYRIMIDINEBOXHVEPB1	TTTTTTCC	S000298	305	278	GA induction
29	RHERPATEXPA7	KCACGW	S000512	-	581	Root hair specific
30	ROOTMOTIFTAPOX1	ATATT	S000098	128, 199, 230, 259, 402	116, 137, 144, 233, 281, 440, 286, 525	Root specific
31	SEF4MOTIFGM7S	RTTTTTR	S000103	353, 382		Seed specific

32	SP8BFIBSP8BIB	TACTATT	S000184		69	Present in the 5' upstream regions of sporamin (storage protein) and beta-amylase of tuberous roots.
33	TAAAGSTKST1	TAAAG	S000387	407		Anther/pollen specific
34	TATABOX2	TATAAAT	S000109	75	55	TATA box elements are critical for accurate initiation
35	TATCCACHVAL21	TATCCAC	S000416		127	GA response
36	WBOXNTCHN48	CTGACY	S000508		382	Elicitor-responsive transcription of defense genes; binding site of WRKY
37	WBOXNTERF3	TGACY	S000457		381	Plant responses to various hormones or environmental cues; binding site of WRKY
38	WRKY71OS	TGAC	S000447	265, 341	381	Response to wounding, pathogens

Figure 2B.5:

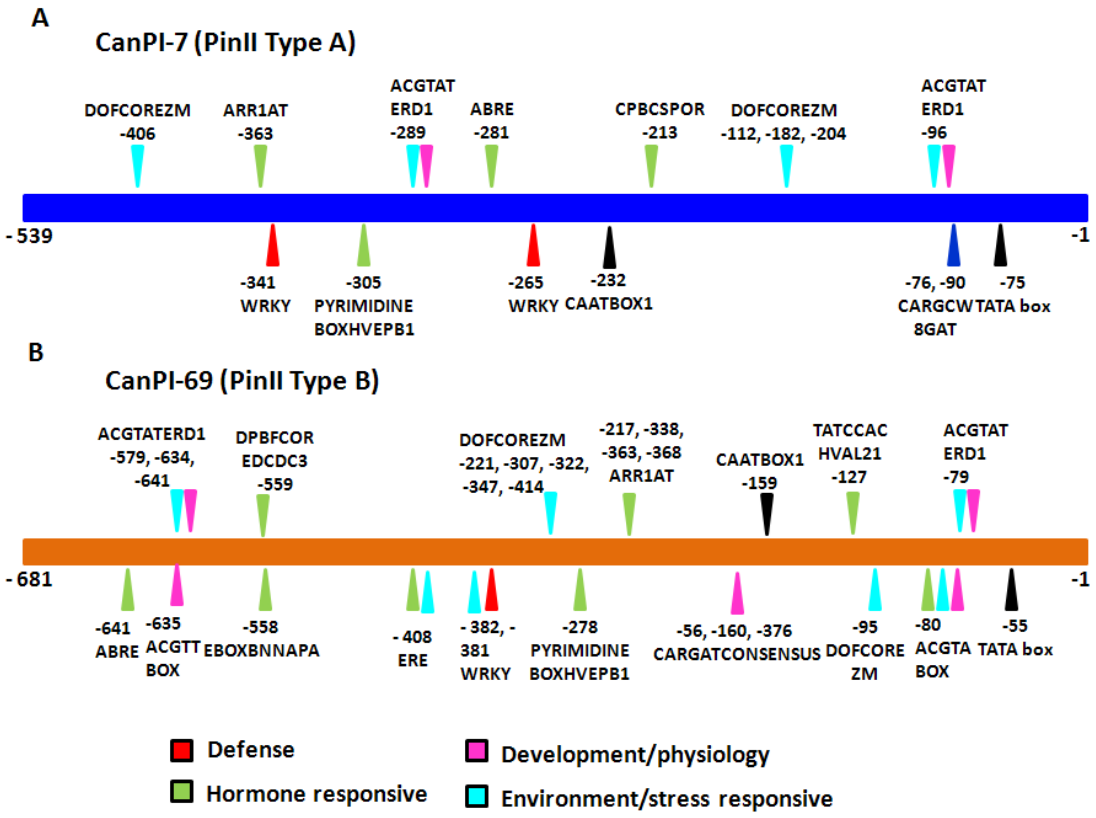


Figure 2B.5: Molecular architecture of (A) *CanPI-7* and (B) *69* promoters. Various *cis*-acting elements were predicted from PLACE. Description of various cis-acting elements marked in the sequence is provided in Table 2B.2.

2B.3.5. Activity of CanPI promoters

Mobility shift assay of promoter fragments with plant NPE indicated the binding of nuclear proteins to the DNA **(Fig. 2B.6A)**. Further MS analysis resulted into the identification of DNA binding proteins **(Table 2B.3)**. These identified proteins are in agreement with the predicted elements mentioned in **Table 2B.2**. Semi-quantitative PCR analysis revealed that type B PIs (*CanPI-69*, *-70*) were induced only in wound-induced leaves while not in uninduced leaves or wounded leaves applied with insect oral secretion **(Fig. 2B.6B)**. However, *CanPI-7* showed expression both in induced as well as uninduced leaves.

Figure 2B.6:

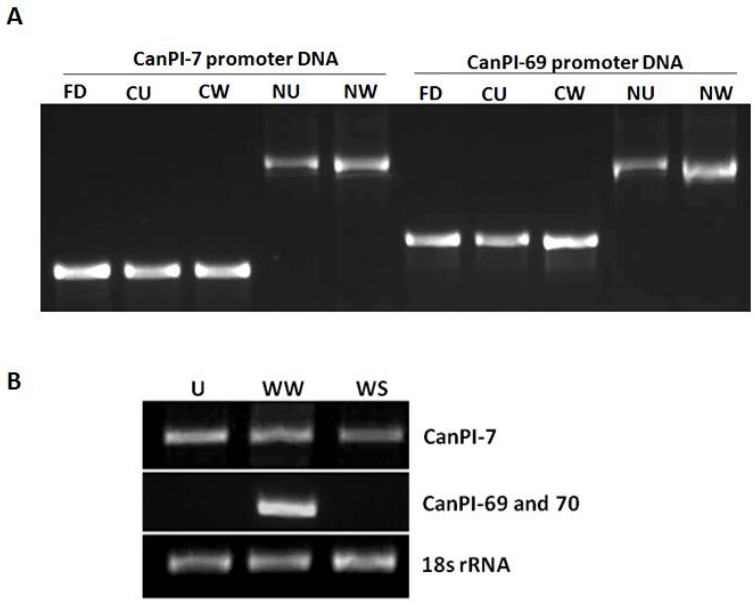


Figure 2B.6: Characterization of promoter elements. (A) Electrophoretic mobility shift assay with *CanPI-7* and *-69* promoters. Free DNA (FD), Cytoplasmic extract from uninduced (CU) and wound induced leaves (CW), Nuclear protein extract (NPE) from uninduced (NU) and wound induced (NW) leaves. Shift in mobility of bands is evident upon incubation with nuclear extracts. (B) Gene expression of Pin II Type A (*CanPI-7*) and Pin II type B (*CanPI-69* and *-70*) in uninduced (U), Wounding + water (WW) and Wounding + oral secretion (WS) systemic leaf tissues.

Table 2B.3: Mass-spectrometric identification of proteins fractionated from the EMSA bands

EMSA Reaction	**Proteins identified**
CanPI-7 promoter + NPE (uninduced)	C2H2 and C2HC zinc finger; DNA-binding storekeeper; Dof family protein; MADS-box; MYB
CanPI-7 promoter + NPE (wound induced)	C2H2 and C2HC zinc finger; DNA-binding storekeeper; MYB family; NAC domain; MADS-box; Zinc finger, Dof-type; Myc-type
CanPI-69 promoter + NPE (uninduced)	C2H2 and C2HC zinc finger; DNA-binding storekeeper; MADS-box; MYB
CanPI-69 promoter + NPE (wound induced)	C3H family protein; C2H2 and C2HC zinc finger; DNA-binding storekeeper; MYB

2B.3.6. Structural attributes of clasped bracelet type CanPIs

The structure of type B CanPI-69 (3-IRD) was predicted by protein modeling and the similar 3-IRD type A CanPI-1 was used for the structural comparisons. Since X-ray crystallographic or NMR structure of any three domain Pin-II PI molecule is not available, we used *de novo* protein modeling approach using I-TASSER server. Generally, I-TASSER considers various template structures of Pin-II PIs such as 1pjuA, 1oyvI and 1fybA to model specific regions of the query sequence. The selected best model was found to have C score of -1.47, T_m score = 0.53±0.15 and RMSD = 8.2±4.5Å and satisfy the range of parameters for molecular modeling **(Fig. 2B.7)**.

Figure 2B.7:

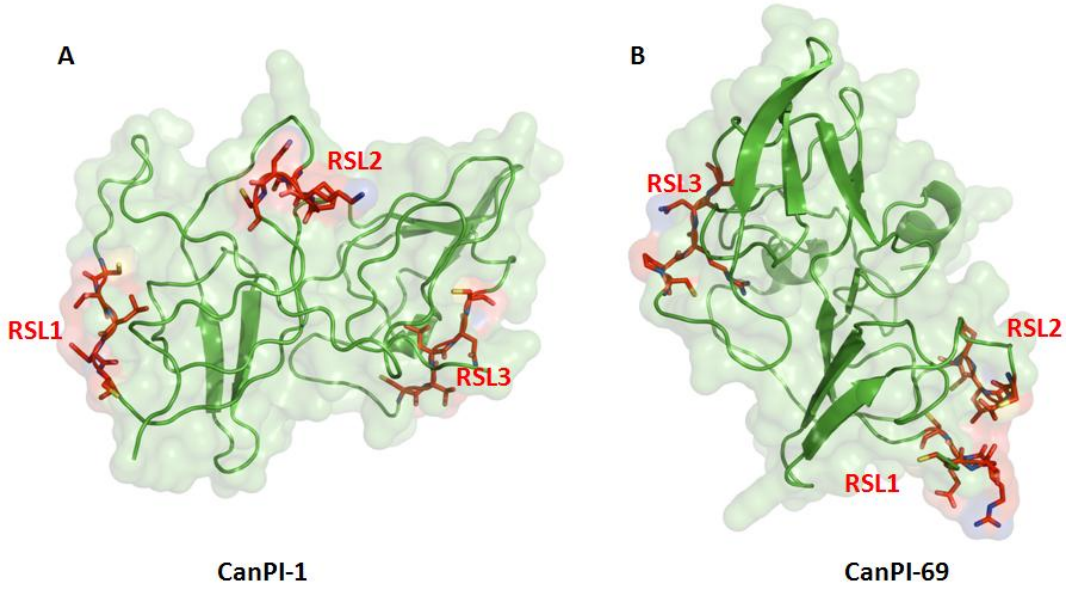


Figure 2B.7: Structure of (A) CanPI-1 and (B) CanPI-69. The predicted structures of CanPI-1 and -69 from I-TASSER server displayed β-sheets and unordered loops as the major secondary structural components and the three reactive sites of PIs displayed in red color (RSL1–RSL3) lied on exposed loops.

β-sheets and unordered loops mainly contribute to the secondary structures of both the CanPIs. CanPI-1 displayed three evenly distributed and solvent exposed reactive site loops (RSL1, 2 and 3) lying on unordered loops as shown in **Fig. 2B.7A**. As in other members of the Pin-II family, the reactive site loops are flanked by disulfide bonds formed by cysteine residues at the P3 and P2 positions (Cys-4to Cys-41 and Cys-8 to

Cys-37 in Domain I, Cys-62 to Cys-99 and Cys-67 to Cys-95 in Domain II and Cys-123 toCys-161 in Domain III). Structure of CanPI-69 showed two complete Pin-II scaffold and intact RSL (2 and 3) and partial RSL1 **(Fig. 2B.7B)**. Disulfide bonding network is dispersed in CanPI-69 and RSL2 and RSL3 is confined by disulfide bridges (Cys-62 to Cys-99 and Cys-67 to Cys-95 in Domain II, Cys-120 to Cys-157 and Cys-123 toCys-161in Domain III). Despite alteration in IRD organization, both CanPI-1 and CanPI-69 have almost similar 3D shape, which might suggest similar mode of geometrical association with proteases.

2B.4. Discussion

Stress accompanies metabolic reorganization in plants, which in turn aids in accumulation of defensive molecules. In *C. annuum*, elicitation of wound response leads to the systemic induction of significant amount of PI proteins. *CanPI* gene expression is regulated spatially, temporally as well as qualitatively and quantitatively upon different elicitation. Detailed investigation of PI genes, in *C. annuum*, revealed an array of different PI genes formed as a result of combinations of various unique IRDs. These genes, possessing 1 to 4 IRDs were characterized for their structural and functional diversity (Tamhane et al., 2009; Mishra et al., 2012). However, complete characterization of *CanPI* genes including promoter, 3' and 5' UTRs had not been performed. Sequence diversity in PI genes possibly contributes in recruiting a diverse pool of PI proteins to counter different biotic stresses and particularly dynamic gut proteases of the insects.

The *CanPI* genes exhibit a typical Pin-II type gene structure and display a remarkably conserved architecture with respect to the intron structure i.e. a single intron interjects the exon 1 and exon 2. Although the intron length of *CanPI* genes (type A- 294 bp and type B- 584 bp) was found greater than previously reported PIs (average length-100-200 bp) from Solanaceae, they retained their conserved features such as presence of GT….AG at the splicing motif and formation of glycine residue by the boundaries of the two exons **(Fig. 2B.3A)** (Kong and Ranganathan, 2008). The introns in CanPIs (type A and B) were found to be located within the signal peptide and rich in AT (A+T ~75%), which are the typical characteristics of Pin-II PIs (Keil et al., 1986; Zhang et al., 2004). In *CanPIs*, exon 1 contributes to 16 aa residues in the signal peptide while 9 to 13 aa

residues of the signal peptide and the IRDs constitute exon 2. Presence of diverse types of domain organization (beads on string or clasped bracelet) in CanPIs certainly raises a question about the evolutionary significance of Pin-II genes i.e. whether the occurrence of these domain organizations is sequential or they have evolved on divergent paths.

Several 3' and 5' UTRs of *CanPI* genes were obtained and they showed typical conserved characteristics. Usually, the 5'UTRs are found to be shorter than 3'UTR. The length of 5'UTR sequences in higher eukaryotes ranges from 20 to 80 nucleotides with an average of 60 nucleotides (Day and Tuite, 1998). The translation start sites in *CanPI-7* as well as *CanPI-69* retain the features such as "A" at -3 and "G" at +4 as described by Kozak, (2001). Transcript profiling has been adopted for comparing gene expression quantitatively. The characteristic distinctiveness of 3'UTR sequences has enabled the distinguishing among closely related transcripts (Eveland et al., 2008). The 3'UTRs of *CanPI* genes were distinct from each other and grouped into 3 clusters. Further functional validations will be required to examine whether the sequence variations among the three groups of UTR sequences influence the transcription and/or translation of the mRNA. In *Solanum tuberosum*, different PI gene families were found to exist in mixed clusters. Pin-II PI genes were observed to be placed as tandem repeats at the *StKI* locus on potato chromosome III. Moreover, this PI-rich region is interspersed with several transposable elements, which can be implicated in generating PI gene diversity (Odeny et al., 2010). Extensive accumulation of retroelements in *Capsicum* genome has previously been described (Park et al., 2012). Hence, it is compelling to speculate the existence of a similar physical organization, Pin-II genes interspersed with retro-elements, in *C. annuum*. This supposition can only be authenticated after an exhaustive genome-wide analysis of *CanPI* genes. Although *C. annuum* draft genome has been published recently, we could not find it suitable for further understanding of genomic architecture of *CanPIs*.

The understanding of *cis*-regulatory elements required for wound-inducible gene expression and interacting transcription factors is crucial for designing adept wound-inducible expression vectors, which can be used in insect resistant transgenics. Inducible character of PI genes and their roles in defense and endogenous functions necessitates the identification of the related promoter elements. Investigation of a tomato Pin-II promoter (accession number: AY129402) yielded several promoter elements such as WUN motif,

ABRE, ELI-box3, TCA-elements, which were responsive to wounding and plant hormones (Zhang et al., 2004). The characterization of another Pin-II promoter from *Solanum americanum* (accession number: AY749108) led to the identification of its role in trichome development. The study further resulted into the identification of elements such as MYB-binding motifs and L1 box (Liu et al., 2006). The PI gene promoters analyzed in the current study showed the presence of wide range of regulatory elements, which have role in hormonal, developmental, physiological and defense response. Detailed characterization of the promoter activity led to the identification of transcription factors like MYB, Myc, MADS-box, Dof type and so on bound to the promoter. These interactions suggest interplay of different regulatory elements in governing CanPI gene expression. Diverse interactions are believed to play a role in differential expression of these genes (Lorberth et al., 1992). Various pathways for the developmental and environmental regulation of Pin-II genes might exist (Pena-Cortis et al., 1991). Also, several components such as abscisic acid, jasmonic acid, salicylic acid and ethylene are reported to play role in wound signal transduction pathways *in planta* (Kim et al., 1992).

Lack of X ray and NMR structures of multidomain Pin-II inhibitors makes difficult to predict structure of CanPI-1 and CanPI-69. In general, 3D structure of both CanPI-1 and CanPI-69 showed conserved basic topology and shape. The structure of basic domain is similar to single domain Pin-II PI structure reported from *Nicotiana alata* (PDB:1TIH). Furthermore, CanPI-1 exhibited three solvent exposed and equi-spaced reactive loops on the disorder loop, which is a prerequisite for maximizing the possible interactions with multiple proteases. All the RSLs were flanked with disulfide bonds. Barrette et al., (2003a) reported the crystal structure of two domain tomato Pin-II inhibitor (PDB: 1PJU), which also showed equi-spaced solvent exposed RSL, confined with disulfide bridges. In case of CanPI-69, only two complete RSL flanked the disulfide bonds and one incomplete RSL was situated on disordered loops. Out of three RSL, two are well disturbed and third RSL was unstable due to absence of complete loop. In such scenario, there is a possibility that CanPI-69 can interact with maximum two proteases and hence it can possess less reactivity as compared to CanPI-1. Disulfide bridges are thought to hold the reactive site loop in a relatively rigid conformation that may help to prevent proteolytic cleavage of the inhibitor upon interaction with proteinases. Reduced

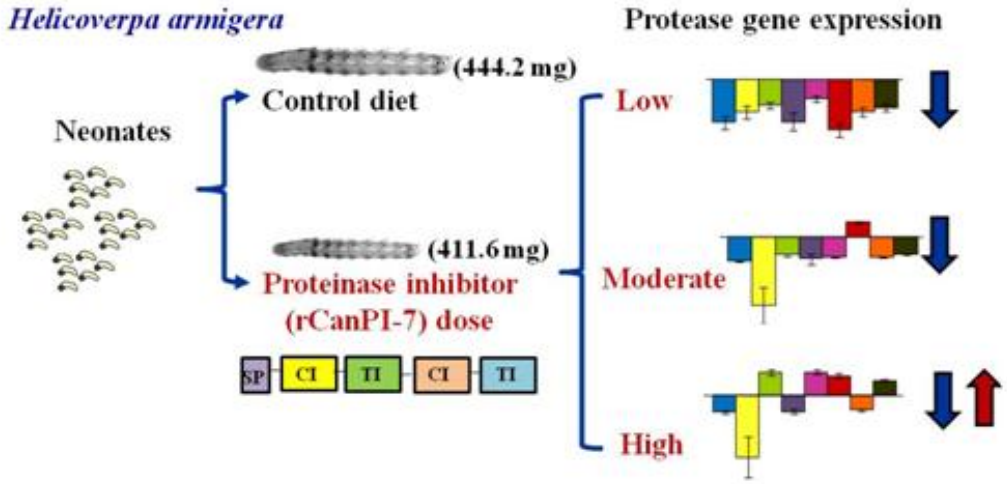
conformational flexibility in the loop is also thought to enhance proteinase binding by reducing entropic loss to achieve tighter binding. These structural aspects of protease and CanPI interactions suggest that PI molecules like CanPI-1 (beads on strings) are more favored in nature due to their higher reactivity against proteases as compared to CanPI-69 (bracelet like structure). Irregular processing of PI precursors can occur upon elicitation if the proteolytic machinery gets saturated (Horn et al., 2005). Under such conditions, above discussed structural attributes might affect the efficacy of PI processing. This can be the reason for expression of multiple wound induced PIs with beads on string like structure, since PIs with bracelet like structure are very few in *C. annuum*.

2B.5. Conclusion

The present study demonstrated the complete gene architecture of two variants of *CanPIs* (*CanPI-7* and *-69*). 3'UTR of *CanPI* genes showed remarkable diversity and distinctiveness. The investigation first time revealed the occurrence of clasped bracelet type PIs in *C. annuum*. The predicted structure of clasped bracelet type CanPI satisfied all structural attributes and found to be in accordance with the previous PI structures. Several regulatory elements were predicted in *CanPI* promoters and the binding of transcription factors was experimentally validated. The presence of diverse regulatory elements indicated a complex network of pathways for the developmental and environmental regulation of Pin-II PI genes in *C. annuum*. Altogether, the results obtained in the present study can be important in the development of PI based transgenic strategy to boost plant defense for insect control.

Chapter 3

Molecular response of *Helicoverpa armigera* upon exposure to multi-domain *Capsicum annuum* protease inhibitor-7



The research work described in this chapter is part of full-length paper, which has been published in Biochimica et Biophysica Acta 1830, 3414 -3420.

Chapter 3: Molecular response of *Helicoverpa armigera* upon exposure to multi-domain *Capsicum annuum* protease inhibitor-7

Chapter 3: Molecular response of *Helicoverpa armigera* upon exposure to multi-domain *Capsicum annuum* protease inhibitor-7

Abstract

A multi-domain Pin-II type protease inhibitor from *Capsicum annuum* (CanPI-7) is known to be effective against the insect pest, *Helicoverpa armigera*. The present study is an attempt to investigate the optimal dose of recombinant CanPI-7 (rCanPI-7) for effective antibiosis to *H. armigera* and further to characterize the responses of digestive proteases upon rCanPI-7 ingestion. The gut protease activity was assessed biochemically and transcript accumulation pattern for selected trypsin and chymotrypsin genes was analyzed by quantitative Real-Time PCR. The growth retardation upon exposure to rCanPI-7 was more prominent in neonates as compared to third instar larvae. Influence of dosage of rCanPI-7 was noticeable on the expression and regulation of candidate trypsin and chymotrypsin genes in *H. armigera*. The transcript accumulation pattern correlated with the protease activity in rCanPI-7 exposed larvae. Moreover, it is also evident that the regulation and expression of *H. armigera* gut proteases are specific to the stage of PI exposure. We conclude that despite the large diversity and plasticity in the expression of protease genes in *H. armigera*, early exposure and specific dose of rCanPI-7 is effective against it. An overall upregulation of the energy metabolism and redox processes was shown in unbiased proteomic analysis. These results highlight the requirement of optimal PI concentration and timely exposure for effective growth retardation and for inhibiting the major gut proteases of *H. armigera*.

3.1. Introduction

Plant protease inhibitors (PIs) are well established as plant defense proteins. PIs bind to and inhibit the activity of insect gut proteases eventually leading to deficient amino acid supply to the insects. However, insect feeding assays with various PIs have demonstrated differential inhibitory potential of the PIs towards individual insects (Johnson et al., 1989; Broadway, 1995; 1997; Bown et al., 2004; Hartl et al., 2010). The growth inhibitory potential of PIs is a function of (i) feeding behavior of the insect, (ii) nature of the PI and (iii) its stability in the insect gut (Jongsma and Bolter, 1997).

Lepidopteran insects display multiple strategies to deal with the PIs. Expression of inhibitor-insensitive proteases, overexpression of constitutive proteases, and induction of proteases capable of degrading the PI or switching to enzymes with altered substrate specificity such as the upregulation of chymotrypsin-like activity to counter trypsin inhibitor are the mechanisms reported (Jongsma and Bolter, 1997). Evidences of insect adaptation towards various PIs suggest that any candidate gene should be stringently validated by analyzing the insect responses. In order to test the efficacy of a PI, primary feeding trials should be undertaken along with the investigation of molecular responses of the insect. These can then be followed by controlled field trials.

One of the major lepidopteran pests, *Helicoverpa armigera* possesses a gut protease complement predominantly consisting of trypsin-like (90%) and chymotrypsin-like (5%) activities (Patankar et al., 2001; Srinivasan et al., 2006). Till date 49 complete or partial serine protease gene sequences have been reported from *H. armigera*. Depending on the developmental stage and diet composition, *H. armigera* is known to regulate gut protease expression. This not only encourages polyphagy but also enables insects to sustain the challenges posed by dietary PIs (Bown et al., 1997; Srinivasan et al., 2006).

Pin-II type PIs of *Capsicum annuum* (CanPIs) are composed of single or multiple inhibitory repeat domains (IRDs), each with trypsin or chymotrypsin inhibitory (TI or CI) specificity. In earlier studies, we have compared *in vitro* enzyme inhibition and *in vivo* growth inhibition potential of four recombinant CanPIs (rCanPIs) against *H. armigera*. rCanPI-7 (with two variant TI and CI IRDs each) was the most potent in retarding the *H. armigera* growth and development (Tamhane et al., 2007). Individual IRDs released after

proteolytic processing in the insect gut could stably and efficiently inhibit the *H. armigera* gut proteases *in vitro* (Mishra et al., 2010).

The present study is an attempt to detail the responses of *H. armigera* upon ingestion of rCanPI-7. We attempt to address the following questions (i) What is the effect of varied doses of rCanPI-7 on larval growth progression? (ii) Does the response offered vary with the larval stage of exposure? (iii) How does the ingestion of varied doses of rCanPI-7 affect the gene expression and enzyme activities of digestive proteases in *H. armigera*? (iv) What is the nature of rCanPI-7 regulated proteases? and (v) What concurrent proteomic changes occur in response to rCanPI-7 ingestion? The study may provide validation for the use of rCanPI-7 in the development of insect resistant transgenic plants.

3.2. Materials and Methods

3.2.1. Insect feeding bioassays

Neonates obtained from laboratory culture of *H. armigera* maintained at 27 ± 2 °C, 60 ± 5% relative humidity were used to carry out the bioassays. The artificial diet (AD) was prepared as described by Nagarkatti and Prakash (1974). Based on the earlier reports (Tamhane et al., 2007; 2005) three concentrations of rCanPI-7 viz., 0.01% (100 µg-low), 0.015% (150 µg-moderate) or 0.03% (300 µg/-high) per gram of AD were incorporated for the feeding bioassays. Each larva was maintained in an individual vial containing control or rCanPI-7 incorporated diet separately. Diet was added as needed and larval mass was recorded every alternate day. For each bioassay, the harvested tissue was snap frozen in liquid nitrogen and stored at −80 °C until further use.

3.2.1.1. Bioassay for dose dependent response

Neonates were fed on control (AD) and rCanPI-7 (0.01%, 0.015% or 0.03%) incorporated artificial diets. Each set containing 100 insects was maintained. Insect tissue was harvested at 4 and 16 days post-exposure. Owing to the very small size of larvae, at day four whole insects were harvested while for later time point (day 16) insects were dissected and whole gut tissue was collected. Equal amount of RNA was used as basis of comparison of the expression patterns between the whole larvae and whole gut.

3.2.1.2. Bioassay for early and late response

Neonates were reared on control diet till they completed the second instar and thereafter the third instar larvae were exposed to rCanPI-7 (0.01%) incorporated diet. Whole gut tissue was harvested by dissecting the insects at variable time points (3, 6, 12, 24, 48 and 96 h) from rCanPI-7 fed as well as the control AD diet fed sets.

3.2.1.3. Bioassay for stage specific response

Neonates were reared on control or rCanPI-7 (0.01%) incorporated diet at variable stages namely days 2, 4, 6 and 8. Larvae were fed on rCanPI-7 incorporated diet for 24 h and whole insects were harvested at the end of the exposure.

3.2.2. HGP extraction and protease activity assays

For preparation of *Helicoverpa* Gut Protease (HGP) 100 mg of gut tissue was homogenized in 100 µl of 0.2 M glycine- NaOH buffer, pH 10, and allowed to stand for 2 h at 4°C. The gut luminal contents were removed by centrifugation at 10,000*g* for 15 min at 4°C. The supernatant was used for protease activity assays and in-gel activity visualization. The proteolytic activity was determined using substrate Azocasein (Sigma-Aldrich Co.). For azocaseinolytic assay 60 µl of diluted enzyme was added to 200 µl of 1% azocasein (in 0.2 M glycine– NaOH, pH 10.0) and incubated at 37°C for 30 min. The reaction was terminated by the addition of 300 µl of 5% trichloroacetic acid. After centrifugation at 14,230*g* for 10 min, an equal volume of 1 M NaOH was added to the supernatant and absorbance was measured at 450 nm. One proteinase unit was defined as the amount of enzyme that increased the absorbance by 1.0 OD under the given assay conditions (Patankar et al., 2001).

3.2.3. In-gel activity visualization of proteases

Amount of HGP corresponding to equal (azocaseinolytic) activity units (0.4 U) was resolved on 8% native-PAGE. The gel was further processed for activity visualization using the gel X-ray film contact print technique (GXCT). GXCT involves incubation of the gel post-electrophoresis on X-ray film at 37°C. The proteases

immobilized in the gel hydrolyze the gelatin coated on the X-ray film. After removing the gel, the X-ray film was washed with warm water (Harsulkar et al., 1998). To analyze the sensitivity of larval gut proteases, 0.4 *H. armigera* gut protease inhibitory units (HGPIUs) of rCanPI-7 were incubated, *in vitro* with 0.4 U of HGP for 20 min at 24 °C preceding GXCT. To explore the trypsin/chymotrypsin-like nature of the protease isoforms visualized, HGP (0.4 U) was incubated with synthetic inhibitors of chymotrypsin and trypsin, N-tosyl-L-phenylalanine chloromethyl ketone (TPCK) (100 μM) or Na-r-tosyl-L-lysine chloromethyl ketone (TLCK) (100 μM) (Sigma, St Louis, MO, USA), respectively (Telang et al., 2005) and protease profiles were visualized by GXCT.

3.2.4. RNA isolation, preparation of cDNA and expression analysis by real time PCR

Total RNA was isolated from the insect tissues using Trizol reagent (Invitrogen, Carlsbad, CA, USA) based on the manufacturer's protocol. Total RNA was treated with RNase-free DNAase I (Promega). The integrity and quantity of RNA were determined by agarose gel electrophoresis and using Nanodrop (Thermo Scientific), respectively. Synthesis of the first strand cDNA was carried out in 20 μl reactions using 2 μg of total RNA with High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Foster, CA, USA) using random primers. Relative transcript abundance of trypsins and chymotrypsins was determined by quantitative Real-Time PCR (qRT-PCR) using 7900HT Fast Real-Time PCR System (Applied Biosystems, Foster, CA, USA) and Faststart Universal SYBR Green Master (Rox) 2× concentrate (Roche Diagnostics, GmbH, Germany). The relative expression of trypsin and chymotrypsin genes (*HaTry 1*, *2*, *3*, *4*, *5*, *6*, *7*, *8*; *HaChy 1*, *2*, *3*, and *4*) was assessed. Accession numbers and details of primers are mentioned in Appendix I Table 5. The trypsin or chymotrypsin genes examined were selected on the basis of phylogenetic analysis of serine protease genes from *H. armigera* reported in NCBI. The sequences analyzed grouped into several clades based on their sequence similarity, at least one representative from each clade was chosen for analysis (Chikate et al., 2013). For each gene, amplification efficiency was assessed by constructing a standard plot using 4 serial dilutions of cDNA pool prepared by combining aliquots from all the cDNA samples under study. α-actin (Accession no.:

AF286059) was used as a reference gene for normalization. Real-time PCR was performed in 10-µl reactions containing 5 µl of 2X concentrate SYBR mix (Fast Universal SYBR Green, Roche, Berlin, Germany), 0.5 µl of forward and reverse primer each (10 µm, i.e. 500 nm in reaction) and 1 µl of cDNA (10 ng) template. The cycling parameters used were, 95 °C for 10 min, followed by 40 cycles of 95 °C for 3 s and 60 °C for 30 s. At the end of each run, dissociation curve analysis of the amplified product was carried out to evaluate the specificity. This involved denaturation at 95 °C for 15 s, cooling to 60°C for 15 s and then gradual heating at 0.01 °C/s to a final temperature of 95 °C. For each treatment/time point, two biological replicates (each constituting three technical replicates), the average transcript abundance and subsequent fold difference with respect to the control were calculated.

3.2.5. Statistical analysis

Average larval mass and the transcript abundance for trypsin and chymotrypsin genes for the larvae fed on rCanPI-7 (0.01, 0.015 or 0.03%) diet were compared to those of control diet fed larvae using Student's t-test (MS Excel). p values obtained were used to assign significance $p > 0.05$ (non-significant), $0.01 < p < = 0.05$ (c), $0.001 < p < = 0.01$ (b) and $p < = 0.001$ (a).

