Insect Biochemistry and Molecular Biology 65 (2015) 10–19

Contents lists available at ScienceDirect

Insect Biochemistry and Molecular Biology

journal homepage: www.elsevier.com/locate/ibmb

Positive selection of digestive Cys proteases in herbivorous Coleoptera

Juan Vorster [a, b, 1], Asieh Rasoolizadeh [a, 1], Marie-Claire Goulet [a], Conrad Cloutier [c], Frank Sainsbury [a, d], Dominique Michaud [a, *]

[a] Département de phytologie, CRIV–Biotechnologie, Université Laval, Québec, QC G1V 0A6, Canada
[b] Department of Plant and Soil Science, Forestry and Agricultural Biotechnology Institute, University of Pretoria, Pretoria, South Africa
[c] Département de biologie, Université Laval, Québec, QC G1V 0A6, Canada
[d] The University of Queensland, Australian Institute for Bioengineering and Nanotechnology, Centre for Biomolecular Engineering, St. Lucia, Queensland 4072, Australia

ARTICLE INFO

Article history:
Received 23 July 2014
Received in revised form
22 July 2015
Accepted 31 July 2015
Available online 8 August 2015

Keywords:
Positive selection
Herbivorous Coleoptera
Digestive cysteine proteases
Plant cystatins
Plant–insect interactions
Coevolutionary arms race
Leptinotarsa decemlineata Say

ABSTRACT

Positive selection is thought to contribute to the functional diversification of insect-inducible protease inhibitors in plants in response to selective pressures exerted by the digestive proteases of their herbivorous enemies. Here we assessed whether a reciprocal evolutionary process takes place on the insect side, and whether ingestion of a positively selected plant inhibitor may translate into a measurable rebalancing of midgut proteases *in vivo*. Midgut Cys proteases of herbivorous Coleoptera, including the major pest Colorado potato beetle (*Leptinotarsa decemlineata*), were first compared using a codon-based evolutionary model to look for the occurrence of hypervariable, positively selected amino acid sites among the tested sequences. Hypervariable sites were found, distributed within –or close to– amino acid regions interacting with Cys-type inhibitors of the plant cystatin protein family. A close examination of *L. decemlineata* sequences indicated a link between their assignment to protease functional families and amino acid identity at positively selected sites. A function-diversifying role for positive selection was further suggested empirically by *in vitro* protease assays and a shotgun proteomic analysis of *L. decemlineata* Cys proteases showing a differential rebalancing of protease functional family complements in larvae fed single variants of a model cystatin mutated at positively selected amino acid sites. These data confirm overall the occurrence of hypervariable, positively selected amino acid sites in herbivorous Coleoptera digestive Cys proteases. They also support the idea of an adaptive role for positive selection, useful to generate functionally diverse proteases in insect herbivores ingesting functionally diverse, rapidly evolving dietary cystatins.

1. Introduction

Protease::inhibitor interactions in plant–insect systems are the result of a long coevolutionary arms race triggering the continuous diversification of herbivore digestive proteases and plant protease inhibitors (Christeller, 2005; Lopes et al., 2004; Zhu-Salzman and Zeng, 2015). On the one side, herbivorous insects have evolved a range of strategies to cope with dietary protease inhibitors, typically involving the secretion of complex midgut protease complements, the overexpression of inhibitor-sensitive proteases to outnumber the ingested inhibitors, the up-regulation of protease isoforms weakly sensitive to inhibition, and degradation of the plant inhibitors with non-target proteases (Broadway, 2000; Zhu-Salzman and Zeng, 2015). On the other side, plants express a range of protease inhibitors upon wounding or insect attack, encoded by gene families responsive to various developmental and environmental stimuli (Ryan, 1990). Protease inhibitor diversity is illustrated by the occurrence of serine (Ser) protease inhibitor gene families in plant genomes (Barta et al., 2002; Kong and Ranganathan, 2008; Li et al., 2011) matching the occurrence of trypsin and chymotrypsin gene families in their lepidopteran insect predators (Srinivasan et al., 2006). Similarly, plants express an array

Abbreviations: Acn, acetonitrile; Cys, cysteine; Int, intestain; Ldp30, 30-kDa cystatin-sensitive Cys protease of Colorado potato beetle (*Leptinotarsa decemlineata*); MCA, methylcoumarin; RCR, relative leaf consumption rate; RGR, relative growth rate; Ser, serine; SlCYS8, eighth domain of tomato (*Solanum lycopersicum*) multicystatin; StCYS1, first domain of potato (*Solanum tuberosum*) multicystatin.

