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

Andrew E. Christie

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

Correspondence to: Dr. Andrew E. Christie, Békésy Laboratory of Neurobiology, Pacific Biosciences Research Center, University of Hawaii at Manoa, 1993 East-West Road, Honolulu, HI 96822 USA. Phone: 808-956-5212; FAX: 808-956-6984; Email: crabman@pbrc.hawaii.edu

Key words: bioinformatics; neurohormone; neuropeptide; transcriptome shotgun assembly; Isoptera; termite

Abstract

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

1. Introduction

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

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

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

With the advent of high throughput nucleotide sequencing, large genomic and/or transcriptomic datasets have been generated for several members of the Hymenoptera, and peptidomes of several eusocial species have been deduced from these resources (*e.g.* the honey bee *Apis mellifera* [Hummon et al., 2006]). In contrast, essentially no information on the native peptide hormones present in termites currently exists. In the study presented here, the publicly accessible transcriptome shotgun assembly (TSA) data for the giant northern termite *Mastotermes darwiniensis* (**BioProject Accession No. PRJNA219576**; Misof et al., 2014) was mined for putative neuropeptide-encoding transcripts, which were subsequently used to predict a peptidome for this species via a well-vetted bioinformatics workflow (*e.g.* Christie, 2008a, 2008b, 2014a, 2014b, 2014c, 2014d, 2014e, 2014f, 2015a, 2015b; Christie and Chi, 2015a, 2015b; Christie et al., 2018a, 2010a, 2011a, 2011b, 2013; Gard et al., 2009; Ma et al., 2009, 2010). One hundred sixty-three distinct mature peptide structures were predicted for *M. darwiniensis*, including member of the adipokinetic hormone (AKH), adipokinetic hormone-corazonin-like peptide (ACP), allatostatin A (AST-A), allatostatin C (AST-C), allatotropin, bursicon β , CCHamide, corazonin, crustacean cardioactive peptide (CCAP), crustacean hyperglycemic hormone (CHH)/ion transport peptide (ITP), diuretic hormone 31 (DH31), diuretic hormone 44 (DH44), FMRFamide-like peptide (FLP), insulin-like peptide (ILP), leucokinin, myosuppressin, neuroparsin, neuropeptide F (NPF), orcokinin, pigment dispersing hormone (PDH), pyrokinin, RYamide, short neuropeptide F (sNPF), SIFamide, sulfakinin and tachykinin-related peptide (TRP) families. Taken collectively, these data not only greatly

expand the catalog of putative peptide hormones known for the Isoptera, but also provide a foundation for examining how peptidergic signaling systems may contribute to the developmental and behavioral control required for eusociality in this and other termite species.

2. Materials and methods

2.1. Database searches

Database searches were conducted on or before January 24, 2015, using methods modified from a well-vetted protocol (*e.g.* Christie, 2008a, 2008b, 2014a, 2014b, 2014c, 2014d, 2014e, 2014f, 2015a, 2015b; Christie and Chi, 2015a, 2015b; Christie et al., 2018a, 2010a, 2011a, 2011b; Gard et al., 2009; Ma et al., 2009, 2010). Specifically, the database of the online program tblastn (National Center for Biotechnology Information, Bethesda, MD; <http://blast.ncbi.nlm.nih.gov/Blast.cgi>) was set to “Transcriptome Shotgun Assembly (TSA)” and restricted to data from *M. darwiniensis* “(taxid:13139)”. Known insect or crustacean peptide precursors were input into tblastn as the query sequences, and all hits returned by a given search were fully translated using the “Translate” tool of ExPASy (<http://web.expasy.org/translate/>) and then checked manually for homology to the target query. The complete list of peptides families searched for in this study, as well as the specific queries used, are provide in Table 1; this table also provides the BLAST-generated maximum score and E-value for each of the transcripts identified as encoding a putative neuropeptide precursor.

2.2. Peptide prediction

The structures of mature peptides were predicted using a well-established freeware workflow (*e.g.* Christie, 2008a, 2008b, 2014a, 2014b, 2014c, 2014d, 2014e, 2014f, 2015a,

2015b; Christie and Chi, 2015a, 2015b; Christie et al., 2018a, 2010a, 2011a, 2011b, 2011c, 2013; Gard et al., 2009; Ma et al., 2009, 2010). Specifically, each of the deduced precursor proteins was assessed for the presence of a signal peptide using the online program SignalP 4.1 (<http://www.cbs.dtu.dk/services/SignalP/>; Petersen et al., 2011); the D-cutoff values for the program were set to “Sensitive”. Prohormone cleavage sites were identified based on the information presented in Veenstra (2000) and/or by homology to known arthropod pre/preprohormone processing schemes. When present, prediction of the sulfation state of tyrosine residues was conducted using the online program “Sulfinator” (<http://www.expasy.org/tools/sulfinator/>; Monigatti et al., 2002). Disulfide bonding between cysteine residues was predicted by homology to known peptide isoforms and/or by using the online program “DiANNA” (<http://clavius.bc.edu/~clotelab/DiANNA/>; Ferrè and Clote, 2005). Other post-translational modifications, *e.g.* cyclization of amino (N)-terminal glutamine/glutamic acid residues and C-terminal amidation at glycine residues, were predicted by homology to known arthropod peptide isoforms. Figure 1 shows three examples of mature peptide structural prediction using the workflow just described; the mature structures of all peptides predicted in this study are provided in Table 2. All protein/peptide alignments were done using the online program MAFFT version 7 (<http://mafft.cbrc.jp/alignment/software/>; Katoh and Standley, 2013).

