1 **Prediction of *Scylla olivacea* (Crustacea; Brachyura) peptide**
2 **hormones using publicly accessible transcriptome shotgun assembly**
3 **(TSA) sequences**

4
5
6 Andrew E. Christie*

7
8
9 Békésy Laboratory of Neurobiology, Pacific Biosciences Research Center, School of Ocean and
10 Earth Science and Technology, University of Hawaii at Manoa, 1993 East-West Road, Honolulu,
11 HI 96822 USA

12
13
14 *Correspondence to: Dr. Andrew E. Christie, Békésy Laboratory of Neurobiology, Pacific
15 Biosciences Research Center, School of Ocean and Earth Science and Technology, University of
16 Hawaii at Manoa, 1993 East-West Road, Honolulu, HI 96822 USA. Phone: 808-956-5212;
17 FAX: 808-956-6984; Email: crabman@pbrc.hawaii.edu

18
19
20 **Key words:** transcriptomics; bioinformatics; neurohormone; neuropeptide
21

Abstract

The aquaculture of crabs from the genus *Scylla* is of increasing economic importance for many Southeast Asian countries. Expansion of *Scylla* farming has led to increased efforts to understand the physiology and behavior of these crabs, and as such, there are growing molecular resources for them. Here, publicly accessible *Scylla olivacea* transcriptomic data were mined for putative peptide-encoding transcripts; the proteins deduced from the identified sequences were then used to predict the structures of mature peptide hormones. Forty-nine pre/preprohormone-encoding transcripts were identified, allowing for the prediction of 187 distinct mature peptides. The identified peptides included isoforms of adipokinetic hormone-corazonin-like peptide, allatostatin A, allatostatin B, allatostatin C, bursicon β , CCHamide, corazonin, crustacean cardioactive peptide, crustacean hyperglycemic hormone/molt-inhibiting hormone, diuretic hormone 31, eclosion hormone, FMRFamide-like peptide, HIGSLYRamide, insulin-like peptide, intocin, leucokinin, myosuppressin, neuroparsin, neuropeptide F, orcokinin, pigment dispersing hormone, pyrokinin, red pigment concentrating hormone, RYamide, short neuropeptide F, SIFamide and tachykinin-related peptide, all well-known neuropeptide families. Surprisingly, the tissue used to generate the transcriptome mined here is reported to be testis. Whether or not the testis samples had neural contamination is unknown. However, if the peptides are truly produced by this reproductive organ, it could have far reaching consequences for the study of crustacean endocrinology, particularly in the area of reproductive control. Regardless, this peptidome is the largest thus far predicted for any brachyuran (true crab) species, and will serve as a foundation for future studies of peptidergic control in members of the commercially important genus *Scylla*.

1. Introduction

Scylla is a genus of portunid (swimming) crabs that are common throughout the Indo-Pacific region. Four species are generally recognized within the genus, *Scylla olivacea*, *Scylla paramamosain*, *Scylla serrata* and *Scylla tranquebarica*, all of which share a common range and are fished commercially; the global wild fishery of *Scylla* species was ~38,000 tons in 2013 (Species Fact Sheet, Fisheries and Aquaculture Department, Food and Agriculture Organization of the United Nations; <http://www.fao.org/fishery/species/2637/en> [accessed January 11, 2016]). Due to habitat loss/degradation and increasing demand for product, there has been growing commercial culture of *Scylla* species in Southeast Asia since the early 1970s. Early farming initiatives focused on low-density natural pond polyculture (typically with either fish or shrimp) using tidally introduced wild seed, with farming today employing high-density artificial monoculture; in 2013, the total global aquaculture of *Scylla* species was estimated to be 180,000 tons (*Scylla serrata* Fact Sheet, Cultured Aquatic Species Information Programme, Fisheries and Aquaculture Department, Food and Agriculture Organization of the United Nations; http://www.fao.org/fishery/culturedspecies/Scylla_serrata/en [accessed January 11, 2016]).

As part of the ongoing effort to support and enhance the aquaculture of *Scylla* species, much work has focused on understanding the biology of these crabs (*e.g.*, Bao et al., 2015; Chen et al., 2015; Girish et al., 2015; Gong et al., 2015; Huang et al., 2014, 2015; Ikhwanuddin et al., 2014; Kornthong et al., 2014; Sun et al., 2015; Wang et al., 2015a, 2015b; Xie et al., 2014; Zeng et al., 2016; Zhao et al., 2015). One outcome of these studies has been the development of extensive transcriptomic resources for several *Scylla* species (*e.g.*, Bao et al., 2015; Kornthong et al., 2014; Ma et al., 2014; Xie et al., 2014), including *S. olivacea*, for which approximately 160,000 transcriptome shotgun assembly (TSA) sequences are now available in GenBank (*e.g.*,

BioProject Accession No. PRJNA289610; M. Ikhwanuddin, unpublished direct GenBank submission). Here, the *S. olivacea* TSA dataset was mined for putative peptide-encoding transcripts; peptides are the largest and most diverse class of hormones and are involved in regulating essentially all aspects of physiology and behavior (*e.g.*, Kastin, 2006), including reproduction, feeding and growth in crustaceans (*e.g.*, Christie et al., 2010a; Chung et al., 2010), all key aspects of *S. olivacea* biology that must be understood in order to increase production of this crab in aquaculture.

2. Materials and methods

2.1. Database searches

Database searches were conducted on or before January 11, 2016, using methods modified from a well-vetted protocol (*e.g.* Christie, 2008a, 2008b, 2014a, 2014b, 2014c, 2014d, 2014e, 2014f, 2015a, 2015b, 2015c, 2015d; Christie and Chi, 2015a, 2015b, 2015c; Christie et al., 2008a, 2010b, 2011a, 2011b, 2013, 2015; 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 *S. olivacea* “(taxid:85551)”. 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, 2014b, 2014c, 2014d, 2014e, 2014f, 2015a, 2015b, 2015c, 2015d; Christie and Chi, 2015a, 2015b, 2015c; Christie et al., 2008a, 2010b, 2011a, 2011b 2011c, 2013, 2015; 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. Figure 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).

3. Results

Thirty-seven distinct peptide families/subfamilies were searched for within the publicly accessible *S. olivacea* TSA dataset (Table 1). In the interest of space, only those searches that resulted in the identification of putative precursor-encoding transcripts are described here, 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-corazonin-like peptide (ACP)

Three *S. olivacea* transcripts were identified as encoding putative ACP precursors (Table 1). Translation of one transcript revealed a 101 amino acid full-length precursor protein (*Scylla olivacea* [Scyol]-prepro-ACP I; Fig. 1A), while translation of the other two sequences revealed identical 74 amino acid full-length preprohormones (Scyol-prepro-ACP II; Fig. 2A). Scyol-prepro-ACP I and II are identical in amino acid sequence over much of their length, differing only at their C-termini (alignment not shown). Four peptides, each distinct in structure, were predicted from Scyol-prepro-ACP I (Table 2 and Fig. 1A), one of which, pQITFSRSWVPQamide, is identical in structure to an isoform of ACP predicted previously from crayfish *Procambarus clarkii* (Christie and Chi, 2015c) and the lobster *Homarus americanus* (Christie et al., 2015). Three distinct peptides were predicted from Scyol-prepro-ACP II (Table 2 and Fig. 2A), two of which, the ACP isoform pQITFSRSWVPQamide and the linker/precursor-related peptide SSSPGDIPEPLDPCRDA, are shared with Scyol-prepro-ACP I.

The remaining linker/precursor-related peptide present in Scyol-prepro-ACP II is distinct in structure from the others derived from Scyol-prepro-ACP I.

3.2. *Allatostatin A (AST-A)*

Two transcripts were identified as encoding putative AST-A precursors (Table 1). Translation of these sequences revealed each to encode an internal fragment of a preprohormone, with one partial protein being 130 amino acids in length and the other 99 amino acids long. Given that there is no region of overlap between the two fragments, they are considered derived from distinct proteins and were named Scyol-prepro-AST-A I (Fig. 2B1) and Scyol-prepro-AST-A II (Fig. 2B2), respectively. Twelve peptides (ten full-length and two partial), all distinct in structure (Table 2 and Fig. 2B1), were predicted from Scyol-prepro-AST-A I. Of these peptides, six possess the C-terminal motif –YXFGLamide (where *X* represents a variable residue), the hallmark of the AST-A family (Christie et al., 2010a). Six peptides (five full-length and one partial), each distinct from one another and from those derived from Scyol-prepro-AST-A I, were predicted from Scyol-prepro-AST-A II (Table 2 and Fig. 2B2); these predicted peptides included four isoforms of AST-A and two linker/precursor-related sequences.

3.3. *Allatostatin B (AST-B)*

Two transcripts were identified as encoding putative AST-B precursors (Table 1). Translation of these sequences revealed one to encode a 184 amino acid full-length preprohormone (Scyol-prepro-AST-B I; Fig. 2C1) and the other a 298 amino acid N-terminal partial precursor (Scyol-prepro-AST-B II; Fig. 2C2); the first 167 amino acids of these two proteins are identically conserved (alignment not shown). Eleven distinct peptides were

predicted from Scyol-prepro-AST-B I (Table 2 and Fig. 2C1), five of which possess –
WX₆Wamide C-termini (where X₆ represents six variable amino acids), the hallmark of the AST-
B family (Christie et al., 2010a). Eighteen distinct peptides were predicted from Scyol-prepro-
AST-B I (17 full-length and one partial), eight being isoforms of AST-B (Table 2 and Fig. 2C2).
Scyol-prepro-AST-B I and II have ten peptides in common, five isoforms of AST-B and five
linker/precursor-related sequences.

3.4. *Allatostatin C (AST-C)*

A single transcript was identified as encoding a putative AST-C precursor (Table 1).
Translation of this sequence revealed a 149 amino acid full-length preprohormone (Scyol-
prepro-AST-C; Fig. 1B). Six distinct peptides were predicted from Scyol-prepro-AST-C (Table
2 and Fig. 1B), one of which, pQIRYHQCYFNPISCF (disulfide bridging between the two
cysteine residues), is a known and broadly conserved decapod AST-C isoform (*e.g.*, Stemmler et
al., 2010).

3.5. *Bursicon β*

A single transcript was identified as encoding a putative bursicon β precursor (Table 1).
Translation of this transcript revealed a 106 amino acid C-terminal partial protein (Scyol-pre-
bursicon β; Fig. 2D). This partial sequence appears to be the C-terminus of an isoform of
bursicon β; the extant portion of the peptide is 91% identical in amino acid sequence to that of an
isoform of bursicon β from the green crab *Carcinus maenas* (Wilcockson and Webster, 2008).

3.6. *CCHamide*

A single transcript was identified as encoding a putative CCHamide precursor (Table 1). Translation of this transcript revealed a 94 amino acid N-terminal partial protein (Scyol-prepro-CCHamide; Fig. 2E). Three distinct peptides (two full-length and one partial) were predicted from Scyol-prepro-CCHamide (Table 2 and Fig. 2E), one of which, HGSVKVGCLNYGHSLGAHamide (disulfide bridging between the two cysteines), possesses the structural features of an N-terminally extended CCHamide family member (Hansen et al., 2011), *i.e.*, the consensus motif XCXXW/Y/FGXXCXGXHamide (disulfide bridging between the position 2 and 9 cysteine residues).

3.7. *Corazonin*

Two transcripts were identified as encoding putative corazonin precursors (Table 1). Translation of these sequences revealed each to encode an identical 117 amino acid full-length precursor protein (Scyol-prepro-corazonin; Fig. 2F). Six distinct peptides were predicted from Scyol-prepro-corazonin (Table 2 and Fig. 2F), including pQTFQYSRGWTNamide, a known and broadly conserved decapod isoform of corazonin (*e.g.*, Christie et al., 2015; Ma et al., 2009, 2010).

3.8. *Crustacean cardioactive peptide (CCAP)*

A single transcript was identified as encoding a putative CCAP precursor (Table 1). Translation of this transcript revealed a 141 amino acid full-length prehormone (Scyol-prepro-CCAP; Fig. 2G). Six distinct peptides were predicted from Scyol-prepro-CCAP (Table 2 and Fig. 2G), one of which, PFCNAFTGCamide (disulfide bridging between the two cysteine residues), is identical in structure to authentic CCAP (Christie et al., 2010a).

3.9. Crustacean hyperglycemic hormone (CHH) superfamily

The CHH superfamily can be subdivided into two subgroups based on peptide structure and on the organization of the precursor proteins from which they are cleaved: the CHHs proper and the molt-inhibiting (MIH)/gonad-inhibiting /vitellogenesis-inhibiting (VIH)/mandibular organ-inhibiting hormone subfamily, commonly referred to as the MIH subgroup (*e.g.*, Böcking et al., 2002; Chan et al., 2003; Chung et al., 2010; Fanjul-Moles, 2006). Members of the CHH subfamily are typically 70-72 amino acids long and possess six identically placed cysteine residues (which allow for the formation of three disulfide bridges, one between the first and fifth cysteines, one between the second and fourth cysteines and one between the third and sixth cysteines). In addition the preprohormones from which members of the CHH subgroup are generated contain an additional peptide, commonly referred to as CHH precursor-related peptide or CPRP, between the signal sequence and the CHH isoform. In contrast, members of the MIH subfamily are typically longer and their precursors lack a CPRP; the arrangement the cysteine residues in MIHs, and their patterns of disulfide bridging, are similar, though not always identical to those of the CHHs proper.

3.9.1. CHH subfamily

Four transcripts were identified as encoding putative CHH subfamily precursors. Translation of two of these transcripts revealed 139 and 132 amino acid full-length preprohormones (Table 1). These two proteins are identical in sequences with the exception of a seven amino acid N-terminal extension present in the longer precursor (contained within the signal peptide) and a single conservative amino acid substitution in their prohormone portions.

Given their extensive amino acid conservation, these two proteins have been named Scyol-prepro-CHH Ia (the longer precursor; Fig. 2H1) and Scyol-CHH Ib (the shorter sequence; Fig. 2H2). Two peptides were predicted from each CHH I preprohormone (Table 2 and Fig. 2H1-2), an isoform of CPRP and an isoform of CHH. The CPRPs differ at a single residue (position 8), an arginine in the peptide produced from Scyol-prepro-CHH Ia vs. a glutamine in the CPRP produced from Scyol-prepro-CHH Ib; the CHHs derived from Scyol-prepro-CHH Ia and Ib are identical in structure. Analysis of the CHH using DiANNA suggest disulfide bridges are present between the first and fifth, second and fourth, and third and sixth cysteines in the peptide. It should be noted that Scyol-prepro-CHH Ia is essentially identical in sequence to a CHH precursor previously identified from *S. olivacea* (i.e., **Accession No. ABP88270**; Tsai et al., 2008); the two proteins differ at just a single position, namely a serine for threonine substitution in what would be position 20 of their respective CPRPs.

