Interactions between diverse proteinase inhibitors from *Capsicum annuum* and insect pests: A biochemical & molecular approach

A thesis submitted to the
University of Pune

For the degree of
DOCTOR OF PHILOSOPHY

In
BIOTECHNOLOGY

By
MANASI MISHRA

Research Guide

Dr. VIDYA S. GUPTA
Plant Molecular Biology Group
Division of Biochemical Sciences
National Chemical Laboratory
Pune 411008 (INDIA)

JULY 2012

This work was performed at

NATIONAL CHEMICAL LABORATORY, PUNE, INDIA

Scientists involved were

Dr. (Mrs.) Vidya S. Gupta (Research supervisor)

Dr. Ashok P. Giri (Co-supervisor)

CERTIFICATE

Certified that the work in the Ph.D. thesis entitled **'Interactions between diverse proteinase inhibitors from *Capsicum annuum* and insect pests: A biochemical & molecular approach'** submitted by **Ms. Manasi Mishra** was carried out by the candidate under my supervision. The material obtained from other sources has been duly acknowledged in the thesis.

Date:

Pune

Dr. Vidya S. Gupta

(Research Guide)

Dr. Ashok P. Giri

(Co-guide)

DECLARATION

I hereby declare that the thesis entitled **'Interactions between diverse proteinase inhibitors from *Capsicum annuum* and insect pests: A biochemical & molecular approach'** submitted for Ph.D. degree to the University of Pune has not been submitted by me for a degree at any other university.

Date: Manasi Mishra

National Chemical Laboratory, Pune

Dedicated to my beloved parents

And

The plant biologists

"It is not the strongest of the species that survives, nor the most intelligent, but the one most responsive to change"

— Charles Darwin

CONTENTS

ACKNOWLEDGEMENTS

This thesis was made possible by the immense and much appreciated help and support of a number of people and I take this opportunity to thank them. First of all, I would like to express my sincere gratitude to my supervisor, Dr. Vidya Gupta for her continuous support, encouragement, guidance and advice throughout the course of my doctoral candidature. She herself by her punctuality, enthusiasm, and devotion to work sets an ideal example before me, which has always given me a boost for carrying out the Ph.D. work. I am especially indebted to my Co-supervisor, Dr. Ashok Giri who has been a true friend and mentor throughout my candidature. His enthusiasm has always inspired me and I thank him for helping me in developing a good research plan and its efficient execution.

I highly appreciate the constructive suggestions offered by Dr. Mahesh Kulkarni and Dr. Sushama Gaikwad in planning the work for two chapters. I express my heartfelt gratitude to Dr. Jochen Muller Dieckmann for his kindness and academic guidance during my research work in the European Molecular Biology Laboratory, Germany. I would also like to acknowledge the critical suggestions provided by Dr. Ian T. Baldwin (Max-Planck institute for Chemical Ecology, Germany) in one of the chapters and improving it to a larger extent. I would like to thank Dr. Narendra Kadoo for his help in DNA sequencing and all timely support throughout my tenure. The valuable and critical inputs in the research work by Dr. Suresh and Dr. Rajesh Singh are highly acknowledged.

I feel happy to express my thanks to Dr. Sourav Pal, Director and Dr. S. Sivaram, ex-Director, NCL for providing me the research facilities.

I take immense pleasure in saluting late Prof. V.S. Jaiswal (Department of Botany, Banaras Hindu University) for his exceptional teaching and motivation to pursue research career.

This work would never be complete without expressing my gratitude to Dr. Vaijayanti Tamhane for her contribution in

developing my PhD objectives and for her suggestions and constant encouragement throughout the work. I am extremely grateful to Neha Khandelwal, Neha Mahajan and Rakesh Joshi for their helping hands in work and support all the way throughout. I would like to extend special thanks to Dr. Saravanan Panneerselvam for his help and timely advices in my research work during my tenure at EMBL, Germany.

I would like to thank Dr. Bhushan Dholakia, Sofia, Pavan, Priya and Yojana for always being helpful. I thank Ajit, Amey, Ashish, Atul, Ashiwini, Hemangi, Gouri, Gayatri, Hemalata, Manje, Medha, Priyanka, Radhika, Ram, Rasika, Ramya, Reema, Rahul, Rebecca, Savita, Sheon, Sachin, Sandip, Vitthal, Vishal, Varsha, Yashwant and all other labmates for their timely help. I reserve special thanks to Charu, Sheetal and Yashshree who continue to manage the scientific requirements of the group. I thank Mr. Jagtap for his refreshing support in the laboratory.

My friends have always stood by my side through good times and rough patches. I owe my thanks to my hostel inmates and friends, who have made my stay immensely enjoyable.

I find no words for my parents who have been a constant support and have always encouraged me to achieve my dreams. Special thanks to my brothers, Vinod and Pramod for providing encouragement and moral support on the bumpy ride of my PhD candidature. Its a pleasure to acknowledge Ankita and Suneela for their encouraging words.

I thank CSIR, India for supporting my research work through Research fellowships and DAAD, for funding my stay at EMBL, Germany for six months.

Manasi

LIST OF ABBREVIATIONS

AA	Amino Acids
ACN	Acetonitrile
AI	Aphid infestation
ANS	1-anilino-8-naphthalenesulfonate
BA*p*NA	Benzoyl-DL-Arginyl-*p*-Nitroanilide
BMMY	Buffered Methanol-complex Medium
Bt	*Bacillus thuringiensis*
CanPI	*Capsicum annuum* Proteinase Inhibitor
CBB-R250	Coomasssie Brilliant Blue-R250
CD	Circular dichroism
cDNA	complementary deoxyribonucleic acid
CI	Chymotrypsin Inhibitor
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
g, mg, µg, ng	gram, milligram, microgram, nanogram
GXCT	Gel- X ray film Contact-print Technique
HCl	Hydrochloric acid
HGPI	*Helicoverpa armigera* Gut Proteinase Inhibitor
HGPs	*Helicoverpa armigera* Gut Proteinases
HIC	Hydrophobic Interaction Chromatography
IC_{50}	Inhibitor concentration at 50% inhibition
IEF	Iso-electric Focussing
IF	Intensity Fading
IMAC	Immobilized metal ion affinity chromatography
IPG	Immobilized pH gradient
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IRD	Inhibitory Repeat Domain
JA	Jasmonic acid
kDa/kD	kilo Dalton
K_i	Inhibition constant
L, mL, µL	liter, milliliter, microliter
LIC	Ligation independent cloning

M, mM, µM	molar, millimolar, micromolar
MALDI-TOF-MS	Matrix assisted laser desorption ionization-Time of flight-Mass spectrometry
MeJA	Methyl Jasmonate
MGY	Minimal Glycerol Medium
MM	Minimal Methanol
mmole, µmole	millimole, micromole
mRNA	messenger RNA
MRE	Mean residual ellipticity
NaCl	Sodium chloride
NaOH	Sodium hydroxide
NaProPI	Pin-II proteinase inhibitor precursor from *N. alata*
NaTPI	*N. alata* trypsin proteinase inhibitor
OD	Optical Density
OS	Oral secretions
PAGE	Poly-acrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
PDB	Protein data bank
PI	Proteinase Inhibitor
PIN-I/II	Potato Proteinase Inhibitor I/II
rCanPI	recombinant *C. annuum* proteinase inhibitor
RDB	Regeneration Dextrose Medium
RNA	Ribonucleic acid
rpm	revolutions per minute
RSL	Reactive site loop
RT-PCR	Reverse Transcriptase-Polymerase Chain Reaction
SA	Salicylic Acid
SDS	Sodium Dodecyl Sulphate
SGP	*S. exigua* gut proteases
TFA	Trifluoroacetic acid
TI	Trypsin inhibitor
TIU	Trypsin inhibitory units
TRIS	Tris-Hydroxymethyl Aminomethane
2-D	2-dimensional gel electrophoresis

LIST OF FIGURES

Chapter 1

Chapter 2

Chapter 3

Chapter 4

LIST OF TABLES

Chapter 1

Chapter 2

Chapter 3

APPENDIX

.

THESIS SUMMARY

Interactions between diverse proteinase inhibitors from *Capsicum annuum* and insect pests: A biochemical and molecular approach

Lepidopteran insect pests are responsible for causing severe losses of several crop plants. Proteinase inhibitor (PI) based approach is extensively in focus, for environment friendly pest control approaches and detailed study of plant-insect interactions. I have carried out my thesis work on Pin-II type PIs from a non-preferred host of *H. armigera*, *Capsicum annuum*. The interaction between these *C. annuum* PIs and insect pests have been studied using various biochemical, molecular and proteomic approaches. The major objectives and the results obtained are summarized below:

Implications of induced proteinase inhibitor diversity in *Capsicum annuum*

Capsicum annuum, has an array of Pin-II type PI genes (*CanPIs*) displaying regulated expression in steady state and induced conditions. Induction experiments were performed on *C. annuum* leaves *viz.* mechanical wounding followed by treatment with oral secretions of *H. armigera* (W+OS), with water (W+W) and aphid infestation (AI). We cloned and sequenced the treatment specific *CanPI* transcripts to identify distinct *CanPI* genes for each treatment, which to our surprise yielded 44 new diverse Pin-II genes. Sixty seven *CanPIs* comprising 55 unique sequence variants of IRDs, illustrating varied distribution across treatments were identified. The overall abundance of 3-IRD PIs and TI domains were prevalent across these treatments whereas distinct equal affluence of 4-IRD PIs was credible in wounding with OS. The *CanPI* expression pattern in response to OS (elicitors) was markedly different from the other two treatments viz. wounding with water and aphid infestation, signifying the defense related role of CanPIs against lepidopteran pests. A differential pattern of induced PI activity accumulation was observed in *C. annuum* leaves upon various inductions. Further characterization of the induced PIs by protease inhibitory activity assays, 1D and 2-dimensional gel electrophoresis followed by TI visualization and MALDI-TOF-MS, corroborated their structural and functional diversity. Across various plant

parts of *C. annuum* also, the qualitative and quantitative variations in PI activity were evident with flower showing the highest accumulation of PI activity.

Interaction of recombinant CanPIs with *Helicoverpa armigera* gut proteases reveals their processing patterns, stability and efficiency

Six diverse representative *C. annuum* genes (*CanPI-13, -15, -19, -22, -5* and *-7*) comprising one, two, three or four IRDs, were selected for cloning and expression in *P. pastoris*. Recombinant proteins were characterized with specific reference to their (i) processing by *H. armigera* gut proteinases (HGP) (ii) stability in proteolytic environment (iii) inhibitory activity against HGP. PAGE and MALDI-TOF-MS techniques revealed presence of multiple, processed repeats in the purified rCanPI proteins. rCanPIs were resolved on native and SDS-PAGE and visualized for TI profiles. Multiple TI activity isoforms were detected for the rCanPIs and indicated presence of heterogeneity at the activity level. rCanPIs were used for inhibition studies against trypsin, chymotrypsin and *H. armigera* gut proteinases of the 4^{th} instar larvae fed on artificial diet (AD-HGP). The interactions of rCanPIs with various proteinases were studied by IF-MALDI-TOF-MS and revealed the processing of multi-IRD proteins at the linker regions by the HGPs. *In vitro* and *in vivo* stability of rCanPIs was analyzed by native in-gel TI activity visualization. The stability of individual rCanPI varied in the proteolytic environment. rCanPI-5 and rCanPI-7 showed maximum inhibition of HGP isoforms and their processed units were also found to be stable in presence of HGP. Even single aa variations in IRDs were found to significantly change the HGP inhibition specificity. Results demonstrated the low efficiency of single IRD CanPIs against HGPs; indicating importance of presence of multiple IRD genes *in planta* for defense.

Structural-functional insights of single and multi-domain *Capsicum annuum* proteinase inhibitors

Owing to the inducibility of 4-IRD CanPIs under insect infestation, higher stability and efficiency against HGPs, CanPI-7 was used for further biochemical and biophysical characterization under various conditions of temperature, pH and against diverse proteases. CanPI-15 and CanPI-7 were cloned in bacterial

expression system for the requirement of more purified protein for structural studies. Recombinant proteins were expressed and purified by Nickel affinity chromatography followed by size-exclusion. The purified protein preparations were analyzed for their interaction with broad range of proteases like trypsin, chymotrypsin, elastase, HGP, *Spodoptera exigua* gut proteases and showed 60-95% inhibition. The inhibition kinetic studies of CanPIs with trypsin revealed higher binding efficiency and potency of CanPI-7 as indicated by IC_{50} and K_i. The proteins were characterized for their optimum activity under varied conditions of temperature and pH and respective secondary structural changes in the proteins were monitored using the fluorescence spectroscopy, ANS binding assays and CD spectroscopy. CanPI proteins exhibited stability over a wide range of conditions. Secondary structure analysis by CD and *in silico* structure prediction for CanPI-7 gave interesting insights *viz*., adaptability of CanPIs to attain polyproline fold and spatial arrangement of multiple domains in precursor molecules. The generated models were validated and assessed using bioinformatics tools. Molecular docking studies on CanPI-7 and the target proteases revealed the probable binding mechanism of CanPI precursors with multiple protease molecules at the same time thus, contributing its higher potency and inhibition efficiency.

The present study on *C. annuum* has brought in to light a diverse array of Pin-II PIs expressed in various tissues and under various conditions in the plant, with emphasis on their defense role. The characterization of representative CanPIs have indicated their high efficiency and promising potential in inhibiting lepidopteran insect gut proteases. Thus, the naturally occurring gene diversity in CanPIs provides an effective starting material to reach the goal of crop protection.




Chapter 1

General Introduction and Review of Literature




CHAPTER 1

Review of Literature: Insights in to plant proteinase inhibitors with special reference to wound inducible PIs based plant defense against insect pests

Stresses imposed by other organisms that adversely affect the growth, development and productivity of plants are biotic stresses. The sources of principle biotic stress are pests, pathogens, other plants and herbivores. Insect pests, diseases and weeds are the major constraints limiting the potential agricultural production. It is estimated that herbivorous insects alone lead to about 26% loss of the food production all over the world while India losses about 18% of the crop yield valued at Rs. 90,000 crore due to pest attacks each year (http://www.livemint.com/2009/02/25181538/India-loses-Rs90000-cr-crop-y.html).

1.1 Lepidopteran insects

Lepidoptera is the second largest order in the class Insecta which includes moths and butterflies. The larvae of many lepidopteran species are herbivores and major pests in agriculture. Some of the major pests belong to the families Tortricidae, Noctuidae, and Pyralidae. The larvae of the Noctuidae genus *Spodoptera* (armyworms) and *Helicoverpa* (corn earworm) can cause extensive damage to certain crops (**Fig. 1.1**). The host plant range for Lepidopteran insects may either be narrow (monophagous or specialist), e.g., *Manduca sexta*, which exhibits preference for Solanaceous plants, or diverse (polyphagous or generalist), e.g., *Helicoverpa armigera*, which feeds on various legumes, vegetables and fruits (Ehrlich and Raven, 1964; Tamhane et al., 2005a; Wu and Baldwin, 2010). Some butterflies and moths affect the economy negatively while; some species are valuable economic resources. The most prominent example is of the domesticated silkworm moth (*Bombyx mori*), the larvae of which make their cocoons out of silk, which can be spun into cloth. A high-quality assembly of the silkworm genome sequence has been completed recently (Xia et al., 2004; Wang et al., 2005) and is being used as a model for insect genetics and genomics. Silkworm is also being used as a bioreactor for proteinaceous drugs and as a source of biomaterials.

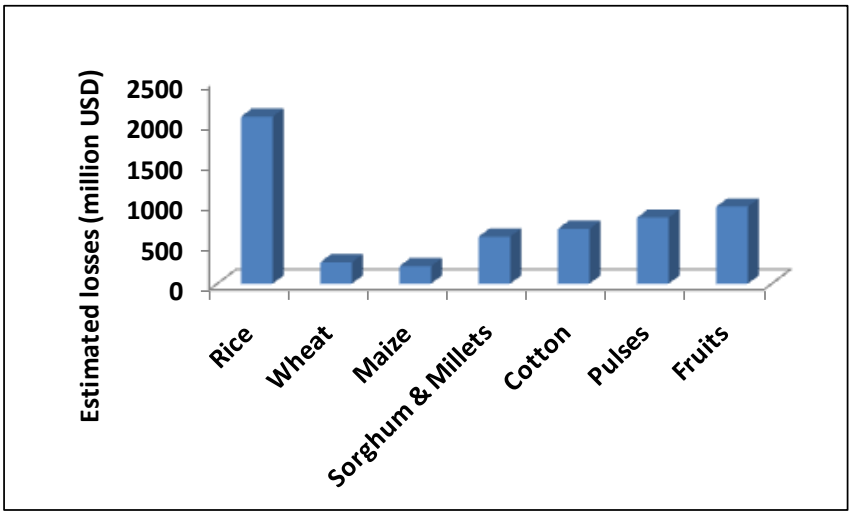


Figure 1.1: Losses in field crops due to insect pests [Adopted from Reddy and Zehr, 2004].

1.1.1 *Helicoverpa armigera* (Hübner)

Members of genus *Heliothis* (Family: Noctuidae) are agricultural pests of worldwide significance. *H. armigera, H. zea, H.virescens* and *H. punctigera* are highly polyphagous and damage a large number of plant species, including food, fiber, oil, fodder and also horticultural and ornamental crops (Fitt, 1989). *H. armigera* infests about 300 plant species and the most important host crops include cotton, legumes, tomato, tobacco, okra, potato, sunflower, safflower, maize, groundnut, etc. (Fitt, 1989; Rajapakse and Walter, 2007) (**Fig. 1.2**). Preferentially larvae of *H. armigera* feeds on the reproductive structures and economically important plant parts like seeds, fruits, pods which lead to prominent reduction in the yield of the crop. The losses due to *H. armigera* infestation are estimated to be 20 to 80% in cotton, 14 to 100% in pigeon pea, 18 to 26% in sorghum, 30 to 60% in sunflower, 15 to 46% in tomato and over 40% in okra (Reddy and Zehr, 2004). It inhabits many countries of Asia, Europe, Australia and Africa and is well known pest because of its wide geographic presence. The characteristic features which make it a successful pest are polyphagy, high reproductive rate, mobility and facultative diapause (Fitt, 1989).

H. armigera has a life span ranging between 25 to 35 days. Each female can lay several hundred eggs on lower surface of leaves, flowers, shoot tips and young pods, where larva feeds preferentially. The eggs are hatched after 3-4 days, while the larval stage lasts for 12-16 days. The larval period is divided into six instars in which

the fourth and the fifth are highly voracious and the most damaging ones (Tamhane et al, 2005a). The caterpillars are aggressive, occasionally carnivorous and, when the opportunity arises, cannibalistic. Pupa is a non feeding stage which lasts normally for 6-10 days and the moth remains alive for 4-5 days (Gowda et al., 2005; Tamhane et al., 2005a). In case of unfavourable conditions, the larvae undergo facultative diapause i.e. a state of suppressed metabolism as pupae in order to survive. 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.

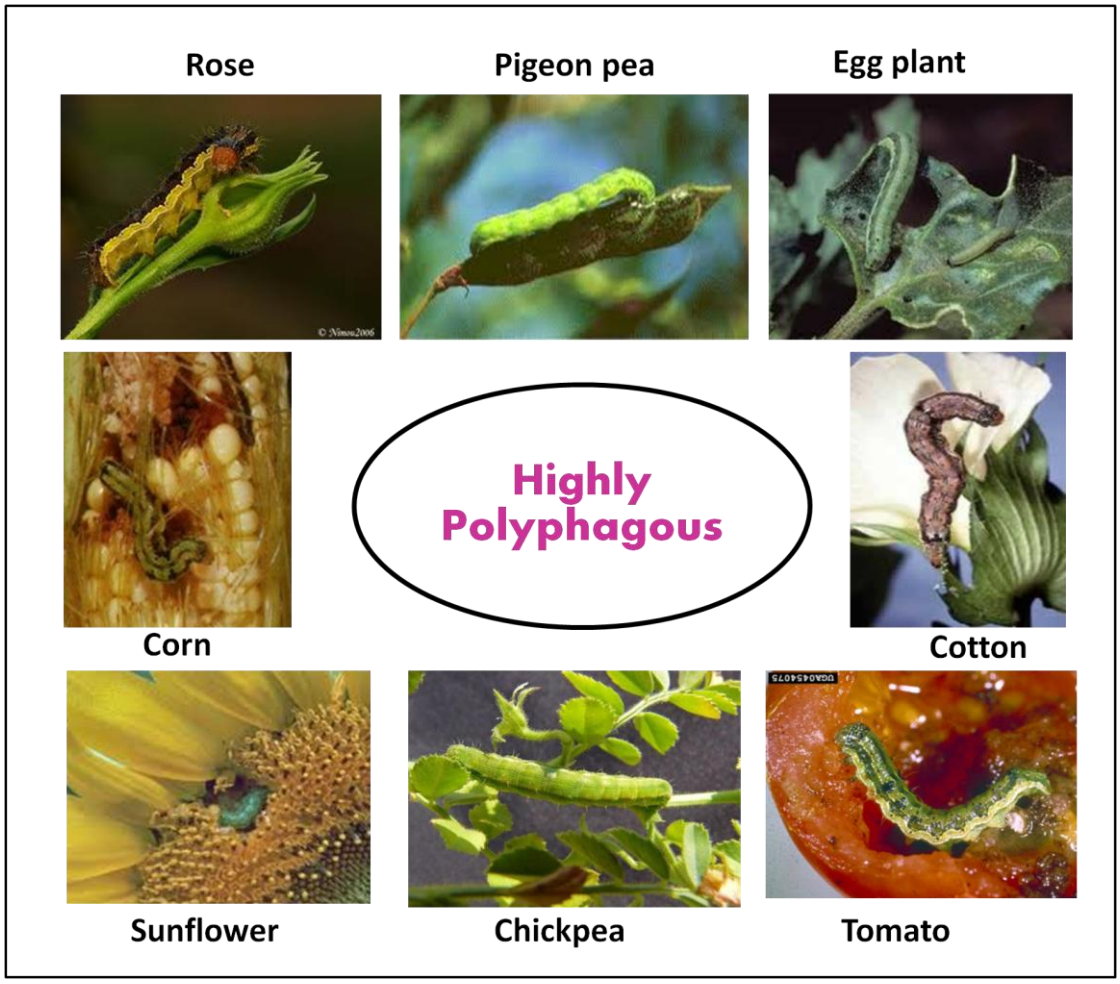


Figure 1.2: Polyphagous insect pest, *Helicoverpa armigera*

1.1.2 Gut protease diversity of *H. armigera*

The insect digestive system can be roughly divided in to fore-, mid-, and hind-gut. The dietary components are completely macerated by mandibles prior to its entering in the foregut. In the midgut it is subjected to enzymatic breakdown, and then

the products are assimilated. Hindgut shows very low digestive function, rather mostly involved in assimilation of the nutrients. The high gut pH and presence of free glycine are other peculiar characteristics of its guts (Johnston et al., 1991). Lepidopteran insects depend mostly on proteases for their digestive processes (Telang et al., 2000).

The digestive complement of *H. armigera* consists of endo-peptidases like serine, metallo-, cathepsin B like proteinases and exopeptidases. Serine proteinases, such as trypsin and chymotrypsin, form the dominant mechanistic class (>95%) in the gut environment (Johnston et al., 1991; Purcell et al., 1992; Harsulkar et al., 1999; Patankar et al., 2001). Gut protease activity increases during larval development with the highest activity seen in the fifth instar larvae, followed by a sharp decline in the sixth instar (**Fig. 1.3A**). The gut protease composition varies according to the developmental stage of the larvae and the dietary components (Patankar et al., 2001; Chougule et al., 2005; Sarate et al., 2012). Many of these proteinases have been isolated, identified and the coding DNA/cDNA have also been well characterized (Gatehouse et al., 1997; Bown et al., 1997, 1998; Mazumdar-Leighton et al., 2000; Patankar et al., 2001; Bayes et al., 2003; Chougule et al., 2005; Telang et al., 2005). The complex proteinaceous food material is broken down in to smaller oligopeptides by endo-peptidases like trypsins and chymotrypsins. These oligo-peptides are further digested by exo-peptidases liberating the free amino acids. Thus, the digestion in the larval gut follows a rational trend as a consequence streamlining the digestive process.

1.1.3 Polyphagy and differential protease gene expression

H. armigera feeds simultaneously on a number of host plants leading to high population build-ups. The insects' ability to survive on diverse host plants is an adaptive mechanism for their survival and propagation in the ecosystem (Tamhane et al., 2005a). Polyphagy requires metabolic adjustments in insects in order to maximize the benefits from protein-rich plant reproductive structures, carbohydrate-rich leaves, and even diverse unbalanced diets (Sarate et al., 2012). The total amount of protein, carbohydrate, and lipid in the diet directly influence the insect growth and development.

Digestive enzymes of *H. armigera* (amylases, proteases and lipases) display variable expression levels, regulated on the basis of macromolecular composition of the diet. Significant correlations between nutritional quality of the diet and larval and

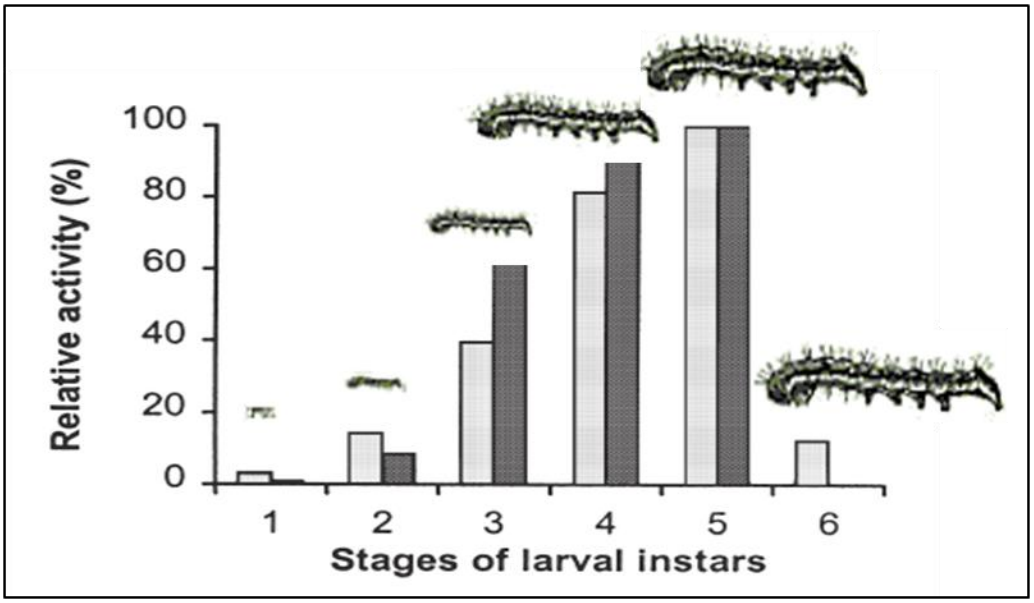


Figure 1.3A: Relative gut proteinase activity of *H. armigera* during the stages of larval development [Reproduced and modified from Patankar et al., 2001].

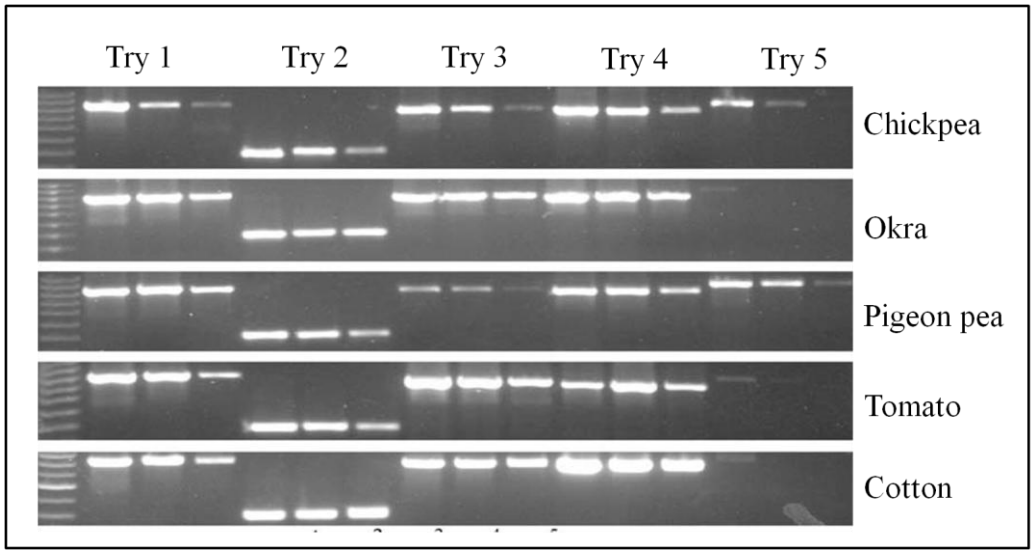


Figure 1.3B: Expression of trypsin-like proteinases of *H. armigera* larvae feeding on various host plants [Adopted from Chougule et al., 2005].

pupal mass have been observed on feeding *H. armigera* larvae on various host plants *viz.* legumes (chickpea and pigeon pea), vegetables (tomato and okra), flowers (rose and marigold), and cereals (sorghum and maize) (Fefelova and Frolov, 2008; Kotkar et al., 2009; Sarate et al., 2012).

The relative levels of expressed gut proteinase and amylase genes also vary between larvae feeding on various host plants (Chougule et al., 2005; Srinivasan et al., 2005; Kotkar et al., 2012) (**Fig. 1.3B**). This interesting feature has been attributed to differential expression of digestive proteinases in response to change in chemical properties of the dietary protein (Broadway et al., 1996; Harsulkar et al., 1999; Patankar et al., 2001). Such 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 proteinase genes (Srinivasan et al., 2006). The cDNA library prepared from the mid-guts of *H. armigera* reared on a high-protein diet has revealed 18 genes encoding trypsin-like proteinases, 14 genes of chymotrypsin-like proteinases and 2 genes of elastase-like proteinases (Gatehouse et al., 1997). Multigene families encoding variants of serine proteases in response to the dietary content have been identified in *H. armigera* (Bown et al., 1997). Variation in digestive complement pertaining to larval stage and diet explains well the polyphagous nature of *H. armigera* which allows it to infest a wide variety of agriculturally important crops simultaneously (Patankar et al., 2001).

1.2 Control of *H. armigera*

Owing to the extensive loss in crop yield due to *H. armigera*, several strategies used for its management include: cultural practices, chemical insecticides, biological control, host resistance, biotechnological approaches and integrated pest management.

Cultural practices include deep ploughing of soil, hand picking of large sized larvae, shaking of plants, weeding, intercropping, use of trap crops, following a good time of sowing (Dahiya et al., 1999) and fertilizer application. The cultural practices are eco-friendly and economical, however, they are mostly very laborious. Using chemical pesticides have been the main strategy for controlling pests since the last few decades. The main advantage of chemical pesticides is that they are effective even when used in an advanced stage of infestation but their continuous usage has led to the development of resistance against many chemical insecticides including organochlorides, organophosphates, carbamates, pyrethroids etc. (Mccaffery, 1998; Gunning et al., 1998). Moreover, chemical pesticides cause a severe threat to the environment by contaminating soil, water, vegetation and ecosystems and exert toxic

effects on the biome including human beings and non-target organisms. Biological control methods involve deploying the natural enemies of the pest *viz*. predators, parasitoids, biopesticides with live nematodes, fungi, bacteria, viruses or products derived from them and other plant products (Gurjar et al., 2011). Being environment friendly, they are recently becoming very popular. For example, entomopathogenic fungus such as *Beauveria bassiana* and *Metarhizium anisopliae*, are treated as natural enemies of *H. armigera.* Among all the bio-pesticides, Nuclear Polyhedrosis Virus (NPV) and *Bacillus thuringiensis* (Bt) have been very popular. Bt produces insecticidal crystal proteins which can lead to death of a wide range of insect orders and importantly these toxins could be specific to insect class (Lepidoptera, Diptera, Coleoptera, Hymenoptera, Homoptera, Orthoptera and Mallophaga) (Sanchis and Bourguet, 2008). Neem products, vegetable oils, crude plant oils are other popular plant products used for pest control with limited success.

Host plant resistance to insect pests relies on the availability of resistance genes in the germplasm and their transfer through conventional breeding procedures or modern biotechnological approaches to high yielding cultivars. Few wild relatives of chickpea (*C. bijugum, C. judaicum* and *C. reticulatum*) have been reported to be potential sources of resistance/tolerance to *H. armigera* which can be used to increase the levels of insect resistance in chickpea (Sharma et al., 2007).

One of the important approaches is the use of recombinant DNA technology to enhance insect resistance of plants by transfer of heterologous genes from various other sources (Ferry et al., 2006). Bt toxin has been the most commonly used gene for generating transgenic plants resistant to insect pests (Barton et al., 1987; Gupta et al., 2000). Commercial introduction of genetically modified maize, potato and cotton plants expressing *Bt* genes in the mid-1990s has been the most important landmark in crop improvement which revolutionized agriculture by increasing its productivity. However, modification of a Bt toxin receptor site in insects has led to the development of resistance against transgenic Bt plants (McGaughey et al., 1998; Frutos et al., 1999). In anticipation of insect adaptation to *Bt*, toxin attention was shifted to other bio-molecules also.

Inhibitors of amylases and proteinases of insects and lectins are few of the most studied plant defense molecules which exhibit potential to use them in insect

resistance strategies (Ryan, 1990; Hilder and Boulter, 1999; Schuler et al., 1998; Sharma et al., 2000; Lawrence and Koundal, 2002; Carlini and Grossi-de-Sa, 2002; Babu et al., 2003; Haq et al., 2004; Giri et al., 2005). Lectins disrupt the gut cell wall of insects by binding to glycoproteins of brush border cells and thus, the nutrient uptake is hampered. α-amylase inhibitors (α-AIs) occur in many plants as a part of the natural defense mechanisms. They complex with the insect gut amylases and need further research to identify potent AIs. Another attractive and alternate molecule is plant protease inhibitor (PI), for producing transgenic plants. PIs act as anti-metabolic proteins by blocking the digestive proteinases in the insect gut, thereby limiting release of amino acids from the food proteins (Hilder and Boulter, 1999). As a result, the growth and development of the insect population is arrested and finally its impact on the crop is reduced (Hilder et al., 1987; Ryan, 1990). 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 (Hilder et al., 1987 and 1993). Making use of combinations of a variety of defense molecules has risen as a futuristic approach to improve insect resistance in crop plants.

Recently several genomic approaches are underway in order to identify new functional genes or sites which 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*. The insect models for comparative genomics will help elucidate the function of genes and homologues. Development of post genomic tools like microarrays for functional genomics, mutant analysis, RNAi/miRNA based silencing (Terenius et al., 2011) of target genes has further broadened the basic and applied aspects of insect science.

This ongoing battle between plants and insects has lasted over million years. During their co-evolution with plants, insects have evolved for feeding and oviposition on many plant species using physical or chemical cues from host plants. Accordingly, plants have evolved intricate defense systems to resist insect herbivores (**Fig. 1.4**).

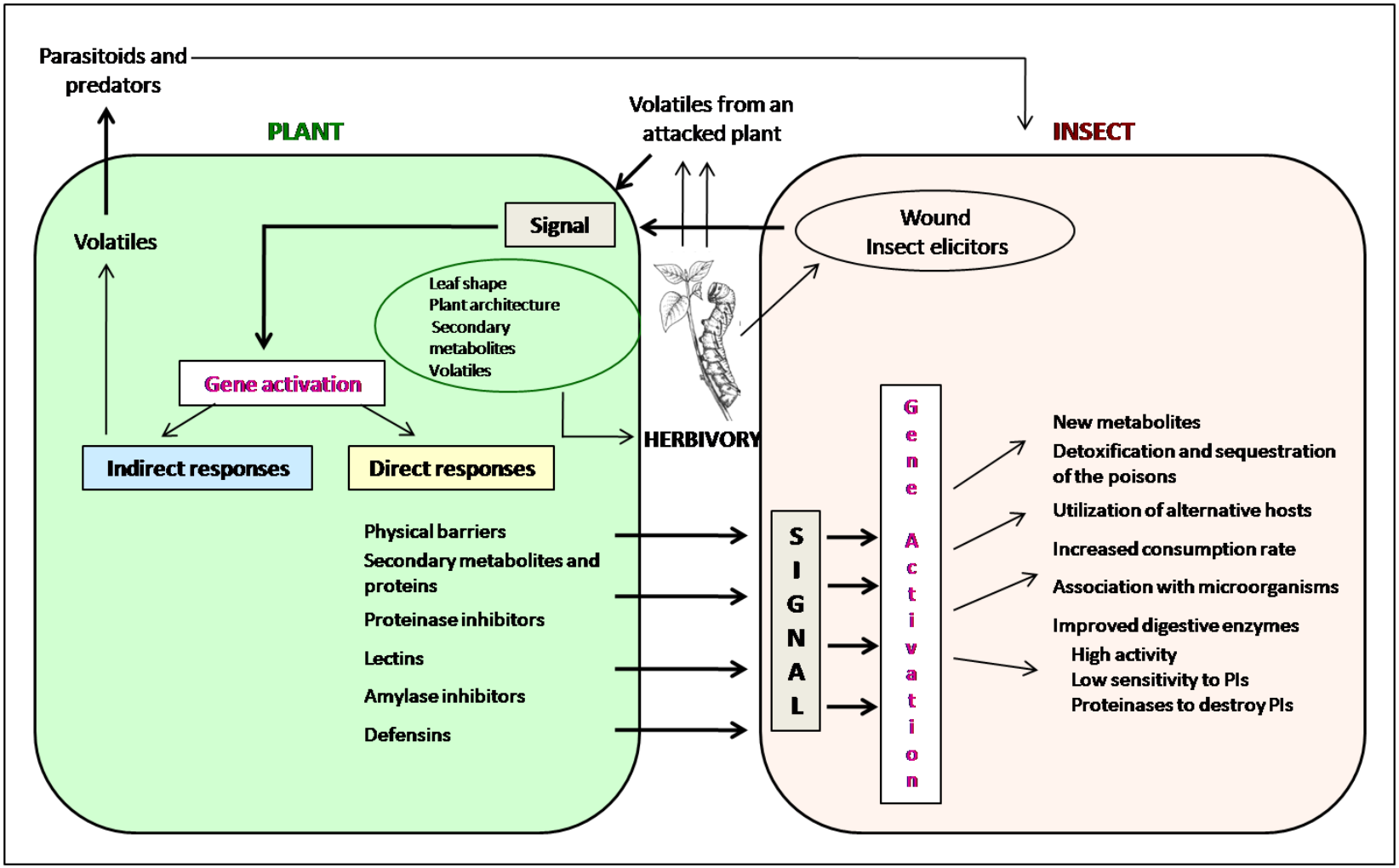


Figure 1.4: Plant-insect interaction: Co-evolution for survival [Modified from Mello et al., 2002].

1.3 Plant defense mechanisms

Plant defense against herbivory includes a range of adaptations evolved by plants in order to reduce the impact of herbivores and improve their survival and reproduction. Constitutive defenses are always present in the plant while induced defenses commence only after the attack by the herbivores. Plants use various direct or indirect strategies to defend themselves against herbivory. Some of these strategies include camouflage, mechanical and chemical defenses. Some plants mimic the presence of insect eggs on their leaves by developing physical structures. Some plants bear many external structural defenses that discourage herbivores; like trichomes on leaves and stems or lignin or silica deposition on the epidermal cells (Freeman et al., 2008). Many plants produce secondary metabolites and proteins that can influence herbivores' metabolism, growth and survival. Plants' indirect defenses-green leaf volatiles, volatile organic compounds, and extrafloral 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. Proteinase inhibitors are ubiquitously found in various plant tissues and also typically produced in response to herbivore attack or damage to the plant tissues.

1.4 Proteinase inhibitors as defensive proteins

PIs are naturally encountered in many plant species; their expression varies between various plant tissues including leaves, flowers and fruits/seeds. PIs are abundantly present in the storage tissues of plants and can represent up to 10% of total protein (Ussuf et al., 2001). Many PIs are produced in response to various biotic and abiotic stress conditions, e.g. pathogen invasion, insect attack, wounding, and environmental stress. The accumulation of trypsin and chymotrypsin-like proteinase inhibitors (PIs) throughout the aerial tissues of tomato and potato plants was demonstrated to be a direct consequence of insect-mediated damage or mechanical wounding (Green and Ryan, 1972). The earliest report on possible role of PIs in plant protection dates back to 1947 when, abnormal development of insect larvae was observed on soybean-based products (Mickel and Sytandish, 1947). Subsequently, the

toxic effects of trypsin inhibitors from soybean were shown on larvae of flour beetle, *Tribolium confusum* (Lipke et al., 1954). The direct evidence for inhibitory effect of PIs in plant leaves against insects was first demonstrated by Hilder et al (1987) by expressing a cowpea trypsin inhibitor gene in transgenic tobacco plants. The leaves produced 1% of the leaf protein as inhibitors and showed more resistance to feeding by larvae of *Heliothis virescens* than the untransformed control tobacco plants. Cowpea inhibitor was established as an anti-nutrient agent against *Heliothis, Spodoptera*, *Diabrotica* and *Tribolium*, all agronomically important insect pests. Following these studies, there has been many examples of antagonistic activity of PIs against insect pests, by *in vitro* assays against insect gut proteases as well as by *in vivo* (insect feeding) assay (Broadway and Duffy, 1986a and 1986b; Johnston et al., 1993; Harsulkar et al., 1999; Bown et al., 2004; Giri et al., 2005; Tamhane et al., 2007). Firm establishment of PIs as antagonists have led the way to transgenic expression of PIs in various crop plants rendering them with higher resistance to pests, as detailed in **Table 1.1**. PIs may adversely affect the proteinases of phytopathogenic microorganisms also. Plant derived trypsin and chymotrypsin inhibitors were found to suppress the activity of proteinases excreted by a fungus *Fusarium solani* and also the growth of hyphae and conidium germination in *F. culmorum*, and *Botrytis cinerea* (Mosolov et al., 1976; Valueva et al., 2004).

Most of the PIs are inhibitors of serine proteases like trypsin or chymotrypsin, although some of them may exhibit inhibitory activity against elastases, subtilisin-like proteases etc. They do not appear to function as inhibitors of endogenous plant proteinases as most of the latter ones are cysteine proteinases (Reeck et al., 1997). Thus, it is widely acknowledged that serine PIs have a defensive role against herbivores (especially, Lepidopterans) which mainly rely on serine proteinases for digesting the dietary protein components. PIs act as substrate mimics and block the digestive proteinases in the larval gut thereby limiting the release of amino acids from food proteins (Broadway and Duffey, 1986a and 1986b; Hilder and Boulter, 1999).

Table 1.1: Transgenic plants harbouring PI genes, with increased resistance to pests

Plant species	PI	Target pest
		Insects:
Tobacco (*Nicotiana tabacum* L.)	CpTI	*Heliothis virescens*
	PI-II	*Manduca sexta* (L.)
		Chrysodeixis eriosoma
	SpTI	*Spodoptera litura*
	MTI-2	*Spodoptera littoralis*
		Mamestra brassicae (L.)
	NaPI	*Helicoverpa punctigera*
		Helicoverpa armigera (Hübner)
Potato (*Solanum tuberosum* L.)	CpTI	*Lecanobia oleracea* (L.)
	OC-I	*Leptinotarsa decemlineata*
Arabidopsis (*Arabidopsis thaliana* L.)	MTI-2	*Spodoptera littoralis*
		Plutella xylostella (L.)
		Mamestra brassicae (L.)
Rice (*Oryza sativa* L.)	CpTI	*Chilo suppressalis*
		Sesamia inferens
	PI-II	*Chilo suppressalis*
		Sesamia inferens
Rape (*Brassia napus* L.)	MTI-2	*Spodoptera littoralis*
	OC-I	*Myzus persicae*
Pea (*Pisum sativum* L.)	NaPI	*Helicoverpa armigera* (Hübner)
		Nematodes:
Arabidopsis (*Arabidopsis thaliana* L.)	OC-Im	*Heterodera schachtii*
		Rotylenchulus reniformis
Banana (*Musa cavendishii* L.)	OC-Im	*Radopholus similis*
		Others:
Arabidopsis (*Arabidopsis thaliana* L.)	OC-Im	*Deroceras reticulatum*
Tobacco (*Nicotiana tabacum* L.)	NaPI	*Botrytis cineria*
		Pseudomonas solanacearum

CpTI, cowpea trypsin inhibitor; PI-II, trypsin and chymotrypsin inhibitor from potato; SpTI, sweet potato trypsin inhibitor; MTI-2, white mustard (*Sinapis alba* L.) trypsin inhibitor; NaPI, serine proteinase inhibitor from *Nicotiana alata* (L.); OC-I, oryzacystatin I; OC-Im, modified oryzacystatin I.

[Adapted from Mosolov et al., 2008]

As a consequence, the depletion of amino acids exerts a profoundly detrimental effect on larval physiology and thereby retards the growth and development of the larvae (Broadway and Duffy, 1986; De Leo et al., 2001; Telang et al., 2003; Damle et al., 2005; Tamhane et al., 2007; Hartl et al., 2010). Further, the decreased fertility and fecundity of the adult moths reduce the overall fitness of the insect populations and thus reduce their impact on crop. Additionally, antagonistic effects of PIs act synergistically with other components of the plant defense mechanism viz. retarded insects become easier targets for their parasites (Lewis et al., 1997).

1.4.1 Protease inhibitor families in plants

PIs have been classified based on their specificity against four mechanistic classes of proteolytic enzymes, i.e. serine, cysteine, aspartic and metallo-protease (Ryan, 1990) (**Fig. 1.5**). Further, each type includes different inhibitor families which have been classified based on their molecular mass, sequence homology, structural characteristics and expression patterns (Garcia-Olmedo et al., 1987; Laskowski et al., 1980; Ryan, 1990). In plants, ten protease-inhibitor families have been recognized. Most inhibitor proteins in these families exhibit molecular mass from 5 to 25 kDa. Many of them are products of multigene families and several isoinhibitors have been found in a single species exhibiting different specificities towards proteases (Wu et al., 2006; Tamhane et al., 2009). Serine proteinase inhibitors are further classified in to various families such as Kunitz, Bowman-Birk, Wound-inducible (potato proteinase inhibitor [PIN] type I and II, Squash, Cereal trypsin /α-amylase inhibitor and Mustard seed trypsin inhibitor. Members of serine and cysteine proteinase inhibitor families have been more relevant to the area of plant defense because of their wide presence across many plants while metallo- and aspartic have limited occurrence (Ryan, 1990).

1.4.2 Serine PIs and their standard mechanism of inhibition

Serine PIs are the most broadly distributed family of protease inhibitors and they all employ the same competitive mechanism of inhibition. They have a rigid, reactive-site loop that is complementary to the substrate binding site of serine proteinases and thus they bind in a substrate like manner (Laskowski et al., 1980; Bode et al., 1992; Krowarsch et al., 2003). Tight binding to the enzymes is achieved by maintaining a conformationally stable reactive-site loop (RSL) which protrudes from the protein scaffold and serves as a recognition motif. The P1 residue of the RSL

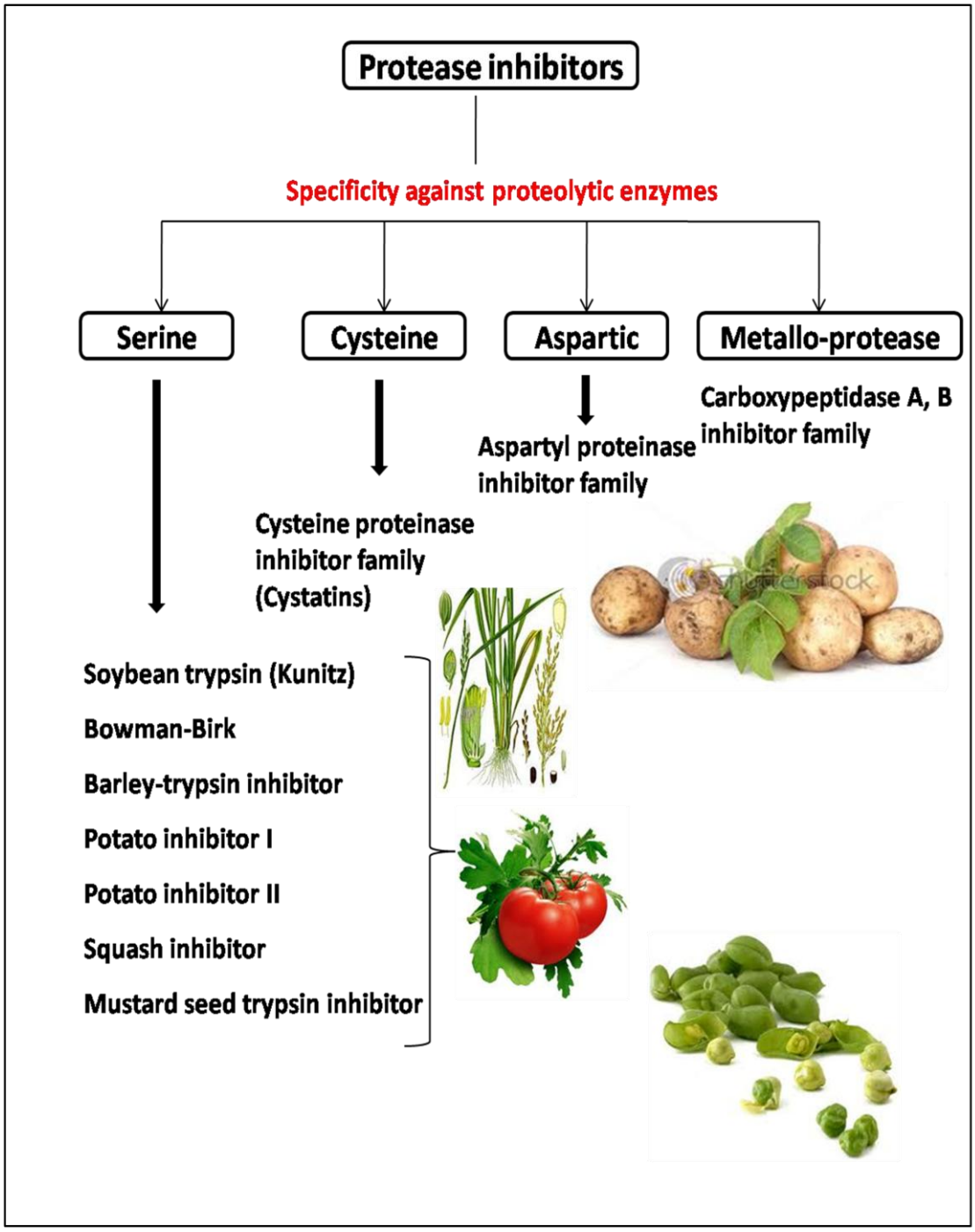


Figure 1.5: Protease inhibitor families in plants.

makes the most contacts with the proteinase in the S1 pocket and the primary determinant of the inhibitory specificity (Laskowski et al., 1987) (**Fig. 1.6A**). Lys or Arg for trypsin–like enzymes, Phe, Tyr or Leu for chymotrypsin and Ala for elastase are usually present at the P1 position of the inhibitor. The protease recognizes the specific peptide bond P1-P1'in the RSL and the two proteins are frozen in to a stable complex in which the enzyme cannot complete the hydrolysis of the peptide bond, nor can the complex easily dissociate. The strength of the protease–PI interaction is determined by the compatibility of all the amino acid residues (P4–P4') which also serve to direct the inhibitor towards the active site cleft of the serine proteinases (Bateman et al., 2011). Tight binding and slow hydrolysis are the peculiar features which are result of stable hydrogen bonding that surround the scissile peptide bond and the Van der Waals contacts at the interface. Disulfide bonds are abundant in many of the proteinase inhibitor families and they provide stability to the exposed RSL by holding it through covalent attachments. These interactions restrict any distortion of the P1-P1' peptide bond and increase the activation-energy barrier for hydrolysis. The inhibitory loops of serine PIs have a characteristic conformation, defined by the torsion angles of the P3–P3' segment, irrespective of the family they belong to, while the remaining part of the molecule, known as scaffold, has widely different folds in different families of inhibitors (**Fig. 1.6B**).

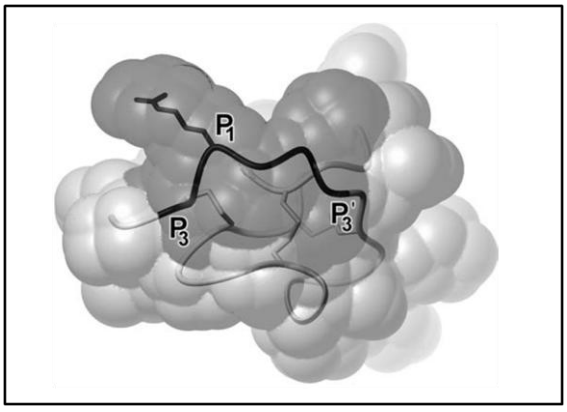


Figure 1.6A: The solvent-exposed protease binding loop (residues P3-P3') of serine PIs.

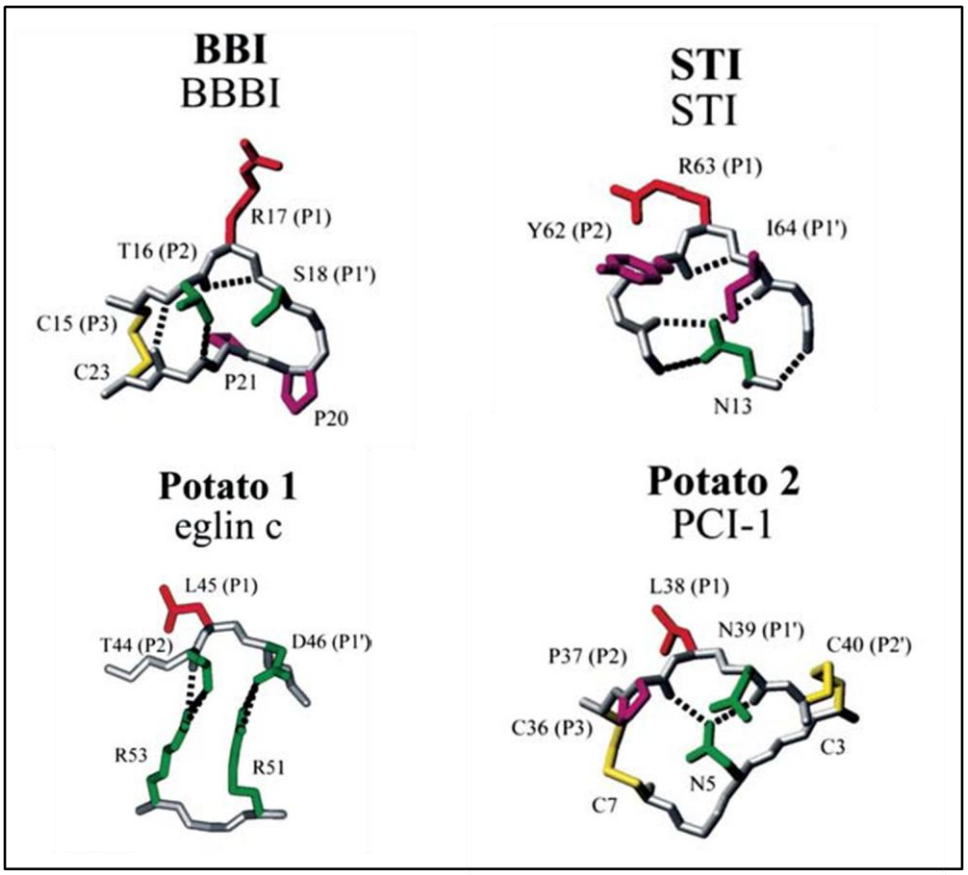


Figure 1.6B: Binding loop structures of representatives of inhibitor families: BBBI, barley Bowman-Birk inhibitor; STI, soybean trypsin inhibitor; Potato 1 and Potato 2 trypsin inhibitor [Adopted from Krowarsch et al., 2003].

1.5 Effect of dietary PIs on insects

Although many plant species express PIs, including host plants, many lepidopteran pests feed and thrive well on them. An answer to this has been attributed to the adaptation of insect pests to host plant PIs by synthesizing proteinases that are either insensitive to inhibitors (Broadway, 1995, 1996, 1997; Jongsma et al., 1995) or over-expression of proteinases (Broadway, 1997; Gatehouse et al., 1997) or have the capacity to degrade them (Michaud, 1997; Giri et al., 1998; Harsulkar et al., 1999). Host plants might possess effective PIs, but their effects could be undermined due to low level of expression (Chougule et al., 2003; Damle et al., 2005; Srinivasan et al., 2005). It can be overcome by expressing the host-PI genes under a stronger promoter to increase the inhibitor protein content. On the other hand, 'non-host' plants PIs have

been found to be effective against insects and responsible for rendering them unsuitable as a food for the insects. *H. armigera* larvae reared on a diet containing non-host PIs (groundnut, winged bean and potato) showed a reduction in total and trypsin-like proteinase activity, and the production of inhibitor-insensitive proteinases (Harsulkar et al., 1999; Giri et al., 2003). Non-host PIs added to larval diet also inhibited growth of *H. armigera* and *Spodoptera litura* larvae (Harsulkar et al., 1999; Telang et al., 2003; Tamhane et al., 2005b; 2007). Thus, stability of PIs to proteolytic degradation and the synergistic interaction of different PIs can drastically influence the efficacy of PIs.

The success or failure of the PI strategy can largely depend on the nature of the insect response to the ingested PIs. When incorporated into insect diets or expressed in transgenic plants and fed to insects, the efficacy of protease inhibitors has proved to be variable. High levels of growth retardation and mortality have been observed in some studies, whereas very minor or no deleterious effects seen even when the inhibitor has been shown to be effective in preventing proteolysis, *in vitro* (Jongsma and Bolter, 1997). A possible explanation for this has been stated as; certain insect species have evolved the capacity to adapt to protease inhibitors present in the diet (Broadway, 1995; Jongsma et al., 1995). *S. exigua* larvae fed on detached leaves of tobacco plants transformed with trypsin/chymotrypsin-specific potato 2 inhibitor (PI2) under the control of constitutive promoter remained unaffected and showed 2.5 fold induction of new tryptic gut activity that was insensitive to inhibition by PI2 (Jongsma et al., 1995). PI gene, from *Solanum americanum* (SaPIN2a) was expressed under the control of the CaMV 35S promoter by Luo et al. (2009). Bioassays for insect resistance showed that transgenic tobacco plants over expressing SaPIN2a were more resistant to *H. armigera* and *S. litura* larvae than the control plants. They also reported an increase in the glandular trichome density along with the promotion of trichome branching in such transgenic tobacco plants. Transgenic cotton plants expressing the *N. alata* proteinase inhibitor (NaPI) which is a Pin-II type inhibitor and/or Pin-I inhibitor isolated from *S. tuberosum* (StPin1A), subjected to *H. armigera* infestation recorded a "statistically significant", increase in number of cotton bolls than the parental line (untransformed) (Dunse et al., 2010a). Transgenic cotton plants, expressing both the types of inhibitors (NaPI-StPin1A) showed better performance than the lines expressing only the individual inhibitors. Combined leaf-specific over

expression of potato PI-II and carboxypeptidase inhibitors in transgenic tomato resulted in increased resistance to *Heliothis obsolete* and *Liriomyza trifolii* larvae (Abdeen et al., 2005). These studies have indicated that the combined expression of defense genes with different modes of action and combination of inhibitors might be more effective for insect control and stable resistance against pests (Dunse et al., 2010a).

1.5.1 Adaptability: Structural and functional diversities in proteases

Lepidopteran insects have been found to employ usually one of the mechanisms for overcoming the effects of dietary PIs: generalized over expression of digestive proteases so as to compensate for the loss of activity due to inhibitor binding (Broadway and Duffey, 1986b; Broadway, 1997; Gatehouse et al., 1997; Chougule et al., 2005) or over-expression of those proteinases which can bind and degrade the ingested PIs (Giri et al., 1998). Studies have indicated biphasic response by insects, characterized by an initial upregulation of all digestive proteases, preceding a simultaneous down-regulation of PI-sensitive proteases and *de novo* synthesis and up-regulation of PI-insensitive proteases (Jongsma et al., 1995; Broadway, 1996; Bown et al., 1997; Mazumdar-Leighton and Broadway, 2001a, 2001b; Volpicella et al., 2003). Reports also indicate a switch to enzymes within the same mechanistic class but of different substrate specificity, for example, chymotrypsin-like activity rather than trypsin-like (Gatehouse et al., 1997; Wu et al., 1997; Dunse et al., 2010b). However, the exact mechanism by which the insect senses the presence and nature of PIs and adjusts the proteinase gene expression accordingly, remains unexplored yet certain neuropeptides have been reported to regulate these switches (Harshini et al., 2002a and b; Bown et al., 2004).

The presence of multiple protease isoforms has been traced back to multi-copy protease genes in the insect genome, that probably have arisen due to gene duplication and diversification events (Barrett et al., 1995; Lopes et al., 2004). The activity of serine proteases is a function of the well conserved “catalytic triad” which results from the spatial proximity of histidine (57), aspartate (102) and serine (195) residues in the polypeptide (Polgar et al., 2005). However, the specificity of the serine proteases is not governed by the catalytic triad, but is rather due to a molecular (S1) pocket which interacts with the side chains of the amino acids lying in the cleavage

(P1-P1') site of the substrate. Variability in the amino acids in the S1 pocket may be linked to functional diversity, as they identify and interact with various substrate polypeptides. Analysis of amino acid variations across representatives of lepidopteran trypsins and chymotrypsins, have illustrated characteristically well conserved S1 pocket in trypsins, unlike chymotrypsins which show much variability (Srinivasan et al., 2006). Thus, it is hypothesized that chymotrypsins might exhibit greater flexibility in substrate recognition and in activity as well. Variations in the key recognition motifs as well as the adventitious contacts may determine the (in)-sensitivity to PIs. A recent study has identified chymotrypsins in *Helicoverpa* larvae which are resistant to PI from *N. alata* (NaPI). Amino acid mutations at selected positions and molecular modelling studies revealed the molecular mechanism of resistance to NaPI (Dunse et al., 2010b). A single Phe→Leu substitution at position 37 in the chymotrypsin resulted in the loss of important binding contacts with NaPI and thus making it resistant to NaPI.

The complete genome sequence project of *H. armigera* (http://insectacentral.org/helicoverpa) would be an invaluable asset to identify varied proteinase genes and their regulation and thus, designing the PI molecules. Such studies provoke an interesting chain of future investigations for understanding how insects develop resistance to PIs which should be the rationale for design of PIs targeted against insect pests. Appropriate PI(s) need to be selected or modified from the available PI pool to inhibit both, constitutively expressed as well as induced proteinases.

In this co-evolutionary race of plant-insect interface, plants have also evolved elaborate defense mechanisms to resist the insect herbivores. Induction of defensive traits in plants helps them to defend against the insect attacks and also prepares them well in advance for future attacks.