* Corresponding author. Département de phytologie, Université Laval, Pavillon Envirotron, 2480 boul. Hochelaga, Québec, QC G1V 0A6, Canada.
E-mail address: dominique.michaud@fsaa.ulaval.ca (D. Michaud).

[1] Equal contributors to this paper.

http://dx.doi.org/10.1016/j.ibmb.2015.07.017
0965-1748/

of stress-inducible cysteine (Cys) protease inhibitors, the so-called cystatins (Benchabane et al., 2010), matching complex sets of digestive Cys proteases in herbivorous Coleoptera (Gruden et al., 2004; Sainsbury et al., 2012a; *Tribolium* Genome Sequencing Consortium, 2008).

Several evolutionary processes have shaped the organization of protease inhibitor complements in biological systems, notably involving gene duplication followed by positive selection of non-synonymous mutations at functionally relevant amino acid sites (Christeller, 2005). Protease inhibitor gene families derived from multiple gene duplications is a common feature of plant genomes, as exemplified by the proteinase inhibitor II family of Solanaceae (Barta et al., 2002; Kong and Ranganathan, 2008; Mishra et al., 2012), the mustard trypsin inhibitor family of Cruciferae (Clauss and Mitchell-Olds, 2004), the Kunitz trypsin inhibitors of *Populus* and *Solanum* species (Neiman et al., 2009; Speranskaya et al., 2012) or the multigene cystatin complex of higher plants (Girard et al., 2007; Martinez et al., 2005; Massonneau et al., 2005; Tan et al., 2014). Inhibitor variants within these families often show sequence hypervariability at specific amino acid positions, presumably indicative of positive selection and functional diversification towards arthropod herbivore digestive proteases (Ingvarsson, 2005; Kiggundu et al., 2006; Kong and Ranganathan, 2008; Li et al., 2011; Neiman et al., 2009; Talyzina and Ingvarsson, 2006).

An example is potato multicystatin, an eight-domain cystatin known to protect endogenous storage proteins in potato tubers (Green et al., 2013; Weeda et al., 2009) which is also upregulated in leaves upon insect attack (Bouchard et al., 2003). The eight domains of this protein likely were the result of multiple gene duplications in close *Solanum* ancestor(s) (Benchabane et al., 2010). Their primary sequence includes hypervariable, rapidly evolving amino acid sites giving the protein a range of inhibitory specificities towards plant and insect Cys proteases (Goulet et al., 2008; Kiggundu et al., 2006). Here we assessed whether a similar evolutionary process involving positive selection is taking place on the insect side, using digestive Cys proteases of herbivorous Coleoptera as a model. We also conducted enzymology and functional proteomics work on midgut Cys proteases of the coleopteran herbivore Colorado potato beetle (*Leptinotarsa decemlineata*) to establish a link between the eventual positive selection of Cys proteases in coleopteran herbivores and the response of these insects to the functionally diverse, rapidly evolving cystatins of their plant hosts.