3.2.6. Proteomic analysis

Larval gut tissue from AD and rCanPI-7 diet fed fourth instar larvae was used for proteomic analysis to identify the differentially accumulated proteins. Proteins from gut and frass were extracted according to Schuster and Davies (1983) with slight modifications. Tissue (500 mg) was ground using a mortar and pestle in liquid nitrogen, mixed with 20 ml extraction buffer (0.7 M sucrose; 0.5 M Tris; 30 mM HCl; 50 mM EDTA; 0.1 M KCl; 2% [v/v] mercaptoethanol) and vortexed for 10 min. The homogenate was centrifuged at 4 °C for 13,000*g* and the supernatant was transferred to a new falcon tube, to this an equal volume of water saturated phenol was added. Phenol phase was re-extracted with an equal volume of extraction buffer by shaking for 10 min and centrifugation at 13,000*g* at 25 °C. The re-extracted phenol phase was precipitated with 5 volumes of 0.1 M ammonium acetate in methanol at 20 °C overnight. The precipitate was

washed thrice with 0.1 M ammonium acetate in methanol and once with 100% acetone. The pellets were air-dried and resuspended in lysis buffer containing 8 M urea, 2 M thiourea, 4% (v/v) CHAPS and 50 mM DTT. The resulting suspension was centrifuged and stored at 80 °C for later use.

Sample proteins (80 µg each) were loaded on 12% SDS polyacrylamide constant separation gel with a 4% stacking gel and electrophoresed at 25 °C using a vertical electrophoresis apparatus. Protein bands were excised from the CBB-stained gels, washed twice with milli-Q water. In-gel digestion was performed as mentioned in section 2A.2.4. Samples were analyzed by LC-MS^E on SYNAPT High Definition Mass spectrometer (Waters Corporation) as mentioned in section 2A.2.5. After each injection, peptides were eluted into the NanoLockSpray ion source at a flow rate of 300 nL/min using a gradient of 2-40% B over 50 min. To analyze the differential expression label-free quantification was performed using BSA (100 fmol) as an external standard. All samples were run in triplictaes and the average of the three was considered to calculate the fold difference. Protein Lynx Global server (PLGS) browser version 2.1.5 software (both from Waters Corporation, USA) were used for database searches (section 2A.2.5). The Lepidoptera database from Uniprot was used for database search. Gene ontology annotation of the identified proteins was obtained from UniProt Knowledgebase (www.uniprot.org).

3.3. Results

3.3.1. Dose dependent responses of H. armigera to rCanPI-7

Significant reduction in larval mass gain was observed till day 14 with all the three doses of the inhibitor **(Fig. 3.1)**. Ingestion of rCanPI-7 led to delayed larval development; low and high doses caused a delay of 2 days while moderate dose caused a delay of 3 days with respect to larval instar, in comparison to that of control diet fed insects.

Figure 3.1:

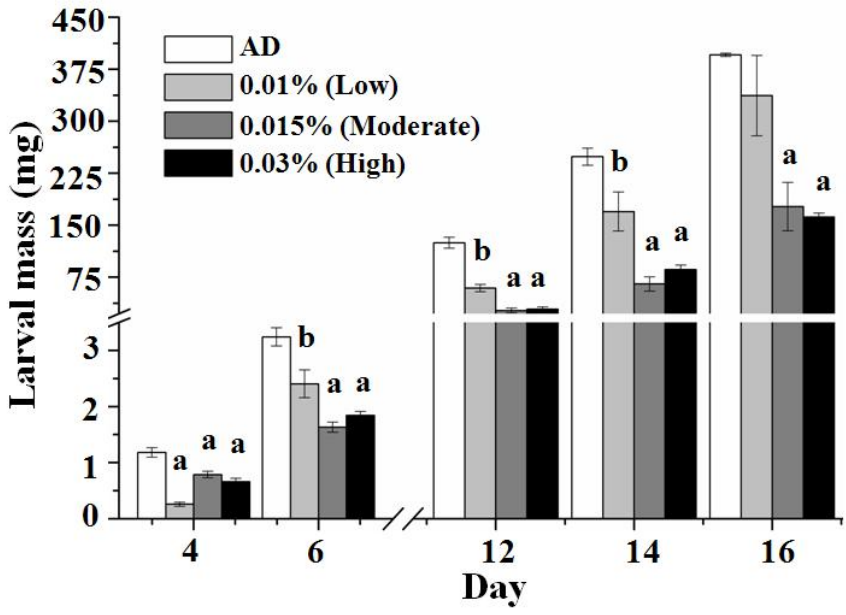


Figure 3.1: *H. armigera* larval growth since neonate stage upon continuous feeding with rCanPI-7 (0.01%, 0.015% and 0.03% w/w) incorporated diet. Antibiosis effect of rCanPI-7 on insects is evident by the reduced larval mass gain in rCanPI-7 diet fed insects.

The available/residual proteolytic activities, at days 4 and 16, were higher in the larvae fed on high dose as compared to those in control diet (by at least 1.5-fold). In case of moderate dose fed larvae, the protease activity was comparable to control at day 4, while it was significantly lower than that in control at day 16. Larvae fed with low dose of rCanPI-7 showed comparable protease activity at day 4 and 16 with respect to control **(Fig. 3.2A)**. Protease profiles with equal activity units for the larvae fed on, moderate and high dose inhibitor diet at 16 days showed reduced intensity of several HGP isoforms (HGP 3, 4, 5, 6, and 9). This indicates reduction in the available/residual protease activity in the larval gut **(Fig. 3.2B)**. Further *in vitro* incubation of gut proteases with rCanPI-7 followed by activity visualization, demonstrated inhibition of most of the HGP isoforms, except for HGP 8 **(Fig. 3.2B)**. Nonetheless, the protease activity induced after ingestion of rCanPI-7 remains sensitive to the PI but the levels might be maintained *in vivo* by overexpression. No new HGP isoforms were visualized in the 1D activity gel; however possibility of the occurrence of additional protease isoforms cannot be neglected.

Figure 3.2:

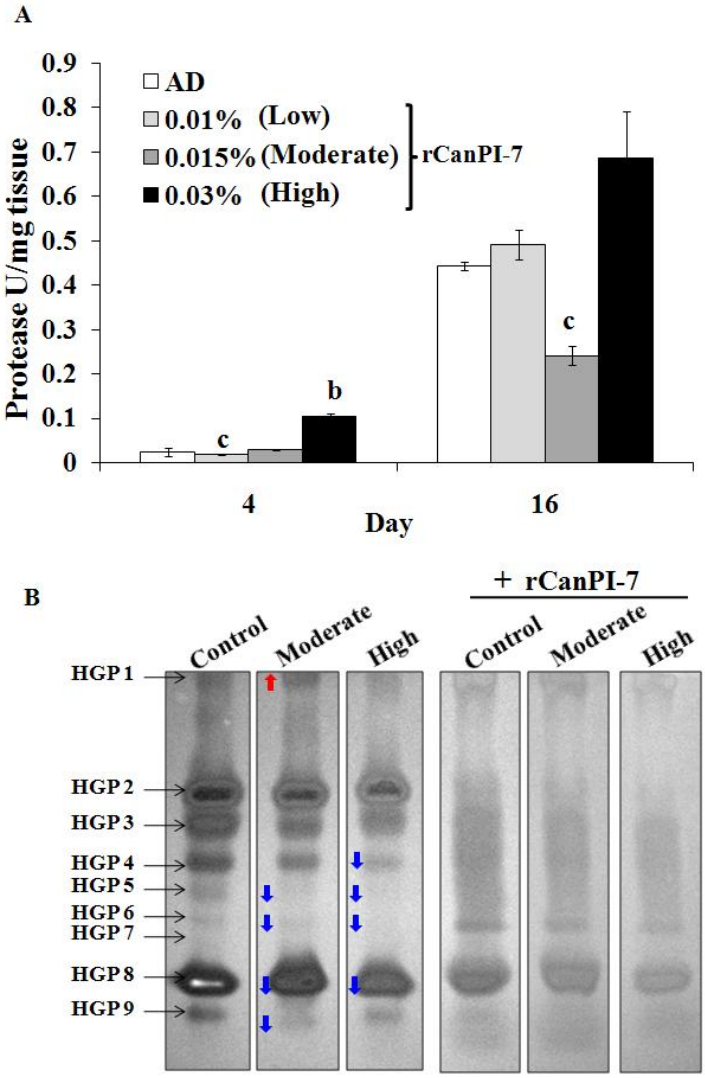


Figure 3.2: (A) Total available proteolytic activity (U/mg tissue) at days 4 and 16 upon continuous exposure to rCanPI-7 (0.01%, 0.015% and 0.03%) incorporated diet since neonate stage. At day 4, the available proteolytic activity is significantly elevated in high dose fed larvae. At day 16, the available proteolytic activity in 0.03% (high) and 0.01% (low) dose fed larvae is comparable to that of control but that in (0.015%) moderate dose diet fed larvae it is depleted. (B) Gut protease profiles (equal units) of larvae exposed to rCanPI-7 (0.015% and 0.03%) incorporated diet since neonate stage, at day 16. Diminished isoform intensity is detected in case of 0.015% (moderate) and 0.03% (high) dose fed larvae. Visualization of HGP isoforms from control, moderate and high dose diet fed larvae after *in vitro* incubation with rCanPI-7, indicates sensitivity of most of induced isoforms to rCanPI-7.

Figure 3.3:

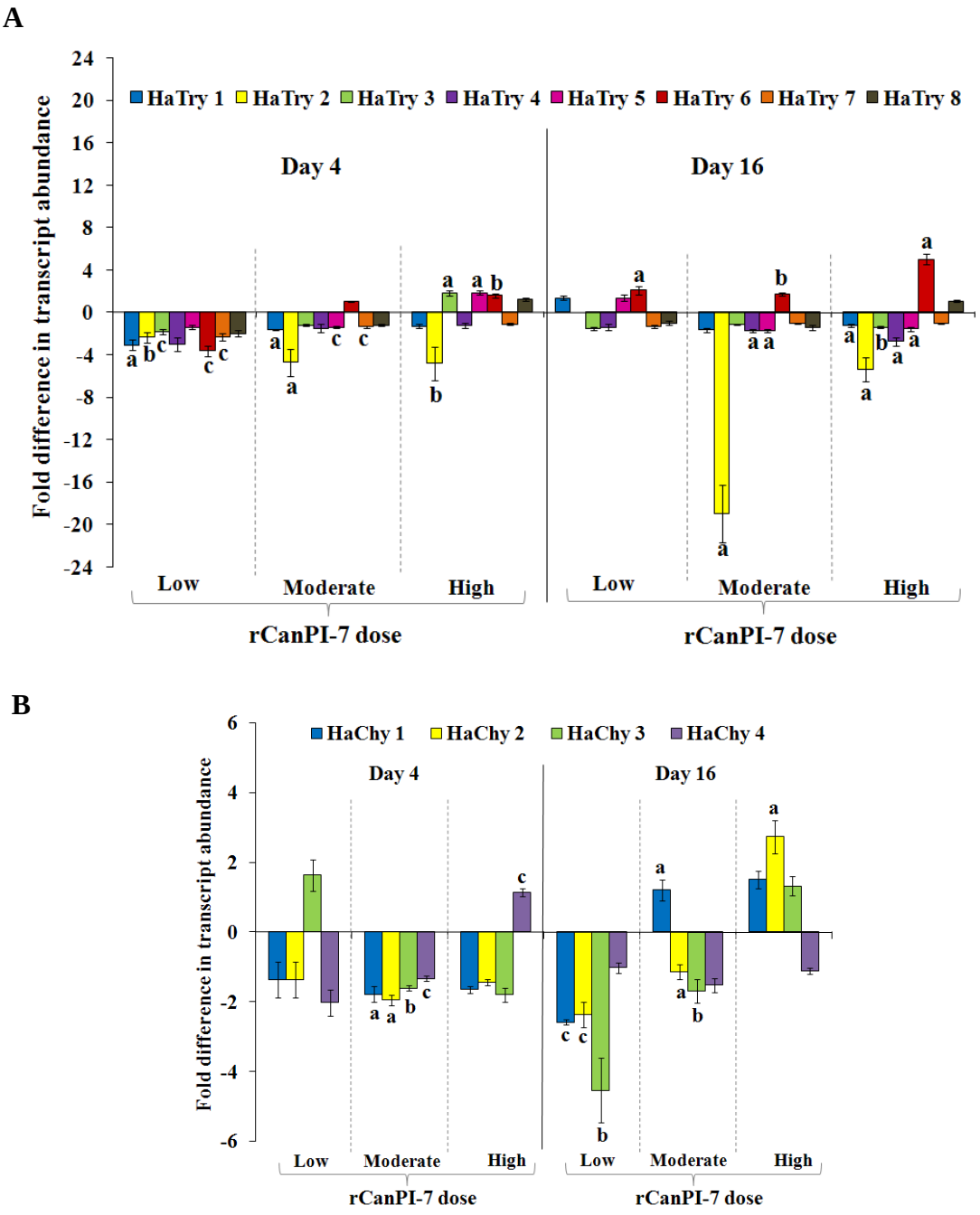


Figure 3.3: Fold difference in transcript abundance of trypsins and chymotrypsins for rCanPI-7 (0.01%, 0.015% and 0.03%) incorporated diet fed larvae, exposed at neonate stage. (A) Fold difference in transcript abundance of trypsin genes at days 4 and 16. (B) Fold difference in transcript abundance of chymotrypsin genes at days 4 and 16. Downregulation of trypsin and chymotrypsin gene expression is observed in case of 0.01% (low) and 0.015% (moderate) dose diet fed larvae but upregulation of the same is detected in case of 0.03% (high) dose diet fed larvae.

After 4 days of exposure, trypsin genes namely, *HaTry 1*, *2*, *3*, *4*, *5*, *7* and *8* were downregulated in case of low and moderate dose diets while only *HaTry 6* appears upregulated in moderate dose diet fed larvae. In case of high dose diet, after 4 days of exposure, upregulation of *HaTry 3*, *5*, *6* and *8* was evident with down regulation of *HaTry 1*, 2, *4* and *7* **(Fig. 3.3A)**. *HaTry 6* which displayed a 3-fold downregulation in low dose diet fed larvae was 1.6-fold upregulated in high dose diet fed larvae. Upregulation of *HaChy 3* and *HaChy 4* was evident in low and high doses, respectively at day 4 while other chymotrypsins showed consistent downregulation **(Fig. 3.3B)**. *HaTry 1*, *2*, and *7* and *HaChy 1* and *2* were downregulated in low, moderate and high dose diet fed larvae by at least 1.5 fold.

At day 16, the trypsin gene transcripts from the moderate and high dose fed larvae revealed similar patterns of expression **(Fig. 3.3A)**. Transcripts of genes *HaTry 3*, *4*, *7* and *8* were downregulated in low dose, *HaTry 1*, *2*, *3*, *4*, *5*, *7* and *8* were downregulated in moderate dose and HaTry 1, 2, 3, 4, 5 and 7 were downregulated in high dose. *HaTry 2* was downregulated by 18-fold in case of moderate dose while 5-fold in case of high dose fed larvae. HaTry 6 showed upregulation up to 2-fold in low and moderate doses whereas 6-fold in case of high dose diet fed larvae. Among chymotrypsins, *HaChy 1* was upregulated and *HaChy 2* and *3* were downregulated in moderate dose diet fed larvae while in high dose diet fed larvae all the three chymotrypsins exhibited upregulation with the highest transcript abundance for *HaChy 2* (3-fold) **(Fig. 3.3B)**.

On comparing the relative transcript abundance at days 4 and 16 in case of moderate and high dose diet fed larvae, *HaTry 6* was significantly upregulated (Fig. 3.3A). In the larvae fed on moderate dose diet, all chymotrypsin gene transcripts except *HaChy1* remained downregulated from days 4 to 16 **(Fig. 3.3B)**. In high dose diet fed larvae at day 16, *HaChy 1*, *2* and *3* were upregulated while *HaChy 4* gene transcripts displayed reduced abundance **(Fig. 3.3B)**.

3.*3.2. Early and late responses of H. armigera to rCanPI-7*

Third instar *H. armigera* larvae fed on rCanPI-7 (low dose) incorporated diet continuously for 7 days showed no reduction in larval mass gain as compared to the control diet fed larvae **(Fig. 3.4A)**. Significantly high gut proteolytic activity was detected at 24 h post exposure which subsequently reduced at 48 h **(Fig. 3.4B)**. Protease profiles visualized after *in vitro* incubation with rCanPI-7 showed inhibition of majority of the HGP isoforms, except HGP 8. Differences in protease profiles of control diet and rCanPI-7 fed larvae were identified by their sensitivity to synthetic inhibitors of trypsin, TLCK and chymotrypsin, and TPCK **(Fig. 3.4C)**. Inhibition of trypsins by TLCK facilitated the visualization of chymotrypsins in the HGP. The absence of isoform HGP 3 in HGP from rCanPI-7 fed larvae indicated *in vivo* inhibition of the chymotrypsin-like isoform in the PI fed larvae. Similarly, diminished activity of isoforms HGP 4 and 5 in HGP from rCanPI-7 fed larvae reveals *in vivo* inhibition of trypsins by rCanPI-7. Presence of all the isoforms after TPCK treatment of HGP from AD or rCanPI-7 fed larvae, affirmed the presence of trypsin isoforms as major proteases in the gut of *H. armigera.*

Transcripts of genes *HaTry 2, 3, 5,* and *6* and *HaChy 1, 2, 3,* and *4* showed high abundance at as early as 3 and/or 6 h, up to 12 h **(Fig. 3.5A)**. All the chymotrypsins under consideration were upregulated within 24 h, except *HaChy 4* and further downregulated gradually by 96 h **(Fig. 3.5B)**. Reviewing the overall trend, most trypsin and chymotrypsin genes under the present study showed an upregulation till 24 h of rCanPI-7 exposure and thereafter downregulation of trypsin and chymotrypsin genes was noticed.

Fig. 3.4:

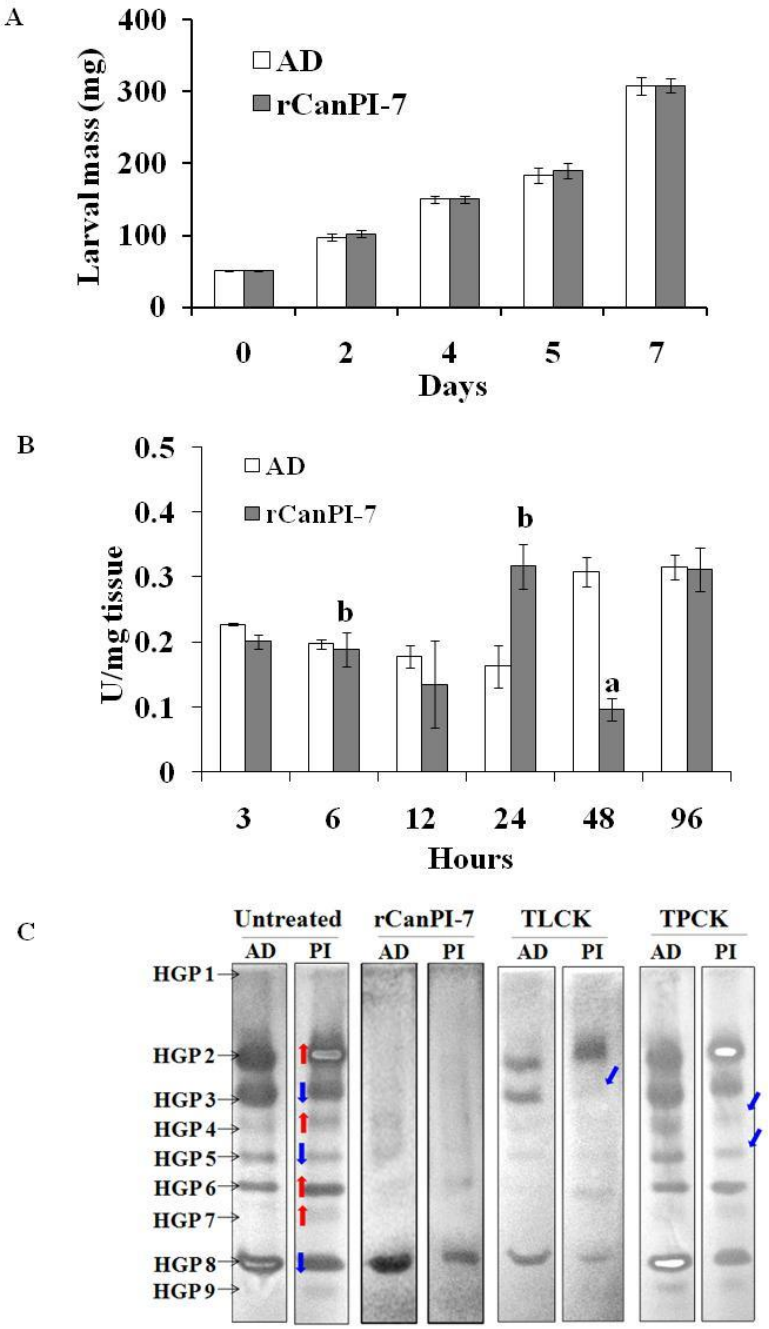


Fig. 3.4: (A) Growth of *H. armigera* larvae third instar onwards when fed with rCanPI-7 (0.01%) incorporated diet. "Day 0" indicating larval mass before exposure. (B) Total available proteolytic activity (U/mg), at different time points post-exposure (3, 6, 12, 24, 48 and 96 h), for control and rCanPI-7 (0.01%) fed larvae, when exposed at the third instar. (C) Gut protease profiles (native PAGE) of larvae exposed to rCanPI-7 (0.01%) at the third instar, after 4 days of exposure. The gut proteases visualized after *in vitro* incubation with rCanPI-7, TLCK and TPCK show differential isoforms.

Fig. 3.5:

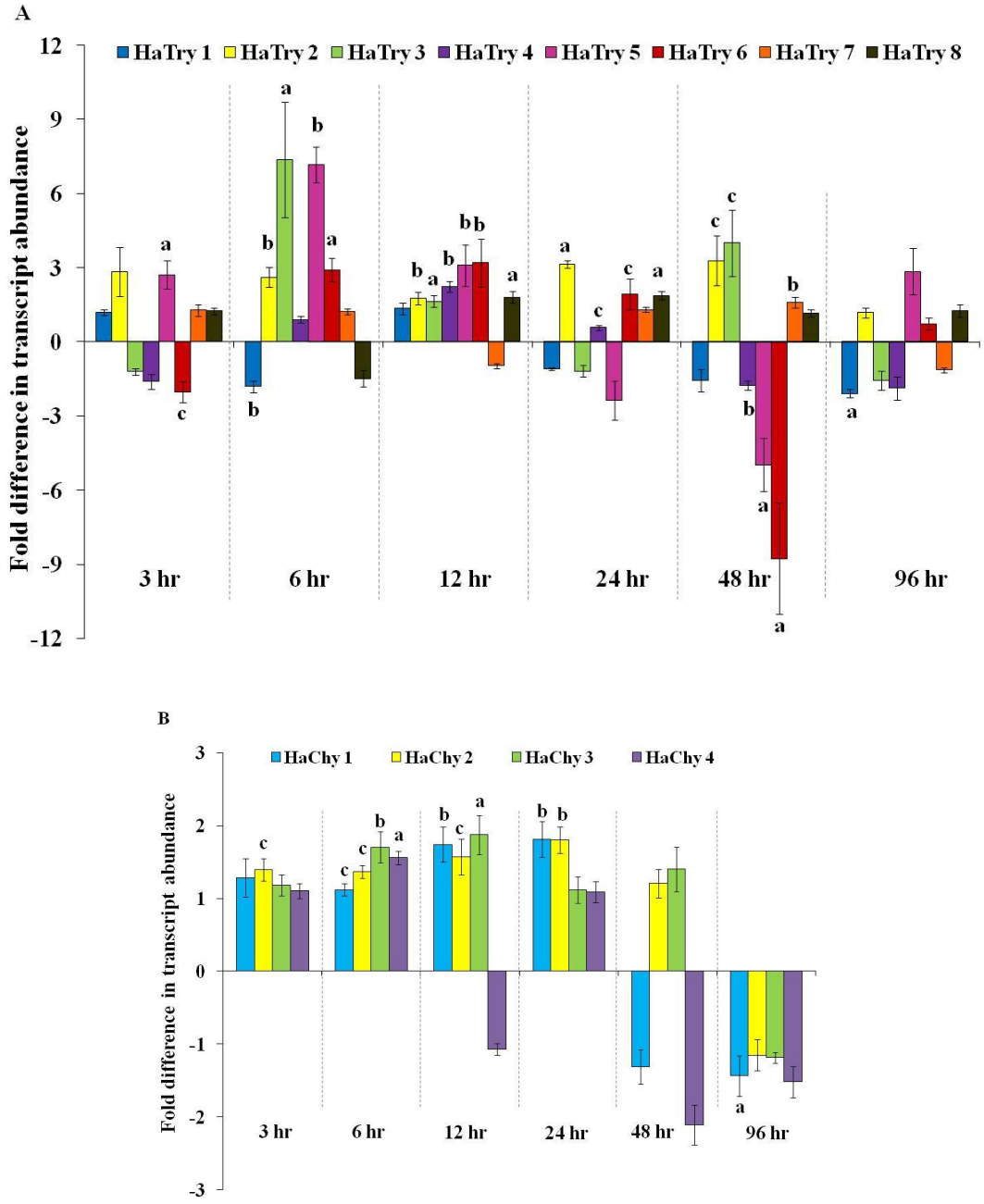


Fig. 3.5: Fold difference in transcript abundance of (A) trypsin and (B) chymotrypsin genes, respectively, at different time points post-exposure (3, 6, 12, 24, 48 and 96 h), for control and rCanPI-7 (0.01%) fed larvae, when exposed at the third instar. Initial upregulation of trypsins and chymotrypsins followed by gradual stabilization is evident.

3.3.3. Stage-specific response of the larvae at initial instars

Subjecting the early stage larvae to acute exposure of rCanPI-7 did not affect the proteolytic activity significantly **(Fig. 3.6)**.

Fig. 3.6:

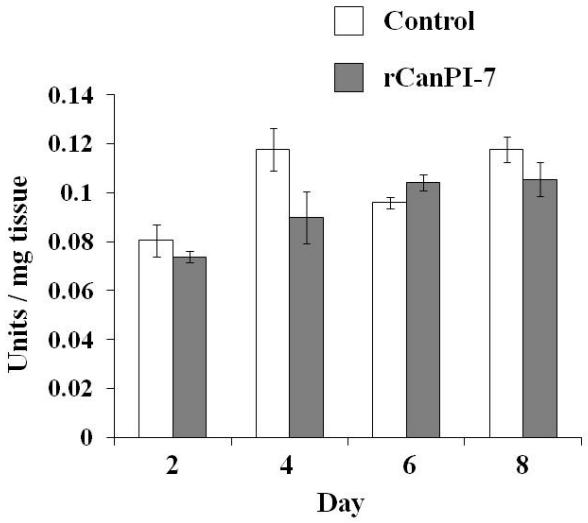


Fig. 3.6: Proteolytic activity (U/mg), upon acute (24 h) exposure at different stages (Day 2, 4, 6 and 8), for AD and rCanPI-7 (0.01%) fed larvae.

Transcript profile of trypsin and chymotrypsin genes demonstrated intriguing responses of the early stage larvae **(Fig. 3.7A and B)**. Upon exposure at day 2, upregulation of *HaTry 1*, *2*, *3*, *4*, *5*, *7*, and *8*, and *HaChy 1*, 2 and *4* was evident. Exposure at day 4 resulted into the upregulation of *HaTry 1*, *3*, *4*, *6*, and *8* and *HaChy 1*, *2*, and *3* whereas that at day 6 caused the upregulation of *HaTry 1*, *3*, *4*, and *6* and *HaChy 1*. Upon exposure at day 8, all the trypsin/chymotrypsin genes under study, except *HaTry 3* were downregulated by at least 2-fold.

Fig. 3.7:

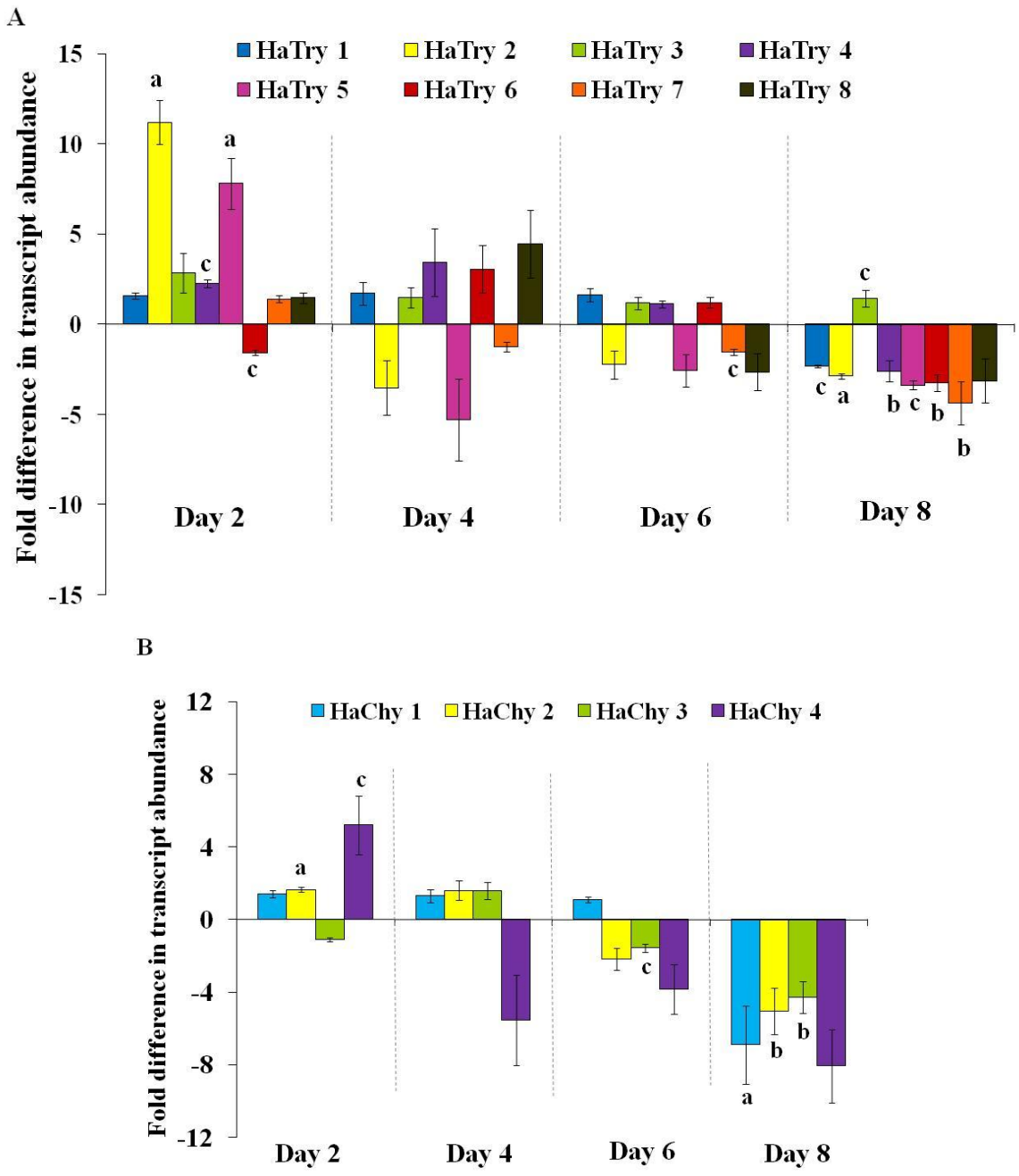


Fig. 3.7: Fold difference in transcript abundance of (A) trypsins and (B) chymotrypsins, respectively, upon acute (24 h) exposure at different stages (Day 2, 4, 6 and 8), for control and rCanPI-7 (0.01%) fed larvae. Varied responses are presented depending upon the stage of exposure.

3.3.4. Proteomic analysis of the response of H. armigera

In proteomic analysis, over 200 proteins were identified and 41 of these consistently showed differential abundance across the replicates. Figure 3.8 shows the SDS-PAGE profiles of larval gut proteins and Table 3.1 lists the fold change (above 1.4) of differentially accumulated proteins. The proteins identified were related to energy metabolism, digestion, redox processes and detoxification.

Figure 3.8:

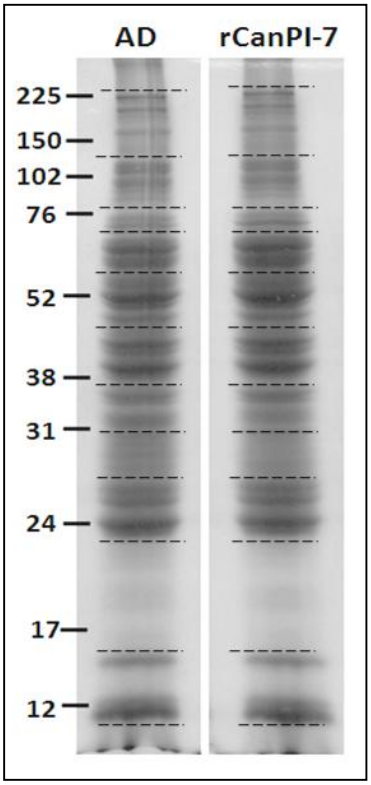


Figure 3.8: The SDS-PAGE protein profile of the larval gut. AD and rCanPI-7 indicate the larval gut proteins from AD and rCanPI-7 diet fed insects. Dotted lines indicate the gel slices excised for tryptic digestion.