3. Results

In this study, 36 distinct peptide families/subfamilies were searched for within the publicly accessible *M. darwiniensis* TSA sequence database (Table 1). In the interest of space, only those searches that resulted in the identification of putative precursor-encoding transcripts are described here (Table 1), with the data presented in alphabetical order based on family name.

All precursor proteins listed as “full-length” exhibit a functional signal sequence (including a “start” methionine) and are flanked on their C-terminus end by a stop codon. Proteins described here as “partial” lacked a start methionine (referred to as C-terminal partial proteins), a stop codon (referred to as N-terminal partial proteins), or both of these features (referred to as internal protein fragments).

3.1. Adipokinetic hormone

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

3.2. Adipokinetic hormone-corazonin-like peptide

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

pQVTFSRDWNAamide, while new to *M. darwiniensis*, is a known insect/crustacean ACP isoform (Christie, 2014a; Hansen et al., 2010).

3.3. Allatostatin A

One *M. darwiniensis* transcript was identified as encoding a putative AST-A precursor using *D. melanogaster* prepro-AST-A (**Accession No. AAF56331**; Adams et al., 2000) as the query sequence (Table 1). Translation of this transcript revealed a 385 amino acid full-length precursor protein (Masda-prepro-AST-A; Fig. 2B). Twenty-one distinct peptides were predicted from Masda-prepro-AST-A (Table 2 and Fig. 2B), nine of which possess the C-terminal motif – YXFGLamide, where *X* represents a variable residue, or a close approximation thereof, the hallmark of members of the AST-A family (Christie et al., 2010b).

3.4. Allatostatin C

A single *M. darwiniensis* transcript was identified as encoding a putative AST-C precursor using *Homarus americanus* prepro-AST-C (Dickinson et al., 2009a) as the query sequence (Table 1). Translation of this transcript revealed a 96 amino acid full-length preprohormone (Masda-prepro-AST-C; Fig. 1B). Two distinct peptides were predicted from Masda-prepro-AST-C, one of which, SYWKQCAFNAVSCFamide (disulfide bridging between the two cysteine residues), while new to *M. darwiniensis*, is a known and broadly conserved arthropod AST-C variant (e.g. Dickinson et al., 2009a; Hummon et al., 2006; Veenstra, 2009).

3.5. Allatotropin

One *M. darwiniensis* transcript was identified as encoding a putative allatotropin precursor using *Manduca sexta* prepro-allatotropin (**Accession No. AAB08759**; Taylor et al., 1996) as the query sequence (Table 1). Translation of this transcript revealed a 122 amino acid full-length preprohormone (Masda-prepro-allatotropin; Fig. 2C). Four distinct peptides were predicted from Masda-prepro-allatotropin, including GFKNVALSTARGFamide, a peptide that possesses the C-terminal motif –TARGFamide, the structural hallmark of members of the allatotropin family (Elekonich and Horodyski, 2003).

3.6. Bursicon β

A single *M. darwiniensis* transcript was identified as encoding a putative bursicon β precursor using *D. melanogaster* pre-partner of bursicon (**Accession No. AAF53396**; Adams et al., 2000) as the query sequence (Table 1); partner of bursicon is a synonym for bursicon β . Translation of this transcript revealed a 143 amino acid full-length prehormone (Masda-pre-bursicon β ; Fig. 2D). A 116 amino acid peptide was predicted from Masda-pre-bursicon β (Table 2 and Fig. 2D). A blastp search of the extant arthropod proteins found the amino acid sequence of this peptide to be most similar to that of a bursicon β predicted from a *Zootermopsis nevadensis* (another termite) precursor (**Accession No. KDR13885**; Terrapon et al., 2014); the *M. darwiniensis* peptide is N-terminally extend by 2 amino acids relative to its *Z. nevadensis* counterpart, with just 10 substituted residues in the remaining sequence, 9 being conservative amino acid substitutions (alignment not shown). Analysis of the *M. darwiniensis* bursicon β by DiANNA suggests disulfide bridging between its first and third, second and eighth, fifth and seventh, sixth and tenth, and ninth and eleventh cysteines (Table 2).

3.7. CCHamide

One *M. darwiniensis* transcript was identified as encoding a putative CCHamide precursor using *D. melanogaster* prepro-CCHamide 1 (**Accession No. AAF55014**; Adams et al., 2000) as the query sequence (Table 1). Translation of this transcript revealed a 132 amino acid full-length preprohormone (Masda-prepro-CCHamide; Fig. 2E). Three distinct peptides were predicted from Masda-prepro-CCHamide, one of which, GCSAFGHSCFGGHamide (disulfide bridging between the position 2 and 9 cysteines), possesses the structural hallmarks of members of the CCHamide family, *i.e.* XCXXW/Y/FGXXCXGXHamide with disulfide bridging between the position 2 and 9 cysteine residues (Hansen et al., 2011).

