General and Comparative Endocrinology 224 (2015) 18–37

Contents lists available at ScienceDirect

General and Comparative Endocrinology

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

Identification of the first neuropeptides from the enigmatic hexapod order Protura

Andrew E. Christie *, Megan Chi

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 4 March 2015
Revised 28 May 2015
Accepted 29 May 2015
Available online 6 June 2015

Keywords:
Bioinformatics
Neurohormone
Neuropeptide
Transcriptome shotgun assembly
Acerentomon sp. AD-2013

ABSTRACT

The Hexapoda consists of two classes, the Entognatha and the Insecta, with the former group considered basal to the latter. The Protura is a basal order within the Entognatha, the members of which are minute soil dwellers first identified in the early 20th century. Recently, a transcriptome shotgun assembly (TSA) was generated for the proturan *Acerentomon* sp., providing the first significant molecular resource for this enigmatic hexapod order. As part of an ongoing effort to predict peptidomes for little studied members of the Arthropoda, we have mined this TSA dataset for transcripts encoding putative neuropeptide precursors and predicted the structures of mature peptides from the deduced proteins. Forty-seven peptide-encoding transcripts were mined from the *Acerentomon* TSA dataset, with 202 distinct peptides predicted from them. The peptides identified included isoforms of adipokinetic hormone, adipokinetic hormone-corazonin-like peptide, allatostatin A, allatostatin B, allatostatin C, allatotropin, bursicon α, bursicon β, CCHamide, corazonin, crustacean cardioactive peptide, crustacean hyperglycemic hormone/ion transport peptide, diuretic hormone 31, diuretic hormone 44, ecdysis-triggering hormone, eclosion hormone, FMRFamide-like peptide, GSEFLamide, insulin-like peptide, intocin, leucokinin, myosuppressin, neuropeptide F, orcokinin, proctolin, pyrokinin, RYamide, short neuropeptide F, SIFamide, sulfakinin and tachykinin-related peptide; these are the first neuropeptides described from any proturan. Comparison of the *Acerentomon* precursors and mature peptides with those from other arthropods revealed features characteristic of both the insects and the crustaceans, which is consistent with the hypothesized phylogenetic position of the Protura within the Pancrustacea, *i.e.* at or near the point of divergence of the hexapods from the crustaceans.

1. Introduction

The Protura is an enigmatic group of arthropods that was first described in the early 20th century (for review see: Pass and Szucsich, 2011). Recent molecular analyses place the proturans at or near the base of the hexapod lineage (*e.g.* Misof et al., 2014), making them an important group of animals for understanding the evolution of neuropeptide signaling in the Arthropoda, particularly within the Pancrustacea (Crustacea + Hexapoda [*e.g.* Dunn et al., 2008; Friedrich and Tautz, 1995; Richter, 2002; Roeding et al., 2007]). This said, the neurochemistry of the Protura is uninvestigated, with no native peptide hormones known from any member of this taxon; as in all multicellular animals (*e.g.* Kastin, 2006), peptides undoubtedly represent the largest single class of hormones in proturans.

The lack of knowledge concerning the peptidergic systems of proturans is due, at least in part, to their small size and difficulty of collection (they are typically very small [<2 mm] and are soil or leaf litter dwellers [*e.g.* Pass and Szucsich, 2011]), which precludes peptide discovery via many conventional means that require large pools of starting material for analysis, for example brute force biochemical isolation/sequencing and biological mass spectrometry (*e.g.* accurate mass matching and *de novo* tandem mass spectrometric sequencing). However, the recent public deposition of a large (100,000^{+}-sequence) transcriptome shotgun assembly (TSA) for a member of the Protura, *Acerentomon* sp. AD-2013 (BioProject No. PRJNA219521; Misof et al., 2014), provides an alternative strategy for neuropeptide discovery in proturans, namely *in silico* transcriptome mining and bioinformatics peptide prediction.