1.6 Induction of plant defenses

Plants have evolved inducible defense mechanisms that are activated or amplified in response to attack or damage. Particular signals can prime the plant tissues, that is prepare them for an improved response to future attack. Plants can perceive various signals which originate from the nature of damage. Various pathogen-associated molecular patterns (PAMP) are recognized by specific receptors

and R-gene mediated defense system initiates the hypersensitive response (Wu and Baldwin, 2010). Similarly, herbivore derived elicitors and the specific chewing patterns function as herbivore associated molecular patterns (HAMPs) (Mithofer et al., 2008). Some of these elicitors are fatty acid amino acid conjugates (FACs), volicitin, glucose oxidase, inceptins, caeliferins, etc. (Walling, 2000; Baldwin et al., 2001; Wu and Baldwin, 2010). Elicitors trigger the generation of the signaling intermediates which further amplify the defense cascade throughout the plant. Although mechanical wounding and herbivory, both result in tissue damage and loss, studies have demonstrated dramatic differences in their plant responses (Korth and Dixon, 1997a and 1997b; Baldwin et al., 2001; Halitschke et al., 2001). Most of the induced responses occur in a timeframe of few minutes to several hours of wounding and include generation, release and spread of specific signals for coherent activation of appropriate defense genes.

An important feature of many induced direct defense responses is their occurrence in undamaged tissues located far from the site of damage. This has implied the existence of specific mechanisms in plants to generate, transport, and perceive alarm signals that are generated at the site of tissue damage. Systemically induced responses can be achieved by two mechanisms: first, by *de novo* expression of defensive traits that are activated by translocated signals from the attacked tissues and second, by systemic transport of defensive metabolites. Wound-inducible PIs in several Solanaceous plants have been widely used as a model system to study the molecular mechanism of local and systemic wound signaling. Green and Ryan (1972) proposed that chemical signals produced at the wound site travel through the plant and activate PI expression in undamaged leaves. In tomato plants, *PI* genes showed expression in distal leaves within 1–2 hrs after insect attack or mechanical wounding (Ryan, 2000; Strassner et al., 2002). Several studies have led to the identification of many phytochemicals governing the regulation of induced defenses: cell-wall-derived oligogalacturonides (OGAs), systemin, jasmonic acid (JA), salicylic acid (SA) and its derivatives, hydrogen peroxide, ethylene and abscisic acid (Ryan, 2000; Gatehouse, 2002). These hormones activate various defense genes through separate signal transduction pathways, depending on the type of damage/ infection/ infestation (Karban and Baldwin, 1997). Physical signals (e.g., hydraulic forces and electrical signals) generated by tissue damage have also been shown taking part in the systemic

signaling process (Wildon et al., 1992; Malone, 1996). The cross–talk between various defense signaling pathways provides the plant with an elaborate regulatory potential that leads to the activation of the most appropriate defense against the invading encountered (Pieterse et al., 2001; Van Wees et al., 2000). For example, the interactions between jasmonate and salicylate pathway are known to strongly influence the defense response of the plant (Thaler et al., 2004). However, negative cross-talk between these two pathways is suggested (Pena-Cortes et al., 1993; Felton et al., 1999). Plants infected with pathogens decrease their ability to induce JA while promote the production of SA. The vice-versa has been observed in case of insect pest infestations.

1.6.1 Systemin signaling and activation of Jasmonate pathway

Extracellular signals such as OGAs and systemin generated in response to tissue damage trigger the intracellular production of JA via the octadecanoid pathway, which in turn activates the expression of defensive genes (Farmer and Ryan, 1992). Wound-induced production of OGAs is catalyzed by polygalacturonases (PGs) and is thought to act as local mediator of the signal transduction (Bergey et al., 1999; Ryan, 2000; Navazio et al., 2002). Systemin was the first bioactive peptide discovered in plants and its key role in induced defense responses was indicated in several studies (Pearce et al., 1991; Howe and Ryan, 1999; Howe, 2004). This 18-amino-acid peptide is derived from proteolytic cleavage of a larger precursor protein (200 aa) Prosystemin. Hydroxyproline rich glycopeptides (20 aa) similar to the activity of systemin peptides have also been isolated from tomato leaves (Pearce et al., 2003; 2009). The systemin signaling pathway is initiated upon binding of the peptide to a 160-kDa plasma membrane-bound receptor (SR160), a member of leucine-rich repeat (LRR) receptor-like kinase family of proteins. Binding of systemin to the cell surface triggers several rapid signaling events, including increased cytosolic Ca^{2+} levels, membrane depolarization, and activation of a MAP kinase cascade (Stratmann and Ryan, 1997; Moyen et al., 1998; Wu and Baldwin, 2010) (**Fig. 1.7)**. Systemin-regulated phospholipase A2 activity releases free linolenic acid (LA), a JA precursor, from lipids in the plasma membrane (Farmer and Ryan, 1992; Narv´aez-V´asquez et al., 1999). After a series of reactions taking place in chloroplast catalysed by, lipoxygenase (LOX), allene oxide synthase (AOS), and allene oxide cyclase (AOC), LA is further converted to 12-oxo-phytodienoic acid (OPDA), which gets transported

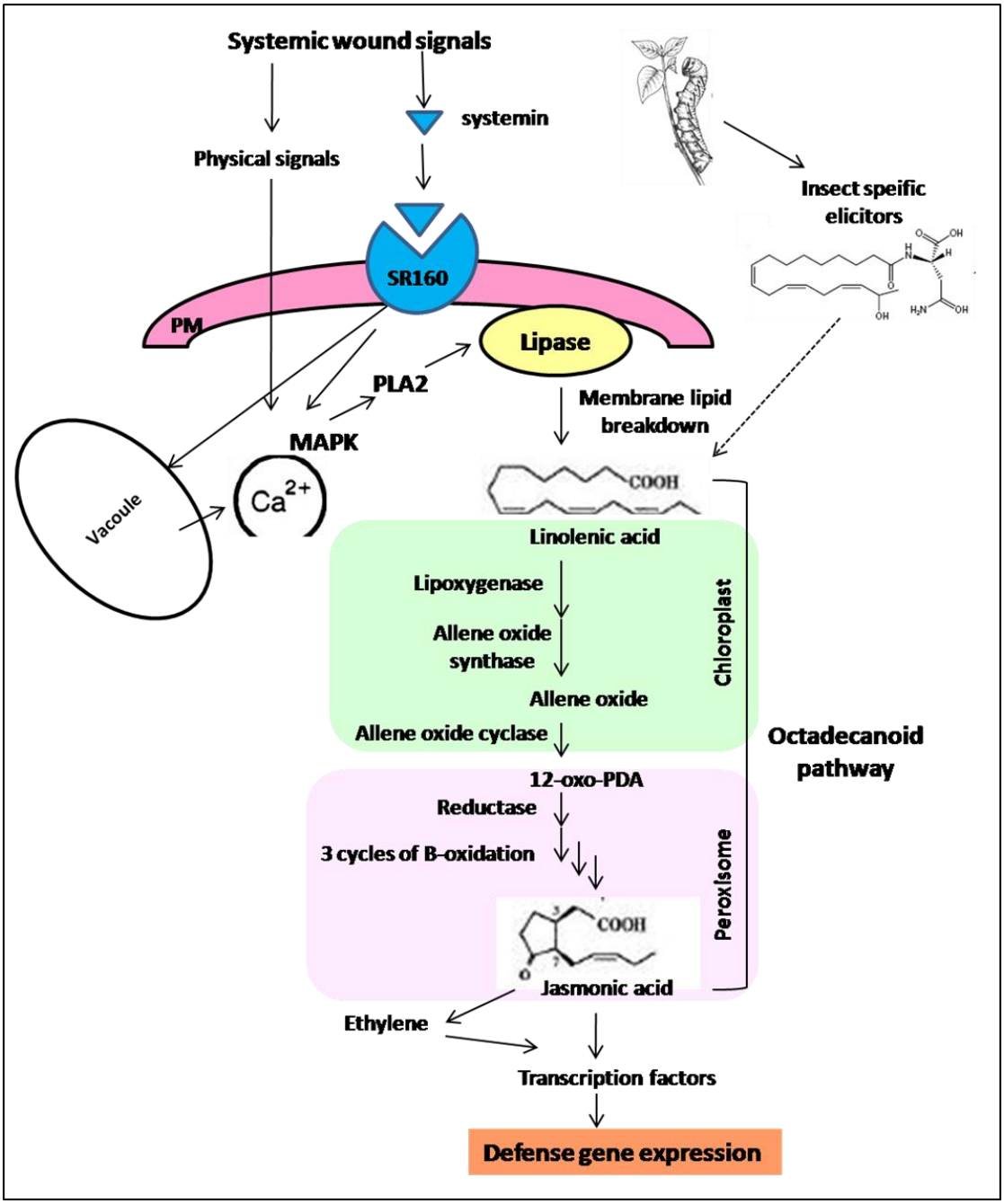


Figure 1.7: Systemin signalling and activation of jasmonate pathway. Intracellular wound signal transduction pathway leading to the induction of defense gene expression.

to peroxisomes. After three steps of β-oxidation, JA is formed (Wasternack, 2007; Wu and Baldwin, 2010). JA-responsive transcription factors like WRKY and several others finally activate the expression of defense genes (Skibbe et al., 2008; Koo et al., 2009). JA acts in accordance to ethylene and hydrogen peroxide to positively regulate the expression of downstream genes (O' Donnell et al., 1996; Howe and Schaller,

2008). Role of JA in major reprogramming of gene expression on wounding/ insect attack is well demonstrated (Reymond et al., 2000). JA also forms conjugates with other amino acids (Leu, Val, and Phe). MeJA and certain jasmonoyl-amino acid conjugates (e.g., JA-Ile) are potent elicitors of defense gene expression (Wasternack et al., 1998). Transgenic tomato plants expressing an antisense *prosystemin* (*Prosys*) cDNA were deficient in wound-induced systemic expression of PIs and as a consequence, were more susceptible to insect herbivores (McGurl et al., 1992; Orozco-C´ardenas et al., 1993). On the contrary, over-expression of prosystemin under a constitutive promoter activated PI expression even in the absence of wounding (McGurl et al., 1994; Li et al., 2002; Chen et al., 2005). This clearly indicated the role of systemin in induction of the plant defenses.

The plant vascular system is involved in long-distance trafficking of a wide range of signaling compounds (Lucas and Lee, 2004). Systemin travels through the phloem to the distal parts of the plant, triggering release of more systemin from prosystemin locally and systemically and in turn, triggers the JA pathway. Studies have also provided direct evidence for the presence and transportation of jasmonates and JA biosynthetic enzymes in phloem (Hause et al., 2003; Stenzel et al., 2003; Wasternack, 2007). Thus, jasmonic acid or its derivatives may act as long-distance transmissible signal for wound responses in plants (Ryan and Moura, 2002; Sun et al., 2011). In *Nicotiana* plants, unknown long-distance mobile signal travels to roots and induces defenses, e.g., nicotine biosynthesis (Wu and Baldwin, 2010). Wound induced accumulation of PI proteins was observed in pepper leaves (Moura and Ryan, 2001). Exposing intact plants to methyl jasmonate vapours induced the accumulation of PI proteins.

Local and systemic expression of a variety of defense molecules or enzymes occurs, which can protect the plant from the invaders and includes Proteinase inhibitors (PI), Polyphenol oxidase (PPO), Peroxidase, Phenylalanine ammonia lyase (PAL), Cathepsin D, Arginase, Threonine deaminase and Leucine aminopeptidase. Out of these inducible defense molecules our laboratory has focused on the study of PI mediated plant defenses and the research work embedded in the present thesis is mainly on the wound-inducible potato inhibitor II (Pin-II) family of PIs.

1.7 Pin-II family of Proteinase inhibitors

PIs of the potato inhibitor II family (PIN2/Pin-II) are the best characterized plant serine PIs with respect to their molecular properties (Bryant et al., 1976; Xu et al., 2001; Xu et al., 2004; Sin and Chye, 2004) and have been interestingly explored at gene, protein and functional level. Pin-IIs have been found in various tissues, e.g. tubers (Bryant et al., 1976), fruits (Richardson, 1979; Damle et al., 2005; Tamhane et al., 2007), wounded leaves (Pearce et al., 1993), and flowers (Sin and Chye, 2004). These Pin-II PIs have a unique single or multi domain repeat structure with variations, also a wound/insect infestation induced up-regulation and expression. All these features have made them a fascinating subject area to study plant-insect interactions as well as to understand the plants' physiology. The focus is on potato inhibitor-II (Pin-II) family of serine PIs which displays a striking genetic and molecular diversity and a significant plant defense related role.

1.7.1 Occurrence

Pin-II PIs are predominantly found in Solanaceous plants and their wound induced up regulation has established their defense related role (Green and Ryan, 1972). Identification of many Pin-II homologs dispersed throughout the whole range of mono- and dicotyledonous plants has indicated more widespread occurrence of this family (Barta et al., 2002). The striking feature of Pin-II PIs is the presence of variable number of inhibitory repeat domains (IRDs), structurally forming multi-domain proteins. The first two well characterized Pin-II PIs from tomato (Graham et al., 1985) and potato (Sanchez-Serrano et al., 1986) had two domains with trypsin and chymotrypsin reactive sites. Since then, PIs containing even up to 8 domains have been reported. Pin-II PIs are encoded by one, two or multi-gene families in *Nicotiana*, *Solanum* and *Capsicum* species, respectively. In *Nicotiana* most of the members show presence of single functional genomic copy of Pin-II PIs, but varying number of IRDs (2 to 8). The numbers of IRDs in various *Nicotiana* species have shown no correlation to the phylogenetic associations, since plants with close phylogenetic relationships had different repeat numbers (Wu et al., 2006). The evolutionary advantages of repetitive IRDs have not been established, but it has been predicted that the repetitive domains provide plants with a more efficient pool of inhibitory units which might play

varied functions. Phylogenetic relationships of Pin-II PIs from various genera of Solanaceae are represented in **Fig. 1.8**.

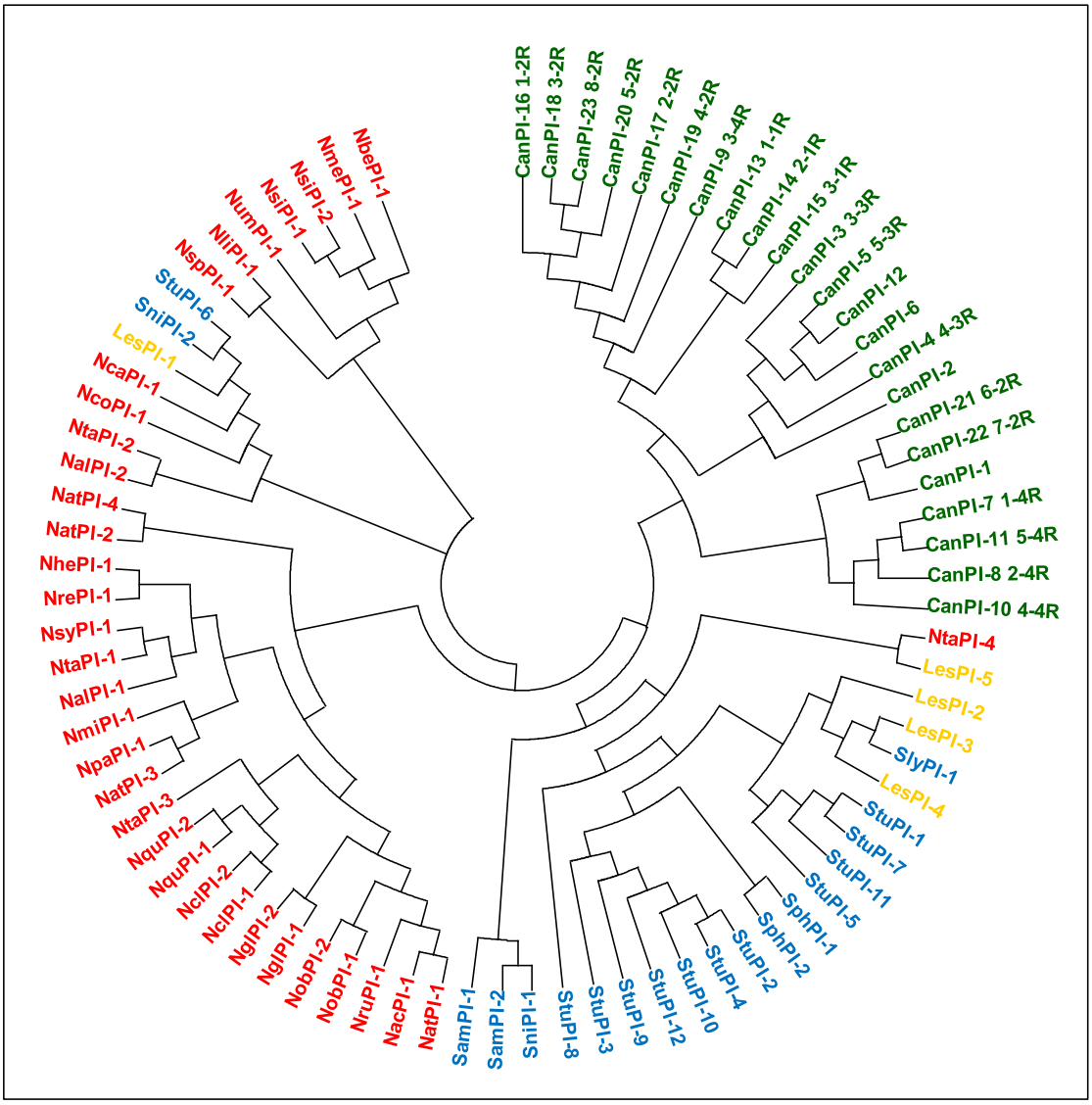


Figure 1.8: Phylogenetic tree of Pin-II PIs from Solanaceae. Genera *Nicotiana* (red colour), *Solanum* (blue colour), *Capsicum* (green colour) and *Lycopersicum* (yellow colour) are shown. The sequences were taken from NCBI and the accession numbers are provided in Appendix I.

Several Pin-II PIs having 1- to 4-IRDs (*CanPIs*) have been isolated from developing fruit and stem tissues of *C. annuum* (Tamhane et al., 2009). Significantly higher level of PI activity as compared to the leaf, stem and fruit tissues have been found in flowers (Damle et al., 2005; Tamhane et al., 2009). Induced up regulation of Pin-II PIs has been noted in various Solanaceae plants like *Nicotiana* sp., *C. annuum*, *Solanum* sp. in response to insect attack, wounding, systemin, methyl jasmonate,

polyethylene glycol, salt, abscisic acid, cold stress and application of electric current (PeñaCortés et al., 1989; Kim et al., 2001; Moura and Ryan, 2001; Tamhane et al., 2009). A 3-IRD Pin-II transcript induced by auxin in tomato roots was characterized by Taylor et al. (1993).

1.7.2 Structure of Pin-II PIs

Pin-II family remains one of the best characterized families of PIs not only for its wound inducibility and defense potential but also with respect to the gene as well as protein structure. Several studies have highlighted its peculiar structural characteristics and are outlined below:

1.7.2.1 Gene structure of Pin-II PIs: accelerated evolution of repeats with variations

The architecture of the Pin-II genes is conserved. The first exon encoding the N-terminus of the signal peptide and the second main exon encoding the C-terminus of the signal peptide followed by variable number of IRDs, are always separated by an intron of 100-200 bp (Barta et al., 2002) (**Fig. 1.9 inset**). The splicing motif is conserved and found to be GT...AG (Kong and Ranganathan, 2008). The Gly residue formed by the boundaries of the two exons in the signal peptide is a conserved feature for Pin- II family PIs. A genomic clone of Pin-II PI gene in tomato showed an upstream regulatory region (~1kb) containing wound and hormone responsive elements similar to box-WUN-motif and transcription region containing a single intron (Zhang et al., 2004).

Presence of variable number of repeat regions has suggested that duplication of Pin-II genes has occurred several times, especially in Solanaceae members. Single repeat genes are also found outside Solanaceae whereas 2 to 3 and 4 to 8 repeat unit PIs are exclusively present in Solanaceae. Single repeat Pin-II PIs are thought to be the ancestral members that have given rise to the other forms by a series of gene duplication and domain duplication events (Barta et al., 2002) (**Fig. 1.9**).

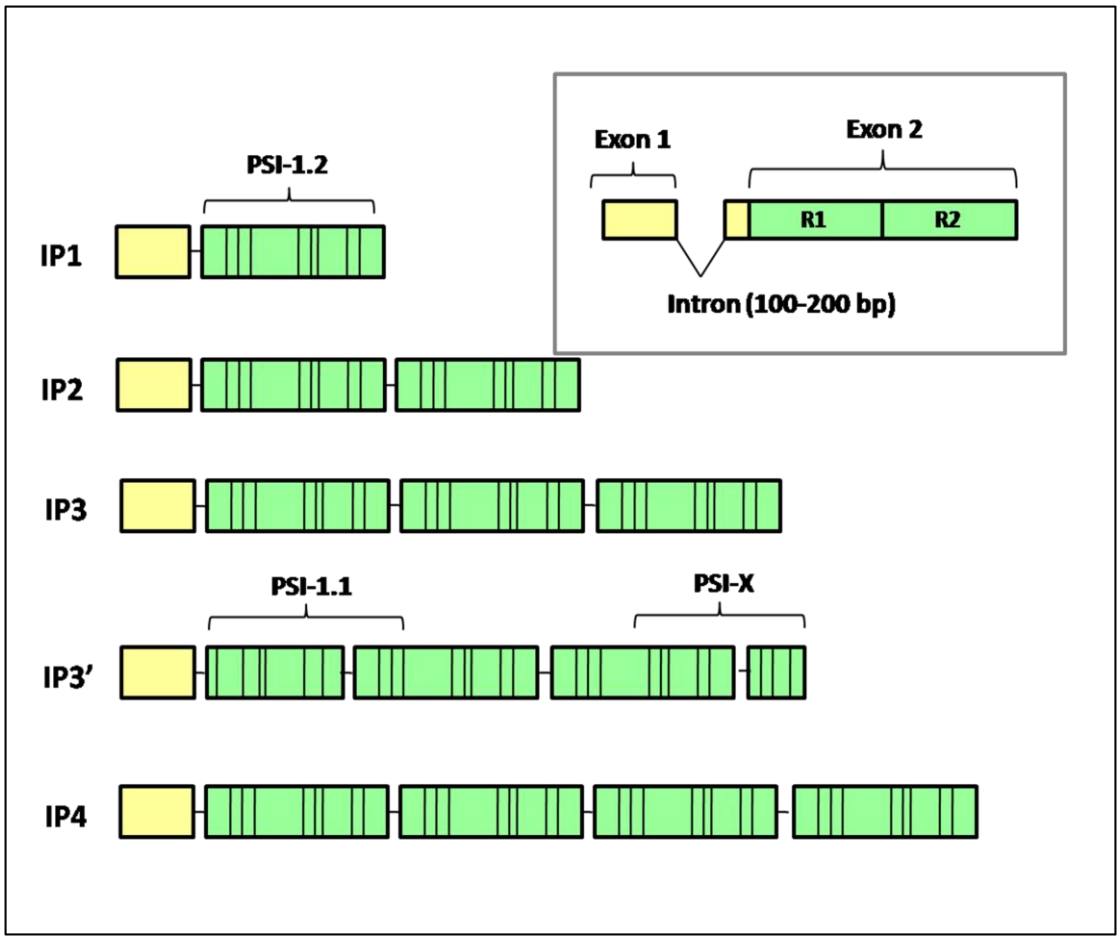


Figure 1.9: Pin-II PI gene structure and evolution. The inset shows the consensus PI structure. IP1...IP4 designate the total number of IP repeats (green boxes) within each precursor. Yellow box, signal peptide; black vertical lines, Cys residues. The presence of adjacent, identical repeats is a recurrent pattern. PSI-1.1. PSI-X and PSI-1.2 are paprika seed inhibitors. [Reproduced with modifications from Barta et al., 2002].

Kong and Ranganathan (2008) also support this gene duplication hypothesis for Pin-II family gene evolution. Unequal crossing over is presumed to be responsible for the expansion of the repeated domains (**Fig. 1.10**). Inhibitory domain multiplication with the domains remaining fused plays a critical role in generating such diversity (Christeller, 2005).

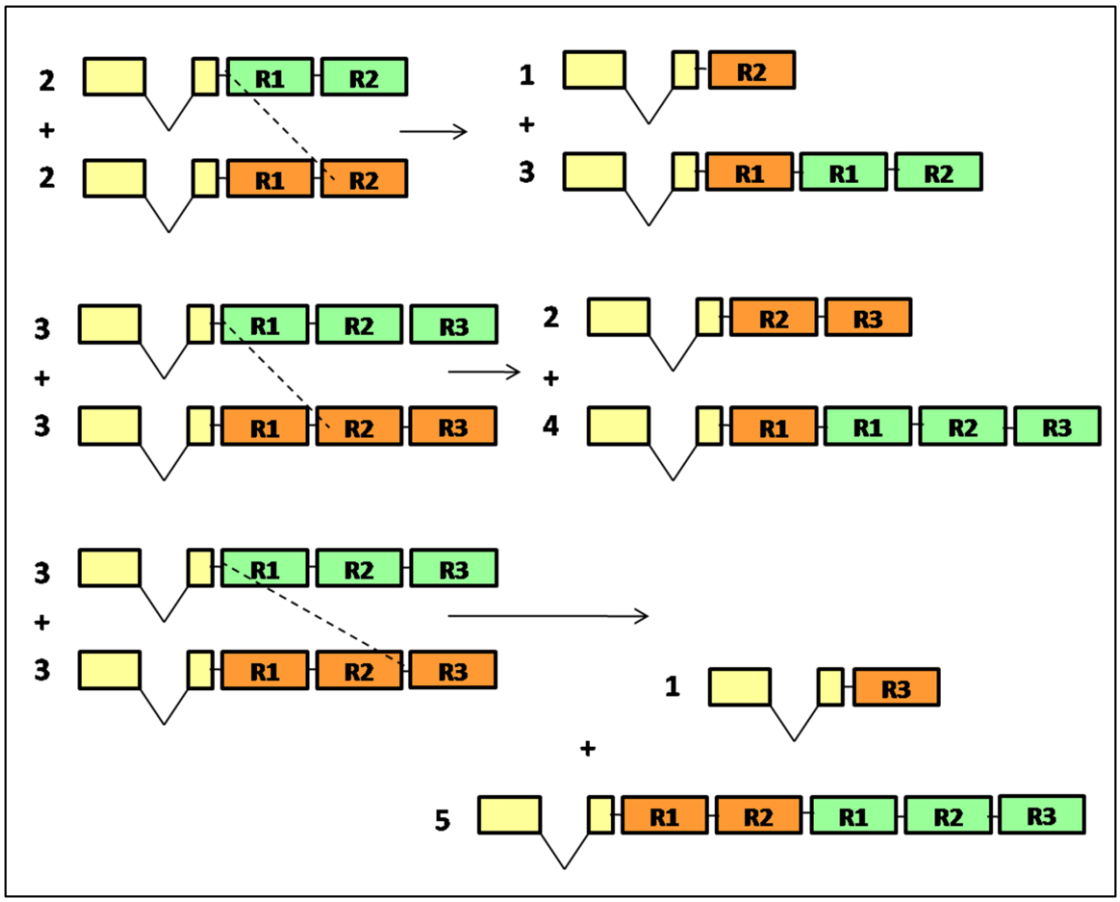


Figure 1.10: Unequal cross-over events in Pin-II PI family. potential unequal cross-over (UECO) events that explain the emergence of sequence identity patterns of the potato II family precursors. The two partners are colored green and orange. Two types of UECO are shown by dashed lines. [Reproduced with modifications from Barta et al., 2002].

1.7.2.2 Architecture of Pin-II PI proteins

The conserved Pin-II PI protein consists of an endoplasmic reticulum signal peptide of 25 aa followed by variable number of IRDs, each of ~55 aa. The aa 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 a conserved feature of each IRD. The IRDs are separated by 5 aa linker regions. A vacuolar sorting signal at the C-terminal region is present in some Pin-II PIs. In most of the Pin-II PI precursors, there are partial IRDs at the N- and C- terminal which form covalent bonding to generate a functional IRD. Exceptionally, *C. annuum* PIs do not possess the N- and C- terminal partial IRDs (Tamhane et al., 2009). For a multi-

IRD Pin-II protein, there are two possible domain organizations: (1) circularly permuted domain organization which is formed by the covalent joining of the two terminal half-repeats to form a repeat domain; or (2) tandem repeat domain organization where domains are arranged in beads-on-a-string way (Schirra and Craik, 2005) (**Fig. 1.11**).

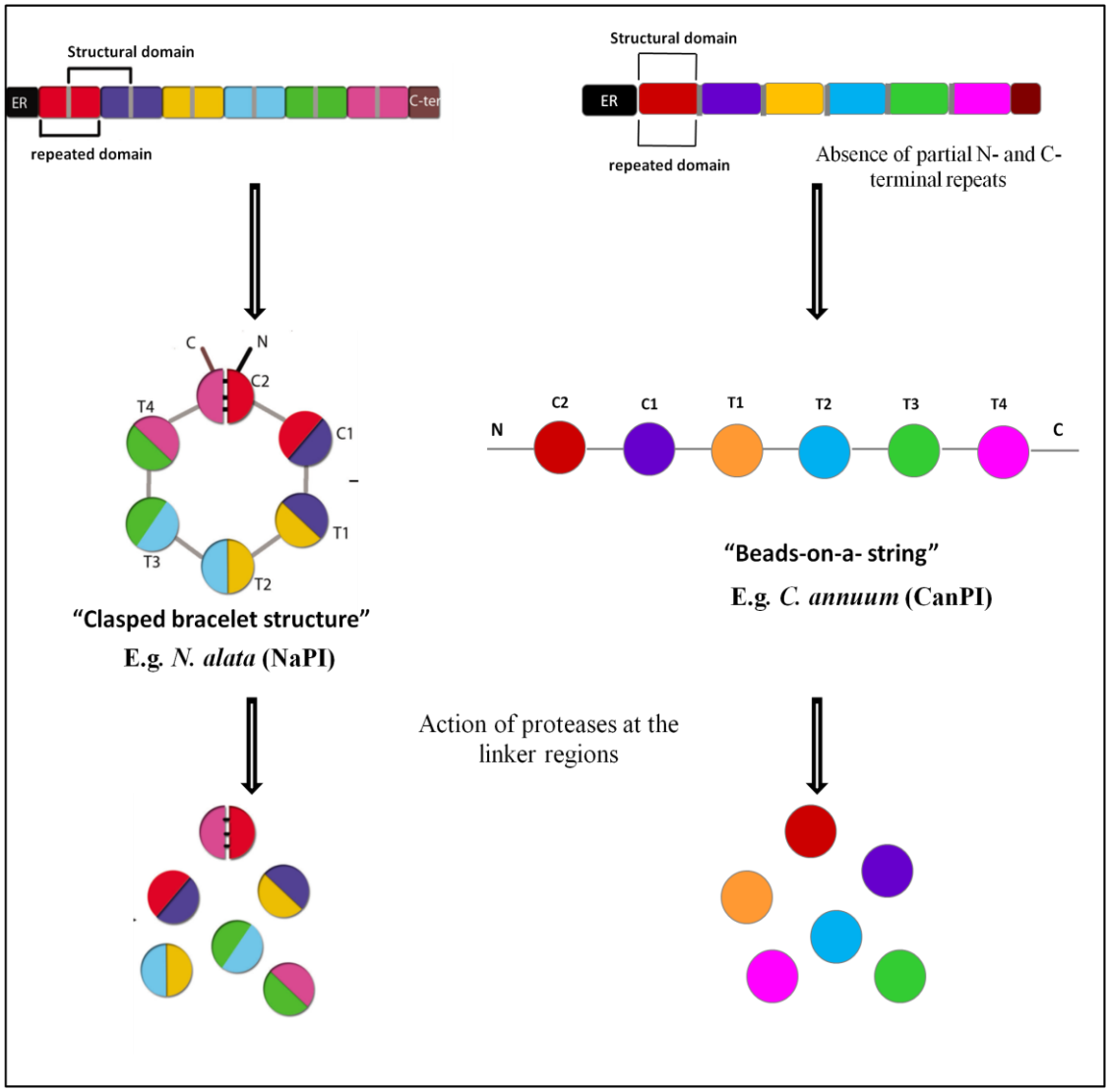


Figure 1.11: Model of Pin-II PI precursors. Structure and processing of the 6 domain Na-proPI. The 40.3 kDa PI precursor consists of 6 repeated domains preceded by an ER signal sequence and terminated by a vacuolar targeting signal. C2 is formed by joining of N- and C- terminal partial ends. In case of lack of these partial ends, beads on a string like structure is expected to be attained by the precursors. [Modified from Dunse and Anderson; ISB news report, June 2011].

The 43 kDa precursor PI of *N. alata* (NaProPI) forms a circular 'clasped bracelet' like structure as a result of formation of disulfide bridges between the partial repeat regions at the N- and C- terminals of the precursor (Scanlon et al., 1999; Schirra and Craik, 2005) (**Fig. 1.11)**. Sequence of the linker regions between the IRDs is almost conserved in *Nicotiana* sp (EEKKN), whereas show sequence variations in other genera. For example, in *C. annuum* sequences are QRNAK, EENAE, EASAE, EGNAE and EETQK. The linker region is very sensitive to proteases and is cleaved by endogenous proteinases *in planta* (Heath et al., 1995) to release individual functional units, IRDs, which are able to inhibit a single proteinase molecule. The linker sequences are assumed to play an important role in stabilizing the cross-repeat folding pattern of multi-domain Pin-II PIs.

In all the precursor Pin-II PI sequences, two 'types' of linkers are present, (i) DPRNP like and (ii) EEKKN like. Due to presence of prolines in the DPRNP linker it can adopt a smaller conformational range and thus prefers its own incorporation into the structure of PI domain while EEKKN linker has no conformational preference (Schirra and Craik, 2005). After an alignment of IRD sequences from Pin-II family Kong and Ranganathan (2008) have suggested that each IRD is formed by a combination of two fragments, namely 'Heavy' (H) and 'Light' (L) which may or may not be connected by a linker. Available repeat domains can be classified into three types based on the existence of linker sequences: 1) H-L type; here the H and L fragments are connected by linker-1 i.e. DPRNP; 2) L-H type; here the H and L fragments are connected by linker-2 i.e. EEKKN; and 3) H+L type; here there is no linker between the H and L fragments (**Fig. 1.12) (**Kong and Ranganathan, 2008). They propose that H-L topology is thermodynamically favourable as compared to L-H type of topology. The striking feature of multi-domain Pin-II PIs is that the sequence repeat does not correspond to the structural repeat. Thus, each PI domain comprises parts from two sequence repeats, and represents a circular permutation with respect to the sequence repeats which is a result of intramolecular domain swapping (Schirra and craik, 2005).

Phylogenetic analysis of Pin-II family grouped IRDs into seven clades on the basis of repeat number and species (Kong and Ranganathan 2008). In spite of high similarity among the PIs of Solanaceae species, repeat units from *C. annuum* stand out from all the others because they are of H-L type, unlike all the other groups. Thus, the

sequence repeat is identical to the structural repeat and lacks N- and C-terminal partial repeats, which form the "bracelet" link domain as in other multi-IRD PIs. As each domain adopts the H-L type topology, multiple-domain PIs from *C. annuum* are likely to adopt tandem structural domains with a "beads-on- a-string" like organization (**Fig. 1.11)**, which is different from all the other multiple-domain PIs in Pin-II family.

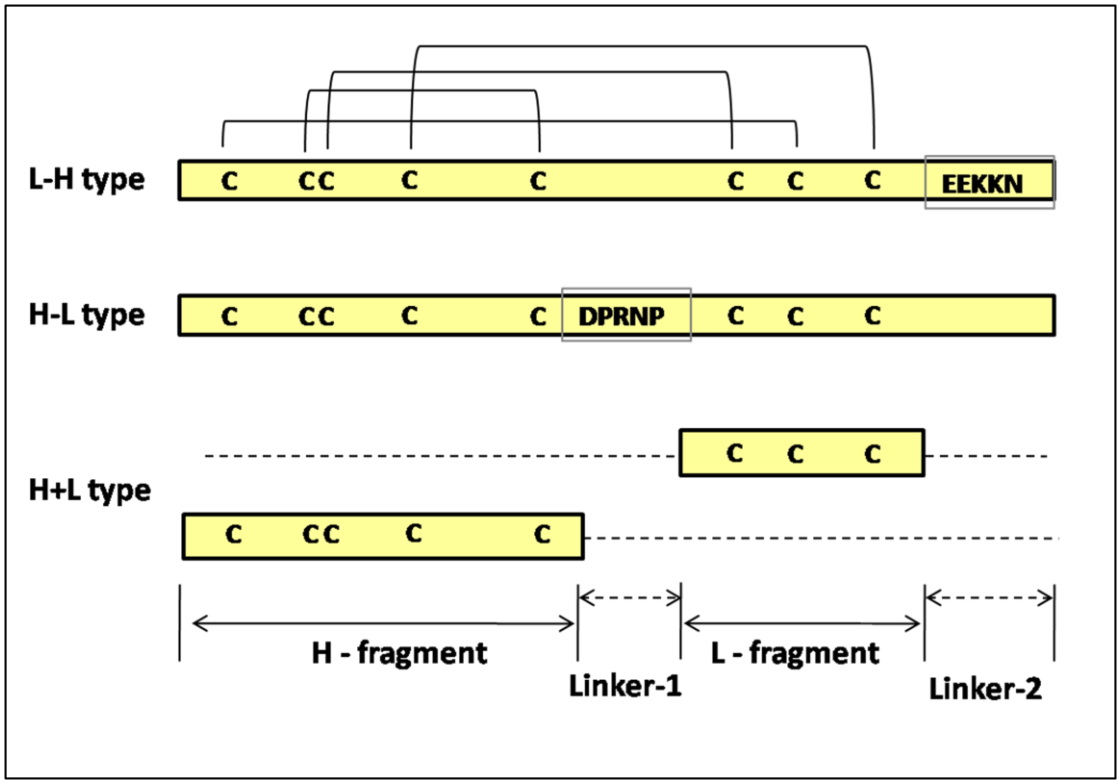


Figure 1.12: Diagrammatic representation of the three types of topologies in IRDs of Pin-II PIs. [Reproduced with modifications from kong and Ranganathan, 2008].

1.7.2.3 Pin-II PI protein structures based on NMR and X-ray crystallographic studies

The three-dimensional structure of several Pin-II PIs has been elucidated, either by X-ray crystallography or NMR. The structure of chymotrypsin-binding domain (PCI-1) from potato PI-II, in complex with *Streptomyces griseus* proteinase B was solved to 2.1Å by X-ray crystallography (Greenblatt et al., 1989). The structures of single IRD from Pin-II PIs of *N. alata* (Nielsen et al., 1994), two domain precursor PIs from tomato individually and in ternary complex with two molecules of subtilisin Carlsberg (Barrette-Ng et al., 2003a) and 6-IRD PIs from *N. alata* (Schirra and Craik, 2005) have been determined.

Single domain PIs having either trypsin inhibitory (TI) or chymotrypsin inhibitory (CI) reactive sites are formed from the multi-domain precursor protein by proteolytic cleavage at the linker regions (Heath et al., 1995; Lee et al., 1999). The eight conserved cysteines in each IRD are involved in formation of four disulphide bonds, which stabilize the repeat structure (**Fig. 1.13A**). The single IRD CI protein contains a triple stranded β-sheet as the dominant secondary structural element, with several turns and a short region of helix. The reactive site (CI) lies on an exposed loop, which is less defined than the rest of the protein. The overall shape of CI is disk like and the N- and C- terminals are exposed, indicating that this protein results from post-translational processing of the precursor protein (Greenblatt et al., 1989; Nielsen et al., 1994, 1995). Due to the high sequence identity between TI and CI IRDs, it has been anticipated that the TI domain also adopts 3D structures similar to CI (**Fig. 1.13B)**.

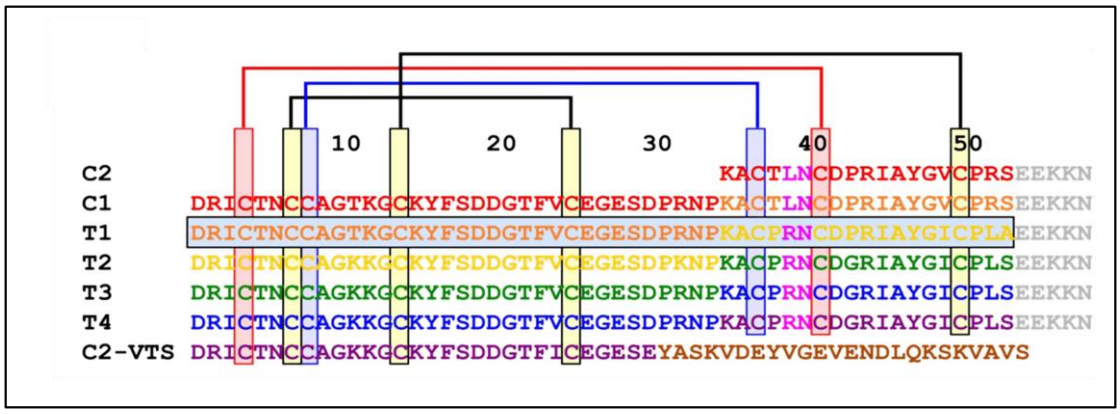


Figure 1.13A: Amino acid sequence of NaProPI. The reactive site of each inhibitor is colored magenta and the disulfide bridging patterns are indicated on top of the sequence. The two disulfide bonds (C8-C37 and C4-C41) anchoring the reactive-site loop to the core of the inhibitor are marked in blue and red, respectively. [From Schirra et al., 2008].

Further structural refinement of the TI or CI domains from *N. alata* by NMR (Schirra et al., 2008) has led to detailed information on the structural framework of single IRDs (**Fig. 1.13B)**. The reactive site of the TI domain containing the scissile peptide bond is positioned as a flexible loop and remains anchored to the core of the inhibitor by two disulfide bonds, C8-C37 and C4-C41. The imperative functional role of the two disulfide bonds anchoring the reactive site loop (RSL) is reflected by their conserved presence among all the known Pin-II inhibitors. Further stabilization of the RSL is accomplished by two additional disulfide bonds, C7-C25 and C14-C50; a prevalent network of hydrogen bonds; and presence of a proline residue at position P2

(Schirra et al., 2008). Apart from these, other highly conserved residues which are structurally important are Pro-18, Gly-38 and Gly-46 as they belong to the three β-turns, respectively (Kong and Ranganathan, 2008).

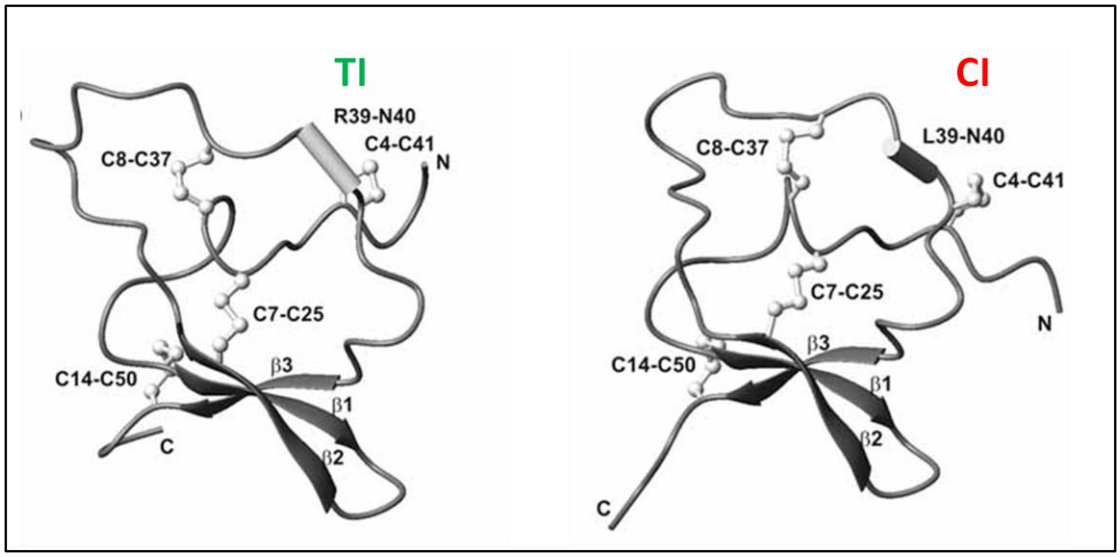


Figure 1.13B: 3-D structure of Pin-II PI proteins. The ribbon drawing of the representative models of solution structure of TI and CI, showing the regular secondary structure. [From Schirra et al., 2010].

Normally in a TI domain, the RSL is constrained by both C8-C37 and C4-C41 disulfide bonds thereby limiting its flexibility in order to provide it specificity for binding. Recent NMR studies on the disulfide bond variant of TI (T1) from *N. alata*, (C4A/C41A-T1) showed similar conformation of the reactive loop as the normal TI, with moderately decreased inhibitory potential (**Fig. 1.14A**). In contrast, the C8A/C37A-T1 variant showed a disordered and highly flexible RSL making it's binding to trypsin more difficult and therefore, suppressing its inhibitory potency (**Fig. 1.14B)** (Schirra et al., 2010). NMR relaxation experiments confirmed the much increased flexibility of binding loop for C8A/C37A-T1 variant and confirmed the indispensability of C8-C37 disulfide bond in maintaining the stability and function of the protein whereas the C4-C41 was not much critical as being supplemented by other stabilizing interactions (**Fig. 1.14C**). Flexibility of the RSL is remarkable in order to allow the binding of Pin-II PIs to wide range of proteinases but also to be specific for efficient binding.

It was believed that there were no natural variants for cysteine residues likely because of the essential role of disulphide bonds in Pin-II PIs. Recent studies by Li et

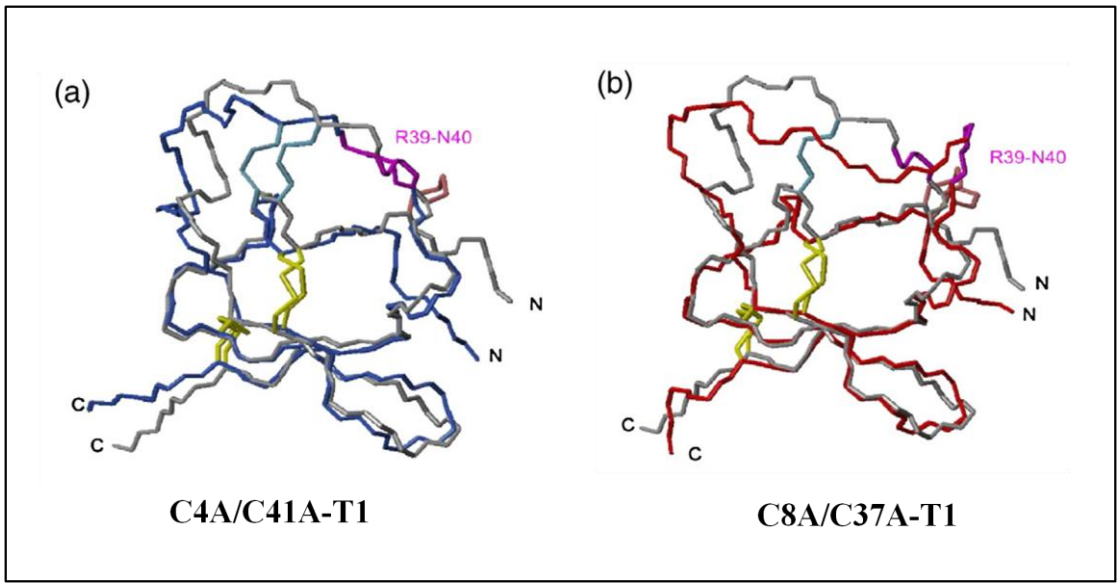


Figure 1.14(a): Backbone superpositions of C4A/C41A-T1 (blue) with the wild-type protein T1 (gray).
Figure 1.14(b): Backbone superpositions of C8A/C37A-T1 (red) with the wild-type protein T1 (gray). C7–C25 and C14–C50 disulfide bonds colored yellow. The C4–C41 disulfide bonds are colored light red, and the C8–C37 disulfide bonds are colored light blue. [From Schirra et al., 2010].

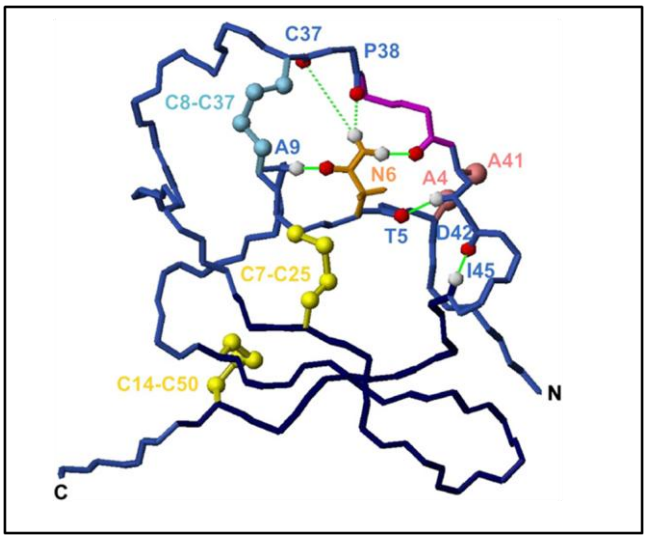


Figure 1.14C: Detailed view of C4A/C41A-T1 showing the stabilization of the reactive-site loop. Hydrogen bonds stabilizing the reactive-site loop are indicated by continuous green lines. [From Schirra et al., 2010].

al. (2011) have identified six natural variants of Pin-II PIs from potato, missing cysteine residues involved in one or two disulphide bonds. The selective loss of cysteine residues at the first reaction centre and in pairs has suggested that the loss of disulphide bonds is specific, and not random. Loss of disulphide bonds is often coupled with functional differentiation of the protein. Analysis of more naturally occurring variants can throw light on the evolution of disulphide bonds and the Pin-II PI protein evolutionary process.

The crystal structure of 2-IRD PI from tomato (TI-II) by Barrette-Ng et al. (2003a, 2003b) reveals significant conformational flexibility in the domains in absence of bound proteinases. Each individual IRD adopts the fold determined previously for the single domain Pin-II inhibitors. The two domains remain arranged in an extended configuration presenting the respective RSLs at the opposite ends of the inhibitor molecule interacting with two subtilisin molecules simultaneously in a single complex (**Fig. 1.15A)**. The unbound form of the inhibitor displays a range of conformations in the orientation of the IRDs. Conformational flexibility seen in the RSLs of unbound TI-II suggests a mechanism by which the inhibitor can balance the need for tight binding along with the broad range of protease specificity (**Fig. 1.15B)** (Barrette-Ng et al., 2003b). In a two domain inhibitor, individually each domain has identical secondary structure and the linker region connecting the two domains acquires form of a distorted loop. A two domain CI-TI, recombinant inhibitor also had shown that conventional domain swapping is dependent on the flexibility of the linker region connecting the swapped domains (Scanlon et al., 1999). The two domains are essentially independent of each other and have no long-lived and highly specific interactions between them. Both RSLs are positioned at the opposite ends, allowing the binding of two proteinases simultaneously without any steric interference. Overall, it appears that inter-domain interactions in Pin-II PIs are only transient (Scanlon et al., 1999; Barrette-Ng et al., 2003a, 2003b; Schirra and Craik, 2005). The lack of strong inter-domain association is likely to be important for individual inhibitors to ensure that there is no masking of reactive sites, especially if the number of domains is more than two in the precursor. For Pin-II PIs with more than two domains, the two binding sites would be less than 180° apart and would lead to steric hindrance upon binding to proteinases. This has highlighted the need for proteolytic processing of the multidomain PI precursor molecules in order to achieve stoichiometric ratio of PI-

proteinase for maximum protease inhibition (Heath et al., 1995, Schirra et al., 2005, 2008). Though a lot of information is available on the structure of single repeats from Pin-II PIs, the multidomain PIs remain unexplored to much extent.

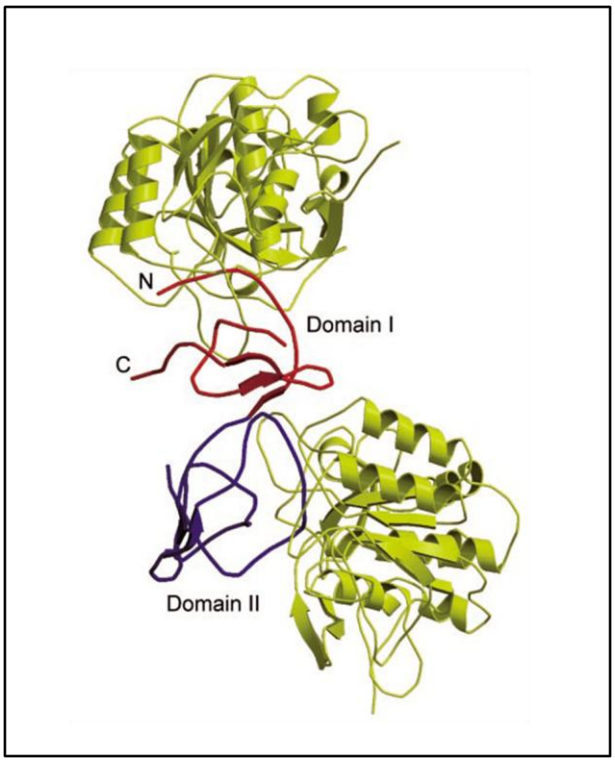


Figure 1.15A: Structure of the TI-II-(subtilisin)2 complex. The two domains lie at the opposite ends of the inhibitor molecule. [From Barette Ng et al, 2003a, 2003b].

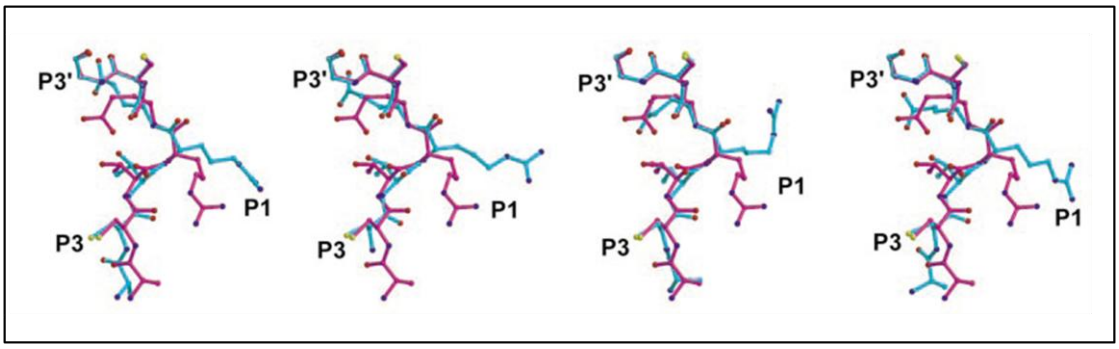


Figure 1.15B: Copies of the RSLs from domain I of unbound TI-II are shown superimposed onto that of bound TI-II. The bound form is shown in magenta and unbound form is shown in cyan colour. [From Barette Ng et al, 2003a, 2003b].

1.7.3 Interaction of Pin-II PIs with proteases

Each IRD of the Pin-II precursor contains a single reactive site either for trypsin or chymotrypsin inhibition. Single IRD PI of Pin-II family can bind to a single protease, while two domain PIs of tomato and potato can simultaneously inhibit two protease molecules (Greenblatt et al., 1989; Barette Ng et al., 2003a). The P1 residue of the reactive site determines its specificity; presence of lysine 'K' or arginine 'R' in the P1 position confers TI potential and Leucine 'L' bestows the inhibitor CI potential (**Fig. 1.13)**. The profound docking of the side chain of P1 residue in the S1 binding pocket of the protease plays an extensive role in the energetics of the specificity of PI-protease interaction (**Fig. 1.15)**. In bovine pancreatic trypsin inhibitor (BPTI, Kunitz type), the substitution of P1 residue with aa other than K/R showed a weaker side chain interaction of the P1 residue in the S1 binding pocket of trypsin (Otlewski et al., 2001). A stretch of 5 aa close to the reactive site, P3 to P2' is important in determining enzyme specificity of the inhibitor. It is bound by two disulfide bonds which gives the reactive site a considerable rigidity, while the aa in the adjoining region of this core segment show a very high sequence variability, conferring flexibility to provide broad inhibitory potential. The P4 to P3' of the inhibitor domain interacts with S6 to S2' of protease pocket to bring about its inhibition by mimicking a substrate. Mutational studies on PI-II from potato have highlighted the importance of secondary contacts also not involving RSL as well, in determining the specificity of protease inhibition. In PI-II from tomato, the inhibition capacity of a TI domain could not be transferred to the other domain by mutating the P1 residue or the residues within the RSL (Beekwilder et al., 2000; Schirra and Craik, 2005).

The effect of mutation/variation in the conserved residues is reflected not only in the structure but also in activity of Pin-II proteins. The selective substitution of conserved cysteine residues responsible for disulfide bond formation flanking the reactive site, resulted in poor inhibitory activity (K_i~ 1.8 µM) by C8A/C37A-T1 variant (Cys-C at positions 8 and 37 were replaced by Ala-A). On the other hand, substantial retention of TI activity by C4A/C41A-T1 variant (K_i~ 350 nM) as compared to the wild-type TI (K_i< 5 nM) affirmed the indispensability of C8-C37 bond in Pin-II proteins (Schirra et al., 2010) (**Fig. 1.14)**. The findings augment the vitality of conserved residues in Pin-II proteins for their structure and thus, their function.

The structural basis of inhibition of a multi-domain Pin-II inhibitor has been shown by its ternary complex with two subtilisin Carlsberg molecules and revealed how it can bind to and simultaneously inhibit two enzyme molecules within a single ternary complex (Barette-Ng et al., 2003a) (**Fig. 1.15)**. The diversity in the number of repeats and diversity within IRD sequences is predominantly observed in Pin-II PIs. Different specificities within a multi repeat protein contribute to a PI cocktail to fight against varied pest/ pathogenic attacks. In various members of Pin-II precursors studied there is a combination of TI/ CI domains. For example, the six domain *N. alata* PI (NaProPI) possesses four TI domains and two CI domains. The two, three and four domain PIs of *C. annuum* possesses varied combinations of inhibitory specificities. In potato PI which has two IRDs, one is TI specific and the other is CI specific. However, the NaProPI cannot bind to six proteases simultaneously because of steric interference. This 6-IRD PI of *N. alata* could inhibit maximum of four chymotrypsin or 2.6 trypsin molecules (Heath et al., 1995; Lee et al., 1999). In order to attain total inhibition potential, individual IRDs must be released from the precursor by proteolytic processing. This also highlights the reason behind the absence of inter-domain interactions in Pin-II PIs.

In a pioneering study by Bryant et al. (1976) Pin-II PIs from potato with CI and TI activity were characterized. Significant differences were noted among the isoinhibitors in their isoelectric points, immuno-electrophoretic mobilities, ion-exchange properties and their inhibitory activities against trypsin/chymotrypsin. Two IRD PIs from *S. tuberosum* (Eddy et al., 1980) and *L. esculantum* (Plunkett et al., 1982) were shown to inhibit chymotrypsin, subtilisin and trypsin. Two *C. annuum* seed PIs of around 50 aa, PSI1.1 and PSI1.2 have been found to inhibit trypsin, chymotrypsin, thrombin and factor Xa with different specificities (Antcheva et al., 2001). Two PIs from *C. annuum* leaves, CapA1 and CapA2 exhibiting a molecular mass of 12 kDa inhibited bovine trypsin and chymotrypsin suggesting presence of two inhibitory sites one each in the two repeats (Tamhane et al., 2005b). *C. annuum* PIs with variable number of IRDs and sequence variations, showed significant changes in their specificity and inhibitory potential towards proteases like trypsin, chymotrypsin and insect gut proteases (Tamhane et al., 2007). A 6 kDa protein with inhibitory activity against trypsin and chymotrypsin was purified from stigmas of *N. alata* rather

than the precursor protein of 41.6 kDa because of the processing at the linkers to generate active inhibitory units (Atkinson et al., 1993).

1.7.4 Effects of Pin-II PIs on insect proteases

Wound induction of Pin-II PIs and their role in plant defense against insect pests has been established by several studies which suggests that induction of such PIs strengthen the defense of plants which are already partially protected by the constitutively expressed PI defense. Tomato PIs inhibited about 50–80% HGP activity of *H. armigera* larvae feeding on various host plants and of various larval instars (Damle et al., 2005). Tomato PIs were found to be highly stable to insect gut proteases and bioassay using *H. armigera* larvae fed on artificial diet containing tomato PIs revealed adverse effect on larval growth and development, pupation, adult formation and fecundity in a dose-dependent manner. This demonstrated that host plant PIs also aid in their defense. The efficacy of *C. annuum* PIs against *H. armigera* gut proteases was demonstrated by Tamhane et al. (2005b). *In vitro* assays showed 68-91% inhibition of trypsin activity and 39-85% inhibition of chymotrypsin-like activity of *H. armigera* gut proteases from larvae fed on various host and non-host plants by *C. annuum* PIs extracted from leaf. *H. armigera* fed on *C. annuum* PI incorporated diet, showed growth retardation as well as reduction in fertility and fecundity for two consecutive generations. Moreover, recombinant expressed diverse CanPIs, with 1- to 4-IRDs were characterized for their insect inhibitory potential. *H. armigera* larvae fed on rCanPI diet showed 30% mortality and 40% lower mass among the survivors, in the early instars (Tamhane et al., 2007). Pupal mass reduction of 12-25% was recorded, leading to decreased fecundity. CanPI-7 with two CI sites and two TI sites showed the strongest anti-metabolic effect on *H. armigera*.

