Chapter IV: Evaluating synthetic and natural protease inhibitors against H. armigera proteases

1. Designing, synthesis and evaluation of synthetic peptide as potent insect gut protease inhibitor

2. Way towards “Dietary Pesticides”- Mechanistic insight into insecticidal action of natural phenols
Chapter IV: Evaluating synthetic and natural protease inhibitors against *H. armigera* proteases

IV.1 Designing, synthesis and evaluation of synthetic peptide as potent insect gut protease inhibitor

Reactive site region of 387 Pin-II PIs was used to design tri-peptides for its potential application as insect protease inhibitor. Out of 23 tripeptides 6 peptides (TLN, PRN, TRE, PKN, PLN, PRY) were predominantly distributed (81%) among the Pin II PIs population, and also showed strong binding score against various insect proteases. These peptides were further synthesized on solid phase platform using F-Moc chemistry. Furthermore, these peptides were characterized and assessed for their *in vitro* protease inhibition potential. Most of the selected peptides showed significant protease activity inhibition; while TRE was most potent protease inhibitor with Ki~24µM. Feeding bioassay showed that 50 to 200 ppm of TRE had significant negative impact on growth and survival of *H. armigera*. Docking analysis of TRE and other RSL tripeptides against various insect protease exhibited variation in the molecular interaction. Differential molecular responses at translational and transcriptional level were observed in TRE fed larvae. These results shed light on the potential of small tailored peptides for further selective and multi-targeted inhibitor design for effective pest control.

IV. 1.1 INTRODUCTION

Basic structure of IRD consists of 50 aa including eight conserved cysteine residues. Reactive site loop (RSL) is part of IRD that interacts with target proteases and it is found to be highly variable. Co-evolution of RSL with their target protease indicates its crucial role in plant-insect interaction (Jongsma & Beekwilder, 2011). Variation in RSL results in deviation in their affinity and specificity toward target protease, thus this information can be explored to tailor small peptide inhibitors against *H. armigera* gut proteases. In second chapter of this thesis, we have showed effect of structural and functional variations in three IRDs namely, IRD-7, -9 and -12 resulted in differential antibiosis effect on *H. armigera* growth. These findings indicate that
IRD-9 exhibits enhanced protease inhibition due to lack of disulfide bond and flexibility in reactive loop as compared other IRDs.

Here, we have in silico screened 23 RSL peptide and analysed for their occurrence frequency. Six peptides were selected, synthesized by F-Moc chemistry on solid phase. Different inhibitory kinetics and in vitro studies of synthesized peptides was performed to understand inhibitory specificities of these inhibitors against trypsin, chymotrypsin and HGPs. In vivo efficacy of these tripeptides was analysed by monitoring growth performance and nutritional parameters. Molecular response of insect digestive proteinases after ingestion of peptide was evaluated by proteinase gene expression, activity and zymography studies. This report demonstrates the approach of exploring RSL sequence variations for designing a potent inhibitor for effective control of insect pests.

**IV. 1.2 MATERIALS AND METHODS**

**IV.1.2.1 Materials**

The protected amino acids were purchased from NovaBiochem (Merck Milipore, Darmstadt, Germany). All solvents used during the synthesis of peptides were of peptide synthesis grade and for HPLC, of HPLC grade. Enzymes and substrates were obtained from Sigma Chemical Co., St. Louis, MO, USA. X-ray films and Kodak163 DA developer were purchased from Kodak (Chennai, India). Highly pure chemicals for AD (AD), natural diets and the rest of the insect rearing materials of highest purity were purchased locally. *H. armigera* larvae were maintained in laboratory as described in earlier section **II.2.1**.

**IV.1.2.2 Synthesis of peptide**

Our interest in synthesis of these peptides is exploration of RSL variation for obtaining effective protease inhibitor, thus we have eliminated the cysteine residues from both end and only three central residues of RSL were considered for further synthesis. The tripeptides obtained from MEME and occurrence frequency analysis were synthesized on MBHA resin as the solid support (1.75 mmol/g). The synthesis
was carried out manually using 3 equivalent of Fmoc-protected amino acids, and HOBr and TBTU as the coupling agents, in the presence of N,N-Diisopropylethylamine. Successive deprotection and coupling steps were carried out as iterative cycles until the desired tripeptide was synthesized. The deprotection of the Fmoc-protected amino group and the coupling reaction were monitored by the Kaiser test. The N-terminal amino group, after Fmoc-deprotection, was capped using acetic anhydride in the presence of dry pyridine. The synthesized tripeptides were cleaved from the solid support using TFA-TFMSA to yield tripeptides with N-terminal acetates and C-terminal amides. They were purified by RP-HPLC using an increasing gradient of acetonitrile in water containing 0.1% TFA, and characterized by MALDI-TOF mass spectrometric analysis.

IV.1.2.3 Motif and distribution analysis

To characterize reactive loop region, we first investigated features represented by conserved regions or motifs. From several algorithms available, we chose the expectation maximization method MEME (Bailey et al., 2009). For MEME, a fixed minimum motif length of 5 and a maximum of 10 was set and 20 motifs were requested using the zero or one occurrence per sequence model. The results obtained with MEME were further used for distribution as well as frequency analysis. The obtained reactive loop motifs with their occurrence frequency in Pin II PIs population were plotted in bar graph.

