The remarkable efficiency of a Pin-II proteinase inhibitor sans two conserved disulfide bonds is due to enhanced flexibility and hydrogen bond density in the reactive site loop

Rakesh S. Joshi, Manasi Mishra, Vaijayanti A. Tamhane, Anirban Ghosh, Uddhavesh Sonavane, C.G. Suresh, Rajendra Joshi, Vidya S. Gupta, and Ashok P. Giri*

*Plant Molecular Biology Unit, Biochemical Sciences Division, CSIR-National Chemical Laboratory, Dr. Homi Bhabha Road, Pune 411 008, MS, India; †Bioinformatics Group, Centre for Development of Advanced Computing (C-DAC), Pune University Campus, Pune 411 007, MS, India; ‡Institute of Bioinformatics and Biotechnology, University of Pune, Pune 411 007, MS, India

Communicated by Ramaswamy H. Sarma

(Received 8 August 2012; final version received 28 October 2012)

 Capsicum annuum (L.) expresses diverse potato type II family proteinase inhibitors comprising of inhibitory repeat domain (IRD) as basic functional unit. Most IRDs contain eight conserved cysteines forming four disulfide bonds, which are indispensible for their stability and activity. We investigated the functional significance of evolutionary variations in IRDs and their role in mediating interaction between the inhibitor and cognate proteinase. Among the 18 IRDs encoded by C. annuum, IRD-7, -9, and -12 were selected for further characterization on the basis of variation in their reactive site loop, number of conserved cysteine residues, and higher theoretical ΔGbind for interaction with Helicoverpa armigera trypsin. Moreover, inhibition kinetics showed that IRD-9, despite loss of some of the disulfide bonds, was a more potent proteinase inhibitor among the three selected IRDs. Molecular dynamic simulations revealed that serine residues in the place of cysteines at seventh and eighth positions of IRD-9 resulted in an increase in the density of intramolecular hydrogen bonds and reactive site loop flexibility. Results of the serine residues chemical modification also supported this observation and provided a possible explanation for the remarkable inhibitory potential of IRD-9. Furthermore, this natural variant among IRDs showed special attributes like stability to proteolysis and synergistic inhibitory effect on other IRDs. It is likely that IRDs have coevolved selective specialization of their structure and function as a response towards specific insect proteases they encountered. Understanding the molecular mechanism of pest protease–plant proteinaceous inhibitor interaction will help in developing effective pest control strategies.

An animated interactive 3D complement (I3DC) is available in Proteopedia at http://proteopedia.org/w/Journal:JBSD:39

Keywords: Capsicum annuum; Helicoverpa armigera; proteases; proteinase inhibitor

Introduction

Plants and insects have coevolved in order to survive in their ever-changing niche. Whereas insects evolved their proteinase enzymes to derive maximum nutritional benefit that plants evolved upregulating the expression of defense related molecules to counter it (Green & Ryan, 1972; Karban, 1989; Kessler & Baldwin, 2002). The agronomically adverse insect pest Helicoverpa armigera (Lepidoptera: Noctuidae) has been widely studied for its polyphagy and adaptability on various host plants. In order to sustain on chemically varied dietary content, insects acquired molecular flexibilities of their gut enzymes (Patankar et al., 2001; Sarate et al., 2012; Zalucki, Daglish, Firempong, & Twine, 1986).

Numerous investigations corroborate the fact that plants rely upon plant proteinase inhibitors (PIs) for protection against attack by phytophagous pests. PIs are ubiquitous in the plant kingdom and have been extensively studied as plant defense molecules, which inhibit hydrolytic enzymes of the insect gut (Green & Ryan, 1972; Ryan, 1990). In some systems, the role of PIs in plant defense is directly evident; whereas in others, the PI diversity and its differential expression may suggest their endogenous functions (Damle, Giri, Sainani, & Gupta, 2005; Johnson, Miller, & Anderson, 2007; Sin & Chye, 2004; Zavala, Patankar, Gase, & Baldwin, 2004; Zavala, Patankar, Gase, Baldwin, & Hui, 2004). Among various PI families, serine PI Pin-II/Pot-II family displays...
a remarkable structural and functional diversity at the gene and protein level (Kong & Ranganathan, 2008). Wound, herbivory, and stress induced upregulation of these PIs clearly link them to plant defense (Green & Ryan, 1972). Previous studies using transgenic systems or in vivo assays have positively correlated the advantage offered by Pin-II PI expression in plants against insect attack (Duan et al., 1996; Johnson, Narvaez, An, & Ryan, 1989; McManus, White, & McGregor, 1994). Recently, a combination of Pin-II PIs from Nicotiana alata and Solana num tuberosum expressed as transgenes in cotton and tested at the field level had proved to enhance the productivity by 30% due to reduction in pest infestation (Dunse, Kaas, Guarino, Barton, & Craik, 2010).

Precursor proteins of Pin-II PIs consist of one to eight inhibitory repeat domains (IRDs) connected by proteolytic-sensitive linkers, which release IRD units upon cleavage. Each IRD is a peptide of around 50 aa length with a molecular mass of ~6 KDa. The aa sequence of IRDs shows variations, at the same time the eight cysteine residues that form disulfide bridge are conserved (Lee, Scanlon, Craik, & Anderson, 1999; Nielsen, Heath, Anderson, & Craik, 1995; Scanlon, Lee, Anderson, & Craik, 1999; Schirra, Scanlon, Lee, Anderson, & Craik, 2001). One structural feature of Pin-II IRD is a disordered loop with triple-stranded β-sheet scaffold. The disordered solvent exposed reactive loop is anchored by the four conserved disulfide bonds (C4–C41, C7–C25, C8–C37, and C14–C50) (Schirra & Craik, 2005; Schirra, Anderson, & Craik, 2008). Among the four disulfide bonds, C8–C37 has been found to be very crucial for maintaining active conformation, whereas C4–C41 has an important role in maintaining the flexibility of the reactive loop (Schirra, Guarino, Anderson, & Craik, 2010). Thus, any selective loss of disulfide bond is expected to have evolutionary significance leading to functional differentiation of inhibitors (Li, Zhang, & Donnelly, 2011).

