General and Comparative Endocrinology 159 (2008) 67–79

Contents lists available at ScienceDirect

General and Comparative Endocrinology

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

In silico analyses of peptide paracrines/hormones in Aphidoidea

Andrew E. Christie *

Békésy Laboratory of Neurobiology, Pacific Biosciences Research Center, University of Hawaii at Manoa, 1993 East–West Road, Honolulu, HI 96822, USA
Center for Marine Functional Genomics, Mount Desert Island Biological Laboratory, P.O. Box 35, Old Bar Harbor Road, Salisbury Cove, ME 04672, USA

ARTICLE INFO

Article history:
Received 2 July 2008
Accepted 29 July 2008
Available online 7 August 2008

Keywords:
Aphid
Aphidoidea
Acyrthosiphon pisum
Aphis gossypii
Myzus persicae
Toxoptera citricida
B-type allatostatin
Crustacean cardioactive peptide (CCAP)
FMRFamide-related peptide
Insect kinin
Myosuppressin
Orcokinin
Periviscerokinin
Pheromone biosynthesis activating neuropeptide (PBAN)
Proctolin
Pyrokinin
Short neuropeptide F (sNPF)
SIFamide
Tachykinin-related peptide (TRP)
Expressed sequence tag (EST)
Neurohormone
Neuropeptide
Bioinformatics
Functional genomics

ABSTRACT

The Aphidoidea is an insect superfamily comprising most of the known aphid species. While small in size, these animals are of considerable economic importance as many members of this taxon are serious agricultural pests, inflicting physical damage upon crop plants and serving as vectors in the transmission of viral plant diseases. In terms of identifying the paracrines/hormones used to modulate behavior, particularly peptides, members of the Aphidoidea have largely been ignored, as it is not tractable to isolate the large pools of tissue needed for standard biochemical investigations. Here, a bioinformatics approach to peptide discovery has been used to overcome this limitation of scale. Specifically, *in silico* searches of publicly accessible aphidoidean ESTs were conducted to identify transcripts encoding putative peptides precursors, with the mature peptides contained within them deduced using peptide processing software and homology to known arthropod sequences. In total, 39 ESTs encoding putative peptides precursors were identified from four aphid species: *Acyrthosiphon pisum* (14 ESTs), *Aphis gossypii* (four ESTs), *Myzus persicae* (20 ESTs) and *Toxoptera citricida* (one EST). These precursors included ones predicted to encode isoforms of B-type allatostatin, crustacean cardioactive peptide, FMRFamide-related peptide (both myosuppressin and short neuropeptide F subfamilies), insect kinin, orcokinin, proctolin, pyrokinin/periviscerokinin/pheromone biosynthesis activating neuropeptide, SIFamide and tachykinin-related peptide. In total, 83 peptides were characterized from the identified precursors, most novel, including two B-type allatostatins possessing the variant $-WX_7$Wamide motif, two N-terminally extended proctolin isoforms and an N-terminally truncated and substituted SIFamide. Collectively, these results expand greatly the number of known/predicted aphid peptide paracrines/hormones, and provide a strong foundation for future molecular and physiological investigations of peptidergic control in this insect group.

1. Introduction

The Aphidoidea (aphids) are one of the most significant groups of agricultural pests, both for the direct damage they inflict upon crops plants during sap-feeding and for the viral diseases they transmit (van Emden and Harrington, 2007). While some aphid species appear to be host-specific, many are capable of feeding from a wide range of plants, making them the most important known vectors of viral-borne plant diseases (van Emden and Harrington, 2007).

An important component in aphid-borne viral transmission is the feeding strategy used by the vector (van Emden and Harrington, 2007). Specifically it has been suggested that different modes of feeding can influence the extent of virus transmission (van Emden and Harrington, 2007). In all multicellular animals, hormones are a common mechanism for modulating behavior (Kastin, 2006). In arthropods, peptide hormones are known to play critical roles in the control and modulation of many behaviors (Kastin, 2006), including feeding (e.g. Lee et al., 2004).