IV.1.2.4 Virtual screening

Structures for predominant tripeptides were predicted using Chimera software. The structure was energy minimized for 10000 steps with steepest gradient. Structures of *H. armigera* trypsins, chymotrypsins, cathepsins were predicted using CPH model server. A docking study was performed to determine the binding energy and interaction of tripeptides with *H. armigera* proteases. Predicted structures of proteases were refined by energy minimization and restraint relaxation using Swiss PDB-Viewer (v4.1.0) (Guex et al., 1997). In order to perform molecular docking,
models of *H. armigera* proteases and inhibitors were submitted to Patchdock online server (http://bioinfo3d.cs.tau.ac.il/PatchDock/) following the standard package protocols and further refined by FireDock online server (Schneidman-Duhovny *et al.*, 2005). Binding energy obtained for each complex was normalized by mean values and represented in heat map format using MeV software packages (http://www.tm4.org/mev/). The gradient ruler from -5 to 5 is an indicator of interaction strength. Data was clustered using hierarchical clustering method (Saeed *et al.*, 2006).

**IV.1.2.5 Inhibition kinetics**

Concentration dependent reduction in trypsin activity by synthesized tripeptides was also estimated through a BApNA assay. Detailed methodology is given in section II.3.4.

**IV.1.2.6 Feeding assay**

Bioassays were conducted by feeding *H. armigera* larvae on AD containing PIs (Tamhane *et al.*, 2005). AD was supplemented with the tripeptides in 50 to 200 ppm/g (25 to 100 µM) of AD. The neonates obtained lab-reared moths were reared for the first 2 days on control diet and then transferred to peptide-containing. Detailed methodology is given in section II.3.7.

**IV.1.2.7 Assessment of nutritional parameters**

Various nutritional parameters of fourth-instar larvae exposed to control and tripeptide-containing diet were compared. Detailed methodology is given in section II.3.8.

**IV.1.2.8 In vivo inhibition of protease activity**

Inhibition of total gut protease, trypsin and chymotrypsin activities of larvae fed on 50 to 200 ppm tripeptide were determined using azocasein and other specific
substrates. Assay procedures were already discussed in earlier sections of thesis (section II.3.4). Minimum three replicates of each experiment were performed.

**IV.1.2.9 Semi-quantitative PCR**

Total RNA was isolated from the TRE-fed insect gut tissues using Trizol reagent (Invitrogen, Carlsbad, CA, USA) and synthesis of the first strand cDNA was carried out with High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Foster, CA, USA) using random primers based on the manufacturer's protocol. Detailed methodology is given in section II.3.10.

**IV.1.2.10 Statistical analysis**

All data were statistically analyzed by independent sample t-test. Asterisks indicate significant differences (* $p<0.05$; ** $p<0.01$).

**IV.1.3 RESULTS AND DISCUSSION**

**IV.1.3.1 Identification of conserved reactive site loop motifs**

To identify reactive site loop motifs in 389 Pin II PIs, we first searched for conserved in regions of 5 residues with cysteines at both terminals. The search was carried using MEME. Twenty-three motifs correspond to RSL were retrieved per PIs sequence. A comprehensive list and sequence of the all motifs is provided in Fig.4.1.

![Fig. 4.1: Occurrence frequency of 23 RSL variants in Pin II PIs population. RSL marked with red block constitute 81%, while remaining 17 RSL contribute to 19% of total population.](image)
Occurrence frequency analysis of 23 unique RSL in 389 Pin II PIs showed that 6 RSL with sequence CTLNC, CPRNC, CPRYC, CPLNC, CPKNC and CTREC were found to be predominant with 81% of total population. These analyses showed that RSL region in Pin II PIs is most prone to natural variations and thus engaged in generation of diversity of PIs against various target proteases. Occurrence analysis indicated that nature promotes specific sequence to propagate in population might be due to their superior functional attributes.

**IV.1.3.2 In vitro assay indicates inhibition of trypsin activity by RSL tripeptides**

Six tripeptides with sequence TLN, PRN, PRY, PLN, PKN and TRE were synthesized and purified on RP-HPLC. They were characterized by MALDI-TOF mass spectrometric analysis. The synthesized tripeptides along with their MALDI-TOF mass characterization data are listed in **Table 4.1**.

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Peptide sequence</th>
<th>MALDI-TOF mass (Da or m/z)</th>
<th>Calculated</th>
<th>Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Ac-TRE-NH₂</td>
<td>445.23</td>
<td>446.89</td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>Ac-TLN-NH₂</td>
<td>387.21</td>
<td>388.12</td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>Ac-PRY-NH₂</td>
<td>475.25</td>
<td>476.66</td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>Ac-PKN-NH₂</td>
<td>398.23</td>
<td>400.55</td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>Ac-PRN-NH₂</td>
<td>426.23</td>
<td>426.99</td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td>Ac-PLN-NH₂</td>
<td>383.21</td>
<td>384.32</td>
<td></td>
</tr>
</tbody>
</table>

Inhibition kinetic studies displayed a sigmoidal pattern with increasing concentrations of the tripeptides suggesting reversible and competitive inhibition with tight binding. TRE turned out to be a potent inhibitor of bovine trypsin (IC₅₀ ~24.05 µM) compared to other 5 tripeptides from this study (Fig. 4.2). The inhibition constant Ki determined directly from IC₅₀ by using the Cheng-Prusoff’s equation. Ki values for all 6 tripeptides were enlisted in **Table 4.2**. Differences in inhibition kinetics of tripeptides exhibited that amino acid variation of RSL tripeptide account for its differential protease binding and inhibition efficiency.
These results suggest that inhibitory property of RSL of Pin II PIs is sequence and conformation dependent.