Unlike mass spectrometric and biochemical-based methods of peptide discovery, the generation of a transcriptome can be achieved using a small amount of starting material, and once assembled, provides a permanent resource for identifying transcripts encoding essentially any protein of interest, including those encoding neuropeptide precursors (e.g. Christie et al., 2010a). With a set of peptide-encoding transcripts in hand, the structures of

* Corresponding author. Tel.: +1 808 956 5212; fax: +1 808 956 6984.
E-mail address: crabman@pbrc.hawaii.edu (A.E. Christie).

http://dx.doi.org/10.1016/j.ygcen.2015.05.015
0016-6480/

mature peptides can be predicted by subjecting the deduced pre/preprohormone sequences to a simple bioinformatics workflow that involves identifying signal sequences, prohormone cleave sites and post-translational modifications using freeware and homology to known peptide processing schemes (*e.g.* Christie et al., 2010a). Via this strategy, large peptidomes have recently been predicted for a number of small, rare and/or geographically inaccessible arthropod species (*e.g.* Christie, 2008a, 2008b, 2014a, 2014b, 2014c, 2014d, 2014e, 2015a, 2015b, 2015c; Christie and Chi, 2015a; Gard et al., 2009).

In the study presented here, the publicly accessible TSA sequences for the Protura were mined for putative neuropeptide -encoding transcripts, with a peptidome for *Acerentomon* sp. subsequently predicted using this data. In total, 202 distinct mature peptide structures were deduced using the identified *Acerentomon* sp. transcripts. The predicted peptides include members of the adipokinetic hormone (AKH), adipokinetic hormone-corazonin -like peptide (ACP), allatostatin A (AST-A), allatostatin B (AST-B), allatostatin C (AST-C), allatotropin, bursicon α, bursicon β, CCHamide, corazonin, crustacean cardioactive peptide (CCAP), crustacean hyperglycemic hormone (CHH)/ion transport peptide (ITP), diuretic hormone 31 (DH31), diuretic hormone 44 (DH44), ecdysis-triggering hormone (ETH), eclosion hormone (EH), FMRFamide-like peptide (FLP), GSEFLamide, insulin-like peptide (ILP), intocin, leucokinin, myosuppressin, neuropeptide F (NPF), orcokinin, proctolin, pyrokinin, RYamide, short neuropeptide F (sNPF), SIFamide, sulfakinin and tachykinin-related peptide (TRP) families, as well as a large collection of linker/precursor-related peptides. While most of the peptides described here possess novel structures, a number of known insect and/or crustacean isoforms were also discovered, supporting the position of the Protura at the base of the hexapod lineage and close to the Hexapoda-Crustacea split. The predicted *Acerentomon* sp. peptidome is the only one currently extant for any member of the Protura; it is also one of the largest peptidomes thus far predicted from any single arthropod species using *in silico* transcriptome mining. This catalog of *Acerentomon* sp. peptide structures represents a new resource for initiating anatomical and physiological investigations of peptidergic signaling in one of the most basal hexapod taxons.

2. Materials and methods

2.1. In silico peptide discovery

2.1.1. Database searches

Database searches were conducted on or before January 29, 2015, using methods modified from a well-vetted protocol (*e.g.* Christie, 2008a, 2008b, 2014a, 2014b, 2014c, 2014d, 2014e, 2014f, 2015a, 2015b, 2015c; Christie and Chi, 2015a, 2015b; Christie et al., 2008, 2010b, 2011a, 2011b; 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 the Protura "(taxid:29999)". Known arthropod 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 peptide families searched for in this study, as well as the specific queries used (typically ones that have been employed successfully for similar bioinformatics analyses; *e.g.* Christie, 2008a, 2008b, 2014a, 2014b, 2014c, 2014d, 2014e, 2014f, 2015a, 2015b, 2015c; Christie and Chi, 2015a, 2015b; Christie et al., 2008, 2010b, 2011a, 2011b; Gard et al., 2009; Ma et al., 2009, 2010), are 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.

2.1.2. Peptide prediction

The structures of mature peptides were predicted using a well-established workflow (*e.g.* Christie, 2008a, 2008b, 2014a, 2014b, 2014c, 2014d, 2014e, 2014f, 2015a, 2015b, 2015c; Christie and Chi, 2015a, 2015b; Christie et al., 2008, 2010b, 2011a, 2011b, 2011c, 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 carboxyl (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).

2.2. Reciprocal blastp analyses

To identify the known arthropod proteins most similar to the full-length *Acerentomon* sp. pre/preprohormones identified in this study, each deduced protein was used as the query sequence in a blastp search of the non-redundant, arthropod protein database curated at NCBI (all searches were conducted on or before February 5, 2015). For these analyses, the database of blastp was set to "non-redundant protein sequences (nr)" and limited to data from the Arthropoda "(taxid:6656)".

3. Results

3.1. In silico prediction of Acerentomon sp. neuropeptides

In this study, 36 distinct peptide families/subfamilies were searched for within the extant TSA sequence database for *Acerentomon* sp. (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 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).