Table 3.1: Differentially accumulated proteins in *H. armigera* larvae upon rCanPI-7 ingestion. GO: Gene ontology; BP: Biological process; CC: Cellular component; MF: Molecular function.The arrows indicate up (↑) or down (↓) regulation.

Accession	Description	Up/Down	Fold change	GO
Q86CZ0	ADP ATP translocase OS *Helicoverpa armigera* PE 2 SV 2	↑	3.1	BP,CC,MF
B4Z1D7	Alkaline phosphatase OS *Helicoverpa armigera* PE 2 SV 1	↑	1.7	BP,MF
B6CMF1	Aminopeptidase N5 OS *Helicoverpa armigera* PE 2 SV 1	↑	2.7	BP,MF
D3Y4D1	Arginine kinase OS *Helicoverpa armigera* PE 2 SV 1	↑	1.8	MF
D5G3E6	Carboxyl choline esterase CCE014a OS *Helicoverpa armigera* PE 2 SV 1	↑	1.5	MF
Q6H962	Carboxypeptidase OS *Helicoverpa armigera* GN ca42 PE 2 SV 1	↑	1.6	BP,MF
H9BEW3	Catalase OS *Helicoverpa armigera* PE 2 SV 1	↑	1.5	BP,MF
O18438	Chymotrypsin like protease OS *Helicoverpa armigera* PE 2 SV 1	↑	1.8	BP,MF
H9ITU2	Citrate synthase OS *Bombyx mori* GN Bmo 2204 PE 3 SV 1	↑	1.5	BP,MF
B6CMF9	Fatty acid binding protein 2 OS *Helicoverpa armigera* PE 2 SV 1	↑	1.4	MF
B6CMG0	Fatty acid binding protein 3 OS *Helicoverpa armigera* PE 2 SV 1	↑	1.9	MF
O76515	Fatty acid binding protein OS *Helicoverpa zea* PE 2 SV 1	↑	1.9	MF
F5BYI0	Heat shock protein 105 Fragment OS *Helicoverpa armigera* PE 2 SV 1	↑	2.1	BP,MF
C7SIR9	Heat shock protein 70 OS *Helicoverpa armigera* PE 2 SV 1	↑	2	BP,MF
A9X7K9	Lipase OS *Helicoverpa armigera* PE 2 SV 1	↑	1.6	BP,CC,MF
H9JU51	Malate dehydrogenase OS *Bombyx mori* PE 3 SV 1	↑	2.8	BP,MF

Q2F5N1	Mitochondrial prohibitin complex protein 2 OS *Bombyx mori* GN Bmo 589 PE 2 SV 1	↑	2.1	CC
Q1HPR4	NADP dependent oxidoreductase OS *Bombyx mori* GN Bmo 81 PE 2 SV 1	↑	1.6	MF
Q2F5J2	Prohibitin protein WPH OS *Bombyx mori* GN Bmo 4294 PE 2 SV 1	↑	2.9	CC
B1NLE3	Protease OS *Helicoverpa armigera* GN SerProx 4 PE 2 SV 1	↑	4.5	BP,MF
K7NSY9	Sterol carrier protein 2 3 oxoacyl CoA thiolase OS *Helicoverpa armigera* GN SCPx PE 2 SV 1	↑	1.5	BP,MF
Q5XUN5	Triosephosphate isomerase OS *Helicoverpa armigera* PE 3 SV 2	↑	1.8	BP
G9LPQ0	UDP glycosyltransferase UGT40D1 OS *Helicoverpa armigera* GN UGT40D1 PE 2 SV 1	↑	1.7	MF
A1E9B3	Vacuolar ATP synthase catalytic subunit A OS *Bombyx mori* PE 2 SV 1	↑	2.3	BP,CC,MF
F5BYI4	Voltage dependent anion selective channel OS *Helicoverpa armigera* PE 2 SV 1	↑	1.4	CC,MF
D1M6Z4	Superoxide dismutase OS *Helicoverpa armigera* GN MnSOD PE 2 SV 1	↑	4.2	BP,MF
B2KSE9	Thioredoxin peroxidase OS *Helicoverpa armigera* PE 2 SV 1	↑	4.1	BP,MF
D5G3E6	Carboxyl choline esterase CCE014a OS *Helicoverpa armigera* PE 2 SV 1	↓	1.5	BP,MF
Q1HPN7	Fructose bisphosphate aldolase OS *Bombyx mori* GN Bmo 1030 PE 2 SV 1	↓	1.4	BP,MF
D5G3H9	Glucosidase Fragment OS *Helicoverpa armigera* PE 2 SV 1	↓	1.4	BP

Proteins like ADP ATP translocase, alkaline phosphatase, arginine kinase, fatty acid binding protein, malate dehydrogenase, sterol carrier protein, triosephosphate isomerase, vacuolar ATP synthase, and voltage dependent anion selective channel

participating in energy metabolism were upregulated. Proteins related to digestion like chymotrypsin, aminopeptidase, carboxypeptidase and lipase; redox processes like catalase, superoxide dismutase and thiredoxin peroxidase were also overexpressed. UDP glycosyltransferase which plays an important role in several processes namely, detoxication of plant allelochemicals, cuticle formation, pigmentation and olfaction was also found to be upregulated. Prohibitin which has roles in cell proliferation, differentiation, apoptosis and cell signaling; and heat shock proteins, having roles in protein metabolism and stress response were also found to be overexpressed in the PI fed larvae.

The downregulated proteins involved in energy metabolism were fructose bisphosphate aldolase and glucosidase. Carboxyl choline esterase which is central to detoxification and neurodevelopment were also downregulated.

3.4. Discussion

H. armigera when countered with rCanPI-7 dynamically regulated its gut proteases in an attempt to overcome the effect of inhibition and acquire nutritional requirements for growth and development. Differential regulation of specific protease isoforms was evident as a result of rCanPI-7 ingestion. Exposure to varied doses of rCanPI-7 suggested an optimal dose with continuous exposure starting from neonatal stage as vital for the efficient inhibition of larval growth.

Poor larval growth and development in *H. armigera,* following ingestion of rCanPI-7 was reported by Tamhane et al., (2007). The retardation in larval growth has been shown to be relative to the dose of PI used for feeding (Johnson et al., 1989; Tamhane et al., 2005; Damle et al., 2005; Edmonds et al., 1996). Zhu-Salzman et al., (2003) suggested the use of high concentration of soybean cysteine protease inhibitor, soyacystatin N for the control of *Callosobruchus maculatus* to eliminate the possibility of recovery of the insect. They noted effective growth retardation by soyacystatin N restricted to the initial instars while fourth instar onwards the larvae were found to recover. In the present study, the growth of *H. armigera* was hindered in the early instars even at the lowest dose of rCanPI-7. The initial lag further led to a delay in progression to subsequent instars, ultimately delaying the life-cycle. As the larvae matured, they seemed

to overcome the inhibitory effect imposed by the PI, particularly in case of low dose inhibitor diet.

Ingestion of moderate and high doses of rCanPI-7 effectively reduced the larval mass gain; however the biochemical analysis indicated strong upregulation of proteases with high dose of rCanPI-7; even more than moderate rCanPI-7 dose. The HGP isoforms visualized in the larvae fed on control diet as well as rCanPI-7 diet, appeared comparable although all the protease isoforms (except HGP 8) were sensitive to rCanPI-7. Hence we infer that these might be maintained *in vivo* by upregulation. Thus, overexpression of specific trypsins and chymotrypsins was the strategy adopted by *H. armigera* to withstand the feeding of rCanPI-7.

Increased expression of proteases in *H. armigera* larvae at high dose of rCanPI-7 signifies that the high concentration of rCanPI-7 may not necessarily lead to greater inhibition of insect gut proteases. The level of PI expression in transgenic plants should be relative to the insect's sensitivity to the PI to achieve effective defense against the insect (De Leo et al., 1998). It is presumed that below a certain threshold expression level of PI, the growth suppression in insects could be overcome by the overexpression of proteases but it was above this threshold that the inhibitory effects reflected into reduced growth and development (De Leo et al., 1998). To successfully establish an insect control strategy using a PI, an optimal dose of PI needs to be implemented taking into consideration the dynamism in insect gut proteases. Exposure of the larvae to low dose of rCanPI-7 at a later stage (third instar) least affected the larval mass gain probably owing to the advanced stage of development. The available/residual proteolytic activity in the insect after feeding on rCanPI-7 diet at 48 h was highly reduced. This drop in proteolytic activity also correlateswith the reduced transcript accumulation at 48 h. However, this did not reflect immediately on the larval mass gain possibly due to the gradual compensation by overexpression. A role of exopeptidases (carboxy-peptidases/aminopeptidases) for efficient nutrient utilization and activity compensation has been also suggested which might aid the process of recovery (Bown et al., 1998; Lomate and Hivrale, 2011). Monitoring the temporal response of trypsin and chymotrypsin in the third instar *H. armigera* larvae upon rCanPI-7 ingestion, depicted a transformation in the expression patterns of the transcripts; the transcripts were upregulated in the early period of feeding

on rCanPI-7 and gradually stabilized through reduced upregulation and downregulation of certain genes. Similar observations were reported by Bown et al., (2004) while investigating the response of the fourth instar *H. armigera* larvae to ingested soybean Kunitz trypsin inhibitor (SKTI) protein. The prominent upregulation of protease transcripts till 24 h in the rCanPI-7 fed third instar *H. armigera* larvae was also reflected in the proteolytic activity.

Feeding behavior (and preferences) as well as the complement of gut proteases are subject to modulation across *H. armigera* larval instars during the life-cycle (Patankar et al., 2001; Srinivasan et al., 2006; Browne et al., 2003; Chougule et al., 2005; Sarate et al., 2012). Consistent downregulation of *HaTry 2* upon exposure at neonate stage as against steady upregulation of the same upon exposure at the third instar suggests the stage specific regulation of gut proteases. The physiological response showed variable expression of proteases depending on the day and dose. Alteration (qualitative and quantitative) in the abundance of trypsin and chymotrypsin gene transcripts from day 4 to day 16 for moderate and high dose rCanPI-7 fed larvae also seemed to be the combined outcome of the stage specific gut protease regulation and response to the ingested PI. This was also reflected by the stage specific response of the early stage larvae. At neonate stage or at later stages/instars, *HaTry 1* was either consistently downregulated or showed non-significant upregulation indicating sensitivity of this isoform to rCanP-7. From the present study *HaTry 2*, *3*, *4*, and *6* and *HaChy 1* to *4* appeared to be upregulated in response to rCanPI-7 ingestion under varied treatments, hence these could be assumed to be playing a role in aiding *H. armigera* to deal with the inhibitory effects of rCanPI-7.

There have been numerous accounts wherein, upregulation of chymotrypsins was noted when the larvae were fed with trypsin inhibitors (Bown et al., 1997; Dunse et al., 2010a; Gatehouse et al., 1997; Mazumdar-Leighton et al., 2001; Wu et al., 1997). rCanPI-7 possesses two chymotrypsin inhibitory domains, hence it was expected that rCanPI-7 would inhibit the chymotrypsins along with trypsins in the insect gut. Absence of isoform HGP 3 and diminished intensity of isoform HGP 8 in the gut extract of rCanPI-7 fed larvae demonstrated the chymotrypsin inhibitory activity of rCanPI-7. Moreover, each of the isoforms visualized in-gel could be a mixture of more than one isoform of either trypsin and/or chymotrypsin. Chymotrypsin gene transcripts were

upregulated during the early period of exposure to rCanPI-7 namely, 3 and 6 h by the third instar larvae as well as at day 16 in case of high dose. Hence, the contribution of chymotrypsins in combating the rCanPI-7 cannot be neglected. The overall low abundance of chymotrypsin against trypsin in the insect gut was also substantiated by the number of isoforms visualized upon TLCK and TPCK treatments. Amongst the visualized protease isoforms HGP 8 appears to be less sensitive to rCanPI-7 and probably requires another PI with a distinct inhibitory specificity.

Elaborate experimental approaches were used by Heath et al., (1997) to explore the inhibitory potential of a 5-domain (4 TIs and 1 CI) *N. alata* PI (NaPI) towards *H. punctigera* and *T. commodus*. Firstly, the *in vitro* protease inhibitory activity of the PI or its TI or CI domains was determined, which ranged from 37 to 79%. Secondly, on feeding *H. punctigera* larvae on artificial diet incorporated with 0.26% (w/w) NaPI, delayed development and reduced larval mass were recorded. Thirdly, transgenic *N. tabacum* plants expressing NaPI (0.2% of soluble protein) were developed. However, the effect on growth in *H. punctigera* larvae fed on such leaves varied among the colonies of the larvae. Above report demonstrates a workflow for establishing the efficacy of candidate PI genes.

The present study includes *in vitro* assay of PI activity and transcriptomic analysis of the selected protease genes in CanPI-7 fed *H. armigera* larvae. Variable protease transcripts and activity were detected depending on the day of exposure and dose. *H. armigera* counteracted rCanPI-7 by the differential regulation of trypsin and chymotrypsin genes. Ingestion of higher PI dose may not always give effective antibiosis; ingestion of the moderate PI dose restricted the growth and development of *H. armigera*. There could be other reasons for this peculiar phenomenon which need to be investigated further; they may be PI/insect interaction specific. High amount of the ingested PI in adapted *H. armigera* may just be more protein for use in metabolic turnover and growth (Burgess et al., 1994). So, while applying the PI strategy for the control of insect pests these factors like (i) dose (ii) exposure time (iii) fluctuations in the insect proteases on ingestion of the PI and (iv) consistency in response across insect generations have to be considered thoroughly for maximizing and sustaining the benefits of crop protection it may render.

Proteomic analysis was carried out to gauge the effect of chronic PI ingestion on protein accumulation in the larvae. In general, increased energy and redox metabolism was observed in the larvae fed on rCanPI-7 diet. So far there has been no such study revealing the proteomic changes in response to PI ingestion in lepidopteran larvae. In one such investigation involving analysis of global proteomic changes induced in *Callosobruchus maculatus* larvae after ingestion of cystatin, showed overexpression of proteins associated with transcription/translation and anti-stress reactions (Nogueira et al., 2012). Besides, there are reports about proteomic analysis of *H. armigera* tissues after ingestion of insecticides and plant allelochemicals (Dawkar et al., 2011; 2014; Konus et al., 2013). In all the cases increases energy and anti-stress metabolism was observed in the treated *H. armigera* larvae.

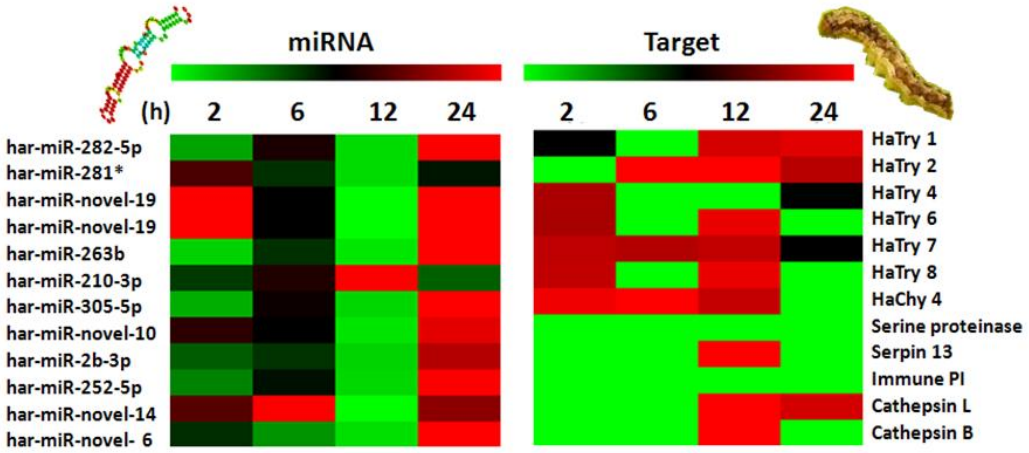
3.5. Conclusion

Effective inhibition of *H. armigera* growth and development was observed upon continuous exposure to rCanPI-7 since neonatal stage. The molecular response offered by the larvae upon exposure to rCanPI-7 varied with the doses of rCanPI-7 as well as the larval stage of exposure. Such a dose-dependent response emphasizes imposition of optimal dosage at early larval stages for effective antibiosis. The stress imposed by the ingestion of rCanPI-7 was evident in the proteomic analysis. Current study suggests 0.015% (150 μg protein) rCanPI-7 per gram of AD as an optimal dose, which could serve as a reference when taking the CanPI-7 gene to field trials. However *in planta* transgene expression of CanPI-7 for effective antibiosis towards *H. armigera* remains subject to further investigation.

Next Generation Genomics

Chapter 4

Identification and expression profiling of *Helicoverpa armigera* microRNAs and their possible role in the regulation of digestive proteases



Expression of *H. armigera* miRNAs and their target mRNAs

The research work described in this Chapter is part of full-length paper, which has been accepted for publication in Insect Biochemistry and Molecular Biology, 2014, (In Press)

Chapter 4: Identification and expression profiling of *Helicoverpa armigera* microRNAs and their possible role in the regulation of digestive proteases

Chapter 4: Identification and expression profiling of *Helicoverpa armigera* microRNAs and their possible role in the regulation of digestive proteases

Abstract

The present investigation is an effort to determine the possible roles of microRNAs (miRNAs) in the regulation of protease gene expression in *Helicoverpa armigera* upon exposure to plant protease inhibitors (PIs). Using Illumina platform, deep sequencing of 12 small RNA libraries was performed from *H. armigera* larvae fed on artificial diet (AD) or recombinant *Capsicum annuum* PI-7 (rCanPI-7) incorporated diet, at various time intervals (0.5, 2, 6, 12, 24, and 48 h). Sequencing data were analyzed with miRDeep2 software; a total of 186 unique miRNAs were identified from all the 12 libraries, out of which 96 were conserved while 90 were novel. These miRNAs showed all the conserved characteristics of insect miRNAs. Homology analysis revealed that most of the identified miRNAs were insect-specific, and more than 50 miRNAs were Lepidoptera-specific. Several candidate miRNAs (conserved and novel) were differentially expressed in rCanPI-7 fed larvae as compared to the larvae fed on AD. *H. armigera* miRNAs were found to have target sites in several protease genes as well as in protease regulation related genes such as serine PI and immune reactive PI. As expected, negative correlation in the relative abundance of miRNAs and their target mRNAs was evident from qRT-PCR analysis. The investigation revealed potential roles of miRNAs in *H. armigera* protease gene regulation.

4.1. Introduction

MicroRNAs (miRNAs) are single stranded, small non-coding RNAs (about 22 nucleotides) regulating gene expression at the post-transcriptional level in almost all biological processes (Bartel, 2004; Kim et al., 2009). They represent a major class of post-transcriptional regulators, which play key roles in development, differentiation, to cope up with different stresses and apoptosis in eukaryotic organisms (Ambros, 2004; Miska, 2005; Asgari, 2013). MiRNAs critically regulate the expression of target genes through binding to complementary sites (usually 3' or 5' -untranslated regions) of the mRNA targets (Lee at el., 1993; Asgari, 2013). Recently, numerous miRNAs have been identified and annotated from several eukaryotic organisms including insects. More than 3000 insect miRNAs have been identified from insect species such as *Drosophila* sp., *Anopheles gambiae*, *Aedes aegypti*, *Bombyx mori*, *Manduca sexta*, *Spodoptera litura*, *Helicoverpa armigera*, *Tribolium castaneum*, *Acyrthosiphon pisum*, *Apis mellifera* and *Locusta migratoria* (Asgari, 2013) (http://www.mirbase.org/). A number of methods have been used to identify insect miRNAs. For example, in *S. litura*, 58 putative miRNAs have been identified using computational methods and their expression patterns have been investigated during various developmental stages (Rao et al., 2012). Several numbers of miRNAs (163 conserved and 13 novel) were identified from *M. sexta* larvae, pupae and adults using small RNA sequencing (Zhang et al., 2012). Homology-based prediction approach was applied to predict miRNAs from *H. armigera* and *H. virescens* (Ge et al., 2013; Chilana et al., 2013). Regulatory functions of miRNAs in metamorphic switch were revealed in *B. mori* (Yu et al., 2008). Although several miRNA families have been isolated from different insect species, their roles are still unexplored particularly in the regulation of protease genes. Insect digestive processes and genes involved therein are tightly regulated, and the involvement of different regulatory factors including miRNAs is apparent.

Digestive proteases serve as an important means for insects to adapt and overcome the effect of plant protease inhibitors (PIs). PIs are one of the major plant defensive molecules and have potential to inhibit feeding, growth and development of herbivorous insects (Giri et al., 2003; Schlüter et al., 2010; Lomate and Hivrale, 2011). However, insects have evolved number of strategies to counteract or evade the effect of

plant PIs. These include (i) regulation of existing gut proteases in such a way that they minimize the PI effect (Bown et al., 1997; Broadway, 1997), (ii) production of enzymes that hydrolyze and thus, inactivate the inhibitors (Giri et al., 1998, Telang et al., 2005), and (iii) secretion of proteases that are insensitive to inhibition by those PIs (Jongsma et al., 1995; Jongsma and Bolter, 1997, Chougule et al., 2005). Altogether, it seems that proteases might be the key player in the adaptation of insects and their expression is systematically regulated upon exposure to PIs. Currently, very little information is available about the regulation of proteases in the midgut of herbivorous insects. Multiple factors have been presumed to be regulating the expression of gut protease genes upon PI ingestion (Bown et al., 2004).

Helicoverpa armigera (Hübner) (Lepidoptera: Noctuidae) is a polyphagous and devastating insect pest of many important crop plants and is responsible for heavy economical losses in agriculture. Earlier studies on *H. armigera* gut proteases showed differential patterns and altered biochemical properties during larval development when exposed to various plants and PIs (Patankar et al., 2001). Furthermore, *H. armigera* digestive proteases revealed differential response to various diets and their levels were regulated based on the PI composition in the diet (Bown et al., 2004; Dawkar et al., 2011). However, the factors or mechanisms underlying the regulation of protease gene expression in *H. armigera* are unknown and need to be uncovered to develop better insect control strategies. In the present work, *H. armigera* larvae were fed on recombinant *Capsicum annuum* PI-7 (rCanPI-7: known to alter the protease gene expression; as discussed in the previous chapter) in artificial diet and subjected to unbiased miRNA analysis at various time intervals (0.5 to 48 h) of PI feeding. Further, we have carried out the identification of miRNAs, their comparative profiling and relative transcript abundance in the larvae fed on control and PI containing diet. Following the approach of high throughput sequencing using Illumina platform, we have attempted to establish a rational correlation between the expression of miRNAs and their target proteases.

4.2. Material and Methods

4.2.1. Insect culture and feeding assays

Neonates obtained from laboratory culture of *H. armigera* maintained at optimum growth conditions (27±2 °C, 60±5% relative humidity and a photoperiod of 14 h light and 10 h dark) were used to carry out the bioassays. The artificial diet (AD) was prepared as described by the method of Nagarkatti and Prakash (1974). Based on the earlier results 150 µg of recombinant *Capsicum annuum* protease inhibitor-7 (rCanPI-7) was incorporated per gram of AD for the feeding bioassays. First instar larvae were allowed to feed for 48 h on AD and AD incorporated with rCanPI-7. Each larva was maintained in an individual vial containing the respective diets. Each set containing 100 insects was maintained. Whole insects were harvested at various time intervals (0.5, 2, 6, 12, 24, and 48 h) from rCanPI-7 fed as well as the AD fed insects. At each stage of bioassay, the harvested insect tissues were snap frozen in liquid nitrogen and stored at −80 °C until further use.

4.2.2. Small RNA sequencing

H. armigera larval samples from each time point of feeding assay were processed for small RNA library preparation. Total RNA was isolated from the whole-body homogenates of insect tissues using Trizol reagent (Invitrogen) based on the manufacturer's protocol. RNA was quantified and checked for purity and integrity using agarose gel electrophoresis, Nanodrop (Thermo Scientific) and the Agilent 2100 Bioanalyzer (Agilent technologies, Palo Alto, CA, USA). Small RNA library was prepared using Tru Seq Small RNA sample preparation kit v2 (Illumina, San Diego, CA, USA) following manufacturer's specified protocols. Total RNA (1 µg) was taken and adapters were ligated to the 5' phosphate and 3' hydroxyl groups that are specific to miRNAs according to manufacturer's instructions. Reverse transcription PCR was done based on manufacturer's protocol to make cDNA constructs that would enrich the fragments that have adapter molecule on both ends. A size selection targeting miRNA's in the range of 20 to 35 nucleotides was performed (The adapter is around 125 bases in length). The library is then validated on the bioanalyzer. Small RNA sequencing was performed using the HiSeqTM 2000 sequencing system from Illumina (Illumina) at Centre for Cellular and Molecular Platforms (C-CAMP), Bangaluru, KA, India. The

obtained sequence tags from the Illumina sequencing were subjected to a primary analysis in which low-quality tags and adaptor contaminants were discarded.

4.2.3. Identification of miRNAs from sequencing data

H. armigera miRNA candidates were identified by using the miRDeep2 software (Friedlander et al., 2012). miRDeep2 uses the Bowtie read mapper internally to map letter space data (Illumina) to genome sequences and detects both known and novel miRNAs (Langmead et al., 2009). Raw sequencing reads from 12 libraries (6 different time intervals each for AD and rCanPI-7 fed larvae) were submitted to miRDeep2 and data from each library were separately analyzed. Briefly, miRDeep2 pre-processed raw sequencing reads by removing the 3' adapter sequence and discarding reads shorter than 18 nucleotides, before aligning reads to the *Bombyx mori* (silkworm) genome. *H. armigera* was designated as the related species and the known miRNA input was from miRBase 20.0 to analyze sequenced miRNAs.

4.2.4. Homology of H. armigera conserved miRNAs

To check similarity of *H. armigera* conserved miRNAs with the miRNAs of other organisms, homology analysis was carried out. miRBase database (Release 20.0) (Griffiths-Jones et al., 2008) was used to compare *H. armigera* miRNA sequences with other species. Each conserved miRNA sequence was searched against miRBase database and the similar miRNAs in other organisms were identified. Nomenclature of *H. armigera* conserved miRNAs was done on the basis of these homologous miRNAs identified in other organisms (Griffiths-Jones et al., 2006).

4.2.5. Expression profile of miRNAs

In addition to identification of mature miRNAs from sequenced small RNA samples, miRDeep2 also generates expression values as a read count for the detected miRNAs. Differential expression of conserved and novel miRNAs was analyzed in the *H. armigera* larvae fed on AD or rCanPI-7 at various time intervals. The expression values produced by miRDeep2 for each sample were normalized. Normalization of read counts in each sample was achieved by dividing the counts by a library size parameter of the corresponding library and finally the values transformed into log2 scale, and fold changes were given in log2 scale (Creighton et al., 2009).

4.2.6. Prediction of miRNA targets

Targets for identified miRNAs were predicted using NCBI and in house *H. armigera* transcriptome database. As we were interested in proteases, we analyzed whether the *H. armigera* conserved and novel miRNAs could target the protease genes. The mRNA sequences of *H. armigera* proteases were downloaded from NCBI. The miRNA targets were predicted in the mRNA sequences of *H. armigera* proteases from NCBI and in the transcript sequences from in house database. Two miRNA target-prediction software were used with default parameters: miRanda (http://www.microrna.org/microrna/getDownloads.do) (Betel et al., 2008) and RNAhybrid (http://bibiserv.techfak.uni-bielefeld.de/rnahybrid/welcome.html) (Rehmsmeier et al., 2004). The miRNA targets commonly predicted by both the software were selected.

4.2.7. Validation of miRNAs and their targets by qRT-PCR

Total RNA isolation, synthesis of the first strand cDNA, Real-Time PCR and determination of the relative expression of miRNA and their target genes was performed as mentioned in section 3.2.4. Details of primers for miRNAs and target genes are mentioned in **Appendix I Table 5 and 6**.

4.3. Results

4.3.1. Deep sequencing of H. armigera small RNA libraries

Total 12 *H. armigera* small RNA libraries were sequenced using Illumina platform. A total of 117-million quality sequence reads were obtained from the small RNA libraries after raw sequencing data was filtered through illumina quality filter. More reads were obtained for AD fed larvae (59-million) than for larvae fed on rCanPI-7 containing diet (57-million) **(Fig. 4.1A)**. These data have been submitted to the NCBI Gene Expression Omnibus (GEO) under series record GSE61503 (Barrett et al., 2009). The length of these small RNAs ranged from 15 to 30 nt. The highest peak for nucleotide length distribution was obtained between 22 to 23 nt (with few exceptions where the highest peak was at 24-25/26-27 nt), which showed the quality of sequencing **(Fig. 4.1B)**.

Figure 4.1:

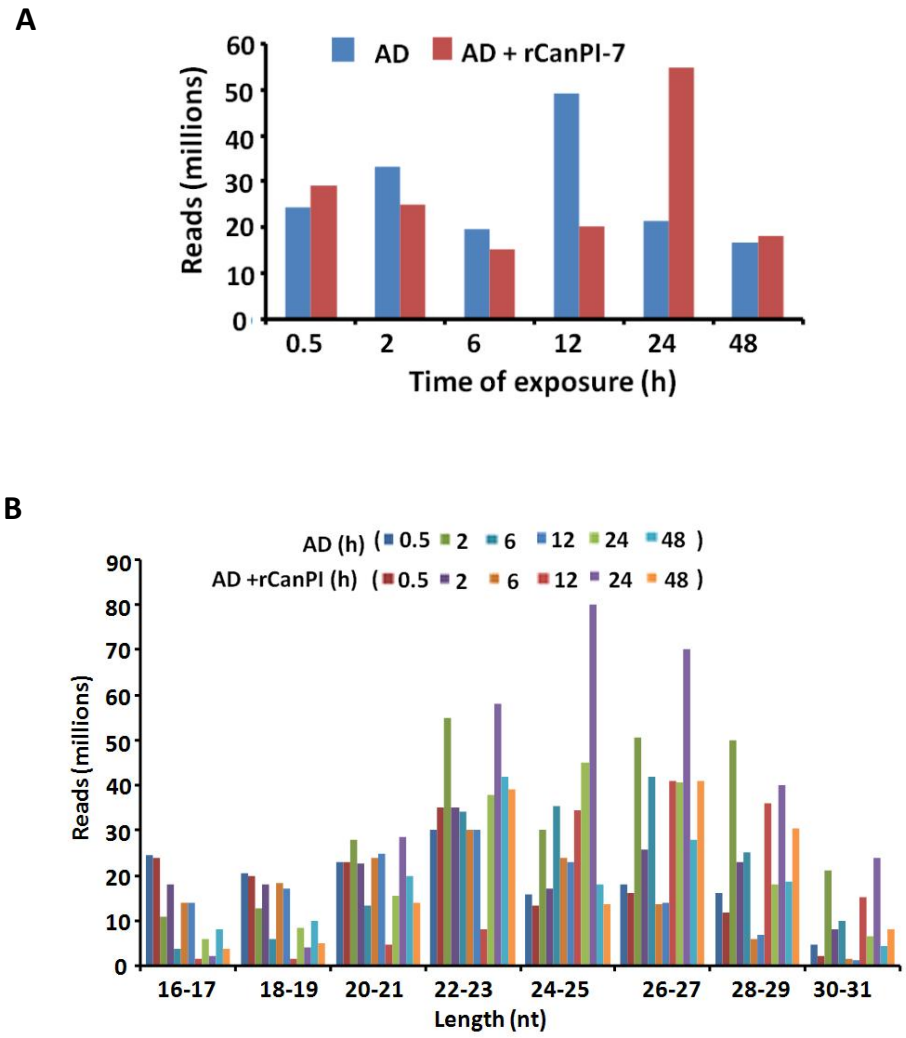


Figure 4.1: Read obtained by smRNA sequencing and length distribution of *H. armigera* smRNAs. (A) Total number of reads obtained from smRNA libraries at different time points. (B) Length distribution of smRNA reads in AD and rCanPI-7 fed larvae at different time points.

4.3.2. H. armigera miRNAs and their specific conserved features

Raw sequencing data were submitted to miRDeep2 software. *B. mori* genome was used as reference since *H. armigera* genome is not available. miRDeep2 predicted more than 2000 miRNAs in the 12 small RNA libraries of *H. armigera*. After filtering, we identified 186 unique miRNA candidates including 90 potentially novel and 96 conserved miRNAs **(Appendix I Table 7 and 8)**. *H. armigera* miRNAs were mainly 22 to 25 nt, which is a similar length distribution of miRNAs identified in other species. Several

miRNAs were found to be common in the libraries obtained from the larvae fed on AD or rCanPI-7 containing diet and also at different time intervals of PI feeding. The specific nucleotide occurrence was analyzed in the obtained miRNA sequences. *H. armigera* miRNAs showed a dominant bias to uracil (U) at the first nucleotide, particularly the miRNAs with a length of 19 to 24 nt **(Fig. 4.2)**. We also examined the percentage of the four nucleotides occurring at each position. The positions showing most dominant bias to U were 1, 9, and 17. The dominance of uracil at 1st position towards 5' end is considered to be one of the conserved features of miRNAs (Lau et al., 2001).