3.8. Corazonin

A single *M. darwiniensis* transcript was identified as encoding a putative corazonin precursor using *D. melanogaster* prepro-corazonin (**Accession No. AAF55046**; Adams et al., 2000) as the query sequence (Table 1). Translation of this transcript revealed a 136 amino acid full-length preprohormone (Masda-prepro-corazonin; Fig. 2F). Three distinct peptides were predicted from Masda-prepro-corazonin (Table 2 and Fig. 2F), including pQTFQYSRGWTNamide, a peptide whose structure is identical to that of authentic corazonin, a broadly conserved arthropod neuropeptide (*e.g.* Christie et al., 2010b; Hauser and Grimmelikhuijzen, 2014). DiANNA analysis of the longer of the two linker/precursor-related peptide predicted from Masda-prepro-corazonin, GGPAPPMFIPSSASGAVLQNLDESSVISNPCSQLQRIRFLLGARNPQQFYYP CDSWRDISE TPSEEVSERF, suggests a disulfide bridge is present between its two cysteine residues (Table 2).

3.9. Crustacean cardioactive peptide

One *M. darwiniensis* transcript was identified as encoding a putative CCAP precursor using *D. melanogaster* prepro-CCAP (**Accession No. AAF56045**; Adams et al., 2000) as the query sequence (Table 1). Translation of this transcript revealed a 167 amino acid full-length preprohormone (Masda-prepro-CCAP; Fig. 2G). Six distinct peptides were predicted from Masda-prepro-CCAP (Table 2 and Fig. 2G), one of which, PFCNAFTGCamide (disulfide bridging between the two cysteine residues), possesses a structure identical to authentic CCAP, a broadly conserved arthropod neuropeptide (*e.g.* Altstein and Nässel, 2010; Christie et al., 2010b; Nässel and Winther, 2010).

3.10. Crustacean hyperglycemic hormone/ion transport peptide

A single *M. darwiniensis* transcript was identified as encoding a putative CHH/ITP precursor using *D. melanogaster* prepro-ITP (**Accession No. ABI31114**; Adams et al., 2000) as the query sequence (Table 1). Translation of this transcript revealed a 115 amino acid full-length preprohormone (Masda-prepro-CHH/ITP; Fig. 2H). Two distinct peptides were predicted from Masda-prepro-CHH/ITP (Table 2 and Fig. 2H), including one, SFFEIQCKGVYDKSIFARLDRICEDCYNLFREPQLHSLCRSKCFSSDYFKGCLEALLLKQE EQKFNQMVVEYVamide, that possesses structural similarity to member of the CHH/ITP family, *e.g.* disulfide bonding between the first and fifth, second and fourth, and third and sixth cysteine residues in this peptide (Table 2), a pattern of disulfide bridging that is characteristic of CHH superfamily members (Böcking et al., 2002).

3.11. Diuretic hormone 31

Two *M. darwiniensis* transcripts were identified as encoding putative DH31 precursors using *D. melanogaster* prepro-DH31 (**Accession No. AAF52685**; Adams et al., 2000) as the query sequence (Table 1). Translation of one of these transcripts revealed a 115 amino acid full-length preprohormone (Masda-prepro-DH31 I; Fig. 2I1), while translation of the other revealed an 82 amino acid N-terminal partial protein (Masda-prepro-DH31 II; Fig. 2I2). Four distinct peptides were predicted from Masda-prepro-DH31 I (Table 2 and Fig. 2I1), one of which, GLDLGLSRGFSGSQAAKHLMGLAAANYAGGPamide, possesses the structural hallmarks of the DH31 family (*i.e.* an overall length of 31 amino acids, an amidated C-terminus, and 13 highly conserved residues, **XXDXGLXRGXSGXXXAKXXXXXXXXANXXXGPamide** [Christie et al., 2010b]). Two peptides were predicted from Masda-prepro-DH31 II (Table 2 and Fig. 2I2), a full-length linker/precursor-related peptide that is distinct in sequence from those derived from Masda-prepro-DH31 I, and a 16 amino acid partial peptide that is identical in sequence to the first 16 amino acids of the DH31 derived from Masda-prepro-DH31 I (given this sequence conservation, the partial peptide has not be included in Table 2).

3.12. Diuretic hormone 44

One *M. darwiniensis* transcript was identified as encoding a putative DH44 precursor using *D. melanogaster* prepro-DH44 (**Accession No. AAF54421**; Adams et al., 2000) as the query sequence (Table 1). Translation of this transcript revealed a 198 amino acid full-length preprohormone (Masda-prepro-DH44; Fig. 2J). Three distinct peptides were predicted from Masda-prepro-DH44 (Table 2 and Fig. 2J), including TGAGPSLSIVNPLDVLQRLLLEIARRRMRSQDQIQANREILQSIamide, a putative DH44

isoform. A search of the extant arthropod protein database identified an isoform of corticotropin-releasing factor-like diuretic hormone (a synonym for DH44) from the cockroach *Periplaneta americana* (**Accession No. P41538**; Kay et al., 1992) as the most similar sequence to the DH44 predicted from Masda-prepro-DH44; the two peptides differ in sequence at just two residues, both conservative amino acid substitutions (alignment not shown).

3.13. *FMRFamide-like peptide*

A single *M. darwiniensis* transcript was identified as encoding a putative FLP precursor using *D. melanogaster* prepro-FLP (**Accession No. AAF58874**; Adams et al., 2000) as the query sequence (Table 1). Translation of this transcript revealed a 170 amino acid N-terminal partial protein (Masda-prepro-FLP; Fig. 2K). Nine distinct peptides were predicted from the extant portion of Masda-prepro-FLP (Table 2 and Fig. 2K), including seven possessing C-termini that are similar to the –FMRFamide consensus motif of the prototypical members of the FLP family (e.g. Christie et al., 2010b), i.e. –FIRFamide (four peptides), –FIRLamide (one peptide), –FVRLamide (one peptide) or –LLRLamide (one peptide).