Translation of the remaining two CHH subfamily-encoding transcripts revealed 116 and 165 amino acid full-length proteins, which are identical in sequence save a 49 amino acid N-terminal extension present in the longer protein. Interestingly, this N-terminal addition generates a situation where no signal peptide is detected in the longer sequence using SignalP. Whether this N-terminal addition is an assembly artifact or is part of the true protein (perhaps some sort of regulatory element) is unknown. Regardless, the two proteins are presented here with the shorter (which does contain a signal sequence) named Scyol-prepro-CHH IIa (Fig. 2H3) and the longer named Scyol-prepro-CHH IIb (Fig. 2H4). The two precursors would generate an identical set of peptides (Table 2 and Fig. 2H3-4), one CPRP and one CHH. DiANNA analysis suggests disulfide bridges between the first and fifth, second and fourth, and third and sixth cysteines in the CHH isoform. It should be noted that Scyol-prepro-CHH IIa is nearly identical to a CHH

precursor identified from *S. paramamosain* (**Accession No. ALQ28582**; Bao et al., 2015). Like Scyol-prepro-CHH IIa, this *S. paramamosain* protein also lacks a signal peptide (as predicted using SignalP), and as all of the variation (five positions) between the two sequences save one occur prior to the CPRP, the peptides predicted from the two precursors are identical (their CPRPs) or nearly so (one substituted residue in the two CHHs).

3.9.2. Molt-inhibiting hormone (MIH) subfamily

A single transcript was identified as encoding a putative MIH subfamily precursor (Table 1). Translation of this sequence revealed a 125 amino acid full-length pre-hormone (Scyol-pre-MIH; Fig. 2H5). A single isoform of MIH was predicted from Scyol-pre-MIH (Table 2 and Fig. 2H5). DiANNA analysis of the MIH isoform produced from Scyol-prepro-MIH suggests disulfide bridging between the first and fifth, second and fourth, and third and sixth cysteines in this peptide. It should be noted that Scyol-pre-MIH is identical in sequence to a putative VIH precursor recently cloned from *S. paramamosain* (**Accession No. AHE40786**; C. Liu, X. Wang, X. Zhang, Z. Zhang and Y. Wang, unpublished direct GenBank submission), hence the MIH/VIH isoforms they contain are also identical in structure.

3.10. Diuretic hormone 31 (DH31)

A single transcript was identified as encoding a putative DH31 precursor (Table 1). Translation of this transcript revealed a 98 amino acid C-terminal partial protein (Scyol-prepro-DH31; Fig. 2I). Four distinct peptides (three full-length and one partial) were predicted from Scyol-prepro-DH31 (Table 2 and Fig. 2I), including GLDLGLGRGFSGSQAAKHLMGLAAANYAGGPamide, which possesses the structural

feature characteristic of DH31 family members, *i.e.*, an overall length of 31 amino acids, an amidated C-terminus, and 13 highly conserved residues, *i.e.*

XXDXGLXRGXSGXXXAKXXXXXXXXANXXXGPamide (Christie et al., 2010a).

3.11. Eclosion hormone (EH)

A single transcript was identified as encoding a putative EH precursor (Table 1). Translation of this sequence revealed a 93 amino acid full-length prehormone (Scyol-pre-EH; Fig. 2J). A single peptide was predicted from Scyol-pre-EH (Table 2 and Fig. 2J). This EH isoform, while N-terminally extended by 13 amino acids, is highly similar to a recently predicted EH from the lobster *Homarus americanus* (Christie et al., 2015); the two EHs are 91% similar/71% identical in amino acid composition in their region of overlap (alignment not shown). Analysis of the *S. olivacea* EH by DiANNA suggests disulfide bridging between its first and fourth, second and third, and fifth and sixth cysteines.

3.12. FMRFamide-like peptide (FLP)

A single transcript was identified as encoding a putative FLP precursor (Table 1). Translation of this sequence revealed a 251 amino acid protein (Scyol-prepro-FLP; Fig. 2K). While bounded on both its N- and C-terminal ends by stop codons, and with a potential start methionine near its N-terminus, SignalP predicted no signal peptide for Scyol-prepro-FLP. Thus, it appears that a portion of the transcript encoding the N-terminus is missing from the assembled sequence. Thus, Scyol-prepro-FLP, while bounded by stop codons is treated here as a C-terminal partial protein. Fourteen peptides were predicted from the extant portion of Scyol-prepro-FLP (Table 2 and Fig. 2K), eight of which possess –F/YLRFamide C-termini, motifs

common in crustacean FLP family members (Christie et al., 2010a). Six of the FLP isoforms are encoded as single copies in Scyol-prepro-FLP, with two copies of DRNFLRFamide predicted from this protein. Each of the linker/precursor-related peptides predicted from Scyol-prepro-FLP is distinct in structure.

3.13. *HIGSLYRamide*

A single transcript was identified as encoding a putative HIGSLYRamide precursor (Table 1). Translation of this sequence revealed a 316 amino acid internal protein fragment (Scyol-prepro-RYamide; Fig. 2L). Twenty-four peptides (23 full-length and one partial) were predicted from the extant portion of Scyol-prepro-RYamide (Table 2 and Fig. 2L), including nine copies of HIGSLYRamide and one copy of a position 3-substituted isoform, *i.e.*, HIVSLYRamide. Eleven of the 13 linker/precursor-related peptides derived from Scyol-prepro-RYamide are present as single copies; the partial preprohormone contains two copies of the peptide DEDNSQNLSed.

3.14. *Insulin-like peptide (ILP)*

A single transcript was identified as encoding a putative ILP precursor (Table 1). Translation of this transcript revealed a 181 amino acid protein (Scyol-prepro-ILP; Fig. 2M). Like the FLP precursor described in 3.12, Scyol-prepro-ILP is bounded on both its N- and C-terminal ends by stop codons, and no signal peptide was predicted by SignalP starting with the first methionine in the protein. However, when the full sequence was used, a signal peptide was predicted. Thus, it appears that a portion of the transcript encoding a section of the N-terminus is missing from the assembled sequence, and it is treated here as a C-terminal partial protein. Four

distinct peptides were predicted from Scyol-prepro-ILP (Table 2 and Fig. 2M), including two that show structural similarity to previously identified crustacean ILP A- and B-chain isoforms, *e.g.*, those deduced from the lobster *H. americanus* preprohormone (Christie et al., 2015) that was used as the query sequence for their identification. Homology to previously described crustacean ILPs (*e.g.*, Dircksen et al., 2011) suggests disulfide bridging to be present within the *S. olivacea* A-chain peptide (the first and third cysteines bridged in GLSAECCRKACSVSELAGYCY; Table 2), and between it and the predicted B-chain peptide (the first cysteine in LCGWRLANELNRVCKGVYNMPTVSSNALFYLKARGamide bridged to the second cysteine in the A-chain and the second cysteine in this peptide bridged to the fourth cysteine in the A-chain; these interpeptide bridged cysteines are not highlighted in the peptide structures shown in Table 2).

3.15. Intocin

A single transcript was identified as encoding a putative intocin precursor (Table 1). Translation of this sequence revealed a 58 amino acid full-length preprohormone (Scyol-prepro-intocin; Fig. 2N). Three peptides were predicted from Scyol-prepro-intocin (Table 2 and Fig. 2N), including one, CFITNCPPGamide (disulfide bridging between the two cysteines), which is identical in structure to a broadly conserved decapod intocin isoform (Christie, 2014f; Christie and Chi, 2015c; Christie et al., 2015).

3.16. Leucokinin

A single transcript was identified as encoding a putative leucokinin precursor (Table 1). Translation of this sequence revealed a 356 amino acid C-terminal partial protein (Scyol-prepro-

leucokinin; Fig. 2O). Twenty-seven peptides (26 full-length and one partial) were predicted from the extant portion of Scyol-prepro-leucokinin (Table 2 and Fig. 2O), 13 of which possess the C-terminal motif –FXXWA/Gamide, the hallmark of leucokinin family members (Christie et al., 2010a). Of the 13 leucokinins, seven are present as single copies within Scyol-prepro-leucokinin, with four copies of pQAFSAWAamide and two copies of pQGFSAWAamide also present in this partial protein. Similarly, 12 of the 14 linker/precursor-related sequences derived from Scyol-prepro-leucokinin are present as single copies, with two copies of the peptide SEDE also contained within this protein.

3.17. *Myosuppressin*

A single transcript was identified as encoding a putative myosuppressin precursor (Table 1). Translation of this sequence revealed a 110 amino acid protein (Scyol-prepro-myosuppressin; Fig. 2P). As with the FLP and ILP proteins described in sections 3.12 and 3.14, respectively, it appears that a portion of the transcript encoding the N-terminus of Scyol-prepro-myosuppressin did not assemble properly, and the protein is treated here as a C-terminal partial sequence. Three distinct peptides were predicted from the extant portion of Scyol-prepro-myosuppressin (Table 2 and Fig. 2P), one of which, pQDLDHVFLRFamide, is identical in structure to a known and broadly conserved decapod myosuppressin isoform (e.g., Stemmler et al., 2007).

3.18. *Neuroparsin*

Five transcripts were identified as encoding putative neuroparsin precursors (Table 1). Translation of three of the transcripts revealed full-length proteins of 95, 101 and 118 amino

acids; these proteins differ from one another only in their signal peptides. As such, they have been named here Scyol-pre-neuroparsin Ia (Fig. 2Q1), Scyol-pre-neuroparsin Ib (Fig. 2Q2) and Scyol-pre-neuroparsin Ic (Fig. 2Q3), respectively. The neuroparsin isoform derived from these three precursors is 74 amino acids long, and is predicted by DiANNA to possess disulfide bridges between its first and twelfth, second and fourth, third and eighth, fifth and sixth, seventh and tenth, and ninth and eleventh cysteine residues (Table 2 and Fig. 2Q1-3). Translation of a fourth transcript revealed a 102 amino acid full-length protein (Scyol-pre-neuroparsin II; Fig. 2Q4). Like the neuroparsin predicted from Scyol-pre-neuroparsin Ia-c, the isoform predicted from Scyol-pre-neuroparsin II is 74 amino acid long, but possesses a distinct amino acid sequence and predicted pattern of disulfide bridges, *i.e.*, bonds between the first and second, third and sixth, fourth and eleventh, fifth and seventh, eighth and tenth, and ninth and twelfth cysteines (Table 2). Translation of the final transcript revealed a 106 amino acid full-length protein (Scyol-pre-neuroparsin III; Fig. 2Q5). The neuroparsin predicted from Scyol-pre-neuroparsin III is 77 amino acids long, with bonding predicted by DiANNA between its first and twelfth, second and fourth, third and eighth, fifth and eleventh, sixth and seventh, and ninth and tenth cysteines (Table 2).

3.19. Neuropeptide F (NPF)

Two transcripts were identified as encoding putative NPF precursors (Table 1). Translation of one of the transcripts revealed a 102 amino acid full-length precursor (Scyol-prepro-NPF; Fig. 2R), while translation of the other revealed an 84 amino acid N-terminal partial protein; the partial protein is identical to the N-terminus of Scyol-prepro-NPF. Three distinct peptide were predicted from Scyol-prepro-NPF (Table 2 and Fig. 2R), one of which,

YLDALNAYFAIAGRPRFamide, shows structural similarity to members of the NPF family (Christie et al., 2010a), *i.e.*, it possess the C-terminal motif –GRPRFamide and contain tyrosine residues at positions 10 and 17 from the C-termini, though it is shorter than stereotypical family members (17 rather than ~30 amino acids).

3.20. *Orcokinin*

Two transcripts were identified as encoding putative orcokinin precursors (Table 1). Translation of one sequence revealed a protein of 145 amino acids (Scyol-prepro-orcokinin I; Fig. 2S1), while translation of the other revealed a 94 amino acids long protein (Scyol-prepro-orcokinin II; Fig. 2S2). Like FLP-, ILP- and myosuppressin-containing proteins described in sections 3.12, 3.14, and 3.17, respectively, it appears that a portion of each of the orcokinin-encoding transcripts did not assemble properly at what would be the N-terminus of each of the encoded preprohormones, and thus they are treated here as an internal protein fragment (Scyol-prepro-orcokinin I) and a C-terminal partial precursor (Scyol-prepro-orcokinin II). Nine peptides (seven full-length and two partial) were predicted from the extant portion of Scyol-prepro-orcokinin I (Table 2 and Fig. 2S1), including eight (seven full-length and one partial) that possess the structural hallmarks of decapod orcokinin family members, *i.e.*, the N-terminal motif NFDEIDR–, an unamidated C-terminus, and an overall length of 13 amino acids (Christie et al., 2010a). The orcokinins predicted from Scyol-prepro-orcokinin I include one copy of NFDEIDRSSFGFA, two copies of NFDEIDRSSFGFN, four copies of NFDEIDRSSFGFN and the partial peptide NFDEIDR+. Five peptides (four full-length and one partial) were predicted from the extant portion of Scyol-prepro-orcokinin II (Table 2 and Fig. 2S2). The first two peptides contained within this partial precursor (the linker/precursor-related peptide

+SWEGKGLSASLYLTLCSSSSASPRTGFGHS and the orcokinin NFDEIDRSSFGFA) are shared with Scyol-prepro-orcokinin I, the remaining peptides, the orcokinins NFDEIDRSSFGFV and NFDEIDRSGFGFV and the linker/precursor-related sequences MLTPRDLANLY and NAE, are unique to Scyol-prepro-orcokinin II.

