Research paper

Prediction of a peptidome for the western tarnished plant bug *Lygus hesperus*

Andrew E. Christie [a,*], J. Joe Hull [b], Josh A. Richer [b], Scott M. Geib [c], Erica E. Tassone [d]

[a] Békésy Laboratory of Neurobiology, Pacific Biosciences Research Center, School of Ocean and Earth Science and Technology, University of Hawaii at Manoa, 1993 East-West Road, Honolulu, HI 96822, USA
[b] Pest Management and Biocontrol Research Unit, US Arid Land Agricultural Research Center, USDA Agricultural Research Services, Maricopa, AZ 85138, USA
[c] Tropical Crop and Commodity Protection Research Unit, Daniel K. Inouye Pacific Basin Agricultural Research Center, USDA Agricultural Research Services, Hilo, HI 96720, USA
[d] Plant Physiology and Genetics Research Unit, US Arid Land Agricultural Research Center, USDA Agricultural Research Services, Maricopa, AZ 85138, USA

ARTICLE INFO

ABSTRACT

Many strategies for controlling insect pests require an understanding of their hormonal signaling agents, peptides being the largest and most diverse single class of these molecules. *Lygus hesperus* is a pest species of particular concern, as it is responsible for significant damage to a wide variety of commercially important plant crops. At present, little is known about the peptide hormones of *L. hesperus*. Here, transcriptomic data were used to predict a peptidome for *L. hesperus*. Fifty-three *L. hesperus* transcripts encoding peptide precursors were identified, with a subset amplified by PCR for sequence verification. The proteins deduced from these transcripts allowed for the prediction of a 119-sequence peptidome for *L. hesperus*. The predicted peptides include isoforms of allatostatin A, allatostatin B (AST-B), allatostatin C, allatotropin, bursicon, CCHamide, corazonin, crustacean cardioactive peptide, crustacean hyperglycemic hormone/ion transport peptide, diuretic hormone 31, GSEFLamide, insulin-like peptide, myosuppressin, neuroparsin, neuropeptide F, orcokinin, orcomyotropin, pyrokinin, short neuropeptide F, SIFamide, sulfakinin and tachykinin-related peptide. Of note were several isoforms of AST-B that possess –WX$_7$Wamide carboxyl-termini rather than the stereotypical –WX$_6$Wamide (*e.g.,* KWQDMQNPGWamide), an allatotropin ending in –SARGFamide rather than –TARGFamide (GLKNGPLNSARGFamide), a GSEFLamide ending in –GTEFLamide (TVGTEFLamide), several orcokinins with PMDEIDR– rather than NFDEIDR– amino-termini (*e.g.,* PMDEIDRAGFTHFV), and an eight rather than 12 amino acid long isoform of SIFamide (PPFNGSIFamide). Collectively, the *L. hesperus* peptidome predicted here provides a resource for initiating physiological investigations of peptidergic signaling in this species, including studies directed at the biological control of this agricultural pest.

1. Introduction

Peptides are the largest and most diverse single class of molecules used for chemical communication (*e.g.,* Kastin, 2006). In insects, peptidergic signaling systems are key components in the control of essentially all aspects of physiology and behavior, and as such, are of growing interest for the development of biological control strategies for many pest species (*e.g.,* Altstein, 2004; Audsley and Down, 2015; Nachman et al., 2009a,b; Scherkenbeck and Zdobinsky, 2009; Van Hiel et al., 2010; Xie et al., 2015; Zhang et al., 2015a,c,). Insecticidal effects have been demonstrated via the knockdown of several peptidergic systems, as well as through the application of native and/or modified peptide analogs,

and targets for biological control include, among others, the peptidergic pathways involved in reproduction, development, growth and feeding (for a recent review see Audsley and Down (2015)).