<table>
<thead>
<tr>
<th>RSL Tripeptide</th>
<th>~Ki (in µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRE</td>
<td>24.05</td>
</tr>
<tr>
<td>PRN</td>
<td>126.66</td>
</tr>
<tr>
<td>PRY</td>
<td>203.12</td>
</tr>
<tr>
<td>PLN</td>
<td>326.96</td>
</tr>
<tr>
<td>PKN</td>
<td>327.32</td>
</tr>
<tr>
<td>TRE</td>
<td>60.9</td>
</tr>
</tbody>
</table>

**IV.1.3.3 TRE exhibit higher binding affinity and broader specificity against various insect proteases**

Docking and relative analysis displayed significant differences in binding energies suggesting that synthesized tripeptides had variable interaction with *H. armigera* proteases. Among the 6 tripeptides, TRE and PRY showed strong interaction with the lowest binding energy with various *H. armigera* proteases supporting our *in vitro* results (Fig. 4.3). Docking studies revealed broad specificity of TRE with *H. armigera* serine proteases. Strong binding of TRE with trypsin and chymotrypsin among all the tripeptides motivated us to access it’s *in vivo* effect on *H. armiger* digestive physiology. Interaction pattern of TRE and PRY with most of the proteases was similar, which led to their clustering for all the analyzed proteases. Binding energy comparison and hierarchical clustering analysis provides wide overview of specific interaction of inhibitor with various proteases. Furthermore, this analysis might give us global overview about effect tripeptide ingestion on *H. armigera* digestive physiology.
Chapter IV: Synthetic and natural PIs against H. armigera proteases

IV.1.3.4 TRE exerts negative impact on the growth and development of H. armiger larvae

To understand the in vivo effect of TRE on the development of H. armiger larvae, feeding experiments were conducted with appropriate controls. Various concentrations (50 to 200 ppm, 25 to 100 µM) of TRE were incorporated into diet to examine their in vivo potential against H. armigera. Development of larvae reared on a control and TRE containing diets is presented in Fig. 4.4A. Feeding of insects on TRE-containing diet caused concentration dependent reduction in larval mass gain and survival rate. On day 11, larvae fed on diets TRE containing 50, 100 and 200 ppm weighed ~10, 15 and 25 % less, than the larvae fed on AD.

Fig 4.3: Interaction energy of binding of all three IRDs with H. armigera proteases is compared by using Heatmap analysis. Heat-map with hierarchical clustering of relative free binding energy (obtained from docking study) normalized by mean values of cumulative free energy obtained from H. armigera serine proteases binding with 6 RSL tripeptides. The gradient ruler is an indicator of interaction strength, where blue colour indicates weak binding and red indicate strong binding.

Fig 4.4: Growth (A) average size of larvae recorded on day 10 (B) and (C) Survival rate of H. armigera larvae raised on AD containing 50 to 200 ppm of TRE. Eggs were hatched, and neonates were transferred to AD containing TRE.
In comparison, larvae fed on AD containing TRE were ~50 to 60% smaller in body size than control larvae (Fig. 4.4B). Furthermore, larvae fed on inhibitor containing diet showed significant ($p$≤0.05) reduction in survival rate. At day 11, there was 30 to 40% reduction in survival rate of larvae fed on 100 and 200 ppm TRE containing AD as compared to control larvae (Fig. 4.4C). TRE fed larvae displays early and sharp decrease in larval survival rates, followed by partial recovery as the feeding period extends. This is might be due to expression of PIs insensitive proteases and overexpression of proteases, which might help insect to overcome the lethal and detrimental effect of inhibitors (Dunse et al., 2010; de Oliveira et al., 2013).

Evaluation of nutritional parameter like Efficiency of Conversion of Ingested Food (ECI), Efficiency of Conversion of Digested Food (ECD) and Approximate Digestibility (Ad) revealed that the ingestion of TRE have deleterious effect on growth and rudimentary metabolism of the insect (Table 4.3). There was direct correlation in the inhibitory potential and reduction in ECI, ECD and Ad. Assessment of these parameters showed that TRE negatively affected the digestive physiology of insect and thus impedes insect growth and survival (Table 4.3). Inhibition of serine protease activities also obstructs normal developmental pathways leading to delay in pupation and molting, which was also evident from data (Fig. 4.4B). Our results indicate that TRE could serve as potent inhibitor molecules against gut proteases from H. armigera.

Table 4.3: Effect of IRDs ingestion on the H. armigera feeding behavior and dietary utilization.

<table>
<thead>
<tr>
<th></th>
<th>50 ppm</th>
<th>100 ppm</th>
<th>200 ppm</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECI</td>
<td>12.31 (±1.5)</td>
<td>8.35 (±1.12)</td>
<td>6.86 (±1.82)</td>
</tr>
<tr>
<td>ECD</td>
<td>29.09 (±2.31)</td>
<td>18.89 (±1.87)</td>
<td>12.42 (±2.3)</td>
</tr>
<tr>
<td>Ad</td>
<td>56.15 (±1.05)</td>
<td>46.55 (±1.36)</td>
<td>40.05 (±1.5)</td>
</tr>
</tbody>
</table>

# Abbreviations: ECI = Efficiency of Conversion of Ingested Food; ECD = Efficiency of Conversion of Digested Food; Ad = Approximate Digestibility.
IV. 1.3.5 In vivo inhibition of gut proteinases in *H. armiger* larvae reared on TRE containing diet

In comparison with control HGP activity, HGP of larvae fed on TRE showed concentration dependent reduction. Larvae fed on 50 and 100 ppm TRE showed ~50 to 70% reduction of total protease activity. Larvae fed on 200 ppm of TRE showed significant inhibition (~80%) of HGP activity as compared to control larvae (Fig. 4.5). In case of trypsin and chymotrypsin activity of HGP from larvae fed on TRE showed ~60 to 80% and 50 to 70% reduction, respectively. Protease activity provides quantitative account about *in vivo* proteases activity upon TRE ingestion.