3.21. Pigment dispersing hormone (PDH)

Six transcripts were identified as encoding putative PDH precursors (Table 1). Translation of two of the transcripts revealed identical 88 amino acid full-length proteins (Scyol-prepro-PDH I; Fig. 2T1). Two distinct proteins were predicted from Scyol-prepro-PDH I (Table 2 and Fig. 2T1), one of which, NSELINSILGLPKVMNDAamide, is identical in structure to authentic β -PDH, a broadly conserved decapod peptide (Christie et al., 2010a). Translation of the remaining transcripts revealed two to encode 59 amino acid proteins and the others 89 amino acid proteins. Like several of the other precursors described earlier, these sequences, while bounded on both sides by stop codons, appear to be missing portions of their N-termini, presumably do to issues in transcript assembly, and as such are treated here as C-terminal partial proteins. All for partial proteins are similar, though distinct in structure and have been named Scyol-prepro-PDH IIa (Fig. 2T2), Scyol-prepro-PDH IIb (Fig. 2T3), Scyol-prepro-PDH IIc (Fig. 2T4) and Scyol-prepro-PDH IId (Fig. 2T5). Each of these partial precursors is predicted to produce two peptides, a partial linker/precursor-related sequence and an isoform of PDH. The partial linker/precursor-related peptides predicted from Scyol-prepro-PDH IIa-d are each distinct in structure (Table 2 and Fig. 2T2-5), with Scyol-prepro-PDH IIa, IIb and IIc sharing the PDH isoform NSELINLLGLSALMNEAamide. The PDH derived from Scyol-prepro-PDH IId,

NSELINSLGLSPLMNEAamide possesses a proline for alanine substitution at position 13 relative to the isoform predicted from Scyol-prepro-PDH IIa, IIb and IIc.

3.22. *Pyrokinin*

A single transcript was identified as encoding a putative pyrokinin precursor (Table 1). Translation of this sequence revealed a 146 amino acid C-terminal partial protein (Scyol-prepro-pyrokinin; Fig. 2U). Ten peptides (nine full-length and one partial) were predicted from the extant portion of Scyol-prepro-pyrokinin (Table 2 and Fig. 2U), eight of which possess the C-terminal motif –FXPRXamide, the hallmark of the pyrokinin family (Christie et al., 2010a). Six of the pyrokinin isoforms are present in the extant portion of Scyol-prepro-pyrokinin as single copies, while SNFAFAPRPamide is present twice in the partial protein. Both of the linker/precursor-related peptides derived from Scyol-prepro-pyrokinin are distinct in structure.

3.23. *Red pigment concentrating hormone (RPCH)*

A single transcript was identified as encoding a putative RPCH precursor (Table 1). Translation of this sequence revealed a 117 amino acid N-terminal partial protein (Scyol-prepro-RPCH; Fig. 2V). Two distinct peptides were predicted from the extant portion of Scyol-prepro-RPCH (Table 2 and Fig. 2V), one of which, pQLNFSPGWamide, is identical in structure to authentic RPCH (Christie et al., 2010a).

3.24. *RYamide*

A single transcript was identified as encoding a putative RYamide precursor (Table 1). Translation of this sequence revealed a 97 amino acid protein (Scyol-prepro-RYamide; Fig. 2W).

Like several of the other precursors described earlier, this sequence, while bounded on both sides by stop codons, appear to be missing a portion of its N-terminus, likely do to the transcript assembly process, and thus is treated here as a C-terminal partial protein. Five distinct peptides (four full-length and one partial) were predicted from the extant portion of Scyol-prepro-RYamide (Table 2 and Fig. 2W), two of which, SGFYANRYamide and FVGGSRYamide, are known decapod RYamide family members (*e.g.*, Ma et al., 2009). It should be noted that an atypical dibasic pairing (HX₂R, where X₂ represents two variable residues) is proposed for the cleavage of SGFYANRYamide from Scyol-prepro-RYamide.

3.25. Short neuropeptide F (sNPF)

Two transcripts were identified as encoding putative sNPF precursors (Table 1). Translation of these sequences revealed one to encode a 154 amino acid N-terminal partial protein, and the other a 121 amino acid N-terminal partial precursor. With the exception of a 33 amino acid insertion present in the longer protein, and a single substituted residue, the two partial preprohormones are identical in sequence and have been named Scyol-prepro-sNPF Ia (Fig. 2X1) and Scyol-prepro-sNPF Ib (Fig. 2X2), respectively. Six peptides (five full-length and one partial) were predicted from the extant portion of Scyol-prepro-sNPF Ia (Table 2 and Fig. 2X1), three of which possess the C-terminal motif –RLRFamide, the hallmark of the sNPF family (Christie et al., 2010a). Four peptides were predicted from the extant portion of Scyol-prepro-sNPF Ib (Table 2 and Fig. 2X2), two isoforms of sNPF and two linker/precursor-related sequences, all but the final partial linker/precursor-related peptide being shared with Scyol-prepro-sNPF Ia.

3.26. *SIFamide*

A single transcript was identified as encoding a putative SIFamide precursor (Table 1). Translation of this sequence revealed a 78 amino acid full-length preprohormone (Scyol-prepro-SIFamide; Fig. 2Y). Two peptides were predicted from Scyol-prepro-SIFamide (Table 2 and Fig. 2Y), one of which, GYRKPPFNGSIFamide, is a broadly conserved decapod SIFamide isoform (*e.g.*, Stemmler et al., 2007).

3.27. *Tachykinin-related peptide (TRP)*

Two transcripts were identified as encoding putative TRP precursors (Table 1). Translation of one sequence revealed a 103 amino acid N-terminal partial protein (Scyol-prepro-TRP Ia; Fig. 2Z1), while translation of the other revealed a 131 amino acid N-terminal partial precursor (Scyol-prepro-TRP Ib; Fig. 2Z2). These two proteins are identical in sequence with the exception of a 28 amino acid N-terminal extension in Scyol-prepro-TRP Ib, an addition that generates a situation where no signal peptide is detected in it using SignalP. Regardless, a common set of five peptides (four full-length and one partial) were predicted from Scyol-prepro-TRP Ia and Scyol-prepro-TRP Ib (Table 2 and Fig. 2Z1-2), two copies of APSGFLGMRamide, a well known and broadly conserved decapod TRP (*e.g.*, Stemmler et al., 2007), and two distinct linker/precursor-related peptides (one full-length and one partial).

4. Discussion

4.1. *Prediction of peptidome for Scylla olivacea*

In silico transcriptome mining has become a popular and highly effective method for peptide discovery in crustaceans (*e.g.*, Bao et al., 2015, Christie, 2014a, 2014b, 2014c 2014d,

2014e, 2014f, 2015b; Christie and Chi, 2015c; Christie et al., 2008a, 2010b, 2013, 2015; Gard et
 al., 2009; Ma et al., 2009, 2010; Toullec et al., 2013, Veenstra, 2015; Ventura et al., 2014; Yan et
 al., 2012). When a large transcriptome is available for a species, BLAST searches using known
 peptide precursors as the input queries can be used to rapidly identify transcripts encoding
 putative homologs. Translation of these transcripts allows for the prediction of the amino acid
 sequences of the proteins they encode, and when the deduced proteins are subjected to a simple
 bioinformatics workflow employing freeware and homology to known crustacean
 pre/preprohormone processing schemes, the structures of the mature peptides putatively liberated
 from the pre/preprohormones can be predicted. Via this strategy, large peptidomes have recently
 been predicted for a number of decapod crustaceans, for example, the 194 distinct peptides
 recently predicted for the American lobster *H. americanus* (Christie et al., 2015). Here, *in silico*
 transcriptome mining was used to predict a peptidome for the brachyuran crab *S. olivacea* using
 a recently deposited and publicly accessible transcriptome (**BioProject Accession No.**
PRJNA289610; M. Ikhwanuddin, unpublished direct GenBank submission). Forty-nine
 transcripts encoding putative peptide precursors were identified with the ~160,000 sequences
 available for *S. olivacea*. These transcripts allowed for the prediction of 187 distinct peptides for
 this species. The predicted peptides included members of the ACP (one isoform), AST-A (ten
 isoforms), AST-B (eight isoforms), AST-C (one isoform), bursicon β (one isoform), CCHamide
 (one isoform), corazonin (one isoform), CCAP (one isoform), CHH/MIH (three isoforms), CPRP
 (three isoforms), DH31 (one isoform), EH (one isoform), FLP (seven isoforms),
 HIGSLYRamide (two isoforms), ILP (two isoforms), intocin (one isoform), leucokinin (nine
 isoforms), myosuppressin (one isoform), neuroparsin (three isoforms), NPF (one isoform),
 orcokinin (six isoforms), PDH (three isoforms), pyrokinin (seven isoforms), RPCH (one

isoform), RYamide (two isoforms), sNPF (three isoforms), SIFamide (one isoform) and TRP (one isoform) families, as well as a large collection of other linker/precursor-related peptides. The peptidome predicted for *S. olivacea* is one of the largest thus far known for any decapod crustacean, and is the largest thus far deduced for any member of the Brachyura (true crabs).

The peptidome presented here for *S. olivacea*, while extensive, is likely incomplete. For example, only partial precursors were identified for many of the peptide families, and other peptides are undoubtedly present in the portions of these proteins that remain unknown. In addition, no precursors were identified for several well known crustacean peptide groups, *e.g.*, the allatotropin, bursicon α , diuretic hormone 44, GSEFLamide, proctolin and sulfakinin families. Given their broad conservation in crustaceans, including decapod species (*e.g.*, Bao et al., 2015; Christie, 2014f; Christie and Chi, 2015c; Christie et al., 2015; Fu et al., 2005), it seems likely that members of these peptide groups are also present in *S. olivacea*, though not discovered in this study. As additional transcriptomic data are generated and deposited for *S. olivacea*, it will be interesting to see what other peptides can be discovered for this crab.

4.2. Conservation of peptide isoforms in members of the genus *Scylla*

In addition to the peptides predicted here for *S. olivacea*, several other studies have focused on peptide identification in member of the genus *Scylla* (*e.g.*, Bao et al., 2015; Huang et al., 2014a, 2014b; Kornthong et al., 2013; Tsai et al., 2008). One investigation conducted on *S. paramamosain* is of particular note (Bao et al., 2015), as it employed essentially the same strategy/workflow for peptide discovery that was used here for *S. olivacea*; in this investigation, the structures of approximately 100 mature *S. paramamosain* peptides were predicted via *in silico* transcriptome mining (Bao et al., 2015). Comparison of the structures of the peptides

predicted for *S. olivacea* with those predicted for *S. paramamosain* suggests that a high degree of peptide structural conservation exists within *Scylla* species. In Table 2, putative mature peptide structures that were predicted for both *S. olivacea* and *S. paramamosain* have been highlighted in yellow. It should be noted that for a few peptides, different prohormone convertase cleavage loci were chosen, and hence slightly different lengths were predicted for these isoforms in one or the other species. Similarly, slightly different post-translational modifications have been reported for peptides with the same amino acid sequences in the two species. In both cases, these peptides too are highlighted in yellow in Table 2.

Members of 20 peptide families were identified from both *S. olivacea* (this study) and *S. paramamosain* (Bao et al., 2015). All of the isoforms of AST-B, AST-C, CCAP, DH31, FLP, myosuppressin, orcokinin, RPCH, SIFamide and TRP that were predicted here for *S. olivacea* are also present in *S. paramamosain* (Table 2). For a number of other peptide groups (HIGSLYRamide, leucokinin, neuroparsin, pyrokinin and sNPF), the majority of isoforms predicted for *S. olivacea* are also present in *S. paramamosain*, e.g., eight of nine leucokinins and five of seven pyrokinins (Table 2). For all but one of the remaining peptide groups that had identifications in both *S. olivacea* and *S. paramamosain* (AST-A, CHH, NPF and PDH), some isoforms are conserved in both species and/or others varied at just a few residues in their sequences (Table 2). Only for EH was there considerable variation seen between the peptide predicted from *S. olivacea* and that predicted from *S. paramamosain*. Given the high degree of peptide structural conservation present in the two species, it seems highly likely that conserved physiological roles will ultimately be shown to exist for these signaling molecules in both crabs.

4.3. Application of peptide hormones in aquaculture

The aquaculture of decapod crustaceans is of growing importance for generating product for human consumption. While penaeid shrimp represent the largest aquaculture initiative involving decapods, there is growing interest in the commercial farming of crabs, in particular members of the genus *Scylla*. For example, in 2013 alone, nearly 200,000 tons of *Scylla* were produced for market via aquaculture (*Scylla serrata* Fact Sheet, Cultured Aquatic Species Information Programme, Fisheries and Aquaculture Department, Food and Agriculture Organization of the United Nations; http://www.fao.org/fishery/culturedspecies/Scylla_serrata/en [accessed January 11, 2016]). To bring any commercially cultured crustacean to market requires raising it through multiple developmental stages, and understanding its intrinsic physiological control systems is a key component for any successful rearing. To increase production per unit area, one can develop new or improve upon existing management infrastructure, can “domesticate” and/or manipulate the physiology of the aquaculture target species, or can combine the two approaches. For the vast majority of crustacean species, there is limited knowledge of the basic physiological processes that control important components of their aquaculture, *i.e.*, reproductive maturation, mating, spawning, larval and adult growth, and disease resistance, which makes the latter two strategies for increasing production difficult. Clearly, all of the basic physiological processes that are important to the commercial farming of crustaceans are under endocrine control, including control by locally released and circulating peptides. For example, recent studies on *Scylla* species have implicated both PDH and RPCH in ovarian development and maturation (Huang et al., 2014b; Zeng et al., 2016), while ILP signaling has been hypothesized to play roles in regulating both growth and metabolism (Huang et al., 2015). One goal of study presented here is to extend our knowledge of the peptidergic control systems of *S. olivacea* in order to contribute to improvement in its aquaculture. The

peptides described here greatly expand the known peptidome for *S. olivacea*, and for the genus *Scylla* as a whole, providing targets for possible endocrine manipulation in *S. olivacea*. The potential for such manipulation has already been piloted in the penaeid *Litopenaeus vannamei*, where oral administration (via laced feed) of one of its native peptide hormones (NPF) has been shown to significantly increase the rates of food intake and growth in juvenile shrimp (Christie et al., 2011d).

4.4. Production of “neuropeptides” in non-neuronal tissues

The peptide families for which precursors were identified in this study are well known neuropeptides, *i.e.*, produced by neurons and used for local modulation of neuronal activity and/or for release into the circulatory system to exert influence on more distantly located target tissues (Christie, 2011; Christie et al., 2010a). Members of at least some of the peptide groups identified here have been shown to be produced by and released from non-neuronal tissues as well, *e.g.*, the gut. For example, epithelial endocrine cells in the foregut and hindgut of the crab *C. maenas* have been shown to produce and release CHH, presumably to regulate water and ion uptake during ecdysis, the only time these cells express CHH (Chung et al., 1999). Similarly, epithelial endocrine cells in the midguts of several species of *Cancer* crabs have been shown to produce SIFamide or TRP family members, which are hypothesized to be involved in the paracrine/hormonal regulation of feeding-related behavior (Christie et al., 2007). Interestingly, and quite surprisingly, the sole tissue reported to have been used for the production of the *S. olivacea* transcripts identified here as encoding putative “neuropeptide” precursors was testis (BioProject Accession No. [PRJNA289610](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA289610); M. Ikhwanuddin, unpublished direct GenBank submission). While neuronal contamination in the tissue samples used for the production of the

transcriptome mined here remains a possibility, if the peptides are truly produced by *S. olivacea* testis, a rethinking of hormonal control of reproduction in this and other crustaceans needs to be considered given the large suite of peptide hormones that appear to be produced and liberated from this reproductive tissue. Interestingly, and in support of the possibility of “neuropeptide” production in crustacean reproductive structures, is the prior finding of PDH-encoding transcript production in the ovaries of *S. paramamosain* (Huang et al., 2014b). Clearly more work needs to be conducted to confirm the production of the peptides reported here from the *S. olivacea* testis, but if true, this tissue adds to a growing list of non-neuronal ones for which “neuropeptide” hormones are likely be produced and secreted.