Each IRD possesses a single reactive site, which inhibits trypsin or chymotrypsin depending on whether a Lys/Arg or a Leu residue is present at the P1 position. In Pin-II precursor, the IRDs are capable of simultaneously inhibiting single or several protease molecules (Barrette et al., 2003a, Barrette, Ng, Cherney, Pearce, & Ryan, 2003b). In a well-known standard mechanism of protease inhibition by Pin-II PIs, the convex-shaped reactive site of the inhibitor (P1 side chain) is recognized by the concave active site (S1 binding pocket) of the enzyme in a substrate like manner and plays a major role in the energetics of recognition (Czapinska & Otlewski, 1999). Protease–PI interaction is an entropy-driven process and is further influenced by noncontact residues of the inhibitor by means of van der Waal’s interaction and hydrogen (H) bonding (Otlewski et al., 2001). The structure of Pin-II IRDs or two domain PIs in complex with protease displays molecular framework of the protease–PI interaction (Barrette et al., 2003a; Dunse et al., 2010). Structural studies on unbound Pin-II inhibitor provide insight into the conformational flexibility of the reactive loop and its role in modulating protease binding efficiency (Barrette et al., 2003b).

Previous studies have shown that Pin-II PIs from C. annuum (CanPIs) and their recombinant forms exhibit antimetabolic effects on H. armigera by inhibiting larval growth and development (Antcheva et al., 2001; Tamhane, Chougule, Giri, Dixit, & Sainani, 2005; Tamhane, Giri, Sainani, & Gupta, 2007). CanPI precursor proteins interact with the gut proteases of H. armigera and get processed into their constituent IRDs (Mishra et al., 2010). Eighteen unique IRDs with aa variations in the reactive loop and/or in the number of cysteine residues have been identified (Tamhane, Giri, Kumar, & Gupta, 2009). Recent findings have led to the emergence of a hypothesis that functional specifications of the Pin-II PIs are closely associated with their sequence and structural variations (Li et al., 2011; Schirra et al., 2010).

For investigating protease–PI interaction and identifying a potent PI, it is important to study inhibition potential of each PI. However, no previous study has specific details about the structure of H. armigera gut trypsin or chymotrypsin or their interaction with CanPIs. Here, we report the recombinant expression, inhibition mechanism, and other biochemical characteristics of CanPI IRD(s). We have also addressed the effect of IRD(s) sequence variation, with special reference to variants with different number of cysteine residues, on the protease–PI interaction by using molecular dynamics (MD) simulation.

**Materials and methods**

**Materials**

All reagents, enzymes and substrates were obtained from Sigma-Aldrich, St Louis, MO. Sterile plastics ware from Abdos, WB, India; expression vector pPIC9 and P. pastoris GS115 from Invitrogen, Carlsbad, CA, USA; Bradford reagent and electrophoresis reagents were from Bio-Rad Laboratories, Hercules, CA; X-ray films and Kodak 163 DA developer were purchased from Kodak, Chennai, India; HIC matrix i.e. Phenyl Sepharose and disposable PD-10 Desalting Columns were from GE Healthcare Life Sciences, Uppsala, Sweden.

**Selection, cloning, expression, and purification of IRD(s)**

Eighteen sequentially unique IRD(s) were identified from 21 CanPI genes, which were reported in our previous study (Tamhane et al., 2009). Phylogenetic analysis of these IRDs was carried out using MEGA5 software (http://www.megasoftware.net/). Depending on sequence analysis IRD-7, -9, and -12 were selected for recombinant protein expression and further characterization. The mature peptide region of selected IRDs were cloned into expression vector pPIC9 for recombinant, extracellular
expression in *P. pastoris* GS115 and purified by hydrophobic interaction chromatography as described previously (Tamhane et al., 2005). The purified proteins were quantified by Bradford reagent and checked for purity on 15% Tricine-SDS-PAGE.

**Inhibition assay and kinetics**

*H. armigera* larvae were reared on artificial diet and complete gut tissue was dissected out from fourth instar larvae. *H. armigera* gut proteases (HGP) were extracted from 2 gm of gut tissue by homogenizing in 0.2 M Glycine-NaOH buffer, pH 10.0 in 1:1 ratios (w/v), and kept at 4°C for 2 h (Tamhane et al., 2005). The suspension was centrifuged at 13,000X g, 4°C for 20 min, and the resulting supernatant was used as a source of HGP. Total proteolytic activity of 50 mM bovine trypsin/HGP and inhibition of their activity by IRDs (5μg) were measured by Azocasein assays. Trypsin-like activity of the HGP was estimated through a Prusoffs classical equation (GXCT) (Pichare & Kachole, 1994). For this, 0.05 HGPI units from each sample were separated on the 15% native-PAGE gel. After electrophoresis, gel was equilibrated with 0.01 M Tris-HCl buffer (pH 7.8) for 10 min followed by incubation in 0.04% trypsin for 10 min and Tris-HCl wash for 2 min. The gel was exposed to X-ray film for the time intervals of 5, 10, and 15 min, respectively. The films were washed with warm water and inhibitory activity bands were visualized as unhydrolyzed gelatin on the X-ray film (Tamhane et al., 2005).

\[
\Delta A_{412} = E_{412}TNB^{2+} \cdot [RSH]
\]

where, \(\Delta A_{412} = A_{f:\text{final}} - (3.1/3.2)(A_{DTNB} - A_{\text{buffer}})\)

\[E_{412}TNB^{2+} = 1.415 \times 104 \text{cm}^{-1}\text{M}^{-1}\]

**Effect of temperature**

Each IRD protein (5 μg) was heated from 60–100°C for 15 min. The treated samples were then used for inhibition assay using BAPNA and trypsin inhibition (TI) was estimated throughout the above mentioned range of temperature.

**Proteolytic stability and HGP inhibition visualization**

To study the interaction and stability of PIs with HGP *in vitro*, 0.5 HGPI units of individual IRDs (IRD-7, -9, and -12) were incubated with 0.5 U HGP for two time points (5 min and 1 h) at 24°C. These HGP-treated PIs were resolved on native-PAGE and processed for TI activity visualization as described above. This mixture of protease and PIs was also used for visualizing the remaining protease activity of HGP in the presence of inhibitor, on 8% native-PAGE using GXCT.