The addition of potato inhibitor II to artificial diets of larvae of *Heliothis zea* and *S. exigua* had shown an elevation of trypsin-like activities in their digestive tracts and inhibition of larval growth at about 10% concentration of the PIs in the artificial diets (Broadway and Duffey, 1986). All the five inhibitors (one 6-kDa CI and four 6-kDa TIs from a single 40.3-kDa precursor protein) from *N. alata,* evaluated for their ability to inhibit gut protease activity in insects, have shown 37 to 79% inhibition (Heath et al., 1997). The CI was less effective than TIs. Incorporation of *N. alata* PIs in the artificial diet of the native budworm (*H. punctigera*) and the black field cricket

(*Teteogryllus commodus*) revealed significant reduction in growth and failed to complete molting than insects on the control diet. *H. punctigera* larvae fed with a cotton leaf-based artificial diet, composed of 0.26% (wt/vol) of *N. alata* proteinase inhibitor (NaPI) showed higher mortality (80%) as well as lower larval mass as compared to those fed with the control diet (Dunse et al., 2010a). Interestingly, the consumption of NaPI by the larvae, led to induction of an insensitive chymotrypsin which was found to be resistant to inhibition by NaPI. The activity of this chymotrypsin was found to be inhibited by a Pin-I inhibitor (StPin1A) isolated from wound-induced leaves of *S. tuberosum*. *H. armigera* larvae fed with diets containing NaPI and StPin1A showed 50% and 40% weight loss, respectively than the larvae fed with control diet. Higher efficacy (90% loss in larval mass) was obtained, when fed with an artificial diet composed of both the types of inhibitor proteins, StPin1A and NaPI (**Fig. 1.16A)**. *In vitro* assays of NaTPI activity inhibited the trypsin-like activity from total gut proteases of 2nd and 3rd instar larvae of *M. sexta* and also affected the larval performance (Zavala et al., 2008). Although, it was suggested that larvae can minimize the effects of high NaTPI levels by feeding on leaves with high protein and low NaTPI activity, the host plant's endogenous NaTPIs are effective against *M. sexta.*

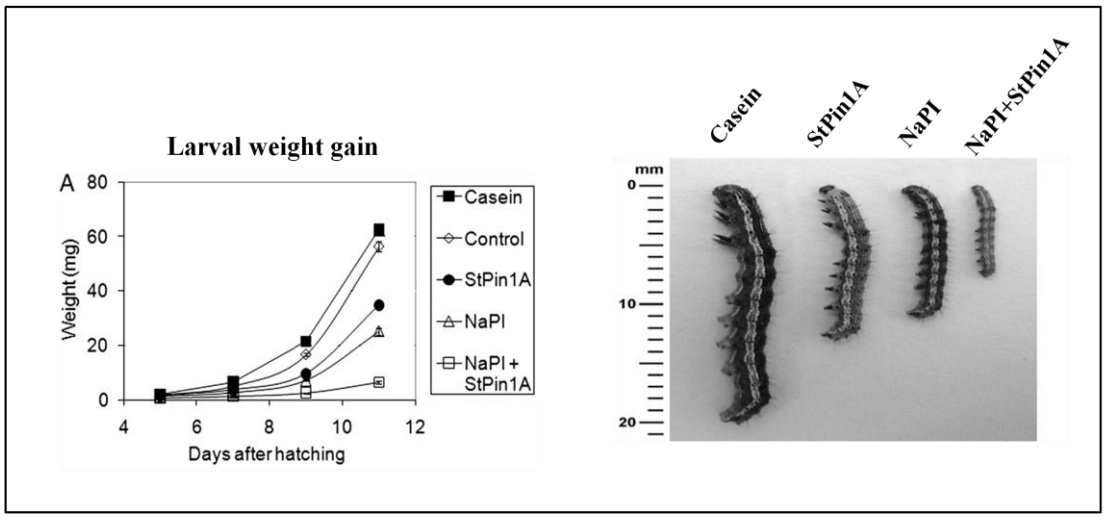


Figure 1.16A: Growth of *H. armigera* larvae on cotton leaf-based diets containing PIs. [From Dunse et al., 2010].

Jasmonate (JA)-elicited trypsin-proteinase inhibitor (TPI) anti-herbivore defense was studied by Wu et al. (2006) and correlation to phylogenetic relationships

was established among species of *Nicotiana*, e.g. *N. quadrivalvis* (Nq) and *N. clevelandii* (Nc), and *N. attenuata* (Na) and *N. obtusifolia* (No). High TPI activity and mRNA transcript accumulation were observed in Na, Nq and Nc, on applying oral secretions (OS) from *Manduca sexta* larvae onto the wounded tissues. It was shown that a network composed of an upstream signaling system, downstream interactions between *cis*- and *trans*- elements and post-transcriptional regulators probably controlled the PI expression. *S. nigrum*'s complete serine-protease-inhibitor (SPI) profile was examined by Hartl et al. (2010) and four PI genes; *SnSPI2c*, *SnSPI2a*, *SnSPI2c-R2* and *SnSPI2c-R3* were identified. SPI2c displayed high specificity for trypsin and chymotrypsin and was most abundant and highly upregulated on application of MeJA, regurgitant from *M. sexta* and mechanical wounding. Lower responses of the other two isoforms indicated functional diversification of SPI2c towards plant defense. In field experiments with transgenic plants silenced for SPIs, all SPIs displayed herbivore and gene-specific defensive properties. SPI2c silenced lines exhibited no effect on development and growth while suffered more overall damage from noctuids, than the wild type.

1.7.5 Improved insect resistance in transgenics expressing Pin-II PIs

Transgenic expression of Pin-II PIs in plants has helped them in improving resistance against insect pests and showed how addition of a defensive trait can provide enhanced protection against the important predators. Pin-II PI genes of tomato (inhibitor I and II) and potato (inhibitor II) were expressed in transgenic tobacco plants (Johnson et al., 1989). *Manduca sexta* larvae (tobacco hornworms) feeding on the leaves of transgenic plants containing inhibitor II, a powerful inhibitor of both trypsin and chymotrypsin, showed significant retardation in growth, compared to that of larvae fed on untransformed leaves. Levels of inhibitor II protein as low as 50 µg/g of tissue moderately affected larval growth, whereas levels above 100 µg/g severely reduced growth. Expression of inhibitor I, a potent chymotrypsin inhibitor, did not affect the growth of hornworm to that extent implying that trypsin inhibitory activity is a primary cause for detrimental effects on larval growth. Interestingly, greater insecticidal effect was observed in tobacco plants transformed with the genomic sequence of tomato PI-II than in those transformed with the cDNA sequence indicating appropriate splicing of exogenous sequences in the transgenic plants to obtain the active protein (Zhang et al., 2004). Tobacco plants transformed with potato

proteinase inhibitor II (*PPI* II) gene, encoding a chymotrypsin inhibitor showed reduced larval growth for *Chrysodeixis eriosoma* (green looper) while no adverse effect on *S. litura* was observed (McManus et al., 1994). Thus, a single proteinase inhibitor gene may not be universally effective against a range of insect pests.

A transgenic expression of multi-domain proteinase inhibitor precursor from *N. alata* (Na-PI) in tobacco resulted in accumulation of 6kDa polypeptides in leaves to a level of 0.3%. *H. armigera* larvae feeding on such tobacco leaves, exhibited higher mortality or were delayed in growth and development relative to the control larvae (Charity et al., 1999). Similar adverse effects of Na-PI were observed on *Epiphyas postvittiana* (light-brown apple moth) when expressed in transgenic apple plants (Maheswaran et al., 2007). Field trials with transgenic cotton plants, expressing the individual inhibitors and combination of inhibitors (NaPI-StPin1A), subjected to *H. armigera* infestation along with the natural prevalence of *H. punctigera* at the field site were conducted by Dunse et al. (2010a). They recorded an increase in number of cotton bolls in the transgenic line expressing both StPin1A and NaPI than the parental untransformed line and also a boost in lint weight per plant for these transgenic lines, when evaluated against the control line (**Fig. 1.16B**).

Pin-II PIs of potato had also been used in transgenic rice and wheat plants to control biotic infestations by *Sesmia inferens* in rice (Duan et al., 1996) and *Heterodera avenae,* a nematode in wheat (Vishnudasan et al., 2005). Combined leaf-specific over expression of potato PI-II and carboxypeptidase inhibitors in transgenic tomato resulted in increased resistance to *Heliothis obsolete* and *Liriomyza trifolii* larvae (Abdeen et al., 2005). Transgenic tobacco plants over expressing SaPIN2a (from *S. americanum*) under the control of CaMV 35S promoter were more resistant to *H. armigera* and *Spodoptera litura* larvae than the control plants (Luo et al., 2009). They also reported an increase in the glandular trichome density along with the promotion of trichome branching in transgenic tobacco plants.

An approach other than over expression of PI genes has also been considered to analyze the potential of PI in insect tolerance. Enzymes lipoxygenase and linolenic acid are involved in JA biosynthesis. Anti-sense lines for lipoxygenase gene in potato plants abolished the accumulation of PIs on wounding to much extent (Royo et al., 1999). As a consequence, the weight gain of Colorado potato beetles and beet

armyworm fed on anti-sense plants was significantly larger than those fed on wild type plants.

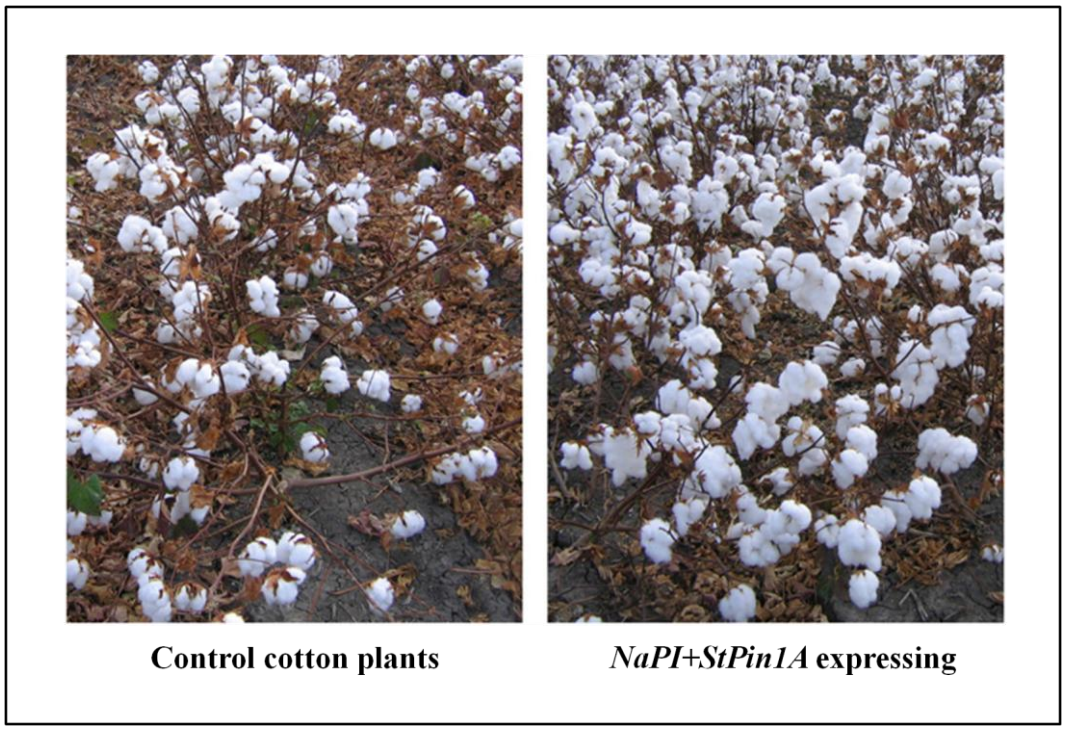


Figure 1.16B: Cotton balls on field-grown transgenic cotton plants. [From Dunse et al., 2010].

1.7.6 Endogenous functions of Pin-II PIs

Demonstration of anti-metabolic effects of plant serine PIs on insects by inhibiting the gut proteases (Hilder et al., 1993; Gatehouse et al., 1999) largely suggests that plant serine PIs have inhibition specificity for animal or microbial enzymes alone and not against plant proteinases. Subsequently, several evidences towards tissue specific accumulation of PIs and their developmental regulation have assigned endogenous functions to them. Plant parts that express Pin-II protein include flowers (PeñaCortés et al., 1991; Atkinson et al., 1993; Pearce et al., 1993; Sin and Chye, 2004; Damle et al., 2005), fruits (Damle et al., 2005; Tamhane et al., 2009), stem (Xu et al., 2001), tubers (Sánchez-Serrano et al., 1986) and roots (Taylor et al., 1993). It is suggested that they can regulate cell proteolysis by their action on endogenous proteinases, there by regulating the protein synthesis and turnover (Horn et al., 2005). Making a distinction between defense and developmental role of PIs

remains challenging and thus, only a few reports have elucidated the endogenous functions of Pin-II PIs (Chye et al., 2006; Hartl et al., 2010).

A novel endogenous role of *S. americanum* Pin-II PIs was suggested in phloem by Xu et al. (2001). Two wound-inducible mRNAs, *SaPIN2a* and *SaPIN2b* were identified which also showed expression in flowers. *SaPIN2a* was found to be abundantly expressed in stems, especially in companion cells (CC) and sieve elements (SE) of phloem. Predominant localization of SaPIN2a in SE suggested its involvement in regulating the proteolysis in phloem development or function. Transgenic lettuce plants expressing *SaPIN2a* gene showed its constitutive expression and the inhibition of endogenous trypsin-like and chymotrypsin-like activities thereby suggesting endogenous regulation of proteolysis by SaPIN2a (Xu et al., 2004). Transgenic *S. americanum* plants, overexpressing *SaPIN2a* under constitutive promoter showed low height and presence of chloroplast-like organelles with thylakoids in the enucleate SEs unlike the wild-type plants, which might impair the transport of various molecules through the phloem (Xie et al., 2007) (**Fig. 1.17**). The possible endogenous roles of SaPIN2a and SaPIN2b in inhibiting trypsin- and chymotrypsin-like activities during flower development were also suggested (Sin and Chye, 2004). Localization of *SaPIN2a* and *SaPIN2b* mRNAs and their proteins were found in flower tissues that would apparently undergo developmental programmed cell death (PCD): the ovules, the stylar transmitting tissue, the stigma, vascular bundles, nuclear cells of the ovule and the outermost cell layer of the placenta and suggested their role in PCD modulation (**Fig. 1.17)**. RNAi mediated silencing of SaPIN2 PIs adversely affected nutritional support to the endosperm and embryo (Sin et al., 2006) and reduced seed set due to seed abortion. Liu et al. (2006) had shown constitutive expression of *SaPIN2b* in glandular trichomes, predominantly in the gland cells. *SaPIN2b* promoter sequence contained six MYB binding motifs and a L1 box that are involved in trichome differentiation and development. Overexpression of *SaPIN2b* in transgenic tobacco resulted in a significant increase in glandular trichome density and trichome branching.

Contrary to these observations on the endogenous role of PIs in *S. americanum*, recent report by Hartl et al (2010) suggested certain degree of functional differentiation but also considerable functional overlap in the PIs from *S. nigrum*. Out of the four Pin-II PI isoforms identified in *S. nigrum*, *SnSPI2a*, *SnSPI2b* and *SnSPI1*

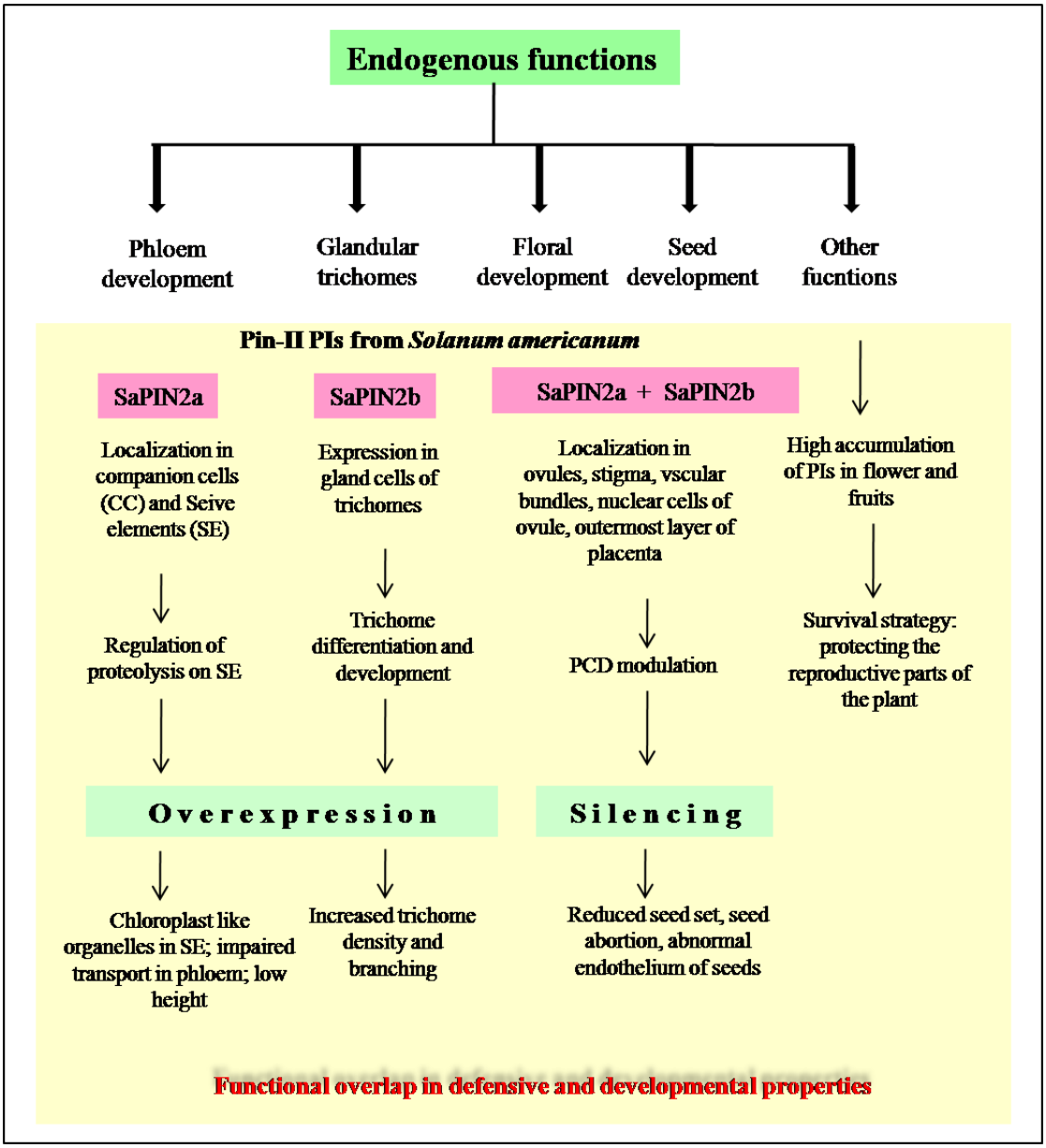


Figure 1.17: Endogenous functions of Pin-II PIs.

were found to be strong subtilisin-inhibitors while *SnSPI2c* was identified as a strong inhibitor of trypsin and chymotrypsin. In *S. nigrum* upon silencing *SPI2a* and *SPI2b*, only 0.7 to 2.8% of the seeds were found aborted or defective in contrast to the 80% seed abortion reported earlier by Sin et al. (2006) in *S. americanum.* The highly abundant PI (*SnSPI2c)* displayed typical characteristics of a defense-related gene where as the other two, *SnSPI2a* and *SnSPI2b*, showed an overlap of defensive and

developmental properties. The specificity of most SnPIs for subtilisin suggests an interaction with plant endogenous subtilases.

NaPIs account for up to 30% of soluble protein in stigma cells of *N. alata* (Atkinson et al., 1993). Johnson et al. (2007) observed two forms of NaPIs synthesized in maturing stigmas; the precursor form targeted for vacuole and the mature PI released from the precursor by proteolytic processing in the endoplasmic reticulum which gets trafficked to the cell surface of the stigma. Hundred times higher accumulation of PI and HGPI activity in flowers as compared to leaves and developing fruit stages, of field grown tomato plants was observed by Damle et al. (2005). This partitioning and diversification of metabolic resources towards the reproductive organ i.e. the flower of the plants was visualized as a survival strategy of the plant. The constitutive expression patterns of *C. annuum* PIs (CanPIs) were shown to vary qualitatively and quantitatively in various tissues of the mature *C. annuum* plant (Tamhane et al., 2009). In stem tissue, 1- and 2-IRD *CanPIs* were strongly expressed while higher expression of 3- and 4-IRD *CanPIs* was observed in various developmental stages of fruit. *CanPI* expression was found to be differentially up-regulated upon wounding and insect attack. But the low protein turnover indicated by the corresponding low PI activities suggested involvement of PI in plants endogenous function(s).

Bezzi et al. (2010) provide evidence for a role of PIs in the processing and secretion of nectar proteins. The nectar from flowers of PI-silenced *N. attenuata* plants contained 3.6-fold more total protein than the nectar of wild-type flowers. The effect of silencing PIs on nectar protein accumulation suggested an endogenous regulatory function for PIs in *N. attenuata* flowers. *N. attenuata* plants silenced for the PI genes do not produce PIs and grow faster, flower earlier, and produce more seed capsules compared with the PI-producing genotypes (Zavala et al., 2004a, b). Though PIs aid in protection from pests, it is a cost effective implementation for the plants.

Several studies have been directed to identify the endogenous functions of Pin-II PIs in plants but, a species dependant functional modification of PIs has made it difficult to demarcate the specific roles. Further, high sequence identity in the IRDs of Pin-II PIs has complicated the identification and assignment of a particular endogenous or defense function to a specific PI variant(s).

In summary, PIs are the naturally found defense proteins of the plant which have shown their potential against insect pests. Though they might have other functions, primarily they seem to be an effective strategy for improving the plant defense. The antibiosis effects of Pin-II PIs on major insect pests and the successful accomplishment of inhibitor-expressing transgenic plants in agricultural fields has paved the way towards development of superior crop varieties with insect resistance and high productivity.

1.8. Genesis of the thesis

Our lab is engaged in development of proteinase inhibitors as an approach against Lepidopteran insect pests. In the past studies, PIs from Chickpea, Bitter gourd, Winged bean and Capsicum have been well characterized and their interaction with insects was studied by *in vitro* and *in vivo* assays. The interesting leads obtained in Capsicum PIs (CanPIs) with respect to their diversity and efficiency against *H. armigera* growth retardation, were thus thought to be pursued further and became the rationale behind my thesis work.

Capsicum annuum ("Chilli pepper") is a domesticated species of genus *Capsicum* from Solanaceae family. It is one of the non-preferred hosts of *H. armigera*. The effects of PI proteins from *C. annuum* were pronounced on *H. armigera* and showed not only reduction in larval and pupal weights but also dramatic reduction in the fertility and fecundity which was carried on up to the second generation also (Tamhane et al., 2005b). Several novel forms of *CanPI* genes were identified which showed homology to Pin-II family and corresponded to 1, 2, 3, or 4-IRD PIs (Shin et al., 2001; Kim et al., 2001; Tamhane et al., 2009). Extensive variability in the *CanPIs* was a striking observation. Simultaneous presence and variable expression patterns were observed for *CanPIs* across the various *C. annuum* tissues screened. 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). Recombinant CanPI proteins (3- and 4-IRD type) inhibited *H. armigera* gut proteinases and exerted anti-metabolic effects on the larval growth and development (Tamhane et al., 2007).

Significantly high diversity in precursor Pin-II PIs, giving rise to several subtypes was a remarkable observation and raised several questions on their presence

specifically in *C. annuum*; (i) Are there specific types of CanPIs upregulated under induced conditions? (ii) Is this diversity amplified at the protein level by post translational processing of the precursor proteins? (iii) How much does the CanPI diversity lead to divergence in their inhibitory potential and interaction with proteases? (iv) Does this variation also lead significantly to structural variation and thus, their binding with proteinases? (v) How different would be the precursor CanPI structures, considering the absence of N-terminal partial repeat in *C. annuum* PIs? and more. In the present thesis, an attempt has been made to deal with the above questions. The screening of various CanPIs may help us in selecting the best insect defensive combination provided by the nature which can be used for crop protection. Furthermore, the study of insect response to ingested recombinant CanPIs might give insights on adaptation mechanisms of insects and help us in improving our approach for development of PIs against insect pests.

1.9. Organization of the thesis

The leads coming out of the earlier studies on *C. annuum* were taken up as a foundation to further investigate the interactions of CanPIs with the insect pests, using various biochemical and molecular approaches and the extensive information available regarding Pin-II PIs as summarized in the present chapter **'Review of Literature'**. The work carried out has been further organized in to the following chapters and the contents are as follows:

Chapter 2: Implications of induced proteinase inhibitor diversity in *Capsicum annuum*

This chapter describes differential expression, processing and tissue-specific distribution of *CanPIs* under steady-state and induced conditions. Cloning and sequence characterization were done to identify specific *CanPIs* expressed under induced conditions and deriving correlations for their defense functions. Characterization of the induced and tissue-specific PI activity by proteomic approaches revealing the role of post-translational modifications has also been reported.

Chapter 3: Interaction of recombinant CanPIs with *Helicoverpa armigera* gut proteases reveals their processing patterns, stability and efficiency

This chapter describes the cloning and recombinant protein expression in *Pichia pastoris* of selected *CanPI* genes. Recombinant proteins were characterized, *in vitro* and *in vivo* for their interaction with proteases using biochemical and proteomic approaches. The significance of aa variations leading to unique specificities against proteases, stability and efficiency of CanPIs has been reported.

Chapter 4: Structural-functional insights of single and multi-domain *Capsicum annuum* proteinase inhibitors

This chapter describes biochemical and biophysical characterization of representative single and multi-domain recombinant CanPI proteins. Inhibition kinetics and assays were done to evaluate relative binding and structural transitions under various conditions were monitored by biophysical techniques. Insights in to conformational stability, structure and binding mechanisms of multi-domain Pin-II PIs have been highlighted.

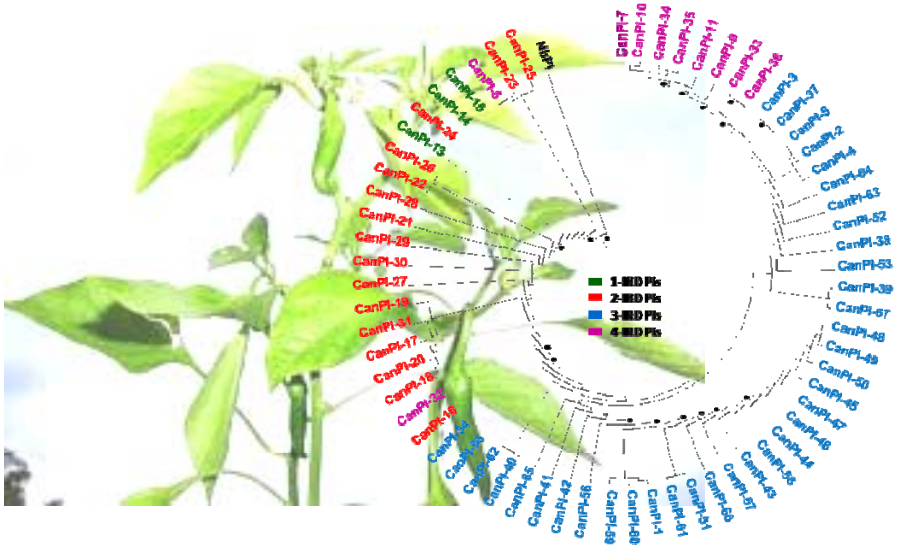
Chapter 5: General discussion and future directions

This chapter describes important findings of the present research work. Several aspects of CanPI expression, regulation, inhibitory efficiencies, stability, processing and structure have been discussed. The outcomes of present work throw light on the significance and advantage of using *C. annuum* PIs for developing insect tolerant transgenic plants. Some future approaches have also been discussed which can aid in further understanding of the CanPI structural and functional diversity and better implementation in development of plant defenses against insect pests.

Finally, all the literature referred, to develop protocols and to infer our results vis-a-vis other systems has been listed in the **Bibliography** chapter.

Chapter 2

Implications of induced proteinase inhibitor diversity in *Capsicum annuum*



The research work described in this chapter is part of a full-length paper, which has been communicated to BMC Plant Biology

CHAPTER 2

Implications of induced proteinase inhibitor diversity in *Capsicum annuum*

Abstract

To explore the functional specialization of an array of *Capsicum annuum* (L.) proteinase inhibitor (*CanPIs*) genes, their expression, processing and tissue-specific distribution were studied under steady-state and induced conditions. Inductions were performed by subjecting *C. annuum* leaves to various treatments, namely aphid infestation or mechanical wounding followed by treatment with either oral secretion (OS) of *Helicoverpa armigera* or water. The elicitation treatments regulated the accumulation of *CanPIs* corresponding to 4-, 3-, and 2-inhibitory repeat domains (IRDs). Sixty-seven *CanPI* genes composed of 55 unique IRDs were identified in total including those reported earlier. The *CanPI* gene pool either from uninduced or induced leaves was dominated by 3-IRD PIs and trypsin inhibitory domains. A major contribution by 4-IRD *CanPI* genes possessing trypsin and chymotrypsin inhibitor domains was specifically revealed in wounded leaves treated with OS. Among sixty clones analysed per treatment, wounding displayed the highest number of unique *CanPIs* while wounding with OS treatment resulted in the high representation of specifically *CanPI-4*, *-7* and *-10*. Characterization of the PI protein activity through two dimensional gel electrophoresis revealed tissue and induction specific patterns. Consistent with transcript abundance, wound plus OS or water treated *C. annuum* leaves exhibited significantly higher PI activity and isoform diversity contributed by 3- and 4-IRD *CanPIs*. *CanPI* accumulation and activity was weakly elicited by aphid infestation yet resulted in the higher expression of *CanPI-26*, *-58* and *-63*. Based on the differentially elicited *CanPI* accumulation patterns, it is intriguing to speculate that generating sequence diversity in the form of multi-IRD PIs is a part of elaborate plant defense strategy to obtain a diverse pool of functional units to confine insect attack. Plants can differentially perceive various kinds of damage by wounding and/or insect attacks and respond appropriately through activating plant defenses including regulation of PIs at transcriptional, translational and post-translational levels.

2.1. Introduction

Plants have evolved elaborate defense strategies composed of constitutive and inducible responses in order to cope with herbivore challenge. The induced defenses commence only when herbivore-derived signals are perceived by the plants. A wide array of studies have reported the induction of direct and indirect plant defenses in response to herbivory and other biotic stresses (Karban and Baldwin, 1997; Haruta et al., 2001; Van Dam et al., 2004; Hartl et al., 2010). Insect damage, mechanical wounding and/or elicitors in insect oral secretions (OS), such as fatty acid amino acid conjugates, volicitin, inceptins, caeliferins, and glucose oxidase, stimulate the local and systemic release of signalling intermediates like systemin and/or jasmonic acid which then amplify the defense cascade throughout the plant (Walling, 2000; Baldwin et al., 2001; Wu and Baldwin, 2010). Though the major consequence of herbivory is wounding, plants' responses to insect feeding are more complex due to the elicitors present in the insect OS (Baldwin et al., 2001). Defense responses entail the regulated activation of plant defense genes and the suppression of growth-related genes (Hermsmeier et al., 2001; Giri et al., 2006). As a result, defensive metabolites and/or proteins accumulate in plants within the local tissues damaged by herbivores as well as systemically in undamaged tissues.

The accumulation of trypsin and chymotrypsin-like proteinase inhibitors (PIs) throughout the aerial tissues of tomato and potato plants was demonstrated to be a direct consequence of insect-mediated damage or mechanical wounding (Green and Ryan, 1972). Thus, serine PIs represent one of the best examples of locally and systemically induced responses in Solanaceous plants (Bryant et al., 1976; Plunkett et al., 1982; Ryan, 1990; Ryan, 2000; Damle et al., 2005; Tamhane et al., 2009). The constitutive expression of PIs, which has been reported to occur in storage organs and the reproductive tissues of plants, may fulfil anti-insecticidal as well as other endogenous functions *in planta* (Garcia-Olmedo et al., 1987; Ryan, 1990; Patankar et al., 1999; Hartl et al., 2010; Tamhane et al., 2009).

Most Solanaceae members contain a multi-gene family encoding Pin-II type PIs (Wu et al., 2006; Tamhane et al., 2009; Hartl et al., 2010), which possesses considerable sequence diversity resulting from variations in tandem sequence repeats, domain duplications and circularly permuted domain organizations (Barta et al., 2002). A distinct feature of these PIs is the presence of tandem repeats of a 50-amino-acid polypeptide called IRDs, which can vary from 1 to 8 with inter-connecting linker

peptides. Horn et al. (2005) isolated a set of IRDs resulting from differential proteolysis at the linker peptide separating the subunits of a 7-domain precursor from methyl-jasmonate-elicited *N. attenuata* leaves.

Several different PI proteins and genes with 1- to 4-IRDs have been identified and characterized from *C. annuum* (CanPIs) tissues (Antcheva et al., 2001; Moura and Ryan, 2001; Kim et al., 2001; Shin et al., 2001; Tamhane et al., 2005; 2007; 2009). There was substantial variability in the induced expression of *CanPIs* upon aphid infestation, virus infection, chewing by insects and mechanical wounding. The abundance of transcripts did not always result in higher CanPI proteins, though they were well correlated in lepidopteran-infested *C. annuum* leaves. Furthermore, these studies indicated that many *CanPIs* expressed simultaneously, but the significance of such PI expression diversity in *C. annuum* remained unclear.

In order to examine the potential functional specificities of various isoforms of the CanPI in *C. annuum*, the following questions were addressed: (i) Does elicitation increase PI isoform diversity? (ii) How specialized is the induction response to a particular treatment? In the present chapter, diversity in the CanPI transcripts and the protein profiles were investigated following experimental inductions of *C. annuum* leaves. Sequencing revealed 44 novel *CanPI* transcripts, increasing the total known to 67. Selective analysis of PI activity in proteomes using 1D and 2D electrophoresis followed by mass spectrometry depicted local and systemic responses in PI activity.

2.2. Materials and methods

2.2.1. Plant material and induction treatments

C. annuum seeds (cv PhuleJyoti) (diploid) were grown in pots with Soilrite (Mixture of horticulture grade expanded perlite, Irish Peat moss and exfoliated vermiculite in equal ratio; M/s Naik Krushi, Pune, MS, India) and supplemented daily with Hoagland solution. 30-day-old seedlings were transferred to individual pots and grown in a growth chamber maintained at 23°C (±2°C) with a 14 h light photoperiod. Leaves, stems, various stages of fruits (early, mid, turning and late), roots and flowers from mature plants (3 months old) were harvested for screening tissue-specific CanPIs.

All induction experiments were performed on 3-month-old plants. Leaves were mechanically wounded by rolling a fabric pattern wheel along the length of the lamina

(4 to 6 rolls depending on the size of leaf) and the resulting puncture wounds were immediately treated with water or OS. These were considered as local tissue, whereas the non-wounded leaves one node above or below were harvested to measure systemic responses. Local and systemic tissues were collected after 30 h of treatment. OS used was collected from *H. armigera* larvae and diluted 1:50 times in MQ water. Plants were kept in an open garden for passive aphid infestation. Natural infestation by *Myzus persicae* was observed on *C. annuum* leaves within a week. The density of aphids was high on the leaf lamina towards the petiole. Leaves with at least 20 nymphs per leaf growing at the same nodes were collected as local tissue whereas non-infested leaves were harvested as systemic tissues. Since the plants were naturally infested by aphids in open conditions, the possible comparison of this treatment with the other two, which were performed in a highly specific and controlled manner, was limited. Leaves from un-elicited control plants growing at the same nodes were harvested from uninduced plants. All the tissue collections were done at the same time and were flash-frozen in liquid nitrogen and stored at -80°C until further use.

2.2.2. Expression profiling, cloning of Pin-II genes and their sequence analysis

Total RNA from *C. annuum* leaf tissues (uninduced and all three inductions-systemic) was isolated using TRIZOL (Invitrogen, Carlsbad, CA, USA) followed by DNAse treatment at 37°C. Purified RNA was quantified by spectrophotometry and 1.5 µg was used for first-strand cDNA synthesis using a reverse transcriptase kit (Promega, Madison, WI, USA). Proof-reading Accuprime Pfx DNA polymerase (Invitrogen) was used to amplify cDNAs from systemic leaf tissues of individual treatments and uninduced leaf in independent PCR reactions using *CanPI* gene (Genbank accession: AF039398) specific primer pair (CanPin-1F and CanPin-1R; **Table 2.1A**). Amplicons were cloned into pGEMT-easy vector (Promega) and 60 cloned fragments from each treatment were sequenced using standard T7 forward and SP6 reverse primers. Sequence editing and analysis were carried out using BioEdit, Clustal X and DNA star (Laser gene, DNASTAR, Inc., Madison, WI, USA). Specific primer pairs were designed for individual *CanPI* genes in order to check the expression of specific *CanPIs*. However, due to a high degree of similarity/homology within *C. annuum PIs*, it was possible to design gene-specific primer pairs for *CanPI-3*, *-5*, *-7*, *-8* and *-10* only. The sequences of oligonucleotide primers and specific pairs used for the internal differentiation of *CanPIs* are stated in **Table 2.1A** and **B**.

Table 2.1: Primers used for *CanPI* expression analysis

[A]: Oligonucleotide primers for RT-PCR and *CanPI* internal differentiation

Primer name	Primer sequence (5'-3')	Region
Can 18s rRNA F	CCG GTC CGC CTA TGG TGT GCA CCG G	-
Can 18s rRNA R	CCT CTG ACT ATG AAA TAC GAA TGC CCC	-
CanPin-1F	ATG GCT GTT CCC AAA GAA G	5' end- Signal Peptide (SP)
CanPin-1R	CTG TTC ATG CTT TTA CTT TTC	3' end of CanPI gene
V49-F	GCT TCC CTA CTT GTA CTT GG	Internal- SP
V51-R	CAA TTG GTG CAT ATG GGT C	Internal- towards 3' end
V52-R	C ACA GTT CAG AGT GCA AGC	Internal- towards 3' end
V53-R	T CTT GGA TCA CAG TTC AGA GTG	Internal- towards 3' end
V57-F	GCC TTC CTA CTT GTT CTT G	Internal- SP
V58-F	GCT TCC TAC TTG TTC TTG G	Internal- SP
V63-F	CAA AGA AAC GCA AAA GAA CC	Internal- linker

[B]: Oligonucleotide pairs for internal differentiation of *CanPIs*

Primer pair	Specific CanPI	Amplicon size (bp)
CanPin-1F, CanPin-1R	All CanPIs	789, 614, 455
V57, CanPin-1R	CanPI-3	600
V58, CanPin-1R	CanPI-5	585
CanPin-1F, V51	CanPI-10	644
V63, V53	CanPI-7	552
V49, V52	CanPI-8	574

2.2.3. HGP extraction, leaf protein extraction and proteinase inhibitor activity assays

Gut tissue was dissected from the laboratory reared 4th instar larvae of *H. armigera* and immediately frozen in liquid nitrogen. *H. armigera* gut proteases (HGP) were extracted from gut tissue by homogenizing in 0.2 M Glycine-NaOH buffer, pH 10.0 in 1:1 ratios (w/v) and kept at 4ºC for 2 h (Patankar et al., 2001). The suspension was centrifuged at 13,000X g, 4ºC for 20 min and the resulting supernatant was used as a source of gut proteases of *H. armigera* (HGP).

Total soluble protein was extracted from 1g of fresh leaf tissue obtained from uninduced and induced (local and systemic for all treatments) *C. annuum*, using a mixture of water and 5% polyvinylpolypyrrolidone (Sigma, St. Louis, MO, USA). Following protein estimation using Bradford reagent (Bio-Rad Laboratories, Hercules, CA, USA), the trypsin inhibitory and *H. armigera* gut protease (HGP) inhibitory activities were estimated enzymatically in the leaf extracts using a synthetic substrate Benzoyl-DL-Arginyl-*p*-Nitroanilide (BA*p*NA) (Sigma) as described earlier (Tamhane et al., 2005). For enzyme inhibitor assay, the protein extract was mixed with the enzyme and premix was incubated at 25-27ºC for 20 min. The residual enzyme activity was then estimated. One proteinase unit was defined as the amount of enzyme that increased the absorbance by 1.0 OD under the given assay conditions. One PI unit is defined as amount of inhibitor required for inhibiting one proteinase activity unit. Protease inhibitor activity was expressed as trypsin inhibitory units per mg of tissue (TIUs/mg). Various concentrations of the leaf protein extracts were titrated against HGP to determine maximum inhibitory activity. All the assays were carried out in triplicate and statistical analysis of the data was performed using single-factor ANOVA followed by Tukey's post hoc analysis (Microsoft Excel).

2.2.4. 1D and 2D electrophoresis for in-gel identification of proteinase inhibitory activity

Equal TIUs of leaf protein extracts were resolved on 12% native-PAGE and further processed in order to visualize trypsin inhibitor isoforms using the previously described gel X-ray film contact print (GXCT) method (Pichare and Kachole, 1994). This involves resolution of PIs in native or SDS-PAGE gels followed by their incubation in a buffer (Tris- HCl, pH 7.8) containing trypsin (0.04% w/v). These gels are further placed on gelatin coated X-ray films for varying time points. Due to the action of the proteinases imbibed in the gel, the gelatin coating on the X-ray film is degraded, except for the corresponding region in the gel where the PIs are present. The PI activity is thus indirectly visualized as a band of undegraded gelatin. Similarly, to visualize protease activity, protein extracts incubated with HGP were resolved on 8% native-PAGE gel and processed by GXCT (Harsulkar et al., 1999).

For 2D electrophoresis, acetone-precipitated proteins were re-suspended in rehydration buffer and separated by isoelectric focusing. For the first dimension, 11-cm IPG strips, pH 3-10 NL (Bio-Rad Laboratories) were rehydrated with 300 µg of

protein (in 200 µl rehydration buffer with DTT excluded) for 14 h at room temperature (RT). The proteins were focused on a Protean IEF Cell (Bio-Rad) at 20°C, rapid voltage ramping, 20,000 volt-hours, and 50 µA current per IPG strip. Preceding to 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% SDS, 20% glycerol and 2% DTT) followed by an iodoacetamide buffer (II) (DTT replaced by 2.5% iodoacetamide), each for 10 min at RT. Second dimension separation was done on 12% SDS-PAGE gel using Hoefer electrophoresis unit (GE Healthcare Bio-sciences AB, Buckinghamshire, UK) at 24°C and 200 V constant. TI activity was visualized by GXCT. The entire gel after 2-DE, was washed thrice for 20 min each by 2.5% triton X-100 followed by equilibration in 0.1 M Tris-HCl buffer, pH 7.8 and incubation in 0.04% trypsin solution and then overlaid on X-ray film. After washing the film, the TI spots were visualized as unhydrolyzed gelatin on X-ray films.

2.2.5. Partial purification and MALDI-TOF-MS analysis of leaf protein extracts

To identify the proteinase inhibitors from *C. annuum* leaf tissues, protein extracts were partially purified in the following manner. Ammonium-sulphate-precipitated proteins (90% saturation) from leaf extracts were resuspended in 50mM Tris-HCl buffer, pH 8.0 and subsequently treated at 65°C for 10 min before being desalted using PD SpinTrap^{TM}G-25 column (GE Healthcare). Proteins were then separated on DEAE-Sephacel (GE Healthcare) equilibrated with 50 mM Tris-HCl buffer, pH 8.0 and flow-through was collected separately from bound proteins that were eluted with a gradient of 0.25 M to 0.4 M NaCl in 50 mM Tris-HCl buffer, pH 8.0. Fractions with trypsin inhibitor activity were concentrated and desalted using a PD Spintrap column. The partially purified protein fractions were pooled and qualitatively analyzed by matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS). The mass spectral analysis was done on Voyager-De-STR MALDI-TOF (Applied Biosystems, Framingham, MA, USA) equipped with 337 nm pulsed nitrogen laser. The mode of operation was in a positive linear mode with an accelerating voltage of 25 kV. All spectra's were acquired by accumulating 50 single laser shots over each sample spot with the following settings: delayed ion extraction time of 1100 ns, grid voltage 93% and low mass ion gate set to 1000 Da. In brief, 2 µg of protein sample was mixed with 20 µl of freshly prepared

sinapinic acid (Sigma) (in 30% acetonitrile [ACN], 0.1% trifluoroacetic acid [TFA]) and spotted on the stainless steel MALDI plate, and spectral profiles were acquired in the range of 1 to 25 kDa. The spectra were analyzed with Data Explorer^{TM} for regions of interest and processed for advanced base-line correction and noise removal.

2.2.6. SDS-PAGE analysis of leaf proteins and their identification by database searches

To obtain protein sequence data, the partially purified proteins were separated on 16% Tricine SDS-PAGE gel (Schagger et al., 2006) to resolve the low molecular mass proteins. Protein bands were excised from the gel and completely destained by washing in 50% acetonitrile (ACN)/50% 50 mM ammonium bicarbonate followed by dehydration with 100% ACN so that the gel pieces appear shrunken and white. Gel pieces were reduced in 10 mM DTT (in 100 mM ammonium bicarbonate) for 60 min at 56°C. Alkylation was done by incubation with 55 mM iodoacetamide (in 100 mM ammonium bicarbonate) for 45 min in dark at RT. Gel pieces were again washed and dehydrated with ammonium bicarbonate and ACN as earlier. In-gel trypsin digestion was carried out overnight with 0.5 µg of trypsin (Porcine trypsin; Promega) in 20 µl of 50 mM ammonium bicarbonate at 37°C. The digest solution was collected in separate tubes. Peptides were extracted with 50 µl extraction buffer (50% ACN/2% formic acid) by vortexing and sonication and the supernatant was combined with the initial digest solution. The extraction step was repeated twice. The 120µl solution containing the peptide mixture was then vacuum dried.

Dry peptides were dissolved in 10 µl of aqueous 30%ACN/ 0.1% trifluoroacetic acid and analyzed by Q-TOF-MALDI-TOF-MS/MS (SYNAPT High Definition Mass spectrometer, Waters Corporation, Milford, MA, USA). Mass spectral acquisition was carried out by MALDI survey method. Protein Lynx Global server version (PLGS) 2.4 software (Waters) was used for data processing and database searches. The MS/MS data was searched against the Pin-II protein database constructed separately using the following parameters: peptide tolerance of 20ppm, fragment tolerance of 0.05D, one missed cleavage, carbamidomethylation of cysteines and possible oxidation of methionine.

2.3. Results

2.3.1. Differential regulation of *CanPIs* upon induction

The amplification of cDNA derived from uninduced and induced leaves with *CanPI* gene-specific oligonucleotides yielded transcripts of 789, 614, 445 and 267 bp, representing 4-, 3-, 2- and 1-IRD *CanPIs*, respectively (in PCR with 35 cycles). Reduction in PCR cycles displayed differential accumulation of transcripts among treatments (**Fig. 2.1)**. In comparison to *CanPIs* with 2-IRDs, those with 4- and 3-IRDs showed higher abundance in wounded leaves treated with water or OS (**Fig. 2.1**). In aphid-infested leaves, while the overall expression of *CanPIs* was low, 3-IRD transcripts were relatively prominent. The amplified transcripts were cloned and 60 representative clones from each treatment (25 in case of uninduced) were sequenced to confirm the identities of the *CanPIs.* This analysis detected novel 4-, 3- and 2-IRD subtypes with variations in amino acid composition. Based on this analysis, in addition to the 23 *CanPIs* previously reported (Kim et al., 2001; Shin et al., 2001; Tamhane et al., 2009), 44 novel *CanPIs* comprising eight 2-IRD, thirty-one 3-IRD, and five 4-IRD *CanPIs* were identified (**Table 2.2**). 1-IRD CanPIs detected in uninduced leaves were not observed under any of the induction treatments. Details regarding the treatment-specific representation of *CanPIs* and their IRD composition are summarized in **Table 2.2**. The frequency of occurrence of individual CanPIs out of the 60 clones sequenced per treatment was analyzed and is referred to as abundance in leaf under a particular treatment (**Table 2.2**).

Variation in the abundance of *CanPIs* was apparent (**Fig. 2.2 A**); with 3-IRD *CanPIs* being the highest (from 50 to 70%) either in uninduced or induced leaves. The abundance of 4-IRD *CanPIs* increased in leaves subjected to wounding and treated with OS (38%), compared to aphid-infested (13%) leaves or in wounded leaves treated with water (17%). The proportion of 2-IRD *CanPIs* ranged from the lowest (10%) in wounded leaves treated with OS to the highest (21%) in wounded leaves treated with water. The differential expression of the various subtypes of *CanPIs* (with respect to their IRD composition; see **Table 2.2**) resulted in an induction-specific *CanPI* profile. *CanPI-4, -7, -10, -24,-58* and *-63* showed the highest representation in leaves across all the induction treatments (**Table 2.2**). In wounded leaves treated with

OS, *CanPI-4 and -7* showed the highest frequency, whereas *CanPI-58 and -63* were represented most highly in aphid-infested leaves.

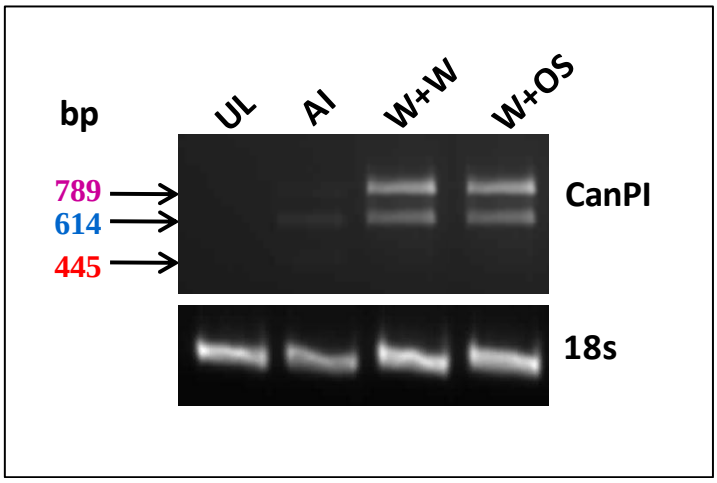


Fig. 2.1 RT-PCR amplification (25 cycles) of mRNA from uninduced (UL) and systemic leaf tissues of induced *C. annuum* infested with aphids (AI), or wounded and treated with water (W+W) or oral secretions (W + OS). Details of primers used are given in **Table 2.1[A]**. cDNA templates were normalized based on *18S rRNA* amplification.

Table 2.2. IRD composition and induction patterns of *C. annuum Pin-II PI* genes

CanPI 

Name	SP	1-IRD	2-IRD	3-IRD	4-IRD	Abundance in leaf			
						UL	AI	W+W	W+S
CanPI-1	5	4	5	10		2	1		
CanPI-2	2	1	16	13					1
CanPI-3	1	1	1	17		3	3	2	1
CanPI-4	1	1	25	17			3	1	11
CanPI-5	4	1	1	17		6	1	3	3
CanPI-6		2	25	17	p	ND	ND	ND	ND
CanPI-7	5	4	14	5	10	1	5	1	12
CanPI-8	5	4	14	3	10			2	3
CanPI-9	1	1	25	5	10				
CanPI-10	5	4	14	5	8	2		3	6
CanPI-11	1	4	14	5	10			3	1
CanPI-12		1	1	11	p	ND	ND	ND	ND

***CanPI*-13**	**1**	17				1			
***CanPI*-14**	**1**	6				ND	ND	ND	ND
***CanPI*-15**	**5**	7				2			
***CanPI*-16**	**1**	1	17			1			
***CanPI*-17**	**1**	12	17			1	1	1	1
***CanPI*-18**	**1**	25	17			1			
***CanPI*-19**	**1**	1	25			ND	ND	ND	ND
***CanPI*-20**	**1**	25	17						3
***CanPI*-21**	**5**	4	17			1			
***CanPI*-22**	**5**	4	9				1	2	2
***CanPI*-23**	**3**	25	17			1			
***CanPI*-24**	1	1	18				5	4	
***CanPI*-25**	3	1	18				2		
***CanPI*-26**	5	4	10				4	1	
***CanPI*-27**	8	20	19					1	
***CanPI*-28**	5	23	18					1	
***CanPI*-29**	8	22	18					1	
***CanPI*-30**	6	21	18					1	
***CanPI*-31**	7	1	17					1	
***CanPI*-32**	5	4	1	12	28		1		
***CanPI*-33**	1	1	12	27	17		1		1
***CanPI*-34**	5	4	14	26	10		1		
***CanPI*-35**	5	4	14	5	10		1		
***CanPI*-36**	8	24	1	12	17			1	
***CanPI*-37**	1	1	12	46				1	
***CanPI*-38**	1	1	12	17				1	
***CanPI*-39**	1	30	17	51				1	
***CanPI*-40**	1	12	41	18				1	
***CanPI*-41**	8	1	12	17				1	
***CanPI*-42**	8	31	1	18				1	
***CanPI*-43**	8	5	35	50				1	
***CanPI*-44**	8	5	36	18				1	
***CanPI*-45**	8	53	37	47				1	
***CanPI*-46**	8	5	37	49				1	
***CanPI*-47**	8	5	37	18				1	
***CanPI*-48**	9	5	38	18				1	
***CanPI*-49**	10	32	39	18				1	
***CanPI*-50**	8	33	37	48		1		2	3
***CanPI*-51**	11	5	40	17				1	
***CanPI*-52**	1	1	34	17				1	
***CanPI*-53**	1	1	54	17				1	
***CanPI*-54**	1	1	43	44				1	
***CanPI*-55**	12	29	25	17			1		
***CanPI*-56**	8	5	12	17			1		
***CanPI*-57**	5	5	37	18			1		
***CanPI*-58**	8	5	37	18			10	4	5
***CanPI*-59**	5	4	1	17			1		
***CanPI*-60**	5	4	42	18			1		

CanPI-61	8	1	40	17			1		
CanPI-62	1	7	52	45			1		
CanPI-63	1	1	12	17		2	11	4	4
CanPI-64	1	1	12	17			1		
CanPI-65	13	1	12	18				2	
CanPI-66	1	4	37	18			1	1	1
CanPI-67	1	1	1	55				1	2

p **Partial sequence**
ND **Not detected**

Table 2.2: The general architecture of a Pin-II PI gene is displayed on the top of the table. A typical *CanPI* gene consists of a signal peptide (SP) followed by 1, 2, 3 or 4 IRDs interconnected by linkers. The *CanPI* genes listed are colour coded as follows: green for 1-IRD type; red for 2-IRD type; blue for 3-IRD type and pink for 4-IRD type. SP denotes signal peptide and the numbers indicate any of the 13 sequence variants. The position of IRDs is indicated at the top of the columns and the numbers indicate the occurrence of any of the 55 IRDs at a given position. *CanPI-1* and *-2* are from Kim et al. (2001) and Shin et al. (2001), respectively while *CanPI-3* to *-23* are from Tamhane et al. (2009). *CanPI-24* to *-67*were identified in the present work. The frequency of occurrence of individual *CanPIs* in leaves from 60 clones analyzed from each induction treatment is given as abundance in leaf on the right side of the table. The nucleotide sequences of *CanPI-24* to *-67* have been submitted to NCBI and the Accession numbers have been given in the Appendix II.

Wounded leaves treated with water showed the highest number of expressed *CanPIs* (40) as well as a wide representation of several unique *CanPIs* (24 in number) in low frequencies (**Table 2.2**; **Fig. 2.2 B**). Responses to wounding with OS appear more specialized as suggested by the expression of a few *CanPIs* (17) in high frequencies (**Table 2.2**, **Fig. 2.2 B**). Both the treatments share more *CanPIs* (3- and 4-IRD type) possibly due to the standardization of the amount of wounding between these two treatments. Aphid-infested leaves also accumulated transcripts of 13 unique *CanPIs* representing a diverse array of PIs. Only transcripts specific for *CanPI-3*, *-5*, *-7*, *-8* and *-10* could be analyzed from these tissues. *CanPI-7* (4-IRD) showed constitutive expression in uninduced as well as induced leaves, while *CanPI-3* and *-5* (3-IRD PIs) showed low accumulation levels in wounded leaves treated with OS.

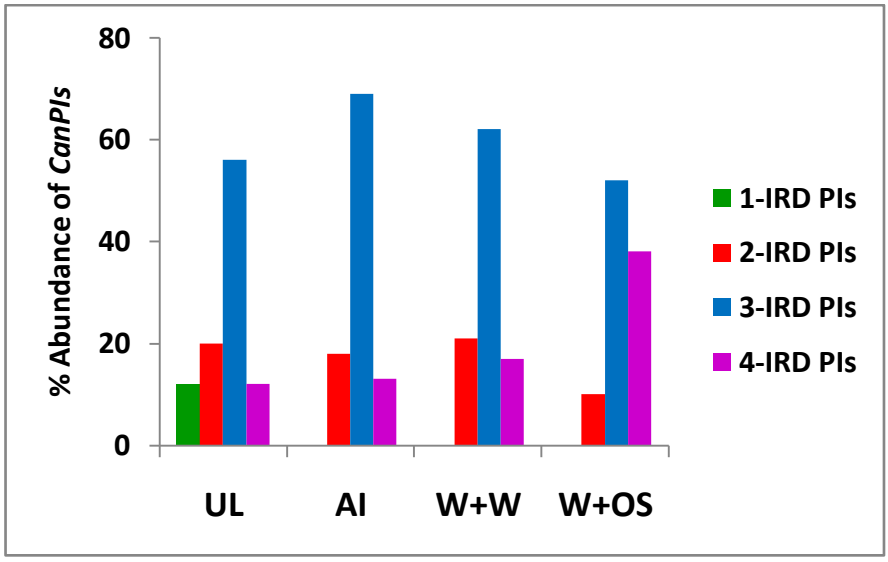


Figure 2.2A: Abundance of 1-, 2-, 3- and 4-IRD *CanPIs* upon each induction treatment to *C. annuum* leaves. Abundance represents the frequency of a specific type of *CanPI* in the total clones sequenced per induction. The up-regulation of 4-IRD *CanPIs* in W+OS is evident.

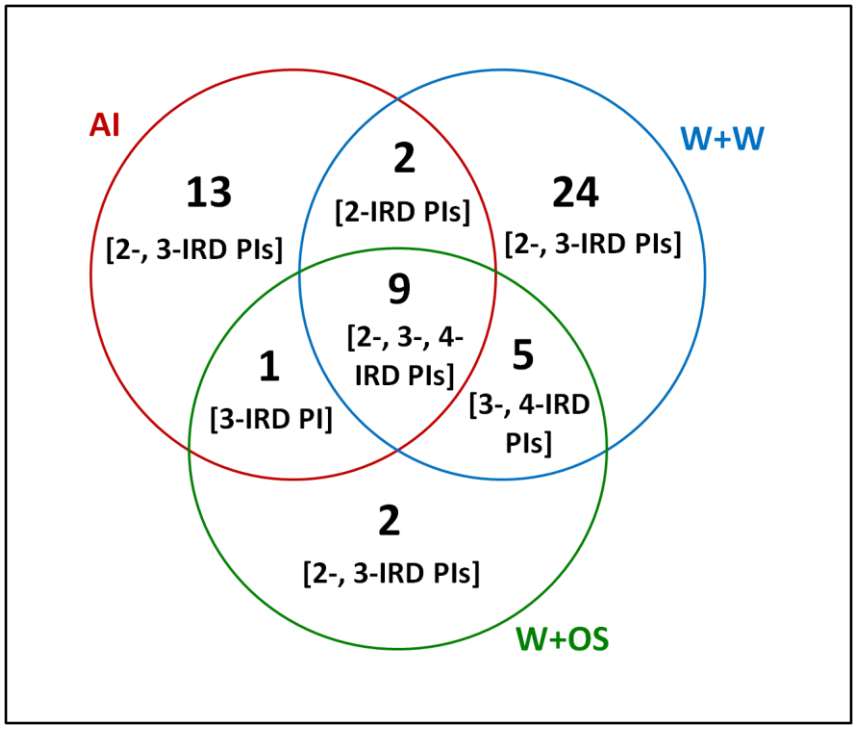


Figure 2.2B: Comparison of *CanPI* expression patterns in *C. annuum* leaves after different inductions. The Venn diagram presents the number of *CanPIs*, common or differentially expressed upon each type of induction; AI, W+W, W+OS. W+W showed the highest number of expressed *CanPIs*. The numbers within bracket represent type of *CanPIs* within that group.

CanPI-8 and *-10* (4-IRD PIs) were differentially expressed in wounded leaves treated with water or OS, and completely absent in aphid-infested leaf tissues (**Fig. 2.2 C**).

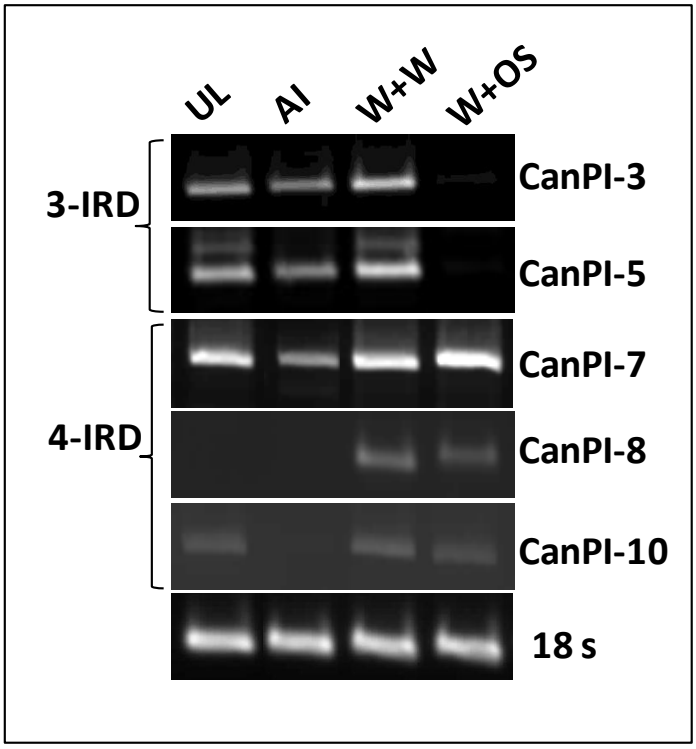


Figure 2.2C: Expression analysis of selected *CanPIs* in un-induced and systemic leaf tissue of induced *C. annuum*. The details of gene-specific primers are given in **Table 2.1[B].**

The Protein Prowler predictor (Boden and Hawkins, 2005) revealed ER signal peptides in all *CanPIs*. The ER signal peptide (SP) sequences of 25 aa at the N-terminal of *CanPIs*, which showed thirteen variants, were named SP-1 to SP-13 (**Fig. 2.2D**). SP-3, a short polypeptide of 16 aa, was present in *CanPI-23* and *-25* (**Table 2.2**). SP-1, -5 and -8 were prevalent in many *CanPIs*, whereas others were present in single *CanPIs*. To investigate the interrelationships and groupings within *CanPIs* and with out-groups, phylogenetic analysis was carried out. *CanPIs* formed a distinct cluster from full-length Pin-II PI (4-IRD) from *Nicotiana benthamiana* which was used as an out-group. The dendrogram revealed clustering based on the number of IRDs and identical component IRDs in *CanPIs* (**Fig. 2.3**). Distinct clusters of 4-, 3-, 2- and 1-IRD PIs formed with some intermixing for e.g. CanPI-8 and -32. CanPIs showing more aa sequence similarity associated close to each other for e.g. CanPI-7 and -10, CanPI-34 and -35, CanPI-33 and -36 and CanPI-1 and -60.

Figure 2.2D: Multiple sequence alignment of deduced aa sequences of signal peptides (SP-1 to SP-13) of *CanPI* genes displaying variations.