Members of the genus *Lygus*, a complex of 28 morphologically similar polyphagous mirid plant bugs (Schwartz and Foottit, 1998; Wheeler, 2001; Schuh, 2013), cause significant damage via their cell-rupture (also referred to as "lacerate-and-flush") method of feeding (Backus et al., 2005, 2007). In this strategy, *Lygus* species inject a mixture of proteolytic enzymes into plant vegetative and reproductive structures via specialized stylets (Shackel et al., 2005; Backus et al., 2007). Crop losses, which are mainly attributed to pre-reproductive adults and *Lygus* nymphs (Cooper and Spurgeon, 2013), arise from a combination of saliva-induced cell rupturing and plant wound responses that result in deformation of developing fruit, reduced vegetative growth, and feeding site necrosis (Strong, 1970; Tingey and Pillimer, 1977). In the western

United States, the predominant *Lygus* species is *Lygus hesperus* Knight (commonly referred to as the western tarnished plant bug), which has an extensive host plant range (>150 species) including many economically important food, fiber, and seed crops (Scott, 1977; Wheeler, 2001; Ritter et al., 2010; Naranjo et al., 2011). Despite its significance as an agricultural pest, essentially nothing is known about the native peptide hormones of *L. hesperus*.

While many methods can be used to determine arthropod peptidomes, *in silico* transcriptome/genome mining provides a rapid means for large-scale peptide discovery when appropriate molecular resources are available for a species (*e.g.*, Dircksen et al., 2011; Hauser et al., 2010; Hummon et al., 2006; Toullec et al., 2013; Ventura et al., 2014; Weaver and Audsley, 2008; Yan et al., 2012). A common strategy for this type of analysis is one based on homology to known peptide precursors, with the known protein being used to probe the target genomic or transcriptomic dataset for genes/transcripts encoding putative homologs (*e.g.* Christie, 2008a,b, 2014a,b,c,d,e,f, 2015a,b,c,d, 2016a,b; Christie and Chi, 2015a,b,c; Christie et al., 2008, 2010a, 2011a,b,c, 2013, 2015; Christie and Pascual, 2016; Gard et al., 2009; Ma et al., 2009, 2010). Once identified, the gene or transcript of interest can be translated, and the deduced pre/preprohormone subjected to a simple bioinformatics workflow that utilizes freeware and homology to known arthropod precursor processing schemes to predict the structures of the mature bioactive peptides that are likely generated from it (*e.g.* Christie, 2008a,b, 2014a,b,c,d,e,f, 2015a,b,c,d, 2016a,b; Christie and Chi, 2015a,b,c; Christie et al., 2008, 2010a, 2011a,b,c, 2013, 2015; Christie and Pascual, 2016; Gard et al., 2009; Ma et al., 2009, 2010).

The development of a transcriptome shotgun assembly (TSA) for *L. hesperus* (BioProject Nos. PRJNA210219 and PRJNA238835 (Hull et al., 2014)) provides a resource for peptide discovery in this species, and here these data were used for such an investigation. As the results that follow will show, numerous peptide-encoding transcripts were identified within the *L. hesperus* TSA dataset, allowing for the prediction of over 100 distinct peptide hormones/modulators for this species. The identified peptides include many involved in the control of reproduction, development, growth and feeding in other insects, and it is likely that they play similar roles in *L. hesperus* as well. Given their putative functional roles, the identified transcripts and peptides provide a resource for initiating physiological investigations of peptidergic signaling in *L. hesperus*, including studies directed at exploring their potential use as targets for biorational-based control of this important agricultural pest.

2. Materials and methods

2.1. Database searches

Database searches were conducted on or before March 22, 2016, using methods modified from a well-vetted protocol (*e.g.* Christie, 2008a,b, 2014a,b,c,d,e,f, 2015a,b,c,d, 2016a,b; Christie and Chi, 2015a,b,c; Christie et al., 2008, 2010a, 2011a,b, 2013, 2015; Christie and Pascual, 2016; Gard et al., 2009; Ma et al., 2009, 2010). Specifically, the database of the online program tblastn (National Center for Biotechnology Information, Bethesda, MD; http://blast.ncbi.nlm.nih.gov/Blast.cgi) was set to "Transcriptome Shotgun Assembly (TSA)" and restricted to data from *Lygus hesperus* "(taxid:30085)". Known insect or crustacean peptide precursors were input into tblastn as query sequences, and all hits returned by a given search were fully translated using the "Translate" tool in ExPASy (http://web.expasy.org/translate/) and then checked manually for homology to the target query. The complete list of peptide families searched for in this study, as well as the

specific queries used, is provided in Table 1; this table also provides the BLAST-generated maximum score and E-value for each of the transcripts identified as encoding a putative neuropeptide precursor.