**Combinatorial inhibition assay**

In nature, PIs comprises of different combination of IRDs. IC\(_{50}\) concentrations of each IRD(s) were used to formulate various combinations of IRDs to check their synergistic effect on HGP/bovine TI potential. The inhibition assay was carried out as already described. Four different formulations i.e. IRD-7+9, IRD-7+12, IRD-9+12, and IRD-7+9+12 were used for inhibition assay.

**Molecular dynamic (MD) simulations of IRD-trypsin complex**

The templates for molecular modeling, Nuclear magnetic resonance (NMR) structure of trypsin inhibitor from *N. alata* (PDB ID: 1TIH_A; Figure S1) for IRD (Schirra et al., 2008), and the crystal structure of bovine trypsin...
(PDB ID: 3MH4; Figure S1) for HaTry (UniProt ID: B6CMF9) were selected based on sequence similarities. The 3D models were generated using MODELLER package (version 9.6). All the models were energy minimized using 1000 steps of the conjugate gradient algorithm and short MD simulations, as part of the MODELLER protocol in order to refine the side chain orientations. Fifty models were generated for each sequence, which were ranked according to the GA141 and DOPE scoring functions. The stereo-chemical properties of the final selected models were validated using PROCHECK and ProSA (https://prosa.services.came.sbg.ac.at/prosa.php) analyses. Structural superimposition of IRD-7, IRD-9, and IRD-12 with NaTI was also done using MODELLER and PyMol (The PyMol Molecular Graphics System, Version 1.2r3 pre, Schrodinger LLC) (Ghosh Sonavane, Andhikra, Aradhya, & Joshi, 2011). MODELLER software was used for in silico point mutation and three variants of IRD-9 namely IRD-9A, -9B, and -9C were designed. These variants comprise mutations S7A, S8A in IRD-9A; C28S, C37S in IRD-9B and -9C were designed. These variants comprise mutations S7A, S8A in IRD-9A; C28S, C37S in IRD-9B and -9C, respectively.

Furthermore, each individualIRD was docked against H. armigera trypsin and binding energy (ΔGbind) was calculated for each complex. N. alata trypsin inhibitor (NaTI) was used as a control for correspondence with previous data (Schirra et al., 2010). The data obtained from binding energy calculation were normalized by ΔGbind of (NaTI+HaTry) complex. A heat map for this analysis was constructed with normalized ΔGbind values using TIGR Multi Experiment Viewer (MeV, http://www.tm4.org/mev.html). IRDs showing stronger binding with proteases and possessing large aa variation in reactive loop sequence were selected for further studies. The aa sequence analysis and multiple sequence alignment were performed using DNA star (Laser gene, DNASTAR, Madison, WI, USA) and Clustal X software.

In order to obtain the structure of trypsin-IRD complex, protein–protein docking was carried out with the predicted models of HaTry and IRDs using the rigid-body docking program ZDOCK (version 3.0.1) (Chen, Li, & Weng, 2003). Six sets of protein–protein docking were carried out viz. HaTry_IRD7, HaTry_IRD9, HaTry_IRD-9A, HaTry_IRD-9B, HaTry_IRD-9C, and HaTry_IRD-12. The binding site residues for HaTry and each of the IRDs were specified for docking, to allow the catalytic triad of HaTry (His69, Asp114, and Ser211) to interact with the reactive loop of the IRDs (CPxNC). After the initial docking, the best complex in each case was chosen based on the ZDOCK scores i.e. ZRANK, which is in the range of 15–31 for small proteins of 100 residues.

Explicit MD simulations were carried out for exploring the molecular mechanism of the dynamic interactions, the importance of the interacting residues in binding, and the stability of the disulfide bridges. A set of six simulations was performed, corresponding to the six protease-IRD complexes using the GROMACS 4.0.7 package with GROMACS ffG43a1 force field for 20 ns each. All the six systems were solvated with single-point charge water model and neutralized with proper counterions. All the six systems were then energy minimized using 10,000 steps of the steepest descent algorithm present in the GROMACS package followed by a 100 ps position restraining simulation – restraining the protein by a 1000 kJ/Mol harmonic constraint to relieve the close contacts with water molecules under canonical (NVT) ensemble conditions, where moles (N), volume (V) and temperature (T) are conserved. V-rescale (modified Berendsen) temperature coupler was used to couple the temperature. Another equilibration run under isothermal–isobaric ensemble (NPT) ensemble conditions was performed for 100 ps, before the final production run of 20 ns each for all the systems, where moles (N), pressure (P) and temperature (T) are conserved. V-rescale temperature and Parrinello-Rahman pressure couplers were used to maintain the temperature (293 K) and pressure (1 bar) values with the protein and nonprotein (water and ions) molecules separately coupled with a coupling constant of Tc = 0.1 picoseconds (ps). The isotropic pressure coupling was set with Tp = 2 ps. A time step of 2 femtoseconds (fs) was used throughout with periodic boundary conditions and LINCS constraint algorithm was used to maintain the geometry of the molecules. The van der Waal’s interactions and Coulomb interactions were cutoff at 12 Å with updates every five steps, while long-range electrostatic interactions were calculated using the particle-mesh Ewald method. All the simulations were performed on a PARAM Yuva cluster at the Centre for Development of Advanced Computing at Pune, India, using 64 Intel Xeon 2.93 GHz Quad Core processors. The results were analyzed using the in-built analysis package of GROMACS, XMGRACE (http://plasmagate.weizmann.ac.il/Grace/) and in-house developed scripts (Ghosh et al., 2011). The trajectories were visualized using Visual Molecular Dynamics (VMD) software and all the images were rendered using PyMol. The overall stability of all the simulated systems was also checked with respect to temperature, pressure, and potential energy. All the six simulated systems were in thermodynamic equilibrium during the production simulation runs confirming the convergence of the individual trajectories (Figure S2).