The deduced amino acid sequences of *CanPIs* (67 in total) consisted of 55 unique IRDs; of these, 11 had chymotrypsin inhibitory (CI) and 44 had trypsin inhibitory (TI) reactive sites. The multiple sequence alignment of IRDs revealed variation mostly within reactive site loops or towards the C-terminal ends (**Fig. 2.4A**). Five TI IRDs possess 'Lys' while 39 had 'Arg' at the P1 position of the reactive site. CI IRDs had 'Leu' at P1, except IRD 22 which had 'Pro'. IRD 48 (TI) had active site variation with 'Asn' being replaced with 'Asp'. Another crucial difference was variation in the number and positioning of cysteine residues per IRD. Seven such cysteine variants, namely IRD 9, 11, 24, 35, 38, and 43, were TIs, and IRD 33 was a CI (**Fig. 2.4A**). The dendrogram analysis of the unique IRDs represented their clustering based on amino acid sequence variation (**Fig. 2.4B**).

The relative abundance of IRDs that resulted from various types of inductions was analyzed (**Fig. 2.5A**). Representation of IRDs 1, 4, 5, 10, 12, 14, 17, 18, 25 and 37 was high in induced tissues and exhibited treatment specificity. IRDs 1, 12, 17, 18 and 37, which are TIs, showed high frequency in aphid-infested leaves and wounded

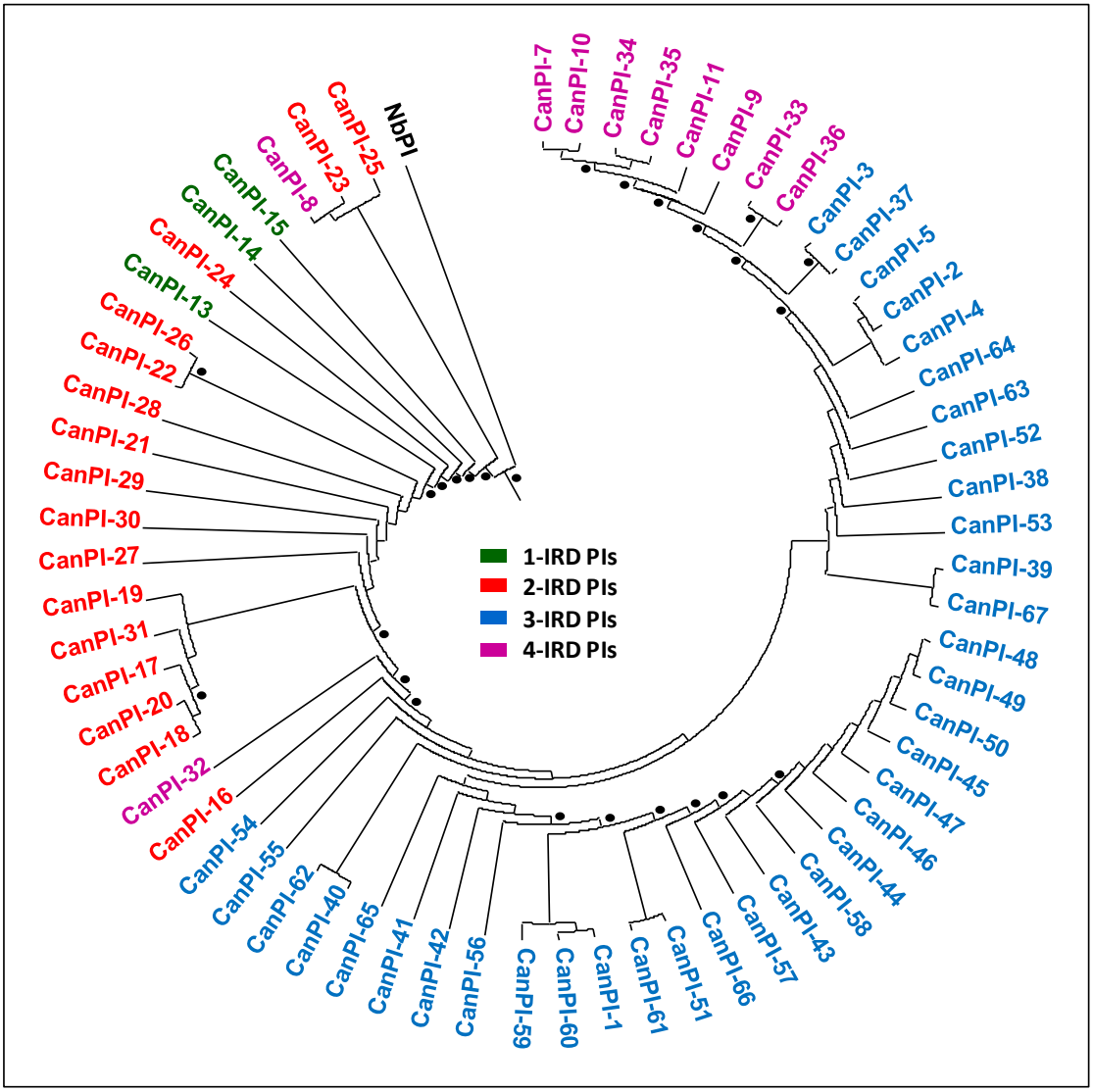


Figure 2.3: Dendrogram of *CanPIs* based on deduced amino acid sequences of full-length genes of 1- to 4-IRDs isolated from *C. annuum*. Pin-II type PI from *N. benthamiana* (*NbPI*, NCBI: ABA42892) is used as an out-group. Details of the *CanPIs* (1 to 67), composition, number of IRDs, tissue-wise representation and color coding are given in Table 2.2. • denotes bootstrap value ≥80 as calculated by Lasergene software.

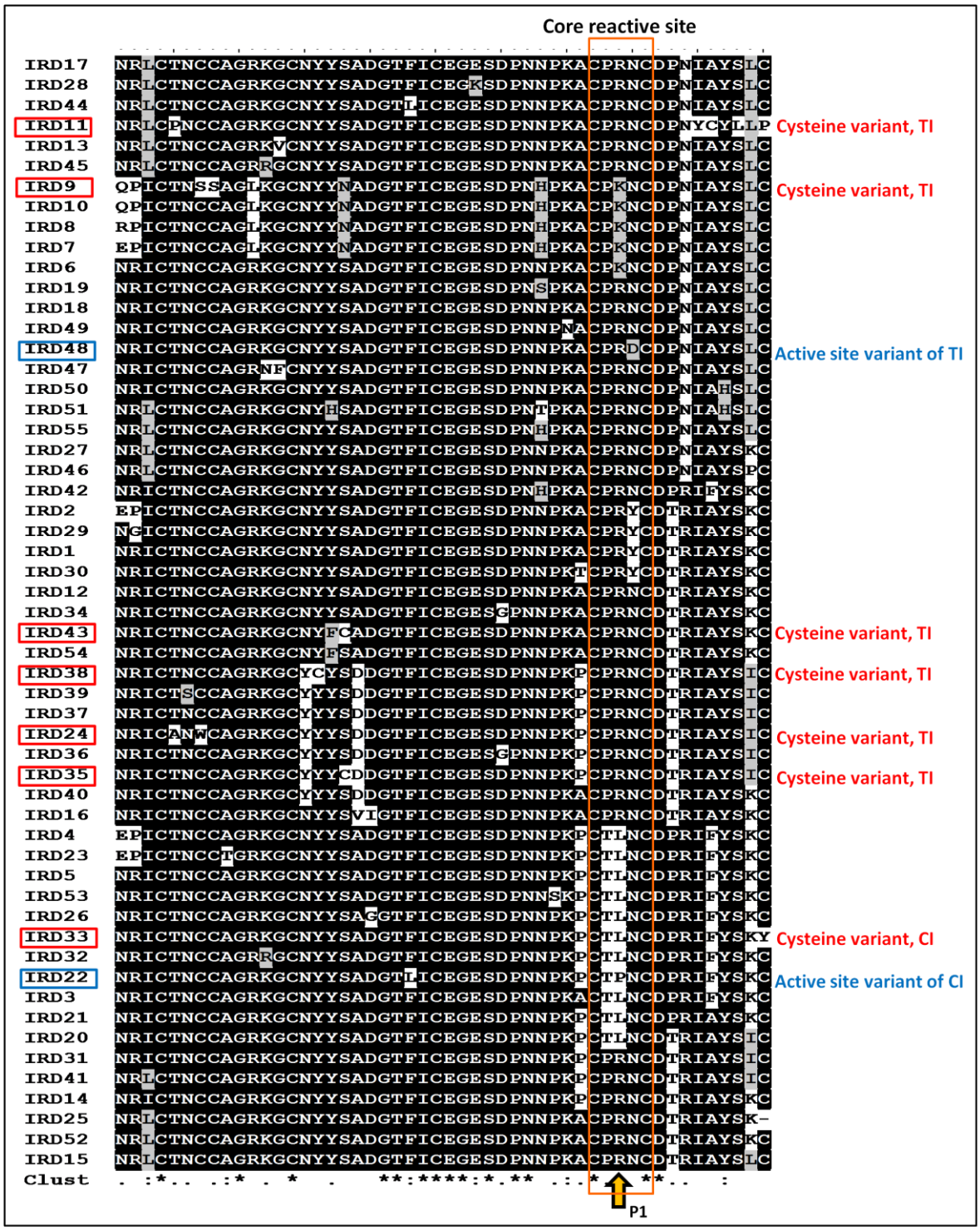


Figure 2.4A: Multiple sequence alignment of deduced aa sequences of IRDs (55 in number) constituting all the *CanPI* genes. The reactive site residue P1 is marked by an arrow. Presence of Lys (K) or Arg (R) at P1 site, indicates trypsin inhibitory site while Leu (L) indicates chymotrypsin inhibitory site. The core reactive site is marked by an orange box. IRD 1/IRD 18 represents a typical trypsin inhibitory domain (TI) and IRD 5 represnets a typical chymotrypsin inhibitory domain (CI). The selected IRDs showing variation in the number of cysteines or active site residues are marked by red and blue boxes, respectively.

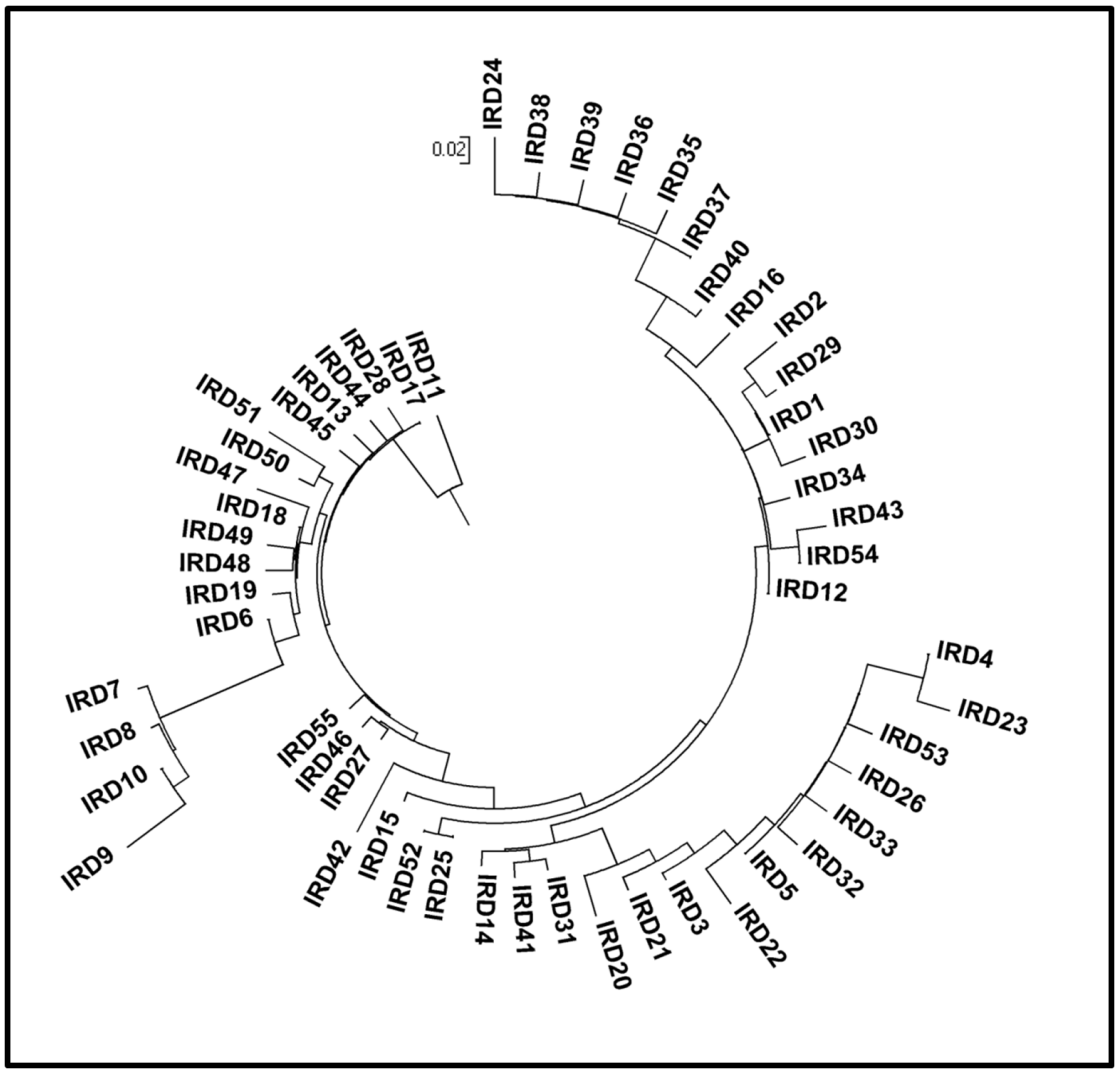


Figure 2.4B: Dendrogram of 55 unique IRDs based on deduced aa sequences from full-length *CanPI* genes. Represents the clustering among them depending on aa sequence variation.

leaves treated with water, whereas the abundance of IRDs 4 and 5, which are CIs, and IRDs 10, 14 and 25, which are TIs, was distinct in wounded leaves treated with OS (**Fig. 2.5A**). Most of the remaining diversity of IRDs is contributed by wounded leaves treated with water, which show presence of several IRDs at lower frequencies (**Fig. 2.5B**).

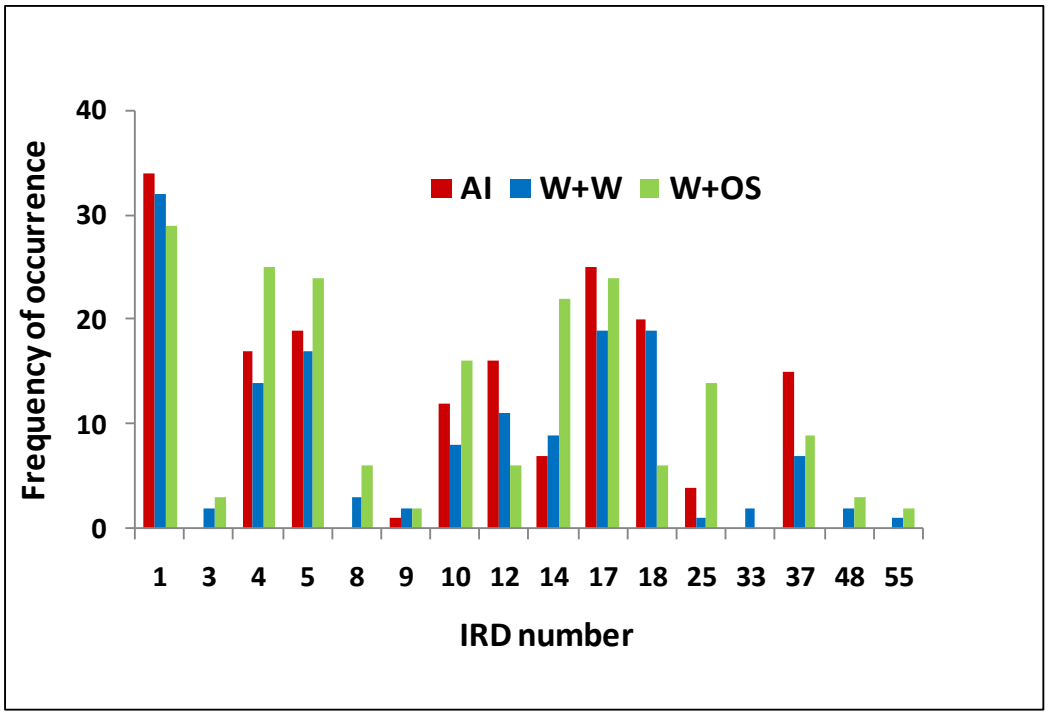


Figure 2.5A: The frequency of occurrence for individual IRDs per treatment. IRDs occurring only once were excluded from this graph. The highest representation of IRDs 1, 4, 5, 17, 18 across all treatments is apparent.

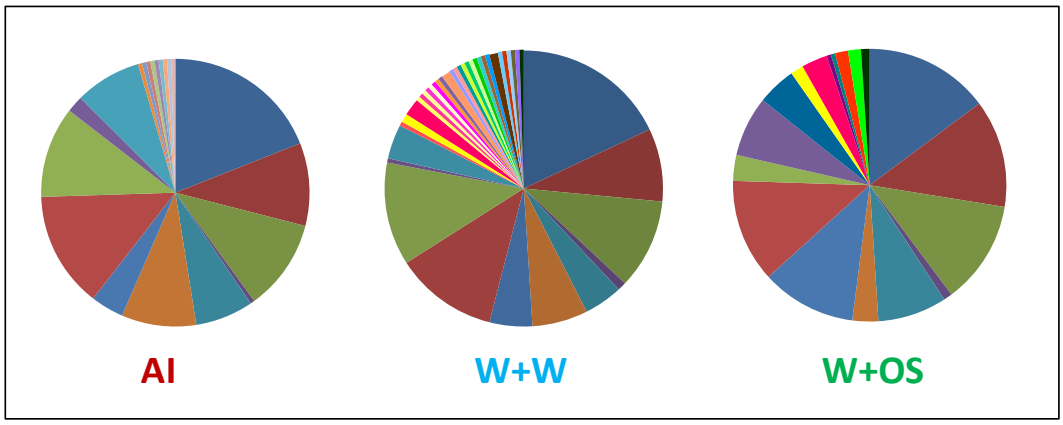


Figure 2.5B: Pie charts represent the distribution of IRDs under each induction. A diverse array of unique IRDs is observed in wounded leaves treated with water.

2.3.2. Wounding and insect damage results in quantitative and qualitative changes in CanPI proteins

C. annuum leaves were found to have increased PI activity upon induction (**Fig. 2.6A**). Significantly higher level of PI activity was evident in the systemic leaves of wounding treated with water induction and both, the local and systemic leaves

induced with wounding treated with OS compared to similar leaves from unwounded control plants (**Fig. 2.6A**). The PI activity in aphid-infested leaves was higher than that in uninduced leaves; however, it was 1.5- and 3.5-fold less than that in wounded leaves treated with water or with OS, respectively. Induced levels of PI activity ranged from 2- to 4-fold in leaves wounded and treated with water and treated with OS, respectively.

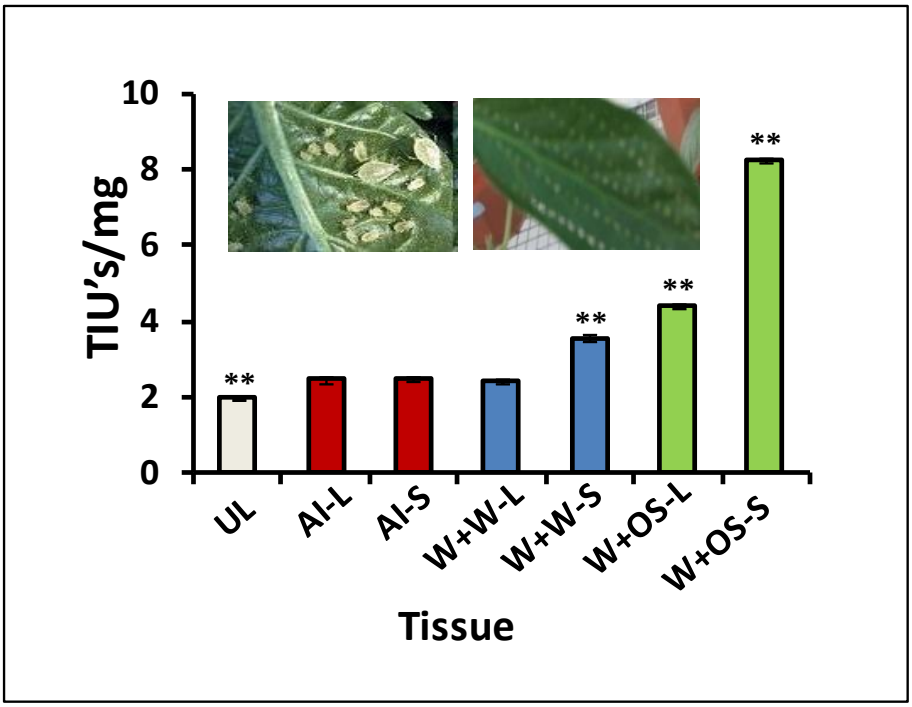


Figure 2.6A: The accumulation of trypsin inhibitory activity in leaves infested with aphids (AI), wounded and treated with water (W+W) or oral secretions (W+OS). Local (L) and systemic (S) tissues corresponding to the respective induction are represented. The W+OS systemic tissue shows statistically significant higher trypsin inhibitory units (TIUs) (Tukey's *t*-test at $p<0.01$).

A differential pattern of PI isoform induction was observed in *C. annuum* leaves in response to various treatments (**Fig. 2.6B**).Three prominent CanPI activity bands were detected in leaves that were wounded and treated with OS or with water, while only two PI isoforms could be detected in case of aphid infested and uninduced leaves. This difference indicated induced qualitative diversity in the CanPIs that resulted from these two treatments. Extracts from leaves wounded and treated with water or OS attained an early saturation of HGP inhibition (70%) unlike aphid-infested and uninduced tissues, consistent with the quantitative/qualitative differences amongst the PIs in their activity (**Fig. 2.6C**). The HGP of fourth-instar larvae

displayed at least seven protease isoforms (HGP-1 to -7), of which only HGP-6 and -7 were able to retain marginal activity in the presence of CanPIs from either uninduced or induced tissues (**Fig. 2.6D**).

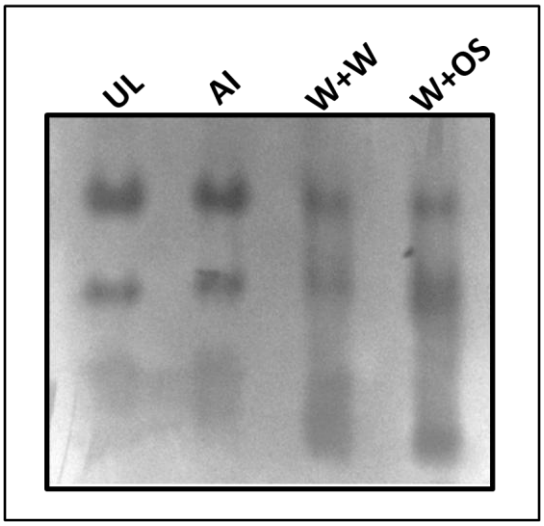


Figure 2.6B: *In-gel TI profiles of induced *C. annuum* leaves. Three TI bands were visualized in W+W and W+OS while only 2 in uninduced and aphid infested leaves.

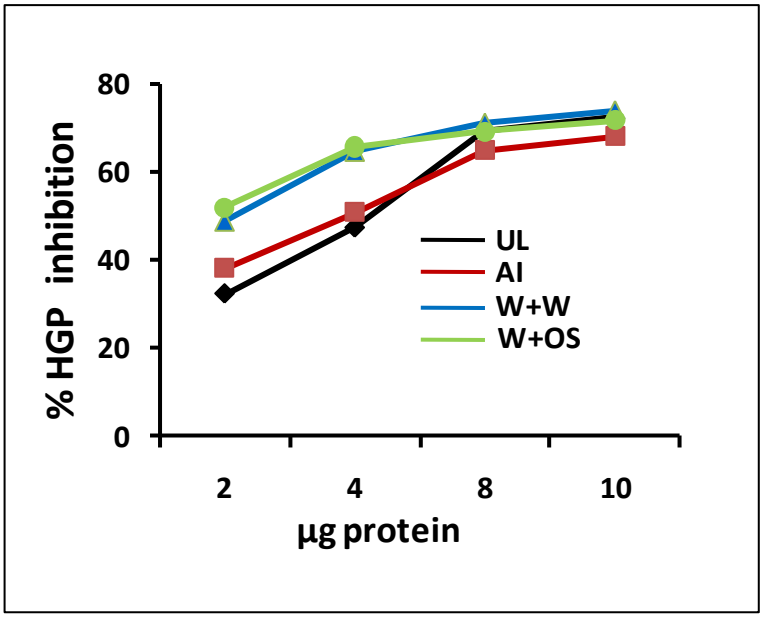


Figure 2.6C: *The percentage inhibition of HGP by PIs from leaf extracts estimated using azocasein as substrate.

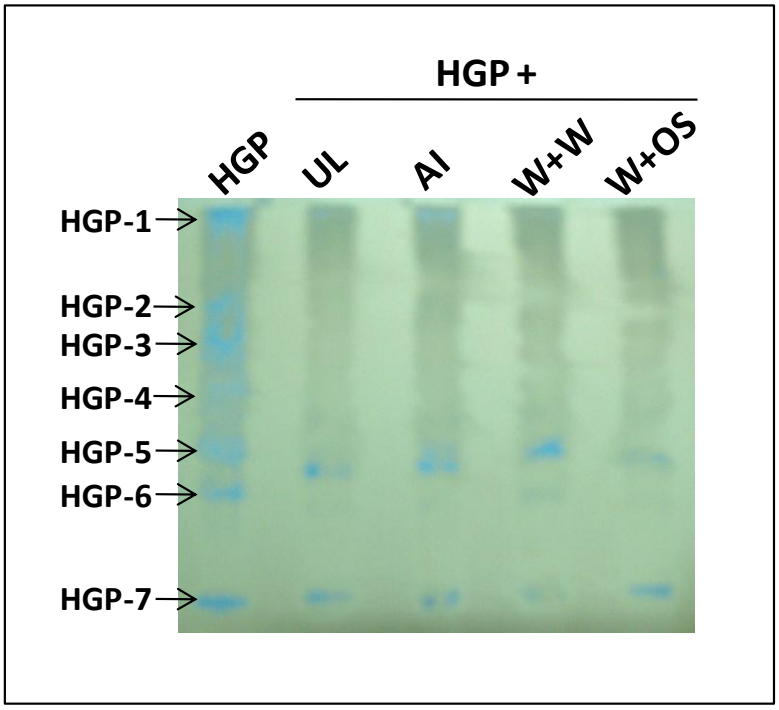


Figure 2.6D: *The inhibition of HGP isoforms by leaf extracts. Equal HGPI units of each of the extracts were incubated with HGP for 1 hr at 25 ºC and visualized for residual protease activity.

* The represented figures of induced tissues belong to systemic tissues of the particular induction treatment. Similar observations for local tissues of induction treatments.

The 2-D activity profiles of induced leaf samples showed the induction of several novel PI isoforms in the range of pI from 4 to 7 and also shift in the isoelectric point (pI) and/or molecular mass of few isoforms (**Fig. 2.7**). Five prominent TI isoforms (TI-1 to TI-5) corresponding to 1-, 2-, 3- and 4-IRD were identified in uninduced leaves, in the pI range of 5 to 7 (**Fig. 2.7**). TI-6 (1-IRD) isoform with major basic shift in pI compared to TI-1, was present in all the three types of induced tissues. TI-3, -4 and -5 were consistently detected in aphid-infested leaves but were absent from wounded leaves treated with water or OS. TI-8 to -13 were present only in wounded leaves treated with water, while TI-14, -15, and -16, corresponding to 4-IRD CanPIs, were present in wounded leaves treated with OS only. The induced PI activity in wounded leaves treated with water showed a distribution of several TI isoforms with low intensity, whereas the up-regulation of only a few TI isoforms was evident in wounded leaves treated with OS. TI-2 was absent in all the induced tissues, consistent with the specificity of some PI isoforms in uninduced leaves. Among the TI

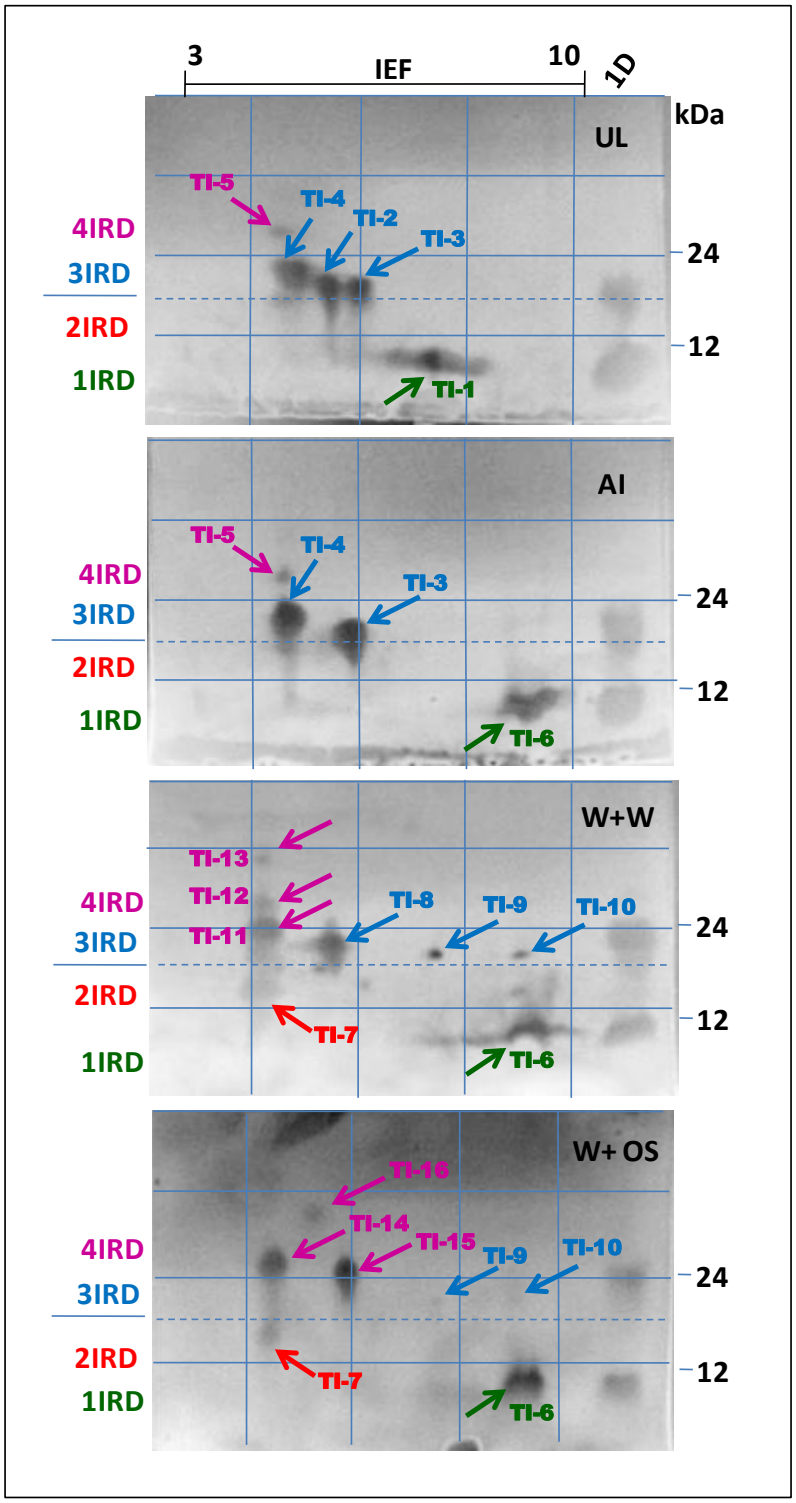


Figure 2.7: 2D-TI activity profiles of uninduced and induced *C. annuum* leaves. Equal TI units from leaf extracts were visualized by GXCT after separation by 2-D. Differential TI activity isoforms, displaying charge and mass variations, were observed in induced leaves. Grids are superimposed across the gel images to help visualize the major shifts in the molecular mass and pI of the induced TI isoforms. The represented figures of induced tissues belong to systemic tissues of the particular induction treatment. Similar observations for local tissues of induction treatments.

isoforms corresponding to 1-, 2-, 3- or 4-IRD CanPIs, more variations were observed for 3-and 4-IRD CanPIs.

Partially purified PIs contained small peptides of about 5.5 to 6.3 kDa equivalent to a single IRD as analyzed on MALDI-TOF-MS (**Fig. 2.8**). High molecular mass proteins exhibited very low intensity in the mass spectra, perhaps due to ion suppression effects, and therefore, are not considered. Uninduced leaf extracts displayed a single major peak of 5583 Da, whereas aphid-infested leaves had major peaks at 5583 Da and 5616 Da and few low intensity peaks. A peak of 5583 Da was most prominent in wounded leaves treated with water, in addition to several peaks of 5616, 5760 and 5832 Da with high intensity. Extract from leaves wounded and treated with OS displayed 6138 and 5961 Da as the major peaks, with low intensity peaks at 5616, 5832, 6036 and 6301 Da and an absence of the 5583 Da peak. This variation in the molecular masses of peptides is likely a result of the proteolytic processing of the precursor PI proteins to generate multiple functional PI species (equivalent to single IRD from multi-IRD CanPIs) (Horn et al., 2005; Huesgen and Overall, 2011).

The partially purified PI protein from uninduced and induced leaves, displaying varying mass spectral profiles on MALDI-TOF-MS, were separated on Tricine gel, and the proteins were individually excised and processed for in-gel digestion followed by the identification of peptides by Q-TOF-MALDI-TOF-MS/MS (**Table 2.3**). The 6 kDa protein in all the extracts showed matches to Pin-II proteinase inhibitors from *C. annuum*. However, due to high homology amongst IRD sequences, the database generated identification to several CanPIs/IRDs (**Table 2.3)**.

2.3.3. Tissue specificity of PI accumulation

Flowers, followed by stems and early fruit showed a significantly higher level of PI activity compared to that from leaves, roots and the different developmental stages of fruit (**Fig. 2.9A**). Flower tissue showed the highest PI activity, while tissue from turning fruit exhibited the least activity with a 7-fold difference. The in-gel PI visualization after resolution on 2-DE for various *C. annuum* plant parts displayed the qualitative variations in the PI activity across these tissues (**Fig. 2.9B**). The clusters corresponding to 1-, 2-, 3- and 4-IRD CanPIs were prominent in stems, early fruit and flowers. In accordance to the very low TIUs in roots, only one TI isoform which corresponded to 2-IRD CanPI was detected. The 1-IRD cluster contained several

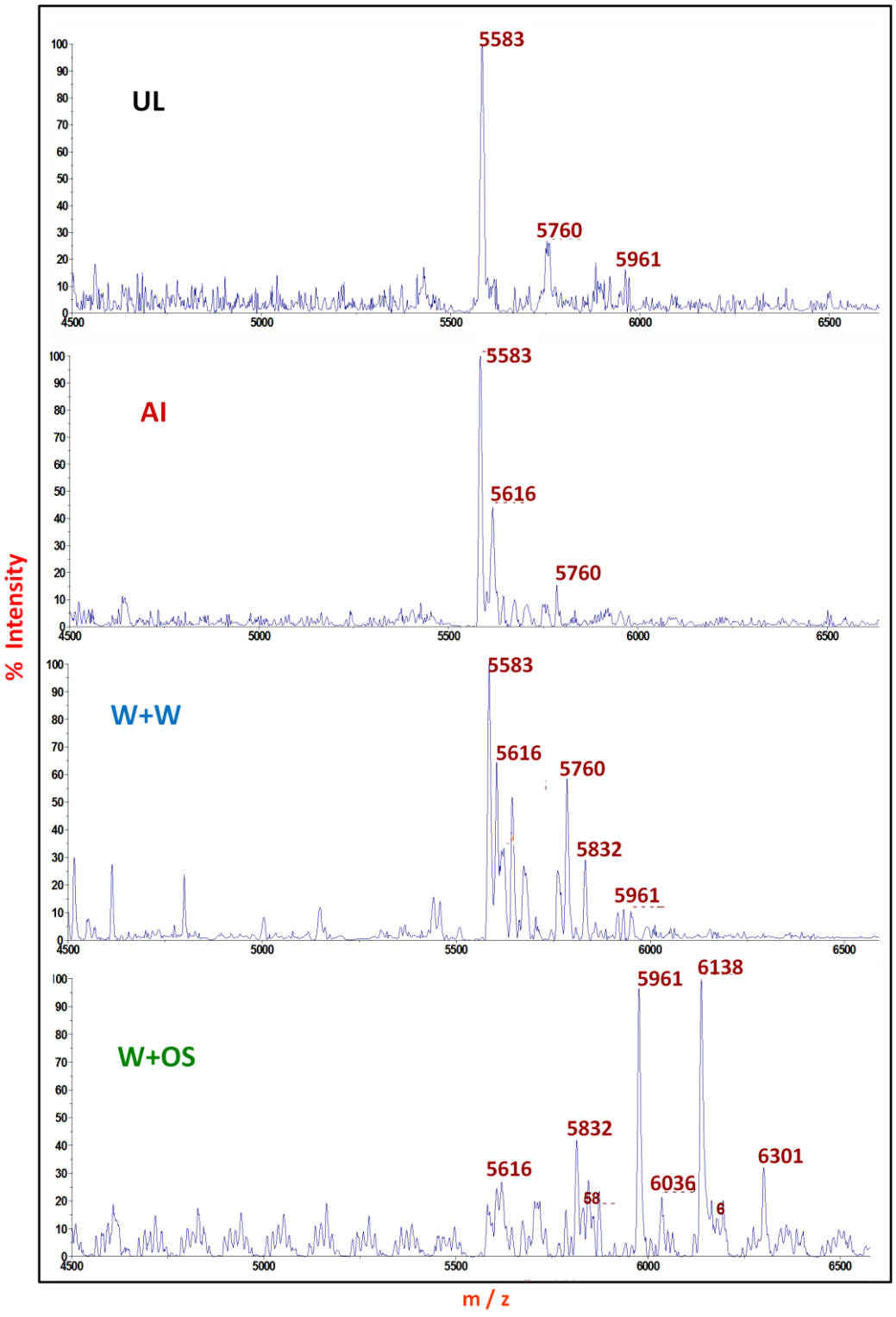


Figure 2.8: MALDI-TOF-MS characterization of partially purified PIs from leaf extracts. Mass spectral analysis revealed peaks of varying masses in the range of 5.5 to 6.3 kDa across treatments. These represent an 1-IRD peptide, which was further confirmed by MS/MS analysis. The increased diversity of processed IRDs is prominent in *C. annuum* leaf tissues wounded and treated with water (W+W) or oral secretions (W + OS) as compared to aphid infested.

Table 2.3: Protein identification by Q-TOF-MALDI-TOF-MS/MS, database searches

Tissue	Accession No.	Identified protein	Protein identification data		
			(No. Pep./coverage %)	PLGS Score	Peptide sequence (MS/MS)
Uninduced	Q4ZIQ3	Pin II type proteinase inhibitor 9 OS Capsicum annuum Pin-II PI 21, 17, 20, 13, 23, 6, 4, 3, 5, 16, 12, 14	15/ 45	2601	(R)KGCNYYSADGTFICEGESDPNNPK(P) (K)GCNYYSADGTFICEGESDPNNPK(P) (R)LCTNCCAGR(K) (R)KGCNYYSADGTFICEGESDPNNPKPCTLNCDPR(I) (K)GCNYYSADGTFICEGESDPNNPKPCTLNCDPR(I) (R)KGCNYYSADGTFICEGESDPNN(P) (R)KGCNYYSADGTFICEGESDPNNPK(P) (K)GCNYYSADGTFICEGESDPNNPK(P) (A)DGTFICEGESDPNNPK(P) (I)CEGESDPNNPK(P) (D)PNNPK(P) (R)LCTNCCAGR(K) (L)CTNCCAGR(K) (C)TNCCAGR(K) (R)KGCNYYSADGTFICEGESD(P)
Aphid infested (systemic tissue)	Q4Z8K3	Pin II type proteinase inhibitor 3 OS Capsicum annuum Pin-II PI 6, 21, 20, 13, 23, 4, 17, 9, 5, 16, 19, 14, 12	18/ 51	718	(N)RICTNCCAGR(K) (R)KGCNYYSADGTFICEGESDPNNPK(A) (R)LCTNCCAGR(K) (K)GCNYYSADGTFICEGESDPNNPK(A) (R)SEGNAENRICTNCCAGR(K) (K)VSFLAFLLVLEILLLHVDAKACSEENAENR(I)

					(R)NCDPNIAYSLCLYEK(-) (N)RICT(N) (N)RICTN(C) (N)RICTNC(C) (T)NCCAGR(K) (N)CCAGR(K) (C)CAGR(K) (Y)YSADGTFICEGESDPNNPK(A) (R)LCTNCCAGR(K) (L)CTNCCAGR(K) (C)TNCCAGR(K) (N)RIC(T)
Wounding + water (systemic tissue)	D2CGT4	Pin II type proteinase inhibitor 17 OS Capsicum annuum Pin-II PI 6, 21, 20, 13, 23, 4, 3, 9, 5, 16, 19, 8, 14, 12	17/ 60	718	(N)RICTNCCAGR(K) (R)KGCNYYSADGTFICEGESDPNNPK(A) (R)LCTNCCAGR(K) (K)GCNYYSADGTFICEGESDPNNPK(A) (K)VSFLAFLLVLEILLLHVDAKACSEENAENR(I) (R)NCDPNIAYSLCLYEK(-) (N)RICT(N) (N)RICTN(C) (N)RICTNC(C) (T)NCCAGR(K) (N)CCAGR(K) (C)CAGR(K) (Y)YSADGTFICEGESDPNNPK(A) (R)LCTNCCAGR(K)

					(L)CTNCCAGR(K) (C)TNCCAGR(K) (N)RIC(T)
Wounding + OS (systemic tissue)	D2CGT4	Pin II type proteinase inhibitor 17 OS Capsicum annuum Pin-II PI 21, 19, 20, 23, 13, 6, 4, 3, 9, 5, 16, 12, 8, 14	12/ 29	603	(R)KGCNYYSADGTFICEGESDPNNPK(A) (K)GCNYYSADGTFICEGESDPNNPK(A) (R)LCTNCCAGR(K) (N)RICTNCCAGR(K) (N)RICTN(C) (N)RICTNC(C) (N)CCAGR(K) (C)CAGR(K) (R)KGCNYYSADGTFICEGESDPNN(P) (R)LCTNCCAGR(K) (C)TNCCAGR(K) (N)RIC(T)

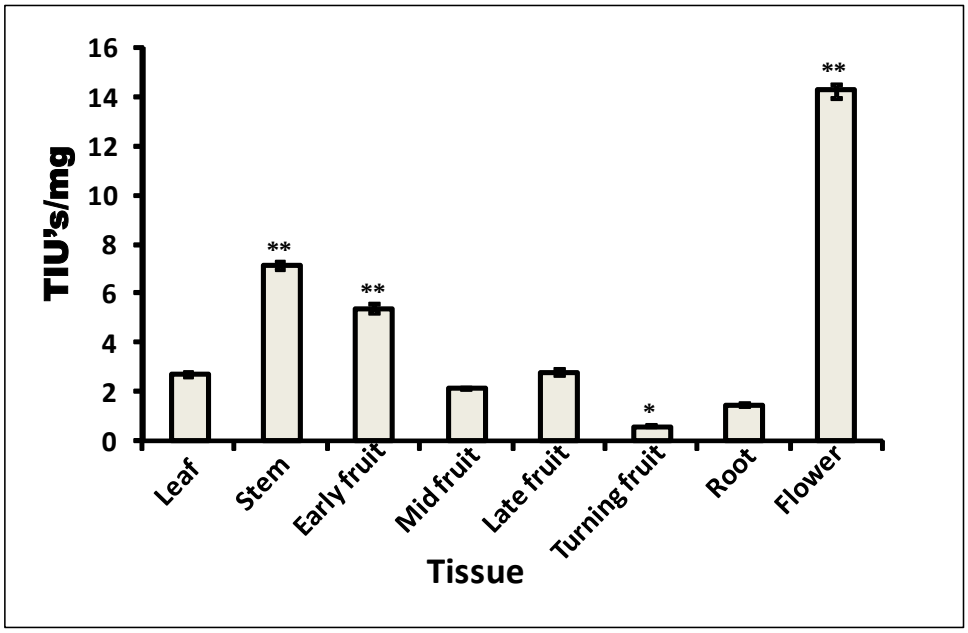


Figure 2.9A: Tissue-specific TI activity in various tissues of a mature *C. annuum* plant. Flower tissue shows the highest TIUs. Stem and early fruit tissue also shows significantly higher TI activity.

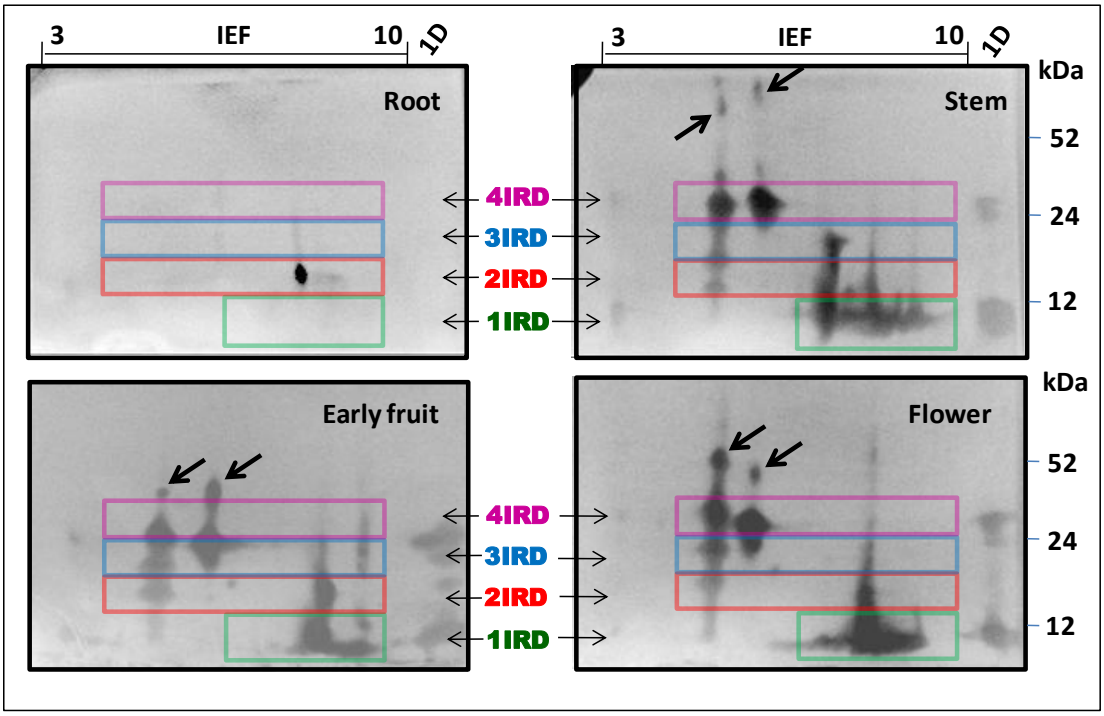


Figure 2.9B: 2D-TI activity profiles of various *C. annuum* plant parts. Qualitative and quantitative variations in TI profiles of tissue from stem, early fruit and flower are quite evident and the differences are highlighted through the clusters of 1-, 2-, 3- and 4-IRD PIs.

merged TI isoforms and indicated multiple charge variants due to amino acid sequence differences. Stem tissue in particular showed less diversity in the 3-IRD cluster while exhibiting the highest number of charge variants in the 1-IRD cluster. Stem, early fruit stage and flower tissues revealed the presence of the high molecular mass isoforms of PIs (**Figure 2.9B**, marked by arrows) that are predicted to be larger than those of 4-IRD CanPIs.

2.4. Discussion

Among Solanaceae, various species of *Nicotiana* display PI genes containing 2 and 4 to 8 IRDs (Wu et al., 2006). In addition, 2-IRD PIs from tomato (*Solanum lycopersicum)* and *S. nigrum* have been well characterized (Graham et al., 1985; Hartl et al., 2010). Simultaneously expressed PIs with varying IRD composition have been reported in *N. alata* stigma (4- and 6-IRDs; Miller et al., 2000), in *N. glutinosa* infected with TMV (6- and 8-IRD; Choi et al., 2000) and in *N. attenuata* in response to herbivory (7-IRD; Horn et al., 2005). In addition to the previously reported 23 PI genes from *C. annuum* pericarp, developing fruit and stem (Kim et al., 2001; Shin et al., 2001; Tamhane et al., 2009), 44 new *CanPIs* were isolated and characterized from the induced leaves in the present study. 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 diversity. Consistent with the previous studies, 3-IRD PIs were highly abundant in *C. annuum* leaves (**Table 2.2**, **Fig. 2.2A**). A strong up-regulation of 3- and 4-IRD *CanPI* transcripts was observed in induced leaves when compared to uninduced leaves (**Fig. 2.1**, **Table 2.2**). Reciprocating patterns obtained at the PI activity level corroborated the induction specific regulation of CanPIs. Other reports on *C. annuum* have also demonstrated an increase in *CanPI* expression in local and systemic leaf tissues upon elicitation by aphids, viruses, insect feeding and mechanical wounding (Kim et al., 2001; Moura and Ryan, 2001; Tamhane et al., 2009).

Despite the difficulties of directly comparing the aphid infestation treatment with other treatments carried out in the present study, an interesting pattern of accumulated *CanPIs* was evident at the transcript and the PI activity level. *CanPI* transcript abundance in aphid infested leaves was much lower than that observed in wounded leaves treated with water or OS (**Fig. 2.1**). Aphids have been found to elicit defense related genes including PIs but the responses are low as compared to that

elicited by chewing insect attack (Voelckel et al., 2004; Zhu-Salzman et al., 2004). However, studies have also found aphid induced transcriptional signatures of salicylic acid signalling and no increase in PI transcripts (Fidantsef et al., 1999). In the present study, transcripts of *CanPI-58* and *-63* were highly accumulated under aphid infestation while *CanPI-8* and *-11* remain either un-elicited or suppressed by aphid damage; the latter being up-regulated by wounding or W+OS treatment (**Table 2.2**). The regulated expression of *CanPI-8* and *-10* (4-IRD type PIs) in wounded leaves treated with water or OS (**Fig. 2.2C**) supported their specialization in plant defense upon high cellular damage. The suppression of selected wound induced responses, rather than the lack of cellular damage, might also be responsible for the low responses to aphid attack. Aphid infestation induced CanPI activity though much less as compared to wounded leaves treated with water or OS (**Fig. 2.6A**) a unique aphid induced CanPI signature was evident in the 2D activity profiles (**Fig. 2.7**).

Particularly, 4-IRD PIs were strongly induced in wounded leaves treated with OS (**Fig. 2.2A**) highlighting the strong and specific effects of insect elicitors on CanPI regulation. Plant responses to wounding/insect feeding are known to be specifically altered by the plant's perception of herbivore-specific elicitors (Baldwin et al., 2001; Voelckel et al., 2001; Thaler, 2004; Diezel et al., 2011). High amounts of jasmonic acid and the rapid accumulation of wound-inducible transcripts have been reported in response to insect damage or insect OS when compared to only mechanical wounding (Korth and Dixon, 1997; Reymond et al., 2000; Walling, 2000). Wounded leaves treated with water showed the highest number (24) of uniquely expressed *CanPIs* and IRDs (**Fig. 2.2B**, **Fig. 2.5B**), though in lower frequencies as compared to a few CanPIs (-4, -7, -10) and IRDs (4, 5, 10, 14, 25) with higher frequencies in wounded leaves treated with OS. Specifically high representation of *CanPIs* with multiple IRDs directed towards enriching the PI blend with both CI and TI activities seemed to be an approach adapted by the plant upon Lepidopteran insect attack, helpful in tackling a wide range of insect proteases. Significantly high PI activity (**Fig. 2.6A**) and detection of three PI activity bands in wounded leaves treated with water or OS in comparison to two activity bands in uninduced and aphid infested leaves (**Fig. 2.6B**) was indicative of quantitative and qualitative variations in the accumulated PI activity. Further characterization by 2D electrophoresis, revealed the presence of multiple charge and/or molecular mass variants observed in wounded leaves treated with water

(**Fig. 2.7**; TI-7 to -13) and with OS (**Fig. 2.7**; TI-14 to TI-16) clearly indicating the induced isoform diversity . The differential isoforms detected mostly corresponded to 3- and 4-IRD PIs and thus correlated with the high *CanPI* transcript accumulation under these two treatments. The absence of certain TI isoforms in OS treated leaves compared to wounding alone, suggested the suppression of some induced responses, resulting in treatment specific patterns. With respect to HGP inhibition potential, the PI activity in all the leaf tissues attained 70% inhibition of HGP (**Fig. 2.6C**) and could inhibit almost all the HGP isoforms (**Fig. 2.6D**). However, an early saturation of HGP inhibition by proteins from leaves induced by wounding and/or treated with OS, as compared to uninduced leaves, is suggestive of the high quantitative accumulation of PI units in such leaf tissues. Additionally it could also be a result of qualitative variations in the induced PI profiles which might have higher specific activity and characteristic role against insect gut proteases (**Fig. 2.6C**).

Multiple IRDs are known to be generated from precursor *N. alata* PI proteins in Me-JA-elicited leaves (Atkinson et al., 1993; Horn et al., 2005) and from CanPI precursor proteins by action of the endogenous proteases at the linker regions (Tamhane et al., 2009). An increase in the number and intensity of variant mass peaks equivalent to single IRDs, in wounding with water and with OS protein fractions (**Fig. 2.8**), suggested the enhanced proteolytic processing of the up-regulated CanPI precursor proteins which was previously suggested by Horn et al. (2005) and Huesgen and Overall (2011). The specific presence of isoform TI-6 and absence of TI-1 (**Fig. 2.7**) under all the inductions also indicated differential accumulation of the 1-IRD isoforms that were generated as a result of processing of CanPI precursors. Thus, our results substantiated the hypothesis of Horn et al. (2005) that elicitation leads to over-production of the CanPI precursors and enhanced, differential processing of the precursors by proteases to IRDs, resulting in structurally and functionally diverse processed products. It was also noticed that induction treatment specificity was maintained even at the level at which precursor proteins were processed. Peaks ranging from 5.9 to 6.3 kDa showed high intensity in wounding with OS treatment whereas the peaks from 5.5 to 5.8 kDa were prominent in uninduced and other treatments, aphid infested and wounding treated with water. These results affirmed that plants can differentially perceive various kinds of biotic stresses and respond appropriately through regulation of PIs at transcriptional, translational and post-translational levels.

Sequence analysis revealed highly homologous *CanPIs* with an average variance of 4%. The clear absence of partial N- and C-terminal repeats in the *CanPI* precursors grouped distinct from *N. benthamiana* Pin-II PIs (**Fig. 2.3**). The explicit clustering of *C. annuum* PIs from all the other Solanaceous Pin-II PIs suggested recent evolutionary origins (Kong and Ranganathan, 2008). Based on their position in the dendrogram, CanPIs with progressively higher number of repeat units appeared to be phylogenetically most recent additions to the gene family. This is in accordance with the study by Barta et al. (2002) which suggested that single repeat PIs are ancestral and a series of gene and domain duplication events lead to the evolution of other forms. 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 (**Fig. 2.4A & 2.4B**). 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). The induction-specific IRD distribution was predominantly biased towards TIs rather than CIs (**Fig. 2.5A**). It is known that in Lepidopteran insects, trypsin-like proteases are predominant which could be correlated to the relatively high abundance of trypsin specific PIs in plants.

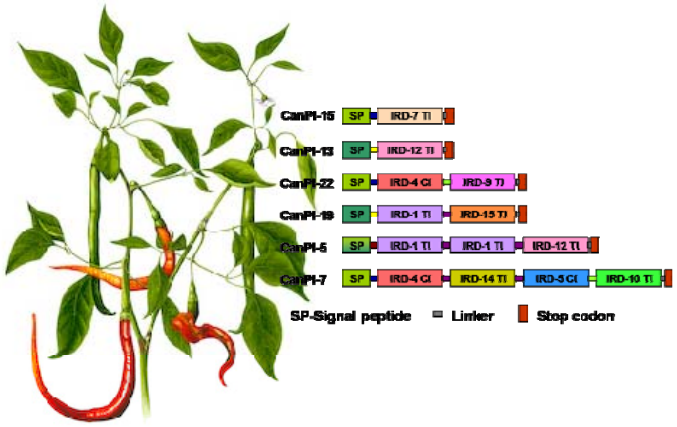
Active site variants of TIs 'CPRNC', 'CPKNC', 'CPRYC' and 'CPRDC' and two types of CI sites 'CTLNC' and 'CTPNC' were present among all the identified 55 IRDs. Interestingly, seven cysteine variants either, missing one or more conserved cysteine residues, change in position of cysteines or having additional cysteines were identified in the present study (**Fig. 2.4A**). Recently, six natural IRD variants with selective losses of cysteine residues were reported in potato (Li et al., 2011). The loss of cysteine residues is often associated with functional differentiation and suggests positive evolutionary gene selection, but what kind of mutations and the associated factors are responsible selective losses of cysteines remains unclear. Studies on the significance of such mutations affecting the proteinase binding affinity and structural stability/integrity of IRDs have initiated an active debate on the evolution of disulfide bonds in the Pin-II family (Schirra et al., 2010; Li et al., 2011). Particularly in *C. annuum*, the identification of nine such IRDs as a result of various inductions suggested plants' elaborate defenses by expressing modified IRDs that improved their overall activity against target proteases. Among various tissues from *C. annuum*, flowers revealed the highest accumulation of PI activity (**Fig. 2.9A**), consistent with a

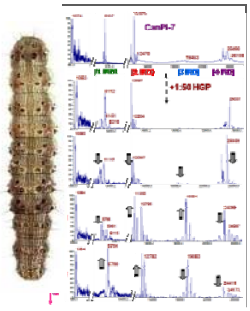
role in protecting the reproductive parts of the plant against pests as reported in tomato (Atkinson et al., 1993; Damle et al., 2005).

It appears that CanPI sequence diversity, tissue specificity and explicit responses to different inductions are part of effective plant defense. The huge complexity of PIs observed specifically in *C. annuum* raises question regarding their significance. Recent reports on the endogenous and/or defensive roles of PIs from various Solanaceous species and simultaneous expression of multiple *CanPIs* constitutively highlight 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). Moreover, up-regulated yet specialized *CanPI* expression upon wounding and insect infestation provides insights into the evolution of PI based plant defense mechanisms against insects and generates many unanswered questions about their regulation. Essentially, more functional studies need to be performed for specific CanPI genes in order to ascertain their roles under a particular treatment and how this variation accounts for the fitness benefit of the plant under specific biotic stress conditions.

Chapter 3

Interaction of recombinant CanPIs with *Helicoverpa armigera* gut proteases reveals their processing patterns, stability and efficiency





The research work described in this chapter is a full-length paper, which has been published in Proteomics (Mishra et al., 2010)

CHAPTER 3

Interaction of recombinant CanPIs with *Helicoverpa armigera* gut proteases reveals their processing patterns, stability and efficiency

Abstract

Six diverse representative *Capsicum annuum* proteinase inhibitor genes, viz CanPI-5, -7, -13, -15, -19, and -22 comprising 1–4 inhibitory repeat domains (IRDs), were cloned and expressed in *Pichia pastoris*. The recombinant proteins were evaluated for their interactions with bovine trypsin, chymotrypsin, and *Helicoverpa armigera* gut proteases (HGP) using electrophoretic (native and denaturing) and mass spectrometric (MALDI-TOF-MS in combination with intensity fading assays) techniques. These techniques allow qualitative and semi-quantitative analysis of multiple and processed IRDs of purified recombinant *C. annuum* proteinase inhibitor (rCanPI) proteins. rCanPIs showed over 90% trypsin inhibition, varying chymotrypsin inhibition depending on the number of respective IRDs and over 60% inhibition of total HGP. rCanPI-15 that has only one IRD showed exceptionally low inhibition of these proteases. Interaction studies of rCanPIs with proteases using intensity fading-MALDI-TOF-MS revealed gradual processing of multi-IRD rCanPIs into single IRD forms by the action of HGP at the linker region, unlike their interactions with trypsin and chymotrypsin. *In vitro* and *in vivo* studies using rCanPI-5 and -7 showed maximum inhibition of HGP isoforms and their processed IRDs were also found to be stable in the presence of HGP. Even single amino acid variations in IRDs were found to change the HGP specificity like in the case of HGP-8 inhibited only by IRD-12. The presence of active PI in insect gut might be responsible for changed HGP profile. rCanPI-5 and -7 enhanced HGP-7 while reduced HGP-4, -5, -10 expression and new protease isoforms were induced. These results signify isoform complexity in plant PIs and insect proteases.

3.1. Introduction

The success or failure of the PI strategy against insects depends mainly on two factors, the nature of insect response to ingested PI and secondly on the stability of PIs to proteolytic degradation. PIs from 'non-host' plants have been found to be more effective against insects by several studies (Harsulkar et al., 1999; Telang et al., 2003; Tamhane et al., 2007). When incorporated into diets and fed to insects, the efficacy of protease inhibitors has proved to be variable probably due to the huge variability in PIs as well as the proteases. The dynamic interaction between gut proteases and the PIs thus, regulates the responses of insects to specific PIs (Bown et al., 2004; Broadway, 1995, 1997; Chougule et al., 2005). The detailed *in vitro* and *in vivo* studies on interaction between varied PI molecules from various sources and insect gut proteases can facilitate the selection of PIs based on their efficacy against insects.

Recently MALDI-MS has been introduced as an alternative strategy to detect non-covalent interactions in solution (Villanueva et al., 2003). In this strategy Intensity fading (IF-) MALDI-TOF-MS, a reduction in the relative abundance of peptide ligands is visualized as a reduction of their relative intensity in the mass spectra, which is observed when their receptor protein is added to the MALDI sample. It has been widely used to study a range of biological interactions like protein-protein, protein-nucleic acid and protein-organic compounds. 'Fading' behaviour of protease inhibitors (potato carboxypeptidase inhibitor (PCI) and bovine pancreatic trypsin inhibitor (BPTI)) interacting with the target enzymes/proteases has been explored (Yanes et al., 2007; Shabab et al., 2008). Another advantage is the requirement of only sub micro molar range of proteins to detect the interactions in solution. However, the detection of intact non-covalent complex ions can be difficult due to poor efficiency of detection of high molecular weight proteins by MALDI-TOF-MS instruments equipped with conventional micro channel plate (MCP) detectors. In the present study, this approach was used to explore the interaction of recombinant CanPI proteins with target proteases like trypsin, chymotrypsin and HGPs, which also yielded interesting insights in to processing patterns and stability of CanPI.