3.14. *Insulin-like peptide*

Four *M. darwiniensis* transcripts were identified as encoding putative ILP precursors using *D. melanogaster* prepro-ILP 2 (**Accession No. AAF50204**; Adams et al., 2000) as the query sequence (Table 1). Translation of these transcripts revealed all to encode full-length preprohormones, with two transcripts encoding identical 147 amino acid preprohormones (Masda-prepro-ILP I; Fig. 2L1), another a 132 amino acid precursor (Masda-prepro-ILP II; Fig. 2L2), and the final transcript, a 135 amino acid protein (Masda-prepro-ILP III; Fig. 2L3). Four

peptides were predicted from both Masda-prepro-ILP I and Masda-prepro-ILP II, with five
 peptides predicted from Masda-prepro-ILP III; all of the peptides predicted from the collective
 set of *M. darwiniensis* precursors possess distinct structures (Table 2 and Fig. 2L). The peptides
 predicted from Masda-prepro-ILP I and Masda-prepro-ILP II include one isoform each with
 sequence homology to the A-, B- and C-chain ILPs derived from known ILP precursors, *e.g.*
 those derived from *Daphnia pulex* prepro-ILP 2 (Dirksen et al., 2011), *i.e.*
 ERSDGGIVDECCIIYKGCTVNELTEY_(SO₃H)CSS,
 DATHRYCGSNLVTTLRFLCRSQYNGIDDE and
 NAEPIVKKSLQVSDPDNSLWTQLQTVDELPEFPPFRSRSIANTFANRVF from Masda-
 prepro-ILP I and HTRGVYDECCRKSCTIQEMASYCamide,
 NTNKYCGRNLANMLRYVCNGNYYPMF and
 ASQDVEDTNDSGIWILSPTEEPQLQFPFRSRSNAASLVPGSF from Masda-prepro-ILP II,
 respectively. Both an A- and B-chain ILP were also among the peptide predicted from Masda-
 prepro-ILP III, *i.e.* pQTRDVGIVEECCMFKGCTLNELTEY_(SO₃H)CAD and
 pEAAFRYCGPNLANILRLLCNGSYHTDEHW, respectively. Homology to known A- and B-
 chain ILPs, *e.g.* those from *D. pulex* (Dirksen et al., 2011), suggests disulfide bridging to be
 present within the A-chain peptide (the first and third cysteines bridged [Table 2]) and between it
 and its corresponding B-chain isoform (*i.e.* the first cysteine in the B-chain bridged to the second
 cysteine in the A-chain and the second cysteine in the B-chain bridged to the fourth cysteine in
 the A-chain [these bridged cysteines are not highlighted in the peptide structures shown in Table
 2]).

3.15. Leucokinin

Two *M. darwiniensis* transcripts were identified as encoding putative leucokinin precursors using *Eucyclops serrulatus* prepro-leucokinin (Christie, 2015b) as the query sequence (Table 1). Translation of these transcripts revealed one to encode a 141 amino acid internal fragment of a preprohormone (Masda-prepro-leucokinin I; Fig. 2M1), and the other a 291 amino acid C-terminal partial protein (Masda-prepro-leucokinin II; Fig. 2M2); there is no region of overlap between these two partial sequences. Eight distinct peptides (six full-length and two partial) were predicted from the extant portion of Masda-prepro-leucokinin I (Table 2 and Fig. 2M1), five of which possess the C-terminal motif –FXXWA/Gamide, the hallmark of members of the leucokinin family (Christie et al., 2010b). Fifteen peptides, each distinct from one another, as well as from those derived from Masda-prepro-leucokinin I, were predicted from Masda-prepro-leucokinin II (Table 2 and Fig. 2M1). Of these peptides, seven (six full-length and one partial) possess structures placing them into the leucokinin family (Table 2 and Fig. 2M1).

3.16. *Myosuppressin*

One *M. darwiniensis* transcript was identified as encoding putative myosuppressin precursor using *D. melanogaster* prepro-myosuppressin (**Accession No. AAF56283**; Adams et al., 2000) as the query sequence (Table 1). Translation of this transcript revealed a 98 amino acid full-length preprohormone (Masda-prepro-myosuppressin; Fig. 2N). Two distinct peptides were predicted from Masda-prepro-myosuppressin (Table 2 and Fig. 2N), one of which, pQDVDHVFLRFamide, possesses the C-terminal motif –HVFLRFamide, the defining feature of the myosuppressin family (Christie et al., 2010b).

3.17. *Neuroparsin*

A single *M. darwiniensis* transcript was identified as encoding putative neuroparsin precursor using *Schistocerca gregaria* pre-neuroparsin 1 (**Accession No. CAC38869**; Janssen et al., 2001) as the query sequence (Table 1). Translation of this transcript revealed a 106 amino acid full-length prehormone (Masda-pre-neuroparsin; Fig. 2O). A single peptide was predicted from Masda-pre-neuroparsin (Table 2 and Fig. 2O). A blastp search of the extant arthropod protein sequences identified neuroparsin A from *Z. nevadensis* (**Accession No. KDR17790**; Terrapon et al., 2014) as the most similar peptide to the neuroparsin derived from Masda-pre-neuroparsin; the *Z. nevadensis* peptide is N-terminally extended by one amino acid relative to the *M. darwiniensis* neuroparsin, but otherwise the two peptides differ in sequence at just ten residues, all but three being conservative amino acid substitutions (alignment not shown). Analysis of the *M. darwiniensis* neuroparsin by DiANNA suggest disulfide bridging between its first and sixth, second and third, fourth and twelfth, fifth and ninth, seventh and eighth, tenth and fourteenth, and eleventh and thirteenth cysteine residues.

