General and Comparative Endocrinology 237 (2016) 53–67

Contents lists available at ScienceDirect

General and Comparative Endocrinology

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

Research paper

Peptidergic signaling in the crab *Cancer borealis*: Tapping the power of transcriptomics for neuropeptidome expansion

Andrew E. Christie *, Micah G. Pascual

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

ARTICLE INFO

Article history:
Received 3 May 2016
Revised 26 July 2016
Accepted 2 August 2016
Available online 3 August 2016

Keywords:
Transcriptomics
Bioinformatics
Neuropeptide
Neurohormone
Crustacea
Decapoda

ABSTRACT

The crab *Cancer borealis* has long been used as a model for understanding neural control of rhythmic behavior. One significant discovery made through its use is that even numerically simple neural circuits are capable of producing an essentially infinite array of distinct motor outputs via the actions of locally released and circulating neuromodulators, the largest class being peptides. While much work has focused on elucidating the peptidome of *C. borealis*, no investigation has used *in silico* transcriptome mining for peptide discovery in this species, a strategy proven highly effective for identifying neuropeptides in other crustaceans. Here, we mined a *C. borealis* neural transcriptome for putative peptide-encoding transcripts, and predicted 200 distinct mature neuropeptides from the proteins deduced from these sequences. The identified peptides include isoforms of allatostatin A, allatostatin B, allatostatin C, CCHamide, crustacean cardioactive peptide, crustacean hyperglycemic hormone, diuretic hormone 31 (DH31), diuretic hormone 44 (DH44), FMRFamide-like peptide, GSEFLamide, HIGSLYRamide, insulin-like peptide (ILP), intocin, leucokinin, neuroparsin, pigment dispersing hormone, pyrokinin, red pigment concentrating hormone, short neuropeptide F and SIFamide. While some of the predicted peptides were known previously from *C. borealis*, most (159) are new discoveries for the species, *e.g.*, the isoforms of CCHamide, DH31, DH44, GSEFLamide, ILP, intocin and neuroparsin, which are the first members of these peptide families identified from *C. borealis*. Collectively, the peptides predicted here approximately double the peptidome known for *C. borealis*, and in so doing provide an expanded platform from which to launch new investigations of peptidergic neuromodulation in this species.

1. Introduction

The nervous systems of crustaceans, particularly members of the Decapoda, have been used for over half a century for investigating the neural control of physiology and behavior. One area where they have proven particularly useful is in elucidating the basic principles that govern the generation, maintenance and modulation of rhythmic behavior, *i.e.*, control of motor systems that produce behaviors such as walking, chewing and breathing in humans. The stomatogastric and cardiac neural circuits of decapods, which control the rhythmic movements of the foregut and heart musculatures, respectively, are arguably the best understood in the animal kingdom (for review see: Blitz and Nusbaum, 2011; Christie, 2011; Christie et al., 2010a; Cooke, 2002; Fénelon et al., 2003; Hooper and DiCaprio, 2004; Marder and Bucher, 2007; Marder et al., 1995; Nusbaum et al., 2001; Selverston, 2005; Selverston and Ayers, 2006; Selverston et al., 1998; Skiebe, 2001; Stein, 2009). Both of these systems consist of very small numbers of relatively large neurons (the stomatogastric network consists of ~25 neurons, depending on the species in question, with the cardiac ganglion containing just nine nerve cells), and this tractability for investigation at the cellular and systems levels has allowed for the connectivity between network elements to be fully worked out for several species. One of the major insights gained from work on the decapod stomatogastric and cardiac systems is that a simple "hardwired" neural circuit is capable of producing diverse motor output, *i.e.*, one circuit driving an essentially infinite array of distinct behaviors. While a number of factors undoubtedly contribute to this functional flexibility, much is attributable to the actions of neuromodulators, locally released and circulating substances that are capable of reconfiguring a neural network by changing the properties of the individual circuit elements. Several classes of neuromodulators have been identified in crustaceans, including small molecule transmitters, biogenic amines and diffusible gases (*e.g.*, Christie, 2011). However, the

* Corresponding author.
E-mail address: crabman@pbrc.hawaii.edu (A.E. Christie).

http://dx.doi.org/10.1016/j.ygcen.2016.08.002
0016-6480/

largest and most diverse group of these compounds is the peptides (*e.g.*, Christie, 2011; Christie et al., 2010a; Skiebe, 2001).