Figure 4.2:

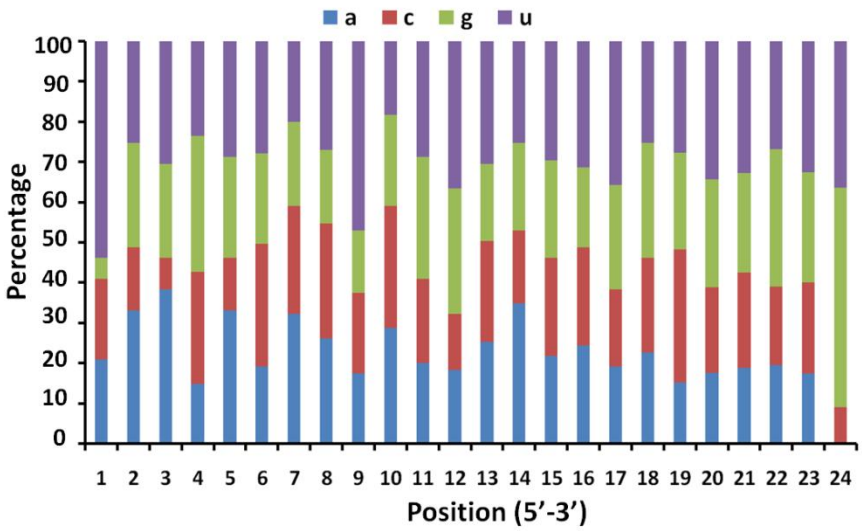


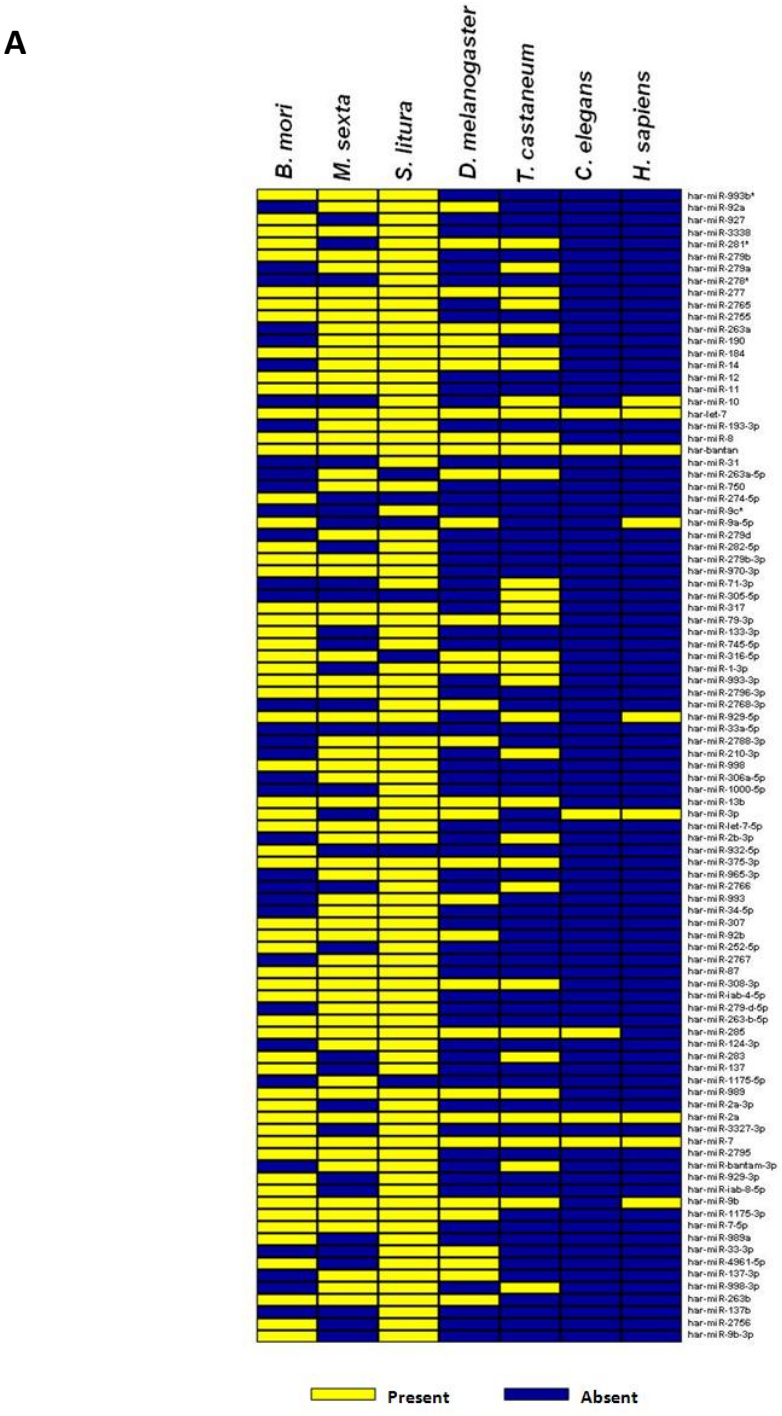
Figure 4.2: Frequency of occurrence of nucleotides at each position in *H. armigera* miRNAs. Bias towards occurrence of specific nucleotides at particular position is evident.

4.3.3. Insect specific miRNAs in H. armigera

Considering that some *H. armigera* miRNAs would be unique to insects, we searched complete miRBase database for the presence of conserved miRNAs within *H. armigera* or insects or other organisms. Only a few *H. armigera* miRNAs were conserved among all the organisms tested using miRBase **(Fig. 4.3A and B)**. As expected, *H. armigera* miRNAs were found to be specifically conserved in closely related species including *S. litura*, *B. mori* and *M. sexta* **(Fig. 4.3A)**. Only few *H. armigera* miRNAs were found to have their orthologs in *H. sapiens* (4%) and *C. elegans* (1%). Moreover,

90% *S. litura*, 60% *B. mori* and 60% *M. sexta* miRNAs were observed to be conserved in *H. armigera* whereas more than 30% of *H. armigera* miRNAs were having their orthologs in *D. melanogaster* or *T. castaneum* **(Fig. 4.3B)**.

Figure 4.3:



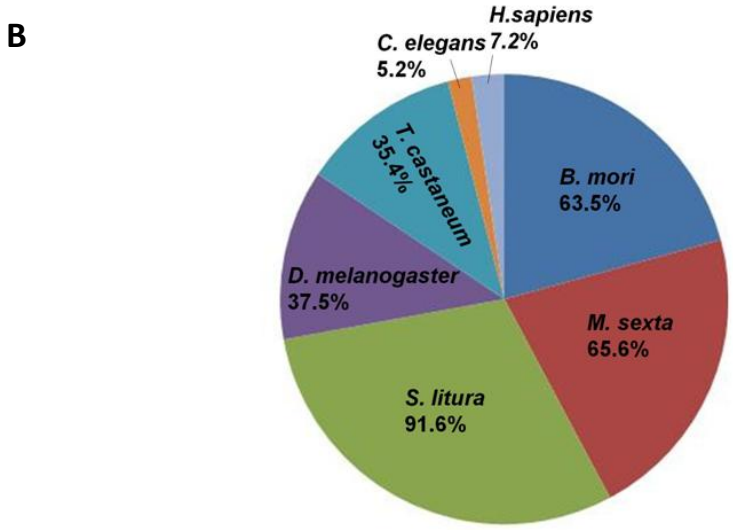


Figure 4.3: *H. armigera* conserved miRNA homologs in other organisms. (A) The occurrence of each conserved miRNA across different organisms. Yellow color denotes presence while blue denotes absence. (B) Percent distribution of conserved miRNAs identified across different organisms [*Bombyx mori, Manduca sexta, Spodoptera litura, Drosophila melanogaster, Tribolium castaneum, Caenorhabditis elegans, Homo sapiens*].

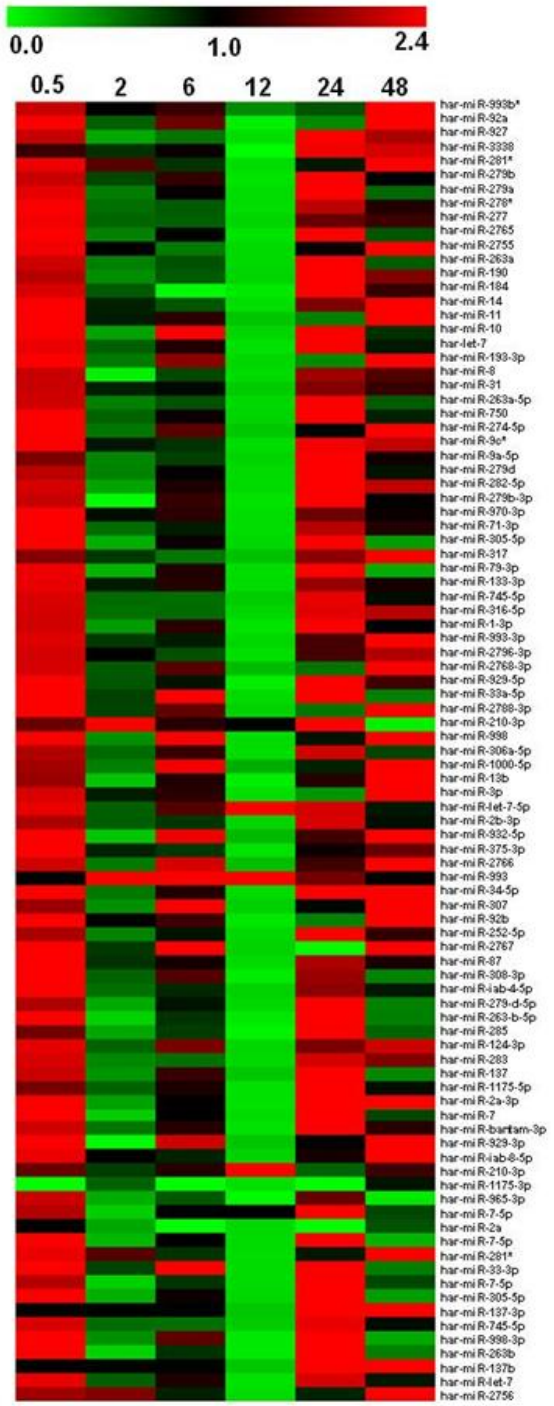
4.3.4. H. armigera miRNAs expressed differentially upon rCanPI-7 exposure

Expression values generated by miRDeep2 were used to analyze differential expression of *H. armigera* miRNAs. The miRNAs having read count more than 10 in minimum three smRNA libraries (94 conserved and 20 novel) were selected for differential expression analysis. Several conserved and novel miRNAs were differentially expressed in the larvae fed on rCanPI-7 as compared to the larvae fed on AD **(Fig. 4.4)**. It clearly suggested the effect of PI feeding on the expression of miRNAs. Almost all the conserved miRNAs (90) were found to be over-expressed in the larvae fed on rCanPI-7 for 0.5 h with respect to the control. The miRNA expression decreased after 2 h and increased slightly at 6 h of CanPI feeding. Surprisingly, almost all conserved miRNAs (91) were down-regulated at 12 h while most of them were up-regulated after 24 h and some at 48 h of CanPI feeding. Although, no specific trend was observed in conserved miRNA expression, few candidates were up or downregulated at all the time points. For example, har-miR-993 and har-miR-210-3p were over-expressed, whereas har-miR-2a

and har-miR-1175-3p were down-regulated at almost all the time points of CanPI feeding **(Fig. 4.4A)**. Novel miRNAs also depicted almost similar pattern **(Fig. 4.4B)**.

Figure 4.4:

A



B

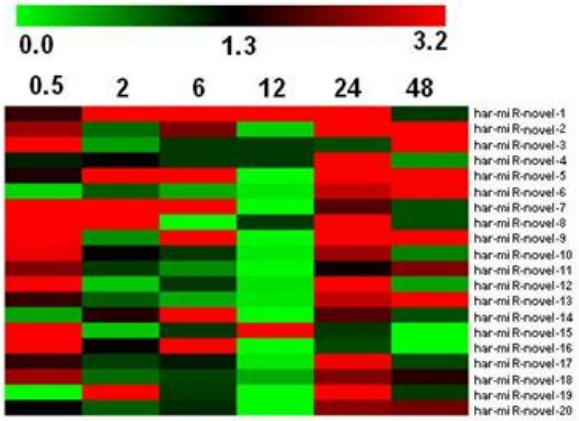


Figure 4.4: Differential expression of *H. armigera* miRNAs after ingestion of rCanPI-7. The miRNAs having read count more than 10 in at least three libraries were selected for this analysis. (A) Expression of conserved miRNA (B) Expression of novel miRNA.

4.3.5. H. armigera miRNAs target several proteases and related genes

The targets of identified miRNAs were computationally predicted from *H. armigera* protease EST sequences and in house transcriptome database. *H. armigera* miRNAs were found to target several genes including important enzymes, transcriptional factors, receptors and hormones. However, the focus of the present study was to get insights on protease gene regulation after the ingestion of PIs. Therefore, we analyzed the miRNAs targeting protease genes or other genes probably involved in protease gene regulation. Several miRNAs were observed to target various protease genes including both endo- and exo-proteases such as trypsin, chymotrypsin, cathepsin, amino and carboxypeptidase **(Table 4.1)**. Moreover, several miRNAs had multiple target genes and numerous genes had mutually common targeting miRNAs. The above-mentioned miRNAs, har-miR-210-3p and har-miR-2a were found to target different proteases. Interestingly, miRNA putative targets also include the protease regulation related genes such as serine proteinase inhibitor 13 (serpin 13) and immune reactive proteinase inhibitor **(Table 4.1)**.

Table 4.1: Proteases and related target genes for *H. armigera* miRNAs.

miRNA	Proteases and related genes
har-miR-novel-2	HaChy 1
har-miR-novel-6	Aminopeptidase 3, serine protease inhibitor 13
har-miR-novel-7	Cathepsin B-S1, cathepsin L-like protease, hemolymph proteinase 18, chymotrypsin 1, caspase-1
har-miR-novel-8	Cathepsin B-S1, protease (SerProx-1)
har-miR-novel-9	Trypsin
har-miR-novel-10	Serine protease inhibitor 002, serine proteinase-like protein 1
har-miR-novel-13	Trypsin, carboxypeptidase B
har-miR-novel-14	Cathepsin B-S1,cathepsin L-like protease, carboxypeptidase B, hemolymph proteinase 1830kP protease A (43k peptide) (Fragment),
har-miR-novel-18	30kP protease A (43k peptide) (Fragment), cathepsin B-S1, hemolymph proteinase 18
har-miR-novel-19	HaTry 4, HaTry 6, HaTry 7, putative trypsin (T1)
har-miR-281*	HaTry 4
har-miR-2755	Tissue plasminogen activator
har-let-7	Putative serine protease
har-bantam	Carboxypeptidase A
har-miR-9c*	HaTry 8, serine protease 4, putative serine protease
har-miR-282-5p	HaTry 1
har-miR-993-3p	Serine protease 4, HaTry 8
har-miR-210-3p	Trypsin, serine protease 4
har-miR-998	Carboxypeptidase A,
har-miR-2b-3p	Immune reactive putative protease inhibitor
har-miR-993	Immune reactive putative protease inhibitor
har-miR-34-5p	HaTry 6, HaTry 7, HaTry 8, carboxypeptidase vitellogenic-like 3
har-miR-252-5p	Immune reactive putative protease inhibitor mRNA
har-miR-2a	Serine proteinase-like protein 1
har-miR-7	Carboxypeptidase A
har-miR-210-3p	Cathepsin L-like protease
har-miR-305-5p	HaChy 4, HaTry 7
har-miR-263b	HaTry 4, HaTry 7, cathepsin B-like cysteine proteinase, serine protease 5, serine protease 6

4.3.6. Validation of miRNAs and their protease related gene targets

The relative transcript abundance of target genes was determined by qRT-PCR. Target genes (12) were selected and their relative transcript abundance was evaluated in the larvae fed on AD or rCanPI-7 diet at 6 to 24 h. The differential expression (calculated by sequencing read count) of selective protease targeting miRNAs was compared with the transcript abundance of target genes. Generally, if miRNA levels are high their respective target levels are expected to be low and *vice versa*. The obtained results revealed that the expression of most of the miRNAs and their respective targets showed the complementary trend of expression with negative correlation in the relative abundance of respective transcripts **(Fig. 4.5A)**. For instance, the expression of har-miR-281* was exactly apposite to that of its target mRNA HaTry 2. Most of the miRNAs were up-regulated and their respective target mRNAs were down-regulated at 24 h of CanPI feeding. Interestingly, the expression of serpin 13 and immune PI was observed to be in accordance with the expression of their respective miRNAs at almost each time interval except at 12 h for immune PI. Selected miRNAs (har-miR-282-5p and har-miR-305-5p) that targeted proteases (one trypsin and one chymotrypsin) were also validated by qRT-PCR. The miRNA expression determined by qRT-PCR correlated with expression calculated using sequencing read count. For example, the pattern of relative abundance of har-miR-282-5p and har-miR-305-5p matches with its expression based on read counts at 2, 6, 12 and 24 h **(Fig. 4.5B)**.

Figure 4.5:

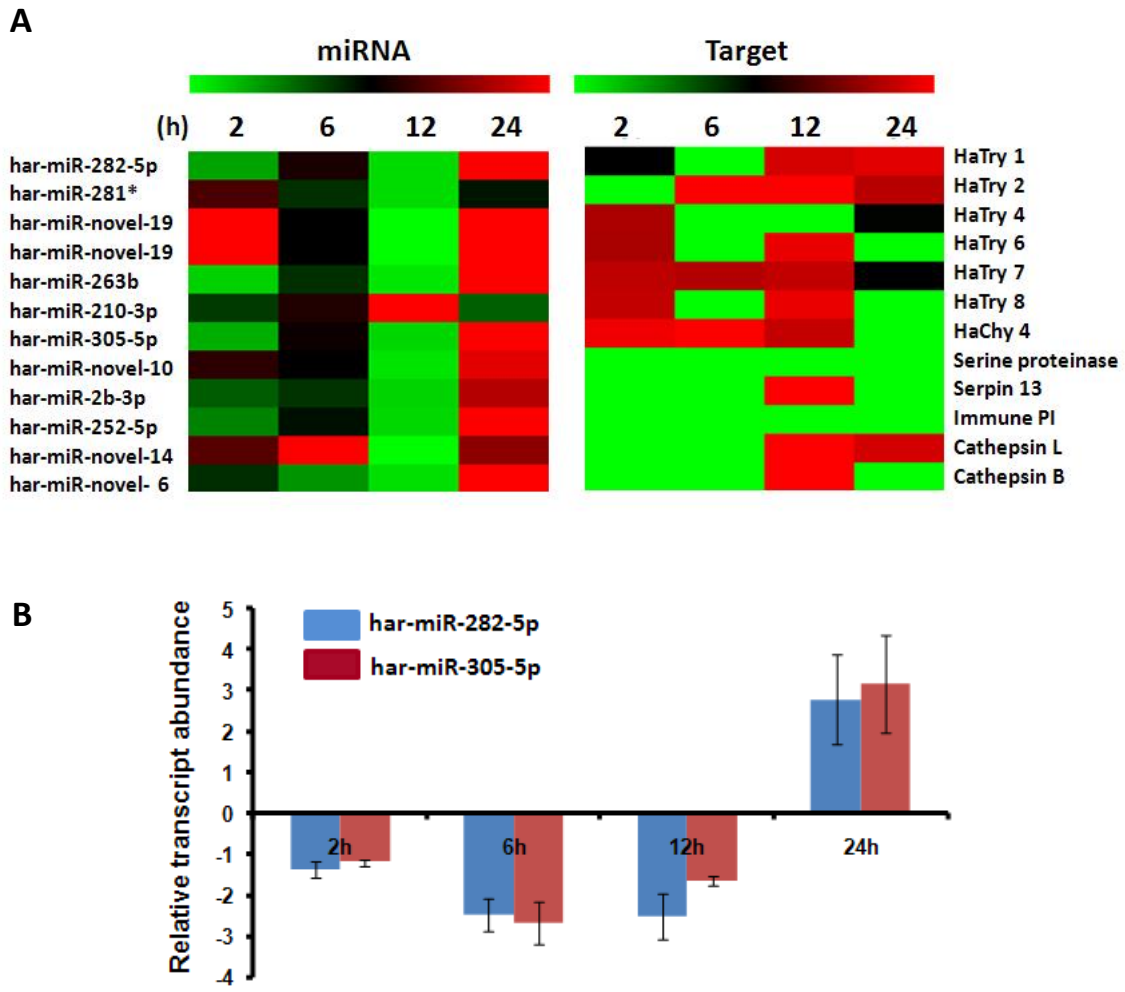


Figure 4.5: Expression of miRNAs and their respective target protease genes. (A) Expression of miRNAs as determined from read counts and expression of target protease genes determined using qRT-PCR. (B) Validation of expression of selected miRNAs (targeting trypsin and chymotrypsin) using qRT-PCR.

4.4. Discussion

Insect digestive physiology is influenced by dietary components and the midgut complement in lepidopteran insects is known to alter upon ingestion of PIs (Bown et al., 2004). Complexity and variable specificity coupled with temporal and spatial expression of proteases is insect's primary mechanism to deal with PIs (Lomate et al., 2013). The present study was undertaken to investigate the role of miRNAs in protease gene

regulation upon exposure to CanPI, while identifying the conserved and novel miRNA candidates, their characteristics and expression patterns in *H. armigera*.

MicroRNAs have been identified in several insects using different methods. For example, the identification of miRNAs was carried out in several lepidopteran insect species using available genomic information or BAC sequences (Jagadeeswaran et al., 2010; Liu et al., 2010; Surridge et al., 2011). Recently, small RNA sequencing has become popular for identifying miRNAs from a range of organisms. Ge et al., (2013) reported identification of *H. armigera* and *S. litura* miRNAs using small RNA sequencing. They identified 74 known and 10 novel miRNA sequences in *H. armigera* using homology based prediction. In the present study, we identified several miRNAs from *H. armigera* larvae fed on AD and rCanPI-7 using small RNA sequencing approach coupled with miRNA analysis by miRDeep2 software. This analysis resulted in the identification of 96 conserved and 90 novel miRNA candidates from *H. armigera* using *B. mori* genome as reference which are much more than the reported ones so far. For each time point tested, only single deep sequencing replicate was used, however, the results obtained through deep sequencing were validated with qRT-PCR for selected transcripts. Furthermore, several time points were used, which can also strengthen our results despite the lack of sequencing replicates.

The specific characteristic of miRNAs is that they generally begin with a U at the 5' terminus (Lau et al., 2001) and accordingly *H. armigera* miRNA sequences dominantly started with uracil (U) as the first nucleotide. Moreover, the percentage of the four nucleotides appearing at each position in identified miRNAs was consistent with that described in previous reports (Zhang et al., 2009; Ge et al., 2013). The results of homology analysis revealed that several *H. armigera* miRNAs were conserved in specifically lepidopteran insect species. Few miRNA candidates such as miR-7, miR-let-7 and miR-bantam were found to be conserved in all the organisms reported so far. MiRNAs have been discovered in virtually all organisms and considered to be highly conserved regulators (Du and Zamore, 2005; Barvkar et al., 2013). Our study gave an understanding of *H. armigera* miRNAs and their conserved orthologs in other organisms. Many of *H. armigera* miRNAs do not have orthologs in mammals and a nematode but were conserved in only insect species. These insect-specific miRNAs might have arisen

after the divergence of insects and other invertebrates, signifying that some miRNAs might have undergone dynamic evolutionary changes. It has been reported that miRNAs are evolutionarily active and undergo a rapid ‘‘birth and death’’ within Drosophila, and some miRNAs even duplicate within the genome (Lu et al., 2008). In homology analysis, we also observed some miR*s such as miR-278*, miR-281*, miR-993b* and miR-9c*. Ge et al., (2013) could identify two miR*s in *H. armigera* and found their homologs in vertebrates. Moreover, miR*s have been identified in other insect species including *D. melanogaster* and *T. castaneum* (Ruby et al., 2007; Marco et al., 2010).

Most of the *H. armigera* miRNA candidates were differentially expressed in CanPI fed larvae as compared to larvae fed on AD. The effect of CanPI on the expression of miRNAs was observed which suggests the possible involvement of miRNAs in the metabolism of PIs in the insect gut. These differentially expressed miRNAs might be involved in different physiological processes such as digestion, detoxification and excretion, which are required to encounter PIs. It is well known that PIs can hamper the growth and development of lepidopteran insects. Our previous reports have demonstrated that CanPIs can reduce the larval growth, pupation and fecundity in *H. armigera* (Tamhane et al., 2005, 2007). In the present study, interestingly the ingestion of CanPI affected the expression of miRNAs such as miR-let-7, which has essential role in the development of diverse animals such as nematode, fly, mouse and human (Bussing et al., 2008). Furthermore, miR-let-7 is required for neuromusculature remodeling during the metamorphosis of flies (Sokol et al., 2008). CanPIs influence the development and metamorphosis of *H. armigera* and hence might influence the expression of miR-let-7.

MiRNA target sites were predicted in *H. armigera* proteases and related genes. Based on the differential expression analysis and qRT-PCR analysis, potential miRNA candidates having target sites in protease or related genes were selected. MiRNAs were found to target several important protease mRNAs such as trypsin, chymotrypsin, and cathepsins. The expression of miRNAs and their respective targets was negatively correlated across several time points. This suggested the possible role of miRNAs in the regulation of protease transcripts. Earlier, several proteases and serine protease inhibitors (Serpins) involved in the regulation of immunity related gene expression have been found

as targets for *M. sexta* miRNAs (Zhang et al., 2014). Although, serpins are known to be necessary for the regulation of several serine protease cascades, they can inhibit insect digestive proteases and ultimately regulate their activity (Gubb et al., 2010). Internal regulation of protease activity is the primary function of protease inhibitors in several organisms (Neurath and Walsh, 1976; Rawlings et al., 2004). Interestingly, *H. armigera* miRNAs target two important serine protease inhibitors, serpin 13 and immune PI. It is possible that the miRNAs might regulate the expression of these PIs and eventually the activity of their target proteases.

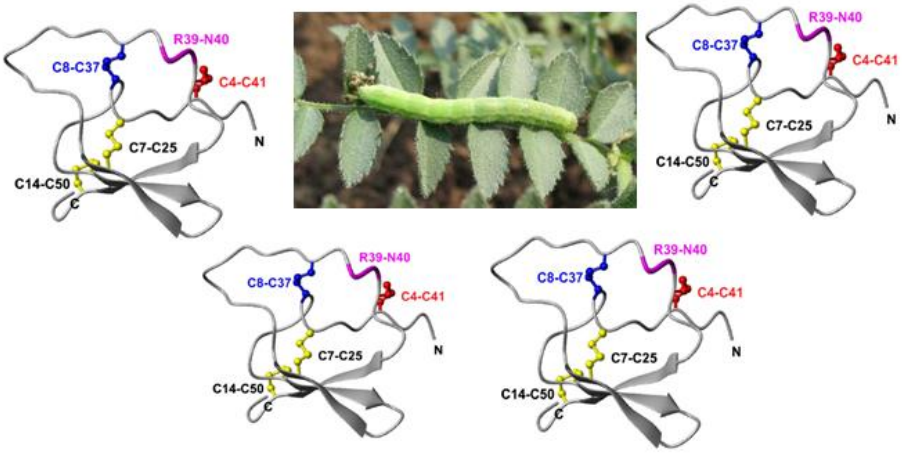
In the current study, several miRNAs were found to target different trypsin mRNAs. Insect trypsins mainly contribute in the digestion of dietary proteins, however, their role in the activation of other proteases is evident (Neurath and Walsh, 1976; Bown et al., 2004). Usually, midgut proteases are secreted in inactive forms (zymogen) and need hydrolysis of some amino acids from the peptide chain for their activation (Neurath and Walsh, 1976; Bown et al., 2004). Trypsins can also modulate the activity of monitor peptides, which are the key regulators of digestive protease activity in humans and presumed to be present in insects (Liddle, 1995; lwai et al., 1998; Huang et al., 1998). Trypsin targeting miRNAs found in the present study perhaps regulate the *H. armigera* protease expression by multiple ways as discussed above. In one of the recent reports, a trypsin targeting har-miR-2002b was cloned from *H. armigera*. Oral administration of har-miR-2002b mimic had significant impact on larval development and survival (Jaychandran et al., 2013). However, we did not identify the har-miR-2002b in our study possibly due to dissimilar diet composition or variation in identification methodology/algorithms. We also identified few cathepsins as the targets for *H. armigera* miRNAs. Cathepsin has functional role in insect metamorphosis and during pupation (Wang et al., 2010). Functional validations will further highlight the specific role of miRNAs in the expression regulation of protease genes. Moreover, protease targeting and other important development related miRNAs could be the best targets to use miRNA mimic inhibitor based insect control strategies. Our studies also provide the clue to further investigate the functional role of miRNAs in protease gene regulation against PIs, which might eventually benefit the PI based insect management approach.

4.5. Conclusion

Using Illumina platform, small RNA sequencing was performed for 12 libraries prepared from *H. armigera* larvae fed on AD or rCanPI-7. Several conserved (96) and novel (90) miRNAs were identified. Bias towards occurrence of specific nucleotides at particular position was evident in the mature miRNA sequences. Most of the miRNAs were observed to be insect-specific, and more than 50% miRNAs were Lepidoptera-specific. Differential expression of several miRNAs after ingestion of CanPI-7 was observed. Negative correlation in the relative abundance of protease targeting miRNAs and their target mRNAs was apparent. The possible role of miRNAs in *H. armigera* protease gene regulation cannot be denied. However, such conclusion can be ascertained after *in vivo* functional validation.

Chapter 5

General discussion and Future directions



Chapter 5: General discussion and future directions

5.1 Biotic stress: Induction of plant defense

5.2 *H. armigera* responses to CanPI ingestion

5.3 Designing competent strategies for *H. armigera* control

5.4 Future directions

Chapter 5: General discussion and future directions

5.1. Biotic stress: Induction of plant defense

In nature, plants are constantly presented with attacking agents like, pathogens, parasites and insect herbivores. After perceiving the attack, plant metabolism must balance demands for resources to support defense and the requirements for cellular maintenance, growth and reproduction. Such induced defenses involve a range of molecules whose synthesis is spatially and temporally controlled (Walling, 2000; Kessler and Baldwin, 2002; Wu and Baldwin 2010). Plant defense against herbivores often relies on the synthesis of phytochemicals (small molecules and proteins) that adversely affect the attacking herbivore. Among numerous defenses, stress inducible proteinaceous protease inhibitors (PI) form an extensively explored class of antinutritive molecules in herbivore defense (Mosolov and Valueva, 2005; Tamhane et al., 2012). However, activation of such defense strategies is associated with fitness costs due to alteration/shift in allocation of resources (Zavala and Baldwin, 2004). Allocating resources from growth to defense leading to a reduction of photosynthetic capacity in leaf tissues characterize the cost for defense. The association between photosynthesis and defense has been well documented (Hermsmeier et al., 2001; Giri et al., 2006).

In the current study, primarily, the characterization of the proteomic changes in the induced *Capsicum annuum* leaves was carried out (Chapter 2A). Upregulation of proteins related to redox and damage repair accompanied by downregulation of photosynthesis related proteins was evident. Similar results have been documented by several researchers while analyzing the defense responses in plants Bilgin et al., 2010). Bilgin et al., (2010) carried out an extensive analysis of transcriptome data from microarray experiments after twenty two different forms of biotic damage on eight different plant species. Photosynthesis-related genes were universally downregulated regardless of the type of biotic attack. However, reduction in gene expression does not necessarily correspond to loss of function. Moreover, abiotic stresses like drought, salinity and unfavourable temperatures, also hamper photosynthesis. Gas-exchange characteristics, photosynthetic pigments, photosystems, components of electron transport

system and activities of different enzymes involved in carbon metabolism are affected in varied intensities by such abiotic stresses (Ashraf and Harris, 2013).

Phytohormones and ROS species play a fundamental role in defense responses against biotic stresses. Plants produce endogenous signal molecules such as jasmonic acid (JA), ethylene (ET), abscisic acid (ABA) and salicylic acid (SA) to regulate signal transduction in plant cells, to activate and/or modulate gene expression upon induction. The jasmonate family of regulators is known to activate defense genes, influence development of physical defenses, interact with pathogen-activated defense signal pathways and control the production of volatiles that participate in indirect defense processes (Wu and Baldwin, 2010). Jasmonic acid or its derivatives may act as long-distance transmissible signal for wound responses in plants and induce the accumulation of PI proteins. Extensive cross-talk between JA, ET and SA signal-transduction pathways allows the plant to fine-tune its defenses against different types of pathogens and insect attackers (Bilgin et al., 2010). Recent studies have also implicated the role of other hormones such as abscisic acid (ABA), auxin, gibberellic acid (GA), cytokinin (CK), brassinosteroids (BR) and peptide hormones in plant defense signaling (Bari and Jones, 2009).

Induction of biotic or abiotic stress disrupts the metabolic balance of cells, resulting in shift in ROS homeostasis. The ROS network responds diversely to each type of stress. ROS has a role in enhancing the expression of different transcription factors like WRKY, Zat, RAV, GRAS and MYB families (Mittler et al., 2004).

To bring about such adaptive and coordinated responses, intricate regulatory mechanisms are required at transcriptional level. Transcription may be coordinated by *cis*-regulatory elements in the promoter region of responsive genes. Promoter elements like SORLIP1, G-box, I-box and T-box were found to be present in most of the stress responsive genes in the analysis conducted by Bilgin et al., (2010). In case of Pin-II gene expression in solanaceous plants, promoter elements like ABA responsive element (ABRE), MYB-binding motifs, L1 box WUN-motif, ELI-box3, TCA-element are considered to be involved (Lorberth et al., 1992; Zhang et al., 2004; Liu et al., 2006).