2.2. Peptide prediction

The structures of mature peptides were predicted using a well-established workflow (*e.g.* Christie, 2008a,b, 2014a,b,c,d,e,f, 2015ab,c,d, 2016a,b; Christie and Chi, 2015a,c; Christie et al., 2008, 2010a, 2011a,b,c, 2013, 2015; Christie and Pascual, 2016; Gard et al., 2009; Ma et al., 2009, 2010). Specifically, each of the deduced precursor proteins was assessed for the presence of a signal peptide using the online program SignalP 4.1 (http://www.cbs.dtu.dk/services/SignalP/; Petersen et al., 2011); the D-cutoff values of SignalP 4.0 were set to "Sensitive" to better match the sensitivity of version 3.0 of this freeware program. Prohormone cleavage sites were identified based on the information presented in Veenstra (2000) and/or by homology to known arthropod pre/preprohormone processing schemes. When present, prediction of the sulfation state of tyrosine residues was conducted using the online program "Sulfinator" (http://www.expasy.org/tools/sulfinator/; Monigatti et al., 2002). Disulfide bonding between cysteine residues was predicted by homology to known peptide isoforms and/or by using the online program "DiANNA" (http://clavius.bc.edu/~clotelab/DiANNA/; Ferrè and Clote, 2005). Other post-translational modifications, *i.e.*, cyclization of amino (N)-terminal glutamine/glutamic acid residues and carboxyl (C)-terminal amidation at glycine residues, were predicted by homology to known arthropod peptide isoforms. Fig. 1 shows two examples of mature peptide structural prediction using the workflow just described; the mature structures of all peptides predicted in this study are provided in Table 2. All protein/peptide alignments were done using the online program MAFFT version 7 (http://mafft.cbrc.jp/alignment/software/; Katoh and Standley, 2013).

2.3. Peptide-encoding transcript cloning

To help validate the *in silico* peptide predictions, a subset of the putative peptide-encoding transcripts were cloned and sequenced. Oligonucleotide primers (Table 3) designed to amplify the putative open reading frames (*i.e.*, the portion of the transcript beginning with the start codon and ending with the stop codon) were designed based on the nucleotide sequences of the transcripts of interest using Primer3 (Untergasser et al., 2012). Total RNA was isolated from 7 to 9 day old mixed sex adult *L. hesperus* heads (obtained from an in-house-reared colony [US Arid Land Agricultural Research Center, USDA Agricultural Research Services, Maricopa, AZ]) using TRI Reagent Solution (Life Technologies/Ambion, Carlsbad, CA) according to the manufacturer's instructions. RNA quality and quantity were determined spectrophotometrically using the Take3 module on a Synergy H4 Hybrid Multi-Mode Microplate Reader (Biotek Instruments, Winooski, VT). To remove any residual genomic DNA, a 1 µg aliquot of RNA was treated with DNase I (New England Biolabs, Ipswich, MA), and then 500 ng of the DNA-free RNA was used as a template for cDNA synthesis using a SuperScript III First-Strand Synthesis System (Life Technologies) with custom-made random pentadecamers (Integrated DNA Technologies, San Diego, CA). Peptide transcripts were amplified in a 20-µL reaction volume using SapphireAmp Fast PCR Master Mix (Clontech Laboratories Inc., Mountain View, CA) with 12.5 ng cDNA and 0.2 µM sense and antisense primers on a Biometra TRIO (Göttingen, Germany) thermocycler. PCR conditions consisted of an initial denaturation at 95 °C for 2 min followed by 40 cycles of 95 °C for 20 s, 53 °C for 20 s, and 72 °C for 60 s, and concluded at 72 °C for 5 min. PCR products were electrophoresed using a