IV.1.3.6 Ingestion of TRE triggers the differential gene expression of the gut proteases

Expression analysis of eight major gut trypsin genes further provides the quantitative evidence of altered digestive physiology of *H. armigera* (Fig. 4.6A). Real time PCR analysis showed that trypsins namely HaTry4, 5 and 6 were found to be downregulated in TRE fed larvae. Downregulation of these trypsins might also correlate with drop in proteolytic activity of numerous HGP isoforms when countered with TRE. HaTry1 and 8 were highly upregulated in TRE fed larvae. HaTry2, 3 and 7 were exclusively upregulated in 200 ppm TRE fed larvae. Overexpression of protease transcripts could help the insect to overcome the
inhibitory effect of PIs, which might in turn help in growth and development of insects.

Among four major chymotrypsins, three transcripts HaChy1, 2 and 3 showed down regulation in TRE-fed larvae (Fig. 4.6B). While HaChy4 showed distinct behavior with overexpression in 50 and 100 ppm TRE fed larvae. Expression dynamics of gut protease genes in an attempt to overcome the effect of inhibition and to acquire optimum nutritional requirements for growth and development provides an excellent survival tactics for the insects.

Fig. 4.6: Fold difference in transcript abundance of major digestive trypsins and chymotrypsins for larvae fed on AD containing 50 to 200 ppm of TRE on the 10th day after feeding. (A) Fold difference in transcript abundance of trypsin genes (B) Fold difference in transcript abundance of chymotrypsin genes. Most of trypsin and chymotrypsin gene showed down regulation; while few of them showed upregulation in TRE fed insects tissues.

**IV.1.4 CONCLUSION**

Exposure of RSL tripeptide ‘TRE’ to neonatal stage of larvae exhibited negative impact on *H. armigera* growth and development. This investigation provided insight
in to the potential of RSL amino acid sequence variations on efficacy and specificity against target proteases. Exposure of larvae to TRE offered variable molecular response and resulted in alteration in digestive physiology, particularly in protease expression. Current study suggested small peptides with 3 aa i.e. TRE proved to be effective growth inhibitors to *H. armigera* larvae. This phenomenon of antibiosis by RSL peptides can be explored for designing effective inhibitor candidate against *H. armigera*.
IV.2 Way towards “Dietary pesticides”- Mechanistic insight into insecticidal action of natural phenols

Insect infestation on various crops leads to reduce food productivity. Existing pest management strategies have raised serious environmental issues thereby resulting into the emergence of pesticide resistance in insects. Present investigation provides mechanistic insights into insecticidal activity of a natural phenol, caffeic acid (CA). In silico and subsequent in vitro screening of several natural phenols indicated CA as an effective inhibitor of *H. armigera* gut serine proteases. Furthermore, CA was found to be responsible for in vivo inhibition of the gut protease activity that led to reduced growth, survival and development of the insects. Upon CA ingestion dynamic alternation in protease expression and activity were apparent in *H. armigera* larvae. In addition, structure-activity relationship of CA highlighted the significance of all functional groups for its potency against target proteases. Various biophysical evidences suggested that binding of CA caused conformational changes in the target enzymes and thus decrease enzyme activity. Furthermore, molecular dynamic simulations and isothermal titration calorimetry results revealed that binding of first CA molecule at active site of trypsin provides a nucleation center for sequential binding of multiple CA and thus disrupts the function. In addition, along with the inhibition of digestive activity, CA showed significant reduction in detoxifying enzyme activities upon ingestion, intensifying the detrimental effect on overall insect physiology. In conclusion, our findings suggest that the natural phenols especially CA can be implicated as highly potent insecticidal.
IV.2.1 INTRODUCTION

Modern agricultural practices are dependent on the extensive use of chemical pesticides which leads to resistance and resurgence in insects and also a reason for substantial environmental and human health problems (Cheng, 1990; Abdollahi et al., 2004). This scenario highlights a need of natural and eco-friendly pesticides. Plants own diverse pool of secondary metabolites possessing insecticidal activity could provide lead molecules for developing ecofriendly insect pest management strategies. Recently, remarkable interest in potential of plants species with insecticidal activity, initiates application of these molecules in biopesticides development. Monomeric and polymeric phenolic compounds constitute one of the most widespread groups in plant secondary metabolites and they are generated biogenetically from the shikimate-phenylpropanoids-flavonoids pathways (Lattanzio et al., 2008). Plants produce phenolic compounds for pigmentation, growth, reproduction and to cope with multiple stresses, etc. Pivotal role of natural phenols in plant-environment interactions especially in plants defense is evidenced from their differential accumulation in response to abiotic stresses (Borejan et al., 2003; Asakawa et al., 2013; Bi et al., 1995; Horvath et al., 2007; Ingersoll et al., 2010; Mehmood et al., 2013). Furthermore, toxic or deterrent activity of phenols against insect pests facilitates the direct defense; while in case of indirect defense they attract natural enemies of insect pests (Green et al., 2003; Magalhaes et al., 2010; Rani et al., 2013; War et al., 2012; War et al., 2013). In phenolic acids, hydroxybenzoic and hydroxycinnamic acid (HCA) serve as basic scaffold for evolution of population of various compounds (Borejan et al., 2003).

In this study we investigated the potential of natural phenols as insect protease inhibitor and molecule(s) as dietary pesticide for insect pest control in agriculture. Protease inhibition potential of selected molecules was evaluated by in silico study, in vitro inhibition kinetics and feeding bioassay. Various growth and development parameters were assessed in support of antibiosis of these natural phenols to H. armigera. Biophysical characterization and multi-scale molecular
dynamic simulations have been performed to elucidate binding mode of natural phenols to proteases and resultant structural changes in vital insect proteins.