Various transcription factors like ethylene-responsive-element-binding factors (ERF), basic-domain leucine-zipper (bZIP), WRKY proteins, MYC, MYB, NAC domain,

bHLH and DOF (DNA binding with one finger) have been linked to biotic stress (Singh et al., 2002; van Verk et al., 2009). MYC transcription factors participate in the regulation of many JA-dependent physiological processes like defense against herbivores/pathogens, drought tolerance, circadian clock, light signaling and root growth (Boter et al., 2004). In *N. attenuata* plants, transcripts of *MYC2* showed transient upregulation in induced leaves locally (Woldemariam et al., 2013). Similarly, *Pieris rapae* infestation induced local expression of *AtMYB102* in *A. thaliana*. ERF1 and MYC2 have been known to regulate JA and ET mediated signaling (Bari and Jones, 2009). A considerable overlap has been observed in the induction characteristics of these transcription factors. Such overlap might be involved in integration of different signals consequently fine-tuning the plant defense responses to attackers activating diverse signal transduction pathways (van Verk et al., 2009).

In the present study, the gene architecture of two CanPI genes was detailed (Chapter 2B). One of the interesting finding of this study was that most of the CanPI genes if not all have unique 3' untranslated regions. Furthermore, we newly identified CanPI genes belonging to bracelet type Pin-II genes in *C. annuum*. Interstingly, investigating the genomic architecture of CanPI genes not only revealed their conserved sequence characteristics but also led to the identification of several regulatory elements in the upstream sequences (promoter region). These predicted elements were previously known to play role in various endogenous and/or defense related functions. Further experimental investigation ascertained the binding of zinc finger proteins, MYB, MYC, MADS-box and DOF type transcription factors to the promoter regions. However, how these transcription factors influence the CanPI gene expression requires elaborate functional analysis.

Further analysis of various aspects of plant defense strategies will help unfold the complex interactions in the expression of defense genes and aid in developing efficient expression vectors for these plant defense molecules.

5.2. H. armigera responses to CanPI ingestion

Plant protease inhibitors are proteinaceous, post-ingestive defense molecules which cause "indigestion" in the insect by inhibiting its gut proteases (Green and Ryan, 1972; Felton, 2005). The efficacy of PIs to reduce insect growth and development has been demonstrated by feeding the larvae on diets containing PIs and their transgenic expression in plants (Hilder et al., 1987; Dunse et al., 2010a). Also, PIs from non-host plants are resistant to degradation; can effectively inhibit gut proteases and thus, hamper the larval growth (Harsulkar et al., 1999; Giri et al., 2003). *C. annuum*, a non-host plant for *H. armigera*, posseses an array of PI genes which can be upregulated upon insect infestation/wounding (Tamhane et al., 2009; Mishra et al., 2012). Proteolytic processing of multi-IRD CanPIs by insect gut proteases to release the individual inhibitory domains (for maximum inhibition) and their stability in insect gut was an encouraging finding (Mishra et al., 2010). Diverse Pin-II PIs from *C. annuum* differentially influenced the growth and development in *H. armigera* (Tamhane et al., 2007).

One of the potent CanPIs, CanPI-7, a 4-domain inhibitor having two variant domains for trypsin and chymotrypsin inhibition each, was shown to be effective in retarding *H. armigera* growth progression (Tamhane et al., 2007). In the present study we investigated the optimal dose of recombinant CanPI-7 (rCanPI-7) for effective antibiosis to *H. armigera*. The biochemical and transcript abundance of proteases, in response to rCanPI-7 ingestion, was also characterized (Chapter 3). Though *H. armigera* dynamically regulated the protease expression, we conclude that early exposure and specific dose of protease inhibitor are essential for effective antibiosis. Proteomic analysis of the rCanPI-7 fed larval gut tissue also showed how insects struggled to obtain nutrition and energy under the "PI-stress". To successfully establish an insect control strategy using PI genes, an optimal dose of inhibitor needs to be implemented (De Leo et al., 1998; Zhu-Salzman et al., 2003). Monitoring the temporal response of trypsin and chymotrypsin in the third instar *H. armigera* larvae upon rCanPI-7 ingestion, depicted a transformation in the expression patterns of the transcripts. The transcripts were upregulated in the early period of feeding on rCanPI-7 and gradually stabilized through reduced upregulation and downregulation of certain genes. Similar observations about temporal response and general upregulation of proteases were reported by Bown et al., (2004) and Srinivasan et

al., (2005) after feeding *H. armigera* larve with soybean and Chickpea Kunitz trypsin inhibitor (KTI), respectively.

Though inhibition of *H. armigera* gut proteases by rCanPI-7 is clearly evident, efficient application of this PI for insect control requires detailed knowledge about the regulation of *H. armigera* gut proteases. The mechanism by which the insect perceives the presence of proteinase inhibitors, and modulates its protease gene expression is not very well understood (Srinivasan et al., 2006; Dunse and Anderson 2011). Peptide hormone systems or peptides hydrolyzed from diet have been hypothesized to play a role in induction of specific proteases (Bown et al., 2004). Nevertheless, the actual mechanism remains to be dissected.

Advances in molecular biology in the last two decades have revolutionized the approaches used to study various aspects of gene function. Development of next generation sequencing platforms, has been a major factor which has allowed deep sequencing of small RNAs with progressive reductions in costs. MicroRNAs (miRNAs) are small non-coding RNAs operating as master regulators of gene expression in various biological processes. MiRNAs regulate gene expression at the post-transcriptional level by translational suppression or mRNA degradation and silencing target gene expression (Bartel, 2004). Levels of miRNAs may vary temporally and spatially thereby fine tuning the target gene expression. In insects, miRNAs are known to regulate diverse biological processes like development, host-pathogen interactions, immunity, regulation of blood-meal activated physiological events, circadian rhythm, olfaction, hormone signaling, regulation of hematopoiesis and hemocyte differentiation, turnover of maternal mRNAs during the maternal-to-zygotic transition, stress resistance and fat metabolism, pigmentation and eclusion and age-dependent behavioral changes (Asgari, 2013). Recently, the gene expression of a trypsin like proteinase (Ha-TLP) gene in *H. armigera*, was shown to be influenced by a miRNA, har-miR-2002b (Jaychandran et al., 2013).

Using high throughput sequencing and bioinformatics approaches, we performed a systematic analysis of miRNAs in *H. armigera* to determine its correlation to protease gene expression (Chapter 4). *H. armigera* larvae were fed on rCanPI-7 diet and subjected to unbiased miRNA analysis at progressive time intervals (0.5 to 48 h). Several miRNAs were found to be differentially expressed in rCanPI-7 fed larvae. *In silico* target

prediction allowed the identification of protease targeting miRNAs. Experimental gene expression analysis revealed a negative correlation in the relative abundance of these miRNAs and their target mRNA transcripts across selected time points, indicating the possible involvement of miRNAs in protease gene regulation. However, *in vivo* functional validation will further substantiate the specific role of miRNAs in regulation of protease gene expression.

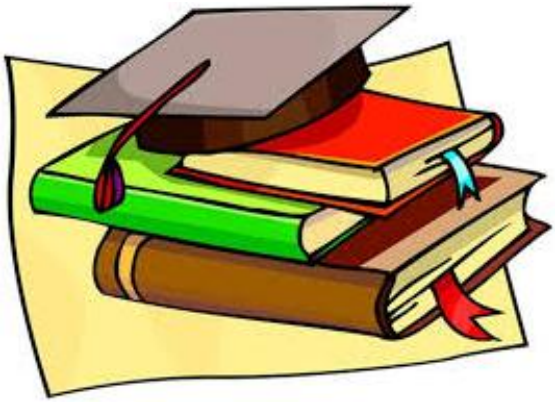
5.3. Designing competent strategies for H. armigera control

Natural plant defense tactics have been explored for devising methods to control insect pests. PIs from non-host plants were shown to be more effective in retarding *H. armigera* growth than that from host plants. *C. annuum* therefore, with an array of Pin-II PI genes presents a diverse pool of inhibitory molecules. Knowledge of the plant response upon herbivory and CanPI gene sequence features (UTRs and promoter) will surely aid in designing adept expression vectors for these genes. Synthetic constructs can be designed by selecting appropriate regulatory elements for tissue specific and induced expression of CanPI genes in transgenics.

On the other hand, comprehensive analysis of insect response at biochemical and molecular level can help in understanding how the target insect counters the ingested PI. This can not only throw light on the regulatory mechanisms operating in the insect but also bring forth novel molecules in the insect that can be targeted for insect control. For instance, the knowledge of protease targeting miRNAs can help in fabricating synthetic dsRNA molecules that can independently or in conjugation with PIs contribute in effective insect control.

5.4. Future directions

The results from the present work have highlighted some aspects that can help in furthering our understanding of plant and insect interactions. Some of the leads that can be followed are: (i) Functional validation of regulatory elements (promoters and UTRs) governing CanPI expression; (ii) Recombinant expression of bracelet type CanPIs and examining their efficacy in affecting insect growth and development; (iii) *in vivo* validation of roles of miRNA in protease regulation and insect control; (iv) detail transcriptomic analysis of *H. armigera* after ingestion of rCanPI-7 (v) *in planta* expression of CanPI-7 protein in host plants of *H. armigera* and measuring level of protection from insect damage in transgenic plants.



Bibliography

BIBLIOGRAPHY

Ambros V (2004) The functions of animal microRNAs. Nature 431:350-355

Asgari S (2013) MicroRNA functions in insects. Insect Biochem. Mol. Biol. 43:388–397

Ashraf M, Harris PJC (2013) Photosynthesis under stressful environments: An overview Review. Photosynthetica 51:163-190

Badger MR, Price GD (1994) The role of carbonic anhydrase in photosynthesis. Annu. Rev. Plant Physiol. Plant Mol. Biol. 45:369-92

Barvkar VT, Pardeshi VC, Kale SM, Qiu S, Rollins M, Datla R, Gupta VS, Kadoo NY, (2013) Genome-wide identification and characterization of microRNA genes and their targets in flax (*Linum usitatissimum*). Planta 237:1149-1161

Bari R, Jones JDG (2009) Role of plant hormones in plant defence responses. Plant Mol. Biol. 69:473-488

Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C, Kim IF, Soboleva A, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Muertter RN, Edgar R (2009) NCBI GEO: archive for high-throughput functional genomic data. Nucleic Acids Res. 37:D885-D890

Barrette-Ng IH, Kenneth KSN, Cherney MM, Pearce G, Ghani U, Ryan CA, James MNG (2003a) Unbound form of tomato inhibitor-II reveals inter-domain flexibility and conformational variability in the reactive site loops. J. Biol. Chem. 278: 31391-31400

Barrette-Ng IH, Kenneth KSN, Cherney MM, Pearce G, Ryan CA, James MNG (2003b) Structural basis of inhibition revealed by 1:2 complex of the two headed tomato inhibitor-II and subtilisin Carlsberg. J. Biol. Chem. 278: 24062-24071

Barta E, Pintar A, Pongor S (2002) Repeats with variations: accelerated evolution of the Pin2 family of proteinase inhibitors. Trends Genet. 18:600-603

Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116:281-297.

Betel D, Wilson M, Gabow A, Marks DS, Sander C (2008) The microRNA.org resource: targets and expression. Nucleic Acids Res. 36:D149–D153.

Bezzi S, Kessler D, Diezel C, Muck A, Anssour S, Baldwin IT (2010) Silencing NaTPI expression increases nectar germin, nectarins, and hydrogen peroxide levels and inhibits nectar removal from plants in nature. Plant Physiol. 152:2232-2242

Bilgin DD, Zavala JA, Zhu J, Clough SJ, Ort DR, DeLucia EH (2010) Biotic stress globally downregulates photosynthesis genes. Plant Cell Environ. 33:1597-1613

Boter M, Ruíz-Rivero O, Abdeen A, Prat S (2004) Conserved MYC transcription factors play a key role in jasmonate signaling both in tomato and Arabidopsis. Genes Dev. 18:1577-1591

Boulter D (1993) Insect pest control by copying nature using genetically engineered crops. Phytochemistry 34:1453-1466

Bown DP, Wilkinson HS, Gatehouse JA (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 DP, Wilkinson HS, Gatehouse JA (1998) Midgut carboxypeptidase from *Helicoverpa armigera* (Lepidoptera: Noctuidae) larvae: enzyme characterisation, cDNA cloning and expression. Insect Biochem. Mol. Biol. 28:739-749

Bown DP, Wilkinson HS, Gatehouse JA (2004) Regulation of expression of genes encoding digestive proteases in the gut of a polyphagous lepidopteran larva in response to dietary protease inhibitors. Physiol. Entomol. 29:278-290

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72:248-254

Broadway RM (1995) Are insects resistant to plant proteinase inhibitors? J. Insect Physiol. 41:107–116.

Broadway RM, (1997) Dietary regulation of serine proteases that are resistant to serine protease inhibitors. J. Insect Physiol. 43:855-874

Browne BL, Raubenheimer D (2003) Ontogenetic changes in the rate of ingestion and estimates of food consumption in fourth and fifth instar *Helicoverpa armigera* caterpillars J. Insect Physiol. 49:63-71

Burgess EPJ, Main CA, Stevens PS, Christeller JT, Gatehouse AMR, Laing WA (1994) Effects of protease inhibitor concentration and combinations on the survival, growth and gut enzyme activities of the black field cricket, *Teleogryllus commodus*, J. Insect Insect Physiol. 40:803-811

Bussing I, Slack FJ, Grosshans H (2008) let-7 microRNAs in development, stem cells and cancer. Trends Mol. Med. 14:400-409

Chikate YR, Tamhane VA, Joshi RS, Gupta VS, Giri AP (2013) Differential protease activity augments polyphagy in *Helicoverpa armigera*. Insect Mol. Biol. 22:258-272

Chilana P, Sharma A, Arora V, Bhati J, Rai A (2013) Computational identification and characterization of putative miRNAs in *Heliothis virescens*. Bioinformation 9:79-83

Chougule NP, Giri AP, Sainani MN, Gupta VS (2005) Gene expression patterns of *Helicoverpa armigera* gut proteases. Insect Biochem. Mol. Biol. 35:355-367

Christeller JT (2005) Evolutionary mechanisms acting upon proteinase inhibitor variability. FEBS J. 272:5710–5722

Corpas FJ, Barroso JB, Sandalio LM, Palma JM, Lupiáñez JA, del Río LA (1999) Peroxisomal NADP-Dependent isocitrate dehydrogenase characterization and activity regulation during natural senescence. Plant Physiol. 121:921-928

Creighton CJ, Reid JG, Gunaratne PH (2009) Expression profiling of microRNAs by deep sequencing. Brief. Bioinform. 10:490-497

Dahiya SS, Chauhan YS, Johansen C, Shanower TG (1999) Adjusting pigeonpea sowing time to manage pod borer infestation. Inter. Chickpea and Pigeonpea Newslett. 6:44-45

Damle MS, Giri AP, Sainani MN, Gupta VS (2005) Higher accumulation of proteinase inhibitors in flowers than leaves and fruits as a possible basis for differential feeding preference of *Helicoverpa armigera* on tomato (*Lycopersicon esculentum* Mill, Cv. Dhanashree). Phytochemistry 66:2659-2667

Dawkar VV, Chikate YR, Gupta VS, Slade SE, Giri AP (2011) Assimilatory potential of *Helicoverpa armigera* reared on host (Chickpea) and nonhost (*Cassia tora*) diets. J. Proteome Res. 10:5128-5138.

Dawkar VV, Chikate YR, Lomate PR, Dholakia B, Gupta VS, Giri AP (2013) Molecular insights into resistance mechanisms of lepidopteran insect pests against toxicants. J. Proteome Res. 12:4727-4737

Dawkar VV, Chikate YR, More TH, Gupta VS, Giri AP (2014) The expression of proteins involved in digestion and detoxification are regulated in *Helicoverpa armigera* to cope up with chlorpyrifos insecticide. Insect Sci. (In Press)

Day DA, Tuite MF (1998) Post-transcriptional gene regulatory mechanisms in eukaryotes: an overview. J. Endocrinol. 157:361-371

De Leo F, Bonade-Bottino MA, Ceci LR, Gallerani R, 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, Volpicella M, Licciulli F, Liuni S, Gallerani R, Ceci LR (2002) PLANT-PIs: a database for plant protease inhibitors and their genes. Nucleic Acids Res. 30:347-348

Delessert C, Wilson IW, Straeten DVD, Dennis S, Dolferus R (2004) Spatial and temporal analysis of the local response to wounding in *Arabidopsis* leaves. Plant Mol. Biol. 55:165-181

Dhaliwal GS, Jindal V, Dhawan AK (2010) Insect pest problems and crop losses: Changing trends. Indian J. Ecol. 37:1-7

Du T, Zamore PD (2005) microPrimer: the biogenesis and function of microRNA. Development 132:4645-4652

Duan X, Li X, Xue Q, Abo el Saad M, Xu D, Wu R (1996) Transgenic rice plants harbouring an introduced potato proteinase inhibitor II gene are insect resistant. Nat. Biotechnol. 14: 494-498

Dunse KM, Stevens JA, Lay FT, Gaspar YM, Heath RL, Anderson MA (2010a) Coexpression of potato type I and II proteinase inhibitors gives cotton plants protection against insect damage in the field. Proc. Natl. Acad. Sci. USA 107:15011-15015

Dunse KM, Kaas Q, Guarino RF, Barton PA, Craik DJ, Anderson MA (2010b) Molecular basis for the resistance of an insect chymotrypsin to a potato type II proteinase inhibitor. Proc. Natl. Acad. Sci. USA 107:15016-15021

Dunse K, Anderson M (2011) Towards the next generation of pest resistant plants. ISB News Report, June 2011

Edmonds HS, Gatehouse LN, Hilder VA, Gatehouse JA (1996) The inhibitory effects of the cysteine protease inhibitor, oryzacystatin, on digestive proteases and on larval survival and development of the southern corn rootworm (*Diabrotica undecimpunctata* howardi). Entomol. Exp. Appl. 78:83–94

Ehrlich PR, Raven PH (1964) Butterflies and Plants: A Study in Coevolution. Evolution 18:586-608

Erb M, Meldau S, Howe GA (2012) Role of phytohormones in insect-specific plant reactions. Trends Plant Sci. 17:250-259

Eveland AL, McCarty DR, Koch KE (2008) Transcript profiling by 3'untranslated region sequencing resolves expression of gene families. Plant Physiol. 146:32-44

Felton GW (2005) Indigestion is a plant's best defense. Proc. Natl. Acad. Sci. USA 102:18771-18772

Ferry N, Edwards MG, Gatehouse J, Capell T, Christou P, Gatehouse AMR (2006) Transgenic plants for insect pest control: a forward looking scientific perspective. Trans. Res. 15:13–19

Fitt GP (1989) The ecology of *Heliothis* in relation to agroecosystems. Annu. Rev. Entomol. 34:17-52.

Foyer CH, Bloom AJ, Queval G, Noctor G (2009) Photorespiratory metabolism: Genes, mutants, energetics, and redox signaling. Annu. Rev. Plant Biol. 60:455-484.

Franco OL, Rigden DJ, Melo FR, Grossi-de-Sá MF (2002) Plant α-amylase inhibitors and their interaction with insect α-amylases Structure, function and potential for crop protection. Eur. J. Biochem. 269:397-412

Friedlander MR, Mackowiak SD, Li N, Chen W, Rajewsky N (2012) miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic Acids Res. 40:37-52

Frutos R, Rang C, Royer M (1999) Managing insect resistance to plants producing *Bacillus thuringiensis* toxins. Crit. Rev. Biotechnol. 19:227-276

Gatehouse LN, Shannon AL, Burgess EPJ, Christeller JT (1997) Characterization of major midgut proteinase cDNAs from *Helicoverpa armigera* larvae and changes in

gene expression in response to four proteinase inhibitors in the diet. Insect Biochem. Mol. Biol. 27:929-944

Ge X, Zhang Y, Jiang J, Zhong Y, Yang X, Li Z, Huang Y, Tan A (2013) Identification of MicroRNAs in *Helicoverpa armigera* and *Spodoptera litura* based on deep sequencing and homology analysis. Int. J. Biol. Sci. 9:1-15

Giri AP, Harsulkar AM, Deshpande VV, Sainani MN, Gupta VS, Ranjekar PK (1998) Chickpea defensive proteinase inhibitors can be inactivated by podborer gut proteinases. Plant Physiol. 116:393-401

Giri AP, Harsulkar AM, Ku MSB, Gupta VS, Deshpande VV, Ranjekar PK, Franceschia VR (2003) Identification of potent inhibitors of *Helicoverpa armigera* gut proteinases from winged bean seeds. Phytochemistry 63: 523-532

Giri AP, Wunsche H, Mitra S, Zavala JA, Muck A, Svatos A, Baldwin IT (2006) Molecular interactions between the specialist herbivore *Manduca sexta* (Lepidoptera, Sphingidae) and its natural host *Nicotiana attenuata*. VII. Changes in the plant's proteome. Plant Physiol. 142:1621-1641

Görg A, Weiss W, Dunn MJ (2004) Current two-dimensional electrophoresis technology for proteomics. Proteomics 4:3665-3685

Graber JH, Cantor CR, Mohr SC, Smith TF (1999) *In silico* detection of control signals: mRNA 3'end-processing sequences in diverse species. Proc. Natl. Acad. Sci. USA 96:14055-14060

Green TR, Ryan CA (1972) Wound induced proteinase inhibitor in plant leaves: A possible defence mechanism. Science 175:776-777

Griffiths-Jones S, Grocock RJ, van Dongen, S, Bateman, A, Enright, AJ (2006). miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res. 34:D140–D144

Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ (2008) miRBase: tools for microRNA genomics. Nucleic Acids Res. 36:D154–D158

Gubb D, Sanz-Parra A, Barcena L, Troxler L, Fullaondo A (2010) Protease inhibitors and proteolytic signalling cascades in insects. Biochimie 92:1749-1759

Gulati J, Kim S-G, Baldwin IT, Gaquerel E (2013) Deciphering herbivore attack-induced gene-to-metabolite dynamics in *Nicotiana attenuata* tissues using a multifactorial approach. Plant Physiol. 162:1042–1059

Gupta AS, Heinen JL, Holaday AS, Burke JJ, Allen RD (1993) Increased resistance to oxidative stress in transgenic plants that overexpress chloroplastic Cu/Zn superoxide dismutase. Proc. Natl. Acad. Sci. USA 90:1629-1633

Gurjar G, Mishra M, Kotkar H, Upasani M, Soni P, Tamhane V, Kadoo N, Giri A and Gupta V (2010) Major Biotic Stresses of Chickpea and Strategies for their Control. Pests and Pathogens: Management Strategies, BS Publications, CRC Press, ISBN: 978-0-415-66576-6 pp 87-134

Hara K, Vagi M, Koizumi N, Kusano T, Sano H, (2000) screening of wound-responsive genes identifies an immediate-early expressed gene encoding a highly charged protein in mechanically wounded tobacco plants. Plant Cell Physiol. 41:684-691

Harrison EP, Willingham NM, Lloyd JC, Raines CA (1998) Reduced sedoheptulose-1,7-bisphosphatase levels in transgenic tobacco lead to decreased photosynthetic capacity and altered carbohydrate accumulation. Planta 204:27-36

Harsulkar AM, Giri AP, Gupta VS, Sainani MN, Deshpande VV, Patankar AG, Ranjekar PK (1998) Characterization of *Helicoverpa armigera* gut proteinases and their interaction with proteinase inhibitors using gel -X-ray film contact print technique. Electrophoresis 19:1397-1402

Harsulkar AM, Giri AP, Patankar AG, Gupta VS, Sainani MN, Ranjekar PK Deshpande VV (1999) Successive use of non-host plant proteinase inhibitors required for effective inhibition of gut proteinases and larval growth of *Helicoverpa armigera*. Plant Physiol. 121:497-506.

Hartl M, Giri AP, Kaur H, Baldwin IT (2010) Serine protease inhibitors specifically defend *Solanum nigrum* against generalist herbivores but do not influence plant growth and development. Plant Cell 22:4158-4175

Hartl M, Giri AP, Kaur H, Baldwin IT (2011) The multiple functions of plant serine protease inhibitors: Defense against herbivores and beyond. Plant Signal. Behav. 6:1-3

Heath RL, Barton PA, Simpson RJ, Reid GE, Lim G, Anderson MA (1995) Characterization of the protease processing sites in a multidomain proteinase inhibitor precursor from *Nicotiana alata*. Eur. J. Biochem. 230: 250-257

Heath RL, McDonald G, Christeller JT, Lee M, Bateman K, West J, Van Heeswijck R, Anderson MA (1997) Proteinase inhibitors from *Nicotiana alata* enhance plant resistance to insect pests. J. Insect Physiol. 43:833-842

Hellman LM, Fried MG (2007) Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. Nat. Protoc. 2:1849-1861

Hermsmeier D, Schittko U, Baldwin IT (2001) Molecular interactions between the specialist herbivore *Manduca sexta* (Lepidoptera, Sphingidae) and its natural host *Nicotiana attenuata*. I. Large-scale changes in the accumulation of growth- and defense-related plant mRNAs. Plant Physiol. 125:683-700

Higo K, Ugawa Y, Iwamoto M, Korenaga T (1999) Plant *cis*-acting regulatory DNA elements (PLACE) database. Nucleic Acids Res. 27:297-300

Hilder VA, Gatehouse AMR, Sheerman SE, Barker RF, Boulter D (1987). A novel mechanism of insect resistance engineered into tobacco. Nature 330, 160–163.

Hilder VA, Boulter D (1999) Genetic engineering of crop plants for insect resistance - a critical review. Crop Prot. 18:1177-1191

Horn M, Patankar AG, Zavala JA, Wu J, Dolečková-Marešová L, Vůjtěchová M, Mareš M, Baldwin IT (2005) Differential elicitation of two processing proteases controls the processing pattern of the trypsin proteinase inhibitor precursor in *Nicotiana attenuata*. Plant Physiol. 139:375-388

Howe GA, Jander G (2008) Plant immunity to insect herbivores. Annu. Rev. Plant Biol. 59:41-66

Huang YQ, Brown MR, Lee TD, Crim JW (1998) RF-amide peptides isolated from the midgut of the com earworm, *Helicoverpa zea*, resemble pancreatic polypeptide. Insect Biochem. Mol. Biol. 28:345-356

Ifuku K, Yamamoto Y, Ono T, Ishihara S, Sato F (2005) PsbP protein, but not PsbQ protein, is essential for the regulation and stabilization of photosystem II in higher plants. Plant Physiol. 139:1175-1184

Iwai K, Fishiki T, Fukuoka SI (1998) Pancreatic enzyme secretion mediated by novel peptide: monitor peptide hypothesis. Pancreas 3:720-728

Jagadeeswaran G, Zheng Y, Sumathipala N, Jiang H, Arrese EL, Soulages JL, Zhang W, Sunkar R (2010) Deep sequencing of small RNA libraries reveals dynamic regulation of conserved and novel microRNAs and microRNA-stars during silkworm development. BMC Genomics. 11:52

Jayachandran B, Hussain M, Asgari S (2013) An insect trypsin-like serine protease as a target of microRNA: Utilization of microRNA mimics and inhibitors by oral feeding. Insect Biochem. Mol. Biol. 43:398-406

Jin JP, Zhang H, Kong L, Gao G and Luo JC (2014) PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. Nucleic Acids Res. 42:D1182-D1187.

Johnson R, Narvaez J, An G, Ryan C (1989) Expression of proteinase inhibitors I and II in transgenic tobacco plants: Effects on natural defense against *Manduca sexta* larvae. Proc. Natl. Acad. Sci. USA 86:9871-9875

Johnson ED, Miller EA, Anderson MA (2007) Dual location of a family of proteinase inhibitors within the stigmas of *Nicotiana alata*. Planta 225: 1265-1276

Johnston KA, Lee MJ, Gatehouse JA, Anstee JH (1991) The partial purification and characterization of serine protease activity in midgut of larval *Helicoverpa armigera.* Insect Biochem. 21:389-397

Jongsma MA, Bakker PL, Peters J, Bosch D, Stiekema WJ (1995) Adaptation of *Spodoptera exigua* larvae to plant proteinase inhibitors by induction of gut proteinase activity insensitive to inhibition. Proc. Nat. Acad. Sci. USA 92:8041-8045

Jongsma MA, Bolter C (1997) The adaptation of insects to plant protease inhibitors. J. Insect Physiol. 43:885–895

Joshi CP (1987) An inspection of the domain between putative TATA box and translation start site in 79 plant genes. Nucleic Acids Res. 15:6643-6653

Joshi RS, Mishra M, Tamhane VA, Ghosh A, Sonavane U, Suresh CG, Joshi R, Gupta VS, Giri AP (2014a) The remarkable efficiency of a Pin-II proteinase inhibitor sans two conserved disulfide bonds is due to enhanced flexibility and hydrogen bond density in the reactive site loop. J. Biomol. Struct. Dyn. 32:13-26

Joshi RS, Gupta VS, Giri AP (2014b) Differential antibiosis against *Helicoverpa armigera* exerted by distinct inhibitory repeat domains of *Capsicum annuum* proteinase inhibitors. Phytochemistry 101:16-22

Keil M, Sanchez-serrano J, Schell J, Willmitzer L (1986) Primary structure of a proteinase inhibitor II gene from potato (*Solanum tuberosum*). Nucleic Acids Res. 14:5641-5650

Kessler A, Baldwin IT (2002) Plant responses to insect herbivory: The emerging molecular analysis. Annu. Rev. Plant Biol. 53:299-328

Kim S, Hong Y-N, An CS, Lee K-W (2001) Expression characteristics of serine proteinase inhibitor II under variable environmental stresses in hot pepper (*Capsicum annuum* L.). Plant Sci. 161:27-33

Kim SR, Choi JL, Costa MA, An G (1992) Identification of G-box sequence as an essential element for methyl jasmonate response of potato proteinase inhibitor II promoter. Plant Physiol. 99:627-631

Kim VN, Han J, Siomi MC (2009) Biogenesis of small RNAs in animals. Nat. Rev. Mol. Cell Biol. 10:126-139

Kong L, Ranganathan S (2008) Tandem duplication, circular permutation, molecular adaptation: how Solanaceae resist pests via inhibitors. BMC Bioinformatics 9 (Suppl 1):S22

Konus M, Koy C, Mikkat S, Kreutzer M, Zimmermann R, Iscan M, Glocker MO (2013) Molecular adaptations of *Helicoverpa armigera* midgut tissue under pyrethroid insecticide stress characterized by differential proteome analysis and enzyme activity assays. Comp. Biochem. Physiol. Part D Genomics Proteomics 8:152-162

Koßmann J, Sonnewald U, Willmitzer L (1994) Reduction of the chloroplastic fructose-l,6-bisphosphatase in transgenic potato plants impairs photosynthesis and plant growth. Plant J. 6:637-650

Kotkar HM, Sarate PJ, Tamhane VA, Gupta VS, Giri AP (2009) Responses of midgut amylases of *Helicoverpa armigera* to feeding on various host plants. J. Insect Physiol. 55:663-670

Kozak M (2001) New ways of initiating translation in eukaryotes? Mol. Cell. Biol. 21:1899-1907

Lammers, JW, Macleod A (2007) Report of a pest risk analysis Distribution (Plant Protection Service (NL) and Central Science Laboratory (UK) joint Pest Risk Analysis for *Helicoverpa armigera*), 1-18

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10:R25

Lau NC, Lim LP, Weinstein EG, Bartel DP (2001) An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. Science 294: 858-862

Lazova G, Naidenova N, Ignatova L, Stefanov D (2009) A pea mutant (costata) expressing higher activity in thylakoid membrane-bound carbonic anhydrase alters PSII downregulation mechanisms. Cell Biol. Int. 33:867-873

Lee MCS, Scalon MJ, Craik DJ, Anderson MA (1999) A novel two-chain proteinase inhibitor generated by circularization of a multidomain precursor protein. Nat. Struct. Biol. 6:526-530

Lee RC, Feinbaum RL, Ambros V (1993) The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 75:843-854

Leterrier M, LUIS A. Delrío LA, Corpas FJ (2007) Cytosolic NADP-isocitrate dehydrogenase of pea plants: Genomic clone characterization and functional analysis under abiotic stress conditions. Free Radical Res. 41:191-199

Liddle RA (1995) Regulation of cholecystokinin secretion by intraluminal releasing factors. Am. J. Physiol. 269:G319-G327

Liu J, Xia KF, Zhu JC, Deng YG, Huang XL, Hu BL, Xu X, Xu ZF (2006) The nightshade proteinase inhibitor IIb gene is constitutively expressed in glandular trichomes. Plant Cell Physiol. 47:1274-1284

Liu S, Li D, Li Q, Zhao P, Xiang Z, Xia Q (2010) MicroRNAs of *Bombyx mori* identified by Solexa sequencing. BMC Genomics. 11:148

Loke JC, Stahlberg EA, Strenski DG, Haas BJ, Wood PC, Li QQ (2005) Compilation of mRNA polyadenylation signals in Arabidopsis revealed a new signal element and potential secondary structures. Plant Physiol. 138:1457-1468

Lomate PR, Hivrale VK (2011) Differential responses of midgut soluble aminopeptidases of *Helicoverpa armigera* to feeding on various host and non-host plant diets. Artho. Plant Intr. 5:359-368

Lomate PR, Jadhav BR, Giri AP, Hivrale VK (2013) Alterations in the *Helicoverpa armigera* midgut digestive physiology after ingestion of pigeon pea inducible leucine aminopeptidase. PLoS One 8:e74889.