MD simulations were used to determine both the internal and the interaction potential for each complex. Free-energy of binding of the IRDs with the trypsin molecule was calculated by solving the linearized Generalized Born equation available with the MM/GBSA module of AMBER11 package. The MM/GBSA module calculates the binding free energy (ΔGbind) between a receptor (here, trypsin) and a ligand (here, IRD) to form a complex by solving the following equation,
ties of these modiﬁcations remove the excess PMSF from samples. Residual activity was measured thrice using disposable PD-10 Desalting Columns to desalt buffer, of pH 7.8 and 5, 10, and 15 mM of PMSF was added to the reaction mixture containing inhibitor (IC\textsubscript{50}) leaving two cysteine residues (C25 and C14) and proteolytic stability (Bronsoms et al., 2011). A single 3\textsubscript{10}-helix of one turn is also present in the structure; bond stretching, angle bending, and dihedral rotation, \( \Delta E_{\text{internal}} \) represents the internal energies due to bond stretching, angle bending, and dihedral rotation, \( \Delta E_{\text{electrostatic}} \) represents the electrostatic energy, and \( \Delta E_{\text{vdW}} \) represents the van der Waal’s energy. \( \Delta G_{\text{GB}} \) represents the polar solvation energy using GB model and \( \Delta G_{\text{SA}} \) represents the nonpolar solvation energy. For GB calculations, the exterior dielectric constant was set at 80 and 1 was used for the solute dielectric constant. Pairwise GB model implemented in AMBER11 was used for calculation, with parameters described by Tsui and Case (Tsui & Case, 2000).

From the experimental inhibition concentrations IC\textsubscript{50} of three IRD(s), the experimental free energy of binding can be approximated using,

\[
\Delta G_{\text{bind}} = k_B T \ln k_i = k_B T \ln IC_{50} + C
\]

where \( k_B \) is the Boltzmann constant and \( T \) the absolute temperature. Here, we have used the Cheng–Prusoff's equation for competitive inhibition to convert an IC\textsubscript{50} value into an inhibition constant. In case of calculating relative free energy differences, C becomes zero. Calculated binding energy was correlated with experimental binding energy and regression coefﬁcient (\( r^2 \)) was calculated (Xu, Li, Zhou, & Hou, 2012).

**Modification of serine residues with phenylmethylsulfonyl ﬂuoride (PMSF)**

Phenylmethylsulfonyl ﬂuoride (PMSF) speciﬁcally binds irreversibly and covalently to serine and thus blocks its function. The reaction mixture containing inhibitor (IC\textsubscript{50} concentration of each IRD in 1 ml) in 100 mM Tris-HCl buffer, of pH 7.8 and 5, 10, and 15 mM of PMSF was incubated at 30°C for 1 h. Aliquots were removed at different time intervals (15, 30, 45, and 60 min) and desalted thrice using disposable PD-10 Desalting Columns to remove the excess PMSF from samples. Residual activities of these modiﬁed inhibitors were determined under standard assay conditions. Inhibitor sample incubated in the absence of PMSF served as control. Pseudo-ﬁrst-order plot for inactivation of IRDs by PMSF and second order plot of pseudo-ﬁrst-order rate constants (K\text{app}) (min\textsuperscript{-1}) as a function of log of PMSF concentration were plotted using the equations from Koller and Kolattukudy (1982). Inhibitory activity of untreated and PMSF treated (10 mM for 2 h at 30°C) inhibitors were also assessed using GXCT method as described above.

**Results and discussion**

**Sequence and structural variation in IRDs**

Phylogenetic (Figure 1(A)) and multiple sequence alignment (Figure S3) analysis of CanPI IRDs showed signiﬁcant divergence due to sequence variations in the reactive loop regions and in the number of cysteine residues. Heat map provides an overview of the binding energetics of all the 18 IRDs with target proteases. The data indicated that IRD-7, -9, and -12 bind more strongly to HaTry compared to the other IRDs and thus selected for further analysis (Figure 1(B)). The multiple sequence alignment of IRD-7, -9, and -12 with N. alata IRD (NaTI) showed over 90% sequence identity (Figure 1(C)). In case of Pin-II PI family, the major variation is found in the reactive loop (Kong & Ranganathan, 2008). The residues in the reactive loop of IRD-7 and -9 is ‘CPKNC,’ whereas in IRD-12 is ‘CPRNC.’ Another crucial variation is in the number of cysteine residues present. The number of conserved cysteine residues in IRD-7 and -12 is eight while the same in IRD-9 is only six, making the latter one unique among IRDs. Two cysteines at seventh and eighth position of IRD-9 are replaced by serine residues, disrupting two disulﬁde bonds.

PROCHECK analysis showed that the predicted models of IRDs and *H. armigera* trypsin had more than 95% residues in allowed and favored regions of the Ramachandran plot (Figure S4). ProSA analysis had also conﬁrmed the quality of predicted models as good. In accordance with the structure of a typical IRD belonging to Pin-II PI family, the predicted structures also have three anti-parallel β sheets joined by disordered loops containing the reactive site and stabilized by four disulﬁde bonds (Figure 1(D)). It is thought that the disulﬁde bonds act as structural scaffold to hold the reactive site in a relatively rigid conformation and provide thermal and proteolytic stability (Bronsoms et al., 2011). A single 3\textsubscript{10}-helix of one turn is also present in the structure; the disordered loop is held by disulﬁde bond in IRD-7 and -12 whereas by a network of intra molecular hydrogen bonds in IRD-9. Superposition of the predicted structures of IRD-7, -9, and -12 on the template structure of NaTI using Ca atom positions gave a root-mean-square deviation (RMSD) of 1.1, 1.55, and 1.1 Å, respectively. The central scaffold superposed well with the larger deviations conﬁned to solvent exposed surface loops. As expected the predicted structures of IRD-7 and -12 with eight cysteines, had four disulﬁde bonds (C4-C41, C7-C25, C8-C37, and C14-C50), whereas IRD-9 with six cysteines had only two disulﬁde bonds (C4-C41 and C14-C50) leaving two cysteine residues (C25 and C37) free. The predicted and validated *H. armigera* trypsin (HaTry) model has classical trypsin-like fold consisting of two β-barrel domains and the juxtaposed catalytic residues. The catalytic triad in HaTry consists of the residues H69, D114, and S211 (Figure S5).
To assess the effect of aa variations on structural stability, a 20 ns MD simulations was performed on IRD structures. The predicted structures remained stable throughout the 20 ns simulation that was performed under NPT conditions at a temperature of 293 K and 1 bar pressure. Postsimulation analysis of the intramolecular hydrogen bonds illustrated that IRD-9 with two disulﬁde bonds (C7-C25 and C8-C37) less, has a relatively higher density of intra-molecular hydrogen bonds as compared to IRD-7 and -12 (Figure 1(D) and 18 R.S. Joshi et al.
These intramolecular hydrogen bonds might be substituting the two lost disulfide bonds of IRD-9 to stabilize the protein structure in the active conformation and also might be protecting the molecules from a hydrophobic collapse (Hansen et al., 2007). The replaced serine residues in the place of two cysteines C7 and C8 in IRD-9 may be contributing to the increased number of hydrogen bonds. This might be a positive natural selection and led to functional differentiation of the inhibitor (Li et al., 2011). The relative orientations of the secondary structural elements were conserved throughout the entire simulations with the RMSD values based on Ca positions remained below 2 Å for IRD-9 and at about 3 Å for IRD-7 and -12, while trypsin had an RMSD value of around 3 Å with bound IRDs (Figure S6).