IV.2.2 MATERIALS AND METHODS

IV.2.2.1 Virtual screening of natural phenols as protease inhibitor

Structures of various hydroxycinnamic acid derivatives were obtained from PubChem database and they were optimized for their 3D coordinates using Marvin Sketch Tool (http://www.chemaxon.com). Three dimensional structures of bovine trypsin (BtTry; PDB ID: 4I8G) and chymotrypsin (BtChy; PDB ID: 1YPH) were accessed from RCSB PDB. Also, *H. armigera* trypsin (HaTry) and chymotrypsin (HaChy) structures were modeled by using CPH 3.0 model server and processed using SwissPDB viewer (Kaplan et al., 2001). Conversion of receptor and ligand from .pdb to .pdbqt format, also grid map and other docking parameters were set using AutoDock 4.2 software (Morris et al., 2009). Virtual screening was carried out using AutoDock Vina software and the Lamarckian genetic algorithm as a searching procedure (Trott et al., 2010). Binding energy obtained for each complex was represented in heat map format using MeV software packages (http://www.tm4.org/mev/). The gradient ruler is an indicator of interaction strength. Represented molecules showing strong binding against all the selected targets, represented by lower free energy were selected for further *in vitro* and simulation study.

IV.2.2.2 Inhibition kinetics study

Concentration dependent reduction in trypsin activity by CA was also estimated through a BApNA assay using chromogenic substrate Benzoyl-L-arginyl p-nitroanilide (BApNA). BApNA assays were performed as described previously (section II.3.4).

IV.2.2.3 Feeding assay and assessment of nutritional parameters

Bioassays were conducted by feeding *H. armiger* larvae on *AD* containing 5 to 200 ppm/g of CA. Detailed method is as described previously (section II.3.7 and II.3.8).
The insect growth, development and survival of insect fed on control and CA containing diet were analyzed on each alternate day for 10 days.

**IV.2.2.4 Semi-quantitative gene expression analysis**

Total RNA was extracted from whole larval body using Trizol reagent (Invitrogen, CA, USA), followed by RQ1 DNase treatment (Promega, Fitchburg, Wisconsin, USA). CDNA was synthesized from 2μg of the DNA-free RNA samples by reverse transcription using oligo dT primers and reverse transcriptase (Promega) following the manufacturer's recommendations. Primers pairs were designed for Acetylcholinesterase, Amylase, Chymotrypsin, CYP450, GSTs and Trypsin genes of *H. armigera*. cDNA was diluted (1: 10) before use in a PCR. Semi-quantitative RT-PCR performed under the following conditions: initial denaturation at 95 °C for 2 min; 30 cycles at 9 5°C for 30 s; 60 °C for 30 s; 72 °C for 30 s and a final extension at 72 °C for 5 min (Kotkar *et al.*, 2012).

**IV.2.2.5 Fluorescence quenching assay**

Fluorometric experiments were carried out on a fluorescence spectrometer (Thermo Scientific, Waltham, Massachusetts, USA). Stock solutions of CA (1 mM) in ethanol and Bovine Trypsin (1 mM) in 100 mM Tris-HCl (pH 7.8) was prepared at room temperature. The final concentration of CA was from 0.2 to 30.0 μM with a constant bovine trypsin content of 1 μM. The fluorescence spectra were recorded at λ<sub>exc</sub> = 280 nm and λ<sub>em</sub> from 300 to 500 nm (Jiang *et al.*, 2004; Kang *et al.*, 2004; Bian *et al.*, 2007; Jin *et al.*, 2012).

**IV.2.2.6 MALDI-TOF MS analysis**

Different molar concentration of CA and trypsin were mixed and incubated for 20 min at 37 °C. The mass spectral analysis of the reaction mix was done on Q-TOF-MALDI-TOF-MS with a standard instrumental protocol (ABSCIIX). Sample preparation, spectral acquisition and processing were done as described earlier (Mishra *et al.*, 2013).

**IV.2.2.7 Isothermal titration calorimetry (ITC)**

ITC was performed using Microcal Auto-iTC instrument (GE Healthcare, Buckinghamshire, UK). 19 injections of 2 μl bovine trypsin (Stock =2.0 mM) was
titrated against 0.1 mM solution of CA. Experiments were carried out at 37 °C in a Tris buffer, pH 7.8. Reference titration was carried out by injecting the same concentration of trypsin into buffer. Reference titration was subtracted from experimental titration. Origin 6.0 software was used to derive affinity constants (Kd), the molar reaction enthalpy (CH) and the stoichiometry of binding (N), by fitting the integrated titration peaks (Weber et al., 2003).

### IV.2.2.8 Circular dichroism spectroscopy

A Jasco J-810 spectro-polarimeter (Jasco, Easton, MD, USA) running the software Jasco J815CD was used to collect spectra at wavelengths 190-240 nm (far-UV) and 240-300 nm (near-UV). Quartz cuvettes of path lengths 5 mm was used for spectral scan. The scanning speed was 20nm/min, the response was 4 s, and the bandwidth was 1 nm. Trypsin (10 µM) was incubated with CA in 1:2, 1:1 and 2:1 ratio for 4 h at room temperature, similar process was done for chymotrypsin. Baseline corrections were performed by subtracting the spectra of the buffer and phenolic compounds from the sample spectra. The data are reported as mean residue ellipticity ([θ]) in deg²cm⁻² 3dmol⁻¹. Details about spectra acquisition and analysis were described previously in section III.1.2.5.