Lorberth R, Dammann C, Ebneth M, Amati S, Sanchez-serrano JJ (1992) Promoter elements involved in environmental and developmental control of potato proteinase inhibitor II expression. Plant J. 2:477-486

Lu J, Shen Y, Wu Q, Kumar S, He B, Shi S, Carthew RW, Wang SM, Wu CI (2008) The birth and death of microRNA genes in *Drosophila*. Nat. Genet. 40:351-355

Luo M, Wang Z, Li H, Xia KF, Cai Y, Xu ZF (2009) Overexpression of weed (*Solanum americanum*) proteinase inhibitor in transgenic tobacco results in increased glandular trichome density and enhanced resistance to *Helicoverpa armigera* and *Spodoptera litura*. Int. J. Mol. Sci. 10:1896-1910

MacIntosh SC, Kishore GM, Perlak FJ, Marrone PG, Stone TB, Sims SR, Fuchs RL (1990) Potentiation of Bacillus thuringiensis insecticidal activity by serine protease inhibitors. J. Agric. Food Chem. 38:1145-1152

Mano J, Torii Y, Hayashi S, Takimoto K, Matsui K, Inzé D, Babiychuk E (2002) The NADPH:Quinone oxidoreductase P1-ζ-crystallin in *Arabidopsis* catalyzes the α,β-hydrogenation of 2-alkenals: Detoxication of the lipid peroxide-derived reactive aldehydes. Plant Cell 43:1445-1455

Marco A, Hui JH, Ronshaugen M, Griffiths-Jones S (2010) Functional shifts in insect microRNA evolution. Genome Biol. Evol. 2:686-696

Mazumdar-Leighton S, Broadway RM (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

Mejía-Teniente L, Durán-Flores F, María C-Oliver A, Torres-Pacheco I, Cruz-Hernández A, González-Chavira MM, Ocampo-Velázquez RV, Guevara-González RG (2013) Oxidative and molecular responses in *Capsicum annuum* L. after hydrogen peroxide, salicylic acid and chitosan foliar applications. Int. J. Mol. Sci.14:10178-10196.

Mishra M, Tamhane VA, Khandelwal N, Kulkarni MJ, Gupta VS, Giri AP (2010) Interaction of recombinant CanPIs with *Helicoverpa armigera* gut proteases reveals their processing patterns, stability and efficiency. Proteomics 10:2845-2857

Mishra M, Mahajan N, Tamhane VA, Kulkarni MJ, Baldwin IT, Gupta VS, Giri AP (2012) Stress inducible proteinase inhibitor diversity in *Capsicum annuum*. BMC Plant Biol 12:217

Mishra M (2012) Interactions between diverse proteinase inhibitors from *Capsicum annuum* and insect pests: A biochemical & molecular approach. Ph.D. dissertation University of Pune, Pune (India)

Mishra M, Joshi RS, Gaikwad S, Gupta VS, Giri AP (2013) Structural-functional insights of single and multi-domain *Capsicum annuum* protease inhibitors. Biochem. Biophys. Res. Commun. 430:1060-1065

Miska EA (2005) How microRNAs control cell division, differentiation and death. Curr. Opin. Genet. Dev. 15:563-568

Mittler R, Vanderauwera S, Gollery M, Van Breusegem F (2004) Reactive oxygen gene network of plants. Trends Plant Sci. 9:490-498

Mosolov VV, Valueva TA (2005) Proteinase inhibitors and their function in plants: A review. Appl. Biochem. Micro. 41: 227-246

Moura DS, Ryan CA (2001) Wound-inducible proteinase inhibitors in pepper. Differential regulation upon wounding, systemin, and methyl jasmonate. Plant Physiol. 126:289-298

Mueller LA, Solow TH, Taylor N, Skwarecki B, Buels R, Binns J, Lin C, Wright MH, Ahrens R, Wang Y, Herbst EV, Keyder ER, Menda N, Zamir D, Tanksley SD (2005) The SOL Genomics Network. A comparative resource for Solanaceae biology and beyond. Plant Physiol. 138:1310-1317

Mulo P (2011) Chloroplast-targeted ferredoxin-NADP+ oxidoreductase (FNR): Structure, function and location. BBA-Bioenergetics 1807:927-934

Nagarkatti S, Prakash A (1974) Rearing *Heliothis armigera* (Hubn) on artificial diet Technical Bulletin 17. Commonwealth Institute of Biological Control, Banglore. pp 169-173

Neurath H, Walsh KH (1976) Role of proteolytic enzymes in biological regulation (A Review). Proc. Natl. Acad. Sci. USA 73:3825-3832

Noctor G (2006) Metabolic signalling in defence and stress: the central roles of soluble redox couples. Plant Cell. Environ. 29:409-425

Nogueira FCS, Silva CP, Alexandre D, Samuels RI, Soares EL, Aragão FJL, Palmisano G, Domont GB, Roepstorff P, Campos FAP (2012) Global proteome changes in larvae of *Callosobruchus maculatus* (Coleoptera:Chrysomelidae:Bruchinae) following ingestion of a cysteine proteinase inhibitor. Proteomics 12: 2704-2715

Odeny DA, Stich B, Gebhardt C (2010) Physical organization of mixed protease inhibitor gene clusters, coordinated expression and association with resistance to late blight at the *StKI* locus on potato chromosome III. Plant Cell Environ. 33:2149-2161

Ölçer H, Lloyd JC, Raines CA (2001) Photosynthetic capacity is differentially affected by reductions in sedoheptulose-1,7-bisphosphatase activity during leaf development in transgenic tobacco plants. Plant Physiol. 125:982-989

Park M, Park J, Kim S, Kwon JK, Park HM, Bae IH, Yang TJ, Lee YH, Kang BC, Choi D (2012) Evolution of the large genome in *Capsicum annuum* occurred through accumulation of single-type long terminal repeat retrotransposons and their derivatives. Plant J. 69:1018-29

Patankar AG, Giri AP, Harsulkar AM, Sainani MN, Deshpande VV, Ranjekar PK, Gupta VS (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

Pena-Cortes H, Willmitzer L, Sánchez-Serrano JJ (1991) Abscisic acid mediates wound induction but not developmental-specific expression of the proteinase inhibitor II gene family. Plant Cell 3: 963-972

Pena-Cortes H, Albrecht T, Prat S, Weiler E Willmitzer L (1993) Aspirin prevents wound-induced gene expression in tomato leaves by blocking jasmonic acid biosynthesis. Planta 191:123-128

Petrov VD, Breusegem FV (2012) Hydrogen peroxide- a central hub for information flow in plant cells, AoB Plants pls014 doi:10.1093/aobpla/pls014.

Raines CA, Lloyd JC, Dyer TA (1999) New insights into the structure and function of sedoheptulose-1,7-bisphosphatase; an important but neglected Calvin cycle enzyme. J. Exp. Bot. 50:1-8

Rajapakse CNK, Walter GH (2007) Polyphagy and primary host plants: oviposition preference versus larval performance in the lepidopteran pest *Helicoverpa armigera*. Arthropod Plant Interact. 1: 17-26

Rao Z, He W, Liu L, Zheng S, Huang L, Feng Q (2012) Identification, expression and target gene analyses of MicroRNAs in *Spodoptera litura*. PLoS One 7: e37730

Rawlings ND, Tolle DP, Barrett AJ (2004) Evolutionary families of peptidase inhibitors Biochem. J. 378:705-716

Rehmsmeier M, Steffen P, Höchsmann M, Giegerich R (2004) Fast and effective prediction of microRNA/target duplexes. RNA 10:1507-1517

Reymond P, Farmer EE (1998) Jasmonate and salicylate as global signals for defense gene expression. Curr Opin Plant Biol.1:404-411

Rojas CM, Senthil-Kumar M, Wang K, Ryu C-M, Kaundal A, Mysore KS (2012) Glycolate oxidase modulates reactive oxygen species–mediated signal transduction during nonhost resistance in *Nicotiana benthamiana* and *Arabidopsis*. Plant Cell 24:336-352

Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. Nat. Protoc. 5:725-738

Ruby JG, Stark A, Johnston WK, Kellis M, Bartel DP, Lai EC (2007) Evolution, biogenesis, expression, and target predictions of a substantially expanded set of *Drosophila* microRNAs. Genome Res. 17:1850–1864

Ruuhola T, Yang S. (2006) Wound-induced oxidative responses in mountain birch leaves. Ann. Bot. 97:29-37

Ryan CA (1990) Proteinase inhibitors in plants, genes for improving defenses against insects and pathogens. Annu. Rev. Phytopath. 28:425-449

Sarate PJ, Tamhane VA, Kotkar HM, Ratnakaran N, Susan N, Gupta VS, Giri AP (2012) Developmental and digestive flexibilities in the midgut of a polyphagous pest, the cotton bollworm, *Helicoverpa armigera*. J. Insect Sci. 12:1-16

Scanlon MJ, Lee MC, Anderson MA, Craik DJ (1999) Structure of a putative ancestral protein encoded by a single sequence repeat from a multidomain proteinase inhibitor gene from *Nicotiana alata*. Structure 7:793-802

Schafleitner R, Wilhelm E (2001) Assessment of stress gene expression in chestnut (*Castanea sativa* Mill.) upon pathogen infection (*C. parasitica* [Murr.] Barr) and wounding. For. Snow Landsc. Res. 76:409-414

Schilmiller AL, Howe GA (2005) Systemic signaling in the wound response. Curr. Opin. Plant Biol. 8:369-377

Schirra HJ, Craik DJ (2005) Structure and folding of potato type II proteinase inhibitors: circular permutation and intramolecular domain swapping. Protein Pept. Lett. 12:421-431

Schlüter U, Benchabane M, Munger A, Kiggundu A, Vorster J, Goulet M-C, Cloutier C, Michaud D (2010) Recombinant protease inhibitors for herbivore pest control: a multitrophic perspective. J. Exp. Bot. 61:4169-4183

Schoonhoven LM, van Loon JJA, Dicke M (2005) Insect-plant biology. Oxford University Press, New York

Schuster AM, Davies E (1983) Ribonucleic acid and protein metabolism in pea epicotyls Plant Physiol. 73:809-816

Sharma HC, Sharma KK, Seetharam N, Ortiz R (2000) Prospects for using transgenic resistance to insects in crop improvement. Electron. J. Biotechnol. 3:76-95

Sharma HC, Sharma KK, Seetharam N, Ortiz R (2001) Genetic transformation of crop plants:Riska and opportunities for the rural poor. Curr. Sci. 80:1495-1508

Sharma P, Jha AB, Dubey RS, Pessarakli M (2012) Reactive oxygen species, oxidative damage, and antioxidative defense mechanism in plants under stressful conditions defense J. Bot. doi:10.1155/2012/217037

Shin R, Lee GJ, Park CJ, Kim TY, You JS, Nam YW, Paek KH (2001) Isolation of pepper mRNAs differentially expressed during the hypersensitive response to tobacco mosaic virus and characterization of a proteinase inhibitor gene. Plant Sci. 161:727-737

Sin S, Yeung EC, Chye M (2006) Down regulation of *Solanum americanum* genes encoding proteinase inhibitor II causes defective seed development. Plant J. 45:58-70

Sin SF, Chye ML (2004) Expression of proteinase inhibitor II proteins during floral development in *Solanum americanum*. Planta 219:1010-1022

Singh KB, Foley RC, Oñate-Sánchez L (2002) Transcription factors in plant defense and stress responses. Curr. Opin. Plant Biol. 5:430-436

Singh TVK, Satyanarayana J, Peshin R (2014) Integrated Pest Management (Pesticide Problems, Vol. 3) Crop loss assessment in India- Past experiences and future strategies. Springer Netherlands pp 227-243

Sokol NS, Xu P, Jan YN, Ambros V (2008) Drosophila let-7 microRNA is required for remodeling of the neuromusculature during metamorphosis. Genes Dev. 22:1591-1596

Sparla F, Tedeschi G, Trost P (1996) NAD(P)H:(Quinone-Acceptor) oxidoreductase of tobacco leaves is a flavin mononucleotide-containing flavoenzyme. Plant Physiol. 112:249-258

Srinivasan A, Giri AP, Gupta VS (2006) Structural and functional diversities in lepidopteran serine proteases. Cell. Mol. Biol. Lett. 11:132-154

Srinivasan A, Giri AP, Harsulkar AM, Gatehouse JA, Gupta VS (2005) A Kunitz trypsin inhibitor from chickpea (*Cicer arietinum* L.) that exerts anti-metabolic effect on podborer (*Helicoverpa armigera*) larvae. Plant Mol. Biol. 57:359-374

Suorsa M, Aro E-M (2007) Expression, assembly and auxiliary functions of photosystem II oxygen-evolving proteins in higher plants. Photosynth. Res. 93:89-100.

Surridge AK, Lopez-Gomollon S, Moxon S, Maroja LS, Rathjen T, Nadeau NJ, Dalmay T, Jiggins CD (2011) Characterization and expression of microRNAs in developing wings of the neotropical butterfly *Heliconius melpomene*. BMC Genomics 12:62.

Tamhane VA, Giri AP, Gupta VS (2005a) *Helicoverpa armigera*: Ecology and Control Using Novel Biotechnological Approaches. Encyclopedia of Pest Management DOI: 10.1081/E-EPM-120041174 1-5

Tamhane VA, Chougule NP, Giri AP, Dixit AR, Sainani MN, Gupta VS (2005b) *In vivo* and *in vitro* effect of *Capsicum annum* proteinase inhibitors on Helicoverpa armigera gut proteinases. Biochim. Biophys. Acta 1722:156-167

Tamhane VA, Giri AP, Sainani MN, Gupta VS (2007) Diverse forms of Pin-II family proteinase inhibitors from *Capsicum annuum* adversely affect the growth and development of *Helicoverpa armigera*. Gene 403:29-38

Tamhane VA, Giri AP, Kumar P, Gupta VS (2009) Spatial and temporal expression patterns of diverse Pin-II proteinase inhibitor genes in *Capsicum annuum* Linn. Gene 442:88-98

Tamhane VA, Mishra M, Mahajan NS, Gupta VS, Giri AP (2012) Plant Pin-II family proteinase inhibitors: Structural and functional diversity. Funct. Plant Sci. Biotechnol. 6:42-58

Telang MA, Giri AP, Sainani MN, Gupta VS (2005) Characterization of two midgut proteinases of *Helicoverpa armigera* and their interaction with proteinase inhibitors. J. Insect Physiol. 51:513-522

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673-4680

Valueva TA, Mosolov VV (2004) Role of inhibitors of proteolytic enzymes in plant defense against phytopathogenic microorganisms. Biochemistry (Mosc.) 69:1305-1309

van Verk MC, Gatz C, Linthorst HJM (2009) Transcriptional regulation of plant defense responses. Adv. Bot. Res. 51:397-438

Vishnudasan D, Tripathi MN, Rao U, Khurana P (2005) Assessment of nematode resistance in wheat transgenic plants expressing potato proteinase inhibitor (PIN2) gene. Transgenic Res. 14:665-675

Walling LL (2000) The myriad plant responses to herbivores. J. Plant Growth Regul. 19:195–216

Wang LF, Chai LQ, He HJ, Wang Q, Wang JX, Zhao XF (2010) A cathepsin L-like proteinase is involved in moulting and metamorphosis in *Helicoverpa armigera*. Insect Mol. Biol. 19:99-111

Woldemariam MG, Dinh ST, Oh Y, Gaquerel E, Baldwin IT, Galis I (2013) NaMYC2 transcription factor regulates a subset of plant defense responses in *Nicotiana attenuata*. BMC Plant Biol. 13:73

Wu J and Baldwin IT (2010) New insights into plant responses to the attack from insect herbivores. Annu. Rev. Genet. 44:1-24

Wu J, Hettenhausen C, Baldwin IT (2006) Evolution of proteinase inhibitor defences in North American allopolyploid species of *Nicotiana*. Planta 224:750-760

Wu Y, Llewellyn D, Mathews A, Dennis ES (1997) Adaptation of *Helicoverpa armigera* (Lepidoptera, Noctuidae) to a proteinase inhibitor expressed in transgenic tobacco. Mol. Breeding 3: 371-380

Xu ZF, Qi WQ, Ouyang XZ, Yeung E, Chye ML (2001) A proteinase inhibitor II of *Solanum americanum* is expressed in phloem. Plant Mol. Biol. 47:727-738.

Xu ZF, Teng WL, Chye ML (2004) Inhibition of endogenous trypsin- and chymotrypsin-like activities in transgenic lettuce expressing heterogeneous proteinase inhibitor SaPIN2a. Planta 218:623-629

Yu X, Zhou Q, Li SC, Luo Q, Cai Y, Lin WC, Chen H, Yang Y, Hu S, Yu J (2008) The silk worm (*Bombyx mori*) microRNAs and their expression in multiple developmental stages. PLoS One 3:e2997

Zavala JA, Baldwin IT (2004) Fitness benefits of trypsin proteinase inhibitor expression in *Nicotiana attenuata* are greater than their costs when plants are attacked. BMC Ecol. 4:11

Zavala JA, Patankar AG, Gase K, Hui D, Baldwin IT (2004) Manipulation of endogenous trypsin proteinase inhibitor production in *Nicotiana attenuata* demonstrates their function as anti herbivore defenses. Plant Physiol. 134: 1181-1190.

Zavala JA, Giri AP, Jongsma MA, Baldwin IT (2008) Digestive duet: Midgut digestive proteinases of *Manduca sexta* ingesting *Nicotiana attenuata* with manipulated trypsin proteinase inhibitor expression. PLoS ONE 3 (4), e2008.

Zhang B, Stellwag EJ, Pan X (2009) Large-scale genome analysis reveals unique features of microRNAs. Gene 443:100-109

Zhang HY, Xie XZ, Xu YZ, Wu NH (2004) Isolation and functional assessment of a tomato proteinase inhibitor II gene. Plant Physiol. Biochem. 42:437-444

Zhang X, Zheng Y, Jagadeeswaran G, Ren R, Sunkar R, Jiang H (2012) Identification and developmental profiling of conserved and novel microRNAs in *Manduca sexta*. Insect Biochem. Mol. Biol. 42:381-395

Zhang X, Zheng Y, Jagadeeswaran G, Ren R, Sunkar R, Jiang H (2014) Identification of conserved and novel microRNAs in *Manduca sexta* and their possible roles in the expression regulation of immunity-related genes. Insect Biochem. Mol. Biol. 47:12-22

Zhang Y (2008) I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 9:40

Zhu-Salzman K, Koiwa H, Salzman RA, Shade RE, Ahn JE (2003) Cowpea bruchid *Callosobruchus maculatus* uses a three-component strategy to overcome a plant defensive cysteine protease inhibitor. Insect Mol. Biol. 12:135–145

Appendix I

Table 1: Accession numbers and sequences of primers used for analysing semi-quantitative gene expression (**F**: Forward primer; **R**: Reverse primer)

No.	Accession No.	Gene	Primer Sequence (5'-3')
1	AA840641	18S rRNA	**F** CCGGTCCGCCTATGGTGTGCACCGG **R** CCTCTGACTATGAAATACGAATGCCCC
2	AF145234	NADPH:quinone oxidoreductase	**F** ATGGAAGCAACGGCAATGGCAGC **R** GCAATTGCCGTCGAGTCTCAAGGC
3	AF061517	Cu/Zn superoxide dismutase	**F** TCGCAAGGCTTGGCAACTCTTC **R** GAGAACATCGTTGTCATTGGCCACAGT
4	NM105265	NADP dependent isocitrate dehydrogenase	**F** GGTGTTTGTGCCAGAAGGATCAGAC **R** GGTGTTCATACCAGATTCCAGCCTCC
5	NM180160	Carbonic anhydrase	**F** GCTCAAGAAGTTGAGGATGAGAGG G **R** GCACTCTCCTTCTGAGCACACTGTC

Table 2: Gene specific primers used for the amplification of UTR and genomic sequences (**F**: Forward primer; **R**: Reverse primer)

No.	Primer name	Primer sequence (5'-3')	Used for
1	CanPin1 F	ATGGCTGTTCCCAAAGAA G	3'UTR
2	CanPin1 R	CTGTTCATGCTTTTACTTTTC	Intron and 5'UTR
3	V63 F	CAAAGAAACGCAAAAGAACC	CanPI-7 semi-quantitative
4	V53 R	TCTTGGATCACAGTTCAGAGTG	
5	N1m F	CCCAAAGAAGTTAGTTTCCTT(A/T/G)CC	3'UTR
6	N5m F	GCTGTTCCCAAAGAAGTTAGTTT(C/T/A)CC	
7	N8 F	GTTAGTTTCCTTGCTTCCCTACTTGTA	
8	N12 F	CAACCCATATGCACCAATAGTA	
9	N17 F	TGCTGTGCAGGCCGTAAGGT	
10	N18 F	AACTATTACAGTGCTGACGGGACCC	
11	V63 F	CAAAGAAACGCAAAAGAACC	
12	N62 F	AACGAAAACAAGGTAAACATGGCTGTT	
13	N32 R	ACATAGATTAGCATGACACCGTTACG	5'UTR, Promoter
14	N27 R	GTATTACGGAGATTAGCATGACACTG	
15	N51 R	AACTCTTACCAAGTACAAGTAGGGAAGC	
16	N54 R	GACATATTTTCCATTGCTTCGCGGTAG	
17	N28 R	CAACATACATCATCATTGATTAGAAATG	
18	N61 F	GGGTGAAAGTGTACTGATTTGCTTGATG	Intron, 5'UTR
19	N29 R	CTATGCTGTGGATATGCTCTGAGACTC	
20	UPM (Long)	CTAATACGACTCACTATAGGGCAAGCAGTG GTATCAACGCAGAGT	3'UTR and 5'UTR
21	UPM (Short)	CTAATACGACTCACTATAGGGC	
22	NUP	AAGCAGTGGTATCAACGCAGAGT	
23	Adaptor	GTAATACGACTCACTATAGGGCACGCGTGG TCGACGGCCCGGGCTGGT	Promoter
24	AP1	GTAATACGACTCACTATAGGGC	
25	AP2	ACTATAGGGCACGCGTGGT	

Table 3: Putative *cis*-acting elements as predicted by PLACE for *CanPI-7*

Factor or Site Name	Signal Sequence	PLACE reference no.	Location (-)	Strand (+/-)
ABRELATERD1	ACGTG	S000414	289	(+)
ACGTATERD1	ACGT	S000415	289	(+)
ACGTATERD1	ACGT	S000415	96	(+)
ARR1AT	NGATT	S000454	363	(+)
BOXLCOREDCPAL	ACCWWCC	S000492	16	(+)
CAATBOX1	CAAT	S000028	232	(+)
CACTFTPPCA1	YACT	S000449	446	(+)
CACTFTPPCA1	YACT	S000449	427	(+)
CACTFTPPCA1	YACT	S000449	191	(+)
CARGCW8GAT	CWWWWWWWWG	S000431	90	(+)
CARGCW8GAT	CWWWWWWWWG	S000431	76	(+)
CPBCSPOR	TATTAG	S000491	213	(+)
CURECORECR	GTAC	S000493	147	(+)
DOFCOREZM	AAAG	S000265	406	(+)
DOFCOREZM	AAAG	S000265	204	(+)
DOFCOREZM	AAAG	S000265	182	(+)
DOFCOREZM	AAAG	S000265	112	(+)
GATABOX	GATA	S000039	403	(+)
GATABOX	GATA	S000039	358	(+)
GT1CONSENSUS	GRWAAW	S000198	526	(+)
GT1CONSENSUS	GRWAAW	S000198	282	(+)
GT1CONSENSUS	GRWAAW	S000198	281	(+)
GT1CONSENSUS	GRWAAW	S000198	186	(+)
GT1CONSENSUS	GRWAAW	S000198	185	(+)
GT1CONSENSUS	GRWAAW	S000198	27	(+)
GT1GMSCAM4	GAAAAA	S000453	185	(+)
GTGANTG10	GTGA	S000378	497	(+)
GTGANTG10	GTGA	S000378	376	(+)
GTGANTG10	GTGA	S000378	120	(+)

GTGANTG10	GTGA	S000378	63	(+)
NODCON1GM	AAAGAT	S000461	406	(+)
NODCON2GM	CTCTT	S000462	294	(+)
OSE1ROOTNODULE	AAAGAT	S000467	406	(+)
OSE2ROOTNODULE	CTCTT	S000468	294	(+)
POLASIG1	AATAAA	S000080	523	(+)
POLASIG1	AATAAA	S000080	167	(+)
POLASIG1	AATAAA	S000080	116	(+)
POLASIG3	AATAAT	S000088	23	(+)
POLLEN1LELAT52	AGAAA	S000245	206	(+)
PREATPRODH	ACTCAT	S000450	389	(+)
PYRIMIDINEBOXHVEPB1	TTTTTTCC	S000298	305	(+)
ROOTMOTIFTAPOX1	ATATT	S000098	402	(+)
ROOTMOTIFTAPOX1	ATATT	S000098	259	(+)
ROOTMOTIFTAPOX1	ATATT	S000098	230	(+)
ROOTMOTIFTAPOX1	ATATT	S000098	199	(+)
ROOTMOTIFTAPOX1	ATATT	S000098	128	(+)
SEF4MOTIFGM7S	RTTTTTR	S000103	382	(+)
SEF4MOTIFGM7S	RTTTTTR	S000103	353	(+)
SORLIP1AT	GCCAC	S000482	99	(+)
TAAAGSTKST1	TAAAG	S000387	407	(+)
TATABOX2	TATAAAT	S000109	75	(+)
TATABOX3	TATTAAT	S000110	127	(+)
TATABOX5	TTATTT	S000203	87	(+)
WRKY71OS	TGAC	S000447	341	(+)
WRKY71OS	TGAC	S000447	265	(+)
CAATBOX1	CAAT	S000028	459	(-)
CAATBOX1	CAAT	S000028	361	(-)
CAATBOX1	CAAT	S000028	274	(-)
GATABOX	GATA	S000039	93	(-)
MARTBOX	TTWTWTTWTT	S000067	167	(-)
POLASIG1	AATAAA	S000080	88	(-)

POLASIG2	AATTAAA	S000081	320	(-)
POLASIG3	AATAAT	S000088	277	(-)
POLASIG3	AATAAT	S000088	215	(-)
ROOTMOTIFTAPOX1	ATATT	S000098	260	(-)
ROOTMOTIFTAPOX1	ATATT	S000098	231	(-)
ROOTMOTIFTAPOX1	ATATT	S000098	200	(-)
SEF4MOTIFGM7S	RTTTTTR	S000103	464	(-)
SEF4MOTIFGM7S	RTTTTTR	S000103	253	(-)
-300ELEMENT	TGHAAARK	S000122	445	(-)
ELRECOREPCRP1	TTGACC	S000142	235	(-)
LTRECOREATCOR15	CCGAC	S000153	240	(-)
GT1CONSENSUS	GRWAAW	S000198	140	(-)
GT1CONSENSUS	GRWAAW	S000198	304	(-)
GT1CONSENSUS	GRWAAW	S000198	303	(-)
TATABOX5	TTATTT	S000203	524	(-)
TATABOX5	TTATTT	S000203	484	(-)
TATABOX5	TTATTT	S000203	117	(-)
TATABOX5	TTATTT	S000203	24	(-)
BOXIIPCCHS	ACGTGGC	S000229	99	(-)
POLLEN1LELAT52	AGAAA	S000245	172	(-)
DOFCOREZM	AAAG	S000265	444	(-)
DOFCOREZM	AAAG	S000265	414	(-)
DOFCOREZM	AAAG	S000265	173	(-)
DOFCOREZM	AAAG	S000265	90	(-)
RAV1AAT	CAACA	S000314	530	(-)
HDZIP2ATATHB2	TAATMATTA	S000373	217	(-)
GTGANTG10	GTGA	S000378	508	(-)
GTGANTG10	GTGA	S000378	467	(-)
GTGANTG10	GTGA	S000378	18	(-)
WBOXATNPR1	TTGAC	S000390	234	(-)
-10PEHVPSBD	TATTCT	S000392	202	(-)

ACGTABREMOTIFA2OSEM	ACGTGKC	S000394	99	(-)
INRNTPSADB	YTCANTYY	S000395	479	(-)
ABREATCONSENSUS	YACGTGGC	S000406	99	(-)
ABRELATERD1	ACGTG	S000414	97	(-)
ACGTATERD1	ACGT	S000415	289	(-)
ACGTATERD1	ACGT	S000415	96	(-)
CARGCW8GAT	CWWWWWWWWG	S000431	90	(-)
CARGCW8GAT	CWWWWWWWWG	S000431	76	(-)
WRKY71OS	TGAC	S000447	234	(-)
CACTFTPPCA1	YACT	S000449	148	(-)
CACTFTPPCA1	YACT	S000449	6	(-)
GT1GMSCAM4	GAAAAA	S000453	304	(-)
ARR1AT	NGATT	S000454	159	(-)
ARR1AT	NGATT	S000454	33	(-)
ARR1AT	NGATT	S000454	20	(-)
WBOXNTERF3	TGACY	S000457	235	(-)
NODCON1GM	AAAGAT	S000461	92	(-)
OSE1ROOTNODULE	AAAGAT	S000467	92	(-)
ANAERO1CONSENSUS	AAACAAA	S000477	84	(-)
CPBCSPOR	TATTAG	S000491	169	(-)
CURECORECR	GTAC	S000493	147	(-)
EECCRCAH1	GANTTNC	S000494	244	(-)
BIHD1OS	TGTCA	S000498	341	(-)
BIHD1OS	TGTCA	S000498	265	(-)

Table 4: Putative *cis*-acting elements as predicted by PLACE for *CanPI-69*

Factor or Site Name	Signal Sequence	PLACE reference no.	Location (-)	Strand (+/-)
-10PEHVPSBD	TATTCT	S000392	524	(+)
-10PEHVPSBD	TATTCT	S000392	398	(+)
-10PEHVPSBD	TATTCT	S000392	232	(+)
-300ELEMENT	TGHAAARK	S000122	326	(+)
ABRELATERD1	ACGTG	S000414	641	(+)
ACGTABOX	TACGTA	S000130	80	(+)
ACGTATERD1	ACGT	S000415	641	(+)
ACGTATERD1	ACGT	S000415	634	(+)
ACGTATERD1	ACGT	S000415	579	(+)
ACGTATERD1	ACGT	S000415	79	(+)
ACGTTBOX	AACGTT	S000132	635	(+)
ARR1AT	NGATT	S000454	368	(+)
ARR1AT	NGATT	S000454	217	(+)
ARR1AT	NGATT	S000454	363	(+)
ARR1AT	NGATT	S000454	338	(+)
BIHD1OS	TGTCA	S000498	567	(+)
BIHD1OS	TGTCA	S000498	263	(+)
CAATBOX1	CAAT	S000028	604	(+)
CAATBOX1	CAAT	S000028	469	(+)
CAATBOX1	CAAT	S000028	376	(+)
CAATBOX1	CAAT	S000028	250	(+)
CAATBOX1	CAAT	S000028	159	(+)
CACTFTPPCA1	YACT	S000449	558	(+)
CACTFTPPCA1	YACT	S000449	384	(+)
CACTFTPPCA1	YACT	S000449	189	(+)
CACTFTPPCA1	YACT	S000449	123	(+)
CACTFTPPCA1	YACT	S000449	542	(+)
CACTFTPPCA1	YACT	S000449	502	(+)
CACTFTPPCA1	YACT	S000449	341	(+)
CACTFTPPCA1	YACT	S000449	199	(+)
CACTFTPPCA1	YACT	S000449	69	(+)
CACTFTPPCA1	YACT	S000449	58	(+)
CARGATCONSENSUS	CCWWWWWWGG	S000404	160	(+)
CARGCW8GAT	CWWWWWWWWG	S000431	376	(+)
CARGCW8GAT	CWWWWWWWWG	S000431	56	(+)
CCAATBOX1	CCAAT	S000030	377	(+)
CCAATBOX1	CCAAT	S000030	160	(+)

CURECORECR	GTAC	S000493	628	(+)
CURECORECR	GTAC	S000493	592	(+)
CURECORECR	GTAC	S000493	543	(+)
CURECORECR	GTAC	S000493	503	(+)
CURECORECR	GTAC	S000493	342	(+)
DOFCOREZM	AAAG	S000265	414	(+)
DOFCOREZM	AAAG	S000265	347	(+)
DOFCOREZM	AAAG	S000265	322	(+)
DOFCOREZM	AAAG	S000265	307	(+)
DOFCOREZM	AAAG	S000265	221	(+)
DOFCOREZM	AAAG	S000265	95	(+)
DPBFCOREDCDC3	ACACNNG	S000292	559	(+)
EBOXBNNAPA	CANNTG	S000144	558	(+)
EECCRCAH1	GANTTNC	S000494	337	(+)
EECCRCAH1	GANTTNC	S000494	216	(+)
ERELEE4	AWTTCAAA	S000037	408	(+)
GATABOX	GATA	S000039	646	(+)
GATABOX	GATA	S000039	531	(+)
GATABOX	GATA	S000039	526	(+)
GATABOX	GATA	S000039	491	(+)
GATABOX	GATA	S000039	487	(+)
GATABOX	GATA	S000039	99	(+)
GATABOX	GATA	S000039	37	(+)
GT1CONSENSUS	GRWAAW	S000198	325	(+)
GT1CONSENSUS	GRWAAW	S000198	242	(+)
GT1CONSENSUS	GRWAAW	S000198	99	(+)
GT1CONSENSUS	GRWAAW	S000198	7	(+)
GT1GMSCAM4	GAAAAA	S000453	325	(+)
GTGANTG10	GTGA	S000378	426	(+)
GTGANTG10	GTGA	S000378	350	(+)
IBOXCORE	GATAA	S000199	99	(+)
INRNTPSADB	YTCANTYY	S000395	175	(+)
MARABOX1	AATAAAYAAA	S000063	468	(+)
MYBCORE	CNGTTR	S000176	511	(+)
MYBCOREATCYCB1	AACGG	S000502	679	(+)
MYBST1	GGATA	S000180	647	(+)
MYBST1	GGATA	S000180	100	(+)
MYBST1	GGATA	S000180	38	(+)
MYCCONSENSUSAT	CANNTG	S000407	558	(+)
NODCON2GM	CTCTT	S000462	178	(+)
OSE2ROOTNODULE	CTCTT	S000468	178	(+)
P1BS	GNATATNC	S000459	647	(+)