Inhibition kinetics and biochemical characterization of IRD-7, -9, and -12
IRD-7, -9, and -12 were extracellularly expressed in Pichia pastoris and the soluble fraction in each case yielded the single protein band in each case corresponding to \(~5.8\) kDa on 15% Tricine-SDS-PAGE (Figure 2(A)). Assays using BApNA and azocasein as substrates showed that IRD-9 and -12 inhibited about 80–85% of HGP activity while inhibition by IRD-7 was only 40–45% (Figure 2(B)). Both the substrates showed low inhibitory efficiency by IRD-7 and highest proficiency by IRD-9.

Furthermore, the kinetic studies displayed a sigmoidal pattern with increasing concentrations of the inhibitors suggesting reversible and competitive inhibition with tight binding. IRD-9 turned out to be a stronger inhibitor of bovine trypsin (IC_{50} \sim 0.0022 mM) than IRD-7 (IC_{50} \sim 0.135 mM) and IRD-12 (IC_{50} \sim 0.065 mM) (Figure 2(C)). The inhibition constant Ki determined directly from IC_{50} by using the Cheng–Prusoff equation also confirmed the same (Table 1). Although the aa and structure variations of IRDs account for their differential binding efficiency, the exact molecular mechanism that contribute to binding efficiency is not understood.

It is known that the disulfide bonds are essential for the folding, function, and stability of IRDs (Schirra et al., 2010). In the present study, the activity of all three IRDs seen on 15% Native-PAGE was lost in the reduced state (Figure 3(A)). Disulfide-rich proteins are also known to show high thermal stability (Bronsoms et al., 2011). Inhibition assays carried out at different temperatures showed that IRD-7 and -12 retained their inhibitory activity against trypsin even at 90 °C for 30 min whereas IRD-9 gradually lost activity starting from 70 °C (Figure 3(B)). The reduced thermal instability of IRD-9 might be due to the decrease in the number of disulfide bonds.

Interestingly, IRD-9 exhibited proteolytic resistance for 60 min when incubated with HGP as compared to IRD-7 and -12, both of which submitted to instantaneous proteolysis (Figure 3(C)). Gut extract of H. armigera, a complex mixture of various trypsin and chymotrypsins like proteases, displayed at least seven isoforms (HGP-1 to -7) (Figure 3(D)). These isoforms of HGP vary in terms of properties and specificity. Interestingly, HGP isoforms were differentially inhibited by various IRDs. The activities of HGP-3 and -4 were inhibited by all IRDs,
whereas that of HGP-5, -6, and -7 were inhibited exclusively by only IRD-9. Protease activity band between HGP-6 and -7 was developed only in the case of IRD-9 treatment and was not present even in untreated HGP, indicating IRD-9 bound protease complex acquiring a different charge state. Thus, IRD-9 presented unique binding property and activity.

The synergistic effect of IRDs was analyzed by performing inhibition assay with combination of different IRDs in IC50 concentration. The presence of IRD-9 in combination with IRD-7 and IRD-12 enhanced their corresponding HGPI activity from 49 to 65% and 51 to 63%, respectively (Figure 3(E)). Results obtained showed that IRD-9 might have a synergistic effect and can lead to higher potentiation of other IRDs. These biochemical evidences support the higher efficiency of varied combination of CanPIs/IRDs in inhibiting insect gut proteases, which signify the biological relevance of sequence variation.

In silico studies indicated that IRD-9 has two free cysteine residues which may be in the form of thiol. This observation is confirmed by Ellman's assay, which estimates free thiol groups in small peptides. In the present study, it showed that 3.8 μM of IRD-9 had ~ 7.9 μM of free thiol (~2 free cysteine residues) whereas a similar amount of IRD-7 and -12 had approximately ~0.155 and ~0.183 μM free thiol content (absence of any free thiol). These results provided additional support to for the in silico predictions that IRD-7 and -12 had four disulfide bonds, whereas IRD-9 had only two leaving two remaining cysteines free.

