General and Comparative Endocrinology xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

General and Comparative Endocrinology

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

In silico prediction of a neuropeptidome for the eusocial insect *Mastotermes darwiniensis*

Andrew E. Christie *

Békésy Laboratory of Neurobiology, Pacific Biosciences Research Center, University of Hawaii at Manoa, 1993 East-West Road, Honolulu, HI 96822, USA

ARTICLE INFO

Article history:
Received 29 January 2015
Accepted 9 June 2015
Available online xxxx

Keywords:
Bioinformatics
Neurohormone
Neuropeptide
Transcriptome shotgun assembly
Isoptera
Termite

ABSTRACT

Mastotermes darwiniensis is the most basal living member of the Isoptera (termites), yet it exhibits an extremely advanced level of eusocial organization. Given the interest in, and the high levels of differential developmental and behavioral control needed for, eusociality, it is surprising that essentially nothing is known about the native peptides of *M. darwiniensis*, which undoubtedly represent the largest and most diverse class of hormones present in this species. The recent public deposition of a 100,000^{+}-sequence transcriptome for *M. darwiniensis* provides a means for peptide discovery in this termite. Here, this resource was mined for putative peptide-encoding transcripts via the BLAST algorithm tblastn and known arthropod neuropeptide precursor queries; mature peptide structures were predicted from the deduced pre/preprohormones using a well-vetted bioinformatics workflow. Thirty-four *M. darwiniensis* peptide-encoding transcripts were identified, with 163 distinct mature peptides predicted from these sequences. These peptides included members of the adipokinetic hormone, adipokinetic hormone-corazonin-like peptide, allatostatin A, allatostatin C, allatotropin, bursicon β, CCHamide, corazonin, crustacean cardioactive peptide, crustacean hyperglycemic hormone/ion transport peptide, diuretic hormone 31, diuretic hormone 44, FMRFamide-like peptide, insulin-like peptide, leucokinin, myosuppressin, neuroparsin, neuropeptide F, orcokinin, pigment dispersing hormone, pyrokinin, RYamide, short neuropeptide F, SIFamide, sulfakinin and tachykinin-related peptide families. This peptidome is the largest thus far predicted for any member of the Isoptera, and provides a foundation for initiating studies of peptidergic signaling in this and other termites, including ones directed at understanding the roles peptide hormones play in the developmental and behavioral control required for eusociality.

1. Introduction

Eusociality is an evolutionarily advanced level of societal existence; it is considered by many to be the highest level of communal organization exhibited by animals (*e.g.* Wilson and Hölldobler, 2005). All eusocial species share four common characteristics (Wilson, 1971): (1) adults live in colonies, (2) there is an overlap of generations within the colony, (3) there is cooperative care of the young by adults (including care of young that are not the offspring of a given caregiver) and (4) there is a physiological/behavior division of labor between the various castes that form the colony, most commonly involving reproduction. Eusociality is distinguished from all others type of social organization in that at least one caste within a colony has lost the ability to perform a subset of the behaviors (*e.g.* reproduction) that is exhibited by individuals comprising another caste (Plowes, 2010). While eusociality is present in a variety of animals, including several mammals, it is most commonly seen in arthropods, and in particular in members of the insect orders Hymenoptera (ants, bees and wasps) and Isoptera (termites), where it has achieved its most advanced states (Plowes, 2010).

Differential developmental and behavioral control in the castes that form eusocial colonies are key to the evolution of eusociality, and are among the trending research themes in the study of eusocial behavior (Plowes, 2010). For example, in some termites, a colony is composed of the reproductive adult male(s) and female(s), the non-reproductive workers and soldiers, and other immature individuals in varying stages of development (Plowes, 2010). Both the developmental commitment of individuals into different caste types, as well as the differences in physiology and behavior seen between castes, is, at least in part, achieved via endocrine modulation of control systems operating at multiple levels. Circulating and locally-released hormones, including peptides, the largest and most diverse class of hormones in all multicellular

* Tel.: +1 808 956 5212; fax: +1 808 956 6984.
E-mail address: crabman@pbrc.hawaii.edu

http://dx.doi.org/10.1016/j.ygcen.2015.06.006
0016-6480/

animals (Kastin, 2006), are undoubtedly significant contributors to this developmental and/or physiological/behavioral control (*e.g.* Cornette et al., 2008; Elliott and Stay, 2008).