POLASIG1	AATAAA	S000080	468	(+)
POLASIG1	AATAAA	S000080	464	(+)
POLASIG1	AATAAA	S000080	318	(+)
POLASIG1	AATAAA	S000080	312	(+)
POLASIG3	AATAAT	S000088	132	(+)
POLLEN1LELAT52	AGAAA	S000245	412	(+)
PYRIMIDINEBOXHVEPB1	TTTTTTCC	S000298	278	(+)
RHERPATEXPA7	KCACGW	S000512	581	(+)
ROOTMOTIFTAPOX1	ATATT	S000098	525	(+)
ROOTMOTIFTAPOX1	ATATT	S000098	486	(+)
ROOTMOTIFTAPOX1	ATATT	S000098	440	(+)
ROOTMOTIFTAPOX1	ATATT	S000098	281	(+)
ROOTMOTIFTAPOX1	ATATT	S000098	233	(+)
ROOTMOTIFTAPOX1	ATATT	S000098	144	(+)
ROOTMOTIFTAPOX1	ATATT	S000098	137	(+)
ROOTMOTIFTAPOX1	ATATT	S000098	116	(+)
SP8BFIBSP8BIB	TACTATT	S000184	69	(+)
SREATMSD	TTATCC	S000470	26	(+)
SV40COREENHAN	GTGGWWHG	S000123	639	(+)
TATABOX2	TATAAAT	S000109	55	(+)
TATABOX3	TATTAAT	S000110	445	(+)
TATABOX3	TATTAAT	S000110	136	(+)
TATABOX3	TATTAAT	S000110	115	(+)
TATABOX5	TTATTT	S000203	207	(+)
TATCCACHVAL21	TATCCAC	S000416	127	(+)
TATCCAOSAMY	TATCCA	S000403	127	(+)
TATCCAYMOTIFOSRAMY3D	TATCCAY	S000256	127	(+)
TBOXATGAPB	ACTTTG	S000383	198	(+)
WBOXNTCHN48	CTGACY	S000508	382	(+)
WBOXNTERF3	TGACY	S000457	381	(+)
WRKY71OS	TGAC	S000447	381	(+)
-10PEHVPSBD	TATTCT	S000392	319	(-)
ABRELATERD1	ACGTG	S000414	579	(-)
ABRERATCAL	MACGYGB	S000507	580	(-)

ACGTABOX	TACGTA	S000130	79	(-)
ACGTATERD1	ACGT	S000415	640	(-)
ACGTATERD1	ACGT	S000415	633	(-)
ACGTATERD1	ACGT	S000415	578	(-)
ACGTATERD1	ACGT	S000415	78	(-)
ACGTTBOX	AACGTT	S000132	634	(-)
ANAERO1CONSENSUS	AAACAAA	S000477	63	(-)
ARR1AT	NGATT	S000454	662	(-)
ARR1AT	NGATT	S000454	598	(-)
ARR1AT	NGATT	S000454	256	(-)
BOXIINTPATPB	ATAGAA	S000296	269	(-)
BOXIINTPATPB	ATAGAA	S000296	229	(-)
CAATBOX1	CAAT	S000028	545	(-)
CAATBOX1	CAAT	S000028	515	(-)
CAATBOX1	CAAT	S000028	505	(-)
CACTFTPPCA1	YACT	S000449	669	(-)
CACTFTPPCA1	YACT	S000449	617	(-)
CACTFTPPCA1	YACT	S000449	426	(-)
CACTFTPPCA1	YACT	S000449	344	(-)
CACTFTPPCA1	YACT	S000449	107	(-)
CARGATCONSENSUS	CCWWWWWWGG	S000404	159	(-)
CARGCW8GAT	CWWWWWWWWG	S000431	375	(-)
CARGCW8GAT	CWWWWWWWWG	S000431	55	(-)
CIACADIANLELHC	CAANNNNATC	S000252	336	(-)
CPBCSPOR	TATTAG	S000491	623	(-)
CURECORECR	GTAC	S000493	627	(-)
CURECORECR	GTAC	S000493	591	(-)
CURECORECR	GTAC	S000493	542	(-)
CURECORECR	GTAC	S000493	502	(-)
CURECORECR	GTAC	S000493	341	(-)
DOFCOREZM	AAAG	S000265	285	(-)
DOFCOREZM	AAAG	S000265	209	(-)
DOFCOREZM	AAAG	S000265	196	(-)
DOFCOREZM	AAAG	S000265	120	(-)
DOFCOREZM	AAAG	S000265	72	(-)
EBOXBNNAPA	CANNTG	S000144	557	(-)
EVENINGAT	AAAATATCT	S000385	487	(-)
GATABOX	GATA	S000039	667	(-)
GATABOX	GATA	S000039	126	(-)
GATABOX	GATA	S000039	75	(-)

GATABOX	GATA	S000039	24	(-)
GT1CONSENSUS	GRWAAW	S000198	483	(-)
GT1CONSENSUS	GRWAAW	S000198	214	(-)
GT1CONSENSUS	GRWAAW	S000198	276	(-)
GT1CONSENSUS	GRWAAW	S000198	275	(-)
GT1GMSCAM4	GAAAAA	S000453	276	(-)
GTGANTG10	GTGA	S000378	596	(-)
GTGANTG10	GTGA	S000378	471	(-)
GTGANTG10	GTGA	S000378	189	(-)
HDZIP2ATATHB2	TAATMATTA	S000373	448	(-)
IBOX	GATAAG	S000124	26	(-)
IBOXCORE	GATAA	S000199	25	(-)
IBOXCORENT	GATAAGR	S000424	27	(-)
LECPLEACS2	TAAAATAT	S000465	439	(-)
LTRE1HVBLT49	CCGAAA	S000250	170	(-)
MARTBOX	TTWTWTTWTT	S000067	314	(-)
MYB2CONSENSUSAT	YAACKG	S000409	510	(-)
MYBCOREATCYCB1	AACGG	S000502	510	(-)
MYBST1	GGATA	S000180	126	(-)
MYBST1	GGATA	S000180	24	(-)
MYCCONSENSUSAT	CANNTG	S000407	557	(-)
NODCON1GM	AAAGAT	S000461	74	(-)
NODCON2GM	CTCTT	S000462	219	(-)
OSE1ROOTNODULE	AAAGAT	S000467	74	(-)
OSE2ROOTNODULE	CTCTT	S000468	219	(-)
P1BS	GNATATNC	S000459	646	(-)
POLASIG1	AATAAA	S000080	399	(-)
POLASIG1	AATAAA	S000080	207	(-)
POLASIG3	AATAAT	S000088	548	(-)
POLASIG3	AATAAT	S000088	458	(-)
POLASIG3	AATAAT	S000088	446	(-)
POLASIG3	AATAAT	S000088	421	(-)
POLLEN1LELAT52	AGAAA	S000245	481	(-)
POLLEN1LELAT52	AGAAA	S000245	212	(-)
POLLEN1LELAT52	AGAAA	S000245	203	(-)
ROOTMOTIFTAPOX1	ATATT	S000098	440	(-)

ROOTMOTIFTAPOX1	ATATT	S000098	432	(-)
ROOTMOTIFTAPOX1	ATATT	S000098	374	(-)
ROOTMOTIFTAPOX1	ATATT	S000098	281	(-)
ROOTMOTIFTAPOX1	ATATT	S000098	238	(-)
ROOTMOTIFTAPOX1	ATATT	S000098	233	(-)
ROOTMOTIFTAPOX1	ATATT	S000098	144	(-)
ROOTMOTIFTAPOX1	ATATT	S000098	128	(-)
SEF4MOTIFGM7S	RTTTTTR	S000103	315	(-)
SP8BFIBSP8BIB	TACTATT	S000184	110	(-)
SREATMSD	TTATCC	S000470	99	(-)
T/GBOXATPIN2	AACGTG	S000458	579	(-)
TAAAGSTKST1	TAAAG	S000387	285	(-)
TATABOX3	TATTAAT	S000110	443	(-)
TATABOX3	TATTAAT	S000110	134	(-)
TATABOX3	TATTAAT	S000110	113	(-)
TATABOX4	TATATAA	S000111	140	(-)
TATABOX5	TTATTT	S000203	464	(-)
TATABOX5	TTATTT	S000203	312	(-)
TATCCACHVAL21	TATCCAC	S000416	101	(-)
TATCCAOSAMY	TATCCA	S000403	100	(-)
TATCCAYMOTIFOSRAMY3D	TATCCAY	S000256	101	(-)
WBOXATNPR1	TTGAC	S000390	261	(-)
WRKY71OS	TGAC	S000447	565	(-)
WRKY71OS	TGAC	S000447	261	(-)

Table 5: Accession numbers and primer sequences of the trypsin and chymotrypsin genes analyzed for transcript abundance (**F**: Forward primer; **R**: Reverse primer)

Name	**Accession**	**Primer sequence (5'-3')**
HaTry 1	EU982841	**F** GAGGACACAGATGTGGAGGGG
		R GAACACACGGAATTCAGCCACG
HaTry 2	EU770391	**F** GCGTAAAGGATGCGGTTGG
		R CAGGATGGCAACCATCCATG
HaTry 3	EU325548	**F** CGACCACACTGACGCGAG
		R GCACGCCACTGGACATGG
HaTry 4	EF600059	**F** GTGCTACCCCTTCTGATTC
		R AACTTGTCGATGGAGGTGAC
HaTry 5	EF600054	**F** GGTCTCTGCTAACCTCCACC
		R CTGGATGCCAGGGACGTGC
HaTry 6	Y12276	**F** CCATCGCCGGTGCCAACTA
		R CTGAACGTGACGCAACTGCTC
HaTry 7	Y12271	**F** CAGAGGATTGTGGGTGGTTCG
		R GCGGTGAGGATAGCCCTGTT
HaTry 8	Y12286	**F** GGGCTACTGGTGCCTTCAACG
		R CAGAGTCATACACGTCACCGACG
HaChy 1	HM209422	**F** CGACTTGTCAGGTGGTCAGGCTG
		R GCGATTCTGGTACCGCCGGAGAAC
HaChy 2	EU325550	**F** GACTTGTCAGGTGGCCAGGCTG
		R GCGATTCTGGTACCGCCGGAGAAC
HaChy 3	GU323796	**F** TGACTTGTCAGGTGGCCAAGCTG
		R GCGATTCTGGTACCGCCGGAGAAC
HaChy 4	Y12273	**F** CACCATCTTCATCTTCCAATCCGTGTGC
		R GTGTTGATACGAGTACCACCGAAGAAC

Table 6: Primer sequences used for analysing transcript abundance (**F**: Forward primer; **R**: Reverse primer)

Transcript	**Description**	**Primer sequence (5'-3')**
Locus_1900_Transcript_1/2_Confidence_0.750_Length_2274	Serine protease inhibitor 13	**F** TGAAGAAGCTGAGTACTTTGGCT **R** GGGTTCTCAGCATTCTCTAGGTC
Locus_4594_Transcript_1/2_Confidence_0.800_Length_2266	Serine proteinase	**F** TTGTGCAGGTGATAAATTGGATGG **R** CAATAGTACCCAGGCACAGTAGAA
AY373973	Cathepsin L	**F** CAACGCCTTCAAGTACATCAAGG **R** GCGTTGTACCTGCACTTATCATC
AF222788.2	Cathepsin B	**F** GGCCACGCTGTTAAGATCCT **R** GTCACCCCAGTCACTGTTCC
DQ875228	Immune reactive protease inhibitor	**F** AATGACGATGGAGAATGCATACCT **R** AACAATCACGACTTATGCTATCGC
har-miR-282-5p		**F** TAGCCTCTCCTTGGCTTTGT **R** CGTAACCTCTATCAGGCTATGTC
har-miR-305-5p		**F** GCACGCCCATTGTACTTCAT **R** CCGACACGGTAAGTGTACTCC

Table 7: Conserved miRNAs identified in *H. armigera*

Sr. No.	miRNA	Mature sequence
1	har-miR-993b*	uacccuguagauccgggcuuu
2	har-miR-92a	uauugcaccagucccggccuau
3	har-miR-927	uuuagaauuccuacgcuuuacc
4	har-miR-3338	auguacuuacuuuguuuguucu
5	har-miR-281*	aagagagcuauccgucgacagu
6	har-miR-279b	ugacuagaucuacacucauug
7	har-miR-279a	ugacuagauccacacucaucca
8	har-miR-278*	ccggacgaacuucccagcucggcc
9	har-miR-277	uaaaugcacuaucugguacgaca
10	har-miR-2765	ugguaacuccaccaccguuggc
11	har-miR-2755	cacccugucagaccauacuuguu
12	har-miR-263a	aauggcacuggaagaauucacgg
13	har-miR-190	agauauguuugauauucuugguug
14	har-miR-184	uggacggagaacugauaagggc
15	har-miR-14	ucagucuuuuucucucuccuau
16	har-miR-12	ugaguauuacuucagguacu
17	har-miR-11	caucacagucagaguucuagcu
18	har-miR-10	uacccuguagauccgaauuugu
19	har-miR-let-7	ugagguaguagguuguauag
20	har-miR-193-3p	uacuggccugcuaagucccaag
21	har-miR-8	uaauacugucagguaaagauguc
22	har-bantam	ugagaucauugugaaagcuaauu
23	har-miR-31	aggcaagaagucggcauagcugu
24	har-miR-263a-5p	aauggcacuggaagaauucacggg
25	har-miR-750	ccagaucuaucuuuccagcuca
26	har-miR-274-5p	uuugugaccgucacuaacgggca
27	har-miR-9c*	ucuuugguauccuagcuguagg
28	har-miR-9a-5p	ucuuugguuaucuagcuguauga
29	har-miR-279d	ugacuagauccauacucgucugc
30	har-miR-282-5p	uagccucuccuuggcuuugucug
31	har-miR-279b-3p	ugacuagaucuacacucauuga
32	har-miR-970-3p	ucauaagacacacgcggcucu
33	har-miR-71-3p	ucucacuaccuugucuuucaug
34	har-miR-305-5p	auuguacuucaucaggugcucugg

35	har-miR-317	ugaacacagcuggugguaucucagu
36	har-miR-79-3p	auaaagcuagauuaccaaagca
37	har-miR-133-3p	uuggucccuucaaccagcugu
38	har-miR-745-5p	cggcucaucguguggcaguuugc
39	har-miR-316-5p	ugucuuuuuccgcuuugcugcug
40	har-miR-1-3p	uggaauguaaagaaguauggag
41	har-miR-993-3p	gaagcucgucucuacagguaucu
42	har-miR-2796-3p	guaggccggcggaaacuacuugc
43	har-miR-2768-3p	auugguuaagauauugcaucgu
44	har-miR-929-5p	aaauugacucuaguagggagu
45	har-miR-33a-5p	gugcauuguaguugcauugca
46	har-miR-2788-3p	caaugcccuuggaaaucccaaa
47	har-miR-210-3p	cuugugcgugugacagcggcuau
48	har-miR-998	uagcaccaugggauucagcuca
49	har-miR-306a-5p	ucagguacuaggugacucuga
50	har-miR-1000-5p	auauuguccugucacagcagua
51	har-miR-13b	uaucacagccauuuuugacgaguu
52	har-miR-3p	ucagguaccugaaguagcgcgcg
53	har-miR-let-7-5p	ugagguaguagguuguauagu
54	har-miR-2b-3p	uaucacagccagcuuuguugagu
55	har-miR-932-5p	ucaauuccguagugcauugcagu
56	har-miR-375-3p	uuuguucgccccggcucgugucg
57	har-miR-965-3p	uaagcguauagcuuuuccccuu
58	har-miR-2766	ucagucuugucgaaugggugggu
59	har-miR-993	cuacccuguagauccgggcu
60	har-miR-34-5p	uggcaguguggguuagcuggguugu
61	har-miR-307	ucacaaccuccuugagugagc
62	har-miR-92b	aauugcaccaaucccggccugc
63	har-miR-252-5p	cuaaguacuagugccgcaggag
64	har-miR-2767	caaguaaaucucgugcgguuug
65	har-miR-87	gugagcaaacuuucaggugugu
66	har-miR-308-3p	aaucacaggauaauacugcga
67	har-miR-iab-4-5p	acguauacugaauguauccuga
68	har-miR-279-d-5p	ugacuagauuuucacuuauccu
69	har-miR-263-b-5p	cuuggcacugggagaauucacag
70	har-miR-285	uagcaccauucgaauucagugc
71	har-miR-124-3p	uaaggcacgcggugaaugcca
72	har-miR-283	aaauaucagcuggguaauucuggg

73	har-miR-137	uuauugcuugagaauacacgu
74	har-miR-1175-5p	ugagauucaacuccuccaacuu
75	har-miR-989	gugugaugugacguaguggaag
76	har-miR-2a-3p	ucacagccagcuuugaugagc
77	har-miR-2a	ucacagccagcuuugaugagca
78	har-miR-3327-3p	auauguaacguuuuuguuguc
79	har-miR-7	uggaagacuagugauuuuguuguu
80	har-miR-2795	caaguuuggugauacgcgggcgc
81	har-miR-bantam-3p	ugagaucauugugaaagcuaau
82	har-miR-929-3p	cucccuaaucgagucagguuga
83	har-miR-iab-8-5p	uuacguauacugaagguauaccgga
84	har-miR-9b	gcuuugguaaucuagcuuuauga
85	har-miR-1175-3p	ugagauucaacuccuccaacuua
86	har-miR-7-5p	uggaagacuagugauuuuguugu
87	har-miR-989a	ugugaugugacguaguggaag
88	har-miR-33-3p	caauaugacuacaaggcaaauc
89	har-miR-4961-5p	uauguauauccauaucagagagc
90	har-miR-137-3p	uauugcuugagaauacacguag
91	har-miR-998-3p	uagcaccaugggauucagcuc
92	har-miR-263b	cuuggcacugggagaauucac
93	har-miR-137b	uuauugcuugagaauacacguag
94	har-miR-2756	acccuguagcugccaaggggcg
96	har-miR-9b-3p	acggagcuaaaucgccaaagcg

Table 8: Novel miRNA identified in *H. armigera*

Sr. No.	miRNA	Mature sequence
1	har-miR-novel-1	auuuuuucuuuacauuuagc
2	har-miR-novel-2	gcuuucgcaauuugccgcaccc
3	har-miR-novel-3	cacucuaaguaugaacaccaagc
4	har-miR-novel-4	gguggugggugggugggugggug
5	har-miR-novel-5	uuugucguuguuuacuauca
6	har-miR-novel-6	cgaugguaguuucugcga
7	har-miR-novel-7	uuuuucuuuacauuuagc
8	har-miR-novel-8	aucugcuuaacggccugcc
9	har-miR-novel-9	augugcgcggcggagagcgc
10	har-miR-novel-10	cgcggauggugaugguggc
11	har-miR-novel-11	aaugcaucgcgacccguucg
12	har-miR-novel-12	ccagcagauguggaucuc
13	har-miR-novel-13	uacugacgugcaaaucgc
14	har-miR-novel-14	acggagcaggauacggagc
15	har-miR-novel-15	gucagagggccggcugua
16	har-miR-novel-16	uaucacagccacuuugauguggu
17	har-miR-novel-17	aggauacauucaguauacguaca
18	har-miR-novel-18	caagaaaucacuaaucugccu
19	har-miR-novel-19	agcgaggcccuggaggacu
20	har-miR-novel-20	ggucaguacuuggaugggu
21	har-miR-novel-21	uucgugcugaggguacuc
22	har-miR-novel-22	caaugaauuggugcguac
23	har-miR-novel-23	ucugcuuaacggccugcc
24	har-miR-novel-24	aaugaagaacgcauuuaacug
25	har-miR-novel-25	uaccagcaguaaugcacc
26	har-miR-novel-26	ugugcgcggcggagagcgc
27	har-miR-novel-27	ugcugcugugcgccggcgc
28	har-miR-novel-28	accuggcguaggagauacc
29	har-miR-novel-29	uuuguaucgauacaaacgacaa
30	har-miR-novel-30	caagaaaucacuaaucugccua
31	har-miR-novel-31	gcccauagagucgcgaau
32	har-miR-novel-32	cgcggucgcugccggccg
33	har-miR-novel-33	ucggggccgcacgcgcgcu
34	har-miR-novel-34	uccacguucaccgcacaa
35	har-miR-novel-35	ggugguggcggcgguggcgg
36	har-miR-novel-36	aaucuugacacagugucg
37	har-miR-novel-37	cauacugacgugcaaauc
38	har-miR-novel-38	ggcggcgaccacguacaugcccgcc
39	har-miR-novel-39	ucguagccaauguuccacaggagag

40	har-miR-novel-40	ggucgugggguucgauuccc
41	har-miR-novel-41	gcgccgagccgucgcugac
42	har-miR-novel-42	aaguggaggugugaucucuucacu
43	har-miR-novel-43	uuucagaagauccagaucaggcu
44	har-miR-novel-44	uccgaagcucgucuggaccu
45	har-miR-novel-45	aggacgagguggacgagaa
46	har-miR-novel-46	ugauuuuucucuuuuauu
47	har-miR-novel-47	cgggcggggcggcgcgcgg
48	har-miR-novel-48	uauugcauuuuuggauug
49	har-miR-novel-49	gugggggcaaauugcgaaagcug
50	har-miR-novel-50	cgaugguaguuucugcgaa
51	har-miR-novel-51	uaaugcugcccgguaagaugcg
52	har-miR-novel-52	gccuugaaacacgcacgguuuug
53	har-miR-novel-53	uucgugcugaggguacucg
54	har-miR-novel-54	ugcucacugcugguucgau
55	har-miR-novel-55	gccuugaaacacgcacgggguuug
56	har-miR-novel-56	augugcgcggcggagagc
57	har-miR-novel-57	ugguugaccccugcgauu
58	har-miR-novel-58	cauguacuuacuuuguuuguucu
59	har-miR-novel-59	gucggcgaaaaagaucgcc
60	har-miR-novel-60	guucgaaucucucaggcgg
61	har-miR-novel-61	aucggcgagaagacugcaucu
62	har-miR-novel-62	gugcugcucgacgaggcg
63	har-miR-novel-63	aacaacaggaauguuauguaca
64	har-miR-novel-64	guggagaaggguuucgcgu
65	har-miR-novel-65	uuuaguacaguguuucccaaccu
66	har-miR-novel-66	aguuauaguuuguuugau
67	har-miR-novel-67	gaggacgagguggacgagaa
68	har-miR-novel-68	gcucacugcugguucgau
69	har-miR-novel-69	guucgaaucucucaggcg
70	har-miR-novel-70	aaauuucaucgggauggg
71	har-miR-novel-71	cgcuggcgcggcgguacgcg
72	har-miR-novel-72	cuguagucgaugaggucug
73	har-miR-novel-73	cgcugaggaacugaagaaggauc
74	har-miR-novel-74	cagguggaccagguggacc
75	har-miR-novel-75	uugaugaauuuucggauu
76	har-miR-novel-76	guacuccaacggcuguccu
77	har-miR-novel-77	guuucguggccuuuagauc
78	har-miR-novel-78	auaaggauuaaauuuaagc
79	har-miR-novel-79	aacaaugaauuggugcgu
80	har-miR-novel-80	cgagcagacgcccgccauc

81	har-miR-novel-81	gccaacguccauaccacg
82	har-miR-novel-82	ccggagagggugagcgcc
83	har-miR-novel-83	auuuuggagauuauugaua
84	har-miR-novel-84	gucaaugaagaacgcauuuaacug
85	har-miR-novel-85	cgcgcuacuucagguaccugacu
86	har-miR-novel-86	aacaacaggaauguuauguacag
87	har-miR-novel-87	ucuauuuuguugguauuc
88	har-miR-novel-88	auacugacgugcaaaucgc
89	har-miR-novel-89	agcuaugccgacuucuugccuuc
90	har-miR-novel-90	uacagcuaggauaccaaagaga

Appendix II

Accession Numbers and nucleotide sequences of *CanPI* genes submitted to NCBI

>CanPI-68 (KM576781)
ATCCATAACAAAAACAAGGTAAATAATCACCACCGATTCATAATATGGCTGT
TCACAAAGAAGTTAGTTTCCTTGCTTTCCTACTTGTTCTTGAAATATTGCTTCT
ACATGTTGATGCCAAGGCTTGTTCAGAAGAAAACGCAGAAAATCGCATATGC
ACCAATTGCTGTGCAGGCCGTAAGGGTTGCAACTATTACAGTGCTGATGGGA
CTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAAGCTTGCCCTCG
GTATTGTGATACAAGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAAC
GCAGAAAATCGCCTATGCACCAATTGCTGTGCAGGCCGTAAGGGTTGCAACT
ATTACAGCGCTGACGGGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAA
CCCAAAAGCTTGCCCTCGGAATTGTGATACAAGAATTGCCTATTCAAAATGTC
CACGTTCCGAAGGAAACGCAGAAAATCGCCTATGCACCAATTGCTGTGCAGG
CCGTAAGGGTTGCAACTATTACAGCGCTGACGGGACTTTCATTTGTGAAGGA
GAGTCTGACCCCAACAACCCAAAAGCTTGCCCTAGGAATTGTGATCCAAATA
TTGCCTATTCACTGTGTCTCTATGAAAAATAA

>CanPI-69 (KM576791)
AAAACGGAGGAAGTATCTAATCCCAACTAAATGAGGATATACGTGGAACGTT
CGTACCTAATAAGTAGAGGATGTAACAATTAATCACCGTACATCGTTGGCAC
GTTTTAGTTGTGTCAGGAACACTTGTCCATTATTGTACTTCAGGTTGATATGA
TATTCTGTATTGTCCGTTATTGTACTTAAACTAGATAGATATTTTCTTCGTTTC
ACAATAAATAAATTATTAAATTAATTATTAATATTTTAATATAAGTGAATTAT
TAAAAAGAAATTTCAAATTTATTCTTATGGATCCACTGACCCAATATATAGAT
TTGATTTTTTGGGGGTGAAAGTGTACTGATTTGCTTGATGAAAAAGAATAAA
AATAAAAAGTTATGTTAAATTTATGTCTTTAATATTTTTTCCTTCTATATGTCA
AAATCTAACAATTCGAGGAAATATAATATTCTATACGAAAGAGATTTTCTTTT
ATTTCTTACTTTGGATCACTCTCGATCCTCTTCATTTCGGGTGGTCCAATTTAG
GCAGCAAATATTATATATTAATAATATCCACTTTTATATTAATAGTAATGTGG
ATAAAAGTTTTAGTTAGCTACGTATCTTTTACTATTTGTTTACTATAAATAGG
AGGAGAAGGATAGACCTCCTTATCCTTAACGAAAACAAGGTAAACATGGCTG
TTCACAAAGAAGTTAGTTTTCTTGGTTATCTACTTGTTCTTGGTAAGACTTTTC
TTTCTCCTTTAATTTTTCTTTTTAATTTTTTCATTGGTATATAGACACGGTTAGA
GAAGAGTTTTTTCTACCGCGAAGCAATGGAAAATATGTCTTAGGAGTATAAA
ACATGATTTTCCCATCTAGTTCCTCTGTTTTATTTTACTTGTTTCAAATTTTTTA
TTAATATTATTTTTAACAAATCAATATGAAATAAAAAAAATTTCATGTTTTTT
TTTTATATTAATTATTTTTTTAAATTAAAATATAAATATTATTTAATAAAAATA
TATATGTTATTAATTATTTTTTTAATTAATTTATCAACTTAAATTAGTACTAGT
AAAATGGGAGTACAAGTAAAATGGGAGGGGGGAGTATTTTCCACCTTACTGA

ATTACTGAAAATTAAACGTATGATGGGAAACAAAAGACACAACTTTTTCAGT
GAACTACTAGGAAGCTTCCTATTAAATATTAAGCACACCTTTCAGTGGGAAA
TTTTACCAGGAAAATATTGCTTTGTGGTAGTGATACTTACCTGGTGTTATGTTT
TTCCTCAATATTTATATATAGTTTCTATATTTGCTGTAGGAATATTTCTACATC
ATGTTGATGGCAAGGCTTGTACCAGAGAATGTGGCAATCTTGGGTATGGGAT
ATGTCCACGTTCAGAAGGAAGTGCAGAAGAACCCATATGCACCAATTGTTGT
GCAGGCTATAAGGGTTGTAAGTATTACAGTGCTGACGGGACTTTCATTTGCG
AAGGAGAGTCTGACCCCAAGAACCCAAAAGGTTGCCCTCGGAATTGTGATCC
AAGAATTGCCTATTCAAAATGTCCACGTTCACAAGGAGAGACGTTAATTTAT
CCCACAGGATGCACCACCTGTTGCACGGGCTACAAAGGTTGCTACTATTTCG
GTAAAGACGGCAAGTTTCTCTGTGAAGGAGAGAGTACTGAACCCAAGGCTTG
TACTCTGGAGTGTGATCCAAGGGTTGCTTACATGACTTGTCCCTCTACTGGAT
TGGCCAAGATTAAGGATGTTTGTGTTAACTGTTGCACTGCGGGAGAGGGTTG
CAAACTGTACGGTCATAATGGATCTCTACTTTGTACTGGAGAGTCTCAGAGCA
TATCCACAGCATAGAGCAACTGTTCATTTCTAATCAATGATGATGTATGTTGT
ACTAGTTTATTAATGTATGAAATAAAAGCATGAACATTGTCATGCTAATCTTT
GTAATATGGGCGTAAGAGCTCTTTTGGGCAAAAAAAAAAAAAAAAAAAAAA
AAAAA

>CanPI-70 (KM576790)
ATGGCTGTTCCCAAAGAAGTTAGTTTTCTTGGTTATCTACTTGTTCTTGGAATA
TTTCTACATCATGTTGATGGCAAGGCTTGTACCAGAGAATGTGGCAATCTTGG
GTATGGGATATGTCCACGTTCAGAAGGAAGTGCAGAAGAACCCATATGCACC
AATTGTTGTGCAGGCTATAAGGGTTGTAAGTATTACAGTGCTGACGGGACTTT
CATTTGCGAAGGAGAGTCTGACCCCAAGAACCCAAAAGGTTGCCCTCGGAAT
TGTGATCCAAGAATTGCCTATTCAAGATGTCCACGTTCACAAGGAGAGACGT
TAATTTATCCCACAGGATGCACCACCTGTTGCACGGGCTACAAAGGTTGCTAC
TATTTCGGTAAAGACGGCAAGTTTCTCTGTGAAGGAGAGAGTACTGAACCCA
AGGCTTGTACTCTGGAGTGTGATCCAAGGGTTGCTTACATGACTTGTCCCTCT
ACTGGATTGGCCAAGATTAAGGATGTTTGTGTTAACTGTTGCACTGCGGGAG
AGGGTTGCAAACTGTACGGTCATAATGGATCTCTACTTTGTACTGGAGAGTCT
CAGAGCATATCCACAGCATAGAGCAACTGTTCATTTCTAATCAATGATGATGT
ATGTTGTACTAGTTTATTAATGTATGAAATAAAAGCATGAACATTGTCATGCT
AATCTTTGTAATATGGGCGTAAGAGCTCTTTTGGACATAAATAGGTGTCCAGT
CTCCAAGTGTTTAATGAAGCACTAGTTTCTTTAGAAAAAAAAAAAAAAAAAA
AAAA

>CanPI-71(KM576782)
TTTCTACATCATGTTGATGGCAAGGCTTGTACCAGAGAATGTGGCAATCTTGG
GTATGGGATATGTCCACGTTCAGAAGGAAGTGCAGAAGAACCCATATGCACC
AATTGTTGTGCAGGCTATAAGGGTTGTAAGTATTACAGTGCTGACGGGACTTT
CATTTGCGAAGGAGAGTCTGACCCCAAGAACCCAAAAGGTTGCCCTCGGAAT
TGTGATCCAAGAATTGCCTATTCAAAATGTCCACGTTCACAAGGAGAGACGT
TAATTTATCCCACAGGATGCACCACCTGTTGCACGGGCTACAAAGGTTGCTAC
TATTTCGGTAAAGACGGCAAGTTTCTCTGTGAAGGAGAGAGTACTGAACCCA
AGGCTTGTACTCTGGAGTGTGATCCAAGGGTTGCTTACATGACTTGTCCCTCT
ACTGGATTGGCCAAGATTAAGGATGTTTGTGTTAACTGTTGCACTGCGGGAG
AGGGTTGCAAACTGTACGGTCATAATGGATCTCTACTTTGTACTGGAGAGTCT
CAGAGCATATCCACAGCATAGAGCAACTGTTCATTTCTAATCAATGATGATGT
ATGTTGTACTAGTTTATTAATGTATGAAATAAAAGCATGAACATTGTCATGCT
AATCTTTGTAATATGGGCGTAAGAGCTCTTTTGGACATAAATAGGTGTCCAGT
CTCCAAGTGTTTAATGAAGCACTAGTTTCTTTTACTAAAAAAAAAAAAAAAA
AAAA