Acknowledgements

Financial support was provided by National Science Foundation grant IOS-1353023 and the Cades Foundation of Honolulu, Hawaii.

References

- Bao, C., Yang, Y., Huang, H., Ye, H., 2015. Neuropeptides in the cerebral ganglia of the mud crab, *Scylla paramamosain*: transcriptomic analysis and expression profiles during vitellogenesis. *Sci. Rep.* 5, 17055.
- Böcking, D., Dirksen, H., Keller, R., 2002. The crustacean neuropeptide of the CHH/MIH/GIH family: structures and biological activities. In: *The Crustacean Nervous System*. K Wiese (Ed.). Springer, Heidelberg. pp. 84–97.

642 Chan, S.M., Gu, P.L., Chu, K.H., Tobe, S.S., 2003. Crustacean neuropeptide genes of the
 643 CHH/MIH/GIH family: implications from molecular studies. *Gen. Comp. Endocrinol.* 134, 214-
 644 219.
 645
 646 Chen, B., Fan, D.Q., Zhu, K.X., Shan, Z.G., Chen, F.Y., Hou, L., Cai, L., Wang, K.J., 2015.
 647 Mechanism study on a new antimicrobial peptide Sphistin derived from the N-terminus of crab
 648 histone H2A identified in haemolymphs of *Scylla paramamosain*. *Fish Shellfish Immunol.* 47,
 649 833-846.
 650
 651 Christie, A.E., 2008a. Neuropeptide discovery in Ixodoidea: an *in silico* investigation using
 652 publicly accessible expressed sequence tags. *Gen. Comp. Endocrinol.* 157, 174-185.
 653
 654 Christie, A.E., 2008b. *In silico* analyses of peptide paracrine/hormones in Aphidoidea. *Gen.*
 655 *Comp. Endocrinol.* 159, 67-79.
 656
 657 Christie, A.E., 2011. Crustacean neuroendocrine systems and their signaling agents. *Cell Tissue*
 658 *Res.* 345, 41-67.
 659
 660 Christie, A.E., 2014a. Prediction of the first neuropeptides from a member of the Remipedia
 661 (Arthropoda, Crustacea). *Gen. Comp. Endocrinol.* 201, 74-86.
 662
 663 Christie, A.E., 2014b. Prediction of the peptidomes of *Tigriopus californicus* and
 664 *Lepeophtheirus salmonis* (Copepoda, Crustacea). *Gen. Comp. Endocrinol.* 201, 87-106.

665

666 Christie, A.E., 2014c. Peptide discovery in the ectoparasitic crustacean *Argulus siamensis*:
667 identification of the first neuropeptides from a member of the Branchiura. Gen. Comp.
668 Endocrinol. 204, 114-125.

669

670 Christie, A.E., 2014d. *In silico* characterization of the peptidome of the sea louse *Caligus*
671 *rogercresseyi* (Crustacea, Copepoda). Gen. Comp. Endocrinol. 204, 248-260.

672

673 Christie, A.E., 2014e. Identification of the first neuropeptides from the Amphipoda (Arthropoda,
674 Crustacea). Gen. Comp. Endocrinol. 206, 96-110.

675

676 Christie, A.E., 2014f. Expansion of the *Litopenaeus vannamei* and *Penaeus monodon*
677 peptidomes using transcriptome shotgun assembly sequence data. Gen. Comp. Endocrinol. 206,
678 235-254.

679

680 Christie, A.E., 2015a. *In silico* characterization of the neuropeptidome of the Western black
681 widow spider *Latrodectus hesperus*. Gen. Comp. Endocrinol. 210, 63-80.

682

683 Christie, A.E., 2015b. Neuropeptide discovery in *Eucyclops serrulatus* (Crustacea, Copepoda): in
684 silico prediction of the first peptidome for a member of the Cyclopoida. Gen. Comp. Endocrinol.
685 211, 92-105.

686

687 Christie, A.E., 2015c. Neuropeptide discovery in *Symphylella vulgaris* (Myriapoda, Symphyla):
 688 *in silico* prediction of the first myriapod peptidome. Gen. Comp. Endocrinol. 223, 73-86.
 689
 690 Christie, A.E., 2015d. *In silico* prediction of a neuropeptidome for the eusocial insect
 691 *Mastotermes darwiniensis*. Gen. Comp. Endocrinol. 224, 69-83.
 692
 693 Christie, A.E., Cashman, C.R., Brennan, H.R., Ma, M., Sousa, G.L., Li, L., Stemmler, E.A.,
 694 Dickinson, P.S., 2008a. Identification of putative crustacean neuropeptides using *in silico*
 695 analyses of publicly accessible expressed sequence tags. Gen. Comp. Endocrinol. 156, 246-64.
 696
 697 Christie, A.E., Cashman, C.R., Stevens, J.S., Smith, C.M., Beale, K.M., Stemmler, E.A.,
 698 Greenwood, S.J., Towle, D.W., Dickinson, P.S., 2008b. Identification and cardiotropic actions
 699 of brain/gut-derived tachykinin-related peptides (TRPs) from the American lobster *Homarus*
 700 *americanus*. Peptides 29, 1909-1918.
 701
 702 Christie, A.E., Chapline, M.C., Jackson, J.M., Dowda, J.K., Hartline, N., Malecha, S.R., Lenz,
 703 P.H., 2011. Identification, tissue distribution and orexigenic activity of neuropeptide F (NPF) in
 704 penaeid shrimp. J. Exp. Biol. 214, 1386-1396.
 705
 706 Christie, A.E., Chi, M., 2015a. Neuropeptide discovery in the Araneae (Arthropoda, Chelicerata,
 707 Arachnida): elucidation of true spider peptidomes using that of the Western black widow as a
 708 reference. Gen. Comp. Endocrinol. 213, 90-109.
 709

710 Christie, A.E., Chi, M., 2015b. Identification of the first neuropeptides from the enigmatic
 711 hexapod order Protura. *Gen. Comp. Endocrinol.* 224, 18–37.
 712
 713 Christie, A.E., Chi, M., 2015c. Prediction of the neuropeptidomes of members of the Astacidea
 714 (Crustacea, Decapoda) using publicly accessible transcriptome shotgun assembly (TSA)
 715 sequence data. *Gen. Comp. Endocrinol.* 224, 38–60.
 716
 717 Christie, A.E., Chi, M., Lameyer, T.J., Pascual, M.G., Shea, D.N., Stanhope, M.E., Schulz, D.J.,
 718 Dickinson, P.S., 2015. Neuropeptidergic signaling in the American lobster *Homarus*
 719 *americanus*: new insights from high-throughput nucleotide sequencing. *PLoS ONE*. 10,
 720 e0145964.
 721
 722 Christie, A.E., Durkin, C.S., Hartline, N., Ohno, P., Lenz, P.H., 2010b. Bioinformatic analyses
 723 of the publicly accessible crustacean expressed sequence tags (ESTs) reveal numerous novel
 724 neuropeptide-encoding precursor proteins, including ones from members of several little studied
 725 taxa. *Gen. Comp. Endocrinol.* 167, 164-178.
 726
 727 Christie, A.E., Kutz-Naber, K.K., Stemmler, E.A., Klein, A., Messinger, D.I., Goiney, C.C.,
 728 Conterato, A.J., Bruns, E.A., Hsu, Y.W., Li, L., Dickinson, P.S., 2007. Midgut epithelial
 729 endocrine cells are a rich source of the neuropeptides APSGFLGMRamide (*Cancer borealis*
 730 tachykinin-related peptide Ia) and GYRKPPFNGSIFamide (Gly¹-SIFamide) in the crabs *Cancer*
 731 *borealis*, *Cancer magister* and *Cancer productus*. *J. Exp. Biol.* 210, 699-714.
 732

733 Christie, A.E., McCoole, M.D., Harmon, S.M., Baer, K.N., Lenz, P.H., 2011c. Genomic
 734 analyses of the *Daphnia pulex* peptidome. Gen. Comp. Endocrinol. 171, 131-150.
 735
 736 Christie, A.E., Nolan, D.H., Garcia, Z.A., McCoole, M.D., Harmon, S.M., Congdon-Jones, B.,
 737 Ohno, P., Hartline, N., Congdon, C.B., Baer, K.N., Lenz, P.H., 2011b. Bioinformatic prediction
 738 of arthropod/nematode-like peptides in non-arthropod, non-nematode members of the
 739 Ecdysozoa. Gen. Comp. Endocrinol. 170, 480-486.
 740
 741 Christie, A.E., Nolan, D.H., Ohno, P., Hartline, N., Lenz, P.H., 2011a. Identification of
 742 chelicerate neuropeptides using bioinformatics of publicly accessible expressed sequence tags.
 743 Gen. Comp. Endocrinol. 170, 144-155.
 744
 745 Christie, A.E., Roncalli, V., Wu, L.S., Ganote, C.L., Doak, T., Lenz, P.H., 2013. Peptidergic
 746 signaling in *Calanus finmarchicus* (Crustacea, Copepoda): *in silico* identification of putative
 747 peptide hormones and their receptors using a *de novo* assembled transcriptome. Gen. Comp.
 748 Endocrinol. 187:117-135.
 749
 750 Christie, A.E., Stemmler, E.A., Dickinson, P.S., 2010a. Crustacean neuropeptides. Cell. Mol.
 751 Life Sci. 67, 4135-4169.
 752
 753 Christie, A.E., Stevens, J.S., Bowers, M.R., Chapline, M.C., Jensen, D.A., Schegg, K.M.,
 754 Goldwaser, J., Kwiatkowski, M.A., Pleasant, T.K. Jr., Shoenfeld, L., Tempest, L.K., Williams,
 755 C.R., Wiwatpanit, T., Smith, C.M., Beale, K.M., Towle, D.W., Schooley, D.A., Dickinson, P.S.,

2010c. Identification of a calcitonin-like diuretic hormone that functions as an intrinsic
 modulator of the American lobster, *Homarus americanus*, cardiac neuromuscular system. J.
 Exp. Biol. 213, 118-127.

Chung, J.S., Dircksen, H., Webster, S.G., 1999. A remarkable, precisely timed release of
 hyperglycemic hormone from endocrine cells in the gut is associated with ecdysis in the crab
Carcinus maenas. Proc. Natl. Acad. Sci. USA. 96, 13103-13107.

Chung, J.S., Wilcockson, D.C., Zmora, N., Zohar, Y., Dircksen, H., Webster, S.G., 2006.
 Identification and developmental expression of mRNAs encoding crustacean cardioactive
 peptide (CCAP) in decapod crustaceans. J. Exp. Biol. 209, 3862-3872.

Chung, J.S., Zmora, N., Katayama, H., Tsutsui, N., 2010. Crustacean hyperglycemic hormone
 (CHH) neuropeptides family: functions, titer, and binding to target tissues. Gen. Comp.
 Endocrinol. 166, 447-454.

Dickinson, P.S., Stemmler, E.A., Barton, E.E., Cashman, C.R., Gardner, N.P., Rus, S., Brennan,
 H.R., McClintock, T.S., Christie, A.E., 2009. Molecular, mass spectral, and physiological
 analyses of orcokinins and orcokinin precursor-related peptides in the lobster *Homarus*
americanus and the crayfish *Procambarus clarkii*. Peptides 30, 297-317.

Dickinson, P.S., Stemmler, E.A., Cashman, C.R., Brennan, H.R., Dennison, B., Huber, K.E.,
 Peguero, B., Rabacal, W., Goiney, C.C., Smith, C.M., Towle, D.W., Christie, A.E., 2008.

779 SIFamide peptides in clawed lobsters and freshwater crayfish (Crustacea, Decapoda, Astacidea):
 780 a combined molecular, mass spectrometric and electrophysiological investigation. Gen. Comp.
 781 Endocrinol. 156, 347-360.
 782
 783 Dickinson, P.S., Stevens, J.S., Rus, S., Brennan, H.R., Goiney, C.C., Smith, C.M., Li, L., Towle,
 784 D.W., Christie, A.E., 2007. Identification and cardiotropic actions of sulfakinin peptides in the
 785 American lobster *Homarus americanus*. J. Exp. Biol. 210, 2278-2289.
 786
 787 Dirksen, H., Neupert, S., Predel, R., Verleyen, P., Huybrechts, J., Strauss, J., Hauser, F.,
 788 Stafflinger, E., Schneider, M., Pauwels, K., Schoofs, L., Grimmlikhuijzen, C.J., 2011.
 789 Genomics, transcriptomics, and peptidomics of *Daphnia pulex* neuropeptides and protein
 790 hormones. J. Proteome Res. 10, 4478-4504.
 791
 792 Fanjul-Moles, M.L., 2006. Biochemical and functional aspects of crustacean hyperglycemic
 793 hormone in decapod crustaceans: review and update. Comp. Biochem. Physiol. C Toxicol.
 794 Pharmacol. 142:390-400.
 795
 796 Ferrè, F., Clote, P., 2005. DiANNA: a web server for disulfide connectivity prediction. Nucleic
 797 Acids Res. 33,W230-W232.
 798
 799 Fu, Q., Kutz, K.K., Schmidt, J.J., Hsu, Y.W., Messinger, D.I., Cain, S.D., de la Iglesia, H.O.,
 800 Christie, A.E., Li, L., 2005. Hormone complement of the *Cancer productus* sinus gland and