**Molecular mechanism of IRD(s)-HGP interaction**

The 20 ns MD simulations were used to predict the binding affinities and hence, the inhibitory effects of the individual IRDs against HaTry. The molecular models of the IRD bound HaTry predicted several atomic interactions with a reactive loop of inhibitors that also explained the contribution of the solvent exposed reactive loop. In IRD-9-HaTry interaction, carbonyl oxygen atoms of MET-92, and SER-207 of HaTry active site formed hydrogen bonds with inhibitor side chain of ASN-40 and LYS-39, while side chain of SER-192 from HaTry formed hydrogen bond with side chain of LYS-39. ARG-39 from IRD-12 reactive site loop formed three hydrogen bonds with SER-207 and HIS-50 of the HaTry active site (Figure 4). In case of IRD-7, LYS-39 amide and carbonyl of reactive site loop interact with carboxyl carbonyl oxygen atom of HIS-50. There are additional hydrogen bonds that exist between CYS-37 from reactive site loop of IRD-9 and -12, with carbonyl oxygen atom of ILE-210 or side chain of ARG-109 residue from HaTry. Although the interaction of active site of enzymes with all the three inhibitors were similar in nature, significant differences were observed in making the weak interaction like hydrogen bonding and van der Waal’s interactions, which resulted in differential binding free energy of the complexes. IRD-9 forms the maximum number of stable hydrogen bonds with the active site residues (HIS-50, MEY-92 and SER-192) of the HaTry and which were maintained for longer duration (Figure S7). Although IRD-12 forms relatively more hydrogen bonds, but they are very unstable as reflected by their fluctuating nature. MD simulations provides structural insight into an importance of inter/intra molecular hydrogen bonds and its effect on the interaction between protease and PIs. The results of this analysis were corroborated with previous reports (Hansen et al., 2007). Postsimulation analysis also

| Inhibitor | Variation | Number of disulfide bonds | Number of free cysteine | Number of intramolecular H bonds | Number of intermolecular H bonds | ΔG<sub>binding</sub> (kcal/mol) |
|-----------|-----------|---------------------------|------------------------|-------------------------------|-------------------------------|-----------------------------|
| IRD-7     | Null      | 4                         | 0                      | 10                            | 2                             | -40.03                      |
| IRD-9<sup>a</sup> | C7S, C8S  | 2                         | 2                      | 20                            | 6                             | -68.63                      |
| IRD-9B<sup>b</sup> | S7A, S8A  | 2                         | 2                      | 8                             | 2                             | -39.88                      |
| IRD-9C<sup>b</sup> | C28S, C37S | 2                         | 2                      | 12                            | 5                             | -74.14                      |
| IRD-12    | Null      | 4                         | 0                      | 10                            | 5                             | -54.27                      |

<sup>a</sup>Natural variant.

<sup>b</sup>Signifies in silico variants.
explained experimentally observed increase in binding affinity, hence activity of IRD-9 towards proteases.

Previous reports suggested the role of C4–C41 disulfide bond in maintaining flexibility of the reactive loop and that of C8–C37 in holding a reactive loop of inhibitor in active and stable form. Interestingly in our study, IRD-9 was found to be a good inhibitor although it lacked a C8–C37 disulfide bond. In silico, analysis of a series of mutations at the seventh and the eighth positions could provide insights into the significance of C7S and C8S variation on IRD-9 inhibitory activity. The three variants and the mutations tried are IRD-9A: S7A & S8A; IRD-9B: C28S & C37S and IRD-9C: S7A, S8A, C28A, and C37A. IRD-9A (~1 hydrogen bond) and -9C (~2 hydrogen bond) showed less number of intermolecular and intramolecular hydrogen bonds as compared to IRD-9B (~3 hydrogen bonds), in complex with HaTry.

This analysis showed that replacement of cysteine with a hydrophilic residue, serine can prevent the hydrophobic collapse of the inhibitor molecule and might provide better flexibility and active conformation to the reactive loop and hence enhancement in the inhibitory potential (Schirra et al., 2010).

Calculations of the free energy of binding between IRDs and HaTry (ΔGbind) pointed to a comparatively more stable complex formed by IRD-9 with the lowest ΔG value of −68.63 kcal/mol, as compared to IRD-7 (−40.03 kcal/mol) and IRD-12 (−54.01 kcal/mol), a trend similar to what observed in inhibition assays. The free energy of binding was also calculated for HaTry_IRD-9 variants complexes, in which the binding of IRD-9B (−74.14 kcal/mol) was found more stable as
Figure 4. The modeled *H.armigera* trypsin (gray) in complex with the predicted structures of the IRDs. The important residues at the interface of IRDs and trypsin in complexes of IRD-7 (orange), -9 (cyan), -12 (violet) are shown separately in boxes. The models were obtained using a combination of homology modeling, loop prediction, and MDs. Thin dotted wheat colored lines represent hydrogen bonds. LYS-39H, ASN-40H in IRD-7, -9, and ARG-40H in IRD-12 form a number of important contacts with active site of HaTry.
compared to IRD-9A (−39.88 kcal/mol) and IRD-9C (−38.04 kcal/mol), respectively. This analysis of the variants has provided valuable insight for carrying out potential site directed mutations of IRDs for higher stability and adaptability. There was good correlation between theoretically calculated and experimentally found ΔG values (Table 1). The high correlation coefficient \( r^2 = .97 \) between the calculated and the experimentally determined binding free energies supports our observation (Figure 5). This implies the reliability of the predicted binding conformations and interaction of the inhibitors with HaTry. The higher conformational flexibility of IRD-9 by the loss of two disulfide bonds has helped it to spatially adapt a better complementary shape suited to the active site of HaTry compared to the more rigid four disulfide containing IRD-7 and IRD-12.

**Effect of serine residues modification in inhibition potential of IRDs**

The effect of PMSF on the activity of IRDs is shown in the Figure 6. Reaction of the inhibitor with PMSF leads to modification of one serine residue (number of residues modified were deduced from graph of Log \( K_{app} \) against concentration 5, 10, and 15 mM for 15, 30, 45, and 60 min (A) IRD-7 (B) IRD-9, and (C) IRD-12. Notes: Inset shows corresponding second order plot of pseudo-first order rate constants (K app) (min\(^{-1}\)) as a function of log (PMSF) concentration.
conc. of PMSF) in IRD-9 and resulted in 35–45% activity loss. Modification of IRD-7 and -12 did not show significant effect on the activity (Figure 6(A)). Furthermore, activity visualization assay showed that PMSF modified IRD-9 has reduced inhibition potential as compared to modified IRD-7 and -12 (Figure 6(B)). These results pointed out that, serine could be involved in holding the reactive loop in proper position through a network of hydrogen bonds which was blocked on treatment with PMSF and resulted in loss of inhibitory activity of IRD-9. Thus, result indicated that serine residues were not directly involved in the interaction, but they significantly affect the binding of inhibitor with a protease molecule.