### IV.2.2.9 Molecular dynamic simulation

The structures with 27 CA molecules (aggregation study) and bovine trypsin-CA complex were relaxed using all-atom MD simulation in explicit water with GROMACS software package using the AMBER99SB force field and the simple point charge (SPC) water model (Ghosh et al., 2012). The time step used in the simulation was 0.002 ps while list of neighbors was updated every 0.01 ps with the grid method and a cut off radius of 14 Å. The coordinates of all atoms in the simulation were saved every 2 ps. The initial velocities were chosen randomly. NPT ensemble was used with the user-defined box dimension. The temperature and pressure were kept at the desired value using the Berendsen method and an isotropic coupling for the pressure (T=300 K, ζT=0.1 ps, P0=1 bar, coupling time = ζT=1 ps). The electrostatic term was calculated using the Particle Mesh Ewald (PME) algorithm with the radius of 16 Å with Fast Fourier Transform (FTT) optimization (with an order of 4 for the
cubic interpolation). The cut off algorithm was applied for the non-coulomb potential with a radius of 9 Å. Initial minimization was done for a total of 50,000 steps, followed by volume and pressure equilibration each for 1 ns. The production MD run for the complexes were then carried out.

Both of these systems were then simulated using the simulation parameters mentioned in the previous section. The parameters for CA were retrieved from the SWISSPARAM server (http://www.swissparam.ch/; Zoete et al., 2011). Both of these simulations were performed and analyzed using the supercomputing facility provided by BRAF cluster using GROMACS (Ghosh et al., 2012).

**IV.2.2.10 Enzymatic assays**

*H. armigera* larvae (500 mg) were ground to a fine powder and extracted in 500 mL of 200 mM Glycine-NaOH buffer (pH 10) for 2 h at 4°C. The extract was then centrifuged at 12000 X g (4°C; 10 min) and the clear supernatant was used as a source of enzyme for all solution assays. Control and CA fed insects protein lysate (50 µg) was used to carry out the following functional assays (a) total protease activity (b) Trypsin-like activity (c) Chymotrypsin like activity (d) Amylase activity (e) Lipase activity (f) Acetylcholinesterase activity (g) GSTs activity and (h) CYP450 activity (Kotkar et al., 2012; Sarate et al., 2012; Ellman et al., 1961; Dwakar et al., 2011). The detailed methodology is described in corresponding references.

**IV.2.2.11. Statistical analysis**

All the experiments were performed three times independently. Student’s t-test was used for statistical analysis. Data were expressed as mean ± SD. A p < 0.05 was considered as statistically significant.

**IV.2.3 RESULTS AND DISCUSSION**

**IV.2.3.1 Protease inhibition potential of hydroxycinnamic acid derivatives**

Comparison of binding energies of various HCA derivatives against bovine and *H. armigera* proteases displayed variability in binding affinity (Fig.4.7A). Caffeic, ferulic, sinapic and p-coumaric acid emerged as strong inhibitors of candidate
proteases (Fig. 4.7B). It was observed that caffeic acid (CA) showed strong competitive inhibition of bovine (Bos taurus) trypsin (Ki ~22 µM) and bovine chymotrypsin (Ki~32 µM), while ferulic acid (FA) showed exclusive inhibition of bovine chymotrypsin with Ki ~32 µM. CA and FA showed significant inhibition of H. armigera trypsin and chymotrypsin activity (Fig. 4.7B).

Molecular docking analysis and binding energy comparison indicated that natural phenols have affinity towards the insect proteases and the interaction amongst the proteases from different insects is conserved (Polyphagous: H. armigera and Plutella xylostella; Monophagous: Manduca sexta) (Fig. 4.7C and D). Analysis of 130 docked complexes revealed that CA had strong interaction (low binding energy) with various insect serine proteases, especially with trypsin like proteases. Elevated concentration of CA and other similar natural phenols in plant tissues upon insect infestation might have direct role in inhibition of digestive function of the insects (Green et al., 2003; Magalhaes et al., 2010; Rani et al., 2013; War et al., 2012; War et al., 2013).

**IV.2.3.2 Molecular response of H. armigera to caffeic acid (CA) ingestion**

Delayed pupation indicates the retardation of development in larvae fed on CA (Fig.4.8A). CA appeared to have significant and concentration dependent negative impact on growth and survival of H. armigera. Larvae fed on diet containing 50 to 200 ppm of CA showed ~ 20 to 50% less body mass (g) as compared to control (Fig. 4.8B).

Furthermore, CA fed leave had ~ 50 to 80% reductions in survival rate (Fig. 4.8C). Evaluation of nutritional parameters like Efficiency of Conversion of Ingested Food (ECI), Efficiency of Conversion of Digested Food (ECD) and Approximate Digestibility (AD) revealed that the ingestion of CA had deleterious effect on growth and rudimentary metabolism of the insect (Fig. 4.8D). Protease activity of insect fed on CA exhibit substantial reduction in trypsin like (~ 20 to 60%) and thus total protease (~15 to 50%) activity (Fig. 4.8E). Inhibition of digestive proteases by CA probably causes starvation in insects. It might led to nutritional scarcity, less pool of free amino acids and also energy required for metabolism and causes the growth
retardation followed by insect death (Green et al., 2003; Magalhaes et al., 2010; War et al., 2012; Rani et al., 2013; War et al., 2013).

Candidate protease gene expression analysis indicated that ingestion of CA also had differential response. Several of the *H. armigera* trypsin’s and chymotrypsin’s (HaTry1 to 8; HaChy1 to 4) were down regulated except HaTry4
Overexpression of HaTry4 and HaChy4 in CA fed insects might have been resulted in response to compensate for reduced the protease activity. This overexpression might be also in response to rescuing the insect from growth retardation by producing inhibitor resistant/insensitive proteases, which could hijack the insect digestive system and fulfill the insect growth requirement (Bown et al., 1997; Broadway et al., 1997; Chikate et al., 2013).