Despite the wealth of knowledge concerning peptide hormones in many insects, little is known about them in aphids. Undoubtedly, the small size of these insects is responsible, at least in part, for their limited study, as tissue from a vast number of individuals would be needed in order to conduct any comprehensive biochemical investigation. To overcome this limitation of scale, the present study has used a bioinformatics strategy for peptide discovery in

* Fax: +1 808 956 6984.
E-mail addresses: crabman@pbrc.hawaii.edu, achristi@mdibl.org.

0016-6480/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygcen.2008.07.022

the Aphidoidea. Specifically, *in silico* searches of the extensive aphidoidean EST database were conducted to identify aphid transcripts encoding putative neuropeptide precursors, with online peptide processing programs and homology to known arthropod sequences used to predict the mature peptides contained within them. In total, 83 peptide hormones were characterized using this strategy, expanding greatly the number of known/predicted peptides in this important group of agricultural pests, and providing a strong foundation for future investigations of peptidergic control in these animals.

2. Materials and methods

2.1. Database searches

Database searches were conducted using methods modified from several recent publications (Christie, 2008; Christie et al., 2008). Specifically, the online program tblastn [National Center for Biotechnology Information (NCBI), Bethesda, MD; http://www.ncbi.nlm.nih.gov/BLAST/] was used to mine for ESTs encoding putative aphid neuropeptide precursors via queries using known arthropod prepro-hormone sequences. For all searches, the program database was set to non-human, non-mouse ESTs (EST_others) and restricted to the Aphidoidea (taxid: 33385). All hits were fully translated (see Section 2.2) and checked manually for homology to the target query, as well as for typical peptide precursor features, including start and stop codons (i.e. a full-length prepro-hormone), the presence of a signal sequence and pro-hormone convertase processing sites. Fig. 1A shows an example transcript, here a putative B-type allatostatin-encoding EST, with it open reading frame translated and typical peptide precursor features highlighted. For each of the putative neuropeptide-encoding transcripts identified, the BLAST score and BLAST-generated *E*-value for significant alignment are provided in the appropriate subsection of the Results.

2.2. Peptide prediction

Translation of the nucleotide sequences of ESTs was performed using the Translate tool of ExPASy (Swiss Institute of Bioinformatics, Basel, Switzerland; http://www.expasy.ch/tools/dna.html). Signal peptide prediction was done via the online program SignalP 3.0, using both Neural Networks and Hidden Markov Models algorithms (Center for Biological Sequence Analysis, Technical University of Denmark, Lyngby, Denmark; http://www.cbs.dtu.dk/services/SignalP/; Bendtsen et al., 2004). Pro-hormone convertase cleavage sites were predicted based on the information presented in Veenstra (2000). Prediction of the sulfation state of Tyr residues was done using the online program Sulfinator (Swiss Institute of Bioinformatics; http://www.expasy.org/tools/sulfinator/; Monigatti et al., 2002). Where applicable, other post-translational modifications, e.g. cyclization of amino (N)-terminal Gln/Glu residues and carboxyl (C)-terminal amidation at Gly residues, were predicted by homology to known peptide isoforms. Fig. 1B shows the predicted prepro-hormone processing for the deduced B-type allatostatin EST illustrated in Fig. 1A.

3. Results

Using the sequences of known insect and decapod crustacean peptide precursors as queries, the Aphidoidea EST database at NCBI was searched for putative peptide-encoding prepro-hormone transcripts. The precursors queried for included those encoding adipokinetic hormone/red pigment concentrating hormone, A-type allatostatin, B-type allatostatin, C-type allatostatin, allatotropin, corazonin, crustacean cardioactive peptide (CCAP), members of the FMRFamide-related peptide (FaRP) superfamily (including the myosuppressin, neuropeptide F [NPF], short NPF [sNPF] and sulfakinin subfamilies), insect kinin, orcokinin, pigment dispersing hormone, proctolin, members of the pyrokinin/periviscerokinin/pheromone biosynthesis activating neuropeptide (PBAN) family, SIFamide and tachykinin-related peptide (TRP). In the interest of space, only those searches that identified putative precursors are described here, with the data presented in alphabetical order based on peptide family name. As Aphidoidean ESTs are continuously being added to the database, those added after February 10, 2008, the last date for which searches were conducted, would not be included in our study.