Previous studies have shown that diverse Pin-II PIs from *Capsicum annuum* differentially influence *H. armigera* growth and development (Tamhane et al., 2007). Furthermore, *C. annuum* showed strong up regulation of multi-IRD PIs upon *S. litura* and aphid attack (Tamhane et al., 2009) and differential expression patterns in various

tissues under steady state or induced conditions. These results indicate significance of CanPI for their defensive and endogenous role which still remains poorly understood. Here we attempt to address the questions (i) what is the fate of CanPI in insect gut? (ii) What is the insect's response to the ingested CanPI? (iii) Whether this reaction is different for specific CanPI? We selected six *CanPI* genes on the basis of sequence variation, specificity and number of IRDs and characterized them with specific reference to their (i) processing by HGP (ii) stability in proteolytic environment and (iii) inhibitory activity against HGP. Using IF-MALDI-TOF-MS, enzyme assays and PI activity gels, interaction of rCanPIs with HGP was analyzed. Since these studies are based on product(s) of particular PI gene(s), it leads to identification of potential PI(s) or IRD(s) effective against constitutive and induced insect gut proteases.

3.2. Materials and methods

3.2.1. Sequence analysis of *CanPI* genes

Out of 21 novel *CanPIs* identified in our previous study (Tamhane et al., 2009), six representative genes were selected for their functional characterization. The criteria for selection included: diversity in number of IRDs, TI or CI specificity of the IRD and aa variation in IRD sequence. The cDNA and aa sequence analysis was performed using the DNA star and Clustal X softwares. *CanPI-13* and *CanPI-15* having 1-IRD; *CanPI-19* and *CanPI-22* having 2-IRDs; *CanPI-5* having 3-IRDs and *CanPI-7* having 4-IRDs were selected for cloning in *Pichia pastoris* for recombinant protein expression.

3.2.2. Cloning of *CanPI* genes in *Pichia pastoris*

The mature peptide regions of *CanPIs* were cloned in expression vector pPIC9 (Invitrogen) for recombinant, extra-cellular expression in *P. pastoris* GS115. The following primers: 5′-AAA AAA *CTC GAG* AAA AGA GAG GCT GAA GCT AAG GCT TGT TCA CAA AG-3′, and 5′-AAA AAA *GCG GCC GC* CTG TTC ATG CTT TTA C-3′ (italicized region indicates *Xho*I and *Not*I sites) were used to carry out amplification of the mature peptide regions from the full-length gene cloned in the pGEMT vector using Accuprime Pfx polymerase (Promega). The amplified product was digested with *Xho*I and *Not*I (both from Promega) and purified using gel elution kit (Sigma). Yeast expression vector pPIC9 (Invitrogen) was prepared by digesting it

with the same restriction enzymes. The digested insert and vector were quantified and ligated in 1:3 ratio using T4 ligase (Promega). The ligated product was used for transformation of competent cells of *E. coli* strain Top10. Positive clones were identified by colony PCR; plasmids were isolated from them and checked by restriction digestion with *Xho*I and *Not*I enzymes. The plasmids were digested with *Sal*I (Promega) for linearization and were used for *P. pastoris* (GS115) transformation. The positive transformants were selected on histidine-deficient regeneration dextrose base (RDB) medium which selectively allowed the growth of the cells having integrated pPIC9 vector.

3.2.3. Recombinant protein expression and purification

Around 10 recombinant *P. pastoris* colonies each from CanPI-15, -13, -22, -19, -5 and -7 were screened for protein expression to arrive at the best clone. The individual recombinant *P. pastoris* colonies were initially grown at 30ºC in 'minimal glycerol (MGY) medium' for ~24 hours. The cell pellet was collected by centrifugation at 3000 g, 10 min at RT and was transferred to 'buffered methanol-complex (BMMY) medium'. Methanol (1%) was supplemented every 24 h and the culture was allowed to grow for ~96 h. It was then centrifuged at 14, 000 g, 10 min at 10ºC to obtain the supernatant. The clear supernatant was saturated with NaCl up to a final concentration of 2 M. The entire volume was loaded on Phenyl-Sepharose column (HIC) pre-equilibrated with 2 M NaCl. The expressed protein was eluted under reverse NaCl gradient (2.0-0.0 M). Protein estimation was carried out spectrophotometrically using Bradford reagent.

3.2.4. PI assays and activity visualization of proteases and PIs

HGP was extracted from insect gut tissues as described in section 2.2.3 of Materials and methods of Chapter 2 and total proteolytic activity of bovine trypsin, chymotrypsin and HGP was determined using azo-caseinolytic assays as given in Tamhane et al. (2005). Inhibitory assays using rCanPIs were performed as detailed in Tamhane et al. (2005) with increasing amount of PIs (0.3 to 4 µg) for rCanPI-5, -7, -19, -22 and 8 µg of PIs for rCanPI-13 and -15. The inhibition potential of all the six rCanPIs against trypsin, chymotrypsin and HGP was estimated.

rCanPIs (30 µg of each) or proteases were resolved in 15% or 8% native-PAGE, respectively and the gel(s) were further processed for activity visualization(s) using the gel X-ray film contact print (GXCT) method as described in section 2.2.4 of Materials and methods of Chapter 2.

3.2.5. *In vitro* and *in vivo* interactions of rCanPI(s) and HGP

To study the interaction of PIs with HGP *in vitro*, 0.5 HGPI units of rCanPIs (CanPI-5, CanPI-7, CanPI-19 and CanPI-22) were incubated with 1U HGP for three time points (5 min, 1 h, and 6 h) at RT. For rCanPI-13 and -15, 2 HGPI units were incubated with 1U of HGP. These HGP- treated PIs were resolved on native PAGE gels and processed for TI activity visualization. Remaining HGP activity was visualized upon incubation with PIs (1PI: 2HGP units) for 1 h.

Seventy µg (amount of inhibitor required for maximum % inhibition of total HGP activity from a single fourth instar larva) of rCanPIs (rCanPI-5 and -7) were incorporated per gram of artificial chickpea flour based diet (Tamhane et al., 2007) to test their *in vivo* potential against *H. armigera*. Larvae were reared on rCanPI-containing diets and control diet in separate sets of 30 larvae each. Each larva was maintained in an individual vial (50 ml). Gut tissue from the 4^{th} instar larvae (from each set) was dissected and stored frozen (-80°C) in polypropylene tubes (1.7 ml) until further use. Proteases from gut tissues (200 mg) of *H. armigera* fed with rCanPI-5 and -7 incorporated artificial diets were extracted as described earlier. A two third portion of this gut extract was heat treated at 70ºC for 15 min to inactivate and precipitate the active enzymes/proteases. The heat treated extracts were clarified by centrifugation at 13,000x g for 10 min at 4 ºC and supernatant resolved by native-PAGE followed by TI activity visualization by GXCT method.

The untreated HGP extracts from the CanPI-5 and -7 fed *H. armigera* guts were separated on native-PAGE and processed for protease activity visualization for identification of CanPI-5 and -7 inhibited/induced HGP isoforms.

3.2.6. Protease-PI interaction assays by MALDI-TOF-MS

rCanPIs and their interactions with HGP and other proteases were monitored by IF-MALDI-TOF analysis where in, reduction in intensity of a ligand (inhibitor) is monitored by MALDI-TOF-MS on addition of a target protease (Yanes et al., 2007).

The mass spectral analysis was done on Voyager-De-STR MALDI-TOF (Applied Biosystems) using standard instrument settings as detailed in section 2.2.5 of Materials and methods of Chapter 2. All spectras were acquired by accumulating 50 single laser shots over each sample spot in the range of 1 to 30 kDa. They were processed for advanced base line correction and noise removal using Data Explorer™ software (Applied Biosystems). The instrument was calibrated using apomyoglobin and bovine serum albumin (both Sigma-Aldrich).

For analysis of rCanPIs, 3 µg of protein sample was mixed with 20 µl of freshly prepared sinapinic acid (Sigma-Aldrich) [30% acetonitrile (CAN), 0.1% trifluoroacetic acid (TFA)]. Aliquots of 2 µl of this mixture were spotted on the stainless steel MALDI plate by dried-droplet method and incubated at 37 ºC for 20 min. The MALDI target plate was further subjected to MALDI-TOF as specified above to get spectral profiles.

For HGP-rCanPI interaction studies, equal (0.05 U) HGPI units of each rCanPI were individually incubated with three dilutions of HGP (0.5 U, 0.1 U, 0.01 U or crude, 1:5, 1:50) for various time points (5 min, 30 min, 1 h, 3 h and 6 h). Total volume of this reaction mixture was 10 µl, out of which 5 µl was mixed with 20 µl of freshly prepared sinapinic acid and processed for obtaining MALDI-TOF-MS profiles as mentioned above. Interaction of rCanPI-7 with 0.5 U of bovine trypsin and chymotrypsin (both Sigma-Aldrich) was also monitored by MALDI-TOF-MS as described above.

3.3. Results

3.3.1. Selection of *CanPI* genes for recombinant protein characterization

The sequence alignments, dendrograms and nine constituent IRDs of six *CanPI* genes selected in the present study are shown in (**Fig. 3.1 & 3.2**). A typical *CanPI* consists of a 25 aa signal peptide followed by 1 to 4-IRDs coupled by 5 aa linker(s). Nine unique IRDs with 2 to 26% sequence divergence were identified from these six *CanPIs* and their multiple sequence alignment revealed major aa substitutions in reactive site loop (**Fig. 3.2**). Each ~50 aa IRD consists of 8 conserved cysteine (C) residues and a single reactive site (P1), either for trypsin (TI) or chymotrypsin (CI) inhibition. Presence of arginine (R) or lysine (K) at the P1 position

is responsible for TI specificity, whereas presence of leucine (L) at the P1 position represents a typical CI activity.

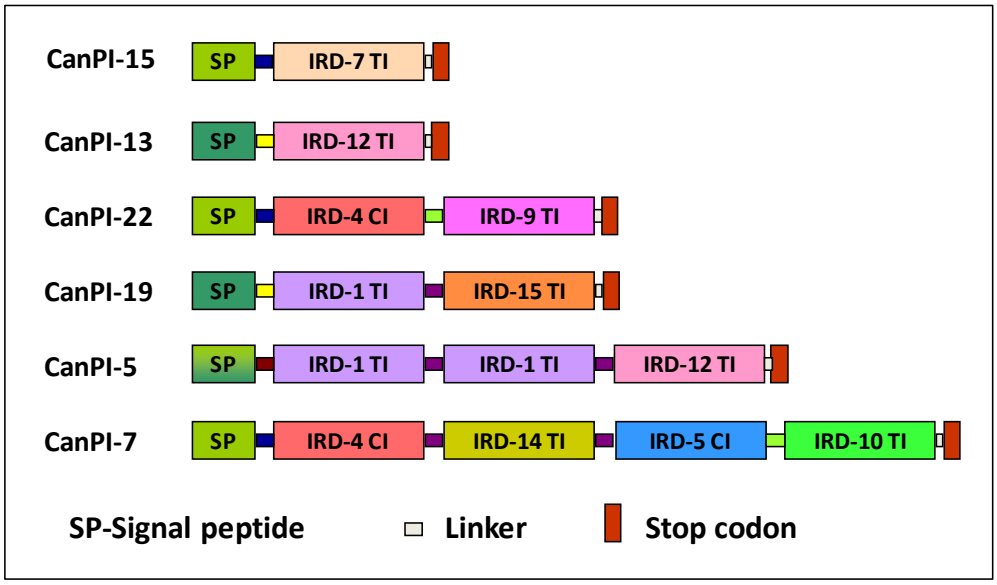


Figure 3.1A: Diagrammatic representation highlighting the gene structure of four types of *CanPI*s found in *C. annuum*, with their signal peptide sequence (SP), various IRD(s), linker region(s) and the stop codon. The signal peptide, IRDs and linker regions varying in the aa sequence are shown in different colors and indicate their positions.

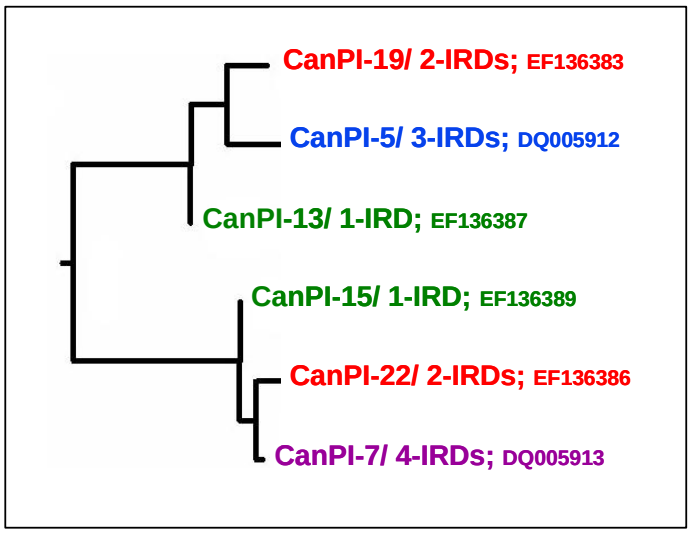


Figure 3.1B: Neighbor-joining tree of CanPIs based on deduced aa sequences of full length genes, number of IRDs and the accession number.

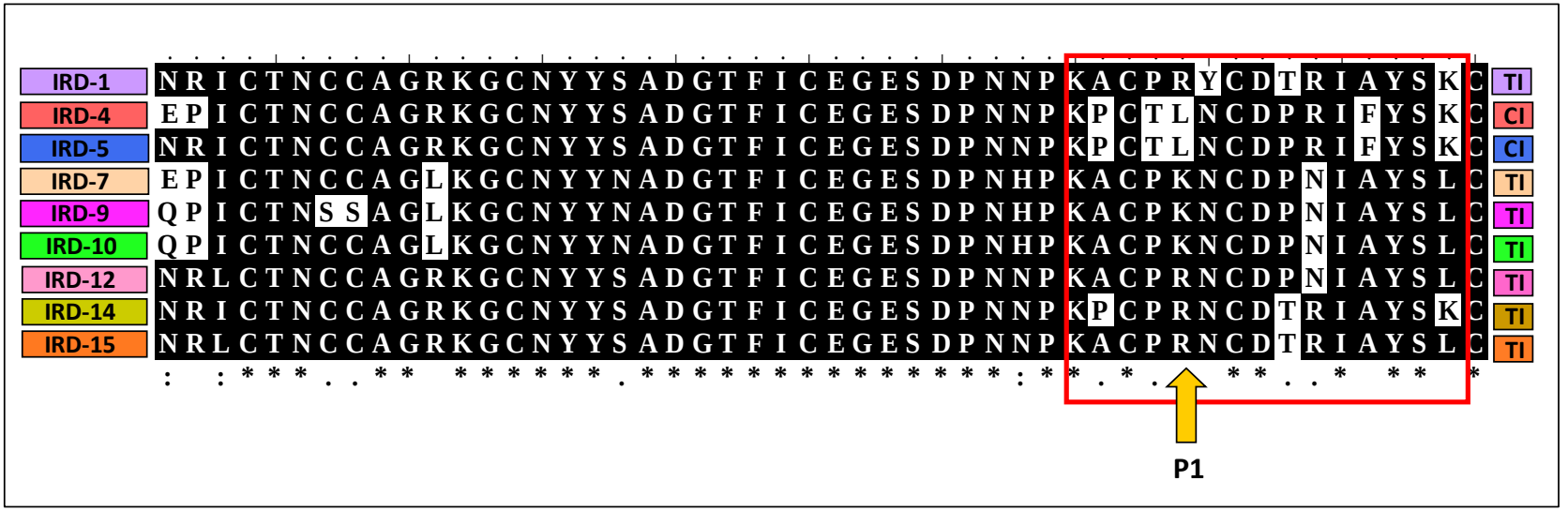


Figure 3.2: Multiple sequence alignment of deduced aa sequences of unique IRDs from the CanPIs selected for the present study. The IRD numbers are according to the details provided in Chapter 2. The inhibitory active site in the particular IRD is referred to as TI for trypsin and CI for chymotrypsin inhibition. The reactive site residue P1 is marked by an arrow and the region close to the active site showing major variation is marked by a box.

3.3.2. MALDI-TOF-MS and electrophoretic analysis of rCanPIs reveals PI isoforms with variable number of active IRDs

SDS-PAGE protein profiles of the individually expressed rCanPIs showed a single ~6 kDa protein in 1-IRD CanPIs, ~12 kDa and ~6 kDa proteins in 2-IRD CanPIs, ~19, ~12 and ~6 kDa proteins in 3-IRD CanPI and four proteins of size ~25, ~19, ~12 and ~6 kDa in 4-IRD CanPI (**Fig. 3.3 inset**). These rCanPIs were further analyzed by IF-MALDI-TOF-MS to study their interaction with HGP as well as to determine their accurate molecular masses. All mass spectras were acquired in the range of 1 to 30 kDa. A single peak of 5229 Da was observed in case of CanPI-15 (1-IRD); peaks of 12231 and 5946 Da for CanPI-22 (2-IRD) and 19214, 12530 and 6190 Da peaks in CanPI-5 (3-IRD) were observed while CanPI-7 (4-IRD) showed four expected size peaks of 25377, 19245, 12070 and 6107 Da (**Fig. 3.3**). These results were in accordance with the SDS-PAGE protein profiles. It is known that the higher molecular mass proteins exhibit low intensity in the mass spectra due to various factors, such as they reach the detector relatively slow, sometimes due to ion suppression, and also because of low sensitivity of MCP detectors. Therefore, low intensity peaks of high (>8kDa) molecular mass proteins were observed by zooming selected region on the X-axis as displayed in insets in **Fig. 3.3**. Mass spectras revealed at least 2 to 3 sub-peaks (or more in some cases) of variable intensities for each multiple or single-IRD isoform (**Fig. 3.3**). MALDI-TOF-MS analysis can qualitatively and semi-quantitatively determine the composition of the processed IRDs in the purified rCanPI pools. Hence this technique was further used to monitor the interactions between rCanPI(s) and various proteases.

3.3.3. Inhibitory activities of rCanPIs against trypsin, chymotrypsin and HGP

rCanPIs were resolved on native and SDS-PAGE and visualized for TI profiles (**Fig. 3.4**). Multiple TI activity isoforms were detected for the rCanPIs including 1-IRD CanPIs indicating the presence of heterogeneity at the activity level. The number of TI activity isoforms was more in case of multi-IRD CanPIs as compared to 1-IRD PIs.

rCanPIs (0.1 to 2.0 µg proteins) were used for inhibition studies against trypsin, chymotrypsin and HGP of 4th instar larvae fed on artificial diet (AD-HGP) to find out a PI concentration required for maximum inhibition. These amounts of

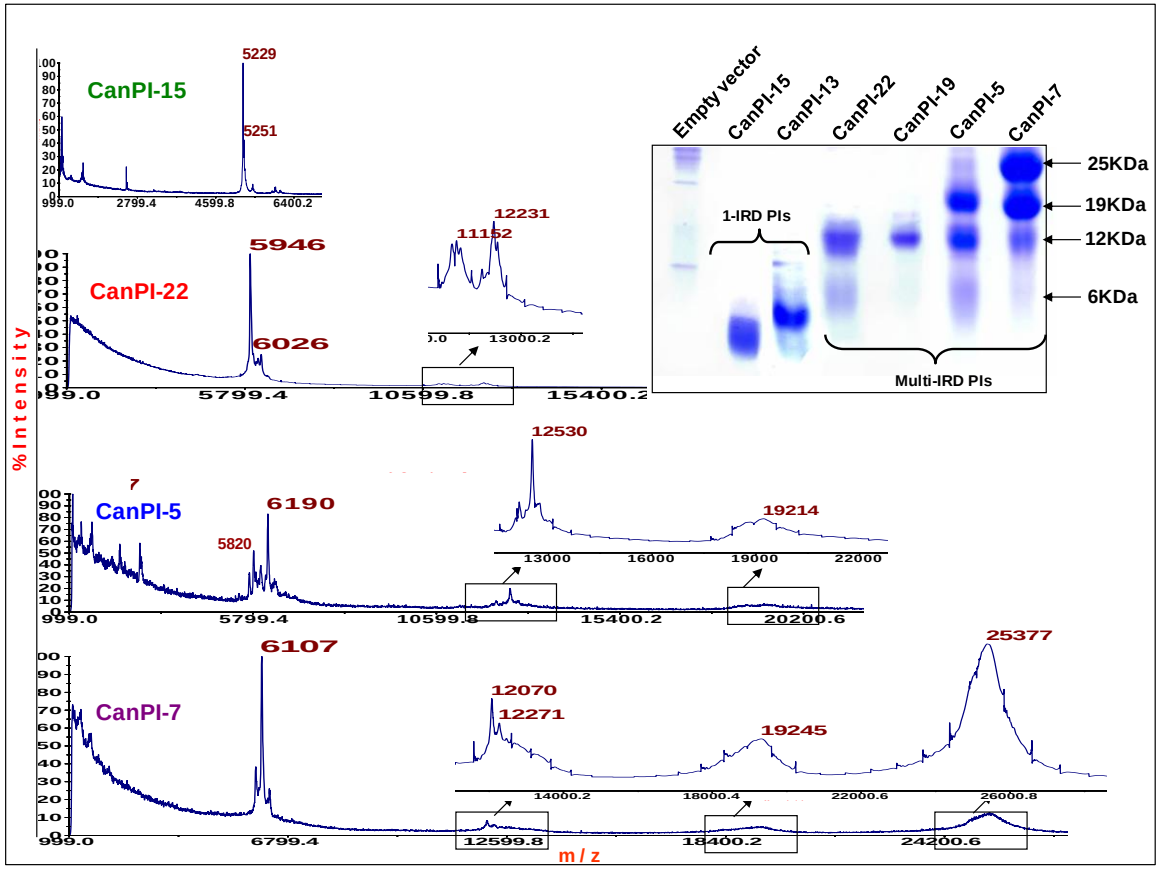


Figure 3.3: Characterization of rCanPIs having either 1, 2, 3 or 4-IRDs each by SDS-PAGE and MALDI-TOF-MS. Proteins stained with Coomassie Blue and mass peaks of 25 kDa (4-IRDs), 19 kDa (3-IRDs), 12 kDa (2-IRDs) and 6 kDa (1-IRD) were detected depending on the number of IRDs present in the rCanPIs. The proteins of increasingly higher molecular mass appear as low intensity peaks in the mass spectra because of the drop-off of the detection efficiency with increasing mass. These peaks were approximately four times magnified in the insets.

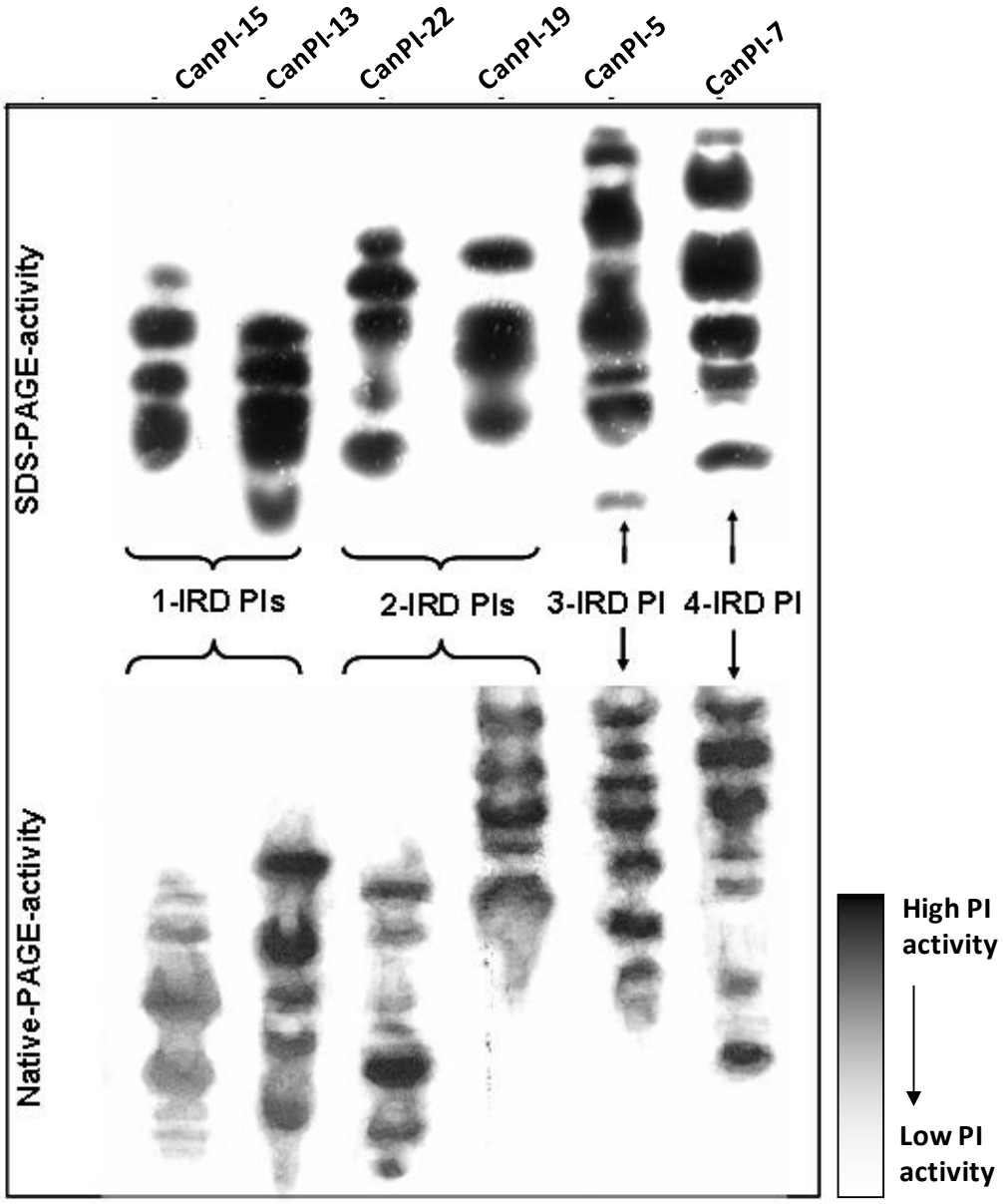


Figure 3.4: Trypsin inhibitory activity (TI) visualization of rCanPIs. rCanPIs were resolved in 15% SDS-PAGE (upper) and 12% Native-PAGE (lower) and subjected to in-gel TI activity visualization using GXCT. Multiple rCanPI TI activity was detected in both the gels. The first TI activity protein in the SDS-PAGE was corresponding to the apparent molecular mass of the predicted mature rCanPI of the respective sequence.

rCanPIs showed 70 to 90% inhibition of bovine trypsin activity; except for rCanPI-15. Chymotrypsin inhibition was in the range of 50 to 90% by various rCanPIs. rCanPIs without CI sites namely rCanPI-5, -13 and -19 exhibited 50 to 84% CI activity whereas rCanPI-7 and rCanPI-22, both of which have 2 and 1 CI sites, respectively,

showed more than 90% CI activity. rCanPIs inhibited 50 to 60% of HGP activity (**Fig. 3.5**). rCanPI-15 showed very low trypsin (10%) and HGP (6%) inhibition and failed to inhibit chymotrypsin activity.

HGP inhibition by all the rCanPIs except rCanPI-15 using BA*p*NA as a substrate was 80% while only 60% by rCanPI-15. The change in the substrate showed remarkable difference in the inhibitory activities of the rCanPIs.

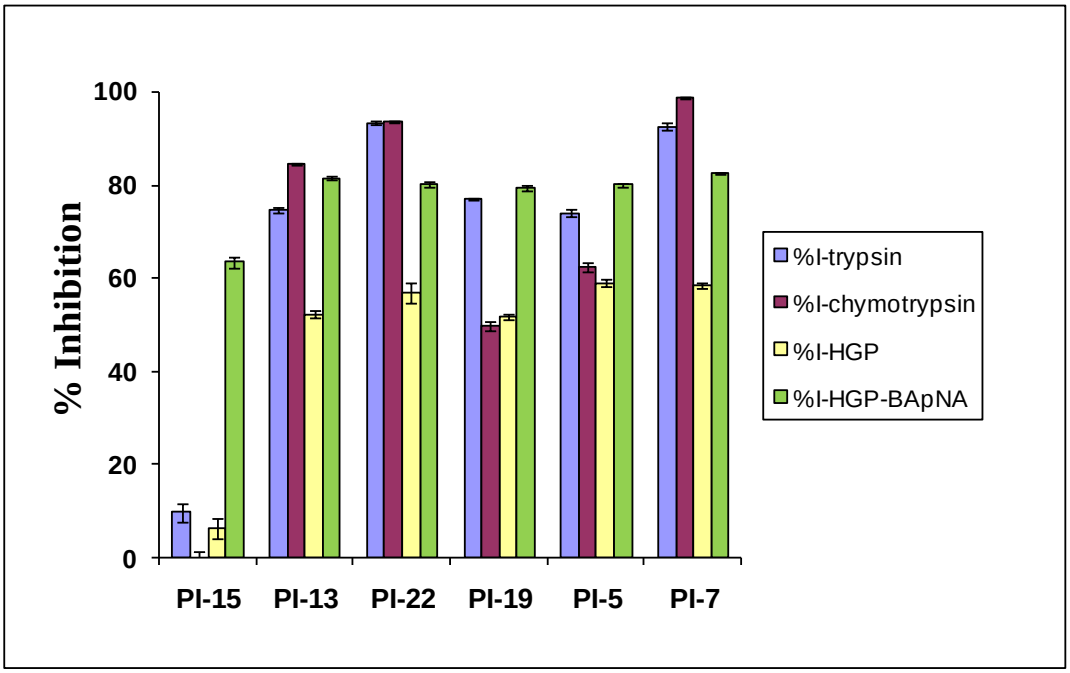


Figure 3.5: Enzyme inhibition by rCanPIs. Maximum percent inhibition of bovine trypsin, chymotrypsin and HGP by minimum amounts (µg) of rCanPIs in azo-caseinolytic assays is represented in the bar graph. Each value is an average of six replicates with bars indicating standard error.

3.3.4. rCanPI-HGP interactions detected by IF-MALDI-TOF-MS

The interactions of rCanPIs with various proteases were studied by IF-MALDI-TOF-MS as described previously (Shabab et al., 2008). The interaction assays between rCanPIs and HGP were optimized with three HGP concentrations (undiluted, 1:5 and 1:50 dilutions) and intensities of the rCanPIs were monitored from 5 min to 6 h (**Fig. 3.6 to 3.8**). In these interaction studies a peak of 1062 Da was considered as an internal standard for relative quantification of CanPI peak because of its consistent presence in all the CanPI sample preparations and secondly there was no other peak that was closer to 5000 - 6000 Da mass range. As shown in **Fig. 3.6**

relative intensity of rCanPI-15 peak at 5229 Da progressively decreased up to 6 h after its interaction with 1:5 HGP.

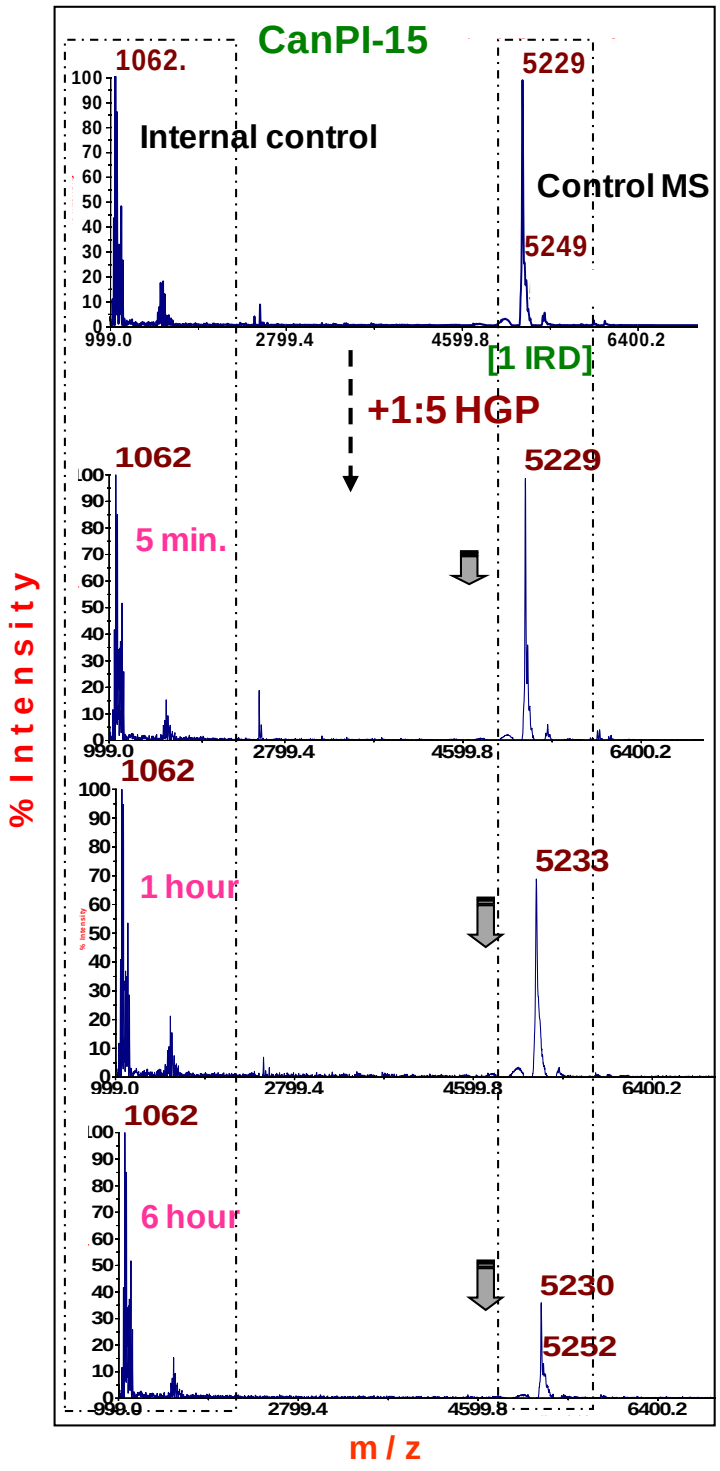


Figure 3.6: IF–MALDI-TOF-MS analysis of rCanPI-15. The decrease in the relative intensity of rCanPI-15 (5.2 kDa) upon addition of target protease, HGP was evident. The internal control (1062 Da) has been used as reference for relative quantification.

HGP and multi-IRD CanPIs interaction showed different patterns in addition to the IF phenomena. In case of rCanPI-22 having 2-IRDs, intensity of both, the 11895 (2-IRD) and the 5950 Da (1-IRD) peaks reduced at 5 min of incubation. However, at later time points (3 h and 6 h) 2-IRD peak was not detectable where as 1-IRD peak (5778 Da) intensity increased. This enrichment of the 1-IRD peak indicated the processing of 2-IRD in to 1-IRD by HGP (**Fig. 3.7A**). Similar interactions were observed between rCanPI-5 and diluted 1:5 HGP (**Fig. 3.7B**). A difference of ~181 Da was noted in the processed 1-IRD peak with respect to control, after 6 h of 1:5 HGP treatment.

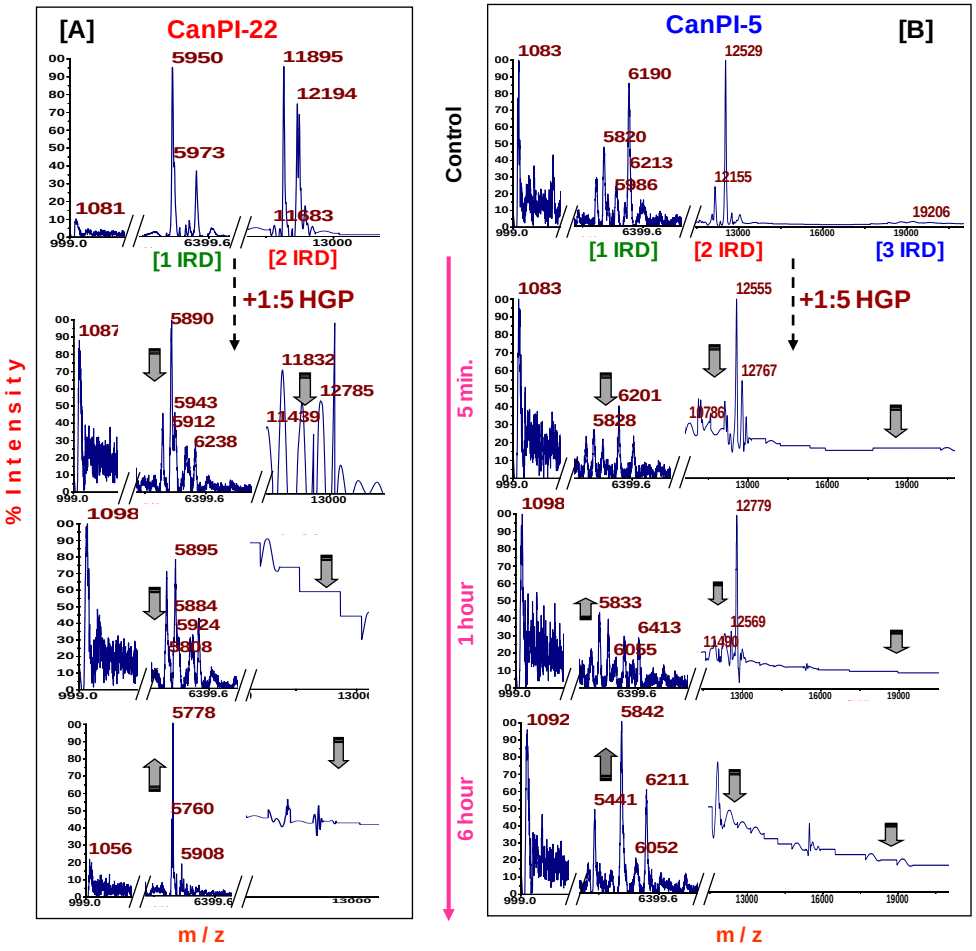


Figure 3.7A: IF-MALDI-TOF-MS analysis of 2-IRD PI interacting with HGP. 1:5 diluted HGP was incubated with 0.05 HGPIU of rCanPI-22 (2-IRD PI) for 5 min, 1 and 6 h at 24 ^{o}C. The changed peak intensities following interactions were monitored and marked by arrows. The 2-IRD peak is not detectable after 1 h and 1-IRD peak shows a progressive increase with time.

Figure 3.7B: IF-MALDI-TOF-MS analysis of 3-IRD PI (CanPI-5) interacting with HGP. Gradual processing of 3-IRD, 2-IRD forms in to 1-IRD form by HGP is indicated by modified intensities of the peaks as compared to original rCanPI pool.

The interaction of rCanPI-7 with two representative concentrations of HGP at four time points is represented in **Fig. 3.8**. The IF of the 4 peaks in rCanPI-7 was prominent with higher concentration of HGP (1:5) (**Fig. 3.8A)** whereas dilute HGP (1:50 HGP) (**Fig. 3.8B)** demonstrated the sequential conversion of 4-IRD to 3-, 2- and finally 1-IRD forms. After 6 h, the intensity of 1-IRD peak (5767 Da) was much higher than the untreated rCanPI-7 peak (6107 Da). A transient increase in the molecular mass diversity of the lower IRD forms followed by appearance of stable and intense 1-IRD peak at 5767 Da with both the HGP concentrations was evident (**Fig. 3.8**). There was no further processing and/or degradation of this peak (5767 Da) up to 6 h of interaction. Comparison of observed and calculated molecular masses of processed repeats was then used for determining the probable sites within the linkers where proteases would have acted (**Table 3.1**).

3.3.5. Comparing interactions of rCanPI-7 with trypsin, chymotrypsin and HGP

Bovine trypsin, chymotrypsin and HGP interactions were monitored by IF-MALDI-TOF-MS in the time interval of 5 min to 6 h. The representative spectra are shown in **Fig. 3.9**. In the rCanPI-7 and trypsin interaction, decrease in the 1-IRD peak intensity was evident at 3 h with a slight reduction in the 2-, 3- and 4-IRD intensities. With chymotrypsin there was a decrease in the intensities of 1-, 2- and 3- IRD peaks up to 3 h. Interesting feature of the rCanPI-7-chymotrypsin interaction was the appearance of peaks at 10 and 15 kDa and increase in 4-IRD peak intensity. rCanPI-7 and HGP interaction is as detailed above in section 3.4.

3.3.6. *In vitro* and *in vivo* stability of CanPIs to HGP

In vitro stability of rCanPIs treated with HGP, was analyzed by native in-gel TI activity visualization of rCanPI isoforms (**Fig. 3.10 & 3.11**). 1-IRD PIs, rCanPI-13 and rCanPI-15 displayed 4 or 5 TI isoforms of which only 1 or 2 remained stable after HGP treatment for 6 h (**Fig. 3.10)**. rCanPI-22 (2-IRD PI) exhibited six TI isoforms, out of which two prominent higher mobility isoforms were stable to HGP. Out of the 5 TI isoforms of rCanPI-19 (2-IRD PI), 3 remained stable up to 1 h in presence of HGP and showed partial degradation at 6 h (**Fig. 3.10)**. Most of the TI isoforms of rCanPI-5 (3-IRD PI) and -7 (4-IRD PI) were found to be stable in presence of HGP even up to 6 h (**Fig. 3.11)**. However, HGP altered the mobility of TI isoforms as in the case of rCanPI-7 where few new lower mobility TIs appeared within 5 min of HGP treatment

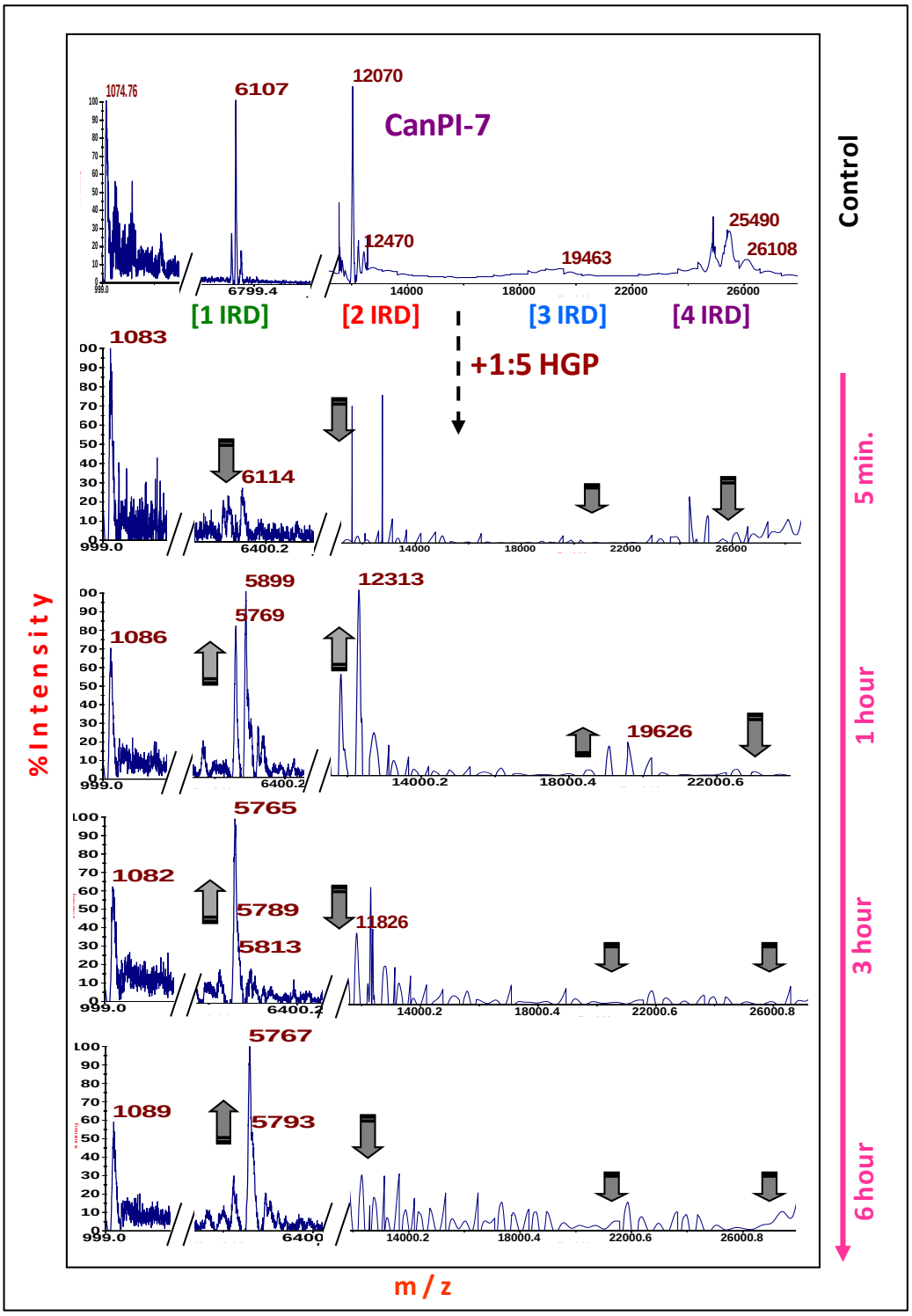


Figure 3.8A: IF-MALDI-TOF-MS analysis of rCanPI-7-HGP interaction. 0.1 U of HGP was incubated with rCanPI-7 (0.05 HGPIU) for 5 min, 1, 3 and 6 h at 24^{o}C. Due to the interactions between the PI and HGP, change in the intensity and diversity of the CanPI peaks were detected and are indicated by arrows. The higher molecular mass range from 8 kDa onwards is enlarged to enhance visibility of peak.

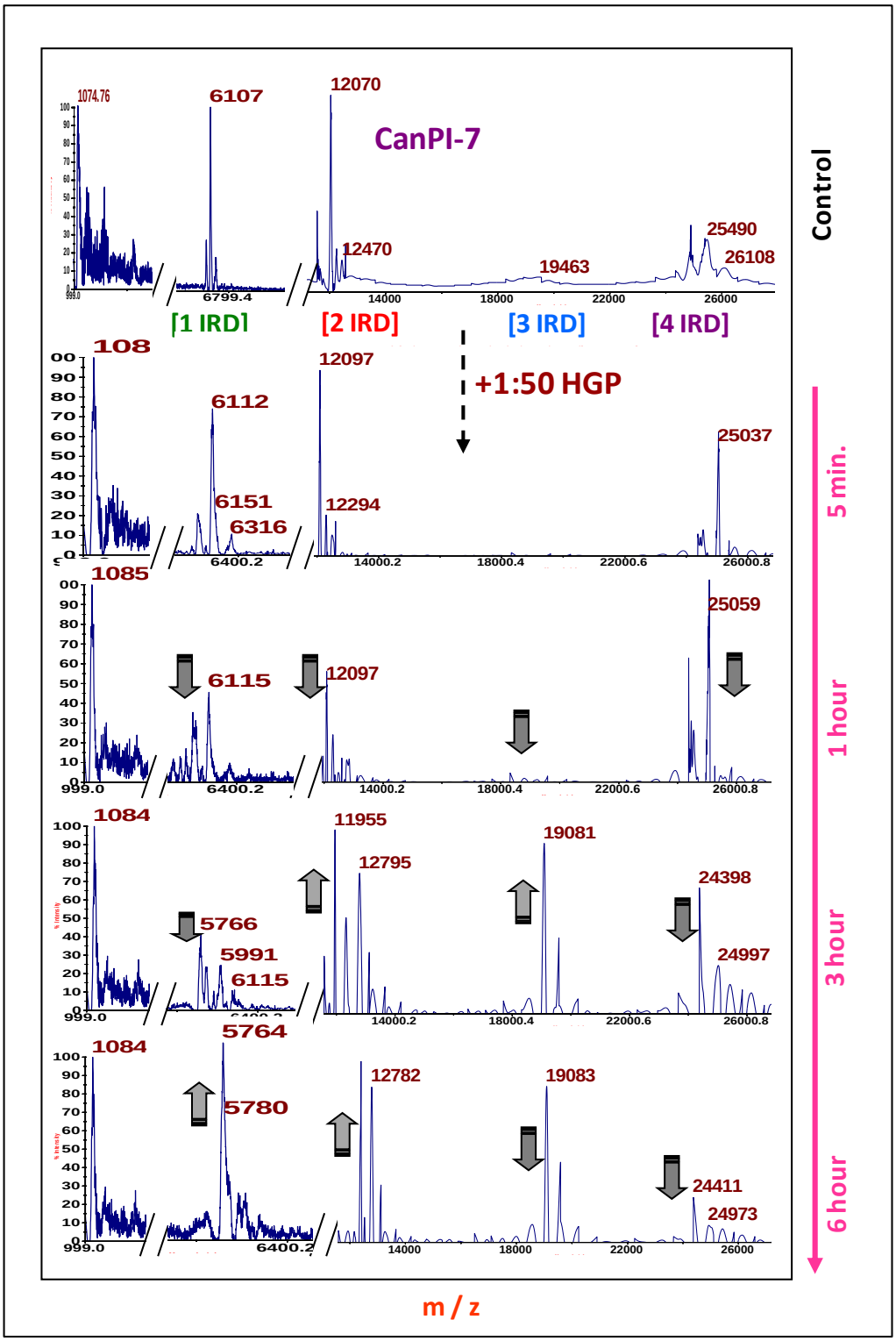


Figure 3.8B: Dilution effect on rCanPI-7-HGP interaction. 0.01 U of HGP was incubated with rCanPI-7 (0.05 HGPIU) for 5 min, 1, 3 and 6 h at 24°C. Due to the interactions between the PI and HGP, change in the intensity and diversity of the CanPI peaks were detected and are indicated by arrows. The processing of multi-IRDs forms was slower with diluted concentration of HGP.

Table 3.1: Molecular mass of processed CanPI-7 IRDs and predicted site of processing.

Isoforms of CanPI-7 generated after proteolytic processing	Observed molecular mass (kDa)	Fragment equivalent to observed molecular mass	Calculated molecular mass (kDa)	Predicted site of processing at the linkers
1R forms	6.111	E-IRD-14-E	6.11	EGNA↓E-IRD-14-E↓GNAE
	6.106	AE-IRD-14-E	6.18	EGN↓AE-IRD-14-E↓GNAE
	5.946	K-IRD-4	5.97	QRNA↓K-IRD-4
	5.767*	AE-IRD-14	5.71	EGN↓AE-IRD-14
	5.767*	AK-IRD-4	5.7	QRN↓AK-IRD-4
	5.767*	AE- IRD-5	5.74	EGN↓AE- IRD-5
	5.767*	ASAE- IRD-10	5.72	E↓ASAE- IRD-10
2R forms	12.270	IRD-14-IRD-5	12.22	EGNAE↓-IRD-14-IRD-5-↓EASAE
	12.096	IRD-4-IRD-14	12.18	QRNAK↓-IRD-4-IRD-14-↓EGNAE
	12.059	GNAE-IRD-5-IRD-10	12.08	E↓GNAE-IRD-5-IRD-10

3R forms	19.245	NAK-IRD-4-IRD-14-IRD-5-EASA	19.2	QR↓NAK- IRD-4-IRD-14-IRD-5- EASA↓E
	18.246	AE-IRD-14-IRD-5-IRD-10	18.25	EGN↓AE- IRD-14-IRD-5-IRD-10
4R forms	25.819	KACSQRNAK-IRD-4-IRD-14-IRD-5-IRD-10-LYEK	25.89	SP-↓KACSQRNAK-IRD-4-IRD-14-IRD-5-IRD-10-LYEK
	25.377	KACSQRNAK-IRD-4-IRD-14-IRD-5-IRD-10	25.36	SP-↓KACSQRNAK-IRD-4-IRD-14-IRD-5-IRD-10
	25.051	SQRNAK-IRD-4-IRD-14-IRD-5-IRD-10	25.06	SP-KAC↓SQRNAK-IRD-4-IRD-14-IRD-5-IRD-10

The molecular masses of 1-, 2-, 3- and 4-IRD isoforms of rCanPI-7 (major peaks in native rCanPI pool as well as on interaction with HGP) observed in the MALDI-TOF-MS spectra were used to compare with the calculated molecular masses of the processed IRDs and thus predict the site of processing at/within the linker region. The link given below was used for calculating masses of processed IRDs by removing aa one by one. [http://www.sciencegateway.org/tools/proteinmw.htm]

* Isoform predominantly observed on processing by HGP

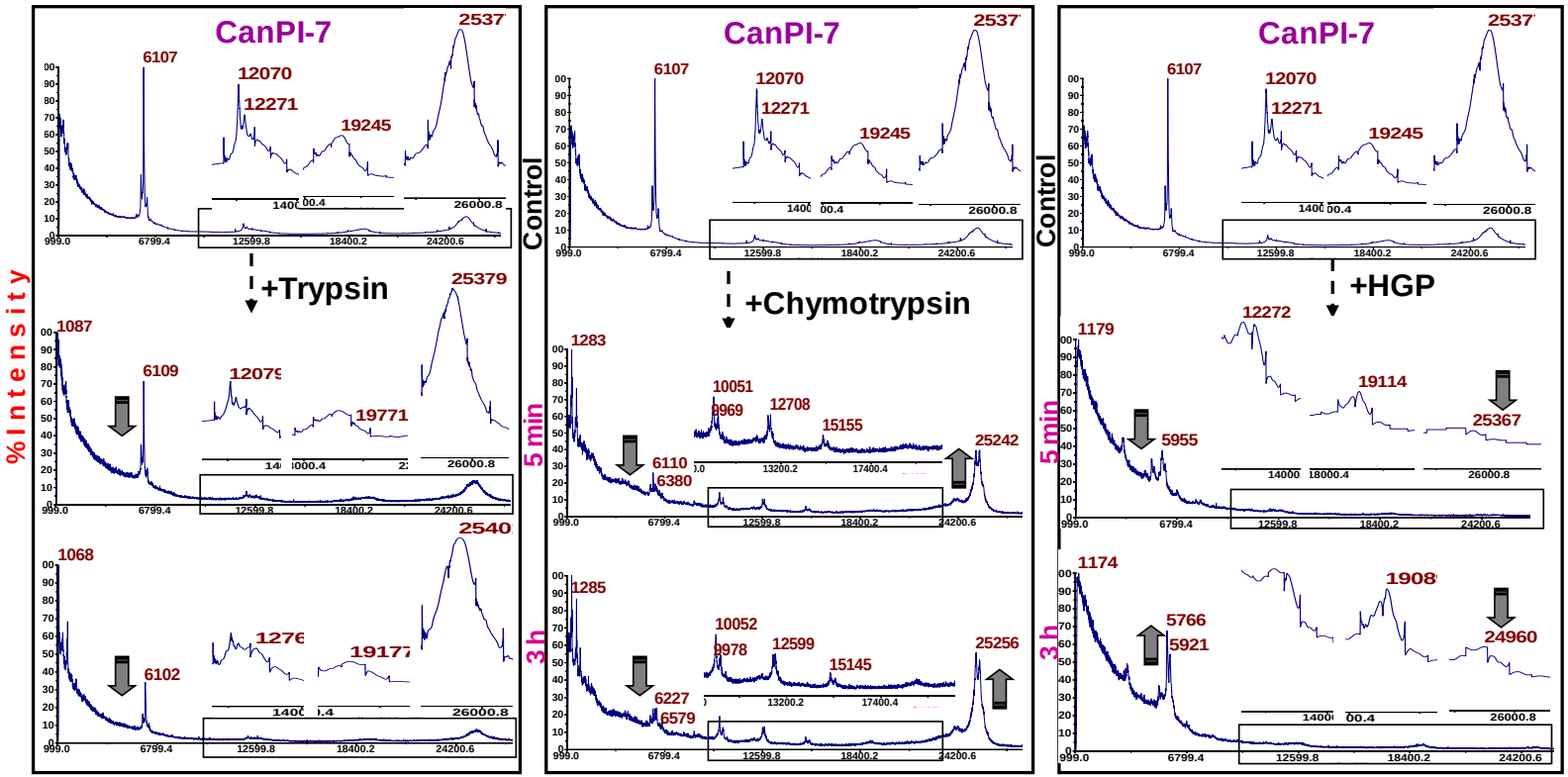


Figure 3.9: IF-MALDI-TOF-MS analysis of the interaction(s) between rCanPI-7 and bovine trypsin (a), chymotrypsin (b) and HGP (c). 0.5 U, 0.5 U, 0.1 U of trypsin, chymotrypsin, HGP, respectively were incubated with 0.05 HGPI units of inhibitor for 5 min and 3 h and analyzed by MALDI-TOF to detect the changes in the four normal peaks of rCanPI. Bovine trypsin and chymotrypsin do not act on the linker regions in the rCanPI-7 whereas HGP cleaves on the linkers in turn processing the multi-IRD forms to lower IRD forms.

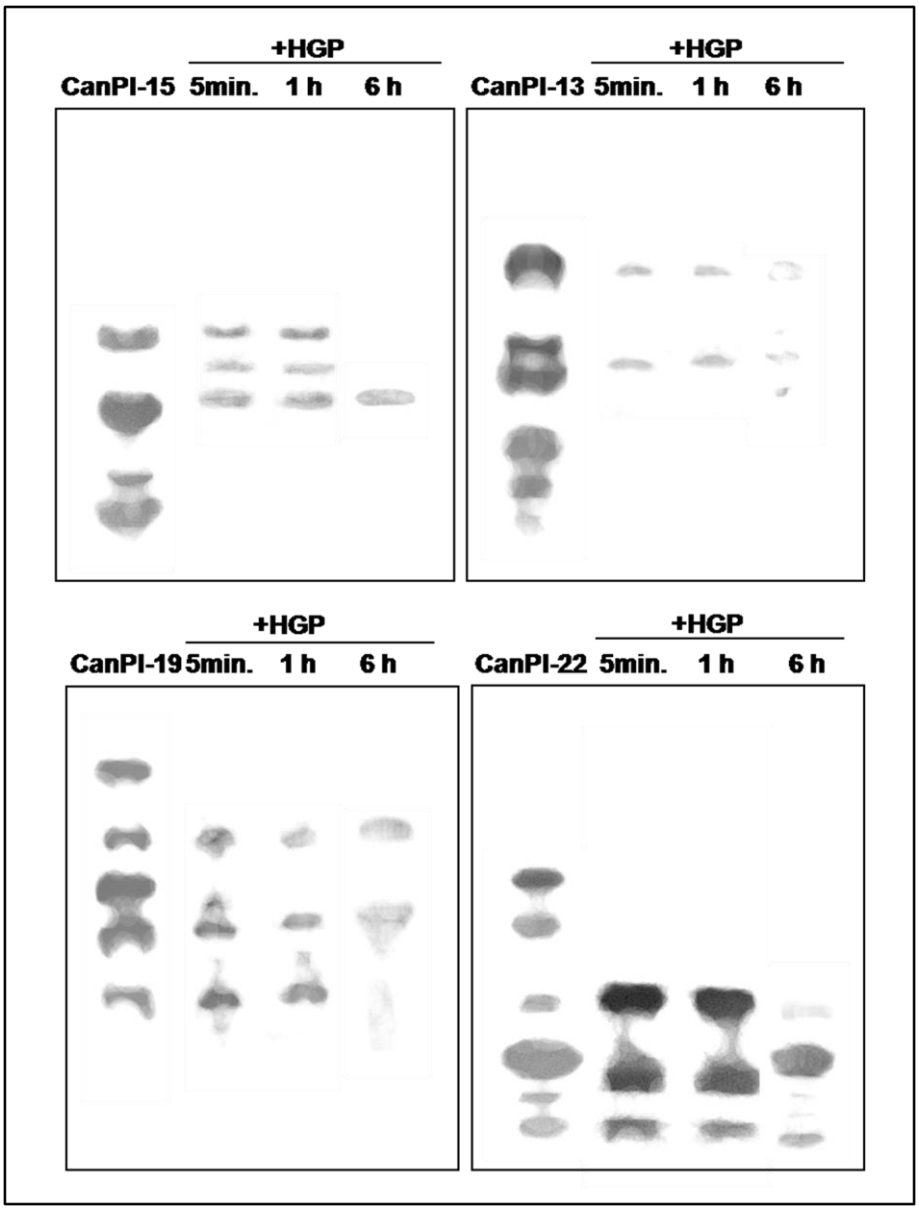


Figure 3.10: *In vitro* stability of one and/or two domain rCanPIs towards HGP. Equal HGPI units (0.5) each of CanPI-15, - 13, -22, -19 were incubated with 1 HGP unit at 24^{o}C for 5 min (lane 2), 1 h (lane 3), 6 h (lane 4) and the reaction mixtures were resolved on 12% native PAGE. Each rCanPI without HGP treatment (lane 1) was loaded as control. The gels were processed for TI activity visualization by GXCT.

and isoforms exhibiting low intensity became stronger. TI activity profiles indicated the processing and partial proteolysis of rCanPIs. The stability of individual rCanPI varied in the proteolytic environment.

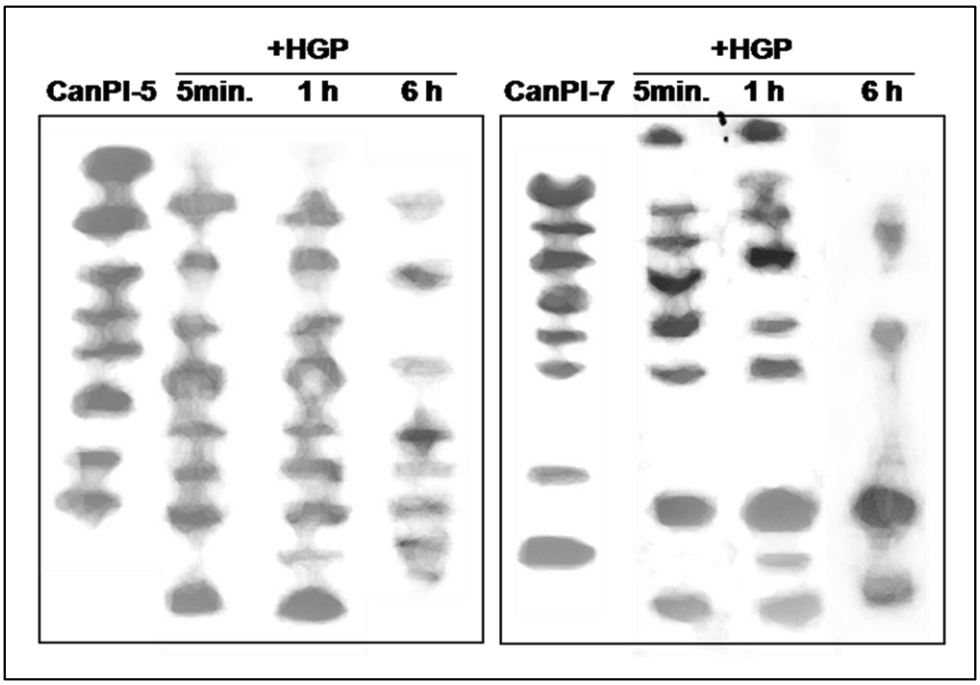


Figure 3.11: *In vitro* stability of multi-IRD rCanPIs towards HGP. Equal HGPI units (0.5) each of CanPI-5 and -7 were incubated with 1 HGP unit at 24^{o}C for 5 min (lane 2), 1 h (lane 3), 6 h (lane 4) and the reaction mixtures were resolved on 12% native PAGE. Each rCanPI without HGP treatment (lane 1) was loaded as control. The gels were processed for TI activity visualization by GXCT.