3.18. Neuropeptide F

Two *M. darwiniensis* transcripts were identified as encoding putative NPF precursors using *D. melanogaster* prepro-NPF (**Accession No. AAF55339**; Adams et al., 2000) as the query sequence (Table 1). Translation of these transcripts revealed each to encode a distinct full-length preprohormone, with one protein being 91 amino acids long (Masda-prepro-NPF I; Fig. 2P1), and the other, 124 amino acids in length (Masda-prepro-NPF II; Fig. 2P2). Three distinct peptides were predicted from each of the *M. darwiniensis* NPF precursors (Table 2 and Fig. 2P), one from each protein possessing structural similarity to member of the NPF family, *i.e.* an overall length of 30 or more amino acids, the C-terminal motif –GRPRFamide (or a close

approximation thereof), and tyrosine residues at positions 10 and 17 from the C-terminus (Christie et al., 2010b): KPTDPEQLAAMADTLKYLQELDRYYSQVARPRFamide from Masda-prepro-NPF I and PKVFTSPDQLRVYLQELGNYYAIEGRPRFamide from Masda-prepro-NPF II.

3.19. Orcokinin

Two *M. darwiniensis* transcripts were identified as encoding putative orcokinin precursors using *H. americanus* prepro-orcokinin I (Accession No. ACB41787; Dickinson et al., 2009b) as the query sequence (Table 1). Translation of these transcripts revealed each to encode an identical 164 amino acid preprohormone (Masda-prepro-orcokinin; Fig. 2Q). Nine distinct peptides were predicted Masda-prepro-orcokinin (Table 2 and Fig. 2Q), two of which possess the N-terminal consensus motif NFDEIDR-, the hallmark of the orcokinin family (Christie et al., 2010b).

3.20. Pigment dispersing hormone/factor

One *M. darwiniensis* transcript was identified as encoding a putative PDH precursor using *D. melanogaster* prepro-pigment dispersing factor (Accession No. AAF56593; Adams et al., 2000) as the query sequence (Table 1). Translation of this transcript revealed a 41 amino acid C-terminal partial protein (Masda-prepro-PDH; Fig. 2R). Two distinct peptides were predicted from the extant portion of Masda-prepro-PDH (Table 2 and Fig. 2R), including NSEIINSILGLPKNLNEAamide, a peptide that differs at just three residues (two of which are conservative amino acid substitutions) from that of a known PDH from the crayfish *Procambarus clarkii* (Accession No. Q9TWW7; McCallum et al., 1991).

3.21. Pyrokinin

A single *M. darwiniensis* transcript was identified as encoding putative pyrokinin precursor using *D. melanogaster* prepro-capability (**Accession No. AAF56969**; Adams et al., 2000) as the query sequence (Table 1); capability (CAPA) is one group of peptides contained within the CAPA/periviscerokinin/pyrokinin superfamily (Predel and Wegener, 2006). Translation of this transcript revealed a 185 amino acid full-length prehormone (Masda-prepro-pyrokinin; Fig. 2S). Eight distinct peptides were predicted Masda-prepro-pyrokinin (Table 2 and Fig. 2S), five of which possess C-termini similar to the –LXXPRXamide consensus motif that defines the CAPA/periviscerokinin/pyrokinin superfamily (Christie et al., 2010b). It should be noted that the five pyrokinins identified here all possess structures distinct from the CAPA isoforms identified previously from *M. darwiniensis* (Roth et al., 2009).

3.22. RYamide

One *M. darwiniensis* transcript was identified as encoding putative RYamide precursor using *P. clarkii* prepro-RYamide (Christie and Chi, 2015b) as the query sequence (Table 1). Translation of this transcript revealed a 115 amino C-terminal partial protein (Masda-prepro-RYamide; Fig. 2T). Five distinct peptides were predicted from the extant portion of Masda-prepro-RYamide (Table 2 and Fig. 2T), two of which possess the C-terminal motif –RYamide, the defining feature of the RYamide family (Christie et al., 2010b). It should be noted that the partial peptide –Yamide was also predicted from Masda-prepro-RYamide; this may represent a portion of an additional RYamide family member (this partial peptide is not included in Table 2 or the total peptide count reported in this study). DiANNA analysis of the longest of the

linker/precursor-related peptide predicted from Masda-prepro-RYamide, SEEPTTTTDETVGFGDMLLAEEEDDNSQVTCLYTGVTNLYRCYY, suggests a disulfide bridge is present between its two cysteine residues (Table 2).

3.23. Short neuropeptide F

A single *M. darwiniensis* transcript was identified as encoding putative sNPF precursor using *D. melanogaster* prepro-sNPF (**Accession No. AAN11060**; Adams et al., 2000) as the query sequence (Table 1). Translation of this transcript revealed a 101 amino full-length preprohormone (Masda-prepro-sNPF; Fig. 2U). Three distinct peptides were predicted Masda-prepro-sNPF (Table 2 and Fig. 2U), including one, ANRSPSLRLRFamide, that possesses the structural hallmarks of the sNPF family, *i.e.* the C-terminal motif–RXRFamide and an overall length of 10 or so amino acids (Christie et al., 2010b).