3.1. B-type allatostatin

Members of the B-type allatostatin family are characterized by the C-terminal motif $-WX_6$Wamide, where X_6 represents six variable amino acids located between the two Trp residues (Stay and Tobe, 2007). Using the sequence of a *Drosophila melanogaster* B-type allatostatin precursor (Accession No. AAK29381; Williamson et al., 2001) as a query, three aphid ESTs were identified as encoding putative prepro-B-type allatostatins, two from the green peach aphid *Myzus persicae* (Accession Nos. EE265136 [BLAST score, 68.6; *E*-value, 5e−12] and DW012587 [BLAST score, 56.6; *E*-value, 2e−08]) and one from the cotton aphid *Aphis gossypii* (Accession No. DR396650 [BLAST score, 69.3; *E*-value, 3e−12]).

Translation of the *M. persicae* EST EE265136 showed it to encode a 219 amino acid, putative full-length precursor protein (Fig. 1A). SignalP analysis of this sequence suggested the first 22 amino acids function as a signal peptide, with a cleavage locus located between Ser^{22} and Ile^{23}. The remaining sequence contains nine pro-hormone convertase target sites (all Lys-Arg), cleavage at which, followed by carboxypeptidase removal of the resulting C-terminal dibasic pairs and subsequent α-amidation at exposed Gly residues, would generate 10 peptides (listed in their order of appearance in the precursor; Fig. 1B and Table 1): IPESTIQAS SIKSSQTEQDNSDYTRSFDEDQEE, AWRDLQTAGWamide, GWQNLKT TWamide, AQDWQNLHSSWamide, QGWQKLHGGWamide, GWKD MQSGGWamide, FKDQPASGQLPQFDEYLDKYEDENPNDVE, SWDNF QGSWamide, AADWTSFRGSWamide and NPVDYMNEYSGYGDSDN YKAYIFPPGYNSYLPNFQTEYEK. Of these peptides, GWQNLKTTWamide, AQDWQNLHSSWamide, QGWQKLHGGWamide, SWDNFQGSWamide and AADWTSFRGSWamide, exhibit the $-WX_6$Wamide motif characteristic of known B-type allatostatins (Table 1). Two other peptides, AWRDLQTAGWamide and GWKDMQSGGWamide, are B-type-like, possessing $-WX_7$Wamide C-termini rather than the standard $-WX_6$Wamide motif (Table 1). Sulfinator analysis of FKDQPASGQLPQFDEYLDKYEDENPNDVE predicted both Tyr residues to be sulfated, resulting in the putative mature structure FKDQPASGQLPQFDEY$_{(SO3H)}$LDKY$_{(SO3H)}$EDENPNDVE for this peptide (Table 1). No sulfation was predicted in any other Tyr-containing sequence.

Translation of DW012587 revealed it to encode a 162 amino acid partial B-type allatostatin precursor (no start codon present), whose sequence was identical to the C-terminal end of EE265136, missing the signal peptide and the first peptide and dibasic site present in the EE265136 prepro-hormone. Predictions of the mature structures of the peptides derived from this deduced precursor would thus be identical to those of EE265136.

Translation of the *A. gossypii* EST DR396650 showed it to encode a 149 amino acid partial B-type allatostatin precursor, missing an unknown number of C-terminal residues. As with EE265136, SignalP analysis of this partial sequence suggested the first 22 amino acids form a signal peptide, with a cleavage locus located between