HGP of 4th instar larvae displayed at least 11 protease isoforms (HGP-1 to HGP-11 in **Fig. 3.12**). Inhibition of HGP isoforms by rCanPIs was studied by performing protease activity visualization of HGP pre-treated with rCanPI (**Fig. 3.12**). HGP-3 to HGP-6 were inhibited by all rCanPIs whereas HGP-2, -7 and -10 remained active in each case. HGP-1 was inhibited by all rCanPIs except 1-IRD rCanPIs, which unexpectedly showed even higher protease activity of this HGP isoform. HGP-9 was differentially inhibited by rCanPI-5, -7 and -19. It was interesting to note that HGP-8 was inhibited effectively only by rCanPI-5 and -13, which share one unique IRD sequence (IRD-12). Overall, most of the isoforms of HGP were inhibited by rCanPI-5 and -7 (**Fig. 3.12-lane 5, lane 6**).

Gut extracts of the 4th instar *H. armigera* larvae fed on AD with rCanPI-5 and -7 were visualized for protease activity (**Fig. 3.13A**). Major reduction in the activity of protease isoforms HGP-4, -5, -9 and -10 was detected in the rCanPI-5 and -7 fed gut extracts as compared to the control (larvae fed on AD). HGP-7 activity was seen to be enhanced in inhibitor fed gut extracts.

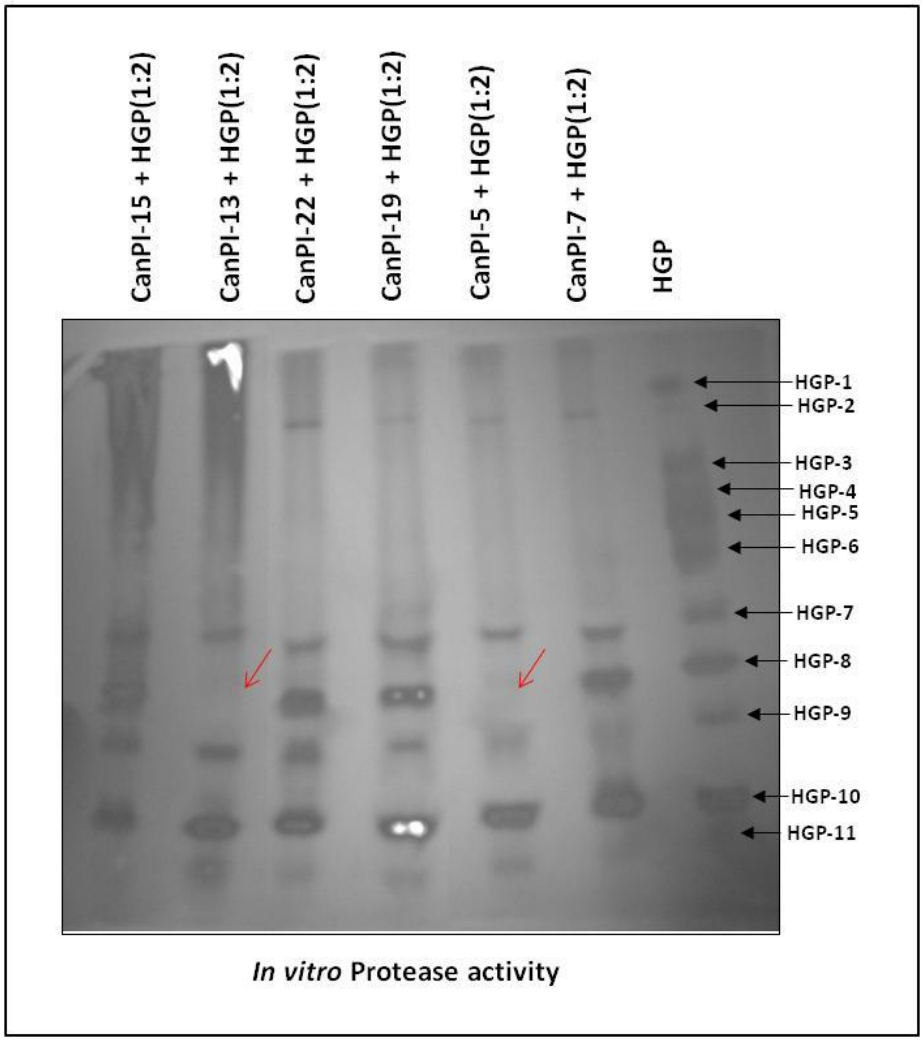


Figure 3.12: Comparative inhibition of HGP isoforms by various rCanPIs. Equal HGPI units of CanPI-15, -13, -22, -19, -5 and -7 were incubated with HGP for 30 min at 24 ^{o}C. The above reaction mixtures were then resolved on 8% native PAGE. The gels were processed for protease activity visualization by GXCT.

The appearance of novel protease isoforms, three in case of CanPI-7 and one in CanPI-5 fed insect gut were detected (indicated by arrows in **Fig. 3.13A**).

The rCanPI fed *H. armigera* gut extracts were visualized for presence of PIs. Three and two prominent PI isoforms were found in rCanPI-5 and -7 fed *H. armigera* gut tissue, respectively (**Fig. 3.13B**). These TIs correlated with *in vitro* profiles of HGP treated rCanPI-5 and -7.

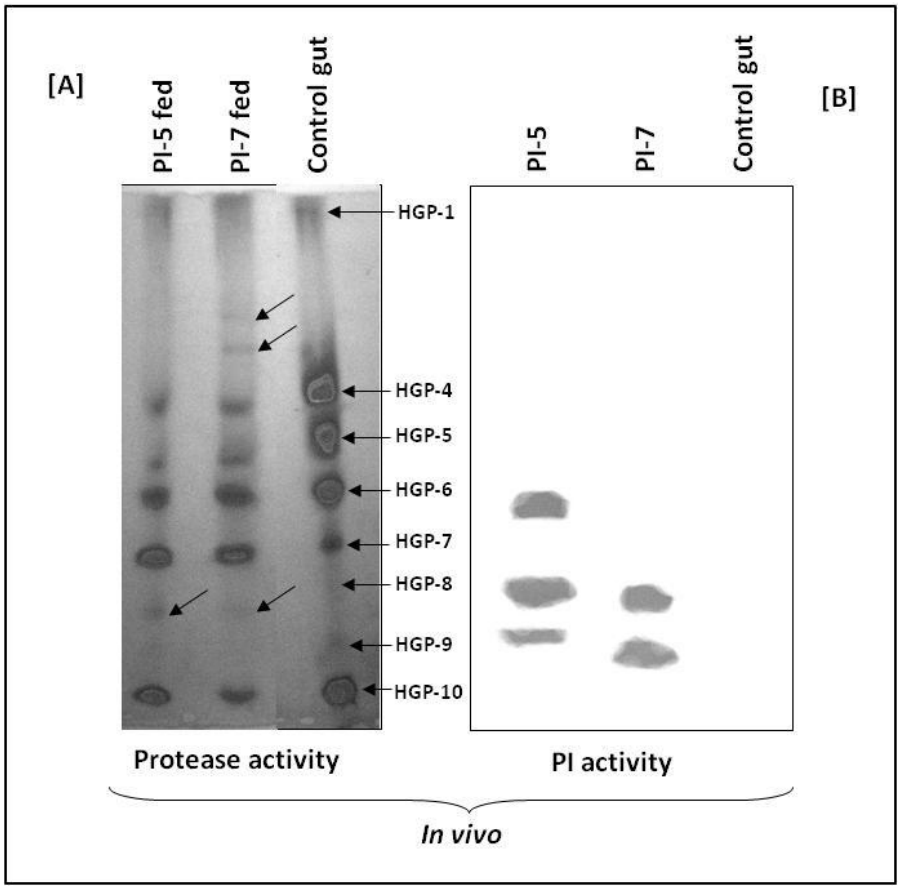


Figure 3.13A: Inhibition of HGP isoforms by rCanPIs. Equal volumes of gut extracts of rCanPI-5 and -7 containing artificial diet fed *H. armigera* were resolved on 8% native PAGE. The gel was processed for protease activity visualization.

Figure 3.13B: *In vivo*: Stability of rCanPIs in *H. armigera* gut. The rCanPI-5 and -7 fed *H. armigera* gut extracts were analyzed for TI activity visualization by GXCT after inactivation of the proteases by heat treatment at 70 ºC for 15 min. TI activity of both rCanPI-5 and -7 was detected in these extracts indicating their *in vivo* stability in *H. armigera* gut.

3.4. Discussion

3.4.1. MALDI-TOF analysis can be used to precisely determine the CanPI-protease interaction

P. pastoris expressed rCanPIs revealed multiple active isoforms formed due to processing of the mature precursor protein (Tamhane et al., 2007). MALDI-TOF-MS of 1-, 2-, 3- and 4-IRD rCanPIs showed 1, 2, 3 and 4 number of peaks, respectively corresponding to processed precursors (**Fig. 3.2**). The proteolytic processing of a precursor is a characteristic feature of Pin-II PIs. Plant endogenous proteinases efficiently processed the 6-IRD *N. alata* precursor into 1-IRD form as indicated by

weak 3 and 4-IRD isoforms and a strong 1-IRD isoform (Atkinson et al., 1993). The 5 aa linker regions in Pin-II PIs and also in CanPIs (QRNAK, EENAE, EASAE, EGNAE, EETQK) frames a hydrophilic loop that presents the protease processing site on the surface of the molecule (Greenblatt et al., 1989). Further processing of CanPIs by proteases towards termini enhances their molecular diversity (Horn et al., 2005).

In the present study, rCanPIs interacting with various proteases over a time period were analyzed by MALDI-TOF-MS. The mass spectra revealed the molecular mass of each of the precursors, processed IRDs and also the molecular diversity before and after rCanPI-protease interaction. On interaction with trypsin, no major variation in the diversity of mass peaks was observed while interaction with chymotrypsin showed additional peaks of 10 kDa and 15 kDa which might be because of cleavage within the IRDs (**Fig. 3.9**). In case of rCanPI-7 and HGP interaction, the 2-, 3-, and 4-IRD isoforms were efficiently processed into single IRDs indicating the processing of the precursor PIs at the linker regions followed by trimming of aa at either or both the termini (**Fig. 3.8**). For multi domain Pin-II PIs, it has been postulated that it is more difficult for simultaneous binding of all the domains to proteases without steric hindrance (Kong and Ranganathan, 2008). In the EGNAE linker, N↓A was the most commonly processed proteolytic cleavage site (**Table 3.1**). Several-fold processing of PIs in suitable proteolytic environment, like in plants and insect gut, leads to increase in its IRD diversity which may have functional significance with respect to modification of its inhibitory potential.

3.4.2. Sequence variation in the CanPIs influences their interactions with proteases

The eight fully conserved cysteines work as structural scaffold to hold the reactive site loop (RSL) in a relatively rigid conformation that helps to prevent proteolytic cleavage of the inhibitor upon interaction with proteases (Barrette-Ng et al., 2003a). The strength of the protease-PI interaction is determined by the compatibility of all aa residues (P4-P4'). It is interesting that the residues outside this loop, referred to as adventitious contacts, can also significantly affect the affinities of the inhibitor for closely related target proteases (Komiyama et al., 2003).

rCanPIs displayed variation in terms of their inhibition potential against proteases like trypsin and chymotrypsin, which correlated with the number of its TI

and CI domains. Over 90% inhibition of these was attained by rCanPI-7 (4-IRD) which had a combination of two TI and two CI IRDs (**Fig. 3.5**). rCanPIs with only TI IRDs (rCanPI-5, -13 and -19) also showed ~84% chymotrypsin inhibition; probably because of the cross reactivity as reported in earlier studies (Moura and Ryan, 2001). With equal rCanPI protein, HGP activity was 6 to 60%; the highest by rCanPI-5 and -7 and the lowest by rCanPI-15, respectively. Higher HGP inhibition by rCanPI-7 (4-IRD PI) at a lower protein concentration than rCanPI-13 (1-IRD PI) could be attributed to multiple and diverse IRDs in rCanPI-7. As an exception, rCanPI-15 (IRD-7; TI) revealed low inhibition efficiency against all proteases tested, when azocasein was used as a substrate. Interestingly, it showed about 60% HGP inhibition and other rCanPIs also showed around 80% HGP inhibition when synthetic substrate BA*p*NA was used. This showed that inhibitor efficiency was strongly influenced by nature of substrates present, and thus among other factors might show further differences in natural conditions.

Gut extract of *H. armigera* is a complex of several enzymes and various proteases (Patankar et al., 2001). Differential inhibition of the total 11 HGP activity isoforms (HGP-1 to HGP-11, **Fig. 3.12**) was performed by various rCanPIs with multi-IRD PIs showing higher suppression of HGP isoforms. Specific inhibition of HGP-8 only by rCanPI-5 or rCanPI-13 was noted which might be because of IRD-12. IRD-12 was not present in any other rCanPI and had high similarity with IRD-14 and IRD-15 (present in CanPI-7 and CanPI-22, respectively) with a difference of only single aa close to the reactive site (**Fig. 3.2**). Even then, CanPI-7 or -22 did not show inhibition of HGP-8 highlighting the functional significance of single aa changes in the IRDs. Similarly, HGP-9 was strongly suppressed by rCanPI-5 or rCanPI-7 compared to that by the other rCanPIs. This implies that one of the component IRDs of these rCanPIs (IRD-1, -4, -5, -10,-12 and -14) might be specific for inhibition of HGP-9. HGP-2 remained uninhibited by any of the rCanPIs indicating its insensitivity to any of the IRDs tested. Differential inhibition potential of Pin-II PIs can also be attributed to the orientation of inhibitory domains in space relative to each other (Lee et al., 1999). The diversity in CanPIs with respect to the number of IRDs per gene, their orientation in a PI and sequence variations in IRDs themselves points towards a more complex interaction of CanPIs with endogenous and/or insect gut proteases.

3.4.3. *In vivo* stability of CanPIs: Significance against *H. armigera*

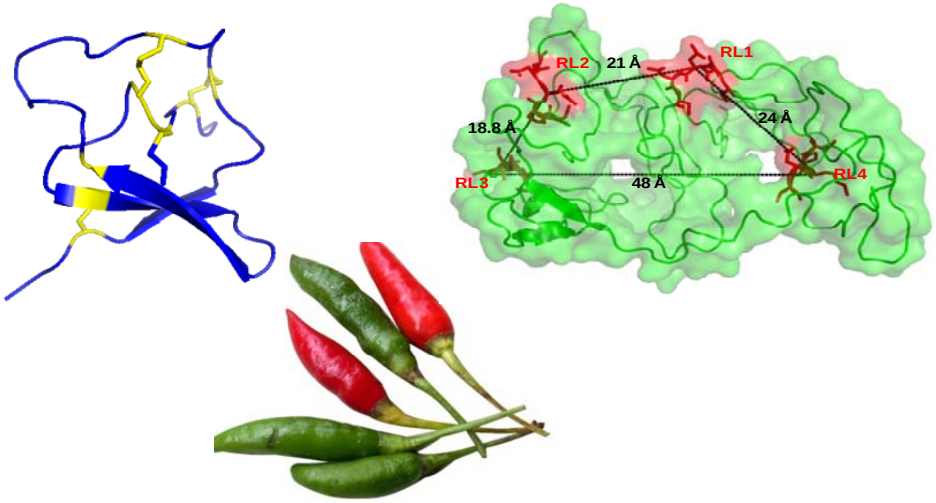
In the present study, the *in vitro* experiments to check the stability of each rCanPI indicated that HGP caused processing. In some cases (rCanPI-15 and -13) degradation of the PIs (**Fig. 3.10**) while in some cases such as, the processed IRDs (from rCanPI-5 and -7) remained stable even after prolonged incubation with HGP (**Fig. 3.11**). Experiments demonstrated the low efficiency of single IRD CanPIs versus multiple-IRD CanPIs against HGP; indicating the importance of presence of multiple IRD genes *in planta* for defense. This could be well correlated to the induced up regulation of higher IRD PI transcripts upon aphid and *S. litura* infestation (Tamhane et al., 2009). *In vivo* studies to determine the fate of CanPIs in insect precisely displayed active CanPIs in the gut (**Fig. 3.13B**). However, in some previous studies host plant PIs were degraded by pest gut enzymes and such PIs are, therefore, ineffective as pest control molecules (Giri et al., 1998). Furthermore, Harsulkar et al. (1999) have reported that PIs from non-host plants can effectively inhibit gut proteases and the larval growth. Stability in the insect gut environment is an important feature of PIs to be used in insect defense.

rCanPI-5 and -7 fed gut extracts showed inhibition of HGP-4, -5 and -10 and HGP-7 was seen to be over expressed (**Fig. 3.13A**). Additionally, induction of novel proteases (marked by arrows in **Fig. 3.13A**) that might be insensitive to rCanPIs was observed. The change in the expression of gut protease isoforms in response to rCanPIs signifies their dynamic interactions.

C. annuum possesses an array of Pin-II PI genes ranging from 1- to 4- IRD *CanPIs* with differential patterns of expression in various tissues. The 1- and 2-IRD PIs are predominantly found to be expressed in the stem tissue (Tamhane et al., 2009). Our study showed their degradation in presence of HGP, which infers that they might not have role in defense against insect, however, these PIs may have physiological role *in planta*. The inhibition potential of multi-IRD CanPIs due to their efficient processing by gut proteolytic machinery to produce a wide spectrum of structurally and functionally divergent IRDs as evidenced in our above studies, can be attributed to their involvement in defense. The up regulation of diverse multi-repeat CanPIs proposes their molecular co-evolution in response to pest attack (Lee et al., 1999). Further divergence within the single genes facilitates the targeting of diverse proteases. Thus, multi-domain protein(s) with various PI specificities may be the plants answer to the gut protease variants expressed by insects.

Chapter 4

Structural-functional insights of single and multi-domain *Capsicum annuum* proteinase inhibitors



CHAPTER 4

Structural-functional insights of single and multi-domain *Capsicum annuum* proteinase inhibitors

Abstract

Pin-II proteinase inhibitors (PIs) are the focus of research because of their large structural-functional diversity and relevance in plant defense. Two representative *Capsicum annuum* PI genes (*CanPI-15* and *-7*) comprising one and four inhibitory repeat domains, respectively, were expressed and recombinant proteins were characterized. β-sheet and unordered structures were found to be predominant in CanPI-15 while CanPI-7 additionally displayed the signatures of polyproline fold, as revealed by circular dichroism studies. Inhibition kinetics against bovine trypsin indicated three times higher potency of CanPI-7 (K_i ~57 µM) than -15 (~184 µM). Both the CanPIs exhibited varying inhibition of trypsin, chymotrypsin, elastase and total gut proteinase activity of two lepidopteran insects. Activity and structural stability of these CanPIs was revealed under various conditions of pH, temperature and denaturing agents. Structure prediction and docking studies of CanPI-7 with proteases exhibited that all the four reactive sites are exposed on unordered loops which facilitate its interaction with multiple target protease molecules, affirming its higher inhibition potency.

4.1. Introduction

The understanding of relationship between protein structure and function remains a primary focus in order to understand how proteins have been engineered for their varied functions. The useful structural information comes from the primary amino acid sequences and the associated secondary and tertiary structures. Usually, the sequence similarity confers the structural similarity and thus, the functional annotation of major fraction of proteins is based on such similarity (Hegyi et al., 2002). The 3-dimensional structure of a protein describes folding of its secondary structure elements and the specific position of each of the atoms in the protein (Voet and Voet, 2010).

The known protein structures have majorly come in to light through X-ray crystallographic or nuclear magnetic resonance (NMR) studies. The atomic coordinates are deposited in a database, Protein Data Bank (PDB) and allow the tertiary structures of variety of proteins to be analyzed and compared. Due to poor feasibility of experimental structural determination of all the proteins, comparative modelling has been a viable strategy to generate reliable models for the proteins based on their sequence similarity. Since the native structure of protein depends only on its amino acid sequence, it is also possible by *ab initio* methods. However, such methods have shown only moderate success as they have sometimes failed with larger polypeptides while in some cases yielded similar models corresponding to the experimentally determined structures. Protein secondary structure includes regular polypeptide folding patterns such as helices, sheets and turns and is widely studied by spectroscopic technique like Circular Dichroism (CD) spectroscopy. It requires dilute protein solutions and is a way to rapidly assess the conformational changes resulting from addition of ligands or other external factors like temperature or pH (Pelton et al., 2000). Such measurements provide valuable information for the confirmation of folding in expressed proteins, protein aggregation and stability. These can provide clues to the structure a protein may adopt. Moreover, the quantitative measurements of secondary structures can provide significant insight into structural features critical for biological function.

The three-dimensional structures of several Pin-II PIs, single- as well as two-domain, have been determined either by X-ray crystallography or NMR and they give a good outline of the structure and dynamics of this class (Greenblatt et al., 1989;

Nielsen et al., 1994 and 1995; Barrette Ng et al., 2003; Schirra et al., 2005). The structure of a single IRD (TI or CI domain) is dominated by triple-stranded β-sheets and turns and the active site is solvent exposed being placed on an unordered loop (Nielsen et al., 1994). In a two-domain PI from tomato, each IRD adopts a similar fold and displays an extended configuration presenting the two binding sites at opposite ends of the inhibitor molecule (Barette Ng et al., 2003a and b). However, there have been few structural studies on PI precursors with more than two domains, giving insights into their domain orientations, binding to proteases and stoichiometry (Heath et al., 1995; Schirra et al., 2005).

C. annuum Pin-II PIs (CanPIs), displaying high isoform diversity with PIs of 1- to 4-IRDs, have been isolated and characterized to assess their defense potential against lepidopteran proteases (Tamhane et al., 2007). Further studies have also revealed the induced accumulation of multi-IRD CanPIs in leaves in response to mechanical wounding and insect infestations (Tamhane et al., 2009). In the present study, single domain CanPI-15 (1-IRD) and multi-domain CanPI-7 (4-IRD) were selected for recombinant expression, biochemical and structural characterization. The specificity of these inhibitors to multiple target proteases was estimated and inhibition kinetics was studied to evaluate binding. PI proteins were further subjected to varied conditions of pH, temperature and denaturing agents and transitions in their structure were monitored by biophysical techniques. CD spectroscopy, *in silico* structure prediction and docking studies rendered insight into the conformational stability and/or flexibility, structure and binding mechanisms of multi-domain Pin-II PIs, respectively. In this chapter biochemical and structural reasons are reported for better efficiency of multi-domain inhibitors against target proteases and their high stability imparted by disulfide bonds assigning them key role in insect control strategies.

4.2. Materials and methods

4.2.1. Cloning of *CanPIs* in bacterial expression system

Based on sequence analysis, *CanPI-15* and *-7* were selected for the present study as representatives of single (1-IRD) and multi-domain (4-IRD) CanPIs, respectively. The cDNA encoding the mature peptide region of *CanPIs* (*CanPI-15* and *-7*) was cloned in ligation-independent cloning (LIC) compatible expression vector pMCSG7 (Eschenfeldt et al., 2009) for recombinant expression in *E. coli* (**Fig. 4.1A**). The vector contains a sequence encoding the tobacco etch virus (TEV) protease

cleavage site next to an *Ssp*I site used for LIC (**Fig. 4.1B**). The base vector, pMCSG7 appends an N-terminal hexahistidine tag to proteins that is followed by the protease recognition sequence.

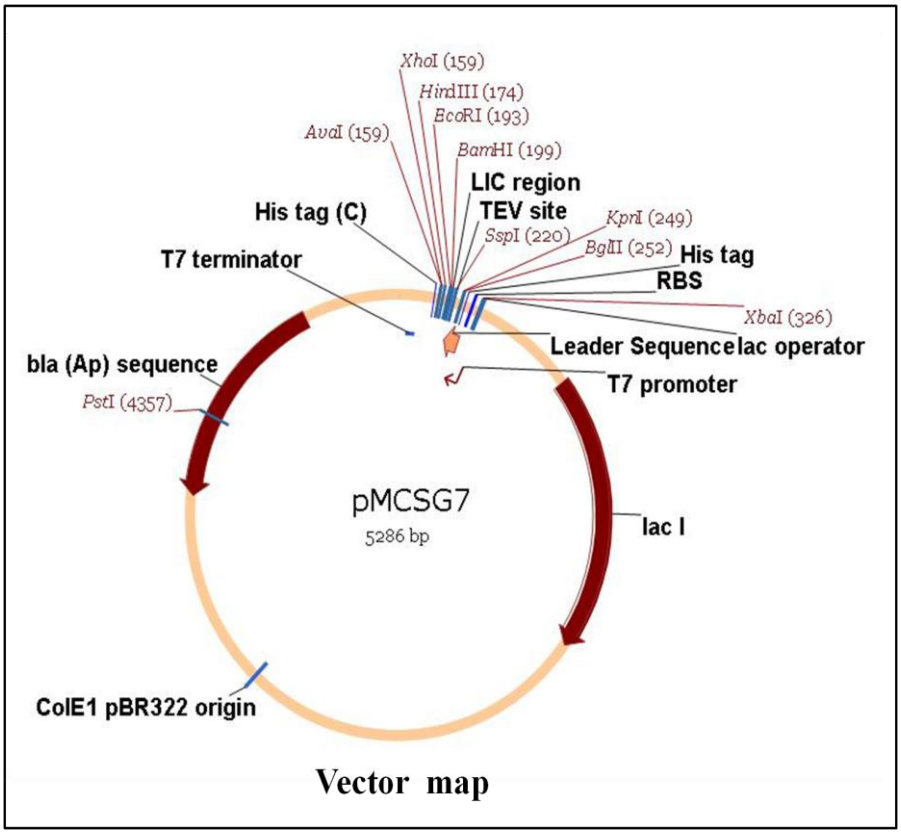


Figure 4.1A: Vector map of pMCSG vector

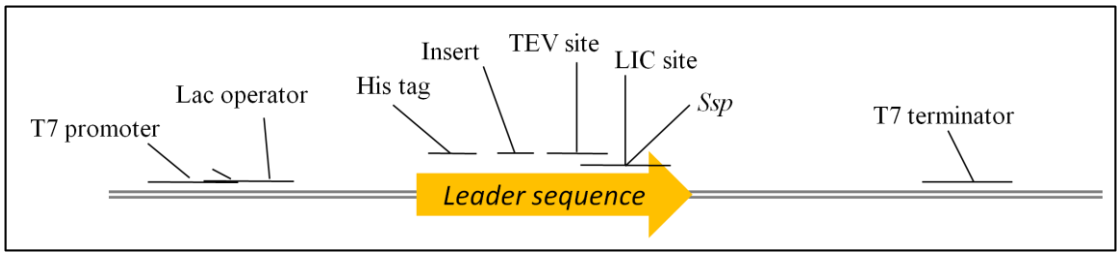


Figure 4.1B: Generalized organization of MCSG vectors. MCSG vectors encode an N-terminal leader sequence (*arrow)* that terminates in an LIC region centered on an *SspI* site.

4.2.2. Recombinant protein expression and purification

E. coli Origami B (DE3) cells were transformed with pMCSG7-CanPI-15 and pMCSG7-CanPI-7. The individual recombinant *E. coli* colonies were initially grown overnight at 37ºC in 10 ml LB medium supplemented with antibiotics (Ampicillin-100 µg/ml, Kanamycin-15 µg/ml, Tetracycline-12.5 µg/ml). The pre-culture was used to inoculate 1 Litre 'Terrific Broth' (TB) medium with appropriate antibiotics and

allowed to grow until the OD (600nm) reached 0.6-0.8. Cells were induced with IPTG (0.5 mM) overnight at 21°C and harvested by centrifugation.

The cell pellet was solubilized in ice cold cell lysis buffer A (50 mM Tris-HCl, pH 8.0; 300 mM NaCl; 2% glycerol) and disrupted by sonication (0.5 s pulse with 0.5 s intervals for 10 min) using an Ultrasonic Disruptor UD-201 (Tomy, Tokyo, Japan). The supernatant and pellets were separately collected by centrifugation for 45 min at 10,000g, 4°C (RS-4S rotor, Kubota, Japan). The supernatant was loaded on Ni-NTA resin (Qiagen, Valencia, California, USA) and purified using standard manufacturer's protocol. The fusion protein eluted with buffer B (50 mM Tris-HCl, pH 8.0; 300 mM NaCl; 2% glycerol; 250 mM imidazole). The expressed recombinant protein in inclusion bodies was purified from the pellet as described earlier (Oganesyan et al., 2004). The inclusion bodies were resuspended in buffer C (buffer A containing 6 M urea and 1mM DTT). The mixture was gently shaken overnight. After centrifugation, the resulting supernatant was loaded on Ni-NTA resin and washed with buffer B containing 20mM imidazole to remove non-specifically bound contaminants. For refolding, the column was washed with a gradient of urea concentration in buffer C (6 M to 0 M in 60 ml). The refolded fusion protein was eluted with buffer B containing 1mM DTT.

The His tag was cleaved using TEV protease at a protease to target protein ratio of 1:100 (w/w) at RT for 12 h. Additional Ni-NTA purification was performed to remove the cleaved tag and collect the protein in flow through. This was applied on pre-equilibrated (buffer: 50 mM tris-HCl, pH 8.0; 200 mM NaCl; 2% glycerol) Sephadex S-75 (size exclusion) for further purification.

4.2.3. Inhibitory activity assays and kinetic analysis

Gut tissues were dissected from the laboratory reared 4th instar larvae of *Helicoverpa armigera* and *Spodoptera exigua* and immediately frozen in liquid nitrogen. Extraction of HGP or SGP was performed as detailed previously (Patankar et al., 2001) and the bovine trypsin, chymotrypsin, elastase, HGP and SGP activities were determined using Azocasein as a substrate as given in Tamhane et al. (2005). Inhibitory assays using rCanPIs against all the above proteases were performed as detailed in Tamhane et al. (2005) with increasing amount of PIs (0.2 to 8 µg) to find out the PI concentrations required for maximum inhibition. One PI unit is defined as the amount of inhibitor required for inhibiting one protease activity unit. Trypsin

inhibitory (TI) activity assays for rCanPI proteins under varying pH and temperature conditions were performed using BA*p*NA as a substrate as described in Tamhane et al. (2005).

Michaelis-Menton constant (Km) for trypsin was calculated by using various concentrations of BA*p*NA substrate (1 to 5 mM), and then plotting double reciprocal curve with 1/ (v) and 1/ (s). Kinetic properties of rCanPIs were analyzed over a range of concentration of inhibitors (0.1 to 50 µg). IC_{50} value for each inhibitor was calculated from the sigmoid curve. K_i value for each inhibitor was calculated directly from the IC_{50} value using Cheng-Prusoff's classical equation (Cheng and Prusoff, 1973).

$$K_i = \frac{IC_{50}}{1 + \frac{[S]}{K_m}}$$

4.2.4. Relative binding free energy (Experimental) calculation

The experimental free energy of binding was approximated using

$$\Delta G_{\text{bind}} = \text{RT ln } K_i$$

Where, R is the Gas constant and T is the absolute temperature. Using Cheng-Prusoff equation for competitive inhibition, IC_{50} value was converted into an inhibition constant K_i.

4.2.5. Fluorescence measurements and ANS binding studies

rCanPI protein samples (100 µg/ml) were incubated in an appropriate buffer over the pH range of 2.0-10.0 for 4 h at 28^{0}C. The following buffers (25 mM) were used for these studies: Glycine –HCl for pH 1-3, acetate for pH 4-5, phosphate for pH 6-7, Tris-HCl for pH 8-9 and Glycine-NaOH for pH 10-12. pH of the reaction remained stable even after 4 h. Appropriate aliquots from samples were used to check for TI activity using BA*p*NA as a substrate and fluorescence measurements were performed on Perkin Elmer LS 50B luminescence spectrometer at 28°C. The rCanPI protein solutions were excited at 280 nm and emission was recorded in the range of 310-400 nm. Slit widths of 7 nm each were set for excitation and emission monochromators and the spectra were recorded at 100 nm/min. To eliminate the background emission, the signal produced by the buffer solution was subtracted.

The intermediate states between denatured and native protein incubated in 25 mM buffers of pH range 2.0-10.0 were analyzed by the hydrophobic dye 1-anilino-8-naphthalenesulfonate (ANS) binding. ANS is a dye which has been shown to bind to solvent-exposed hydrophobic regions in a protein and shows increased fluorescence intensity and blue shift in the λ_{max} of emission (Semisotnov et al., 1991). The final ANS concentration used was 50 µM, excitation wavelength was 375 nm and total fluorescence emission was monitored between 400 and 550 nm. Reference spectrum of ANS in buffer was subtracted from the spectrum of the sample.

4.2.6. Circular dichroism (CD) measurements

The CD spectra of the proteins were recorded on a J-815 spectropolarimeter (Jasco, Tokyo, Japan) at 28°C in a quartz cuvette. Each CD spectrum was accumulated from three scans at 100 nm/min with a 1 nm slit width and a time constant of 1 s for a nominal resolution of 0.5 nm. Far UV CD spectra of the rCanPI proteins (125 µg/ml) were collected in the wavelength range of 195-300 nm using a cell path length 0.1 cm for monitoring the secondary structure. All the spectra were corrected for buffer contributions and observed values were converted to mean residual ellipticity (MRE) in deg cm^2 $dmol^{-1}$ defined as

$$MRE = M\theta_\lambda / 10dcr$$

Where, M is the molecular weight of the protein, θ_λ is CD in millidegree, d is the path length in cm, c is the protein concentration in mg/ml and r is the number of amino acid residues in the protein. Secondary structure elements were calculated by using CD pro software.

rCanPI protein samples (125 µg/ml) were incubated in 25 mM concentration of above mentioned buffers of varying pH and used for CD measurements. Effect of temperature on proteins was carried out by incubating the protein samples (125 µg/ml) in the temperature range of 24-99^{0}C for 2 h and measuring the CD spectra. Protein samples (125 µg/ml) were incubated in 0-8 M GdnHCl solution for 2 h. CD spectra were measured as described above.

4.2.7. Structure prediction and assessment for CanPIs

Structure of CanPI-15 was predicted by homology modelling, based on its maximum sequence identity (79%) with chymotrypsin inhibitor-1 structure from Russet Burbank potato tubers at 2.1 Å resolution (PDB ID: 4SGB_I). Due to absence of any such homologous templates for CanPI-7, various modelling servers like Quark, Rossetta and I-TASSER were used for *in silico* modelling of CanPI-7. 3D structure of CanPI-7 was obtained using the automated I-TASSER service available at the site (http://zhang-lab.ccmb.med.umich.edu/I-TASSER/). The online procedure yielded the 3D model on the basis of multiple-threading alignments by LOMETS and iterative TASSER simulations where it uses multiple PDB structures depending on its structural conservation, to model different parts of proteins (Zhang, 2008; Roy et al., 2010 and 2011). The best model was selected from the output based on C-score (C-score is a confidence score for estimating quality of the predicted models by I-TASSER). An initial structural model was generated, and subjected to an energy minimization procedure with GROMOS96, implemented in PyMol (The PyMol Molecular Graphics System, Version 1.2r3 pre, Schrödinger LLC) to reduce poor Van der Waals contacts and correct the stereochemistry of the model. The quality of the model produced was assessed by checking the protein sterology using the PROCHECK and the energy was checked by ProSA (www.ebi.ac.uk).

4.2.8. Macromolecular docking of CanPI-7 with bovine trypsin and chymotrypsin

A docking study was conducted to evaluate the binding conformation and interaction of CanPI-7 with trypsin and chymotrypsin. Modelled structure of CanPI- 7 was refined by energy minimization and restrain relaxation. In order to perform protein-protein docking, models of CanPI-7 and bovine trypsin (PDB ID: 1S0Q) and chymotrypsin (PDB ID: 1YPH) were submitted separately to online server GRAMMX (Tovchigrechko et al., 2006) following the standard protocols for each package. GRAMMX is an interactive protein docking and molecular superposition program which includes a grid-projection of a smoothed Lennard-Jones potential, combined with a post docking procedure of rigid-body energy minimization and structure clustering. Correlation technique by Fast Fourier Transformation was used to determine the best surface match between the molecules. An important feature of GRAMMX is the ability to smooth the protein surface representation to account for

possible conformational change upon binding within the rigid body docking approach. Successively, two trypsin and two chymotrypsin molecules were docked using standard protocol of GRAMMX and the complexes were analysed to study the probable mechanism of binding.

4.3. Results and Discussion

4.3.1. Recombinant expression of disulfide rich *CanPI-15* and *-7*

The diagrammatic representation of the selected single (*CanPI-15*) and multi-domain (*CanPI-7*) PI genes from *C. annuum* and sequence alignments of the constituent five unique IRDs are shown in **Fig. 4.2A**. CanPI-15 comprised a single trypsin inhibitory domain (TI) while CanPI-7 had two TI and two chymotrypsin inhibitory (CI) domains. Each IRD was 50 aa polypeptide and consisted of 8-conserved cysteine (C) residues contributing to four disulfide bonds and a single reactive site (P1). The core reactive site of an IRD, 'PKN'or 'PRN' for TIs and 'TLN' for CIs was confined by two conserved cysteine residues.

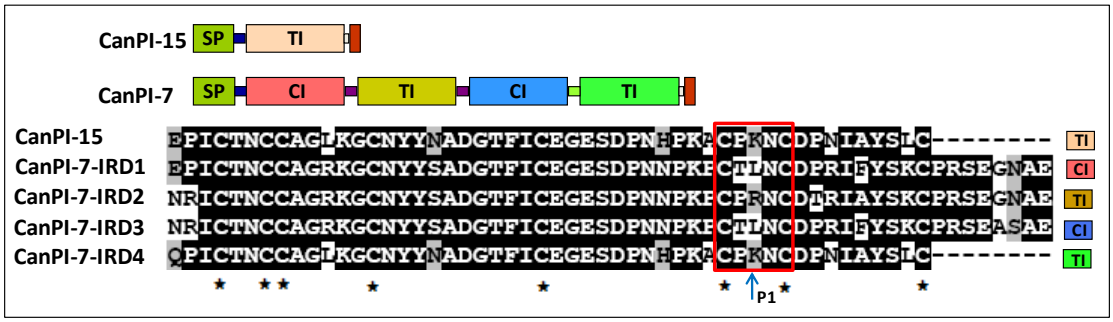


Figure 4.2A: Diagrammatic representation highlighting the gene structure of the two *CanPI*s from *C. annuum*, with their signal peptide sequence (SP), various IRD(s), linker region(s) and the stop codon. The multiple sequence alignment of deduced aa sequences of unique IRDs from these CanPIs are further shown.

E. coli Origami B (DE3) was chosen as a host to produce the recombinant proteins since each IRD had eight conserved cysteine residues forming four intramolecular disulfide bonds (Venturi et al., 2002; Oganesyan et al., 2004; Xiong et al., 2005). Origami B (DE3) is a modified strain that has mutations in both, the thioredoxin reductase (*trxB*) and the glutathione reductase (*gor*) genes. Thus, it maintains the oxidizing cytoplasm facilitating the S-S bond formation in the

recombinant protein in a proper order. Simultaneously, the expression of *CanPIs* in BL21 (DE3) pLys and Rosetta (DE3) pLys cells resulted in production of inactive recombinant proteins. Results indicate that proper formation of S-S bond is required for retaining the inhibitor activity and that could depend on the choice of host strain. rCanPI protein expressed in Origami B (DE3) cells was obtained in both, soluble fraction and in inclusion bodies. The soluble fraction was purified and final preparations yielded single protein corresponding to 6 kDa for CanPI-15 and 24 kDa for CanPI-7 (**Fig. 4.2B**). The proteins in insoluble fraction were purified from inclusion bodies under denaturing conditions.

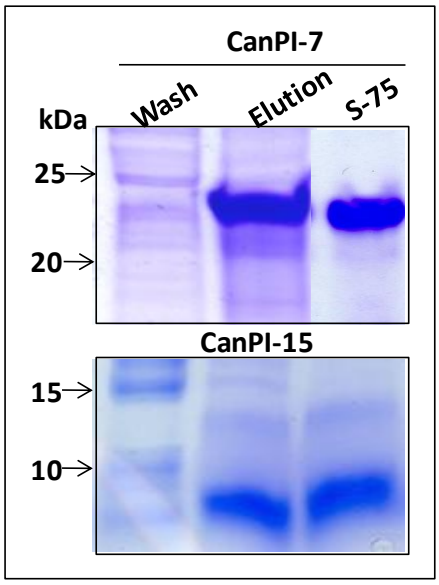


Figure 4.2B: Purification of recombinant CanPI-15 and CanPI-7. Proteins were separated on 15% SDS-PAGE and stained with Coomassie Blue R-250

CD Pro analysis of the far UV CD spectrum (**Fig. 4.3**) yielded the values of secondary structure elements as: α-helix-3.8%, β-sheet-41.4%, turns-21.2% and unordered- 33.4% for CanPI-15. The single negative band between 208 to 210 nm in CD spectrum and the CDPro analysis demonstrated that CanPIs were β-sheet rich and unordered proteins. CanPI-7 exhibited signatures of a polyproline fold (PPII) structure, bearing a small positive band near 227 nm and a large negative band between 208 to 210 nm with a cross over at 223 nm (**Fig. 4.3**). A typical PPII structure is characterized by negative band around 200 nm and a weaker positive band at about 217 nm. Typically this kind of structure is adopted by proteins rich in proline/ hydroxyproline but, even sequences not rich in proline can also take up this structure.

A shift in the positive band towards 225 nm is observed for peptides rich in proline/ hydroxyproline (Young et al., 1975; Bochicchio et al., 2002). Earlier, several residues with PPII structure within the long unordered loops in Bowman-Birk protease inhibitor have been reported by Raman optical activity spectroscopy (Smyth et al., 2001). Previous studies on the structure of Pin-II PIs, either 2-domain precursor or individual IRD(s) have shown a disordered loop containing the reactive site, a triple stranded beta sheet at its base and anchored by four conserved disulfide bonds (Nielsen et al., 1995; Barrette-Ng et al., 2003; Schirra et al., 2005).

The CD spectrum of the refolded CanPI-7 displayed similar secondary structural features as native CanPI-7 except a slight rearrangement (**Fig. 4.3**), indicating proper refolding through On-Column method. Earlier report, using Origami B (DE3) cells for production of disulfide rich protein have found that chymotrypsin inhibitors can maintain their activity in urea-containing buffer (Oganesyan et al., 2004). Owing to the easier refolding of rCanPI-7 and similar CD spectrum obtained in the present study, we speculated that purification under denaturing conditions kept the structure of rCanPI-7 intact, only with slight distortion.

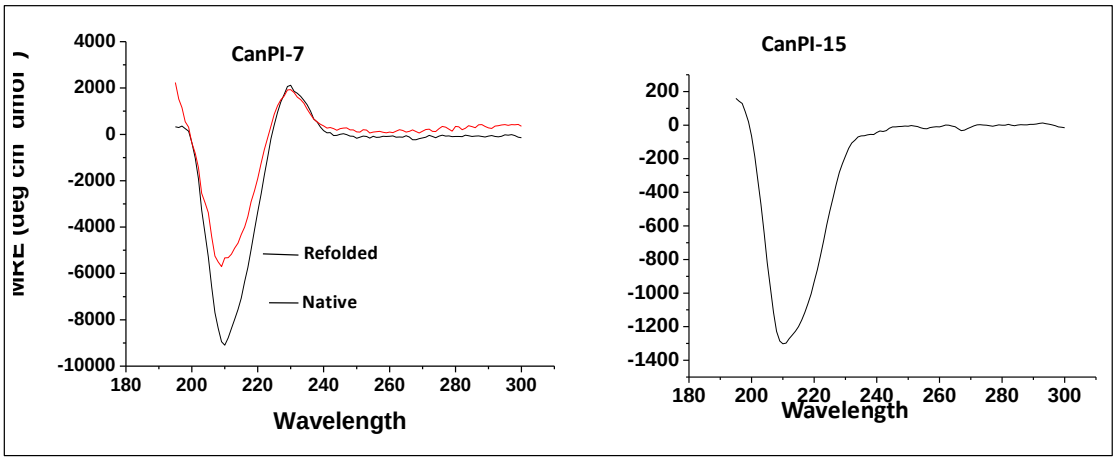


Figure 4.3: Circular dichroism study of recombinant proteins. Far UV CD spectra of rCanPI-7 (125 µg/ml) (native-purified from soluble fraction; refolded-purified from inclusion bodies) and rCanPI-15. rCanPI-15 shows a lower MRE than rCanPI-7.

4.3.2. Inhibitory activities and kinetic analysis of CanPIs

As expected, kinetic studies revealed stronger inhibition of trypsin by CanPI-7 (IC_{50} , 89 µM) than CanPI-15 (IC_{50} , 277 µM). The inhibition of trypsin followed a

sigmoidal pattern with increasing concentrations of the inhibitors (**Fig. 4.4A**). The inhibition constant (K_i), was determined by using the classical Cheng and Prusoff equation where CanPI-7 with low K_i (~57.42 µM) turned out to be a more potent inhibitor and indicated 'tight binding' than CanPI-15 (K_i ~184.66 µM). The IC_{50} and K_i values reflected 1:3 ratio of the molar quantities required for inhibition of equal amounts of trypsin and might be because of more binding sites in CanPI-7 (4 in number) than that in CanPI-15 (1 in number). Relative free energies of binding for CanPI-15 and -7 were -2.29 and -2.60 Kcal/mol, respectively. These values indicated that CanPI-7 forms tight complex with proteases as compared to CanPI-15. This highlights the exponential inhibitory effects that could be attained by use of multi domain PIs against proteases.

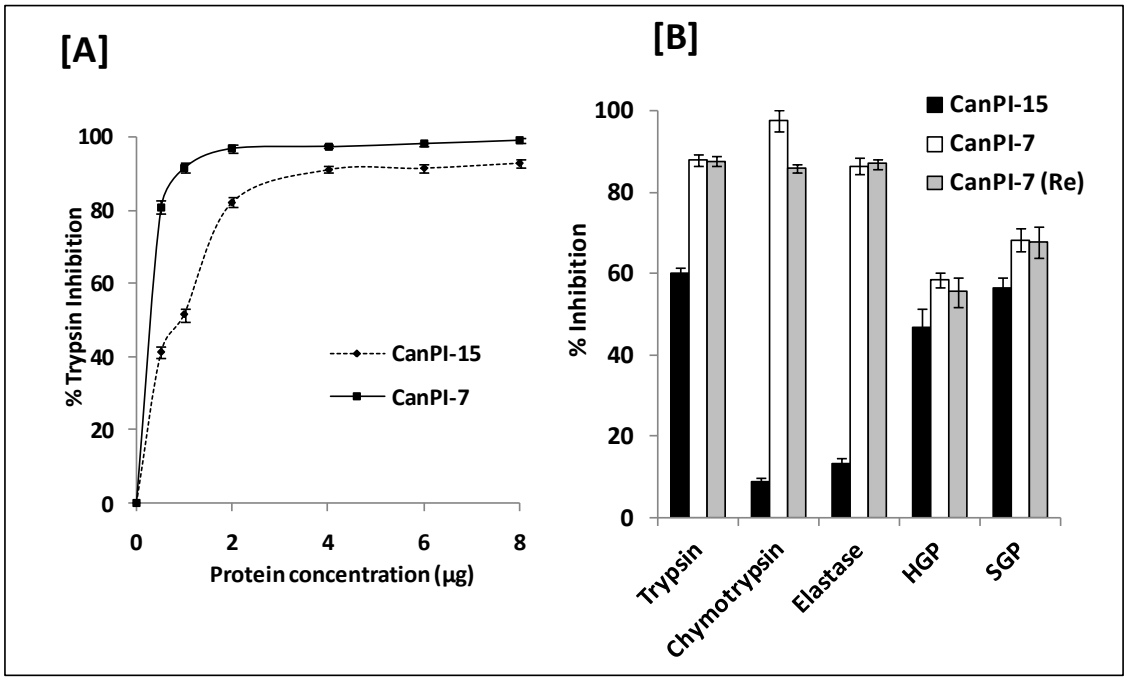


Figure 4.4A: Inhibition kinetics of CanPI-15 and -7 against bovine trypsin. The inhibition of trypsin follows a sigmoidal pattern with increasing concentration of the inhibitors.

Figure 4.4B: Maximum percent inhibition of bovine trypsin, chymotrypsin, elastase, HGP and SGP by optimum concentration of rCanPIs in azo-caseinolytic assays. Each value is an average of six replicates. CanPI-7 showed saturating inhibition at three times lower concentration than that of CanPI-15 against all proteases.

rCanPI proteins displayed variation in terms of their inhibitory potential against trypsin, chymotrypsin, elastase and proteases of 4th instar larvae of *H. armigera* and *S. exigua* (**Fig. 4.4B**). The inhibition patterns correlated to the number

of TI and CI domains in the individual protein. CanPIs showed 60 to 90% inhibition of bovine trypsin activity. As expected, rCanPI-15 without CI site, showed marginal inhibition of chymotrypsin (8%) whereas CanPI-7 which has 2 CI sites, showed 95% chymotrypsin inhibition. Elastase was inhibited to 85% by CanPI-7 while, CanPI-15 could inhibit only 13% of its activity. All the PI proteins exhibited 55 to 70% inhibition of insect proteases *viz*. HGP and SGP activity (**Fig. 4.4B**). CanPI-7 (4-IRD) attained higher inhibition of all the proteases at 3-fold lower concentrations than that of CanPI-15. The refolded CanPI-7 also showed equivalent inhibitory properties as native CanPI-7, against all the proteases (**Fig. 4.4B**) confirming that not only the structure but activity is also absolutely retained. It is important to note that CanPIs could effectively inhibit the gut proteases of *H. armigera* and *S. exigua* which are complex mixtures of several isoforms of trypsin, chymotrypsin and elastases. These results demonstrate that, PIs with multiple and varying specificities are better approach for tackling a wide blend/range of insect gut proteases.

4.3.3. Stabilities of CanPI-15 and -7 to pH, temperature and denaturing agents

CanPI-15 and -7 showed stability over a wide range of pH conditions varying from 2.0 to 10.0, based on activity and secondary structure while pH 7.0 to 8.0 was the optimum range for inhibitory activity (**Fig. 4.5A**). CanPIs exhibited ~10% loss of inhibitory activity against trypsin at acidic pH 2.0. The ellipticity and the minima at 210 nm, for both the rCanPIs at pH 2.0 and 8.0, remained unaltered indicating no loss of the native secondary structure (**Fig. 4.6A and B**). Further, rCanPIs were checked for ANS binding to look for any partial unfolding of the proteins under acidic pH conditions. CanPI proteins incubated in buffer of pH 2.0, showed two fold increase in fluorescence intensity (**Fig. 4.5B**) with blue shift of λ_{max} to 500 nm in presence of ANS, indicating exposure of hydrophobic patches on the surface of the protein. However, this alteration in the structure did not seem to affect the activity.

CanPI-7 retained its inhibitory activity against trypsin up to 80°C of treatment for 2 h while CanPI-15 began gradual loss of activity starting from 60°C (**Fig. 4.5C**). At higher temperatures, 90°C to 99°C, both the proteins displayed progressive loss of activity resulting into complete denaturation. In accordance to the retained activity, the MRE of both the CanPIs at 210 nm was unaltered up to 60°C and showed 15 to 25% gradual loss of secondary structure at 70 to 90 °C of treatment

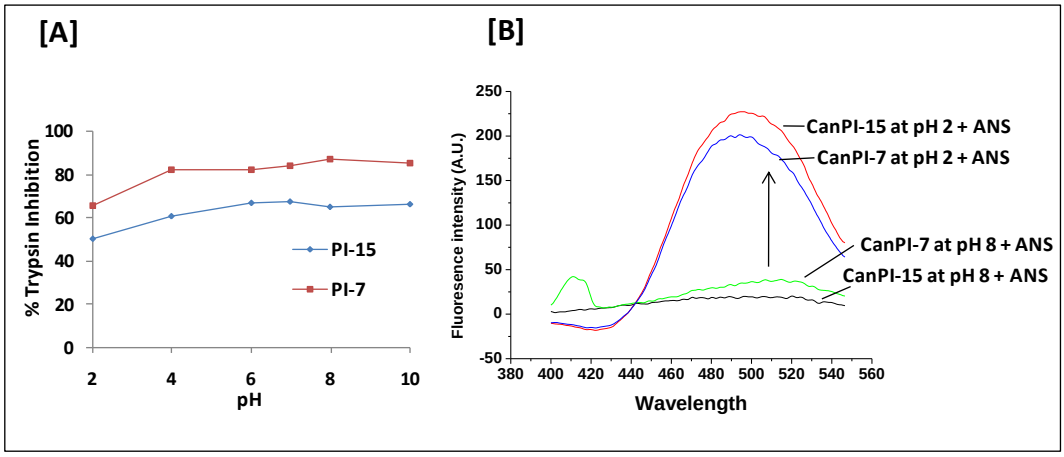


Figure 4.5A: Activity of rCanPIs under varying conditions of pH. Percent inhibition of bovine trypsin by rCanPIs incubated for 4 hrs in buffers of varying pH from 2.0 to 10.0.

Figure 4.5B: ANS binding studies of rCanPIs under varying conditions of pH. ANS binding of the protein samples (100 µg/mL) after incubation in buffers of pH 2.0 and 8.0. Increase in ANS intensity and blue shift is observed in rCanPIs at pH 2.0.

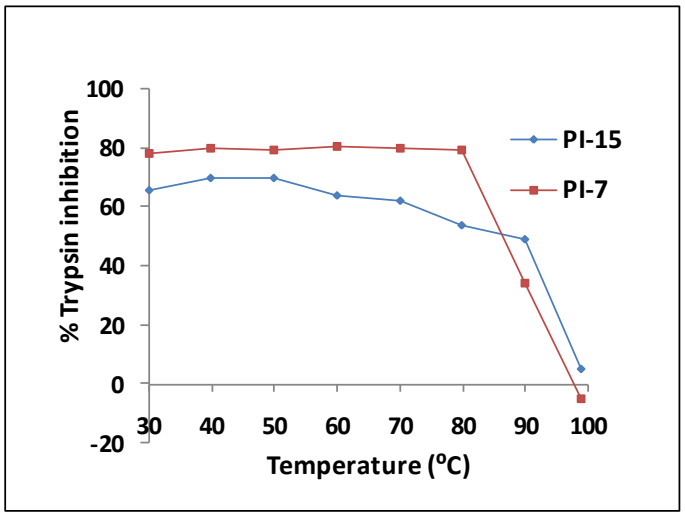


Figure 4.5C: Percent inhibition of bovine trypsin by rCanPIs incubated for 2 hr at varying temperatures from 30^0C to 99^0C.

(**Fig. 4.6C and D**). The minima at 210 nm shifted to 212 nm indicating formation of more of a β-sheet structure. The decrease in the positive band at 227 nm observed with increasing temperature supports the assignment of PPII fold in CanPI-7 (**Fig. 4.6D**). The high thermal stability of CanPIs could be due to the presence of four disulfide bonds in a single IRD. The unusual heat resistance of various proteins has been

attributed to the presence of cysteines resulting in to intra-domain or inter-domain disulfide bonds (Wetzel et al., 1988; Luckey et al., 1991; Fass et al., 2011).

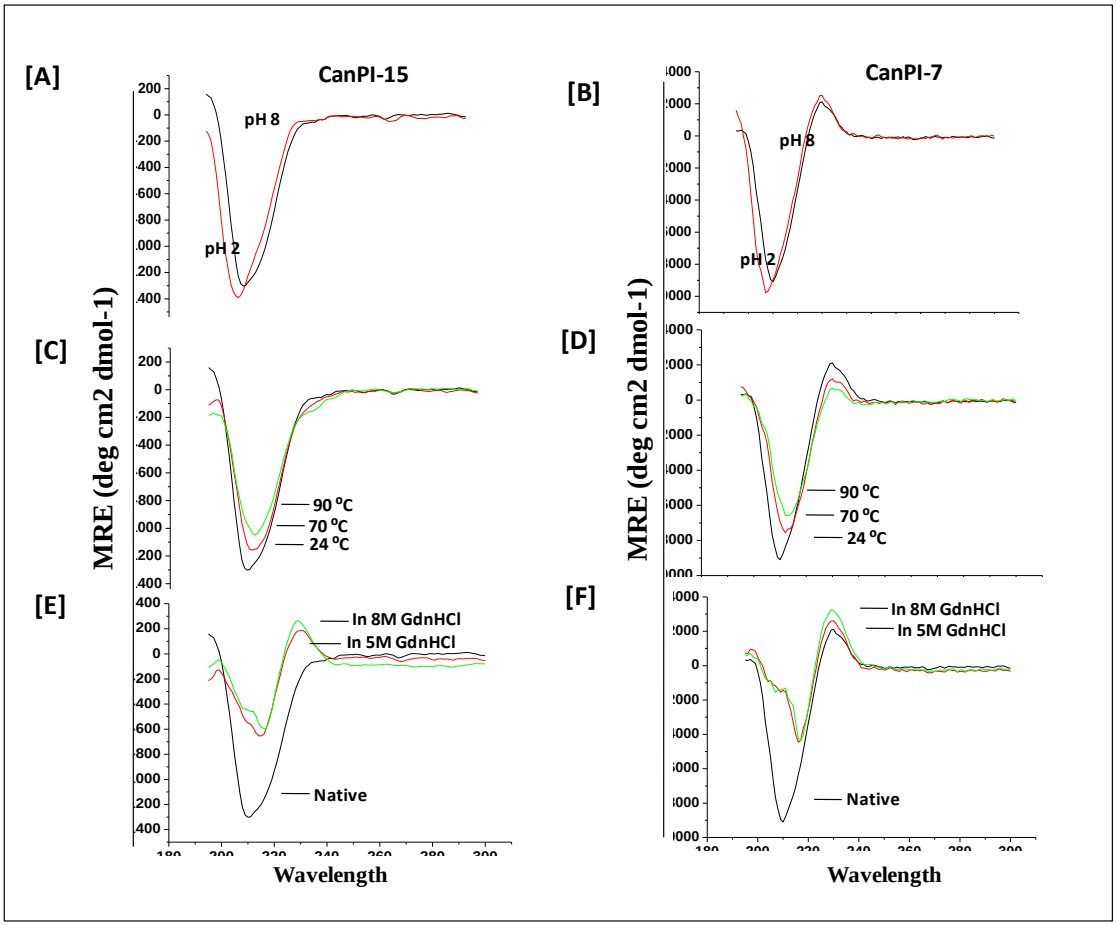


Figure 4.6: CD-spectra of rCanPIs under varying conditions

[A], [B]: Far UV CD spectra of rCanPI-15 and -7 (125 µg/ml) incubated in buffers of pH 2.0 and 8.0, respectively.

[C], [D]: Far UV CD spectra of rCanPI-15 and -7 (both 125 µg/ml) incubated at varying temperature, respectively.

[E], [F]: Far UV CD spectra of native rCanPI-15 and -7 (both 125 µg/ml) and 5 M and 8 M GdnHCl treated proteins.

The change in secondary structure of CanPIs treated with Guanidine HCl (GdnHCl) was monitored using CD spectroscopy in presence of various concentrations of the denaturant. Both, CanPI-15 and -7 showed apparently 50% loss of secondary structure in presence of GdnHCl from 5 M to 8 M concentration as indicated by the loss of ellipticity at 210 nm (**Fig. 4.6E and F**). CD spectra of CanPI-

15 showed positive band at 227 nm with higher concentrations of GdnHCl indicating the generation of polyproline fold (**Fig. 4.6E**). This implied that although CanPI-15 did not show any indication of PPII fold in its native CD profile, but is prone to adopt such kind of structure under specific conditions. CanPI-7 also displayed an increase in the 227 nm peak at progressively higher GdnHCl concentrations supporting the earlier hypothesis that formation of PPII like structure is promoted in presence of denaturing agent (Bochicchio et al., 2002) (**Fig. 4.6F**). This structural rearrangement is unique and has been observed in very few proteins like Human tropoelastin, Titin, Bowman-Birk inhibitor (Smyth et al., 2000; Bochicchio et al., 2002).

4.3.4. *In silico* structure prediction of CanPI-7

Predicted structure of single domain CanPI-15 was found to be similar with the NMR structure of *Nicotiana alata* trypsin inhibitor (1TIH) and some other studies in our laboratory showed that it forms complex with a single molecule of trypsin. CanPI-7 is a multi domain PI (4 IRDs), with high potencies of protease inhibition. The results of the secondary structure prediction showed that β-sheets and unordered loops were the major secondary structures of CanPI-7, representing 19% and 80%, respectively. We had used *de novo* protein modelling approach due to absence of X-ray crystallographic or NMR structure of any multiple domain Pin-II PI molecule possessing more than two domains. Out of the five models generated by I-TASSER, the most significant and acceptable model was selected based on the C-score and Ramchandran plot analysis of all the models. I-TASSER used various PDB structures like 1fybA, 1pjuA and 1oyvI to model specific parts of the query molecule, all these PDB(s) belonged to structure of Pin-II PIs. The selected best model (**Fig. 4.7**) had C score of -3.83, Tm score = 0.35±0.12 and RMSD = 13.4±4.0Å which are within the acceptable range for molecular modelling. This model represented all the four reactive sites (RL1, RL2, RL3 and RL4; **Fig. 4.7**) lying on unordered loops and exposed towards the outer surface of the CanPI-7 molecule. Distances between reactive loops are shown in **Fig. 4.7**. RL1 and RL2 were in close proximity to each other with a distance of only 21Å while RL3 and RL4 appeared distant with 48Å distance. Earlier structural studies on Pin-II PIs have been limited to 2-domain precursors from tomato (Barette Ng et al., 2003) and have shown the orientation of the two domains directly facing to each other leading to binding with two proteinase molecules independently. In case of multi domain PIs, the role of interdomain interactions has been speculated

as a key determinant of orientations of the domains relative to each other (Schirra et al., 2005). Presence of two to six TI domains in many Pin-II PIs has suggested the functional significance of the combination of IRDs within a single PI molecule. It was intriguing to speculate about the orientation and binding of the four domain inhibitor as in case of CanPI-7. PROCHECK and ProSA analysis, also validated that the modelled structure ID was acceptable for the further analysis (**Fig. 4.8**).