801 pericardial organ: an anatomical and mass spectrometric investigation. J. Comp. Neurol. 493,
 802 607-626.
 803
 804 Gard, A.L., Lenz, P.H., Shaw, J.R., Christie, A.E., 2009. Identification of putative peptide
 805 paracrine/hormones in the water flea *Daphnia pulex* (Crustacea; Branchiopoda; Cladocera)
 806 using transcriptomics and immunohistochemistry. Gen. Comp. Endocrinol. 160, 271-287.
 807
 808 Girish, B.P., Swetha, C.H., Reddy, P.S., 2015. Induction of ecdysteroidogenesis, methyl
 809 farnesoate synthesis and expression of ecdysteroid receptor and retinoid X receptor in the
 810 hepatopancreas and ovary of the giant mud crab, *Scylla serrata* by melatonin. Gen. Comp.
 811 Endocrinol. 217-218, 37-42.
 812
 813 Gong, J., Ye, H., Xie, Y., Yang, Y., Huang, H., Li, S., Zeng, C., 2015. Ecdysone receptor in the
 814 mud crab *Scylla paramamosain*: a possible role in promoting ovarian development. J.
 815 Endocrinol. 224, 273-287.
 816
 817 Hansen, K.K., Hauser, F., Williamson, M., Weber, S.B., Grimmelikhuijzen, C.J., 2011. The
 818 *Drosophila* genes CG14593 and CG30106 code for G-protein-coupled receptors specifically
 819 activated by the neuropeptides CCHamide-1 and CCHamide-2. Biochem. Biophys. Res.
 820 Commun. 404, 184-189.
 821

822 Huang, X., Ye, H., Huang, H., Liu, A., Feng, B., 2015. Implication for the regulation of
823 catabolism drawn from the single insulin-like growth factor binding domain protein (SIBD) gene
824 in the mud crab, *Scylla paramamosain*. Gen. Comp. Endocrinol. 216, 24-32.
825

826 Huang, X., Ye, H., Huang, H., Yang, Y., Gong, J., 2014a. An insulin-like androgenic gland
827 hormone gene in the mud crab, *Scylla paramamosain*, extensively expressed and involved in the
828 processes of growth and female reproduction. Gen. Comp. Endocrinol. 204, 229-238.
829

830 Huang, X., Ye, H., Huang, H., Yu, K., Huang, Y., 2014b. Two beta-pigment-dispersing
831 hormone (β -PDH) isoforms in the mud crab, *Scylla paramamosain*: implication for regulation of
832 ovarian maturation and a photoperiod-related daily rhythmicity. Anim. Reprod. Sci. 150, 139-
833 147.
834

835 Ikhwanuddin, M., Muhd-Farouk, H., Memon, A.J., Wendy, W., Abol-Munafi, A.B., 2014.
836 Sperm Viability Assessment over elapsing time maintained at 2 degrees C of orange mud crab,
837 *Scylla olivacea* (Herbst, 1796). Pak. J. Biol. Sci. 17, 1069-1073.
838

839 Kastin, A.J., 2006. Handbook of Biologically Active Peptides, First Edition. Academic Press.
840

841 Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7:
842 improvements in performance and usability. Mol. Biol. Evol. 30, 772-780.
843

844 Klein, J.M., de Kleijn, D.P., Keller, R., Weidemann, W.M., 1992. Molecular cloning of
 845 crustacean pigment dispersing hormone precursor. *Biochem. Biophys. Res. Commun.* 189,
 846 1509-1514.
 847
 848 Kornthong, N., Chotwiwatthanakun, C., Chansela, P., Tinikul, Y., Cummins, S.F., Hanna, P.J.,
 849 Sobhon, P., 2013. Characterization of red pigment concentrating hormone (RPCH) in the female
 850 mud crab (*Scylla olivacea*) and the effect of 5-HT on its expression. *Gen. Comp. Endocrinol.*
 851 185, 28-36.
 852
 853 Kornthong, N., Cummins, S.F., Chotwiwatthanakun, C., Khornchatri, K., Engsusophon, A.,
 854 Hanna, P.J., Sobhon, P., 2014. Identification of genes associated with reproduction in the Mud
 855 Crab (*Scylla olivacea*) and their differential expression following serotonin stimulation. *PLoS*
 856 *One* 9, e115867.
 857
 858 Linck, B., Klein, J.M., Mangerich, S., Keller, R., Weidemann, W.M., 1993. Molecular cloning
 859 of crustacean red pigment concentrating hormone precursor. *Biochem. Biophys. Res. Commun.*
 860 195, 807-813.
 861
 862 Ma, H., Ma, C., Li, S., Jiang, W., Li, X., Liu, Y., Ma, L., 2014. Transcriptome analysis of the
 863 mud crab (*Scylla paramamosain*) by 454 deep sequencing: assembly, annotation, and marker
 864 discovery. *PLoS One* 9, e102668.
 865

866 Ma, M., Bors, E.K., Dickinson, E.S., Kwiatkowski, M.A., Sousa, G.L., Henry, R.P., Smith,
 867 C.M., Towle, D.W., Christie, A.E., Li, L., 2009. Characterization of the *Carcinus maenas*
 868 neuropeptidome by mass spectrometry and functional genomics. *Gen. Comp. Endocrinol.* 161,
 869 320-334.
 870
 871 Ma, M., Gard, A.L., Xiang, F., Wang, J., Davoodian, N., Lenz, P.H., Malecha, S.R., Christie,
 872 A.E., Li, L., 2010. Combining *in silico* transcriptome mining and biological mass spectrometry
 873 for neuropeptide discovery in the Pacific white shrimp *Litopenaeus vannamei*. *Peptides* 31, 27-
 874 43.
 875
 876 Monigatti, F., Gasteiger, E., Bairoch, A., Jung, E., 2002. The Sulfinator: predicting tyrosine
 877 sulfation sites in protein sequences. *Bioinformatics* 18, 769-770.
 878
 879 Petersen, T.N., Brunak, S., von Heijne, G., Nielsen, H., 2011. SignalP 4.0: discriminating signal
 880 peptides from transmembrane regions. *Nat. Methods* 8, 785-786.
 881
 882 Sharp, J.H., Wilcockson, D.C., Webster, S.G., 2010. Identification and expression of mRNAs
 883 encoding bursicon in the plesiomorphic central nervous system of *Homarus gammarus*. *Gen.*
 884 *Comp. Endocrinol.* 169, 65-74.
 885
 886 Stemmler, E.A., Bruns, E.A., Cashman, C.R., Dickinson, P.S., Christie, A.E., 2010. Molecular
 887 and mass spectral identification of the broadly conserved decapod crustacean neuropeptide

888 pQIRYHQCYFNPISCF: the first PISCF-allatostatin (*Manduca sexta*- or C-type allatostatin)
889 from a non-insect. Gen. Comp. Endocrinol. 165, 1-10.
890
891 Stemmler, E.A., Cashman, C.R., Messinger, D.I., Gardner, N.P., Dickinson, P.S., Christie, A.E.,
892 2007a. High-mass-resolution direct-tissue MALDI-FTMS reveals broad conservation of three
893 neuropeptides (APSGFLGMRamide, GYRKPPFNGSIFamide and pQDLDHVFLRFamide)
894 across members of seven decapod crustacean infraorders. Peptides 28, 2104-2115.
895
896 Stevens, J.S., Cashman, C.R., Smith, C.M., Beale, K.M., Towle, D.W., Christie, A.E.,
897 Dickinson, P.S., 2009. The peptide hormone pQDLDHVFLRFamide (crustacean
898 myosuppressin) modulates the *Homarus americanus* cardiac neuromuscular system at multiple
899 sites. J. Exp. Biol. 212, 3961-3976.
900
901 Sun, W., Li, Z., Wang, S., Wan, W., Wang, S., Wen, X., Zheng, H., Zhang, Y., Li, S., 2015.
902 Identification of a novel clip domain serine proteinase (Sp-cSP) and its roles in innate immune
903 system of mud crab *Scylla paramamosain*. Fish Shellfish Immunol. 47, 15-27.
904
905 Toullec, J.Y., Corre, E., Bernay, B., Thorne, M.A., Cascella, K., Ollivaux, C., Henry, J., Clark,
906 M.S., 2013. Transcriptome and peptidome characterisation of the main neuropeptides and
907 peptidic hormones of a euphausiid: the Ice Krill, *Euphausia crystallorophias*. PLoS One. 8,
908 e71609.
909

910 Tsai, K.W., Chang, S.J., Wu, H.J., Shih, H.Y., Chen, C.H., Lee, C.Y., 2008. Molecular cloning
 911 and differential expression pattern of two structural variants of the crustacean hyperglycemic
 912 hormone family from the mud crab *Scylla olivacea*. Gen. Comp. Endocrinol. 159, 16-25.
 913

914 Veenstra, J.A., 2000. Mono- and dibasic proteolytic cleavage sites in insect neuroendocrine
 915 peptide precursors. Arch. Insect. Biochem. Physiol. 43, 49-63.
 916

917 Veenstra, J.A., 2015. The power of next-generation sequencing as illustrated by the
 918 neuropeptidome of the crayfish *Procambarus clarkii*. Gen. Comp. Endocrinol. 224, 84-95.
 919

920 Ventura, T., Cummins, S.F., Fitzgibbon, Q., Battaglione, S., Elizur, A., 2014. Analysis of the
 921 central nervous system transcriptome of the eastern rock lobster *Sagmariasus verreauxi* reveals
 922 its putative neuropeptidome. PLoS One 9, e97323.
 923

924 Wang, J., Jiang, K.J., Zhang, F.Y., Song, W., Zhao, M., Wei, H.Q., Meng, Y.Y., Ma, L.B.,
 925 2015a. Characterization and expression analysis of the prophenoloxidase activating factor from
 926 the mud crab *Scylla paramamosain*. Genet. Mol. Res. 14, 8847-8860.
 927

928 Wang, J., Zhang, F.Y., Song, W., Fang, Y.B., Hu, J.H., Zhao, M., Jiang, K.J., Ma, L.B., 2015b.
 929 Characterization of hemocyanin from the mud crab *Scylla paramamosain* and its expression
 930 analysis in different tissues, at various stages, and under *Vibrio parahaemolyticus* infection.
 931 Genet. Mol. Res. 14, 16639-16651.
 932

933 Weidemann, W., Gromoll, J., Keller, R., 1989. Cloning and sequence analysis of cDNA for
 934 precursor of a crustacean hyperglycemic hormone. *FEBS Lett.* 257, 31-34.
 935
 936 Wilcockson, D.C., Webster, S.G., 2008. Identification and developmental expression of mRNAs
 937 encoding putative insect cuticle hardening hormone, bursicon in the green shore crab *Carcinus*
 938 *maenas*. *Gen. Comp. Endocrinol.* 156, 113-125.
 939
 940 Xie, C., Chen, Y., Sun, W., Ding, J., Zhou, L., Wang, S., Wang, S., Zhang, Y., Zhu, D., Wen, X.,
 941 Hu, S., Li, S., 2014. Transcriptome and expression profiling analysis of the hemocytes reveals a
 942 large number of immune-related genes in mud crab *Scylla paramamosain* during *Vibrio*
 943 *parahaemolyticus* infection. *PLoS One* 9, e114500.
 944
 945 Yan, X.C., Chen, Z.F., Sun, J., Matsumura, K., Wu, R.S., Qian, P.Y., 2012. Transcriptomic
 946 analysis of neuropeptides and peptide hormones in the barnacle *Balanus amphitrite*: evidence of
 947 roles in larval settlement. *PLoS One* 7, e46513.
 948
 949 Yasuda-Kamatani, Y., Yasuda, A., 2006. Characteristic expression patterns of allatostatin-like
 950 peptide, FMRFamide-related peptide, orcokinin, tachykinin-related peptide, and SIFamide in the
 951 olfactory system of crayfish *Procambarus clarkii*. *J. Comp. Neurol.* 496, 135-147.
 952
 953 Zeng, H., Bao, C., Huang, H., Ye, H., Li, S., 2016. The mechanism of regulation of ovarian
 954 maturation by red pigment concentrating hormone in the mud crab *Scylla paramamosain*. *Anim.*
 955 *Reprod. Sci.* 164, 152-161.

956

957 Zhao, M., Jiang, K., Song, W., Ma, C., Wang, J., Meng, Y., Wei, H., Chen, K., Qiao, Z., Zhang,
958 F., Ma, L., 2015. Two transcripts of HMG-CoA reductase related with developmental regulation
959 from *Scylla paramamosain*: Evidences from cDNA cloning and expression analysis. IUBMB
960 Life. 67, 954-965.

961

962 **Figure legends**

963 Figure 1. Two examples of the *in silico* workflow used for the prediction of putative mature
964 *Scylla olivacea* peptide structures. **(A)** Predicted processing scheme for prepro-adipokinetic
965 hormone-corazonin-like peptide (ACP) I. The structure of the putative mature ACP isoform is
966 shown in red, with those of three putative mature linker/precursor-related peptides shown in
967 blue. In this panel, “pQ” indicates the pyroglutamic acid residue present in the putative mature
968 ACP. **(B)** Predicted processing scheme for prepro-allatostatin C (AST-C). The structure of the
969 putative mature AST-C isoform is shown in red, with those of five putative mature
970 linker/precursor-related peptides shown in blue. In this panel, “pQ” indicates pyroglutamic acid
971 and the inverted red bracket indicates a disulfide bridge present between the two cysteine
972 residues in the putative mature AST-C.