![Fig. 4.8](image)

**Fig. 4.8** *H. armigera* larvae were raised on AD containing 5 to 200 ppm/g of CA (A) Average size of larvae recorded on day 10 (B) Insect growth and (C) Survival rate normalized with control (D) Nutritional parameters ECI, ECD and Ad and (E) In vivo proteolytic activity of insect fed on control and CA containing diet were assessed. Bars represent means (±) S.E. from three independent experiments at *p* 0.05 (F) Protease gene expression analysis at the transcript level by performing semi-quantitative RT-PCR using total RNA isolated from control and CA fed larvae. All samples were analyzed on 2% agarose gels containing gel red. This analysis showed differential expression of protease gene in CA fed larvae.

### IV.2.3.3 Fate of CA in insect gut environment

Stability of CA in insect gut environment and probability of formation of subsequent metabolites were investigated by LC-ESI(-)-HRMS analysis (Fig. 4.9). Chromatograms of gut metabolite extract of CA fed larvae in 1:1 acetonitrile and
methanol showed the presence of intact CA molecule and no additional peaks in comparison to control. Thus, it excluded the possibility of the formation of degraded products of CA in digestive track of insect. Stability of CA in insect gut led to maximum availability and hence caused to be constitutively active as protease inhibitor. *In vivo* stability of CA made it a lucrative molecule to be used as an insecticidal agent (Yang *et al.*, 2013).

**Fig. 4.9**: HRMS analysis of insect gut metabolites indicated the fate of CA in insect gut environment.

### IV.2.3.4 Structure-activity relationship of CA as protease inhibitor

Functional group-activity relationships provide a detailed insight into the essentiality of individual functional groups to maintain the inhibitory potency *(Table 4.4)*. Semi-synthesized derivatives were structurally characterized by NMR and mass spectrometric data which were in well agreement with the previous reports (Barontini *et al.*, 2014; Percec *et al.*, 2006; Roche *et al.*, 2005; Takahashi *et al.*, 2010; Uwai *et al.*, 2008; Zhu *et al.*, 2010). Masking of 3-hydroxyl group as methoxy on the phenolic ring (FA, 1b) resulted in the considerable reduction in gut protease inhibition (40 to 60%). Even the addition of another methoxy group at 5-position (1c, sinapic acid) on the phenolic ring led to further reduction in inhibitory
potencies (20 to 30%). Removal of the 3-hydroxyl group (1d, cinnamic acid) from the CA structure resulted in drastic reduction in the inhibitory activity (15 to 25%). Masking of the terminal carboxylic acid as a methyl (1e) ester also caused a significant lowering in inhibitory activities (35 to 50%). Further increment in the chain length as an ethyl ester (1f) showed far more reduction in inhibitory potency (25 to 30%). Hydrogenation (reduction) of the double bond in the side chain of CA (1g) reduced the activity extensively (12 to 20%). Similar effect was observed when methyl ester of CA was subjected to hydrogenation (1h) (inhibition 25 to 50%). Protection of both the hydroxyl groups on the phenolic ring by acetylation (1i and 1j) drastically attenuated its inhibitory potential (10 to 20%). In brief, modification in any of the functional groups (phenolic hydroxyl, carboxylic acid and double bond) on CA structure led to reduction in the inhibitory activity against proteases with variable extent and all of them were found to be essentially responsible for the activity. Therefore, the parent natural product (1a) found to be most potent among all the derivatives screened.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Inhibition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Trypsin ($\sim IC_{50}$ in µM)</td>
</tr>
<tr>
<td>1a</td>
<td>22</td>
</tr>
<tr>
<td>1b</td>
<td>120</td>
</tr>
<tr>
<td>1c</td>
<td>650</td>
</tr>
<tr>
<td>1d</td>
<td>780</td>
</tr>
<tr>
<td>1e</td>
<td>80</td>
</tr>
</tbody>
</table>

**Table 4.4:** Various derivatives of CA and their inhibition values against Trypsin, Chymotrypsin and HGP
IV.2.3.5 Binding of CA causes structural and conformational change in target protein

Bovine trypsin showed reduction in tryptophan fluorescence in response to titration with CA suggesting the change of the molecular environment of tryptophan (Fig. 4.10A). Red shift is observed in this case indicated the conformational change in protease on ligand binding. Furthermore, it was found that multiple CA molecules were interacting non-covalently with proteases. It was validated by mass spectrometric analysis, where shift in mass (m/z) was observed in comparison to native enzymes due to binding of multiple CA molecules to the proteases (Fig. 4.10B). Circular Dichorism (CD) spectroscopy of proteins in near and far UV regions suggested the secondary and tertiary structure alteration in the protease respectively, on CA binding (Fig. 4.10C and D). Fluorescence quenching and CD Spectroscopy analysis together indicated that binding of CA caused environmental alternation around tryptophan residue resulting in structural and conformational change in the protease, which might further be attributed to the reduced enzyme activity (Jiang et al., 2004; Kang et al., 2004; Bian et al., 2007; Jin et al., 2012).
Chapter IV: Synthetic and natural PIs against H. armigera proteases

IV.2.3.6 CA shows sequential binding to protease and exhibit heterogeneous mode of inhibition