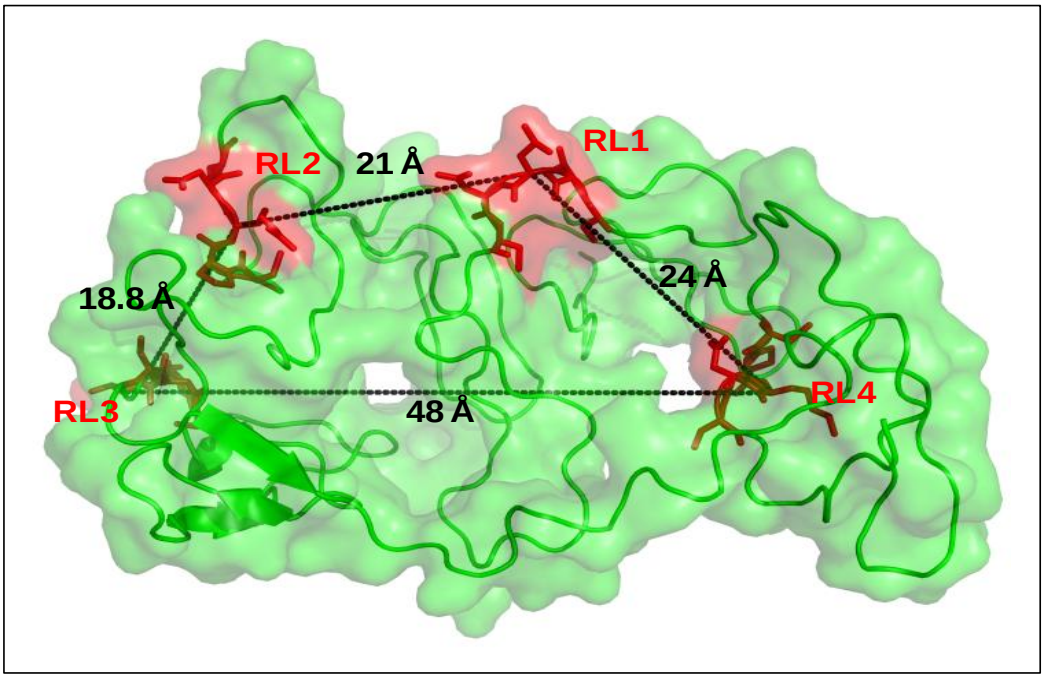


Figure 4.7: *In silico* structure prediction of CanPI-7. The predicted structure from I-TASSER server displays the four reactive sites (RL1, RL2, RL3, RL4) on the exposed loops and major part of the structure is unordered. The distances between the active site loops are marked.

4.3.5. Probable mechanism of CanPI-7 binding with multiple target protease molecules

In order to understand the inhibitory mechanism of CanPI-7 against target proteases, primary docking simulations were performed with GRAMMX webserver. The goal of the initial stage of docking was to generate as many near-native complex structures (hits) as possible. Four complexes (Complex1 to Complex4) (**Fig. 4.9**) were obtained from multiple docking analysis with two trypsin and two chymotrypsin molecules, respectively. Complex1 showed the close proximity of trypsin molecule with the first reactive loop (CTLNC) with distance of 4.6 Å between Cα atom of central residue of the reactive loop and Ser195 of trypsin. Binding of the first trypsin

molecule at RL1, caused the steric hindrance for binding of the next trypsin molecule at second reactive loop (RL2) which was only 21Å away from RL1. However, further

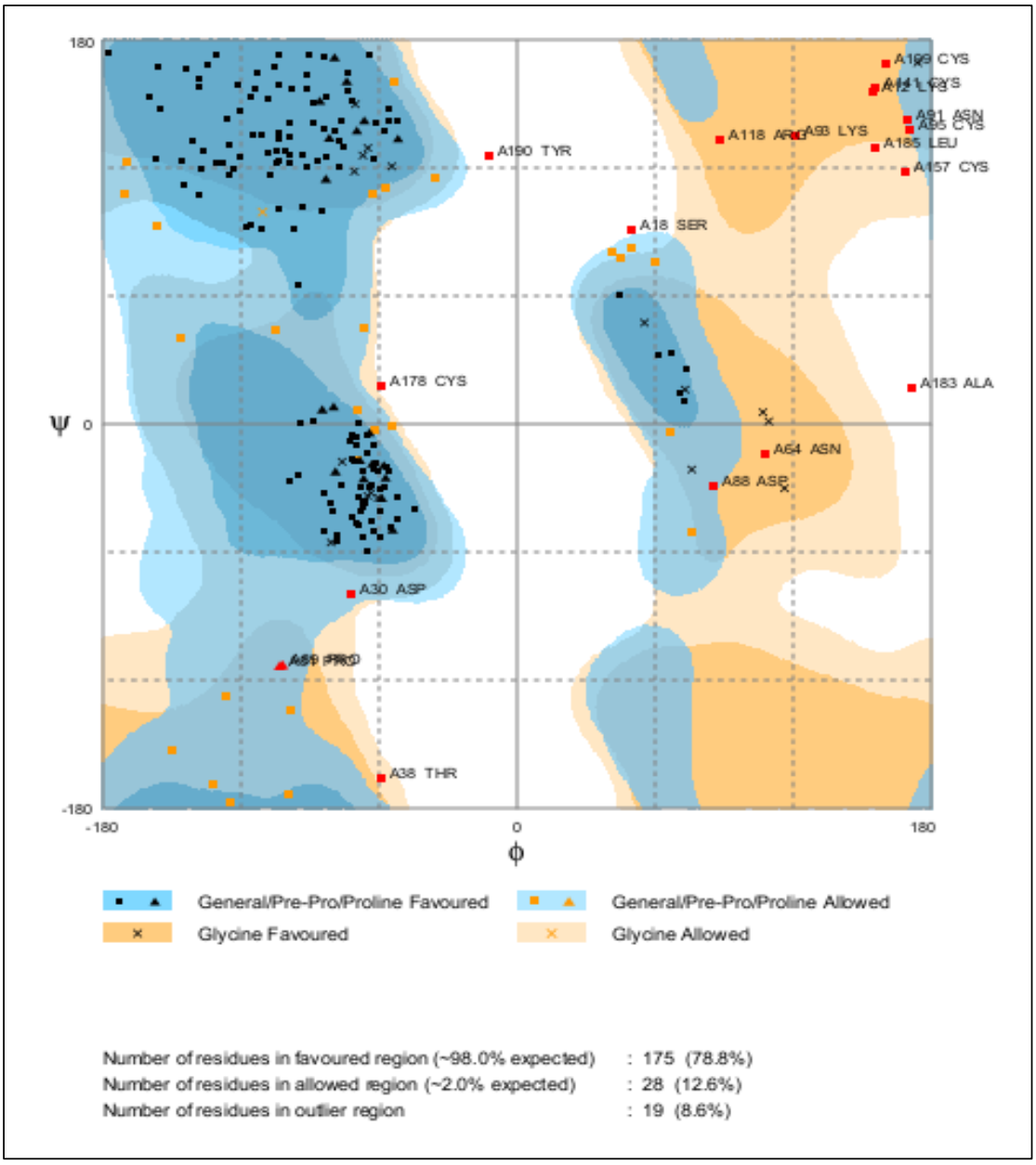


Figure 4.8: Ramchandran plot analysis for CanPI-7 predicted structure.

two chymotrypsin molecules showed close proximity with the third (RL3) and the fourth (RL4) reactive loops with intermolecular distances of 8.1 and 10.4 Å, respectively. The static complex of CanPI-7 representing at least three protease molecules in close proximity to the RLs, supports its higher potency corroborating with the enzymatic assays. Earlier studies have suggested inter-domain flexibility in the unbound form of two-domain Tomato-inhibitor-II as compared to the bound form

of the inhibitor, mediated mostly by slight changes in the interdomain surface (Barette Ng et al., 2003; Schirra et al., 2005). Thus, it is possible that both, the bound and the unbound forms of CanPI-7 could exhibit conformational flexibility within the

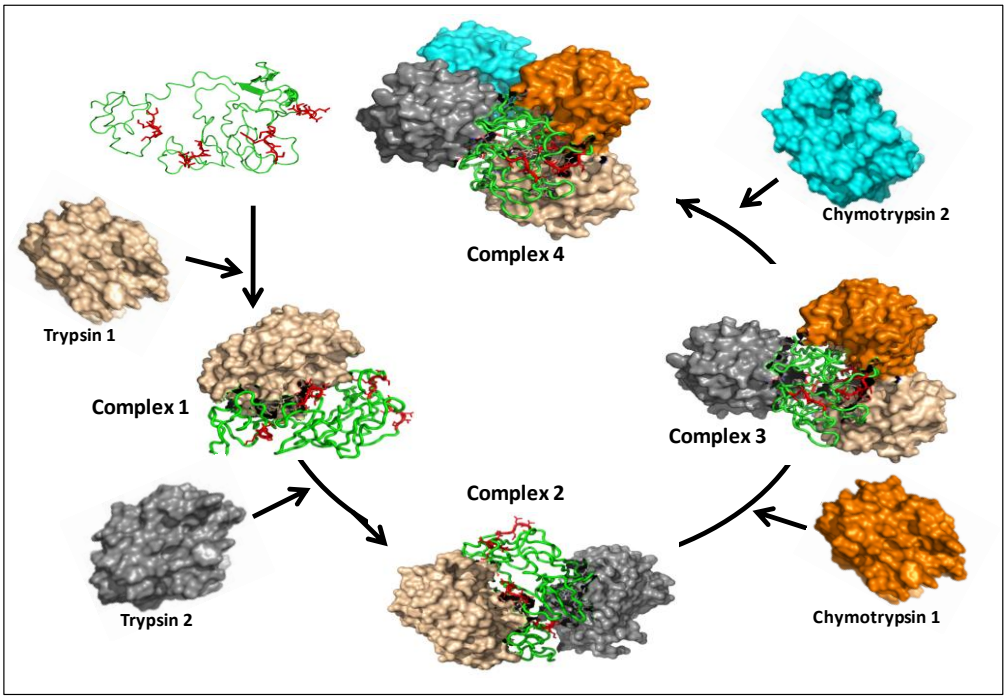


Figure 4.9: Docking of CanPI-7 with two molecules of trypsin and two molecules of chymotrypsin. Close proximity to RSL 1, 3 and 4 was observed with protease molecules while binding to RSL2 appeared obstructed because of steric hindrance.

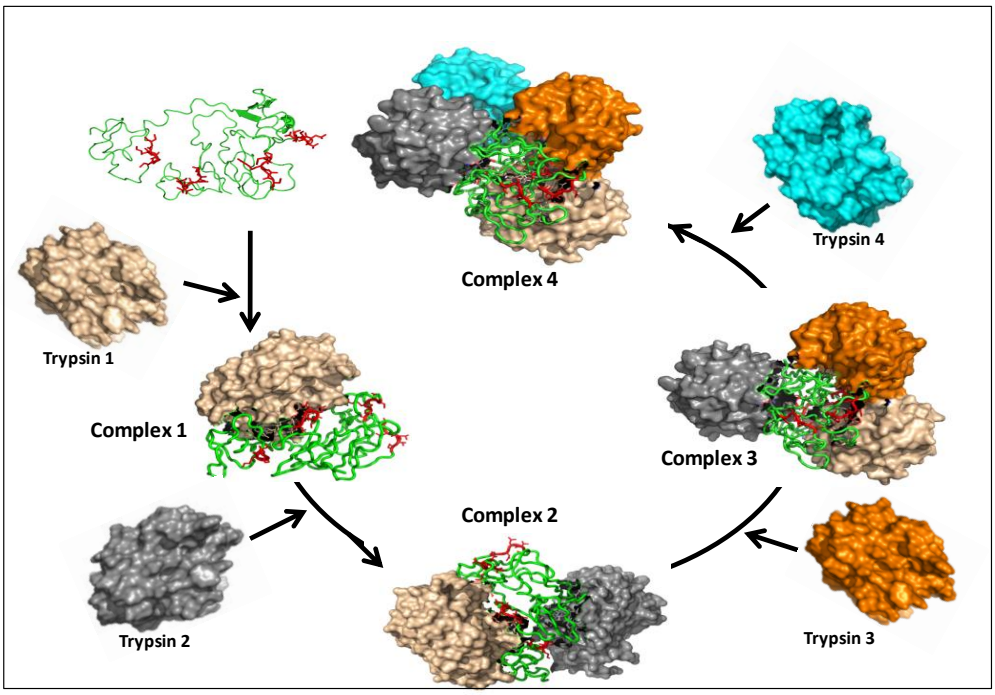


Figure 4.10: Docking of CanPI-7 with four molecules of trypsin. Cross reactivity in binding to either TI or CI site was evident.

domains relative to each other or in reactive site loops leading to either strong interaction with more number of protease molecules or less than those observed in the static model of the complex. Molecular dynamic simulations of each of the complexes or crystal structures would deliver the dynamic picture of this interaction. Cross reactivity in binding of trypsin or chymotrypsin molecules to either TI or CI sites was also observed (**Fig. 4.9 and Fig. 4.10**) which suggested that the reactive site loops retained adequate conformational flexibility to allow recognition by a variety of proteinase molecules. PIs being an innate part of the plant defense system, for protecting them from insects, fungi and bacteria, must have enough variability to interact with a wide range of proteases they come across.

4.4. Conclusions

Based on the present studies, single and multi domain PIs vary in their biochemical properties like inhibition potential and binding efficiencies against target proteases whereas show similar biophysical properties such as high stability and similar biophysical patterns under various conditions. The structural studies put forth the probable mechanism of action for multi-domain Pin-II PIs, revealing the organization of its multiple reactive site loops in space as well as the steric hindrances imposed while binding to proteases due to their close proximity.

Chapter 5

General discussion and future directions

CHAPTER 5

General discussion and future directions

Plant-insect interaction is a dynamic system which is a subject to continual variation and change. They have co-evolved from past millions of years and developed distinctive strategies to overcome each other's defenses or resistance mechanisms. The study of this interaction has become very interesting owing to the new discoveries of plant defense mechanisms and also the pest adaptations or aggressions. Plants have evolved with, physical barriers, secondary metabolites and defensive proteins to resist insect herbivores. On the other hand, insects have also developed several strategies to overcome plant defenses making themselves capable of feeding, growing and reproducing on their host plants. Among numerous defenses, causing 'indigestion' in insects remains the plants' best defense for resisting insects through an array of proteinase inhibitors as defensive proteins thereby depriving insects of nutrients and affecting their growth and development. Several PIs from host and non-host plants have been studied and characterized for their antibiosis effects on insect pests. Pros and cons of these interactions have been discussed in detail in the first chapter.

PIs from *Capsicum annuum* (Solanaceae) showed promising control on *Helicoverpa armigera* (Lepidoptera). Diverse PI genes (*CanPIs*) were identified from various *C. annuum* tissues and the PI proteins showed inhibition of gut proteinases of *H. armigera*. Feeding bioassays with PI proteins had shown not only reduction in larval and pupal weights but also dramatic reduction in fertility and fecundity in a dose dependent manner. Moreover, significantly high expression levels of *CanPIs* under naturally infested conditions had indicated the involvement of *CanPIs* in plant defense. The several questions raised out of these previous studies, as detailed in "Genesis of thesis" were tried to address in the present work as follows: Identification of differentially expressed *CanPIs* and their functional correlation; recombinant expression and biochemical characterization of diverse CanPIs; interaction studies of rCanPIs with HGPs for their stability and *in vitro* and *in vivo* inhibition efficiency; and detailed insights in to structure of representative CanPIs. The present study on *C. annuum* PIs and their interaction with insect proteases have shed light into some

interesting aspects of *C. annuum* PIs and has further raised questions on their diversity and regulation mechanisms.

5.1 Induced PI diversity in *C. annuum*

To explore the functional specialization of an array of *Capsicum annuum* (L.) PI genes (CanPIs), their expression, processing and tissue-specific distribution were studied under steady-state and induced conditions. Various treatments namely, aphid infestation or mechanical wounding followed by treatment with either oral secretion (OS) of *H. armigera* or water resulted in regulated accumulation of diverse *CanPI* genes. The quest to identify differentially expressed *CanPIs* under inductions, yielded 44 novel *CanPIs* comprising eight 2-IRD, thirty-one 3-IRD, and five 4-IRD isoforms. Tissue and induction specific patterns were observed for *CanPI* expression and accumulated PI activity. Although the PI expression was upregulated under all inductions, specifically 4-IRD isoforms showed significantly high expression levels in wounded leaves treated with OS, unlike the uninduced leaves or other inductions. Consistent with the transcript abundance, significantly higher PI activity and isoform diversity contributed by 3- and 4-IRD *CanPIs* was observed for wound plus OS or water treated *C. annuum* leaves. This not only indicated the involvement of *CanPIs* in plant defense but also specifically emphasized on the importance of 4-IRD *CanPIs* in insect resistance. Simultaneous presence of 1, 2, 3 and 4-IRD *CanPIs* and variability in expression across the various *C. annuum* tissues screened raises a question about the significance of such high diversity particularly found only in *C. annuum* within Solanaceae. The expression variability in *CanPIs* and their correlation with the tissue type or induction treatment provides an excellent example of temporal, spatial, qualitative and quantitative gene regulatory mechanism(s) operating in *C. annuum* plants.

Pin-II genes in Solanaceae are important due to their evolution into multiple inhibitory repeat types from the ancestral single repeat Pin-II PI precursor (Barta et al., 2002). Potato expresses two repeat Pin-II PIs, tomato possess both two and three repeat Pin-II PIs, while various *Nicotiana* sp. express Pin-II PIs with two to eight repeat domains (Kim et al., 2001; Xu et al., 2001; Wu et al., 2006). *C. annuum* shows presence of 1 to 4 repeat Pin-II PIs, with significantly high diversity in precursor Pin-II PIs resulting in to several subtypes of each type of the precursors. Other species of

Solanaceae have not shown such a high diversity in the precursor subtypes except for few representatives like SaPin-II a and b from *S. americanum*, which show more than 20% sequence divergence though both possess two repeat Pin-II precursors (Xu et al., 2001). Various combinations of IRDs are present in the *CanPI* precursor molecules from which several unique IRDs were identified in the present study. The diversity in unique IRDs from *C. annuum* that are expressed simultaneously is very high and contributes to a diverse inhibitor cocktail in the plant. The amplification of all IRD type *CanPI* isoforms from signal peptide to stop codon further strengthened the presence of full length *CanPI* genes in *C. annuum* suggesting presence of multiple gene copies. Thus, the huge diversity comprising 67 *CanPI* genes expressed across varied tissues and conditions was a striking observation. Confirmation of copy number of these *CanPIs* by Southern analysis and the transcriptional confirmation by Northern analysis would serve to validate this hypothesis. Pin-II PI gene characterization from genomic DNA of *C. annuum* with respect to introns would endow with the role of post-transcriptional modifications in CanPI diversity, if any. This extraordinarily high gene diversity in *CanPIs* has raised many questions; to list a few; (i) why this high diversity is found only in *C. annuum* among the Solanaceae plants studied so far? and if so, (iii) does this huge diversity exists to serve multiple functions either, *in planta* or defense? Probable clues to these questions may lie specifically in the evolutionary divergence of *C. annuum* as compared to other other members of Solanaceae.

In the present study, all the CanPI isoforms were not ubiquitously found in steady state or under all induction treatments, rather their expression was regulated by inducers like aphid, wounding or OS from *H. armigera*. The qualitative and quantitative variations in the *CanPI* expression and PI protein accumulation patterns suggested that the regulatory mechanisms in *C. annuum* governed the PI expression for induction of specific IRD types of CanPIs *vis-a-vis* a stress. The *CanPI* gene pool either from uninduced or induced leaves was dominated by 3-IRD PIs and trypsin inhibitory domains whereas OS application to wounded leaves specifically upregulated expression of 4-IRD type PIs. Interestingly, 1- and 2-IRD CanPI precursor subtypes were insignificant under induced conditions in leaves. This indicated that, biotic stress to *C. annuum* upregulated induction of CanPIs of higher IRD type, which were otherwise not expressed at steady state levels. Moreover, differential post-translational processing of multi-IRD PIs at the linkers resulted in to

varied processed products (1-IRD isoforms) under different induction treatments which further suggested amplification of the induced diversity. Thus, defense role of induced higher IRD type PIs upon herbivory while tissue-specific endogenous physiological functions of CanPIs expressed under steady state were revealed in our studies. It is quite likely that further diverse forms of CanPI precursors might be observed in different tissues under different physiological conditions. However, the important question which arises is, what is the mechanism by which *C. annuum* brings about the co-ordinated regulation of CanPI expression under steady state and induced conditions as well as across various tissues? 3'UTR regions possessing regulatory sequences and other transcriptional factors might play important role in regulating the spatial and temporal CanPI expression.

5.2 Pin-II protein diversity and functional significance of sequence variation in PIs

Recombinant protein expression and biochemical characterization of diverse CanPIs, led to discovery of further complexity in the protein products of these genes. Multiple, active inhibitory fragments were obtained from the individual, one, two, three and four-IRD precursors. The type of PI protein fragmentation obtained in the present study as detailed in Chapter 3, suggested action of proteases at the linker regions and also trimming at the terminal ends of IRDs. This type of processing is probably very similar to the one that occurs *in planta*, by action of plant proteinases on the linker regions (Heath et al., 1995, Horn et al., 2005), generating IRDs with ragged ends which, in turn, leads to variability for defense functions (Horn et al., 2005).

From the previous reports, a total of 77 different IRDs were identified contributing to the diverse inhibitory pool (Barta et al., 2002). In the present study (Chapter 3), individual *CanPI* genes were chosen for functional characterization based on: diversity in number of IRDs, TI or CI specificity of the IRD and aa variation in the IRD sequence. Thus, the six *CanPI* genes comprised nine unique IRDs with 2 to 26% sequence divergence, with major aa substitutions in reactive site loop. Individual CanPIs displayed variation in their inhibitory potential against proteases like, trypsin, chymotrypsin or HGP. *CanPIs* with higher number of IRDs and having diverse specificities, for example, combination of both TI and CI showed higher inhibition of the above mentioned proteases. Not only quantitative but qualitative variations were

also observed with respect to the interaction of diverse PIs with proteases. HGP is a complex mixture of several enzymes and proteases; differential inhibition of the total 11 HGP protease isoforms by various rCanPIs suggested the role of minor sequence variations affecting their interaction with the respective proteases. Specific inhibition of HGP isoforms by CanPIs bearing common IRDs and the functional significance of even single aa changes in the IRDs were the remarkable observation. An effect of number of IRDs and sequence variations in constituent IRDs of CanPIs was also visible in their stability and interaction with diverse proteases within HGP. Moreover, the insensitivity of few HGP isoforms to the CanPIs under study, recommended the need for a more diverse pool of PIs in order to tackle the entire protease diversity in the insect gut. Thus, the diversity in CanPIs with respect to the number of IRDs per gene, their orientation in a PI and sequence variations within IRDs point towards their functional significance and a more complex interaction of CanPIs with endogenous and/or insect gut proteases.

5.3 Diverse CanPIs: real challenge to explore protease-PI interaction

Diverse CanPIs with varied aa sequences and functional specificities suggest their structural variation. The three-dimensional structure of a protein defines not only its size and shape, but also its function (Hames and Hooper, 2000). The folding of a protein allows for interactions between amino acids that may be distant from each other in the primary sequence of the protein. Therefore, any alteration in amino acids whether, close or distant to active site and the binding loop, can lead to changes in folding and eventually to changes in chemical interactions among amino acids at the active site, which might alter the enzyme activity or binding of the ligands to receptor proteins.

In a Pin-II PI protein, the eight fully conserved cysteines work as structural scaffold to hold the reactive site loop (RSL) in a relatively rigid conformation that helps to prevent proteolytic cleavage of the inhibitor upon interaction with proteases (Barrette-Ng et al., 2003a). The strength of the protease-PI interaction is determined by the compatibility of all aa residues (P4-P4') within the RSL. It was striking to note that the major aa variations in the unique IRD sequences lied mostly within reactive site loops or towards the C-terminal ends of the inhibitory domain. These variations thus, suggest a subsequent/corresponding change in the individual inhibitor structure

which can be the most promising factor affecting its binding or activity against a target protease. It is interesting that the residues outside this loop, referred to as adventitious contacts, can also significantly affect the affinities of the inhibitor for closely related target proteases (Komiyama et al., 2003). Thus, the *C. annuum* PIs represent an array of diverse inhibitor molecules with identical framework and structural scaffold, which have minor aa substitutions, in turn leading to varying binding efficiencies and specificities to proteases. This proposes a molecular co-evolution of CanPIs in response to pest attack as such a mechanism would eventually facilitate the targeting of diverse or wide range of proteases. Therefore, the structural studies on CanPIs and their interaction with proteases would be interesting in order to realize/explore the molecular details of this fascinating interaction which maintains specificity and also the enough flexibility at the same time. Furthermore, the absence of N- and C-terminal partial repeats in *C. annuum* PI precursors puts forth anticipation towards a significant structural variation with respect to orientation of domains in the CanPI precursor proteins. Although several structures of single domain PIs are available, multidomain PIs remain unexplored and need further exploration. The multidomain PI structures, can throw light on the domain orientations in space, interdomain interactions and thus, the resulting effect on their binding with proteases.

5.4 CanPI cocktail to control *H. armigera*: an ideal, sustainable and flexible blend

In the present study, *in vitro* experiments to evaluate the stability of diverse CanPIs in presence of insect gut proteases revealed their remarkable potential against *H. armigera*. Degradation of some CanPIs and proteolytic processing of the precursor PI proteins was observed in presence of HGPs. Proteolytic processing of multi-IRD CanPIs at the linker regions suggested the probable mechanism for processing of precursors in the insect midgut thereby, releasing the inhibitory domains separately in order to achieve maximum inhibition of proteases. The processed IRDs, equivalent to single IRD forms further showed stability in presence of HGP and indicated that CanPIs are non-degradable by insect proteases. The high efficiency and stability of multi-IRD CanPIs can be well correlated to the induced up regulation of higher IRD PI transcripts upon wounding and insect infestation thus, signifying the role of multidomain PIs *in planta* for defense. *In vivo* studies to determine the fate of CanPIs in insect precisely displayed active CanPIs in the *H. armigera* midgut. These observations are similar to the reports suggesting that PIs from non-host plants are

resistant to degradation; can effectively inhibit gut proteases and thus, the larval growth.

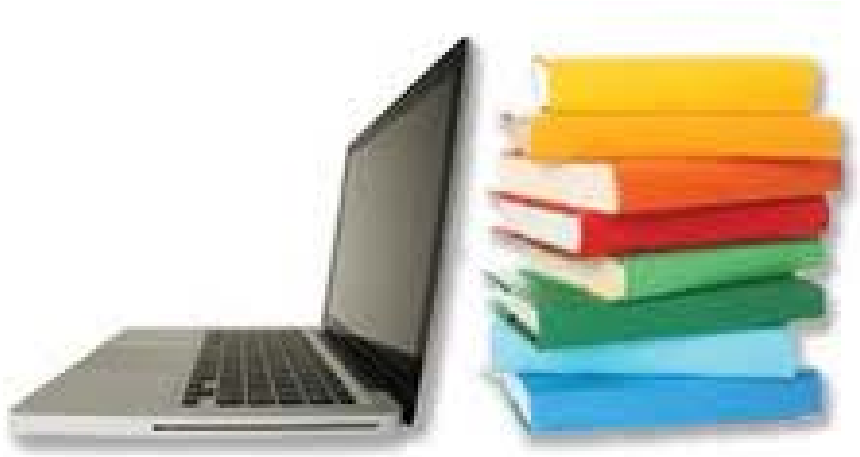
The effects of PIs on insects are very much dependent on the amount of PI ingested by the insect. To exert maximum effect on the insect, the PI amounts must be higher than the threshold values (Johnston et al., 1993; Wu et al., 1997). Thus, an optimization of the effective dosage of inhibitor to be used for insect control is also significant before implementation of PI based approach. Before hand, the molecular dissection of insect responses to CanPI ingestion would be important in order to anticipate the adaptation mechanisms of *H. armigera*. Insect pests are known to adapt to PIs by altering the expression levels of their gut proteases or by *de novo* synthesis of proteases which are capable of degrading the PIs. Identification of induced insensitive proteases, if any, in response to CanPI ingestion can help in effective manipulation of the present pool of inhibitor molecules using modern genetic engineering tools. Synthetic constructs could be designed by selecting and combining novel IRDs and increasing the precursor PI size by fusing selected IRDs to achieve maximum plant protection against insects.

The present study on *C. annuum* highlights the plant's machinery to regulate the expression of PIs as per the requirement of both, endogenous and defense, functions thus exerting minimum cost on plant productivity. Future research in elucidating these mechanisms and their application in PI transgenics would be advantageous to make use of tissue specific/induction via specific promoters for PI expression. The choice of appropriate PIs for insect defense is certainly a long lasting and sustainable approach of plant defense, if proper expression and effect are to be ensured. Due to a high diversity in CanPI IRDs it is possible to identify combinations of IRDs, which are best suited for a particular insect infestation.

Thus, in conclusion, the naturally occurring gene diversity in CanPIs provides a very effective starting material to select the best insect defensive combinations as provided by the nature which could be implemented to reach the goal of crop protection through productive, sustainable and environment friendly insect resistance strategy. As per the discussion stated above, pursuing of the following leads in future would be worth in order to effectively implement CanPIs in plant protection: (i) Identification of regulatory sequences in 3'UTR or transcription factors governing the PI expression; (ii) Detailed molecular analysis of the *H. armigera* responses to CanPI ingestion; (iii)

Identification of protease genes playing significant role in insect adaptation to CanPIs, for example insensitive proteases, if any; (iv) X-ray crystallography of CanPI proteins and/or in complex with standard enzymes like trypsin or *H. armigera* gut proteases; (v) Structural and functional characterization of natural variant IRDs; and (vi) Site-directed mutagenesis studies for selected residues within IRDs in order to understand the basis of Pin-II PI protein evolution.

Bibliography



BIBLIOGRAPHY

Abdeen A, Virgos A, Olivella E, Villanueva J, Avile´s X, Gabarra R, Prat S (2005) Multiple insect resistance in transgenic tomato plants over-expressing two families of plant proteinase inhibitors. Plant Molecular Biology 57: 189–202.

Antcheva N, Pintar A, Patthy AS, Simoncsits A et al. (2001) Proteins of circularly permuted sequence present within the same organism: the major serine proteinase inhibitor from *Capsicum annuum* seeds. Protein Science 10: 2280-2290.

Atkinson AH, Heath RL, Simpson RJ, Clarke AE, Anderson MA (1993) Proteinase inhibitors in *Nicotiana alata* stigmas are derived from a precursor protein which is processed into five homologous inhibitors. Plant Cell 5: 203-213.

Babu RM, Sajeena A, Seetharaman K, Reddy MS (2003) Advances in genetically engineered (transgenic) plants in pest management–an overview. Crop Protection 22: 1071-1086.

Baldwin IT, Halitschke R, Kessler A, Schittko U (2001) Merging molecular and ecological approaches in plant–insect interactions. Current Opinion in Plant Biology 4: 351–358.

Barrett AJ and Rawlings ND (1995) Families and clans of serine peptidases. Archives of Biochemistry and Biophysics 318: 247-250.

Barrette-Ng IH, Ng KK, Cherney MM, Pearce G, Ghani U, Ryan CA, James MN (2003b) Unbound form of tomato inhibitor-II reveals interdomain flexibility and conformational variability in the reactive site loops. Journal of Biological Chemistry 278: 31391-31400.

Barrette-Ng IH, Ng KK, Cherney MM, Pearce G, Ryan CA, James MN (2003a) Structural basis of inhibition revealed by a 1:2 complex of the two-headed tomato inhibitor-II and subtilisin Carlsberg. Journal of Biological Chemistry 278: 24062-24071.

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

Barton KA, Whiteley HR, Yang NS (1987) *Bacillus thuringiensis* delta-endotoxin expressed in transgenic *Nicotiana tabacum* provides resistance to lepidopteran insects. Plant Physiology 85: 1103-1109.

Bateman KS, James MNG (2011) Plant protein proteinase inhibitors: Structure and mechanism of inhibition. Current Protein and Peptide Science 12: 341-347.

Bayes A, Sonnenschein A, Daura X, Vendrell J, Aviles FX (2003) Procarboxypeptidase A from the insect pest *Helicoverpa armigera* and its derived enzyme. European Journal of Biochemistry 270: 3026–3035.

Beekwilder J, Schipper B, Bakker P, Bosch D, Jongsma M (2000) Characterization of potato proteinase inhibitor II reactive site mutants. European Journal of Biochemistry 267: 1975-1984.

Bergey DR, Orozco-Cardenas M, de Moura DS, Ryan CA (1999) A wound- and systemin- inducible polygalacturonase in tomato leaves. Proceedings of National Academy of Sciences, USA 96:1756–1760.

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 Physiology 152: 2232-2242.

Bochicchio B and Tamburro AM (2002) Polyproline II structure in proteins: Identification by Chiroptical spectroscopies, stability and functions. Chirality 14: 782-792.

Bode W and Huber R (1992) Natural protein proteinase inhibitors and their interaction with proteinases. European Journal of Biochemistry 204: 433-451.

Boden M and Hawkins J (2005) Prediction of subcellular localisation using sequence-biased recurrent networks. Bioinformatics 21: 2279-2286.

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 Biochemistry and Molecular Biology 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 Biochemistry and Molecular Biology 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. Physiological Entomology 29: 278-290.

Broadway RM (1995) Are insects resistant to plant proteinase inhibitors. Journal of Insect Physiology 41: 1107-116.

Broadway RM (1996) Plant dietary proteinase inhibitors alter complement of midgut proteases. Archives of Insect Biochemistry and Physiology 32: 39–53.

Broadway RM (1997) Dietary regulation of serine proteinases that are resistant to serine proteinase inhibitors. Journal of Insect Physiology 43: 855-874.

Broadway RM and Duffey SS (1986a) The effect of dietary protein on the growth and digestive physiology of larval *Heliothis zea* and *Spodoptera exigua*. Journal of Insect Physiology 32: 673-680.

Broadway RM and Duffey SS (1986b) Plant proteinase inhibitors: mechanism of action and effect on the growth and digestive physiology of larval *Heliothis zea* and *Spodoptera exigua*. Journal of Insect Physiology 32: 827–833.

Bryant J, Green TR, Gurusaddaiah T, Ryan CA (1976) Proteinase inhibitor II from potatoes: isolation and characterization of its promoter components. Biochemistry 15: 3418–3424.

Carlini CR and Grossi-de-Sa MF (2002) Plant toxic proteins with insecticidal properties. A review on their potentialities as bioinsecticides. Toxicon 40: 1515-1539.

Charity JA, Anderson MA, Bittisnich DJ, Whitecross M, Higgins TJV (1999) Transgenic tobacco and peas expressing a proteinase inhibitor from *Nicotiana alata* have increased insect resistance. Molecular Breeding 5: 357–365.

Chen H, Wilkerson CG, Kuchar JA, Phinney BS, Howe GA (2005) Jasmonate-inducible plant enzymes degrade essential amino acids in the herbivore midgut. Proceedings of National Academy of Sciences, USA 102:19237-19242.

Cheng YC and Prusoff WH (1973) Relationship between the inhibition constant (K_i) and the concentration of inhibitor which causes 50 per cent inhibition (I_{50}) of an enzymatic reaction. Biochemical Pharmacology 22: 3099-3108.

Choi D, Park JA, Seo YS, Chun YJ, Kim WT (2000) Structure and stress-related expression of two cDNAs

encoding proteinase inhibitor II of *Nicotiana glutinosa* L. *Biochimica et Biophysica Acta* 1492: 211-215.

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

Chougule NP, Hivrale VK, Chhabda PJ, Giri AP, Kachole MS (2003) Differential inhibition of *Helicoverpa armigera* gut proteinases by proteinase inhibitors of pigeonpea (*Cajanus cajan*) and its wild relatives. Phytochemistry 64: 681-687.

Christeller JT (2005) Review article: Evolutionary mechanisms acting on proteinase inhibitor variability. FEBS Journal 272: 5710-5722.

Chye ML, Sin SF, Xu ZF, Yeung EC (2006) Serine proteinase inhibitor proteins: Exogenous and endogenous functions. *In Vitro* Cellular and Developmental Biology - Plant 42: 100-108.

Dahiya SS, Chauhan YS, Johansen C, Shanower TG (1999) Adjusting pigeonpea sowing time to manage pod borer infestation. International Chickpea and Pigeonpea Newsletter 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.

De Leo F, Bonade-Bottino M, Ceci LR, Gallerani R, Jouanin L (2001) Effects of a mustard trypsin inhibitor expressed in different plants on three lepidopteran pests. Insect Biochemistry and Molecular Biology 31: 593-602.

Fass D (2012) Disulfide Bonding in Protein Biophysics. Annual Review of Biophysics 41: 43–59.

Diezel C, Kessler D, Baldwin IT: Pithy protection (2011) *Nicotiana attenuata*'s jasmonic acid-mediated defenses are required to resist stem-boring weevil larvae. Plant Physiology155: 1936-1946.

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. Nature Biotechnology 14: 494-498.

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

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. Proceedings of the National Academy of Sciences, USA 107: 15016-15021.

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. Proceedings of the National Academy of Sciences, USA 107: 15011-15015.

Eddy JL, Derr JE, Hass GM (1980) Chymotrypsin inhibitor from potatoes: interaction with target enzymes. Phytochemistry 19: 757-761.

Eschenfeldt WH, Stols L, Millard CS, Joachimiak A, Donnelly MI (2009) A Family of LIC Vectors for High-Throughput Cloning and Purification of Proteins. Methods Molecular Biology 498: 105–115.

Farmer EE and Ryan CA (1992) Octadecanoid precursors of jasmonic acid activate the synthesis of wound-inducible proteinase inhibitors. Plant Cell 4:129–134.

Fefelova Y and Frolov A (2008) Distribution and mortality of corn earworm (*Helicoverpa armigera*, Lepidoptera, Noctuidae) on maize plants in Krasnodar territory. Entomological Review 88(4): 480-484.

Felton GW, Korth KL, Wesley SV, Huhman DV, Mathews MC, Murphy JB, Lamb C, Dixon RA (1999) Inverse relationship between systemic resistance of plants to microorganisms and to insect herbivory. Current Biology 9: 317–320.

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

Fidantsef AL, Stout MJ, Thaler JS, Duffey SS, Bostock RM (1999) Signal interactions in pathogen and insect attack: expression of lipoxygenase, proteinase inhibitor II, and pathogenesis-related protein P4 in the tomato, *Lycopersicon esculentum*. Physiological and Molecular Plant Pathology 54: 97-114.

Fitt GP (1989) The ecology of *Heliothis* in relation to agroecosystems. Annual Review of Entomology 34: 17-52.

Freeman BC and Beattie GA (2008) An overview of plant defenses against pathogens and herbivores. The Plant Health Instructor. DOI: 10.1094/PHI-I-2008-0226-01.

Frutos R, Rang C, Royer M (1999) Managing insect resistance to plants producing *Bacillus thuringiensis* toxins. Critical Reviews in Biotechnology 19: 227–276.

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

Gatehouse AMR, Norton E, Davison GM, Babbe´ SM, Newell CA, Gatehouse JA (1999) Digestive proteolytic activity in larvae of tomato moth, *Lacanobia oleracea*, effects of plant protease inhibitors *in vitro* and *in vivo.* Journal of Insect Physiology 45: 545-558.

Gatehouse JA (2002) Plant resistance towards insect herbivores: a dynamic interaction. New Phytologist 156:145–169.

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 Biochemistry and Molecular Biology 27: 929-944.

Giri AP, Chougule NP, Telang MA, Gupta VS (2005) Engineering insect tolerant plants using plant defensive proteinase inhibitors. In: Recent research developments in phytochemistry, (Pandalai, S. G., Ed) Research Signpost, India, 8: 117-137.

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 Physiology 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, and 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 Physiology 142: 1621–1641.

Gowda CLL and Sharma HC (2005) Legume pod borer/cotton bollworm, *Heliothis*/*Helicoverpa*- the global problem. In souvenir national symposium on *Helicoverpa* management, National challenge, Kanpur, India.

Graham JS, Pearce G, Merryweather J, Titani K, Ericsson LH, Ryan CA (1985) Wound-induced proteinase inhibitors from tomato leaves. II. The cDNA-deduced primary structure of pre-inhibitor II. Journal of

Biological Chemistry 260: 6561-6564.

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

Greenblatt HM, Ryan CA, James MNG (1989) Structure of the complex of *Streptomyces griseus* proteinase B and polypeptide chymotrypsin inhibitor-1 from Russet Burbank potato tubers at 2.1 Å resolution. Journal of Molecular Biology 205: 201-228.

Gunning RV, Moores GD, Devonshire AL (1998) Insensitive acetycholinesterase causes resistance to organophosphates in Australian *Helicoverpa armigera* (Hübner) (Lepidoptera: Noctuidae). Pesticide Science 54: 319-320.

Gupta GP, Mahapatro GK, Chandra A (2000) Bio-potency of insecticidal crystal proteins of *Bacillus thuringiensis* against cotton (*Gossypium hirsutum*) bollworms. Indian Journal of Agricultural Sciences 70: 194-196.

Gurjar G, Mishra M, Kotkar H, Upasani M, Soni P, Tamhane V, Kadoo N, Giri A and Gupta V (2011) 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.

Halitschke R, Schittko U, Pohnert G, Boland W, Baldwin IT (2001) Molecular interactions between the specialist herbivore *Manduca sexta* (Lepidoptera, Sphingidae) and its natural host *Nicotiana attenuata* III. Fatty acid-amino acid conjugates in herbivore oral secretions are necessary and sufficient for herbivore -specific plant responses. Plant Physiology 125: 711-717.

Haq SK, Atif SM, Khan RH (2004) Protein proteinase inhibitor genes in combat against insects, pests and pathogens: natural and engineered phytoprotection. Archives of Biochemistry and Biophysics 431: 145-159.

Hames BD and Hooper NM (2000) Instant notes in Biochemistry. Bios Scientific Publishers Ltd.; 2nd revised edition.

Harshini S, Nachman RJ, Sreekumar S (2002a) Inhibition of digestive enzyme release by neuropeptides in larvae of *Opisina arenosella* (Lepidoptera: Cryptophasidae). Comparative Biochemistry and Physiology Part B 132: 353-358.

Harshini S, Nachman RJ, Sreekumar S (2002b) *In vitro* release of digestive enzymes by FMRF amide related neuropeptides and analogues in the lepidopteran insect *Opisina arenosella* (Walk.). Peptides 23: 1759-1763.

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 Physiology 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 Signalling and Behaviour 6: 1-3.

Haruta M, Major IT, Christopher ME, Patton JJ, Constabel CP **(**2001) A Kunitz trypsin inhibitor gene family from trembling aspen (*Populus tremuloides* Michx.): cloning, functional expression, and induction by wounding and herbivory. Plant Molecular Biology 46: 347–359.

Hause B, Hause G, Kutter C,Miersch O,Wasternack C (2003) Enzymes of jasmonate biosynthesis occur in tomato sieve elements. Plant Cell Physiology 44:643-648.

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*. European Journal of Biochemistry 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. Journal of Insect Physiology 43: 833-842.

Hegyi H and Gerstein M (2002) The relationship between protein structure and function: a comprehensive survey with application to the yeast genome. Journal of Molecular Biology 288: 147-164.

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 Physiology 125: 683–700.

Hilder VA and Boulter D (1999) Genetic engineering of crop plants for insect resistance - a critical review. Crop Protection 18: 177-191.

Hilder VA, Gatehouse AMR, Boulter D (1993) Transgenic plants conferring insect tolerance. Proteinase inhibitor approach. In: Transgenic plants vol. 1. Engineering and utilization. (Eds. SD Kung and R Wu). Academic Press, Inc.California, 317-338.

Hilder VA, Gatehouse AMR, Sherman SE, Barker RF, Boulter D (1987) A novel mechanism for insect resistance. Nature 330: 160–163.

Horn M, Patankar AG, Zavala JA, Wu J, Doleckova-Maresova L, Vujtechova M, Mares M, Baldwin IT (2005) Differential elicitation of two processing proteases controls the processing pattern of the trypsin proteinase inhibitor precursor in *Nicotiana attenuata*. Plant Physiology 139: 375-388.

Howe GA (2004) Jasmonates as signals in the wound response. Journal of Plant Growth Regulation 23:223-237.

Howe GA and Schaller A (2008) Direct defences in plants and their induction by wounding and insect herbivores. Induced Plant Resistance to Herbivory, Springer Science+Business Media B.V. pp 7-29.

Howe GA and Ryan CA (1999) Suppressors of systemin signaling identify genes in the tomato wound response pathway. Genetics 153:1411–1421.

http:// www.ebi.ac.uk

http://www.livemint.com/2009/02/25181538/India-loses-Rs90000-cr-crop-y.html

http://papilio.ab.a.u-tokyo.ac.jp/lep-genome/index.html

http://www.rcsb.org/pdb

http://www.sciencegateway.org/tools/proteinmw.htm

http://zhang-lab.ccmb.med.umich.edu/I-TASSER

Huesgen PF and Overall CM (2011) N- and C-terminal degradomics: new approaches to reveal biological roles for plant proteases from substrate identification. Physiologia Plantarum, 145(1): 5-17.

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.

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. Proceedings of National Academy of Sciences, USA 86: 9871-9875.

Johnston KA, Gatehouse JA, Anstee JH (1993) Effect of soybean protease inhibitors on the growth and development of larval *Helicoverpa armigera*. Journal of Insect Physiology 39: 657-664.

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 Biochemistry 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. Proceedings of the National Academy of Sciences, USA 92: 8041-8045.

Jongsma MA and Bolter C (1997) The adaptation of insects to plant protease inhibitors. Journal of Insect Physiology 43: 885-895.

Karban R and Baldwin IT: Induced responses to herbivory. Chicago: University of Chicago Press 1997, 1–319.

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 Science 161: 27-33.

Komiyama T, VanderLugt B, Fugère M, Day R et al. (2003) Optimization of protease-inhibitor interactions by randomizing adventitious contacts. Proceedings of the National Academy of Sciences, USA 100(14): 8205-10.

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

Koo AJK, Gao X, Jones AD, Howe GA (2009) A rapid wound signal activates the systemic synthesis of bioactive jasmonates in Arabidopsis. The Plant Journal 59: 974–986.

Korth K and Dixon RA (1997a) Evidence for chewing insect-specific molecular events distinct from a general wound response in leaves. Plant Physiology 115: 1299-1305.

Korth KL and Dixon RA (1997b) Differential transcript accumulation near wound sites in response to insect or mechanical damage. Plant Physiology 11: 4-26.

Kotkar HM, Sarate PJ, Tamhane VA, Gupta VS, Giri AP (2009) Responses of midgut amylases of *Helicoverpa armigera* to feeding on various host plants. Journal of Insect Physiology 55(8): 663-670.

Krowarsch D, Cierpicki T, Jelen F, Otlewski J (2003) Canonical protein inhibitors of serine proteases. Cellular and Molecular Life Sciences 60: 2427–2444.

Laskowski M and Jr, Kato I (1980) Protein inhibitors of proteinases. Annual Reviews in Biochemistry 49: 593-626.

Laskowski M. Jr, Kato I., Ardelt W., Cook J., Denton A., Empie M. W. et al. (1987) Ovomucoid third domains from 100 avian species: isolation, sequences, and hypervariability of enzyme inhibitor contact residues. Biochemistry 26: 202–221.

Lawrence PK and Koundal KR (2002) Plant proteinase inhibitors in control of phytophagous insects. Electronic Journal of Biotechnology 5: 93-109.

Lee MC, Scanlon MJ, Craik DJ, Anderson MA (1999) A novel two-chain proteinase inhibitor generated by circularization of a multidomain precursor protein. Nature Structural and Molecular Biology 6: 526-530.

Lewis WJ, van Lenteren JC, Phatak SC, Tumlinson III JH (1997) A total system approach to sustainable pest management. Proceedings of the National Academy of Sciences, USA 94: 12243-12248.

Li L, Li C, Lee GI, Howe GA (2002) Distinct roles for jasmonate synthesis and action in the systemic wound response of tomato. Proceedings of National Academy of Sciences, USA 99:6416-6421.

Li XQ, Zhang T, Donnelly D (2011) Selective loss of cysteine residues and disulphide bonds in a potato proteinase II family. PLoS ONE 6(4): e18615. doi:10.1371/journal.pone.0018615.

Lipke H, Fraenkel GS, Liener IE (1954) Effects of soybean inhibitors on growth of *Tribolium confusum*. Journal of the Science of Food and Agriculture 2:410-415.

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 and Cell Physiology 47: 1274-1284.

Lopes AR, Juliano MA, Juliano L, Terra WR (2004) Coevolution of insect trypsins and inhibitors. Archives of Insect Biochemistry and Physiology 55: 140-152.

Lucas WJ and Lee JY (2004) Plant cell biology- plasmodesmata as a supracellular control network in plants. Nature Reviews Molecular Cell Biology 5:712-726.

Luckey M, Ling R, Dose A and Malloy B (1991). Role of a Disulfide Bond in the Thermal Stability of the Lam B Protein Trimer in *Escherichia coli* Outer Membrane. Journal of Biological Chemistry 266(3): 1866-1871.

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

Maheswaran G, Pridmore L, Franz P, Anderson MA (2007) A proteinase inhibitor from *Nicotiana alata* inhibits the normal development of light-brown apple moth, *Epiphyas postvittana* in transgenic apple plants. Plant Cell Reports 26:773–782.

MaloneM (1996). Rapid, long-distance signal transmission in higher plants. Advances in botanical research. Academic Press, San Diego, pp 163–228.

Mazumdar-Leighton S, Babu CR, Bennett J (2000) Identification of novel serine protease gene transcripts in the midguts of two tropical insect pests, *Scirpophaga incertulas* (Wk.) and *Helicoverpa armigera* (Hb.). Insect Biochemistry and Molecular Biology 30: 57-68.

Mazumdar-Leighton S and Broadway RM (2001a) Transcriptional inductioron of diverse midgut trypsins in larval *Agrotis ipsilon* and *Helicoverpa zea* feeding on the soybean trypsin inhibitor. Insect Biochemistry and Molecular Biology 31: 645-657.

Mazumdar-Leighton S and Broadway RM (2001b) Identification of six chymotrypsin cDNAs from larval mid guts of *Helicoverpa zea* and *Agrtis ipsilon* feeding on the soybean (Kunitz) trypsin inhibitor. Insect Biochemistry and Molecular Biology 31: 633-644.

McCaffery AR (1998) Resistance to *heliothine* Lepidoptera: a global view. Philosophical Transactions of the Royal Society of London B 353: 1735-1750.

McGaughey WH, Gould F, Gelernter W (1998) *Bt* resistance management: a plan for reconciling the needs of the many stakeholders in *Bt*–based products. Nature Biotechnology 16:144–146.

McGurl B, Orozco-Cardenas M, Pearce G, Ryan CA (1994) Overexpression of the prosystemin gene in transgenic tomato plants generates a systemic signal that constitutively induces proteinase inhibitor synthesis. Proceedings of National Academy of Sciences, USA 91:9799-9802.

McGurl B, Pearce G, Orozco-Cardenas M, Ryan CA (1992) Structure, expression and antisense inhibition of the systemin precursor gene. Science 255:1570-1573.

McManus MT, White DWR, McGregor PG (1994) Accumulation of the chymotrypsin inhibitor in transgenic tobacco can affect the growth of insect pests. Transgenic Research 3: 50-58.

Mello MO and Silva-Filho MC (2002) Plant-insect interactions: an evolutionary arms race between two distinct defense mechanisms. Brazilian Journal of Plant Physiology 14(2): 71-81.

Michaud D (1997) Avoiding protease-mediated resistance in herbivorous pests. Trends in Biotechnology 15: 4-6.

Mickel CE and Sytandish J (1947) Susceptibility of processed soy flour and soy grits in storage to attack by *Tribolium castaneum*. University of Minnesota Agricultural Experimental station Technical Bulletin 178: 1-20.

Miller EA, Lee MC, Atkinson AH, Anderson MA (2000) Identification of a novel four-domain member of the proteinase inhibitor II family from the stigmas of *Nicotiana alata*. Plant Molecular Biology 42: 329-333.

Mithofer A and Boland W (2008) Recognition of herbivory-associated molecular patterns. Plant Physiology 146:825-31.

Mosolov VV, Loginova MD, Fedurkina NV, Benken II (1976) The biological significance of proteinase inhibitors in plants. Plant Science Letters 7(2): 77-80.

Mosolov VV and Valueva TA (2008) Proteinase Inhibitors in Plant Biotechnology: A Review. Applied Biochemistry and Microbiology 44(3): 233–240.

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

Moyen C, Hammond-Kosack KE, Jones J, Knight MR, Johannes E (1998) Systemin triggers an increase of cytoplasmic calcium in tomato mesophyll cells: Ca2+ mobilization from intra- and extracellular compartments. Plant Cell Environ 21:1101–1111.

Narv´aez-V´asquez J, Florin-Christensen J, Ryan CA (1999) Positional specificity of a phospholipase a activity induced by wounding, systemin, and oligosaccharide elicitors in tomato leaves. Plant Cell 11:2249–2260.

Navazio L, Moscatiello R, Bellincampi D, Baldan B, Meggio F, Brini M, Bowler C, Mariani P (2002) The role of calcium in oligogalacturonide-activated signalling in soybean cells. Planta 215: 596–605.

Nielsen KJ, Heath RL, Anderson MA, Craik DJ (1994) The Three-dimensional solution structure by H NMR of a 6-kDa proteinase inhibitor isolated from the stigma of *Nicotiana alata*. Journal of Molecular Biology 242: 231-243.

Nielsen KJ, Heath RL, Anderson MA, Craik DJ (1995) Structures of a series of 6-kDa trypsin inhibitors isolated from the stigma of *Nicotiana alata.* Biochemistry 34: 14304–14311.

O'Donnell PJ, Calvert C, Atzorn R, Wasternack C, Leyser HMO, Bowles DJ (1996) Ethylene as a signal mediating the wound response of tomato plants. Science 274:1914–1917.

Oganesyan N, Kim SH and Kim R (2004) On-column Chemical Refolding of Proteins. Pharmagenomics pp 22-25.

Orozco-C´ardenas M, McGurl B, Ryan CA (1993) Expression of an antisense prosystemin gene in tomato plants reduces the resistance toward *Manduca sexta* larvae. Proceedings of National Academy of Sciences, USA 90:8273-8276.

Otlewski J, Jaskolski M, Buczek O, Cicerpicki T, Czapinska H, Krowarsch D, Smalas AO, Stachowiak D, Szpineta A, Dadlez M (2001) Structure-function relationship of serine protease-protein inhibitor interaction. Acta Biochimica Polonica 48: 419-428.

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 Biochemistry and Molecular Biology 31: 453-464.

Patankar AG, Harsulkar AM, Giri AP, Gupta VS, Sainani MN, Ranjekar PK, Deshpande VV (1999) Diversity in inhibitors of trypsin and *Helicoverpa armigera* gut proteinases in chickpea (*Cicer arietinum*) and its wild relatives. Theoretical and Applied Genetics 99:719–726.

Pearce G, Bhattacharya R, Chen Y, Barona G, Yamaguchi Y, Ryan CA (2009) Isolation and characterization of Hydroxyproline-rich glycopeptides signals in black nightshade leaves. Plant Physiology 150(3): 1422-1433.

Pearce G, Johnson S, Ryan CA (1993) Purification and characterization from tobacco (*Nicotiana tabacum*) leaves of six small, wound-inducible, proteinase isoinhibitors of the potato inhibitor II family. Plant Physiology 102, 639-644.

Pearce G, Ryan CA (2003) Systemic signaling in tomato plants for defense against herbivores: isolation and characterization of three novel defense-signaling glycopeptide hormones coded in a single precursor gene. Journal of Biological Chemistry 278: 30044–30050.

Pearce G, Strydom D, Johnson S, Ryan CA (1991) A polypeptide from tomato leaves induces wound-inducible proteinase inhibitor proteins. Science 253:895–898.

Pelton JT and McLean LR (2000) Spectroscopic methods for analysis of protein secondary structure. Analytical Biochemistry 277: 167-176.

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.

Peña-Cortés H, Sanchez-Serrano J, Mertens R, Willmitzer L, Prat S (1989) Abscisic acid is involved in the wound-induced expression of the proteinase inhibitor II gene in potato and tomato. Proceedings of National Academy of Sciences, USA 86: 9851-9855.

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.

Pichare MM and Kachole MS (1994) Detection of electrophoretically separated protease inhibitors using X-ray film. Journal of Biochemical and Biophysical Methods 28: 215-224.

Pieterse CMJ, Ton J, Van Loon LC (2001) Cross-talk between plant defence signalling pathways: boost or burden? AgBiotechNet 3**:** ABN 068.

Plunkett G, Senear DF, Zuroske G, Ryan CA (1982) Proteinase inhibitors I and II from leaves of wounded tomato plants: Purification and properties. Archives of Biochemistry and Biophysics 213: 463-472.

Polgar L (2005) The catalytic triad of serine peptidases. Cellular and Molecular Life Sciences 62: 1-12.

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

Reddy KVS and Zehr UB (2004) In New directions for a diverse planet. Proceedings of the 4th International C rop Science Congress. Brisbane, Australia. pp 1-8.

Reeck GR, Kramer KJ, Baker JE, Kanost MR, Fabrick JA, Behnke CA (1997) Proteinase inhibitors and resistance of transgenic plants to insects. In: Carozzi N and Koziel M, eds. Advances in insect control: the role of transgenic plants. London, Taylor and Francis, 157-183.

Reymond P, Weber H, Damond M, Farmer EE (2000) Differential gene expression in response to mechanical wounding and insect feeding in *Arabidopsis.* Plant Cell 12: 707–719.

Richardson M (1979) The complete amino acid sequence and trypsin reactive (inhibitory) site of the major proteinase inhibitor from the fruits of aubergine (*Solanum melongena L.).* FEBS Letters 104: 322-326.

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

Roy A, Xu D, Poisson J, Zhang Y (2011) A Protocol for Computer-Based Protein Structure and Function Prediction. Journal of Visualized Experiments 57: e3259.

Royo J, Leo J, Vancanneyt G, Albar JP, Rosahl S, Ortego F, Castan EP, Sanchez Serrano J (1999) Antisense -mediated depletion of a potato lipoxygenase reduces wound induction of proteinase inhibitors and increases weight gain of insect pests. Proceedings of the National Academy of Sciences, USA 96: 1146-1151.

Ryan CA (1990) Proteinase inhibitors in plants, genes for improving defenses against insects and pathogens. Annual Reviews on Phytopathology 28: 425-449.

Ryan CA (2000) The systemin signaling pathway: differential activation of plant defensive genes. Biochimia et Biophysicia Acta 1477: 112-121.

Ryan CA and Moura DS (2002) Systemic wound signaling in plants: a new perception. Proceedings of National Academy of Sciences, USA 99: 6519-6520.

Sánchez-Serrano JJ, Schmidt R, Schell J, Willmitzer L (1986) Nucleotide sequence of proteinase inhibitor II encoding cDNA of potato (*Solanum tuberosum*) and its mode of expression. Molecular and General Genetics 203: 15-20.

Sanchis V and Bourguet D (2008) *Bacillus thuringiensis*: applications in agriculture and insect resistance management: A review. Agronomy for Sustainable Development 28:11-20.

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.* Journal of Insect Science 12(42):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.

Schägger H (2006) Tricine–SDS-PAGE. Nature Protocols 1:16-22.

Schirra HJ, Anderson MA, Craik DJ (2008) Structural refinement of insecticidal plant proteinase inhibitors from *Nicotiana alata*. Protein and Peptide Letters 15: 903-909.

Schirra HJ and Craik DJ (2005) Structure and folding of potato type II proteinase inhibitors: Circular permutation and intramolecular domain swapping. Protein and Peptide Letters 12: 421-431.

Schirra HJ, Guarino RF, Anderson MA, Craik DJ (2010) Selective removal of individual disulfide bonds within a potato type II serine proteinase inhibitor from *Nicotiana alata* reveals differential stabilization of the reactive-site loop. Journal of Molecular Biology 395: 609-626.

Schuler TH, Poppy GM, Kerry BR, Denholm I (1998) Insect resistant transgenic plants. Trends in Biotechnology 16: 168–175.

Semisotnov GV, Rodionova NA, Razgulyaev OI, Uversky VN, Gripas AF, Gilmanshin RI (1991) Study of the 'molten globule' intermediate state in protein folding by a hydrophobic fluorescent probe. Biopolymers 31: 119–128.

Shabab M, Kulkarni MJ, Khan MI (2008) Study of Papain–Cystatin Interaction by Intensity Fading MALDI-TOF-MS. Protein Journal 27: 7–12.

Sharma HC, Gowda CLL, Stevenson PC, Ridsdill-Smith TJ, Clement SL, Ranga Rao, GV, Romeis, J, Miles M, EL Bouhssini M (2007) Host plant resistance and insect pest management in chickpea In: Chickpea breeding and management. Eds. Shyam S.Yadav, Redden RJ, Chen W, Sharma B. (CAB International, UK) pp. 520-537.

Sharma HC, Sharma KK, Seetharam N, Ortiz R (2000) Prospect for using transgenic resistance to insects in crop improvement. Electronic Journal of Biotechnology 3: 76-95.

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 Science 161: 727-737.

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

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

Skibbe M, Qu N, Galis I, Baldwin IT (2008) Induced plant defenses in the natural environment: *Nicotiana attenuata* WRKY3 and WRKY6 coordinate responses to herbivory. Plant Cell 20:1984–2000.

Smyth E, Syme CD, Blanch EW, Hecht L, Vasak M, Barron LD (2001) Solution Structure of Native Proteins with Irregular Folds from Raman Optical Activity. Biopolymers 58: 138–151.

Srinivasan A, Giri AP, Gupta VS (2006) Structural and functional diversities in Lepidopteran serine proteases. Cellular and Molecular Biology Letters 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 Molecular Biology 57: 359-374.

Stenzel I, Hause B,Maucher H, Pitzschke A,Miersch O, Ziegler J, Ryan CA,Wasternack C (2003) Allene oxide cyclase dependence of the wound response and vascular bundle-specific generation of jasmonates in tomato- amplification in wound signalling. Plant Journal 33:577-589.

Strassner J, Schaller F, Frick UB, Howe GA, Weiler EW, Amrhein NA, Macheroux P, Schaller A (2002) Characterization and cdna-microarray expression analysis of 12-oxophytodienoate reductases reveals differential roles for octadecanoid biosynthesis in the local versus the systemic wound response. Plant Journal 32:585–601.

Stratmann JW and Ryan CA (1997) Myelin basic protein kinase activity in tomato leaves is induced systemically by wounding and increases in response to systemin and oligosaccharide elicitors. Proceedings of National Academy of Sciences, USA 94:11085–11089.

Sun JQ, Jiang HL, Li CY (2011) Systemin/Jasmonate-mediated systemic defense signalling in tomato. Molecular Plant 4(4): 607-615.

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. Biochimica et Biophysica Acta - General Subjects 1722: 156-167.

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, 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, 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.