3.24. SIFamide

Two *M. darwiniensis* transcripts were identified as encoding putative SIFamide precursors using *D. melanogaster* prepro-SIFamide (**Accession No. AFH08249**; Adams et al., 2000) as the query sequence (Table 1). Translation of one transcript revealed a 72 amino full-length preprohormone (Masda-prepro-SIFamide I; Fig. 2V1), while translation of the other revealed an 80 amino acid full-length precursor protein (Masda-prepro-SIFamide II; Fig. 2V2). Two distinct peptides were predicted Masda-prepro-SIFamide I (Table 2 and Fig. 2V1), including TYRKPPFNGSIFamide, a peptide that possesses the –SIFamide C-terminus that is the hallmark of the SIFamide family (Christie et al., 2010b). Three distinct peptide were predicted from Masda-prepro-SIFamide II (Table 2 and Fig. 2V2), one of which LPFNGSIYamide shows

some sequence similarity to members of the SIFamide family. DiANNA analysis of the longer of the two linker/precursor-related peptides derived from Masda-prepro-SIFamide II, AASALPMDYDSSSKSLSSLCETVAEVCAASFPQQTEN, suggests the presence of disulfide bridge between its two cysteine residues (Table 2).

3.25. Sulfakinin

One *M. darwiniensis* transcript was identified as encoding a putative sulfakinin precursor using *D. melanogaster* prepro-drosulfakinin (**Accession No. AAF52173**; Adams et al., 2000) as the query sequence (Table 1). Translation of this transcript revealed a 124 amino full-length preprohormone (Masda-prepro-sulfakinin; Fig. 1C). Four distinct peptides were predicted Masda-prepro-sulfakinin (Table 2 and Fig. 1C), including two, pQSEDY_(SO₃H)GHMRFamide and pEQFDDY_(SO₃H)GHMRFamide, that possess the C-terminal motif –Y_(SO₃H)GHM/LRFamide, the hallmark of members of the sulfakinin family (Christie et al., 2010b).

3.26. Tachykinin-related peptide

A single *M. darwiniensis* transcript was identified as encoding putative TRP precursor using *D. melanogaster* prepro-TRP (**Accession No. AAF54735**; Adams et al., 2000) as the query sequence (Table 1). Translation of this transcript revealed a 337 amino full-length preprohormone (Masda-prepro-TRP; Fig. 2W). Nineteen distinct peptides were predicted Masda-prepro-TRP (Table 2 and Fig. 2W), including ten that possess –FXGXRamide C-termini, or a close approximation thereof, the hallmark of members of the TRP family (Christie et al., 2010b).

4. Discussion

4.1. Transcriptome shotgun assemblies: a growing resource for peptidome prediction in little studied arthropod taxa

Technological advancement and decreasing costs for high-throughput nucleotide sequencing has lead to the ability to produce extensive genomic and/or transcriptomic datasets for a large number of organisms, including many arthropod species. Interest in establishing phylogenetic relationships within the Arthropoda has necessitated deep sequencing from a number of species, including ones from taxa that have until now had few molecular resources available for them (*e.g.* Misof et al., 2014; von Reumont et al., 2012). While the impetus behind these large-scale sequencing projects rarely has been neuropeptide discovery, the generation and public deposition of nucleotide sequence data from whole organisms, or from tissues that include neural tissue, often allows for the identification of extensive collections of transcripts encoding neuropeptide precursor proteins, and from these sequences, the structures of putative mature peptide isoforms can readily be predicted using simple bioinformatics workflows that employ software programs freely accessible via the internet. Via this strategy, large peptidomes have recently been predicted for members of many little studied arthropod taxa, including, but not limited to, the Remipedia, Copepoda, Branchiura, Amphipoda, and Araneae (*e.g.* Christie, 2014a, 2014b, 2014c, 2014d, 2014e, 2015a, 2015b; Christie and Chi, 2015a). Here, the publicly accessible TSA sequences for *M. darwiniensis* (Misof et al., 2014) were used to predict the first large-scale peptidome from a member of the Isoptera; *M. darwiniensis* is the most basal living member of the Isoptera and is a species with highly derived eusociality (*e.g.* Inward et al., 2007; Shellman-Reeve, 1997). Given the interest in termites as pest species, as well as in understanding the hormonal contributions to the developmental and behavioral control that

underlie the evolution of eusociality, the data presented here serve as a template for future peptide discovery in the Isoptera, and as a foundation for physiological investigations of peptidergic signaling, including its contribution to differential developmental and behavioral control, in *M. darwiniensis*.

4.2. The predicted peptidome of *Mastotermes darwiniensis*

The data presented in this study represent the first large-scale peptidome predicted for any member of the Isoptera. Using the BLAST algorithm `tblastn` and known arthropod protein queries, 34 *M. darwiniensis* transcripts encoding putative neuropeptide precursor proteins were identified. By translating these sequences, a set of pre/preprohormones was obtained, and using a well-vetted workflow, the mature structures of the peptides likely liberated from them were predicted. In total, the structures of 163 distinct peptides were obtained. These peptides included members of the AKH, ACP, AST-A, AST-C, allatotropin, bursicon β , CCHamide, corazonin, CCAP, CHH/ITP, DH31, DH44, FLP, ILP, leucokinin, myosuppressin, neuroparsin, NPF, orcokinin, PDH, pyrokinin, RYamide, sNPF, SIFamide, sulfakinin and TRP families. The vast majority of the peptides predicted here are novel discoveries for *M. darwiniensis*.