In set of simulation system to study the impact of CA binding on the structure of protease, system containing H. armigera trypsin 4 (HaTry4, predominant trypsin from H. armigera gut) with multiple CA in vicinity was simulated for 1µS and trajectory was analyzed. The CA molecules in this simulation were initially restrained at discrete position on the protease molecule and the restrain was released till 10 ns of the simulation. Out of these 5 CA molecules, one was found to remain in active site for the rest of simulation, and interacting with Ser200 of HaTry4 (Fig. 4.11A). Hydroxyl groups of CA can form multiple transient interactions with protease active site, however, H-bond between OH group of CA and Tyr45 and Ser200 was found to be stable throughout simulation time (Fig. 4.11B). Further, after around 208 and 300 ns of the simulation, two more CA molecules were found to be located near mouth of active site (Fig. 4.11C). These three CA clubbed near active site to form stacked triad via Pi-Pi interactions and
maintains its position through another set of Pi-Pi interaction with lining residues of active site mouth i.e. Tyr45, Trp144 and Tyr154 (Fig. 4.11D). In this way CA binding to H. armigera trypsin might cause heterogeneous type of inhibition by (i) competitively binding to the active site residue i.e. Ser200 (ii) blocking the mouth of the active site, leads to inaccessibility to the substrate. This suggests that CA might be having some crucial interactions with the active site residues and longer time-scale simulation might be required to explore major structural changes in the protein.

Thermodynamic studies of trypsin-CA interaction was carried out using ITC as it is one of the most widely used quantitative technique for direct measurement of the enthalpy change when two species interact, allowing the determination of heat of association, stoichiometry, and binding affinity from a single experiment. The raw data and corresponding to the thermogram of the binding experiment is depicted in Fig. 4.11E. Binding was strongly exothermic and showed 1:3 stoichiometry for trypsin and CA. The spontaneity of the process was evidenced by a negative change in the enthalpy, CH, and a positive change in the entropy, CS. Binding of first CA molecule showed lower ΔG as compared to the second and third molecule binding. This thermodynamic pattern of binding indicates that binding of first CA molecule was strong as compared to the second and third CA molecule.

To investigate the aggregation propensity of CA in solution and its effect on protein binding, 27 CA molecules (aggregation study) were relaxed using all-atom MD simulation in explicit water with GROMACS software package using the AMBER99SB force field and the simple point charge (SPC) water model (Ghosh et al., 2012). The 50 ns trajectory of this system was analyzed to see how individual CA molecules display the phenomenon of aggregation. At around 1 to 2 ns of simulation, formation of aggregates was initiated (Fig. 4.11F). These aggregates were observed in different clusters with varying number of CA molecules in individual clusters. Initially after 2 ns of the simulation, two major clusters were observed having 12 CA molecules each and remaining 3 CA in none of them.
Likewise a big cluster of 21 CA molecules was observed after 6.5 ns and just after 30 ns of the simulation, a cluster of 27 CA molecules was also observed.

**Fig.4.11:** Simulation of bovine trypsin and CA in one system illustrates (A) binding of single CA molecule at the active site residue SER200. (B) 2D interaction map showed establishment of multiple interactions of CA with trypsin binding pocket. Strong binding of CA at active site leads to (C) formation of nucleation center for aggregation of other two CA molecules in the active site. (D) Aggregate formation in binding pocket further blocks complete access to the active site, which is showed in surface model. (E) Simulation of only CA in solvation box showed that CA has intrinsic propensity to form aggregate and this might enhance its chances to interact with the proteases.

Some of these aggregated CA molecules also formed stacked orientation with one another during the simulation. This suggested that the CA had an intrinsic
property of displaying stacking interactions as well as self-aggregation. Formation of aggregates might enhance the probability of CA to interact with proteases. Kinetic, Simulation and ITC data demonstrated the binding of single molecule of CA to the active site of trypsin initiates the sequential binding of CA molecules at the active site.

IV.2.3.7 CA inhibits the insect detoxification machinery

Estimation of different enzyme activities and semi-quantitative gene expression analysis indicated that other than protease inhibition activity, CA also had adverse effects on various enzymes primarily involved in detoxification mechanism. It was observed that activity of Glutathion S-tranferase (GST) and Cytochrome P450 (CYP450) reduced drastically in insects fed with CA containing diet (Fig. 4.12A), and insect attempted to compensate this activity reduction by overexpression of these genes at transcript level (Fig. 4.12B).

Reduction in these enzyme activities was concentration dependent and almost 60% GST and CYP450 enzyme activity was reduced at 200 ppm concentration. Other than these enzymes, various digestive enzymes activities were found to be maintained in presence of CA even at concentration of 200 ppm. Inhibition of detoxification enzymes by natural phenols might intensify the detrimental effect on insect growth and development.
Fig. 4.12: The enzymatic activities of (A) AChE, Amylase, lipase, GST and CYP450 were analyzed for control and CA treated insects using protocols described in corresponding references. In CA fed insects, significant and concentration dependent reduction in GST and CYP450 activity was observed. Bars represent means (±) S.E. from three independent experiments at *p 0.05 (B) Gene expression analysis of AChE, Amylase, lipase, GST and CYP450 genes at the transcript level by performing semi-quantitative RT-PCR using total RNA isolated from control and CA fed larvae.

IV.2.4 CONCLUSION

Natural phenol, CA has been found out and validated as a potential insecticidal molecule. All the functional groups contributed to its protease inhibitor property and it also remain stable in insect gut environment. Ingestion of CA causes differential molecular responses at transcriptional and translational level. Structural investigation showed that CA showed sequential binding at the active site of major insect protease, thus inhibits its activity. Although overexpression of these proteases (e.g. HaTry4) were observed in CA fed larvae, but strong inhibition potential of CA overcome the effect of enhanced expression of insect proteases. All these finding suggests that exploration of natural phenols could be effective approach to develop "Dietary pesticide" against H. armiger infestation.