Taylor BH, Young RJ, Scheuring CF (1993) Induction of a proteinase inhibitor II class gene by auxin in tomato roots. Plant Molecular Biology 23: 1005-1014.

Telang A, Booton V, Chapman RF, Wheeler DE (2000) How female caterpillars accumulate their nutrient reserves? Journal of Insect Physiology 47: 1055-1064.

Telang MA, Srinivasan A, Patankar A, Harsulkar A, Joshi V, Damle A, Deshpande V, Sainani MN, Ranjekar PK, Gupta G, Birah A, Rani S, Kachole M, Giri AP, Gupta VS (2003) Bitter gourd proteinase inhibitor: potential growth inhibitors of of *Helicoverpa armigera* and *Spodoptera litura*. Phytochemistry 63: 643-652.

Terenius O, Papanicolaou A, Garbutt JS, et al. (2011). RNA interference in Lepidoptera: An overview of successful and unsuccessful studies and implications for experimental design. Journal of Insect Physiology 57: 231-245.

Thaler JS, Bostock RM (2004) Interactions between abscisic-acid-mediated responses and plant resistance to pathogens and insects. Ecology 85: 48–58.

Thaler JS, Owen B, Higgins VJ (2004) The role of the jasmonate response in plant susceptibility to diverse pathogens with a range of lifestyles. Plant Physiology 135: 530-538.

Tovchigrechko A, Vakser IA (2006) GRAMM-X public web server for protein-protein docking. Nucleic Acids Research 34(W3): 10-4.

Ussuf KK, Laxmi NH, Mitra R (2001) Proteinase inhibitors: Plant-derived genes of insecticidal protein for developing insect-resistant transgenic plants. Current Science 80: 847-853.

Valueva TA, Mosolov VV (2004) Role of inhibitors of proteolytic enzymes in plant defense against phytopathogenic microorganisms. Biokhimiya 69(11): 1600-1606.

Van Dam NM, Witjes L, Svatos A (2004) Interactions between aboveground and belowground induction of glucosinolates in two wild Brassica species. New Phytologist 161: 801–810.

Van Wees SCM, De Swart EAM, Van Pelt JA, Van Loon LC, Pieterse CMJ (2000) Enhancement of induced disease resistance by simultaneous activation of salicylate- and jasmonate-dependent defense pathways in *Arabidopsis thaliana*. Proceedings of National Academy of Sciences, USA 97**:** 8711–8716.

Venturi M, Seifert C, Hunte C (2002) High level production of functional antibody fab fragments in an oxidizing bacterial cytoplasm. Journal of Molecular Biology 315: 1–8.

Villanueva J, Yanes O, Querol E, Serrano L, Aviles FX (2003) Identification of protein ligands in complex biological samples using Intensity-fading Maldi-TOF Mass spectrometry. Analytical Chemistry 75: 3385-3395.

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

Voelckel C, Schittko U, Baldwin IT (2001) Herbivore-induced ethylene burst reduces fitness costs of jasmonate- and oral secretion induced defenses in *Nicotiana attenuata*. Oecologia 127:274-280.

Voelckel C, Weisser WW, Baldwin IT (2004) An analysis of plant-aphid interactions by different microarray hybridization strategies. Molecular Ecology 2004, 13: 3187-3195.

Voet D and Voet JG (2010) Biochemistry, 4th edition. John Wiley and Sons Inc., Corporate Headquarters, 111 River street, Hoboken, NJ 07030-5774.

Volpicella M, Ceci LR, Cordewener J, America T, Gallerani R, Bode W, Jongsma MA, Beekwilder J (2003) Properties of purified gut trypsin from *Helicoverpa zea* adapted to proteinase inhibitors. European Journal of Biochemistry 270: 10-19.

Wallling LL (2000) The myriad plant responses to herbivores. Journal of Plant Growth Regulation 19:195-216.

Wang J, Xia Q, He X, Dai M, Ruan J, et al. (2005) Silk DB: a knowledgebase for silkworm biology and genomics. Nucleic Acids Research 33: D399-D402.

Wasternack C (2007) Jasmonates: an update on biosynthesis, signal transduction and action in plant stress response, growth and development. Annals of Botany 100(4): 1-17.

Wasternack C, Ortel B, Miersch O, Kramell R, Beale M, Greulich F, Feussner I, Hause B, Krumm T, Boland W et al (1998) Diversity in octadecanoid-induced gene expression of tomato. Journal of Plant Physiology 152:345-352.

Wetzel R, Perry LJ, Baase WA and Becktel WJ (1988) Disulfide bonds and thermal stability in T4 lysozyme. Proceedings of National Academy of Sciences, USA 85: 401-405.

Wildon DC, Thain JF, Minchin PEH, Gubb IR, Reilly AJ, Skipper YD, Doherty HM, O Donnell PJ, Bowles DJ (1992) Electrical signalling and systemic proteinase inhibitor induction in the wounded plant. Nature 360: 362-65.

Wu J and Baldwin IT (2010) New insights into plant responses to the attack from insect herbivores. Annual Review of Genetics 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. Molecular Breeding 3: 371-380.

Xia Q, Zhou Z, Lu C, Cheng D, Dai F, et al. (2004) A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). Science 306: 1937-1940.

Xie J, Ouyang XZ, Xia KF, Huang YF, Pan WB, Cai YP, Xu X, Li B, Xu ZF (2007) Chloroplast-like organelles were found in enucleate sieve elements of transgenic plants overexpressing a proteinase inhibitor. Bioscience, Biotechnology and Biochemistry 71: 2759-2765.

Xiong S, Wang Y, Rong Ren X, Li B, Zhang MY, Luo Y, Zhang L, Xie QL, Su KY (2005) Solubility of disulfide-bonded proteins in the cytoplasm of *Escherichia coli* and its "oxidizing" mutant. World Journal of Gastroenterology 11(7): 1077-1082.

Xu ZF, Qi WQ, Ouyang XZ, Yeung E, Chye ML (2001) A proteinase inhibitor II of *Solanum americanum* is expressed in phloem. Plant Molecular Biology 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.

Yanes O, Aviles FX, Roepstorff P, Jørgensen TJD (2007) Exploring the "Intensity Fading" Phenomenon in the Study of Noncovalent Interactions by MALDI-TOF Mass Spectrometry. American Society for Mass Spectrometry 18: 359-367.

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

Young MA and Pysh ES (1975) Vacuum ultraviolet circular dichroism of Poly (L-proline) I and II. Journal of the American chemical Society 97(18): 5100-5103.

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.

Zavala JA, Patankar AG, Gase K, Baldwin IT (2004a) Constitutive and inducible trypsin proteinase inhibitor production incurs large fitness costs in *Nicotiana attenuata*. Proceedings of the National Academy of Sciences, USA 101: 1607-1612.

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

Zhang H-Y, Xie X-Z, Xu Y-Z, Wu N-H (2004) Isolation and functional assessment of a tomato proteinase inhibitor II gene. Plant Physiology and Biochemistry 42: 437-444.

Zhu-Salzman K, Salzman RA, Ahn JE, Koiwa H (2004) Transcriptional regulation of sorghum defense determinants against a phloem-feeding aphid. Plant Physiology 134: 420-431.

APPENDIX I

The sequences were taken from NCBI as follows: *N. attenuata* : NatPI-1 (AY426751), NatPI-2 (DQ158200), NatPI-3 (AY297103), NatPI-4 (AF542547); *N. alata*: NalPI-1 (U08219), NalPI-2 (AF105340); *N. clevelandii*: NclPI-1 (DQ158203), NclPI-2 (DQ158203); *N. quadrivalvis*: NquPI-1 (DQ158202), NquPI-2 (DQ158198); *N. obtusifolia*: NobPI-1 (DQ158201), NobPI-2 (DQ158197); *N. rustica*: NruPI-1 (DQ158196); NcoPI-1 (*N. corymbosa*, DQ158195); NacPI-1 (*N. acuminate*, DQ158194); NpaPI-1 (*N. pauciflora*, DQ158193); NmiPI-1 (*N. miersii*, DQ158192); NspPI-1 (*N. spegazzinii*, DQ158191); NliPI-1 (*N. linearis*, DQ158190); *N. tabacum*: NtaPI-1 (Q158189), NtaPI-2 (DQ071272), NtaPI-3 (EF408803), NtaPI-4 (Z29537); NsyPI-1 (*N. sylvestris*, DQ158188); NrePI-1 (*N. repanda*, DQ158187); NumPI-1 (*N. umbratica*, DQ158186); NsiPI-1 (*N. simulans*, DQ158185); NmePI-1 (*N. megalosiphon*, DQ158184); NhePI-1 (*N. hesperis*, DQ158183); NbePI-1 (*N. benthamiana*, DQ158182); NcaPI-1 (*N. cavicola*, DQ158181); NglPI-1 (*N. glutinosa*, AF205852); NglPI-2 (*N. glutinosa*, AF205851); NglPI-3 (*N. glutinosa*, AF208020); *L. esculentum*: LesPI-1 (L21194), LesPI-2 (K03291), LesPI-3 (AY129402), Le sPI-4 (AY007240), LesPI-5 (X94946); SphPI-1 (*S. phureja*, AY517498), SphPI-2 (*S. phureja*, AY247794); *S. nigrum*: SniPI-1 (AY422686); SniPI-2 (GU133372); *S. tuberosum*: StuPI-1 (U45450), StuPI-2 (L37519), StuPI-3 (DQ168323), StuPI-4 (DQ168321), StuPI-5 (DQ168313), StuPI-6 (EF469204), StuPI-7 (Z13992), StuPI-8 (Z12753), StuPI-9 (X03779), StuPI-10 (X03778), StuPI-11 (X04118); SlyPI-1 (*S. lycopersicum*, AB110700); SamPI-1 (*S. americanum*, AF174381), SamPI-2 (*S. americanum*, AF209709); *C. annuum*: CanPI-7(DQ005913), CanPI-1 (AF039398), CanPI-2 (AF221097), CanPI-3 (AY986465), CanPI-4 (AY986466), CanPI-5 (DQ005912), CanPI-8 (DQ005914), CanPI-9 (DQ005915), CanPI-10 (DQ005916), CanPI-11 (DQ008950), CanPI-13 (EF136387), CanPI-14 (EF136388), CanPI-15 (EF136389), CanPI-16 (EF125182), CanPI-17 (EF136381), CanPI-18 (EF136382), CanPI-19 (EF136383), CanPI-20 (EF136384), CanPI-21 (EF136385), CanPI-22 (EF136386), CanPI-23 (EF144129).

APPENDIX II

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

CanPI-24 [JX106474]

ATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTTCCTACTTGTTCTTGAAATATTGCTTCTACATGTT
GATGCCAAGGCTTGTTCAGAAGAAAACGCAGAAAATCGCATATGCACCAATTGCTGTGCAGGCCGTA
AGGGTTGCAACTATTACAGTGCTGATGGGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCC
AAAAGCTTGCCCTCGGTATTGTGATACAAGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAAC
GCAGAAAATCGCATATGCACCAATTGCTGTGCAGGCCGTAAGGGTTGCAACTATTACAGCGCTGACG
GGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAAGCTTGCCCTAGGAATTGTGATCC
AAATATTGCCTATTCACTGTGTCTCTATGAAAAGTAA

CanPI-25 [JX106475]

ATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTTCCTACTTGTTCTTGGAGAGTCTGACCCCAACAAC
CCAAAAGCTTGCCCTCGGTATTGTGATACAAGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAA
ACGCAGAAAATCGCATATGCACCAATTGCTGTGCAGGCCGTAAGGGTTGCAACTATTACAGCGCTGA
CGGGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAAGCTTGCCCTCGGTATTGTGATA
CAAGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAACGCAGAAAATCGCATATGCACCAATTG
CTGTGCAGGCCGTAAGGGTTGCAACTATTACAGCGCTGACGGGACTTTCATTTGTGAAGGAGAGTCT
GACCCCAACAACCCAAAAGCTTGCCCTAGGAATTGTGATCCAAATATTGCCTATTCACTGTGTCTCTA
TGAAAAGTAA

CanPI-26 [JX106476]

ATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTCCCTACTTGTACTTGGAATATTGCTTCTACATGTT
GATGCCAAGGCTTGTTCACAAAGAAACGCAAAAGAACCCATATGCACCAATTGTTGTGCAGGCCGTA
AGGGTTGCAACTATTACAGTGCTGATGGGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCC
AAAACCTTGCACTCTGAACTGTGATCCAAGAATTTTCTATTCAAAGTGTCCACGTTCCGAAGCAAGTG
CAGAACAACCCATATGCACCAATTGTTGTGCAGGCCTCAAGGGTTGCAACTATTACAATGCTGACGG
GACTTTCATTTGTGAGGGAGAGTCTGACCCCAACCACCCAAAAGCTTGCCCCAAGAATTGTGATCCTA
ATATTGCCTATTCACTATGTCTGTATGAAAAGTAA

CanPI-27 [JX106477]

ATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTTCGTACTTGTTCTTGGAATATTGCTTCTACATGTT
GATGCCAAGGCTTGTTCAGAAGAAAACGCAGAAAATCGCATATGCACCAATTGCTGTGCAGGTCGTA
AGGGTTGCAACTATTACAGTGCTGACGGGACCTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCC
AAAGCCTTGCACTCTGAACTGTGATACAAGAATTGCCTATTCAATATGTCCACGTTCCGAAGGAAAC
GCAGAAAATCGCATATGCACCAATTGCTGTGCAGGCCGTAAGGGTTGCAACTATTACAGTGCTGACG
GGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAGCCCAAAAGCTTGCCCTAGGAATTGTGATCC
AAATATTGCCTATTCACTGTGTCTCTATGAAAAGTAA

CanPI-28 [JX106478]

ATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTCCCTACTTGTACTTGGAATATTGCTTCTACATGTT
GATGCCAAGGCTTGTTCACAAAGAAACGCAAAAGAACCCATATGCACCAATTGTTGTACAGGCCGTA
AGGGTTGCAACTATTACAGTGCTGATGGGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCC
AAAACCTTGTACTTTGAACTGTGATCCAAGAATTTTCTATTCAAAGTGTCCACGTTCCGAAGGAAACG
CAGAAAATCGCATATGCACCAATTGCTGTGCAGGCCGTAAGGGTTGCAACTATTACAGTGCTGATGG
GACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAAGCTTGCCCTAGGAATTGTGATCCA
AATATTGCCTATTCACTGTGTCTCTATGAAAAGTAA

CanPI-29 [JX106479]

ATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTTCGTACTTGTTCTTGGAATATTGCTTCTACATGTT
GATGCCAAGGCTTGTTCAGAAGAAAACGCAGAAAATCGCATATGCACCAATTGCTGTGCAGGTCGTA
AGGGTTGCAACTATTACAGTGCTGACGGGACCCTCATTTGTGAAGGAGAGTCTGACCCCAACAACCC
AAAACCTTGCACTCCGAACTGTGATCCAAGAATTTTCTATTCAAAATGTCCACGTTCCGAAGGAAACG
TAGAAAATCGCATATGCACCAATTGCTGTGCAGGCCGTAAGGGTTGCAACTATTACAGTGCTGACGG
GACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAAGCTTGCCCTAGGAATTGTGATCCA
AATATTGCCTATTCACTGTGTCTCTATGAAAAGTAA

CanPI-30 [JX106480]

ATGGCTGTTCCCAGAGAAGTTAGTTTCCTTGCTTTCGTACTTGTTCTTGGAATATTGCTTCTACATGTT
GATGCCAAGGCTTGTTCAGAAGAAAACGCAGAAAATCGCATATGCACCAATTGCTGTGCAGGTCGTA
AGGGTTGCAACTATTACAGTGCTGACGGGACCTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCC
AAAACCTTGCACTCTGAACTGTGATCCAAGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAAC
GCAGGAAATCGCATATGCACCAATTGCTGTGCAGGCCGTAAGGGTTGCAACTATTACAGTGCTGACG
GGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAAGCTTGCCCTAGGAATTGTGATCC
AAATATTGCCTATTCACTGTGTCTCTATGAAAAGTAA

CanPI-31 [JX106481]

ATGGCTGTTCCCAAAGAAGTTAGTTTCCCTGCTTTCCTACTTGTTCTTGAAATATTGCTTCTACATGTT
GATGCCAAGGCTTGTTCAGAAGAAAACGCAGAAAATCGCATATGCACCAATTGCTGTGCAGGCCGTA
AGGGTTGCAACTATTACAGTGCTGATGGGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCC
AAAAGCTTGCCCTCGGTATTGTGATACAAGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAAC
GCAGAAAATCGCCTATGCACCAATTGCTGTGCAGGCCGTAAGGGTTGCAACTATTACAGCGCTGACG
GGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAAGCTTGCCCTAGGAATTGTGATCC
AAATATTGCCTATTCACTGTGTCTCTATGAAAAGTAA

CanPI-32 [JX106482]

ATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTCCCTACTTGTACTTGGAATATTGCTTCTACATGTT
GATGCCAAGGCTTGTTCACAAAGAAACGCAAAAGAACCCATATGCACCAATTGTTGTGCAGGCCGTA
AGGGTTGCAACTATTACAGTGCTGATGGGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCC
AAAACCTTGTACTTTGAACTGTGATCCAAGAATTTTCTATTCAAAGTGTCCACGTTCCGAAGGAAACG
CAGAAAATCGCATATGCACCAATTGCTGTGCAGGCCGTAAGGGTTGCAACTATTACAGTGCTGATGG
GACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAAGCTTGCCCTCGGTATTGTGATACAA
GAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAACGCAGAAAATCGCATATGCACCAATTGCTG
TGCAGGCCGTAAGGGTTGCAACTATTACAGCGCTGACGGGACTTTCATTTGTGAAGGAGAGTCTGAC
CCCAACAACCCAAAAGCTTGCCCTCGGAATTGTGATACAAGAATTGCCTATTCAAAATGTCCACGTTC
CGAAGGAAACGCAGAAAATCGCCTATGCACCAATTGCTGTGCAGGCCGTAAGGGTTGCAACTATTAC
AGCGCTGACGGGACTTTCATTTGTGAAGGAAAGTCTGACCCCAACAACCCAAAAGCTTGCCCTAGGA
ATTGTGATCCAAATATTGCCTATTCACTGTGTCTCTATGAAAAGTAA

CanPI-33 [JX106483]

ATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTTCCTACTTGTTCTTGAAATATTGCTTCTACATGTT
GATGCCAAGGCTTGTTCAGAAGAAAACGCAGAAAATCGCATATGCACCAATTGCTGTGCAGGCCGTA
AGGGTTGCAACTATTACAGTGCTGATGGGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCC
AAAAGCTTGCCCTCGGTATTGTGATACAAGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAAC
GCAGAAAATCGCATATGCaCCAATTGCTGTGCAGGCCGTAAGGGTTGCAACTATTACAGCGCTGACG
GGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAAGCTTGCCCTCGGAATTGTGATAC
AAGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAACGCAGAAAATCGCCTATGCACCAATTGC
TGTGCAGGCCGTAAGGGTTGCAaCTATTACAGCGCTGACGGGACTTTCATTTGTGAAGGAGAGTCTGA
CCCCAACAACCCAAAAGCTTGCCCTAGGAATTGTGATCCAAATATTGCCTATTCTAAATGTCCACGTT
CCGAAGGAAACGCAGAAAATCGCCTATGCACCAATTGCTGTGCAGGCCGTAAGGGTTGCAACTATTA
CAGCGCTGACGGGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAAGCTTGCCCTAGG
AATTGTGATCCAAATATTGCCTATTCACTGTGTGCTTATGAAAAGTAA

CanPI-34 [JX106484]

ATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTCCCTACTTGTACTTGGAATATTGCTTCTACATGTT
GATGCCAAGGCTTGTTCACAAAGAAACGCAAAAGAACCCATATGCACCAATTGTTGTGCAGGCCGTA
AGGGTTGCAACTATTACAGTGCTGATGGGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCC
AAAACCTTGTACTTTGAACTGTGATCCAAGAATTTTCTATTCAAAGTGTCCACGTTCCGAAGGAAACG
CAGAAAATCGCATATGCACCAATTGCTGTGCAGGCCGTAAGGGTTGCAACTATTACAGTGCTGATGG
GACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAACCTTGCCCTCGGAATTGTGATACA
AGAATTGCCTATTCAAAATGTCCATGTTCCGAAGGAAACGCAGAAAATCGCATATGCACCAATTGCT
GTGCAGGTCGTAAGGGTTGCAACTATTACAGTGCTGGCGGGACCTTCATTTGTGAAGGAGAGTCTGA
CCCCAACAACCCAAAACCTTGCACTCTGAACTGTGATCCAAGAATTTTCTATTCAAAGTGTCCACGTT
CCGAAGCAAGTGCAGAACAACCCATATGCACCAATTGTTGTGCAGGCCTCAAGGGTTGCAACTATTA
CAATGCTGACGGGACTTTCATTTGTGAGGGAGAGTCTGACCCCAACCACCCAAAAGCTTGCCCCAAG
AATTGTGATCCTAATATTGCCTATTCACTATGTCTGTATGAAAAGTAA

CanPI-35 [JX106485]

ATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTCCCTACTTGTACTTGGAATATTGCTTCTACATGTT
GATGCCAAGGCTTGTTCACAAAGAAACGCAAAAGAACCCATATGCACCAATTGTTGTGCAGGCCGTA
AGGGTTGCAACTATTACAGTGCTGATGGGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCC
AAAACCTTGTACTTTGAACTGTGATCCAAGAATTTTCTATTCAAAGTGTCCACGTTCCGAAGGAAACG
CAGAAAATCGCATATGCACCAATTGCTGTGCAGGCCGTAAGGGTTGCAACTATTACAGTGCTGATGG
GACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAACCTTGCCCTCGGAATTGTGATACA
AGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAACGCAGAAAATCGCATATGCACCAATTGCT
GTGCAGGTCGTAAGGGTTGCAACTATTACAGTGCTGACGGGACCTTCATTTGTGAAGGAGAGTCTGA
CCCCAACAACCCAAAACCTTGCACTCTGAACTGTGATCCAAGAATTTTCTATTCAAAGTGTCCACGTT
CCGAAGCAAGTGCAGAACAACCCATATGCACCAATTGTTGTGCAGGCCTCAAGGGTTGCAACTATTA
CAATGCTGACGGGACTTTCATTTGTGAGGGAGAGTCTGACCCCAACCACCCAAAAGCTTGCCCCAAG
AATTGTGATCCTAATATTGCCTATTCACTATGTCTGTATGAAAAGTAA

CanPI-36 [JX106486]

ATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTTCGTACTTGTTCTTGGAATATTGCTTCTACATGTT
GATGCCAAGGCTTGTTCAGAAGAAAACGCAGAAAATCGCATATGCGCCAATTGGTGTGCAGGCCGTA
AGGGTTGCTACTATTACAGTGATGATGGGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCC
AAAACCTTGCCCTCGGAATTGTGATACAAGAATTGCCTATTCAATATGTCCACGTTCCGAAGGAAAC
GCAGAAAATCGCATATGCACCAATTGCTGTGCAGGCCGTAAGGGTTGCAACTATTACAGCGCTGACG
GGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAAGCTTGCCCTCGGTATTGTGATACA
AGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAACGCAGAAAATCGCATATGCACCAATTGCT
GTGCAGGCCGTAAGGGTTGCAACTATTACAGCGCTGACGGGACTTTCATTTGTGAAGGAGAGTCTGA
CCCCAACAACCCAAAAGCTTGCCCTCGGAATTGTGATACAAGAATTGCCTATTCAAAATGTCCACGTT
CCGAAGGAAACGCAGAAAATCGCCTATGCACCAATTGCTGTGCAGGCCGTAAGGGTTGCAACTATTA
CAGCGCTGACGGGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAAGCTTGCCCTAGG
AATTGTGATCCAAATATTGCCTATTCACTGTGTCTCTATGAAAAGTAA

CanPI-37 [JX106487]

ATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTTCCTACTTGTTCTTGAAATATTGCTTCTACATGTT
GATGCCAAGGCTTGTTCAGAAGAAAACGCAGAAAATCGCATATGCACCAATTGCTGTGCGGGCCGTA
AGGGTTGCAACTATTACAGTGCTGATGGGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCC
AAAAGCTTGCCCTCGGTATTGTGATACAAGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAAC
GCAGAAAATCGCATATGCACCAATTGCTGTGCAGGCCGTAAGGGTTGCAACTATTACAGCGCTGACG
GGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAAGCTTGCCCTCGGAATTGTGATAC
AAGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAACGCAGAAAATCGCCTATGCACCAATTGC
TGTGCAGGCCGTAAGGGTTGCAACTATTACAGCGCTGACGGGACTTTCATTTGTGAAGGAGAGTCTG
ACCCCAACAACCCAAAAGCTTGCCCTAGGAATTGTGATCCAAATATTGCCTATTCACCGTGTCTCTAT
GAAAAGTAA

CanPI-38 [JX106488]

ATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTTCCTACTTGTTCTTGAAATATTGCTTCTACATGTT
GATGCCAAGGCTTGTTCAGAAGAAAACGCAGAAAATCGCATATGCACCAATTGCTGTGCAGGCCGTA
AGGGTTGCAACTATTACAGTGCTGATGGGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCC
AAAAGCTTGCCCTCGGTATTGTGATACAAGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAAC
GCAGAAAATCGCATATGCACCAATTGCTGTGCAGGCCGTAAGGGTTGCAACTATTACAGCGCTGACG
GGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAAGCTTGCCCTCGGAATTGTGATAC
AAGAATTGCCTATTCAAAATGTCCACGTTCCGGAGGAAACGCAGAAAATCGCCTATGCACCAATTGC
TGTGCAGGCCGTAAGGGTTGCAACTATTACAGCGCTGACGGGACTTTCATTTGTGAAGGAGAGTCTG
ACCCCAACAACCCAAAAGCTTGCCCTAGGAATTGTGATCCAAATATTGCCTATTCACTGTGTCTCTAT
GAAAAGTAA

CanPI-39 [JX106489]

ATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTTCCTACTTGTTCTTGAAATATTGCTTCTACATGTT
GATGCCAAGGCTTGTTCAGAAGAAAACGCAGAAAATCGCATATGCACCAATTGCTGTGCAGGCCGTA
AGGGTTGCAACTATTACAGTGCTGATGGGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCC
AAAAACTTGCCCTCGGTATTGTGATACAAGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAAC
GCAGAAAATCGCATATGCACCAATTGCTGTGCAGGCCGTAAGGGTTGCAACTATTACAGCGCTGACG
GGACTTTCATTTGCGAAGGAGAGTCTGACCCCAACAACCCAAAAGCTTGCCCTCGGAATTGTGATAC
AAGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAACGCAGAAAATCGCCTATGCACCAATTGC
TGTGCAGGCCGTAAGGGTTGCAACTATCACAGCGCTGACGGGACTTTCATTTGTGAAGGAGAGTCTG
ACCCCAACACCCCAAAAGCTTGCCCTAGGAATTGTGATCCAAACATTGCCCATTCACTGTGTCTCTAT
GAAAAGTAA

CanPI-40 [JX106490]

ATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTTCCTACTTGTTCTTGAAATATTGCTTCTACATGTT
GATGCCAAGGCTTGTTCAGAAGAAAACGCAGAAAATCGCATATGCACCAATTGCTGTGCAGGCCGTA
AGGGTTGCAACTATTACAGTGCTGATGGGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCC
AAAAGCTTGCCCTCGGAATTGTGATACAAGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAAC
GCAGAAAATCGCCTATGCACCAATTGCTGTGCAGGCCGTAAGGGTTGCAACTATTACAGCGCTGACG
GGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAACCTTGCCCTCGGAATTGTGATAC
AAGAATTGCCTATTCAATATGTCCACGTTCCGAAGGAAACGCAGAAAATCGCATATGCACCAATTGC
TGTGCAGGCCGTAAGGGTTGCAACTATTACAGTGCTGACGGGACTTTCATTTGTGAAGGAGAGTCTG
ACCCCAACAACCCAAAAGCTTGCCCTAGGAATTGTGATCCAAATATTGCCTATTCACTGTGTCTCTAT
GAAAAGTAA

CanPI-41 [JX106491]

ATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTTCGTACTTGTTCTTGGAATATTGCTTCTACATGTT
GATGCCAAGGCTTGTTCAGAAGAAAACGCAGAAAATCGCATATGCACCAATTGCTGTGCAGGCCGTA
AGGGTTGCAACTATTACAGTGCTGATGGGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCC
AAAAGCTTGCCCCCGGTATTGTGATACAAGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAAC
GCAGAAAATCGCATATGCACCAATTGCTGTGCAGGCCGTAAGGGTTGCAACTATTACAGCGCTGACG
GGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAAGCTTGCCCTCGGAATTGTGATAC
AAGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAACGCAGAAAATCGCCTATGCACCAATTGC
TGTGCAGGCCGTAAGGGTTGCAACTATTACAGCGCTGACGGGACTTTCATTTGTGAAGGAGAGTCTG
ACCCCAACAACCCGAAAGCTTGCCCTAGGAATTGTGATCCAAATATTGCCTATTCACTGTGTCTCTAT
GAAAAGTAA

CanPI-42 [JX106492]

ATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTTCGTACTTGTTCTTGGAATATTGCTTCTACATGTT
GATGCCAAGGCTTGTTCAGAAGAAAACGCAGAAAATCGCATATGCACCAATTGCTGTGCAGGTCGTA
AGGGTTGCAACTATTACAGTGCTGACGGGACCTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCC
AAAACCTTGCCCTCGGAATTGTGATACAAGAATTGCCTATTCAATATGTCCACGTTCCGAAGGAAAC
GCAGAAAATCGCATATGCACCAATTGCTGTGCAGGCCGTAAGGGTTGCAACTATTACAGTGCTGATG

GGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAAGCTTGCCCTCGGTATTGTGATACA
AGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAACGCAGAAAATCGCATATGCACCAATTGCT
GTGCAGGCCGTAAGGGTTGCAACTATTACAGTGCTGACGGGACTTTCATTTGTGAAGGAGAGTCTGA
CCCCAACAACCCAAAAGCTTGCCCTAGGAATTGTGATCCAAATATTGCCTATTCACTGTGTCTCTATG
AAAAGTAA

CanPI-43 [JX106493]

ATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTTCGTACTTGTTCTTGGAATATTGCTTCTACATGTT
GATGCCAAGGCTTGTTCAGAAGAAAACGCAGAAAATCGCATATGCACCAATTGCTGTGCAGGTCGTA
AGGGTTGCAACTATTACAGTGCTGACGGGACCTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCC
AAAACCTTGCACTCTGAACTGTGATCCAAGAATTTTCTATTCAAAATGTCCACGTTCCGAAGGAAACG
CAGAAAATCGCATATGCACCAATTGTTGTGCAGGCCGTAAGGGTTGCTACTATTACTGTGATGATGG
GACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAACCTTGCCCTCGGAATTGTGATACA
AGAATTGCCTATTCAATATGTCCACGTTCCGAAGGAAACGCAGAAAATCGCATATGCACCAATTGCT
GTGCAGGCCGTAAGGGTTGCAACTATTACAGTGCTGACGGGACTTTCATTTGTGAAGGAGAGTCTGA
CCCCAACAACCCAAAAGCTTGCCCTAGGAATTGTGATCCAAATATTGCCCATTCACTGTGTCTCTATG
AAAAGTAA

CanPI-44 [JX106494]

ATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTTCGTACTTGTTCTTGGAATATTGCTTCTACATGTT
GATGCCAAGGCTTGTTCAGAAGAAAACGCAGAAAATCGCATATGCACCAATTGCTGTGCAGGTCGTA
AGGGTTGCAACTATTACAGTGCTGACGGGACCTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCC
AAAACCTTGCACTCTGAACTGTGATCCAAGAATTTTCTATTCAAAATGTCCACGTTCCGAAGGAAACG
CAGAAAATCGCATATGCACCAATTGTTGTGCAGGCCGTAAGGGTTGCTACTATTACAGTGATGATGG
GACTTTCATTTGTGAAGGAGAGTCTGGCCCCAACAACCCAAAACCTTGCCCTCGGAATTGTGATACA
AGAATTGCCTATTCAATATGTCCACGTTCCGAAGGAAACGCAGAAAATCGCATATGCACCAATTGCT
GTGCAGGCCGTAAGGGTTGCAACTATTACAGTGCTGACGGGACTTTCATTTGTGAAGGAGAGTCTGA
CCCCAACAACCCAAAAGCTTGCCCTAGGAATTGTGATCCAAATATTGCCTATTCACTGTGTCTCTATG
AAAAGTAA

CanPI-45 [JX106495]

ATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTTCGTACTTGTTCTTGGAATATTGCTTCTACATGTT
GATGCCAAGGCTTGTTCAGAAGAAAACGCAGAAAATCGCATATGCACCAATTGCTGTGCAGGTCGTA
AGGGTTGCAACTATTACAGTGCTGACGGGACCTTCATTTGCGAAGGAGAGTCTGACCCCAACAACTC
AAAACCTTGCACTCTGAACTGTGATCCAAGAATTTTCTATTCAAAATGTCCACGTTCCGAAGGAAACG
CAGAAAATCGCATATGCACCAATTGTTGTGCAGGCCGTAAGGGTTGCTACTATTACAGTGATGATGG
GACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAACCTTGCCCTCGGAATTGTGATACA
AGAATTGCCTATTCAATATGTCCACGTTCCGAAGGAAACGCAGAAAATCGCATATGCACCAATTGCT
GTGCAGGCCGTAATTTTTGCAACTATTACAGCGCTGACGGGACTTTCATTTGTGAAGGAGAGTCTGAC
CCCAACAACCCAAAAGCTTGCCCTAGGAATTGTGATCCAAATATTGCCTATTCACTGTGTCTCTATGA
AAAGTAA

CanPI-46 [JX106496]

ATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTTCGTACTTGTTCTTGGAATATTGCTTCTACATGTT
GATGCCAAGGCTTGTTCAGAAGAAAACGCAGAAAATCGCATATGCACCAATTGCTGTGCAGGTCGTA
AGGGTTGCAACTATTACAGTGCTGACGGGACCTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCC
AAAACCTTGCACTCTGAACTGTGATCCAAGAATTTTCTATTCAAAATGTCCACGTTCCGAAGGAAACG
CAGAAAATCGCATATGCACCAATTGTTGTGCAGGCCGTAAGGGTTGCTACTATTACAGTGATGATGG
GACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAACCTTGCCCTCGGAATTGTGATACA
AGAATTGCCTATTCAATATGTCCACGTTCCGAAGGAAACGCAGAAAATCGCATATGCACCAATTGCT
GTGCAGGCCGTAAGGGTTGCAACTATTACAGTGCTGACGGGACTTTCATTTGTGAAGGAGAGTCTGA
CCCCAACAACCCAAACGCTTGCCCTAGGAATTGTGATCCAAATATTGCCTATTCACTGTGTCTCTATG
AAAAGTAA

CanPI-47 [JX106497]

ATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTTCGTACTTGTTCTTGGAATATTGCTTCTACATGTT
GATGCCAAGGCTTGTTCAGAAGAAAACGCAGAAAATCGCATATGCACCAATTGCTGTGCAGGTCGTA
AGGGTTGCAACTATTACAGTGCTGACGGGACCTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCC
AAAACCTTGCACTCTGAACTGTGATCCAAGAATTTTCTATTCAAAATGTCCACGTTCCGAAGGAAACG
CAGAAAATCGCATATGCACCAATTGTTGTGCAGGCCGTAAGGGTTGCTACTATTACAGTGATGATGG
GACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAACCTTGCCCTCGGAATTGTGATACA
AGAATTGCCTATTCAATATGTCCACATTCCGAAGGAAACGCAGAAAATCGCATATGCACCAATTGCT
GTGCAGGCCGTAAGGGTTGCAACTATTACAGTGCTGACGGGACTTTCATTTGTGAAGGAGAGTCTGA
CCCCAACAACCCAAAAGCTTGCCCTAGGAATTGTGATCCAAATATTGCCTATTCACTGTGTCTCTATG
AAAAGTAA

CanPI-48 [JX106498]

ATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTTCGTACTTGTTCCTGGAATATTGCTTCTACATGTT
GATGCCAAGGCTTGTTCAGAAGAAAACGCAGAAAATCGCATATGCACCAATTGCTGTGCAGGTCGTA
AGGGTTGCAACTATTACAGTGCTGACGGGACCTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCC
AAAACCTTGCACTCTGAACTGTGATCCAAGAATTTTCTATTCAAAATGTCCACGTTCCGAAGGAAACG
CAGAAAATCGCATATGCACCAATTGTTGTGCAGGCCGTAAGGGTTGCTACTGTTACAGTGATGATGG
GACTTTCATTTGCGAAGGAGAGTCTGACCCCAACAACCCAAAACCTTGCCCTCGGAATTGTGATACA
AGAATTGCCTATTCAATATGTCCACGTTCCGAAGGAAACGCAGAAAATCGCATATGCACCAATTGCT
GTGCAGGCCGTAAGGGTTGCAACTATTACAGTGCTGACGGGACTTTCATTTGTGAAGGAGAGTCTGA
CCCCAACAACCCAAAAGCTTGCCCTAGGAATTGTGATCCAAATATTGCCTATTCACTGTGTCTCTATG
AAAAGTAA

CanPI-49 [JX106499]

ATGGCTGTTCCCAAAGAAGTTAGTTCCCTTGCTTTCGTACTTGTTCTTGGAATATTGTTTCTACATGTT
GATGCCAAGGCTTGTTCAGAAGAAAACGCAGAAAATCGCATATGCACCAATTGCTGTGCAGGTCGTA
GGGGTTGCAACTATTACAGTGCTGACGGGACCTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCC
AAAACCTTGCACTCTGAACTGTGATCCAAGAATCTTCTATTCAAAATGTCCACGTTCCGAAGGAAACG
CAGAAAATCGCATATGCACCAGTTGTTGTGCAGGCCGTAAGGGTTGCTACTATTACAGTGATGATGG
GACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAACCTTGCCCTCGGAATTGTGATACA
AGAATTGCCTATTCAATATGTCCACGTTCCGAAGGAAACGCAGAAAATCGCATATGCACCAATTGCT
GTGCAGGCCGTAAGGGTTGCAACTATTACAGTGCTGACGGGACTTTCATTTGTGAAGGAGAGTCTGA
CCCCAACAACCCAAAAGCTTGCCCTAGGAATTGTGATCCAAATATTGCCTATTCACTGTGTCTCTATG
AAAAGTAA

CanPI-50 [JX106500]

ATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTTCGTACTTGTTCTTGGAATATTGCTTCTACATGTT
GATGCCAAGGCTTGTTCAGAAGAAAACGCAGAAAATCGCATATGCACCAATTGCTGTGCAGGTCGTA
AGGGTTGCAACTATTACAGTGCTGACGGGACCTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCC
AAAACCTTGCACTCTGAACTGTGATCCAAGAATTTTCTATTCAAAATATCCACGTTCCGAAGGAAACG
CAGAAAATCGCATATGCACCAATTGTTGTGCAGGCCGTAAGGGTTGCTACTATTACAGTGATGATGG
GACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAACCTTGCCCTCGGAATTGTGATACA
AGAATTGCCTATTCAATATGTCCACGTTCCGAAGGAAACGCAGAAAATCGCATATGCACCAATTGCT
GTGCAGGCCGTAAGGGTTGCAACTATTACAGTGCTGACGGGACTTTCATTTGTGAAGGAGAGTCTGA
CCCCAACAACCCAAAAGCTTGCCCTAGGGATTGTGATCCAAATATTGCCTATTCACTGTGTCTCTATG
AAAAGTAA

CanPI-51 [JX106501]

ATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTTCCTACTTGTTCTTGGAATATTGCTTCTACATGTT
GATGCCAAGGCTTGTTCAGAAGAAAACGCAGAAAATCGCATATGCACCAATTGCTGTGCAGGTCGTA
AGGGTTGCAACTATTACAGTGCTGACGGGACCTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCC
AAAACCTTGCACTCTGAACTGTGATCCAAGAATTTTCTATTCAAAATGTCCACGTTCCGAAGGAAACG
CAGAAAATCGCATATGCACCAATTGTTGTGCAGGCCGTAAGGGTTGCTACTATTACAGTGATGATGG

GACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAAGCTTGCCCTCGGAATTGTGATACA
AGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAACGCAGAAAATCGCCTATGCACCAATTGCT
GTGCAGGCCGTAAGGGTTGCAACTATTACAGTGCTGATGGGACTTTCATTTGTGAAGGAGAGTCTGA
CCCCAACAACCCAAAAGCTTGCCCTAGGAATTGTGATCCAAATATTGCCTATTCACTGTGTCTCTATG
AAAAGTAAGAGCATGAACAGTGTCATGCTAATCTCTGTAATACGCATGGGCGTCTATATGAAAAGTA
A

CanPI-52 [JX106502]

ATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTTCCTACTTGTTCTTGAAATATTGCTTCTACATGTT
GATGCCAAGGCTTGTTCAGAAGAAAACGCAGAAAATCGCATATGCACCAATTGCTGTGCAGGCCGTA
AGGGTTGCAACTATTACAGTGCTGATGGGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCC
AAAAGCTTGCCCTCGGTATTGTGATACAAGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAAC
GCAGAAAATCGCATATGCACCAATTGCTGTGCAGGCCGTAAGGGTTGCAACTATTACAGCGCTGACG
GGACTTTCATTTGTGAAGGAGAGTCTGGCCCCAACAACCCAAAAGCTTGCCCTCGGAATTGTGATAC
AAGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAACGCAGAAAATCGCCTATGCACCAATTGC
TGTGCAGGCCGTAAGGGTTGCAACTATTACAGCGCTGACGGGACTTTCATTTGTGAAGGAGAGTCTG
ACCCCAACAACCCAAAAGCTTGCCCTAGGAATTGTGATCCAAATATTGCCTATTCACTGTGTCTCCAT
GAAAAGTAA

CanPI-53 [JX106503]

ATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTTCCTACTTGTTCTTGAAATATTGCTTCTACATGTT
GATGCCAAGGCTTGTTCAGAAGAAAACGCAGAAAATCGCATATGCACCAATTGCTGTGCAGGCCGTA
AGGGTTGCAACTATTACAGTGCTGATGGGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCC
AAAAGCTTGCCCTCGGTATTGTGATACAAGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAAC
GCAGAAAATCGCATATGCACCAATTGCTGTGCAGGCCGTAAGGGTTGCAACTATTTCAGCGCTGACG
GGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAAGCTTGCCCTCGGAATTGTGATAC
AAGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAACGCAGAAAATCGCCTATGCACCAATTGC
TGTGCAGGCCGTAAGGGTTGCAACTATTACAGCGCTGACGGGACTTTCATTTGTGAAGGAGAGTCTG
ACCCCAACAACCCAAAAGCTTGCCCTAGGAATTGTGATCCAAATATTGCCTATTCACTGTGTCTCTAT
GAAAAGTAA

CanPI-54 [JX106504]

ATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTTCCTACTTGTTCTTGAAATATTGCTTCTACATGTT
GATGCCAAGGCTTGTTCAGAAGAAAACGCAGAAAATCGCATATGCACCAATTGCTGTGCAGGCCGTA
AGGGTTGCAACTATTACAGTGCTGATGGGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCC
AAAAGCTTGCCCTCGGTATTGTGATACAAGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAAC
GCAGAAAATCGCATATGCACCAATTGCTGTGCAGGCCGTAAGGGTTGCAACTATTTCTGCGCTGACG
GGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAAGCTTGCCCTCGGAATTGTGATAC
AAGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAACGCAGAAAATCGCCTATGCACCAATTGC
TGTGCAGGCCGTAAGGGTTGCAACTATTACAGCGCTGACGGGACTCTCATTTGTGAAGGAGAGTCTG
ACCCCAACAACCCAAAAGCTTGCCCTAGGAATTGTGATCCAAATATTGCCTATTCACTGTGTCCTCTA
TGAAAAGTAA

CanPI-55 [JX106505]

ATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTTCGTACTTGTTCTTGAAATATTGCTTGTACATGTA
GATGCCAAGGCGTGTTCAGAAGAAAACGCAGAAAATGGCATATGCACCAATTGCTGTGCAGGCCGTA
AGGGTTGCAACTATTACAGTGCGGATGGGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCC
AAAAGCTTGCCCTCGGTATTGTGATACAAGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAGGC
GCAGAAAATCGCCTATGCACCAATTGCTGTGCAGGCCGTAAGGGTTGCAACTATTACAGCGCTGACG
GGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAAGCTTGCCCTCGGAATTGTGATAC
AAGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAACGCAGAAAATCGCCTATGCACCAATTGC
TGTGCAGGCCGTAAGGGTTGCAACTATTACAGCGCTGACGGGACTTTCATTTGTGAAGGAGAGTCTG
ACCCCAACAACCCAAAAGCTTGCCCTAGGAATTGTGATCCAAATATTGCCTATTCACTGTGTCTCTAT
GAAAAGTAA

CanPI-56 [JX106506]

ATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTTCGTACTTGTTCTTGGAATATTGCTTCTACATGTT
GATGCCAAGGCTTGTTCAGAAGAAAACGCAGAAAATCGCATATGCACCAATTGCTGTGCAGGTCGTA
AGGGTTGCAACTATTACAGTGCTGACGGGACCTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCC
AAAACCTTGCACTCTGAACTGTGATCCAAGAATTTTCTATTCAAAATGTCCACGTTCCGAAGGAAACG
CAGAAAATCGCATATGCACCAATTGCTGTGCAGGCCGTAAGGGTTGCAACTATTACAGCGCTGACGG
GACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAAGCTTGCCCTCGGAATTGTGATACA
AGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAACGCAGAAAATCGCCTATGCACCAATTGCT
GTGCAGGCCGTAAGGGTTGCAACTATTACAGCGCTGACGGGACTTTCATTTGTGAAGGAGAGTCTGA
CCCCAACAACCCAAAAGCTTGCCCTAGGAATTGTGATCCAAATATTGCCTATTCACTGTGTCTCTATG
AAAAGTAA

CanPI-57 [JX106507]

ATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTCCCTACTTGTACTTGGAATATTGCTTCTACATGTT
GATGCCAAGGCTTGTTCAGAAGAAAACGCAGAAAATCGCATATGCACCAATTGCTGTGCAGGTCGTA
AGGGTTGCAACTATTACAGTGCTGACGGGACCTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCC
AAAACCTTGCACTCTGAACTGTGATCCAAGAATTTTCTATTCAAAATGTCCACGTTCCGAAGGAAACG
CAGAAAATCGCATATGCACCAATTGTTGTGCAGGCCGTAAGGGTTGCTACTATTACAGTGATGATGG
GACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAACCTTGCCCTCGGAATTGTGATACA
AGAATTGCCTATTCAATATGTCCACGTTCCGAAGGAAACGCAGAAAATCGCATATGCACCAATTGCT
GTGCAGGCCGTAAGGGTTGCAACTATTACAGTGCTGACGGGACTTTCATTTGTGAAGGAGAGTCTGA
CCCCAACAACCCAAAAGCTTGCCCTAGGAATTGTGATCCAAATATTGCCTATTCACTGTGTCTCTATG
AAAAATAA

CanPI-58 [JX106508]

ATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTTCGTACTTGTTCTTGGAATATTGCTTNTACATGTT
GATGCCAAGGCTTGTTCAGAAGAAAACGCAGAAAATCGCATATGCACCAATTGCTGTGCAGGTCGTA
AGGGTTGCAACTATTACAGTGCTGACGGGACCTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCC
AAAACCTTGCACTCTGAACTGTGATCCAAGAATTTTCTATTCAAAATGTCCACGTTCCGAAGGAAACG
CAGAAAATCGCATATGCACCAATTGTTGTGCAGGCCGTAAGGGTTGCTACTATTACAGTGATGATGG
GACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAACCTTGCCCTCGGAATTGTGATACA
AGAATTGCCTATTCAATATGTCCACGTTCCGAAGGAAACGCAGAAAATCGCATATGCACCAATTGCT
GTGCAGGCCGTAAGGGTTGCAACTATTACAGTGCTGACGGGACTTTCATTTGTGAAGGAGAGTCTGA
CCCCAACAACCCAAAAGCTTGCCCTAGGAATTGTGATCCAAATATTGCCTATTCACTGTGTCTCTATG
AAAAGTAA

CanPI-59 [JX106509]

ATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTCCCTACTTGTACTTGGAATATTGCTTCTACATGTT
GATGCCAAGGCTTGTTCACAAAGAAACGCAAAAGAACCCATATGCACCAATTGTTGTGCAGGCCGTA
AGGGTTGCAACTATTACAGTGCTGATGGGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCC
AAAACCTTGTACTTTGAACTGTGATCCAAGAATTTTCTATTCAAAGTGTCCACGTTCCGAAGGAAACG
CAGAAAATCGCATATGCACCAATTGCTGTGCAGGCCGTAAGGGTTGCAACTATTACAGTGCTGATGG
GACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAAGCTTGCCCTCGGTATTGTGATACAA
GAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAACGCAGAAAATCGCCTATGCACCAATTGCTG
TGCAGGCCGTAAGGGTTGCAACTATTACAGCGCTGACGGGACTTTCATTTGTGAAGGAGAGTCTGAC
CCCAACAACCCAAAAGCTTGCCCTAGGAATTGTGATCCAAATATTGCCTATTCACTGTGTCTCTATGA
AAAGTAA

CanPI-60 [JX106510]

ATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTCCCTACTTGTACTTGGAATATTGCTTCTACATGTT
GATGCCAAGGCTTGTTCACAAAGAAACGCAAAAGAACCCATATGCACCAATTGTTGTGCAGGCCGTA
AGGGTTGCAACTATTACAGTGCTGATGGGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCC
AAAACCTTGTACTTTGAACTGTGATCCAAGAATTTTCTATTCAAAGTGTCCACGTTCCGAAGGAAACG
CAGAAAATCGCATATGCACCAATTGCTGTGCAGGCCGTAAGGGTTGCAACTATTACAGTGCTGACGG

GACTTTCATTTGTGAAGGAGAGTCTGACCCCAACCACCCAAAAGCTTGCCCTCGGAATTGTGATCCAA
GAATTTTCTATTCAAAGTGTCCACGTTCCGAAGGAAACGCAGAAAATCGCATATGCACCAATTGCTGT
GCAGGCCGTAAGGGTTGCAACTATTACAGTGCTGACGGGACTTTCATTTGTGAAGGAGAGTCTGACC
CCAACAACCCAAAAGCTTGCCCTAGGAATTGTGATCCAAATATTGCCTATTCACTGTGTCTCTATGAA
AAGTAA

CanPI-61 [JX106511]

ATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTTCGTACTTGTTCTTGGAATATTGCTTCTACATGTT
GATGCCAAGGCTTGTTCAGAAGAAAACGCAGAAAATCGCATATGCACCAATTGCTGTGCAGGTCGTA
AGGGTTGCAACTATTACAGTGCTGACGGGACCTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCC
AAAACCTTGCACTCTGAACTGTGATCCAAGAATTTTCTATTCAAAATGTCCACGTTCCGAAGGAAACG
CAGAAAATCGCATATGCACCAATTGTTGTGCAGGCCGTAAGGGTTGCTACTATTACAGTGATGATGG
GACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAAGCTTGCCCTCGGAATTGTGATACA
AGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAACGCAGAAAATCGCCTATGCACCAATTGCT
GTGCAGGCCGTAAGGGTTGCAACTATTACAGCGCTGACGGGACTTTCATTTGTGAAGGAGAGTCTGA
CCCCAACAACCCAAAAGCTTGCCCTAGGAATTGTGATCCAAATATTGCCTATTCACTGTGTCTCTATG
AAAAGTAA

CanPI-62 [JX106512]

ATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTTCCTACTTGTTCTTGAAATATTGCTTCTACATGTT
GATGCCAAGGCTTGTTCAGAAGAAAACGCAGAAAATCGCATATGCACCAATTGCTGTGCAGGCCGTA
AGGGTTGCAACTATTACAGTGCTGATGGGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCC
AAAAGCTTGCCCTCGGTATTGTGATACAAGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAAC
GCAGAAAATCGCCTATGCACCAATTGCTGTGCAGGCCGTAAGGGTTGCAACTATTACAGCGCTGACG
GGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAAGCTTGCCCTCGGAATTGTGATAC
AAGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAACGCAGAAAATCGCCTATGCACCAATTGC
TGTGCAGGCCGTAGGGGTTGCAACTATTACAGCGCTGACGGGACTTTCATTTGTGAAGGAGAGTCTG
ACCCCAACAACCCAAAAGCTTGCCCTCGGAATTGTGATCCAAATATTGCCTATTCACTGTGTCTCTAT
GAAAAGTAA

CanPI-63 [JX106513]

ATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTTCCTACTTGTTCTTGAAATATTGCTTCTACATGTT
GATGCCAAGGCTTGTTCAGAAGAAAACGCAGAAAATCGCATATGCACCAATTGCTGTGCAGGCCGTA
AGGGTTGCAACTATTACAGTGCTGATGGGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCC
AAAAGCTTGCCCTCGGTATTGTGATACAAGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAAC
GCAGAAAATCGCATATGCACCAATTGCTGTGCAGGCCGTAAGGGTTGCAACTATTACAGCGCTGACG
GGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAAGCTTGCCCTCGGAATTGTGATAC
AAGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAACGCAGAAAATCGCCTATGCACCAATTGC
TGTGCAGGCCGTAAGGGTTGCAACTATTACAGCGCTGACGGGACTTTCATTTGTGAAGGAGAGTCTG
ACCCCAACAACCCAAAAGCTTGCCCTAGGAATTGTGATCCAAATATTGCCTATTCACTGTGTCTCTAT
GAAAAGTAA

CanPI-64 [JX106514]

ATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTTCcTACTTGTTcTTGAAATATTGcTTcTACATGTTGA
TGCCAAGGCTTGTTCAGAAGAAAACGCAGAAAATCGCATATGCACCAATTgCTGTGCAGgCCGTAAG
GGTTGCAACTATTACAGTGCTGATGGGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCaAA
AGCTTGCCCTCGGTATTGTGATACAAGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAACGCA
GAAAATCGCATATGCACCAATTGCTGTGCAGGCCGTAAGGGTTGCAACTATTACAGCGCTGACGGGA
CTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAAGCTTGCCCTCGGAATTGTGATACAAG
AATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAACGCAGAAAATCGCCTATGCACCAATTGCTGT
GCAGGCCGTAAGGGTTGCAACTATTACAGCGCTGACGGGACTTTCATTTGTGAAGGAGAGTCTGACC
CCAACAACCCAAAAGCTTGCCCTAGGAATTGTGATCCAAATATTGCCTATTCACTGTGTCTCTATCAA
AAGTAA

CanPI-65 [JX106515]

ATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTTCGTACTTGTTCTTGAAATATTGCTTCTACATGTT
GATGCCAAGGCTTGTTCAGAAGAAAACGCAGAAAATCGCATATGCACCAATTGCTGTGCAGGCCGTA
AGGGTTGCAACTATTACAGTGCTGATGGGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCC
AAAAGCTTGCCCTCGGTATTGTGATACAAGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAAC
GCAGAAAATCGCATATGCACCAATTGCTGTGCAGGCCGTAAGGGTTGCAACTATTACAGCGCTGACG
GGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAAGCTTGCCCTCGGAATTGTGATAC
AAGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAACGCAGAAAATCGCATATGCACCAATTGC
TGTGCAGGCCGTAAGGGTTGCAACTATTACAGCGCTGACGGGACTTTCATTTGTGAAGGAGAGTCTG
ACCCCAACAACCCAAAAGCTTGCCCTAGGAATTGTGATCCAAATATTGCCTATTCACTGTGTCTCTAT
GAAAAGTAA

CanPI-66 [JX106516]

ATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTTCCTACTTGTTCTTGAAATATTGCTTCTACATGTT
GATGCCAAGGCTTGTTCAGAAGAAAACGCAGAAAATCGCATATGCACCAATTGCTGTGCAGGTCGTA
AGGGTTGCAACTATTACAGTGCTGACGGGACCTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCC
AAAACCTTGCACTCTGAACTGTGATCCAAGAATTTTCTATTCAAAATGTCCACGTTCCGAAGGAAACG
CAGAAAATCGCATATGCACCAATTGTTGTGCAGGCCGTAAGGGTTGCTACTATTACAGTGATGATGG
GACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAACCTTGCCCTCGGAATTGTGATACA
AGAATTGCCTATTCAATATGTCCACGTTCCGAAGGAAACGCAGAAAATCGCATATGCACCAATTGCT
GTGCAGGCCGTAAGGGTTGCAACTATTACAGTGCTGACGGGACTTTCATTTGTGAAGGAGAGTCTGA
CCCCAACAACCCAAAAGCTTGCCCTAGGAATTGTGATCCAAATATTGCCTATTCACTGTGTCTCTATG
AAAAGTAA

CanPI-67 [JX106517]

ATGGCTGTTCCCAAAGAAGTTAGTTTCCTTGCTTTCCTACTTGTTCTTGAAATATTGCTTCTACATGTT
GATGCCAAGGCTTGTTCAGAAGAAAACGCAGAAAATCGCATATGCACCAATTGCTGTGCAGGCCGTA
AGGGTTGCAACTATTACAGTGCTGATGGGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCC
AAAAGCTTGCCCTCGGTATTGTGATACAAGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAAC
GCAGAAAATCGCATATGCACCAATTGCTGTGCAGGCCGTAAGGGTTGCAACTATTACAGCGCTGACG
GGACTTTCATTTGTGAAGGAGAGTCTGACCCCAACAACCCAAAAGCTTGCCCTCGGTATTGTGATACA
AGAATTGCCTATTCAAAATGTCCACGTTCCGAAGGAAACGCAGAAAATCGCCTATGCACCAATTGCT
GTGCAGGCCGTAAGGGTTGCAACTATTACAGCGCTGACGGGACTTTCATTTGTGAAGGAGAGTCTGA
CCCCAACCACCCAAAAGCTTGCCCTAGGAATTGTGATCCAAATATTGCCTATTCACTGTGTCTCTATG
AAAAGTAA

Manasi Mishra

Mobile No. +91 8888696497
Div. of Biochemical Sciences
Email: mr.mishra@ncl.res.in
NCL, Pune-411008
manasimishra1@gmail.com

Academic Qualifications

Master of Science (M.Sc.)	Botany, 2004, First Class, Banaras Hindu University, Varanasi, India
Bachelor of Science (B.Sc.)	Botany (Hons.), 2001, First Class, Banaras Hindu University, Varanasi, India

Other Qualifications

- International English Language Testing System [IELTS] 2008, Score: 7.5
- National Eligibility Test for Junior Research Fellowship (NET, JRF) (June 2006), conducted by Council of Scientific and Industrial Research and University Grant Commission (CSIR-UGC), Government of India.
- National Eligibility Test for Junior Research Fellowship (NET, JRF) (December 2006), conducted by Council of Scientific and Industrial Research and University Grant Commission (CSIR-UGC), Government of India.
- Nomination for Shyama Prasad Mukherjee Fellowship (SPM) by CSIR, New Delhi through CSIR (June, 2006) & CSIR (December, 2006).
- Graduate Aptitude Test for Engineering (GATE) 2006 in Life Sciences; Percentile Score: 95.8

Honors/Awards

- Financial assistance from International travel support scheme (ITS) of Department of science and technology (DST) for participation (poster presentation) in International conference in Amsterdam, Netherlands.
- Short-Term Scholarship (June-November 2010) from German Academic Exchange (DAAD) for carrying out part of the Ph.D. work with Dr. Jochen Muller Dieckmann at the European Molecular Biology Laboratory (EMBL), Hamburg, Germany.
- Junior Research Fellowship (2007-2012) 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 "Biotic stress induced *CanPI* diversity in *Capsicum annuum*" presented at National Science Day, 2010 at the National Chemical Laboratory, Pune.

- Best Poster Award for the poster "Stability and efficiency of diverse rCanPIs from *Capsicum annuum* in presence of *Helicoverpa armigera* gut proteases" presented at National Science Day, 2009 at the National Chemical Laboratory, Pune.

- Ranked second in the Biology Course Work examination held in December, 2007 as a part of Ph.D. studies at the National Chemical Laboratory, Pune.

Publications

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

2- **Mishra M**, Mahajan N, Tamhane VA, Kulkarni MJ, Baldwin IT, Gupta VS and Giri AP (2012). Functional implication of induced proteinase inhibitor diversity in *Capsicum annuum.* (under revision: **BMC Plant Biology**)

3- **Mishra M,** Joshi RS, Gaikwad S, Dieckmann JM, Gupta VS and Giri AP. Structural-functional insights on single and multi-domain proteinase inhibitors from *Capsicum annuum*. (Under preparation)

4- Book Chapter: Gurjar G, **Mishra M**, Kotkar H, Upasani M, Pradeep Kumar, Tamhane V, Kadoo N, Giri A and Gupta V (2011). 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.

5- Review: Tamhane VA, **Mishra M**, Mahajan N, Gupta VS and Giri AP (2012). Plant Pin-II family protease inhibitors: Structural and functional diversity. **Functional Plant science and Biotechnology** (Accepted)

6- Joshi RS, **Mishra M**, Tamhane VA, Ghosh A, Sonavane U, Joshi R, Gupta VS and Giri AP (2012). Reduced disulphide bonding in Pin-II PIs provides flexibility for effective protease inhibition. (under review: **PLoS ONE)**

7- Mahajan N, **Mishra M**, Tamhane VA, Gupta VS and Giri AP (2012) Early exposure and optimal dosage of proteinase inhibitors are essential for effective antibiosis against *H. armigera.* (Under preparation)

Conferences/Workshops

- Poster presentation: Mishra M, Tamhane VA, Khandelwal N, Giri AP, Kulkarni MJ and Gupta VS (2011) "Interaction of recombinant CanPIs with *Helicoverpa armigera* gut proteinases reveals their processing patterns,

stability and efficiency" In: Sixth International Symposium on Molecular Insect Science held at NH Grand Krasnapolsky, Amsterdam, The Netherlands during 2-5 October, 2011.

- Attended "RNA-2010" A National Symposium of the RNA group of India, Department of Biotechnology, University of Pune, India

- Poster presentation: Mishra M, Tamhane VA, Khandelwal N, Giri AP, Kulkarni MJ and Gupta VS (2009) "Stability and efficiency of diverse rCanPIs from *Capsicum annuum* against *Helicoverpa armigera* gut proteases" In: The Sixth Solanaceae Genome Workshop held at Hotel Le Meridien, New Delhi during November 8-13, 2009.

- Poster presentation: Mishra M, Tamhane VA, Khandelwal N, Giri AP, Kulkarni MJ and Gupta VS (2009) "Stability and efficiency of diverse rCanPIs from *Capsicum annuum* against *Helicoverpa armigera* gut proteases" In: International Symposium on: Mass Spectrometry in Life Sciences.

Personal details

Date of Birth:	07/07/1984
Marital Status:	Single
Nationality:	Indian
Place of Birth:	Gorakhpur, India