Interestingly, and despite the identification of over 160 peptides encompassing 26 different families, several well-known arthropod peptide groups were not identified here. For example, no transcripts encoding members of the allatostatin B (AST-B), bursicon α , ecdysis-triggering hormone (ETH), eclosion hormone (EH), intocin or proctolin families were found. Whether or not these peptide groups are truly absent in *M. darwiniensis* remains an open question. While it is possible that one or more of these families may have been lost in this species, the high level of conservation of each of these peptide groups in the Arthropoda (*e.g.*

Altstein and Nässel, 2010; Christie et al., 2010b; Nässel and Winther, 2010) would suggest that the most likely explanation for their lack of detection is that the TSA dataset for *M. darwiniensis* does not provide full coverage for this species, and that transcripts encoding the AST-B, bursicon α , ETH, EH, intocin and proctolin precursors are simply not present in the extant TSA data.

4.3. Peptide hormones and eusociality

Circulating hormones, including a wide array of peptides, are key components in most developmental and physiological/behavioral control systems. A fundamental characteristic of eusociality is the development of physiologically/behaviorally distinct castes within a colony of individuals. In *M. darwiniensis*, larvae initially enter one of two developmental pathways: the alate pathway or the worker pathway (e.g. Watson et al. 1977). In the alate pathway, larvae transition through several nymph stages, ultimately molting into sclerotized, reproductive adults (alates) that function to provide a means of dispersal, and hence an opportunity for the development of new colonies (Watson et al. 1977). Larvae that enter the worker pathway cannot become alates, but rather transition into workers, and either remain within this caste or subsequently enter into developmental program generating soldiers or one that produces reproductive neotenics (Watson et al. 1977). While as of yet untested, molting and metamorphosis in *M. darwiniensis* are likely to involve hormone control by a number of peptides including bursicon and CCAP, two peptides identified here and ones that are highly conserved in the regulation of these developmental processes in other insects (White and Ewer, 2014). Similarly, the process of sclerotization in alates may well involve hormonal control by bursicon (Dewey et al., 2004), while the differential control of aggression between caste types may be modulated by NPF and/or SIFamide (Dierick and Greenspan, 2007; Vázquez-Acevedo, 2009), as

they are in other arthropods. Clearly the wealth of sequence data presented here provide an opportunity to begin to directly test the roles played by the native peptides of *M. darwiniensis*, and as such investigations are undertaken it will interesting to see how differential hormonal control of development and/behavior may contribute to eusociality in this and other species exhibiting this advanced level of communal organization.

Acknowledgements

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

References

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N., Sutton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X., Brandon, R.C., Rogers, Y.H., Blazej, R.G., Champe, M., Pfeiffer, B.D., Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor, G.L., Abril, J.F., Agbayani, A., An, H.J., Andrews-Pfannkoch, C., Baldwin, D., Ballew, R.M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y., Benos, P.V., Berman, B.P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M.R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K.C., Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B., Davies, P., de Pablos, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M., Dodson, K., Doup, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C., Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferriera, S., Fleischmann, W., Fosler, C., Gabrielian, A.E., Garg, N.S., Gelbart,

547 W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z., Guan, P., Harris, M., Harris,
 548 N.L., Harvey, D., Heiman, T.J., Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland,
 549 T.J., Wei, M.H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z., Kennison, J.A.,
 550 Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasko, P.,
 551 Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C.,
 552 McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Mobarry, C., Morris, J., Moshrefi,
 553 A., Mount, S.M., Moy, M., Murphy, B., Murphy, L., Muzny, D.M., Nelson, D.L., Nelson, D.R.,
 554 Nelson, K.A., Nixon, K., Nusskern, D.R., Pacleb, J.M., Palazzolo, M., Pittman, G.S., Pan, S.,
 555 Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K., Saunders, R.D., Scheeler, F.,
 556 Shen, H., Shue, B.C., Sidén-Kiamos, I., Simpson, M., Skupski, M.P., Smith, T., Spier, E.,
 557 Spradling, A.C., Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R., Venter,
 558 E., Wang, A.H., Wang, X., Wang, Z.Y., Wassarman, D.A., Weinstock, G.M., Weissenbach, J.,
 559 Williams, S.M., Woodage, T., Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F.,
 560 Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhong, W.,
 561 Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M., Venter, J.C.,
 562 2000. The genome sequence of *Drosophila melanogaster*. Science 287, 2185-2195.
 563
 564 Altstein, M., Nässel, D.R., 2010. Neuropeptide signaling in insects. Adv. Exp. Med. Biol. 692,
 565 155-165.
 566
 567 Böcking, D., Dirksen, H., Keller, R., 2002. The crustacean neuropeptide of the CHH/MIH/GIH
 568 family: structures and biological activities. In: The Crustacean Nervous System. K Wiese (Ed.).
 569 Springer, Heidelberg. pp. 84–97.

570

571 Christie, A.E., 2008a. Neuropeptide discovery in Ixodoidea: an *in silico* investigation using
572 publicly accessible expressed sequence tags. Gen. Comp. Endocrinol. 157, 174-185.

573

574 Christie, A.E., 2008b. *In silico* analyses of peptide paracrine/hormones in Aphidoidea. Gen.
575 Comp. Endocrinol. 159, 67-79.

576

577 Christie, A.E., 2014a. Prediction of the first neuropeptides from a member of the Remipedia
578 (Arthropoda, Crustacea). Gen. Comp. Endocrinol. 201, 74-86.