973

974 Figure 2. Putative *Scylla olivacea* pre/preprohormones deduced from transcriptome shotgun
975 assembly sequence data. **(A)** Prepro-adipokinetic hormone-corazonin-like peptide II. **(B1)** An
976 internal fragment of prepro-allatostatin A I. **(B2)** An internal protein fragment of prepro-
977 allatostatin A II. **(C1)** Prepro-allatostatin B I. **(C2)** The amino (N)-terminal portion of prepro-
978 allatostatin B II. **(D)** The carboxyl (C)-terminal portion of pre-bursicon β . **(E)** The N-terminal

979 portion of prepro-CCHamide. **(F)** Prepro-corazonin. **(G)** Prepro-crustacean cardioactive
 980 peptide. **(H1)** Prepro-crustacean hyperglycemic hormone Ia. **(H2)** Prepro-crustacean
 981 hyperglycemic hormone Ib. **(H3)** Prepro-crustacean hyperglycemic hormone IIa. **(H4)** Prepro-
 982 crustacean hyperglycemic hormone IIb. **(H5)** Pre-molt-inhibiting hormone. **(I)** The C-terminal
 983 portion of prepro-diuretic hormone 31. **(J)** Pre-eclosion hormone. **(K)** The C-terminal portion
 984 of prepro-FMRamide-like peptide. **(L)** An internal protein fragment of prepro-
 985 HIGSLYamide. **(M)** The C-terminal portion of prepro-insulin-like peptide. **(N)** Prepro-
 986 intocin. **(O)** The C-terminal portion of prepro-leucokinin. **(P)** The C-terminal portion of prepro-
 987 myosuppressin. **(Q1)** Pre-neuroparsin Ia. **(Q2)** Pre-neuroparsin Ib. **(Q3)** Pre-neuroparsin Ic.
 988 **(Q4)** Pre-neuroparsin II. **(Q5)** Pre-neuroparsin III. **(R)** Prepro-neuropeptide F. **(S1)** An internal
 989 protein fragment of prepro-orcokinin I. **(S2)** The C-terminal portion of prepro-orcokinin II. **(T1)**
 990 Prepro-pigment dispersing hormone I. **(T2)** The C-terminal portion of prepro-pigment
 991 dispersing hormone IIa. **(T3)** The C-terminal portion of prepro-pigment dispersing hormone IIb.
 992 **(T4)** The C-terminal portion of prepro-pigment dispersing hormone IIc. **(T5)** The C-terminal
 993 portion of prepro-pigment dispersing hormone IId. **(U)** The C-terminal portion of prepro-
 994 pyrokinin. **(V)** The N-terminal portion of prepro-red pigment concentrating hormone. **(W)** The
 995 C-terminal portion of prepro-RYamide. **(X1)** The N-terminal portion of prepro-short
 996 neuropeptide F Ia. **(X2)** The N-terminal portion of prepro-short neuropeptide F Ib. **(Y)** Prepro-
 997 SIFamide. **(Z1)** The N-terminal portion of prepro-tachykinin-related peptide Ia. **(Z2)** The N-
 998 terminal portion of prepro-tachykinin-related peptide Ib. In this figure, signal peptides are shown
 999 in gray, while all mono/dibasic cleavage loci are shown in black. For each sequence, the
 1000 isoform(s) of the peptide for which the precursor is named is/are shown in red, with all
 1001 linker/precursor related peptides shown in blue. The “+” symbol indicate the presence of

1002 additional, unknown, amino acid residues at the N- and/or C-termini of the protein in question.
1003 It should be noted that two of the precursor proteins shown here (**H4** and **Z2**) are N-terminally
1004 extended variants of shorter sequences (**H3** and **Z1**). These N-terminal extensions (highlighted
1005 in yellow) generate as situation where no signal peptide is predicted for the protein using
1006 SignalP, while one is predicted for the shorter sequence. The extension may be the result of an
1007 artifact in the process of assembling the transcript encoding the protein in question, or may have
1008 true biological significance, *e.g.*, it may function as a potential regulatory element.

Table 1

Table 1. Putative <i>Scylla olivacea</i> peptide-encoding transcripts identified via <i>in silico</i> transcriptome shotgun assembly (TSA) mining						
Peptide target family	TSA transcript/protein identifications				tblastn search statistics	
	Transcript accession number	Transcript length*	Deduced protein length [†]	Deduced protein type [‡]	BLAST score	E-value
Adipokinetic hormone-corazonin-like peptide	GDRN01070310	775	101	F	70.5	1e-12
	GDRN01070309	741	74	F	62.4	9e-10
	GDRN01070311	888	74	F	62.4	2e-09
Allatostatin A	GDRN01047849	392	130	I	88.2	5e-17
	GDRN01047848	299	99	I	77.0	2e-13
Allatostatin B	GDRN01076322	1157	184	F	75.9	4e-14
	GDRN01076320	1436	298	N	76.3	5e-14
Allatostatin C	GDRN01030956	995	149	F	107	7e-25
Allatotropin						
Bursicon α						
Bursicon β	GDRN01116356	450	106	C	171	3e-51
CCHamide	GDRN01012126	493	94	N	43.9	0.020
Corazonin	GDRN01069484	712	117	F	54.3	1e-06
	GDRN01069483	1038	117	F	55.1	1e-06
Crustacean cardioactive peptide	GDRN01107608	569	141	F	243	5e-79
Crustacean hyperglycemic hormone/molt-inhibiting hormone	GDRN01059541	1668	139	F	147	6e-39
	GDRN01059542	1603	132	F	145	3e-38
	GDRN01079185	1440	116	F	69.7	3e-11
	GDRN01079183	1676	165	F	69.7	5e-11
	GDRN01038236	709	125	F	39.7	0.31
DENamide						
Diuretic hormone 31	GDRN01042203	1078	98	C	61.6	1e-08
Diuretic hormone 44						
DXXRLamide						
Ecdysis-triggering hormone						
Eclosion hormone	GDRN01025034	799	93	F	97.8	1e-22
FMRFamide-like peptide	GDRN01060744	1801	251	C	72.4	8e-11
FXGGXamide						
GSEFLamide						
HIGSLYRamide	GDRN01056813	948	316	I	362	4e-122
Insulin-like peptide	GDRN01009614	868	181	C	112	2e-26
Intocin	GDRN01127322	323	158	F	38.1	0.45
Leucokinin	GDRN01058507	1161	356	C	159	8e-40
Myosuppressin	GDRN01080972	747	110	C	52.0	6e-06
Neuroparsin	GDRN01069880	2913	95	F	107	3e-24

	GDRN01069883	2580	101	F	110	2e-25
	GDRN01069884	2777	118	F	108	6e-25
	GDRN01062204	2331	102	F	106	4e-24
	GDRN01109875	1163	106	F	65.9	2e-10
Neuropeptide F	GDRN01070465	1033	102	F	103	6e-24
	GDRN01070466	487	84	N	84.0	7e-18
Orcokinin	GDRN01068489	612	145	I	174	2e-50
	GDRN01068488	3747	94	C	109	1e-23
Pigment dispersing hormone	GDRN01052874	1621	88	F	123	4e-31
	GDRN01052876	1721	88	F	123	6e-31
	GDRN01096105	448	59	C	58.9	4e-09
	GDRN01096108	493	59	C	55.1	1e-07
	GDRN01096107	446	89	C	53.1	5e-07
	GDRN01096104	733	89	C	54.3	4e-07
Proctolin						
Pyrokinin	GDRN01096339	739	146	C	65.9	2e-11
Red pigment concentrating hormone	GDRN01016537	356	117	N	48.5	4e-05
RYamide	GDRN01027614	483	97	C	106	6e-26
Short neuropeptide F	GDRN01012387	568	154	N	161	3e-47
	GDRN01012386	469	121	N	112	2e-28
SIFamide	GDRN01087866	393	78	F	89.7	1e-20
Sulfakinin						
Tachykinin-related peptide	GDRN01093289	758	103	N	61.2	5e-08
	GDRN01093287	457	131	N	61.2	1e-08

*Length in nucleotides

†Length in amino acids

‡Deduced protein type: F, full-length; N, amino-terminal partial; I, internal protein fragment; C, carboxyl-terminal partial.

Query proteins: adipokinetic hormone-corazonin-like peptide (ACP), *Procambarus clarkii* prepro-ACP (deduced from **GBEV01011002**; Christie and Chi, 2015c); allatostatin A (AST-A), *P. clarkii* prepro-AST-A (**BAE45266**; Yasuda-Kamatani and Yasuda, 2006); allatostatin B (AST-B); *P. clarkii* prepro-allatostatin B (deduced from **GBEV01040422**; Christie and Chi, 2015c); allatostatin C (AST-C), *Litopenaeus vannamei* prepro-AST-C (deduced from **FE182974**; Stemmler et al., 2010); allatotropin, *Tigriopus californicus* prepro-allatotropin (deduced from **JW513825**; Christie, 2014b); bursicon α , *Homarus gammarus* pre-bursicon α (**ADI86242**; Sharp et al., 2010); bursicon β , *Homarus americanus* pre-bursicon β (deduced from **CN854188**; Christie et al., 2010b); CCHamide, *P. clarkii* prepro-CCHamide II (deduced from **GBEV01015793**; Christie and Chi, 2015c); corazonin, *H. americanus* prepro-corazonin (from DS01-Homarus1_Transcript_40929; Christie et al., 2015); crustacean cardioactive peptide (CCAP), *Carcinus maenas* prepro-CCAP (**ABB46291**; Chung et al., 2006); crustacean hyperglycemic hormone (CHH), *C. maenas* prepro-CHH (**CAA35605**; Weidemann et al., 1989); DENamide, *Daphnia pulex* prepro-DENamide (Dirksen et al., 2011); diuretic hormone 31, *H. americanus* prepro-calcitonin-like diuretic hormone (**ACX46386**; Christie et al., 2010c); diuretic hormone 44 (DH44), *H. americanus* prepro-DH44 (deduced from DS01-Homarus1_Transcript_7828; Christie et al., 2015); DXXRLamide, *T. californicus* prepro-DXXRLamide Ia (deduced from **JW528324**; Christie, 2014b); ecdysis-triggering hormone (ETH), *D. pulex* prepro-ETH (Dirksen et al., 2011); eclosion hormone (EH), *P. clarkii* pre-EH (deduced from **GBEV01101142**; Christie and Chi, 2015c); FMRamide-like peptide; *P. clarkii* prepro-FLRFamide A (**BAE06262**; Yasuda-Kamatani and Yasuda, 2006); FXGGXamide, *T. californicus* prepro-FXGGXamide Ia (deduced from **JV193177**; Christie, 2014b);

GSEFLamide, *P. clarkii* prepro-GSEFLamide (deduced from **GBEV01013249**; Christie and Chi, 2015c); HIGSLYRamide, *C. maenas* prepro-HIGSLYRamide (deduced from **DV111329**; Christie et al., 2008a); insulin-like peptide (ILP), *H. americanus* prepro-insulin-like peptide (deduced from DS01-Homarus1_Transcript_51972; Christie et al., 2015); intocin, *Pontastacus leptodactylus* prepro-intocin (deduced from **GAFS01000331**; Christie and Chi, 2015c); leucokinin, *P. clarkii* prepro-leucokinin (deduced from the combination of **GBEV01013648** and **GBEV01006537**; Christie and Chi, 2015c); myosuppressin, *H. americanus* prepro-myosuppressin (**ACX46385**; Stevens et al., 2009); neuroparsin, *P. clarkii* pre-neuroparsin II (deduced from **GBEV01000414**; Christie and Chi, 2015c); neuropeptide F (NPF), *P. clarkii* prepro-NPF I (deduced from **GBEV01005906**; Christie and Chi, 2015c); orcokinin, *H. americanus* prepro-orcokinin I (**ACB41787**; Dickinson et al., 2009); pigment dispersing hormone (PDH), *C. maenas* prepro-PDH (**AAA27885**; Klein et al., 1992); proctolin, *Penaeus monodon* prepro-proctolin (deduced from **JR22035**; Christie 2014f); pyrokinin, *P. clarkii* prepro-pyrokinin (deduced from **GBEV01049409**; Christie and Chi, 2015c); red pigment concentrating hormone (RPCH), *C. maenas* prepro-RPCH (**AAB28133**; Linck et al., 1993); RYamide, *P. clarkii* prepro-RYamide (deduced from **GBEV01010112**; Christie and Chi, 2015c); short neuropeptide F (sNPF), *P. clarkii* prepro-sNPF (deduced from **GBEV01004780**; Christie and Chi, 2015c); SIFamide, *H. americanus* prepro-Val¹-SIFamide (**ABV21807**; Dickinson et al., 2008); sulfakinin, *H. americanus* prepro-sulfakinin (**ABQ95346**; Dickinson et al., 2007); tachykinin-related peptide (TRP), *H. americanus* prepro-TRP (**ACB41786**; Christie et al., 2008b).

Table 2

Table 2. Predicted peptidome of <i>Scylla olivacea</i>	
Peptide group	Predicted peptide structure
Adipokinetic hormone-corazonin-like peptide	pQITFSRSWVPQa
Adipokinetic hormone-corazonin-like peptide linker/precursor-related peptide	SSSPGDIPEPLDPCRDA AATLTSLAGHLLDMMNDVAAADHRPLSDDGTTALRLRNAMLD VA AATLTSLAGHLLVSLLLCNY
Allatostatin A	PRAYSFGLa PRDYAFGLa PQSYSFGLa pQRPYSFGLa GYEDEDEDKTLYAFGLa PRTYSFGLa pQTPDHYSFGLa TPQPYAFGLa GGVYSFGLa SGNYNFGLa
Allatostatin A linker/precursor-related peptide	+CEEEEEED NSHYDTYS KQEDMGFS EYMDLD GDDQDLE ADDLTEEEE+ +LQILQQYEAAAAEAVAEAEAAVEELEEEENYSYa SVRDVASEEEEEV
Allatostatin B	AGWSSMRGAWa NNNWRKFQGSWa GNWNKFQGSWa AGWNKFQGSWa TSWGKFQGSWa NNWSKFQGSWa AGWSSLQGAWa AWSNLQGAWa
Allatostatin B linker/precursor-related peptide	QDEGSGLAVAQA DDSSDQGLQVSED GEEIQDAEE

	<p> ADDMTEEAALQAAED GDEVASEDDLQDAED QDDLIQLQVRLFLPHTLASFLLSFI QDDLIQLQDLED SPNSEDIEDEALEEEEELQVSPEALARMVAASPV GWALWa PDYPAVSPRSTNWSSLRGAARPLAQ GGLFNNTYFIESNKHSS+ </p>
Allatostatin C	
Allatostatin C linker/precursor-related peptide	<p> pQIRYHQCYFNPISCF LPAKEVPKAQKEVSSAHEGGRLQ AAGPSSDTSQEELAALKDLILSRLAAELDDTYQDLPSF DLLKTEVDGEEDDEGNEEa EEa MFAPLSGLPGNLRIT </p>
Bursicon β	
	+DDDRYNVSAEEYDDTGRLVRVCEEDVAVNKCEGACVSKVQPSVNTPSGFLKDCRCCREVHLRARDITLTHCYDGDGARLTGAKATQQVKLREPADCQCFKCGDSTR
CCHamide	
CCHamide linker/precursor-related peptide	<p> HGSVKVGCLNYGHSC_LGAHa GSWPPSAPRSPPAVALLAPFLQSLAAPRAAARSWQDKIMSQ QQ+ </p>
Corazonin	
Corazonin linker/precursor-related peptide	<p> QTFQYSRGWTNa STELGGVGVGVAa AGVDFVAADP SLQQQTPAP PLPRDIEE LRAVEAGVSALLRAAQQNPEAAAAAVAGEQDY_(SO3H) YAQN </p>
Crustacean cardioactive peptide	
Crustacean cardioactive peptide linker/precursor-related peptide	<p> PFCNAFTGCa GPVA DIDSLLDGKI SDPELEGLASGSELNDITKHVLAEARLWEQLQNKMEAMRMLAS MESRPMF SLTQPQHDHVHSAAALEHKGDVEKQ </p>
Crustacean hyperglycemic hormone/molt-inhibiting hormone	<p> pQIFDSSCKGVYDRAIFNELEHVCNDCYNLYRTSHVASGCRVNCFENHVFDDCVYELLLHNPEEVLLMRDAIRa ATDLVDTSCKGYDRDTWNELHHL_CED_CDNLYREFHFQAKCRNDC_FASDVFAT_CLTDLGKNVEIYQAMAASLRGS </p>