With the advent of high throughput nucleotide sequencing, large genomic and/or transcriptomic datasets have been generated for several members of the Hymenoptera, and peptidomes of several eusocial species have been deduced from these resources (*e.g.* the honey bee *Apis mellifera* [Hummon et al., 2006]). In contrast, vanishingly little information on the native peptide hormones present in termites currently exists (*i.e.* Veenstra, 2014). In the study presented here, the publicly accessible transcriptome shotgun assembly (TSA) data for the giant northern termite *Mastotermes darwiniensis* (**BioProject Accession No. PRJNA219576**; Misof et al., 2014) was mined for putative neuropeptide-encoding transcripts, which were subsequently used to predict a peptidome for this species via a well-vetted bioinformatics workflow (*e.g.* Christie, 2008a,b, 2014a,b,c,d,e,f, 2015a,b; Christie and Chi, 2015a,b; Christie et al., 2008, 2010a, 2011a,b, 2013; Gard et al., 2009; Ma et al., 2009, 2010). One hundred sixty-three distinct mature peptide structures were predicted for *M. darwiniensis*, including member of the adipokinetic hormone (AKH), adipokinetic hormone-corazonin -like peptide (ACP), allatostatin A (AST-A), allatostatin C (AST-C), allatotropin, bursicon β, CCHamide, corazonin, crustacean cardioactive peptide (CCAP), crustacean hyperglycemic hormone (CHH)/ion transport peptide (ITP), diuretic hormone 31 (DH31), diuretic hormone 44 (DH44), FMRFamide-like peptide (FLP), insulin-like peptide (ILP), leucokinin, myosuppressin, neuroparsin, neuropeptide F (NPF), orcokinin, pigment dispersing hormone (PDH), pyrokinin, RYamide, short neuropeptide F (sNPF), SIFamide, sulfakinin and tachykinin-related peptide (TRP) families. Taken collectively, these data not only greatly expand the catalog of putative peptide hormones known for the Isoptera, but also provide a foundation for examining how peptidergic signaling systems may contribute to the developmental and behavioral control required for eusociality in this and other termite species.

2. Materials and methods

2.1. Database searches

Database searches were conducted on or before January 24, 2015, using methods modified from a well-vetted protocol (*e.g.* Christie, 2008a,b, 2014a,b,c,d,e,f, 2015a,b; Christie and Chi, 2015a,b; Christie et al., 2008, 2010a, 2011a,b; 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 *M. darwiniensis* "(taxid:13139)". Known insect or crustacean peptide precursors were input into tblastn as the query sequences, and all hits returned by a given search were fully translated using the "Translate" tool of ExPASy (http://web.expasy.org/translate/) and then checked manually for homology to the target query. The complete list of peptides families searched for in this study, as well as the specific queries used, are provide 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 freeware workflow (*e.g.* Christie, 2008a,b, 2014a,b,c,d,e,f, 2015a,b; Christie and Chi, 2015a,b; Christie et al., 2008, 2010a, 2011a,b,c, 2013; 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 for the program were set to "Sensitive". 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, *e.g.* cyclization of amino (N)-terminal glutamine/glutamic acid residues and C-terminal amidation at glycine residues, were predicted by homology to known arthropod peptide isoforms. Fig. 1 shows three 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).

3. Results

In this study, 36 distinct peptide families/subfamilies were searched for within the publicly accessible *M. darwiniensis* TSA sequence database (Table 1). In the interest of space, only those searches that resulted in the identification of putative precursor-encoding transcripts are described here (Table 1), with the data presented in alphabetical order based on family name. All precursor proteins listed as "full-length" exhibit a functional signal sequence (including a "start" methionine) and are flanked on their C-terminus end by a stop codon. Proteins described here as "partial" lacked a start methionine (referred to as C-terminal partial proteins), a stop codon (referred to as N-terminal partial proteins), or both of these features (referred to as internal protein fragments).

3.1. Adipokinetic hormone

Using the sequence of an AKH preprohormone from *Drosophila melanogaster* (**Accession No. AAF47846**; Adams et al., 2000) as a query, a single *M. darwiniensis* transcript was identified as encoding a putative AKH precursor (Table 1). Translation of this transcript revealed a 70 amino acid, full-length prepro-hormone (Masda-prepro-AKH; Fig. 1A). Two distinct peptides were predicted from Masda-prepro-AKH (Table 2 and Fig. 1A), one of which, pQVNFSPNWamide, is a known *M. darwiniensis* AKH isoform (Liebrich et al., 1995). Analysis by DiANNA of the linker/precursor-related peptide predicted from Masda-prepro-AKH, SGLQDGPCKTSTDSLMYIYKLIQSEAQKLVDCEKFGAN, suggests the presence of a disulfide bridge between the two cysteine residues in this peptide (Table 2).

3.2. Adipokinetic hormone-corazonin-like peptide

A single *M. darwiniensis* transcript was identified as encoding a putative ACP precursor using an ACP preprohormone from *Tribolium castaneum* (**Accession No. ADF28807**; Hansen et al., 2010) as the query sequence (Table 1). Translation of this transcript revealed a 68 amino acid, N-terminal partial precursor (Masda-prepro-ACP; Fig. 2A). Two distinct peptides were predicted from the extant portion of Masda-prepro-ACP (Table 2 and Fig. 2A),

Please cite this article in press as: Christie, A.E. *In silico* prediction of a neuropeptidome for the eusocial insect *Mastotermes darwiniensis*. Gen. Comp. Endocrinol. (2015), http://dx.doi.org/10.1016/j.ygcen.2015.06.006