Conclusion
This study employed a combination of experimental and theoretical approaches to investigate the molecular details of HaTry–IRD interaction. Expression and biochemical characterization of IRD-7, -9, and -12 revealed IRD’s sequence-dependant variation of inhibition. Furthermore, IRD-9 lacking two disulfide bonds shows phenomenal inhibition activity compared to other IRDs. This natural variant also exhibits special attributes like stability to proteolysis and inhibitory synergistic effect on other IRDs etc., which make this molecule unique among the members of Pin-II inhibitor family. Explicit MD simulation of protease-inhibitor complex suggests that the loss of disulfide bonds in IRD-9 might be compensated by higher density of intramolecular hydrogen bonds and reactive loop flexibility to bind tightly to target proteases. Chemical modification studies of serine residues combined with MD simulation confirm the role of serine residues that replaced cysteine in increasing the inhibition potential of IRD-9. This insight into interaction mechanisms and role of certain residue changes in unexpected increase of IRD inhibition potential will lead to development of new range of potent PIs for effective pest management strategies.

Acknowledgements
Rakesh S. Joshi, Manasi Mishra, and Vaijayanti A. Tamhane are supported by CSIR research fellowships. The Council of Scientific and Industrial Research (CSIR), Government of India, New Delhi supported this project under network project grants to National Chemical Laboratory, Pune (NWP0003). AG, UBS, and RRJ are grateful to acknowledge the financial support provided by the Department of Information Technology (DIT), Government of India, New Delhi. The authors acknowledge Dr. Mahesh Kulkarni, Dr. Sushma Gaikwad, and Dr. Rajesh Kumar Singh from NCL, Pune for their suggestions in experiment also Prof. M. Edelman, Weizmann Institute of Science, Rehovot, Israel and Prof. M. Levitt, Stanford University, CA, USA for their critical comments. The authors have declared no conflict of interest.

Supplementary material
The supplementary material for this paper is available online at http://dx.doi.org/10.1080/07391102.2012.745378.

References
Aitken, A., & Learmonth, M. (1996). Estimation of disulfide bonds using Ellman’s reagent. In J. M. Walker (Ed.), The protein protocols handbook (pp. 487–488). Totowa, NJ: Human Press.
Antcheva, N., Pintar, A., Pathy, A., Simoncsits, A., Barta, E., Tchorbanov, B., & Pongor, S. (2001). Proteins of circularly permuted sequence present within the same organism, the major serine proteinase inhibitor from Capsicum annuumseeds. Protein Science, 10, 2280–2290.
Barrette-Ng, I. H., Ng, K. K., Cherney, M. M., Pearce, G., Ghanii, U., Ryan, C. A., & James, M. N. (2003a). 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, I. H., Ng, K. K., Cherney, M. M., Pearce, G., & Ryan, C. A. (2003b). 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.
Bronsons, S., Pantoja-Uceda, D., Gabrijelcic-Geiger, D., Sanglas, L., Aviles, F. X., Santoro, J., ... Arolas, J. L. (2011). Oxidative folding and structural analyses of a Kunitz-related inhibitor and its disulfide intermediates, functional implications. Journal of Molecular Biology, 414, 427–441.
Chen, R., Li, L., & Weng, Z. (2003). ZDOCK predictions for the CAPRI challenge. Proteins, 52, 80–87.
Copeland, R. A., Lombardo, D., Giannaras, J., & Decicco, C. P. (1995). Estimating Ki values for tight binding inhibitors from dose response plot. Bioorganic & Medicinal Chemistry Letters, 5, 1947–1952.
Czapinska, H., & Otlewski, J. (1999). Structural and energetic determinants of the S1-site specificity in serine proteases. European Journal of Biochemistry. 260, 671–695.
Damlé, M. S., Giri, A. P., Sainani, M. N., & Gupta, V. S. (2005). Higher accumulation of protease inhibitors in flowers than leaves and fruits as a possible basis for differential feeding preference of Helicoverpa armigera on tomato (Lycopersiconesculentum Mill, Cv. Dhanashree). Phytochemistry, 66, 2659–2667.
Duan, X., Li, X., Xue, Q., Abo-El-Saad, M., Xu, D., & Wu, R. (1996). Transgenic rice plants harboring an introduced potato proteinase inhibitor II gene are insect resistant. *Nature Biotechnology, 14*, 494–498.

Dunse, K. M., Kaas, Q., Guarino, R. F., Barton, P. A., & Craik, D. J. (2010). 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, K. M., Steven, J. A., Lay, F. T., Gaspar, Y. M., Heath, R. L., & Anderson, M. A. (2010). Coexpression of potato type I and II proteinase inhibitors gives cotton plants protection against insect damage in the field. *Proceedings of the National Academy Science USA, 107*, 15011–15015.

Ghosh, A., Sonavane, U., Andhirk, S., Aradhityam, G., & Joshi, R. (2012). Structural insights into human GPCR protein OAI, a computational perspective. *Journal of Molecular Modeling, 18*, 2117–2133.

Green, T. R., & Ryan, C. A. (1972). Wound induced proteinase inhibitor in plant leaves, a possible defense mechanism against insects. *Science, 175*, 776–777.

Hansen, D., Macedo-Ribeiro, S., Verissimo, P., Yoo, Im, S., Sampaio, M.U., & Oliva, M.L.V. (2007). Crystal structure of a novel cysteineless plant Kunitz-type protease inhibitor. *Biochemical and Biophysical Research Communications, 360*, 735–740.

Johnson, E. D., Miller, E. A., & Anderson, M. A. (2007). Dual location of a family of proteinase inhibitors within the stigmas of *Nicotiana alata*. *Planta, 225*, 1265–1276.

Johnson, R., Narayan, A. G., & Ryan, C. A. (1989). Expression of proteinase inhibitors I and II in transgenic tobacco plants, Effects on natural defense against insects. *Molecular Modeling, 18*, 2117–2133.

Karban, R. (1989). Induced plant response to herbivory. *Annual Review of Ecology Systematics, 20*, 331–348.