Crustacean hyperglycemic hormone precursor-related peptide	<p>RMIDDEC<u>PN</u>LIGNRDLYEKVDWICE<u>D</u>CANIFRKIRMGYRCRRNCFYNEEFTWCISATENTGHKKELMQWVAILRATPERHRSFLTSIQGGAAIRPHADDGTAY</p> <p>RSAEGFGRMGRLLASLKADTLGPVQDFGVEGAAHPLE RSAEGFGQMGRLLASLKADTLGPVQDFGVEGAAHPLE GPVKPRHNPVSTGGDAGQQH</p>
Diuretic hormone 31	
Diuretic hormone 31 linker/precursor-related peptide	<p>GLDLGLGRGFSGSQAAKHLMGLAAANYAGGPa</p> <p>+ACVCVCVTASS FVRSSSGT ESQAAPLALHHDDHPAAQEHAAVAADAAAAGLQQHSS</p>
Eclosion hormone	VLLLP <u>S</u> <u>A</u> <u>A</u> <u>V</u> <u>A</u> <u>A</u> <u>N</u> <u>R</u> <u>K</u> <u>V</u> <u>S</u> <u>I</u> <u>C</u> <u>I</u> <u>K</u> <u>N</u> <u>C</u> <u>G</u> <u>O</u> <u>C</u> <u>K</u> <u>M</u> <u>Y</u> <u>T</u> <u>D</u> <u>Y</u> <u>F</u> <u>N</u> <u>G</u> <u>G</u> <u>L</u> <u>C</u> <u>G</u> <u>D</u> <u>F</u> <u>C</u> <u>L</u> <u>Q</u> <u>T</u> <u>E</u> <u>G</u> <u>R</u> <u>F</u> <u>I</u> <u>P</u> <u>D</u> <u>C</u> <u>N</u> <u>R</u> <u>P</u> <u>D</u> <u>I</u> <u>L</u> <u>I</u> <u>P</u> <u>F</u> <u>F</u> <u>L</u> <u>Q</u> <u>R</u> <u>L</u> <u>E</u>
FMRFamide-like peptide	<p>NRNFLRFa DRNFLRFa SGHRNYLRFa GNRNFLRFa DDSRNFLRFa GYNRSFLRFa APQRNFLRFa</p>
FMRFamide-like peptide linker/precursor-related peptide	<p>+DRHMNLIFPTRFa GDISSIEDTDMLPETEDSLE SDAEEFGLPGGPLAFSNLQEDDTEDY_(S03B) PVEE SVDRQLKEQKAEAALAPTVVPHSPAKTQDSHRS SASPYNYVMPSHGPAAWAQDFQPDQEDEDLMAIDGPEGAVT NDDSASDVVVVEPASYPRYQ</p>
HIGSLYRamide	<p>HIGSLYRa HIVSLYRa</p>
HIGSLYRamide linker/precursor-related peptide	<p>+SQGLSED SEDNRAED DEDNSQNILED DEDNSQNLSED DEDNANSFGED DDSDSDSPSED DEDNTNSLSED DEDSANSLSED DEDDAHELSED</p>

	DQQVFPTDLIQDEDDIERGLHDLPD YFASLLRDRQSSLGPWNPPD HIGSFFQNRQVPFGMQDSAE HLGSLQHRPSAFAPHDLTAD
Insulin-like peptide	LCGWRLANELNRVCKGVNMP TVSSNALFYLKARGa GLSAECCRKACSVSELAGYCY
Insulin-like peptide linker/precursor-related peptide	SPRTLPGGLVREGES VDLWPVGRELQFTSWTQAPVDDL RAGQLSEPRLLHRPVSSEYQRLPLLTGAEASQVVGGSPRV
Intocin	CFITNCPPGa
Intocin linker/precursor-related peptide	SGGLMSTLGRARTVSF PPVAKAPEC
Leucokinin	PRFLPWGa AFSAWAa VGFSAWAa pQAFSAWAa pEAFNAWAa PFSAWAa pQAFNAWAa pQGFSAWAa FNAWAa
Leucokinin linker/precursor-related peptide	+EGQEDWTARLDPVAALSSLERTEMASRLVPLLPPPMGHVRV SIKEQPHTQVDSYRQES SGMERGKFEDVLHAWATEGDSTDWPASIQMNADM NGDS EAFGPWKa SEDE SSNDND SNNDDDE SDENND SNSDY _(SO3H) D SNNDN SNNED SDDDY _(SO3H) Y _(SO3H) QEQEKEEKSRDQLSSLQQQHQQEHQQASSLLQHSPDSLTHWDANWD
Myosuppressin	pQDLDHVFLRFa
Myosuppressin linker/precursor-related peptide	

	VLSRLTSWSISQVISHSSIQSVSYIILHPVFDGYFQDNLFLSHRALISHIFITSKHNPARNHFPSPQTNTPLSLCPAAIKNSMGVNEPEV AQQ
Neuroparsin	APRCEKHDQEAPKNCRYGTTLDWCKNGVCAKGPGETCGGYRWSEGKCGEGTFCSCGICGGCSPFDGKCGPTSIC IPRCVYDQPPPKNCKYGTALDWCSNGVCAKGPGETCGGYRREEGICGEGTYCEGRCRGCSPFDA SCHDAQLC GPICSSSLNEVLPEMLQAPCRHGVMDWCGNARCAKGPGETCGGRWNVKGSCGKGMVCVCGYCAGCSWDLQCALGRFC
Neuropeptide F	YLDALNAYFAIAGRPRFa
Neuropeptide F linker/precursor-related peptide	AQLPSRQEGGALDAIQALHEAALAGTLSSGEVPYPSRPNVFKSKGEL GEQVRQPEELYDY
Orcokinin	NFDEIDRSSFGFA NFDEIDRSSFGFN NFDEIDRSGFGFA NFDEIDR+ NFDEIDRSSFGFV NFDEIDRSGFGFV
Orcokinin linker/precursor-related peptide	+SWEKGKLSASLYLTLCSSSSASPRTGFGHS MLTPRDLANLY NAE
Pigment dispersing hormone	NSELINSILGLPKVMNDAA NSELINSLLGLSALMNEAA NSELINSLLGLSPLMNEAA
Pigment dispersing hormone linker/precursor-related peptide	ALLTQGQELKYQEREMVAELAQQIYRVAQAPWAAAVGPH +SQVTLRPTPTQAVATLAARILKVHAPQQAATGLPH +SQVTLRPTPTQAVATLAARILKVHGSQEAAAGLPH +PIGLQVQFPSAGIPPIVHLYTYTSVSYTRIITGYPPPHATQAVATLAARILKVHGSQEAAAGLPH +PPGLEVQFPSPGIPPIVHLYTYTSVSYTRIITVYSPPHATQAVATLAARILKVHGSQEAAAGLPH
Pyrokinin	DSAFAFSPRLa VQKSAFAFAPRPa SNFAFAFAPRPa TGTVFAFAFAPRPa TNFAFAFAPRPa TTFAFSPRLa ADFAFAFAPRPa
Pyrokinin linker/precursor-related peptide	+VLQD

	SDSPSETEDRQTGETWWIGEGASSTVTTQPPFLPPRLE
Red pigment concentrating hormone	pQLNFSPGWa
Red pigment concentrating hormone linker/precursor-related peptide	AAATSSSNNGGVGEAVSGLHPSVGGAPGGVPPGSSSGDSCGPIPVSAVMHIYRLI
RYamide	SGFYANRYa FVGGSRYa
RYamide linker/precursor-related peptide	+KGISRFBVLIFLHAE SQUIPEIKVRSS SGVPALAEVPLPVVAPESSEEGEVGASLLLGDVSVCLLDVDPDIYRCL
Short neuropeptide F	APPSMRLRFa DMGWQVAQRSMPTRLRLRFa DARTPALRLRFa
Short neuropeptide F linker/precursor-related peptide	TPVDYDTLNDMYDLLAVHEVE NVDEADPILDHHDLI GASFGEEDMASQE+ NVDEEDMASQE+
SIFamide	GYRKPPFNGSIFa
SIFamide linker/precursor-related peptide	SGADAVYEPGKSQALASVCQVALEACTLWFPGAEE
Tachykinin-related peptide	APSGFLGMRa
Tachykinin-related peptide linker/precursor-related peptide	ASGAGQGEAGSDAPRP EAPSPPLQDAPAPPEDLLPALYQLDVPLRa SEEEEEDE+
Abbreviations: a, amidated C-terminus; +, likely partial peptide with additional unknown amino acids; pQ/pE, pyroglutamic acid; C, disulfide bonded cysteine (see text for descriptions of bonding partners); Y _(SO₃H) , sulfated tyrosine.	
Prediction of sulfation state of tyrosine residues and disulfide bridging between cysteines was done only for full-length peptides.	
Peptides highlighted in yellow have been predicted from both <i>Scylla olivacea</i> (this study) and <i>Scylla paramamosain</i> (Bao et al., 2015).	

Figure 1

A. Prepro-adipokinetic hormone-corazonin-like peptide I (from GDRN01070310)

MASWMLVALVSVCLVGSVT**PQ**ITFSRSWVPQGRSSSPGDIPEPLDPCRDA**RA**ATLTSLAGHLLDMMNDVAAADHRPLSDDGTALRLRNAMLD**RR**VA

↓ Signal peptidase (cleavage locus highlighted above)

QITFSRSWVPQ**GKR**SSSPGDIPEPLDPCRDA**RA**ATLTSLAGHLLDMMNDVAAADHRPLSDDGTALRLRNAMLD**RRRR****VA**

↓ Prohormone convertase (cleavage loci highlighted above)

QITFSRSWVPQ**GKR**SSSPGDIPEPLDPCRDA**R**AATLTSLAGHLLDMMNDVAAADHRPLSDDGTALRLRNAMLD**RRRR****VA**

↓ Carboxypeptidase (cleavage loci highlighted above)

QITFSRSWVPQ**G**SSSPGDIPEPLDPCRDA**A**ATLTSLAGHLLDMMNDVAAADHRPLSDDGTALRLRNAMLD

↓ Peptidylglycine- α -amidating monooxygenase (amidation locus highlighted above)

QITFSRSWVPQamide

↓ Glutaminyl cyclase (cyclization locus highlighted above)

pQITFSRSWVPQamide

B. Prepro-allatostatin C (from GDRN01030956)

MMPCPGHLVVALALVLTLSH**AL**PAKEVPKAQKEVSSAHEGGRLQKRAAGPSSDTSQEELAALKDLILSRLAAELDDTYQDLPSFKRDLLKTEVDGEE
DDEGNEEGRRREEGKKKRMFAPLSGLPGNLRTIK**RQ**IRYHCYFNPISCFRRK

↓ Signal peptidase (cleavage locus highlighted above)

LPAKEVPKAQKEVSSAHEGGRLQ**RA**AGPSSDTSQEELAALKDLILSRLAAELDDTYQDLPSFK**RD**LLKTEVDGEEDEGNEEG**RE**EGKKK**RM**FAP
LSGLPGNLRTIK**RQ**IRYHCYFNPISCFRRK

↓ Prohormone convertase (cleavage loci highlighted above)

LPAKEVPKAQKEVSSAHEGGRLQ**KR**AAGPSSDTSQEELAALKDLILSRLAAELDDTYQDLPSF**KR**DLLKTEVDGEEDEGNEEG**RR**EEG**KKKR**
MFAPLSGLPGNLRTI**KR**QIRYHCYFNPISCF**RRK**

↓ Carboxypeptidase (cleavage loci highlighted above)

LPAKEVPKAQKEVSSAHEGGRLQ**A**AGPSSDTSQEELAALKDLILSRLAAELDDTYQDLPSF**F**DLLKTEVDGEEDEGNEEG**G**EE**G**
MFAPLSGLPGNLRTI**I**QIRYHCYFNPISCF

↓ Peptidylglycine- α -amidating monooxygenase (amidation loci highlighted above)

DLLKTEVDGEEDEGNEEamide**EE**amide**Q**IRYHCYFNPISCF

↓ Glutaminyl cyclase (cyclization locus highlighted above)

pQIRYHCYFNPISCF

↓ Enzymatic disulfide bond formation (bridged cysteines highlighted above)

pQIRYHCYFNPISCF

A. Prepro-adipokinetic hormone-corazonin-like peptide II (from GDRN01070309 and GDRN01070311)

MASWMLVALVSVCLVGSVTPQITFSRSWVPQGRSSSPGDIPEPLDPCRDARAATLTSLAGHLLVSLLLCNYK

B1. Prepro-allatostatin A I (internal fragment protein from GDRN01047849)

+CEEEEEEDKRNSHYDTYSKRPRAYSFGLGKRKQEDMGFSKRPRDYAFGLGKREYMDLDKRPQSYSFGLGKRGGDQDLEKRRPYSFGLGKRGYEDEE
DEDKTLYAFGLGKRPRYSFGLGKRADDLTEEEE+

B2. Prepro-allatostatin A II (internal fragment protein from GDRN01047848)

+LQILQQYEAAAAEAVAEAAAAVEEEEEENYSYGRKQTPDHYSFGLGKRTPQPYAFGLGKRGGVYSFGLGKKSGNYNFGLGKRSVRDVASEEEEE
VKK+

C1. Prepro-allatostatin B I (from GDRN01076322)

MQLATLTATLLTLMAAAAAQDEGSGLAVAQAKRAGWSSMRGAWGKRDDSSDQGLQVSEDKRNNNWRKFQGSWGKRGEETQDAEEKRGNWNKFQGSWG
KRADDMTTEEAALQAAEDKRAGWNKFQGSWGKRGEVASEDDLDQAEDKRTSWGKFQGSWGKRQDDLIQLQVRLFLPHTLASFLLSFI

C2. Prepro-allatostatin B II (N-terminal partial protein from GDRN01076320)

MQLATLTATLLTLMAAAAAQDEGSGLAVAQAKRAGWSSMRGAWGKRDDSSDQGLQVSEDKRNNNWRKFQGSWGKRGEETQDAEEKRGNWNKFQGSWG
KRADDMTTEEAALQAAEDKRAGWNKFQGSWGKRGEVASEDDLDQAEDKRTSWGKFQGSWGKRQDDLIQLQDLEDKRNNWSKFQGSWGKRAGWSSSLQG
AWGKRAWSNLQGAWGKRSPNDSIEDALEEEELQVSPALARMVAASPVKRGWALWGKRDPYPAVSPRSTNWSLGAARPLAQRGGLFNNTYFI
ESNKHSS+