2. Methods

2.1. Sequence variability inferences

Hyperviariable, positively selected amino acids were searched for in the coding sequences of 25 midgut Cys protease genes from the coleopteran herbivores cowpea weevil (*Callosobruchus maculatus*), Western corn rootworm (*Diabrotica virgifera virgifera*) and *L. decemlineata* (Table S1, Fig. S1) using the codon-based mechanistic–empirical combination (MEC) evolutionary model of Doron-Faigenboim and Pupko (2006). K_a/K_s ratios –or ω values– were calculated online using the Selecton server for the identification of codon site-specific positive selection and purifying selection, v. 2.4 (http://selecton.tau.ac.il) (Stern et al., 2007), after uploading the codon-aligned sequences of (Fig. S2), and an unrooted phylogenetic tree for (Fig. S3), the 25 proteases. A codon sequence alignment was first produced using the Multiple Sequence Comparison by Log-Expectation (MUSCLE) tool (Edgar, 2004), and then edited manually based on the protein alignment (Fig. S1) to eliminate indel mismatches between the two sequence datasets. An unrooted phylogenetic tree was inferred from the resulting sequences by the neighbor-joining distance method of Saitou and Nei (1987), after generating a sequence similarity matrix using Kimura's two-parameter model (Kimura, 1983). Because no null model is nested in the MEC model, an Akaike's Information Criteria (AIC_c) score calculated with this model was compared with the AIC_c score obtained under the M8a null model of Swanson et al. (2003), which assumes no positive selection in the tested sequences. The MEC model represents a better fit to the data when the calculated AIC_c score is lower than the AIC_c score calculated for the M8a model. In such a case, a confidence interval between the 5th and 95th percentiles is estimated for amino acid positions showing a ω value greater than 1. When the lower bound of the confidence interval is greater than 1, the inference of positive selection at this position is considered statistically significant.

2.2. Intestain::cystatin interactions

Cys protease::cystatin docking simulations were run using as an example the midgut Cys protease of *L. decemlineata* Intestain D4 (IntD4) (GenBank Accession No. EF154436) (Sainsbury et al., 2012a) and the first inhibitory domain of potato multicystatin StCYS1 (GenBank L16450) (Kiggundu et al., 2006), after building structure homology models for the two proteins. Twenty tentative models were built for each protein using Modeller, v. 9.7 (Eswar et al., 2006), with the crystal structure of human cathepsin L (PDB 1SC8; http://www.wwpdb.org/) as a template for IntD4 and the NMR structure of oryzacystatin I (PDB 1EQK) as a template for StCYS1. Stereochemical quality of the models was assessed by comparison to the template structures using the MetaMQAPII algorithm (https://genesilico.pl/toolkit/unimod?method=MetaMQAPII), and the best models were selected for further analysis. Intestain::cystatin interactions were simulated using the Z-Dock algorithm of Discovery Studio, v. 3.0 (Accelrys Software), to generate 2000 tentative poses of the resulting complex (Chen et al., 2003). Poses with the highest Z-scores were compared to the solved crystal structure of human cathepsin L interacting with human stefin A (PDB 3KSE) to validate the relative binding position and orientation of the two proteins in the predicted complex. Five complexes were chosen and refined through energy minimization using the R-Dock algorithm (Li et al., 2003). The top ranking model was selected for complex visualization and inference of protease and cystatin interacting residues at the binding interface (Sainsbury et al., 2012a).

2.3. Recombinant cystatins

Wild-type tomato cystatin SlCYS8 and two single variants, P2I and T6R, exhibiting different inhibitory spectra against insect and plant Cys proteases were expressed in –and purified from– BL21 *Escherichia coli* cells using the glutathione *S*-transferase gene fusion system (GE Healthcare) as described earlier (Goulet et al., 2008). Biotinylated versions of the cystatins were also produced, which included a 17-amino acid 'AviTag' peptide (Beckett et al., 1999) at the C-terminus for subsequent immobilization on avidin-embedded resin (see *2.7 Mass spectrometry*, below). AviTagged wild-type SlCYS8 was produced using a recently devised Golden-Gate cloning construct (Sainsbury et al., 2012a). AviTag constructs for P2I and T6R were devised by site-directed mutagenesis of the AviTagged SlCYS8 template, using the Quickchange mutagenesis kit (Stratagene) and appropriate DNA primers to drive the selected mutations (Table S2). The AviTagged cystatins were biotinylated *in vivo* by expression in AVB101 *E. coli* cells coexpressing a biotin ligase (Avidity LLC), with 50 μM D-biotin added in the growth medium.