Kessler, A., & Baldwin, I. T. (2002). Plant responses to insect herbivory, the emerging molecular analysis. *Annual Review of Plant Biology, 53*, 299–328.

Koller, W., & Kolattukudy, P. E. (1982). Mechanism of action of cutinase, chemical modification of the catalytic triad characteristic for serine hydrolases. *Biochemistry, 21*, 3083–3090.

Kong, L., & Ranganathan, S. (2008). Tandem duplication, circular permutation, molecular adaptation, how Solanaceae resist pests via inhibitors. *BMC Bioinformatics, 9*, 522.

Lee, M. C., Scanlon, M. J., Craik, D. J., & Anderson, M. A. (1999). A novel two-chain proteinase inhibitor generated by circularization of a multidomain precursor protein. *Natural Structural Biology, 6*, 526–530.

Li, X.-Q., Zhang, T., & Donnelly, D. (2011). Selective loss of cysteine residues and disulfide bonds in a potato proteinase inhibitor II family. *PLoS ONE, 6*, e18615.

McManus, M. T., White, D. W. R., & McGregor, P. G. (1994). Accumulation of a chymotrypsin inhibitor in transgenic tobacco can affect the growth of insect pests. *Transgenic Research, 3*, 50–58.

Mishra, M., Tamhane, V. A., Khandelwal, N., Kulkarni, M. J., Gupta, V. S., & Giri, A. P. (2010). Interaction of recombinant CanPis with *Helicoverpa armigera* gut proteases reveals their processing patterns, stability and efficiency. *Proteomics, 10*, 2845–2857.

Nielsen, K. J., Heath, R. L., Anderson, M. A., & Craik, D. J. (1995). Structures of a series of 6-kDa trypsin inhibitors isolated from the stigma of *Nicotiana alata*. *Biochemistry, 34*, 14304–14311.

Otlewski, J., Jaskolski, M., Buszek, O., Cierpicki, T., Czapinska, H., Krowarsch, D., … Dadlez, M. (2001). Structure-function relationship of serine proteinase–protein inhibitor interaction. *Acta Biochimica Polonica, 48*, 419–428.

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

Pichare, M. M., & Kachole, M. S. (1994). Detection of electrophoretically separated proteinase inhibitors using X-ray film. *Journal of Biochemical and Biophysical Methods, 28*, 215–224.

Ryan, C. A. (1990). Protease inhibitors in plants, genes for improving defenses against insects and pathogens. *Annual Review Phytopathology, 28*, 425–449.

Sarate, P. J., Tamhane, V. A., Kotkar, H. M., Ratnakaran, N., Susan, N., Gupta, V. S., & Giri, A. P. (2012). Development and digestive flexibilities in midgut of a polyphagous pest *Helicoverpa armigera*. *Journal of Insect Science, 12*, 42.

Scanlon, M. J., Lee, M. C., Anderson, M. A., & Craik, D. J. (1999). Structure of a putative ancestral protein encoded by a single sequence repeat from a multidomain proteinase inhibitor gene from *Nicotiana alata*. *Structure and Folding Design, 7*, 793–802.

Schirra, H. J., Anderson, M. A., & Craik, D. J. (2008). Structural refinement of insecticidal plant proteinase inhibitors from *Nicotiana alata*. *Protein and Peptide Letters, 15*, 903–909.

Schirra, H. J., & Craik, D. J. (2005). Structure and folding of potato type II proteinase inhibitors, circular permutation and intramolecular domain swapping. *Protein and Peptide Letters, 12*, 421–431.

Schirra, H. J., Guarino, R. F., Anderson, M. A., & Craik, D. J. (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, 393*, 609–626.

Schirra, H. J., Scanlon, M. J., Lee, M. C., Anderson, M. A., & Craik, D. J. (2001). The solution structure of C1–T1, a two-domain proteinase inhibitor derived from a circular precursor protein from *Nicotiana alata*. *Journal of Molecular Biology, 306*, 69–79.

Sin, S. F., & Chye, M. L. (2004). Expression of proteinase inhibitor II proteins during floral development in *Solanum americanum*. *Planta, 219*, 1010–1022.

Tamhane, V. A., Chougule, N. P., Giri, A. P., Dixit, A. R., & Sainani, M. N. (2005). *In vivo and in vitro effect of Capsicum annuum proteinase inhibitors on Helicoverpa armigera* gut proteasines. *Biochimica et Biophysica Acta, 1722*, 156–167.

Tamhane, V. A., Giri, A. P., Kumar, P., & Gupta, V. S. (2009). Spatial and temporal expression patterns of diverse Pin-II proteinase inhibitor genes in *Capsicum annuum* Linn. *Gene, 442*, 88–98.

Tamhane, V. A., Giri, A. P., Sainani, M. N., & Gupta, V. S. (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.

Tsui, V., & Case, D. A. (2000). Theory and applications of the generalized born solvation model in macromolecular simulations. *Biopolymers, 56*, 275–291.
Xu, L., Li, Y., Zhou, S., & Hou, Y. (2012). Understanding microscopic binding of macrophage migration inhibitory factor with phenolic hydrazones by molecular docking, molecular dynamics simulations and free energy calculations. *Molecular BioSystems, 8*, 2260–2273.

Zalucki, M., Daglish, G., Firempong, S., & Twine, P. (1986). The biology and ecology of *Heliothis armigera* (Hubner) and *Heliothis punctigera* Wallengren (Lepidoptera, Noctuidae) in Australia - what do we know. *Australian Journal of Zoology, 34*, 779–814.

Zavala, J. A., Patankar, A. G., Gase, K., & Baldwin, I. T. (2004). Constitutive and inducible trypsin proteinase inhibitor production incurs large fitness costs in *Nicotiana attenuata*. *Proceedings of the National Academy Science USA, 101*, 1607–1612.

Zavala, J. A., Patankar, A. G., Gase, K., Hui, D., & Baldwin, I. T. (2004). Manipulation of endogenous trypsin proteinase inhibitor production in *Nicotiana attenuata* demonstrates their function as anti herbivore defenses. *Plant Physiology, 134*, 1181–1190.