D. Pre-bursicon β (C-terminal partial protein from GDRN01116356)

+DDDRYNVSAEEYDDTGRLVRVCEEDVAVNKCEGACSVKVPQSVNTPSGFLDKCRCCREVHLRARDITLTHCYDGDGARLTGAKATQQVKLREPADC
QCFKCGDSTR

E. Prepro-CCHamide (N-terminal partial protein from GDRN01012126)

MQTWARISTLLLLLLPFLVLVLPAAEAHGSVKGCLNHYGHSCLAGHGKRGSWPPSAPRSPPAVALLAPFLQSLAAPRAAARSWQDKIMSQRQQ+

F. Prepro-corazonin (from GDRN01069484)

MVRRVTVVVLVASLALAACQTFQYSRGWTNGRKRSTELGGVGVASGRRAGVDFVAADPRRSLQQQTAPRRPLPRDIEERLRAVEAGVSALLRAAQ
QNPEAAAAVAGEQDYAQN

G. Prepro-crustacean cardioactive peptide (from GDRN01107608)

MYFTSLSGRAGLVTAATILLFLAADTEAGPVAKRIDISLLDGKIKRPFCAFTGCGKKRSDPELEGLASGSELNDITKHVLAEARLWEQLQNKME
AMRMLASRMESRPMFRKRSLTQFQHDHVHSAALAEHKGDVEKQ

H1. Prepro-crustacean hyperglycemic hormone Ia (from GDRN01059541)

MSALTSIMQMAVLVACITMATLPDTQARSAGFGMRGRLASLKADTLGPVQDFGVEGAHPLEKRQIFDSSCKGVYDRAIFNELEHVCNDCYNLYR
TSHVASGCRVNCFENHVFDDCVYELLLHNPEEVLMLRDAIRG

H2. Prepro-crustacean hyperglycemic hormone Ib (from GDRN01059542)

MQMAVLVACITMATLPDTQARSAGFGMRGRLASLKADTLGPVQDFGVEGAHPLEKRQIFDSSCKGVYDRAIFNELEHVCNDCYNLYRTSHVASG
CRVNCFENHVFDDCVYELLLHNPEEVLMLRDAIRG

H3. Prepro-crustacean hyperglycemic hormone IIa (from GDRN01079185)

MLVVMVIALNIDGGAAQGVKPRHNPVSTGGDAQQHRRSRRTDLVDTSCKGYDRDTWNELHHLCECDNLYREFHFQAKCRNDCFASDVVFATCLT
DLGKNVEIYQAMAASLRGS

H4. Prepro-crustacean hyperglycemic hormone IIb (from GDRN01079183)

MQLAQSVSRHATENNVEERLFEHLTLASCTAPSVSLINMFDKITTLAVVMLVVMVIALNIDGGAAQGVKPRHNPVSTGGDAQQHRRSRRTDLVDT
SCKGYDRDTWNELHHLCECDNLYREFHFQAKCRNDCFASDVVFATCLTDLGKNVEIYQAMAASLRGS

H5. Pre-molt-inhibiting hormone (from GDRN01038236)

MWLLAVVMAALWSLSVQRTAARMIDDECPNLIGNRDLYEKVDWICEDCANIFRKIRMGYRCRRNCFYNEEFTWCISATENTGHKKELMQWVAILRA
TPERHRSFLTSTQGGAAIRPHADDGTAY

I. Prepro-diuretic hormone 31 (C-terminal partial protein from GDRN01042203)

+ACVCVCVTASSRRRFRVSSSGTKRGLDLGLGRGFSGSQAAKHLMGLAAANYAGGPGRRKRRESQAAPLALHHDHPAAQEHAAVAADAAAAGLQQHS
SR

J. Pre-eclosion hormone (from GDRN01025034)

MVGSRKVVVSALLVLSVALVLA~~V~~LLLPSPSASAAVAANRKVSIKNCGQCKKMYTDYFNGGLCGDFCLQTEGRFIPDCNRPDILIPFFLQRL

K. Prepro-FMRamide-like peptide (C-terminal partial protein from GDRN01060744)

+DRHMLNLIPTTRFGRGDISSIEDTDLMPETEDSLEKRNRNFLRFGRDRNFLRFGRSDAEFGLPGGPLAFSNLQEDDTEDEYFVEEKRSGHRNYLRFGRGNRNFLRFGRDDSRNFLRFGRSVDRQLKEQKAHEAALAPTVPVPHSPAKTQDSHRSKRSASPYNVVMPSHGPAAWAQDFQPDQDEDEDLMAIDGPEGAVTKRGYNRSFLRFGRDRNFLRFGRNRDSDASDVVVVEPASYPYQRAPQRNFLRFGR

L. Prepro-HIGSLYRamide (internal fragment protein from GDRN01056813)

+SQGLSEDKRHIGSLYRGKKSEDNRAEDKRHIVSLYRGKKDEDNSQNILEDKRHIGSLYRGKKDEDNSQNLSEDKRHIGSLYRGKKDEDNSQNLSEDKRHIGSLYRGKKDEDNANSFGEDKRHIGSLYRGKKDDSDSPSEDKRHIGSLYRGKKDEDNTNSLSEDKRHIGSLYRGKKDEDSANSLSEDKRHIGSLYRGKKDEDDAHELSEDKRHIGSLYRGKKDQQVFPTDLIQDEDDIERGLHDLDPKRYFASLLRDRQSSLGPWNPPDKRHIGSFFQNRQVFFGMQDSA EKRHLGSLQHRPSAFAPHDLTADKR+

M. Prepro-insulin-like peptide (C-terminal partial protein from GDRN01009614)

+GSSSHTPLQYSLTQSDPPCWRQVFLLLVVTAMQTGRTRGSPRTLPGGLVREGERRLCGWRLANELNRVCKGVYNMPTVSSNALFYLKARGGKRVDLWPGVRELQFTSWTQAPVDDLRAGQLSEPRLLHRPVSSEYQRLPLLTGAEASQVVGGSPPRVKRLSAECCRKACSVSELAGYCY

N. Prepro-intocin (from GDRN01127322)

MQSGVAVTVVVTTLLVGGAAA~~CFITNCPGG~~KRSGGLMSTLGRARTVSFRPPVAKAPEC

O. Prepro-leucokinin (C-terminal partial protein from GDRN01058507)

+EQQEDWTARLDPVAALSSLERTEMASRLVPLPPPMGHVVRRSIKEQPHTQVDSYRQESRRPRFLPWGGRKSGMERGKFEDVLHAWATEGDSTDWPASIQMNADMKRKAFAAWAGKRNGDSKRVGFAWAGKREAFGPWKGRSEDEKRAFAAWAGKRSEDEKRAFAAWAGKREAFNAWAGKRSSNDNDKRQAFSAWAGKRNNDDKRRKPFSAWAGKRSDENNDKRAFAWAGKRNSNDYDKRAFAWAGKRNNNDKRRKQFAWAGKRNNEDKRQFAWAGKRKRFNAWAGKRSDDDYYQEQQEEEEKSRDQLSSLLQQQHQQEHQQASSLLQHSFDSLTHWDANWDR

P. Prepro-mysuppressin (C-terminal partial protein from GDRN01080972)

+VRRVLSRLTSWSISQVISHSSISQSVSYIHLHPVDFGYFDNLFSLHRALISHIFITSKHNPNARNHFPSPQTNTPLSLCPAAIKNSMGVNEPEVKRQDLDHVFLRFGR~~AAQQ~~

Q1. Pre-neuroparsin Ia (from GDRN01069880)

MSKLCVVCVIVVSVFPREASGAPRCEKHDQEAPKNCRYGTTLDWCNGVCAKGPGETCGGYRWSEGKCGEGTFCSCGICGGCSPFDGKCGPTSI~~C~~

Q2. Pre-neuroparsin Ib (from GDRN01069883)

MASFSCRATTLVLVCSCLLLLLQEASGAPRCEKHDQEAPKNCRYGTTLDWCNGVCAKGPGETCGGYRWSEGKCGEGTFCSCGICGGCSPFDGKCGPT~~SI~~C

Q3. Pre-neuroparsin Ic (from GDRN01069884)

MTEKTDMSLVSSLPGTKMASFSCRATTLVLVCSCLLLLLQEASGAPRCEKHDQEAPKNCRYGTTLDWCNGVCAKGPGETCGGYRWSEGKCGEGTFCSCGICGGCSPFDGKCGPTSI~~C~~

Q4. Pre-neuroparsin II (from GDRN01062204)

MTPRARPATLILASCLLLLLLLQKGSATIPRCVSYDQPPPKNCYGTALDWCNNGVCAKGPGETCGGYRREEGICGEGTYCECGRRCRGCSPFDASCH~~DAQLC~~

Q5. Pre-neuroparsin III (from GDRN01109875)

MEMTTRS~~Y~~IFFFIVSSSTTL~~LL~~LLLPGRCEG~~GP~~ICSS~~LN~~EVLP~~EM~~LQAPCRHGVMDWCGNARCAKGPGETCGGRWNVKGSCGKGMVCVGCYACGCSWD~~LQ~~CALGRFC

R. Prepro-neuropeptide F (from GDRN01070465 and GDRN01070466)

MRGTLTVVAVVVVSVAAAWAAQLPSRQEGGALDAIQALHEAALAGTLSSGEVPYPSRPNVFKSKGELRRYLDALNAYFAIAGRPRFGKRGEQVRQPE~~ELYDY~~

S1. Prepro-orcokinin I (internal fragment protein from GDRN01068489)

+SWEGKGLSASLYLTLCSSSSASPRTGFGHSKRNFDEIDRSSFGFAKRNFDEIDRSSFGFNKRNFDEIDRSSGFGFAKRNFDEIDRSSGFGFAKRNFDEIDRSSFGFNKRNFDEIDRSSGFGFAKRNFDEIDRSSGFGFAKRNFDEIDR+

S2. Prepro-orcokinin II (C-terminal partial protein from GDRN01068488)

+SWEGKGLSASLYLTLCSSSSASPRTGFGHSKRNFDEIDRSSFGFAKRNFDEIDRSSFGFVKRMLTPRDLANLYKRNFDEIDRSSGFGFVRRNAE

T1. Prepro-pigment dispersing hormone I (from GDRN01052874 and GDRN01052876)

MIDPVTLPVWMRS~~AV~~IVAVLVVVALAALLTQGQELKYQEREMVAELAQQIYRVAQAPWAAAVGPHKRNSELINSILGLPKVMNDAGR

T2. Prepro-pigment dispersing hormone IIa (C-terminal partial protein from GDRN01096105)

+SQVTLRPTPTQAVATLAARILKVHAPQQAATGLPHKRNSELINSLGLSALMNEAGKR

T3. Prepro-pigment dispersing hormone IIb (C-terminal partial protein from GDRN01096108)

+SQVTLRPTPTQAVATLAARILKVHGSQEAAGLPHKRNSELINSLGLSALMNEAGRR

T4. Prepro-pigment dispersing hormone IIc (C-terminal partial protein from GDRN01096107)

+PIGLQVQFSPSAGIPPIVHLYTYSVSYTRIITGYPPPHATQAVATLAARILKVHGSQEAAGLPHKRNSELINSLGLSALMNEAGKR

T5. Prepro-pigment dispersing hormone IId (C-terminal partial protein from GDRN01096104)

+PPGLEVQFSPSAGIPPIVHLYTYSVSYTRIITVYSPPHATQAVATLAARILKVHGSQEAAGLPHKRNSELINSLGLSPLMNEAGRR

U. Prepro-pyrokinin (C-terminal partial protein from GDRN01096339)

+VLQDKRDSAFAFSPRLGKRVQKSAFAFAPRPGKRSNFAFAPRPGKKTGTVFAFAPRPGKRTNFAFAPRPGKKSNAFAFAPRPGKKTTFAFSPRLGKK
ADFAFAPRPGKRSDSPSETEDRQTGETWWIGEGASSTVTTQPPFLPPRLE

V. Prepro-red pigment concentrating hormone (N-terminal partial protein from GDRN01016537)

MGLLPFACIMGSPITCTHSLPLLLPQVRRAGVTLLVVALVVVALVSSVSAQLNFSPGWGKRAAATSSNNGGVEAVSGLHPSVGGAPGGVVPSSSS
GDSCGPVPVSAMHIIYRLIR+

W. Prepro-RYamide (C-terminal partial protein from GDRN01027614)

+KGISRFHVLIFLHAERSGFYANRYGRSQGIPEIKVRSRFFVGGSRYGKRSGVPALAEVPLPVVAPESEEVEGVGASLLLGDSVVCLLVDVPDIYRCL
K

X1. Prepro-short neuropeptide F Ia (N-terminal partial protein from GDRN01012387)

MGHATLYVPRSPGCGREPALGMPNTRGEEMGVNGVKCWVALVCCLLLCQLTTATPVDYDTLNDMYDLLAVHEVERRAPPSMRLRFGKRDMGWQVAQ
RSMPTLRLRFGKRNVDDEADPILDHDLIRKDARTPALRLRFGKRGASFGEDMASQE+

X2. Prepro-short neuropeptide F Ib (N-terminal partial protein from GDRN01012386)

MGHATLYVPRSPGCGREPALGMPNTRGEEMGVNGVKCWVALVCCLLLCQLTTATPVDYDTLNDMYDLLAVHEVERRAPPSMRLRFGKRDMGWQVAQ
RSMPTLRLRFGKRNVDDEEDMASQE+

Y. Prepro-SIFamide (from GDRN01087866)

MSMQMRVVAVAVVVVVALLLSSPVSAGYRKPPFNCSI FGKRSGADAVYEPGKSQALASVCQVALEACTLWFPGAEEK

Z1. Prepro-tachykinin-related peptide Ia (N-terminal partial protein from GDRN01093289)

MSRMWVWAAMVVVGVAVAASGAGQGEAGSDAPRPRRAPSGFLGMRGKKEAPSPPLQDAPAPPEDLLPALYQLDVPLRGKKAPSGFLGMRGKKEE
EEEEDE+

Z2. Prepro-tachykinin-related peptide Ib (N-terminal partial protein from GDRN01093287)

MPLAYVLRHQHLHSSRPAPASTPAGVRRMSRMWVWAAMVVVGVAVAASGAGQGEAGSDAPRPRRAPSGFLGMRGKKEAPSPPLQDAPAPPEDLLP
ALYQLDVPLRGKKAPSGFLGMRGKKEEEEEDE+