The brachyuran crab *Cancer borealis* is one of the most commonly used decapods for studies of neuromodulation (*e.g.*, Blitz and Nusbaum, 2011; Christie, 2011; Christie et al., 2010a; Fénelon et al., 2003; Hooper and DiCaprio, 2004; Marder and Bucher, 2007; Marder et al., 1995; Nusbaum et al., 2001; Selverston, 2005; Selverston and Ayers, 2006; Selverston et al., 1998; Skiebe, 2001; Stein, 2009). Not surprisingly, much work has focused on identifying and characterizing the native neuropeptides present in this species (*e.g.*, Christie et al., 1997; Cruz-Bermúdez et al., 2006; Dickinson et al., 2009a; Fu et al., 2005a, 2007; Huybrechts et al., 2003; Li et al., 2002, 2003; Ma et al., 2009a,b,c,d; Marder et al., 1986; Saideman et al., 2007; Stemmler et al., 2005, 2007a,b,c, 2010; Szabo et al., 2011; Weimann et al., 1993), with most studies using mass spectrometry, in particular accurate mass matching and/or de novo tandem mass spectrometric sequencing, for peptide discovery (*e.g.*, Cruz-Bermúdez et al., 2006; Dickinson et al., 2009a; Fu et al., 2005a, 2007; Huybrechts et al., 2003; Li et al., 2002, 2003; Ma et al., 2009a,b,c,d; Saideman et al., 2007; Stemmler et al., 2005, 2007a,b,c, 2010; Szabo et al., 2011). This work has identified a diverse neuropeptidome for *C. borealis*, which, prior to the study presented here, consisted of 150 or so distinct neuropeptides and encompassed approximately 20 different peptide families, *e.g.*, allatostatin A (AST-A), allatostatin B (AST-B), allatostatin C (AST-C), corazonin, crustacean cardioactive peptide (CCAP), crustacean hyperglycemic hormone (CHH)/molt-inhibiting hormone (MIH), crustacean hyperglycemic hormone precursor-related peptide (CPRP), FMRFamide-like peptide (FLP), HIGSLYRamide, leucokinin, myosuppressin, orcokinin, orcomyotropin, pigment dispersing hormone (PDH), proctolin, pyrokinin, red pigment concentrating hormone (RPCH), RYamide, short neuropeptide F (sNPF), SIFamide and tachykinin-related peptide (TRP).

While a powerful tool, peptide identification by mass spectrometry can, in some cases, be limited by the abundance of a peptide, its ionization efficiency, the post-translational modifications present in it, and/or by its length. For these "problematic" peptides, *in silico* prediction from high-throughput nucleotide sequence data offers an alternative means of discovery (*e.g.*, Christie et al., 2010a). Via this strategy, large peptidomes that include low abundance, difficult to ionize, heavily modified and/or long peptides have recently been predicted for a variety of species, including many crustaceans (*e.g.*, Bao et al., 2015; Christie, 2014a, 2014b, 2014c, 2014d, 2014e, 2014f, 2015a, 2016a, 2016b; Christie and Chi, 2015a; Christie et al., 2013, 2015; Gard et al., 2009; Ma et al., 2009e, 2010; Suwansa-Ard et al., 2015; Toullec et al., 2013; Veenstra, 2015; Ventura et al., 2014; Yan et al., 2012). Not surprisingly, the current *C. borealis* peptidome is missing peptides from families that might well be problematic to discover via mass spectral means. Thus, to complement and augment the extant peptidome for this species, we mined a publicly accessible *C. borealis* neural transcriptome (BioProject No. PRJNA310325; Schulz and Marder, unpublished direct GenBank submission) for sequences encoding putative neuropeptide precursors, and then used the proteins deduced from these transcripts to predict the mature structures of peptide isoforms.

As the data that follow will show, 25 transcripts encoding putative peptide precursors were identified within the *C. borealis* transcriptome shotgun assembly (TSA) dataset. The proteins deduced from these sequences allowed for the prediction of 200 distinct mature peptides, 159 being novel discoveries for this species. The identified peptides included members of 20 different families, seven of which, CCHamide, diuretic hormone 31 (DH31), diuretic hormone 44 (DH44), GSEFLamide, insulin-like peptide (ILP), intocin and neuroparsin, are new to *C. borealis*. The peptides predicted here approximately double the known *C. borealis* neuropeptidome, and in so doing provide an expanded platform for investigating peptidergic control of physiology and behavior in this important biomedical model.

2. Materials and methods

2.1. Database searches

Database searches were conducted on or before April 14, 2016, using methods modified from a well-vetted protocol (*e.g.*, Christie, 2008a,b, 2014a,b,c,d,e,f, 2015a,b, 2015c,d, 2016a,b; Christie and Chi, 2015a,b,c; Christie et al., 2008a, 2010b, 2011a,b, 2013, 2015; Gard et al., 2009; Ma et al., 2009e, 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 "*Cancer borealis* (taxid:39395)". Known 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 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, 2008b, 2014a,b,c,d,e,f, 2015a,b,c,d, 2016a,b; Christie and Chi, 2015a,b,c; Christie et al., 2008a, 2010b, 2011a,b; Christie et al., 2011c; Christie et al., 2013; Christie et al., 2015; Gard et al., 2009; Ma et al., 2009e, 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 4.1 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 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

Transcripts encoding 37 peptide families were searched for within the publicly accessible *C. borealis* TSA dataset (Table 1). In the interest of space, only those searches that resulted in the