>CanPI-72 (KM576783)
CCCAAAGAAGTTAGTTTCCTTGCCTTCCTACTTGTTCTTGAAATATTGCTTCTA
CATGTTGATGCCAAGGCTTGTTCAGAAGAAAACGCAGAAAATCACATATGCA
CCAATTGCTGTGCAGGCCGTAAGGGTTGCAACTATTACAGTGCTGATGGGAC
TTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAAGCTTGCCCTCGGT
ATTGTGATACAAGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAACGC
AGAAAATCGCCTATGCACCAATTGCTGTGCAGGCCGTAAGGGTTGCAACTAT
TACAGCGCTGACGGGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACC
CAAAAGCTTGCCCTCGGAATTGTGATACAAGAATTGCCTATTCAAAATGTCC
ACGTTCCGAAGGAAACGCAGAAAATCGCCTATGCACCAATTGCTGTGCAGGC
CGTAAGGGTTGCAACTATTACAGCGCTGACGGGACTTTCATTTGTGAAGGAG
AGTCTGACCCCAACAACCCAAAAGCTTGCCCTAGGAATCGTGATCCAAATAT
TGCCTATTCACTGTGTCTCTATGAAAAATAAAAGCATGAACAGTGTCATGCTA
ATCTCCGTAATACGCATGGGCGTCTATATGAAAAATAAAAGCATGAACGTAA
CGGTGTCATGCTAATCTATGTAATATGGGCGTTGGAGCCCTTTTGGACGCATA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

>CanPI-73 (KM576784)
CTACTTGTACTTGGAATATTGCTTCTACATGTTGATGCCAAGGCTTGTTCACA
AAGAAACGCAAAAGAACCCATATGCACCAATTGTTGTGCAGGCCGTAAGGGT
TGCAACTATTACAGTGCTGATGGGACTTTCATTTGTGAAGGAGAGTCTGACCC
CAACAACCCAAAACCTTGTACTTTGAACTGTGATCCTAATATTGCCTATTCAC
TATGTCTGTATGAAAAGTAAAAGCATGAACAGTGTCATGCTAATCTCTGTAGT

ATGCTTGAGCGTCTATTTATGAAATAAAAGCGTGAACAGTGTCACGCTAATCT
ATGTAATATGGGCGTTGGAGCCCTTTTGGACTCGAAAAAAAAAAAAAAAAAA
AAAAA

>CanPI-74 (KM576785)
GCTGTTCCCAAAGAAGTTAGTTTTCCTGCTTTCCTACTTGTTCTTGAAATATTG
CTTCTACATGTTGATGCCAAGGCTTGTTCAGAAGAAAACGCAGAAAATCGCA
TATGCACCAATTGCTGTGCAGGTCGTAAGGGTTGCAACTATTACAGTGCTGAT
GGGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAAGCTTGCC
CTAGGAATTGTGATCCAAATATTGCCTATTCACTGTGTCTCTATGAAAAATAA
AAGCATGAACAGTGTCATGCTAATCTCTGCAATATGCTTGAGCGTCTATTTAT
GAAATAAAAGCGTGAACAGTGTCACGCTAATCTATGTAATATGGGCGTTGGA
GCCCTTTTGGACTTAAATCATGTGTTTAAGTGTGGCAAAAAAAAAAAAAAAA
AAAAAAAAAAAAA

>CanPI-75 (KM576786)
GAAATATTGCTTCTACATGTTGATGCCAAGGCTTGTTCAGAAGAAAACGCAG
AAAATCGCATATGCACCAATTGCTGTGCAGGCCGTAAGGGTTGCAACTATTA
CAGTGCTGATGGGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCA
AAAGCTTGCCCTCGGTATTGTGATACAAGAATTGCCTATTCAAAATGTCCACG
TTCCGAAGGAAACGCAGAAAATCGCATATGCACCAATTGCTGTGCAGGCCGT
AAGGGTTGCAACTATTACAGCGCTGACGGGACTTTCATTTGTGAAGGAGAGT
CTGACCCCAACAACCCAAAAGCTTGCCCTCGGTATTGTGATACAAGAATTGC
CTATTCAAAATGTCCACGTTCCGAAGGAAACGCAGAAAATCGCCTATGCACC
AATTGCTGTGCAGGCCGTAAGGGTTGCAACTATTACAGCGCTGACGGGACTT
TCATTTGTGAAGGAGAGTcTGACCCCAACAACCCAAAAGCTTGCCCTAGGAA
TTGTGATCCAAATATTGCCTATTCACTGTGCCTTTATGAAAAATAAAAGCATG
AACAGTGTCATGCTAATTTCTGTAATATGCTTGAGCGTCTATTTATGAAATAA
AAGCGTGAACAGTGTCACGCTAATTTATGTAATATGGGCGTTGGAGCCCTTTT
GGACTGAAAAAAAAAAAAAAAAAAAAAAAAAAAA

>CanPI-76 (KM576787)
GTTAGTTTCCTTGCTTCCCTACTTGTTCTTGAAATATTGCTTCTACATGTTGAT
GCCAAGGCTTGTTCAGAAGAAAACGCAGAAAATCGCATATGCACCAATTGCT
GTGCAGGCCGTAAGGGTTGCAACTATTACAGTGCTGATGGGACTTTCATTTGT
GAAGGAGAGTCTGGCCCCAACAACCCAAAAGCTTGCCCTCGGTATTGTGATA
CAAGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAACGCAGAAAATCG
CATATGCACCAATTGCTGTGCAGGCCGTAAGGGTTGCAACTATTACAGCGCT
GACGGGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAAGCTT
GCCCTCGGTATTGTGATACAAGAATTGCCTATTCAAAATGTCCACGTTCCGAA

GGAAACGCAGAAAATCGCCTATGCACCAATTGCTGTGCAGGCCGTAAGGGTT
GCAACTATTACAGCGCTGACGGGACTTTCATTTGTGAAGGAGAGTCTGACCC
CAACAACCCAAAAGCTTGCCCTAGGAATTGTGATCCAAATATTGCCTATTCAC
TGTGTCTCTATGAAAAATAAAAGCATGAACAGTGTCATGCTAATCTCTGTAAT
ATGCTTGAGCGTCTATTTATGAAATAAAAGCATGAACAGTGTCATGCTAATCT
CTGTAATATGCTTGAGCGTCTATTTATGAAATAAAAGCGTGAACAGTGTCAC
GCTAATCTATGTAATATGGGCGTTGGAGCCCTTTTGGACATAAAAAAAAAAA
AAAAAAAAAAAAA

>CanPI-77 (KM576788)
CTACTTGTTCTTGAAATATTGCTTCTACATGTTGATGCCAAGGCTTGTTCAGA
AGAAAACGCAGAAAATCGCATATGCACCAATTGCTGTGCAGGCCGTAAGGGT
TGCAACTATTACAGTGCTGATGGGACTTTCATTTGTGAAGGAGAGTCTGACCC
CAACAACCCAAAAGCTTGCCCTCGGTATTGTGATACAAGAATTGCCTATTCA
AAATGTCCACGTTCCGAAGGAAACGCAGAAAATCGCCTATGCACCAATTGCT
GTGCAGGCCGTAAGGGTTGCAACTATTACAGCGCTGACGGGACTTTCATTTGT
GAAGGAGAGTCTGACCCCAACAACCCAAAAGCTTGCCCTCGGAATTGTGATA
CAAGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAACGCAGAAAATCG
CCTATGCACCAATTGCTGTGCAGGCCGTAAGGGTTGCAACTATTACAGCGCT
GACGGGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAAGCTT
GCCCTAGGAATTGTGATCCAAATATTGCCTATTCACTGTGTCTCTATGAAAAA
TAAAAGCATGAACAGTGTCATGCTAATCTCTGTAATATGCTTGAGCGTCTATT
TATGAAATAAAAGCGTGAACAGTGTCACGCTAATCTATGTAATATGGGCGTT
GGAGCCCTTTTGGACTTAAATCATGTGTTTAAGTGTGTTAAGCTCTCTAAAAA
AAAAAAAAAAAAAAAAAA

>CanPI-78 (KM576789)
CTACTTGTTCTTGAAATATTGCTTCTACATGTTGATGCCAAGGCTTGTTCAGA
AGAAAACGCAGAAAATCGCATATGCACCAATTGCTGTGCAGGCCGTAAGGGT
TGCAACTATTACAGCGCTGATGGGACTTTCATTTGTGAAGGAGAGTCTGACCC
CAACAACCCAAAAGCTTGCCCTCGGTATTGTGATACAAGAATTGCCTATTCA
AAATGTCCACGTTCCGAAGGAAACGCAGAAAATCGCCTATGCACCAATTGCT
GTGCAGGCCGTAAGGGTTGCAACTATTACAGCGCTGACGGGACTTTCATTTGT
GAAGGAGAGTCTGACCCCAACAACCCAAAAGCTTGCCCTCGGAATTGTGATA
CAAGAATTGCCTATTCAAAATGTCCACGTTCCGAAGAAAACGCAGAAAATCG
CCTATGCACCAATTGCTGTGCAGGCCGTAAGGGTTGCAACTATTACAGCGCT
GACGGGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAAGCTT
GCCCTAGGAATTGTGATCCAAATATTGCCTATTCACTGTGTCTCTATGAAAAA
TAAAAGCATGAACAGTGTCATGCTAATCTCTGTAATATGCTTGAGCGTCTATT

TATGAAATAAAAGCGTGAACAGTGTCACGCTAATCTATGTAAACAAAAAAAA
ATAAAAAAAAAAAAAAA

>CanPI-79 (KM576792)
ATGGCTGTTCCCAAAGAAGTTAGTTTTCTTGGTTATCTACTTGTTCTTGGTAAG
ACTTTTCTTTCTCCTTTAATTTTTCTTTTTAATTTTTTCATTGGTATATAGACAC
GGTTAGAGAAGAGTTTTTTCTACCGCGAAGCAATGGAAAATATGTCTTAGGA
GTATAAAACATGATTTTCCCATCTAGTTCCTCTGTTTTATTTTACTTGTTTCAA
ATTTTTTATTAATATTATTTTTAACAAATCAATATGAAATGAAAAAAATTTCA
TGTTTTTTTTTTATATTAATTATTTTTTTAAATTAAAATATAAATATTATTTAAT
AAAAATATATATGTTATTAATTATTTTTTTAATTAATTTATCAACTTAAATTAG
TACTAGTAAAATGGGAGTACAAGTAAAATGGGAGGGGGGAGTATTTTCCACC
TTACTGAATTACTGAAAATTAAACGTATGATGGGAAACAAAAGACACAACTT
TTTCAGTGAACTACTAGGAAGCTTCCTATTAAATATTAAGCACACCTTTCAGT
GGGAAATTTTACCAGGAAAATATTGCTTTGTGGTAGTGATACTTACCTGGTGT
TATGTTTTTCCTCAAATATTTATATATAGTTTCTATATTTGCTGTAGGAATATT
TCTACATCATGTTGATGGCAAGGCTTGTACCAGAGAATGTGGCAATCTTGGGT
ATGGGATATGTCCACGTTCAGAAGGAAGTGCAGAAGAACCCATATGCACCAA
TTGTTGTGCAGGCTATAAGGGTTGTAAGTATTACAGTGCTGACGGGACTTTCA
TTTGCGAAGGAGAGTCTGACCCCAAGAACCCAAAAGGTTGCCCTCGGAATTG
TGATCCAAGAATTGCCTATTCAAAATGTCCACGTTCACAAGGAGAGACGTTA
ATTTATCCCACAGGATGCACCACCTGTTGCACGGGCTACAAAGGTTGCTACTA
TTTCGGTAAAGACGGCAAGTTTCTCTGTGAAGGAGAGAGTACTGAACCCAAG
GCTTGTACTCTGGAGTGTGATCCAAGGGTTGCTTACATGACTTGTCCCTCTAC
TGGATTGGCCAAGATTAAGGATGTTTGTGTTAACTGTTGCACTGCGGGAGAG
GGTTGCAAACTGTACGGTCATAATGGATCTCTACTTTGTACTGGAGAGTCTCA
GAGCATATCCACAGCATAG

Genomic and/or UTR sequences of CanPI-3,-4 and -7

>CanPI-4 (AY986466)
ATGGCTGTTCCCAAAGAAGTTAGtTTCCTTGCTTTCCTACtTGTTCTTGAAATAT
tGCTTCTACATGTtGATGCCAAGGCTTGTTCAGAAGAAAACGCAGAAAATCGC
ATATGCACCAAtTGCTGTGCAGGCCGTAAGGGTTGCAACTATTACAGTGCTGA
TGGGAcTtTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAAGCTTGCC
CTCGGTATTGTGATACAAGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGA
AACGCAGAAAATCGCCTATGCACCAATTGCTGTGCAGGCCGTAAGGGTTGCA
ACTATTACAGCGCTGACGGGACTTTCATTTGTGAAGGAGAGTCTGACCCCAA
CAACCCAAAAGCTTGCCCTCGGAATTGTGATACAAGAATTGCCTATTCAAAA
TGTCCACGTTCCGAAGGAAACGCAGAAAATCGCCTATGCACCAATTGCTGTG

CAGGCCGTAAGGGTTGCAACTATTACAGCGCTGACGGGACTTTCATTTGTGA
AGGAGAGTCTGACCCCAACAACCCAAAAGCTTGCCCTAGGAATTGTGATCCA
AATATTGcCTATTCACTGTGTCTCTATGAAAAGTAAAAGCATGAACAGTGTCA
TGCTAATCTCTGTAATATGCTTGAGCGTCTATTTATGAAATAAAAGCGTGAAC
AGTGTCACGCTAATCTATGTAATATGGGCGTTGGAGCCCTTTTGGACTTAAAT
CATGTGTTTAAGTGTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA

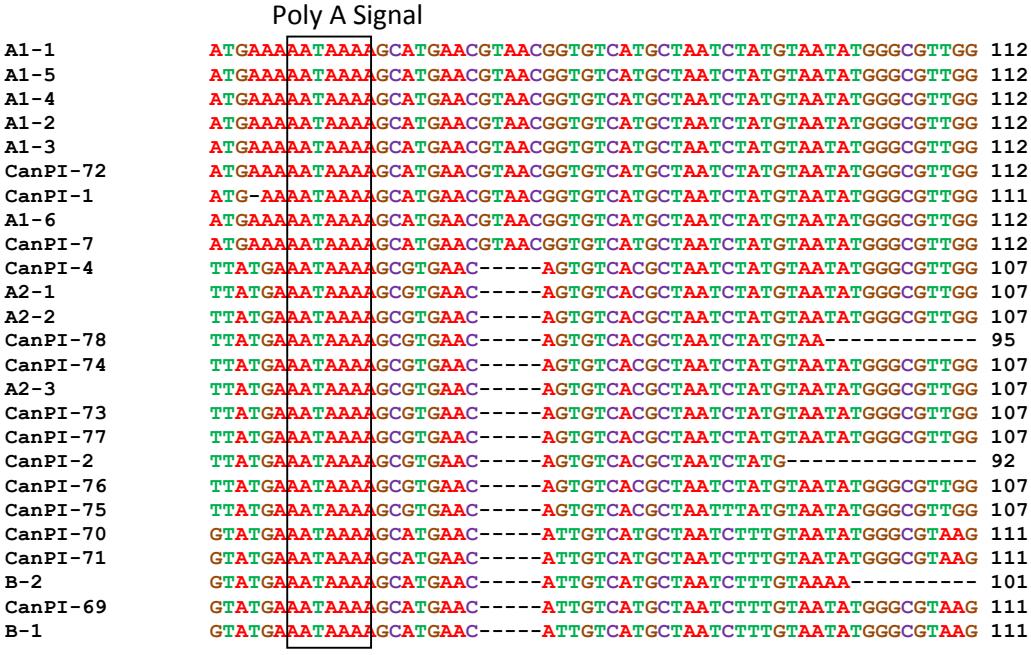
>CanPI-3 (AY986465)
ATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTTCCTACTTGTTCTTGAAATA
TTGCTTCTACATGTTGATGCCAAGGCTTGTTCAGAAGAAAACGCAGAAAATC
GCATATGCACCAATTGCTGTGCAGGCCGTAAGGGTTGCAACTATTACAGTGC
TGATGGGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAAGCT
TGCCCTCGGTATTGTGATACAAGAATTGCCTATTCAAAATGTCCACGTTCCGA
AGGAAACGCAGAAAATCGCATATGCACCAATTGCTGTGCAGGCCGTAAGGGT
TGCAACTATTACAGCGCTGACGGGACTTTCATTTGTGAAGGAGAGTCTGACC
CCAACAACCCAAAAGCTTGCCCTCGGTATTGTGATACAAGAATTGCCTATTCA
AAATGTCCACGTTCCGAAGGAAACGCAGAAAATCGCCTATGCACCAATTGCT
GTGCAGGCCGTAAGGGTTGCAACTATTACAGCGCTGACGGGACTTTCATTTGT
GAAGGAGAGTCTGACCCCAACAACCCAAAAGCTTGCCCTAGGAATTGTGATC
CAAATATTGCCTATTCACTGTGTCTCTATGAAAAATAAAAGCATGAACAGTGT
CATGCTAATCTCTGTAATATGCTTGAGCGTCTATTTATGAAATAAAAGCGTGA
ACAGTGTCACGCTAATCTATGTAATATGGGCGTTGGAGCCCTTTTGGACTTAA
ATCATGTGTTTAAGTGTGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA

>CanPI-7 (DQ005913)
AAAAATTTATGTTGAAAATAAAATGCATAATTCACATTTTTTGTGAGTTATGT
TAAAATAAGAATGAATAACTCACAAAAATTGTAAGTTATGCACTTTTCAAAA
AATTTAACACTGTAAACTACCTTTCGTTAAAGATATTTATGCATAACTCATCA
TTTTTGTGAGTTAGGCATAGATTGATACATTTTTAATACATGACATAATTTGA
TGTTTTTTTTTAATTTGATGGCATTTTTTCCGAACTCTTACGTGGAGGAAAATT
ATTGCAAGATGACAAATATTACAAAAACAAGCAAGTCGGGGTCAATATTAAC
TTGCTTAATTATTAGGAGAAAGAATATTAACTACTAGGAAAAAGACAAGCTT
TCTAATAAAAAAATCAAGCGTTAGTACCTAATTTCCCATTTTATATTAATGTG
AAATAAAAGTTTTAGTTAGCCACGTATCTTTTATTTGTTTGCTATAAATAGGT
GGTGATGGGTAGACTACCTCATCCATAACGAAATCAAGGTAAATAATCACCA
TCCATCAGTAATATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTCCCTACTT
GTACTTGGTAAGAGTTTCCTTTCTCCTTTTTCTATTTAATTTCTTCATTAGTATA
TACACTGTTAGAGATAGAGGAAAACATGATTTTTTCATCCATTTCTCACCTTA
CTGATTACTGGAAAATGTATGGTGAGAAACAGTTTTATATTCAAACACTAGA
AAGCTTCCTAATTTTAAATATTAAGTATATTTTTCTGCGGCAAATTTTAGTAG

GAAAATATTTGATTTTGTAGTAGTGATACACAGAAGTGGTGATATGTTTATTC
TTAATATTTATATATAGCTTATATATTTATTGTAGGATTAGTAGGAAAATATT
TGATTTGTAGTAGTGATACAACAGAAGTGGTGATATGTTTATTCTTAATATTT
ATATATAGCTTATATATTATTGTAGGAATATTGCTTCTACATGTTGATGCCAA
GGCTTGTTCACAAAGAAACGCAAAAGAACCCATATGCACCAATTGTTGTGCA
GGCCGTAAGGGTTGCAACTATTACAGTGCTGATGGGACTTTCATTTGTGAAG
GAGAGTCTGACCCCAACAACCCAAAACTTTGTACTTTGAACTGTGATCCAAG
AATTTTCTATTCAAAGTGTCCACGTTCCGAAGGAAACGCAGAAAATCGCATA
TGCACCAATTGCTGTGCAGGCCGTAAGGGTTGCAACTATTACAGTGCTGATG
GGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAACCTTGCCC
TCGGAATTGTGATACAAGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGA
AACGCAGAAAATCGCATATGCACCAATTGCTGTGCAGGTCGTAAGGGTTGCA
ACTATTACAGTGCTGACGGGACCTTCATTTGTGAAGGAGAGTCTGACCCCAA
CAACCCAAAACCTTGCACTCTGAACTGTGATCCAAGAATTTTCTATTCAAAGT
GTCCACGTTCCGAAGCAAGTGCAGAACAACCCATATGCACCAATTGTTGTGC
AGGCCTCAAGGGTTGCAACTATTACAATGCTGACGGGACTTTCATTTGTGAG
GGAGAGTCTGACCCCAACCACCCAAAAGCTTGCCCCAAGAATTGTGATCCTA
ATATTGCCTATTCACTATGTCTGTATGAAAAGTAAAAGCATGAACAGTGTCAT
GCTAATCTCCGTAATACGCATGGGCGTCTATATGAAAAATAAAAGCATGAAC
GTAACGGTGTCATGCTAATCTATGTAATATGGGCGTTGGAGCCCTTTTGGACT
TAAATAATATGTGTTCAAGTGTGTTAAGCTCTCAAAAAAAAAAAAAAAAAAA
AAAAAAAAAA

Appendix III

Alignment of 25 3'UTR sequences of *CanPI* genes obtained from analysis described in section 2B.2.2. Three groups of 3'UTRs of Pin-II type A1, A2 and B are evident



```
A1-1          AGCCCTTTTGGACTTAAATA----ATATGTGTTCAA---------------------- 144
A1-5          AGCCCTTTTGGACTTAAATA----ATATGTGTTCAAGTGTGTTAAGCTCTCTAATAAAGA 168
A1-4          AGCCCTTTTGGACTTAAATA----ATATGTGTTCAAGTGG------------------- 148
A1-2          AGCCCTTTTGGACTTAAATA----ATATGTGTTCAAGTGTGTTAAG------------- 154
A1-3          AGCCCTTTTGGACTTAAATA----ATATGTGTTCAAGTGTGTTAAGCTCCCTG------- 161
CanPI-72      AGCCCTTTTGGACGCATAAA----AAAAAA------------------------------ 138
CanPI-1       AGCCCTTTTGGACTTAAATA----ATATGTGTTCAAGTGTGTTAAG------------- 153
A1-6          AGCCCTTTTGGACTTAAATA----ATATGTGTTCAAGTGTGTTAAG------------- 154
CanPI-7       AGCCCTTTTGGACTTAAATA----ATATGTGTTCAAGTGTGTTAAG------------- 154
CanPI-4       AGCCCTTTTGGACTTAAATC----ATGTGTTTAAGTGTG---------------------- 142
A2-1          AGCCCTTTTGGACTTAAATC----ATGTGTTTAAGTGTG---------------------- 142
A2-2          GGCCCTTTTGGACTTAAATC----ATGTGTTTAAGTGTG---------------------- 142
CanPI-78      ------------------------------------------------------------
CanPI-74      AGCCCTTTTGGACTTAAATC----ATGTGTTTAAGTGTG---------------------- 142
A2-3          AGCCCTTTTGGAC------------------------------------------------ 120
CanPI-73      AGCCCTTTTGGACT----------------------------------------------- 121
CanPI-77      AGCCCTTTTGGACTTAAATC----ATGTGTTTAAGTGTGTTAAGC--------------- 148
CanPI-2       ------------------------------------------------------------
CanPI-76      AGCCCTTTTGGAC------------------------------------------------ 120
CanPI-75      AGCCCTTTTGGACTG---------------------------------------------- 122
CanPI-70      AGCTCTTTTGGACATAAATAGGTGTCCAGTCTCCAAGTGTTTAATGAAGCACTA------ 165
CanPI-71      AGCTCTTTTGGACATAAATAGGTGTCCAGTCTCCAAGTGTTTAATGAAGCACTA------ 165
B-2           ------------------------------------------------------------
CanPI-69      AGCTCTTTTGGGCAAAA-------------------------------------------- 128
B-1           AGCTCTTTTGGACATAAATAGGTGTCCAGTCTCCAAGTGTTTAATGAAGCACTA------ 165
```

```
A1-1          -------------------AAAAAAAA----------------------AAAAAAAAA 161
A1-5          AGCAACCGTACGTACCTCCAAAAAAAA----------------------AAAAAAAAA 204
A1-4          -------------------AAAAAAAAA---------------------AAAAAAAAA 166
A1-2          ------------CTCTCTGAAAAAAAA----------------------AAAAAAAAA 178
A1-3          ------------CAAAAAAAAAAAAAA----------------------AAAAAAAAA 185
CanPI-72      -------------------AAAAAAAAA---------------------AAAAAAAAA 156
CanPI-1       ------------CTCTCTAATAAAGAAGCAACCGTACGTACCTTCACACAGAAAAAAAAA 201
A1-6          ------------CTCTCAAAAAAAAAA----------------------AAAAAAAAA 178
CanPI-7       ------------CTCTCAAAAAAAAAA----------------------AAAAAAAAA 178
CanPI-4       ----------------TTAAAAAAAAA----------------------AAAAAAAAA 162
A2-1          ----------------TTAAAAAAAAA----------------------AAAAAAAAA 162
A2-2          ----------------TTAAAAAAAAA----------------------AAAAAAAAA 162
CanPI-78      -------------------ACAAAAAAA----------------------AATAAAAAA 113
CanPI-74      ----------------GCAAAAAAAAA----------------------AAAAAAAAA 162
A2-3          ----------------GCAAAAAAAAA----------------------AAAAAAAAA 140
CanPI-73      -------------------CGAAAAAA----------------------AAAAAAAAA 138
CanPI-77      ----------------TCTCTAAAAAA----------------------AAAAAAAAA 168
CanPI-2       --------------------TAAAAAA----------------------AAAAAAAAA 108
CanPI-76      -------------------ATAAAAAA----------------------AAAAAAAAA 137
CanPI-75      ----------------AAAAAAAAAAA----------------------AAAAAAAAA 142
CanPI-70      ------------GTTTCTTTAGAAAAA----------------------AAAAAAAAA 189
CanPI-71      ------------GTTTCTTTTACTAAA----------------------AAAAAAAAA 189
B-2           ---------------------AAAAAA----------------------AAAAAAAAA 116
CanPI-69      ---------------------AAAAAA----------------------AAAAAAAAA 143
B-1           ------------GTTTCTTTTATAAAA----------------------AAAAAAAAA 189
```

```
A1-1            AAAAAAA--------- 168
A1-5            AAAAAAAAA------- 213
A1-4            AAAAAAAAA------- 175
A1-2            AAAAAAAAA------- 187
A1-3            AAAAAAAAA------- 194
CanPI-72        AAAAAAAAA------- 165
CanPI-1         AAAAAAAAAA------ 211
A1-6            AAAAAAAAA------- 187
CanPI-7         AAAAAAAAAA------ 188
CanPI-4         AAAAAAAAAAA----- 173
A2-1            AAAAAAAAAA------ 172
A2-2            AAAAAAAAA------- 171
CanPI-78        AAAAAAAAA------- 122
CanPI-74        AAAAAAAAAAAA---- 174
A2-3            AAAAAAA--------- 147
CanPI-73        AAAAAAAA-------- 146
CanPI-77        AAAAAAAAA------- 177
CanPI-2         AAA------------- 111
CanPI-76        AAAAAAAAA------- 146
CanPI-75        AAAAAAAAA------- 151
CanPI-70        AAAAAAAA-------- 197
CanPI-71        AAAAAAAA-------- 197
B-2             AAAAAAAA-------- 124
CanPI-69        AAAAAAAA-------- 151
B-1             AAAAAAAAAAAAAAAA 205
```

RESUME

Name : **Ms. Neha Sandeep Mahajan**

Address : 406 B- wing, Sneha Paradise, Near Warje Police station, Warje Malwadi, Pune, Maharashtra, India, Pin: 411 058

Date of Birth : 6th Nov 1986

Contact No. : +91 9890061063

Email : neha.m611@gmail.com

Marital Status : Married (Spouse: Mr. Vivek S. Deshpande)

Nationality : Indian

Place of Birth : Ahmednagar, Maharashtra, India

Academic Qualifications

Degree	Year of Passing	University / Board	Class
Secondary School Certificate	March 2002	Maharashtra State Board	Distinction
Higher Secondary Certificate	March 2004	Maharashtra State Board	Distinction
Bachelor of Science (B.Sc.) Biotechnology	April 2007	University of Pune	Distinction
Master of Science (M.Sc.) Biotechnology	May 2009	University of Pune	First Class
Ph.D. Biotechnology (pursuing) Biotechnology	-	University of Pune	-

Ph.D.thesis title: Efficacy of diverse *Capsicum annuum* protease inhibitors against the adaptive plasticity of *Helicoverpa armigera* proteases.

Other Qualifications

- National Eligibility Test for Junior Research Fellowship (NET, JRF) (December 2008), conducted by Council of Scientific and Industrial Research and University Grant Commission (CSIR-UGC), Government of India.
- Graduate Aptitude Test for Engineering (GATE) 2009 in Life Sciences; Percentile Score: 92.95

Honors/Awards

- Financial assistance from International travel support scheme (ITS) of Department of science and technology (DST) for participation (poster and oral presentation) in International symposium at Guangzhou, China.

- Junior Research Fellowship (2009-2014) from the Council of Scientific and Industrial Research (CSIR), India for doing Ph.D. at the National Chemical Laboratory, Pune, India.
- Best Poster award for the poster "Proteomic view of the responses of *Helicoverpa armigera* to *Capsicum annuum* protease inhibitor-7" presented during National Science Day, 2013 at the CSIR-National Chemical Laboratory, Pune.

Publications

1) Tamhane VA, Mishra M, **Mahajan NS**, Gupta VS and Giri AP (2012) Plant Pin-II family proteinase inhibitors: Structural and functional diversity. *Functional Plant Science and Biotechnology* 6 , 42-58
2) Mishra M, **Mahajan NS**, Tamhane VA, Kulkarni MJ, Baldwin IT, Gupta VS and Giri AP (2012) Stress inducible proteinase inhibitor diversity in *Capsicum annuum. BMC Plant Biology* 12, 217
3) **Mahajan NS**, Mishra M, Tamhane VA, Gupta VS and Giri AP (2013) Plasticity of protease gene expression in *Helicoverpa armigera* upon exposure to multi-domain *Capsicum annuum* protease inhibitor. *Biochimica et Biophysica Acta* 1830, 3414-3420
4) **Mahajan NS**, Mishra M, Tamhane VA, Gupta VS and Giri AP (2014) Stress inducible proteomic changes in *Capsicum annuum* leaves. *Plant Physiology and Biochemistry* 74, 212-217
5) **Mahajan NS**, Deewangan V, Joshi RS, Lomate P, Gupta VS and Giri AP (2014) Structural features of diverse *Capsicum annuum* protease inhibitor genes. *Planta* (In Press)
6) Lomate PR, **Mahajan NS**, Kale SM, Gupta VS and Giri AP (2014) Identification and expression profiling of *Helicoverpa armigera* microRNAs and their possible role in the regulation of digestive proteases. *Insect Biochemistry and Molecular Biology* (In Press)

Conferences/Workshops

- **Oral presentation:** Giri AP, Mahajan NS, "*Molecular responses of Helicoverpa armigera to plant protease inhibitors*" In: "The Second International Symposium on Insect Midgut Biology" at South China Normal University, Guangzhou, China during September 24-28, 2012.
- **Poster presentation:** Mahajan NS, Mishra M, Tamhane VA, Gupta VS and Giri AP "*Helicoverpa armigera* countered with proteinase inhibitor from *Capsicum annuum*: Sensitivity or tolerance?" In: "The Second International Symposium on

Insect Midgut Biology" at South China Normal University, Guangzhou, China during September 24-28, 2012.

- **Poster presentation:** Mahajan NS, Lomate P, Gupta VS and Giri AP "Unraveling the mechanisms regulating protease gene expression in *Helicoverpa armigera*" In: Indo-Mexico Workshop on Biotechnology: Beyond Borders at CSIR-National Chemical Laboratory, Pune, India during October 7-9, 2013.
- **Poster presentation**: Mahajan NS, Deewangan V, Joshi RS, Lomate P, Gupta VS and Giri AP "Characterisation of diverse protease inhibitor genes from *Capsicum annuum*" In: 35th Annual meeting of Plant Tissue Culture Association & National Symposium on "Advances in Plant Molecular Biology and Biotechnology" at Indian Institute of Science Education and Research, Pune, India during March 10-12, 2014.
- Attended the 95th Indian Science Congress Conference at Vishakhapatnam in January 2008, India.
- Attended symposium at IIT Bombay on "Emerging Areas in Bioscience and Bioengineering" in February 2009, India.
- Attended "RNA-2010" A National Symposium of the RNA group of India, Department of Biotechnology, University of Pune, India.