579

580 Christie, A.E., 2014b. Prediction of the peptidomes of *Tigriopus californicus* and
581 *Lepeophtheirus salmonis* (Copepoda, Crustacea). Gen. Comp. Endocrinol. 201, 87-106.

582

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

586

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

589

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

592

593 Christie, A.E., 2014f. Expansion of the *Litopenaeus vannamei* and *Penaeus monodon*
 594 peptidomes using transcriptome shotgun assembly sequence data. Gen. Comp. Endocrinol. 206,
 595 235-254.
 596
 597 Christie, A.E., 2015a. *In silico* characterization of the neuropeptidome of the Western black
 598 widow spider *Latrodectus hesperus*. Gen. Comp. Endocrinol. 210, 63-80.
 599
 600 Christie, A.E., 2015b. Neuropeptide discovery in *Eucyclops serrulatus* (Crustacea, Copepoda): in
 601 silico prediction of the first peptidome for a member of the Cyclopoida. Gen. Comp. Endocrinol.
 602 211, 92-105.
 603
 604 Christie, A.E., Cashman, C.R., Brennan, H.R., Ma, M., Sousa, G.L., Li, L., Stemmler, E.A.,
 605 Dickinson, P.S., 2008a. Identification of putative crustacean neuropeptides using *in silico*
 606 analyses of publicly accessible expressed sequence tags. Gen. Comp. Endocrinol. 156, 246-64.
 607
 608 Christie, A.E., Chi, M., 2015a. Neuropeptide discovery in the Araneae (Arthropoda, Chelicerata,
 609 Arachnida): elucidation of true spider peptidomes using that of the Western black widow as a
 610 reference. Gen. Comp. Endocrinol. In press.
 611
 612 Christie, A.E., Chi, M., 2015b. Prediction of the neuropeptidomes of members of the Astacidea
 613 (Crustacea, Decapoda) using publicly accessible transcriptome shotgun assembly (TSA)
 614 sequence data. Gen. Comp. Endocrinol. In press.
 615

616 Christie, A.E., Durkin, C.S., Hartline, N., Ohno, P., Lenz, P.H., 2010a. Bioinformatic analyses
 617 of the publicly accessible crustacean expressed sequence tags (ESTs) reveal numerous novel
 618 neuropeptide-encoding precursor proteins, including ones from members of several little studied
 619 taxa. *Gen. Comp. Endocrinol.* 167, 164-178.
 620
 621 Christie, A.E., McCoole, M.D., Harmon, S.M., Baer, K.N., Lenz, P.H., 2011c. Genomic
 622 analyses of the *Daphnia pulex* peptidome. *Gen. Comp. Endocrinol.* 171, 131-150.
 623
 624 Christie, A.E., Nolan, D.H., Garcia, Z.A., McCoole, M.D., Harmon, S.M., Congdon-Jones, B.,
 625 Ohno, P., Hartline, N., Congdon, C.B., Baer, K.N., Lenz, P.H., 2011b. Bioinformatic prediction
 626 of arthropod/nematode-like peptides in non-arthropod, non-nematode members of the
 627 Ecdysozoa. *Gen. Comp. Endocrinol.* 170, 480-486.
 628
 629 Christie, A.E., Nolan, D.H., Ohno, P., Hartline, N., Lenz, P.H., 2011a. Identification of
 630 chelicerate neuropeptides using bioinformatics of publicly accessible expressed sequence tags.
 631 *Gen. Comp. Endocrinol.* 170, 144-155.
 632
 633 Christie, A.E., Roncalli, V., Wu, L.S., Ganote, C.L., Doak, T., Lenz, P.H., 2013. Peptidergic
 634 signaling in *Calanus finmarchicus* (Crustacea, Copepoda): *in silico* identification of putative
 635 peptide hormones and their receptors using a *de novo* assembled transcriptome. *Gen. Comp.*
 636 *Endocrinol.* 187:117-35.
 637

638 Christie, A.E., Stemmler, E.A., Dickinson, P.S., 2010b. Crustacean neuropeptides. *Cell. Mol.*
 639 *Life Sci.* 67, 4135-4169.
 640
 641 Cornette, R., Gotoh, H., Koshikawa, S., Miura, T., 2008. Juvenile hormone titers and caste
 642 differentiation in the damp-wood termite *Hodotermopsis sjostedri* (Isoptera: Termopsidae). *J.*
 643 *Insect Physiol.* 54, 922-930.
 644
 645 Dewey, E.M., McNabb, S.L., Ewer, J., Kuo, G.R., Takanishi, C.L., Truman, J.W., Honegger,
 646 H.W., 2004. Identification of the gene encoding bursicon, an insect neuropeptide responsible for
 647 cuticle sclerotization and wing spreading. *Curr. Biol.* 14, 1208-1213.
 648
 649 Dickinson, P.S., Stemmler, E.A., Barton, E.E., Cashman, C.R., Gardner, N.P., Rus, S., Brennan,
 650 H.R., McClintock, T.S., Christie, A.E., 2009a. Molecular, mass spectral, and physiological
 651 analyses of orcokinins and orcokinin precursor-related peptides in the lobster *Homarus*
 652 *americanus* and the crayfish *Procambarus clarkii*. *Peptides* 30, 297-317.
 653
 654 Dickinson, P.S., Wiwatpanit, T., Gabranski, E.R., Ackerman, R.J., Stevens, J.S., Cashman, C.R.,
 655 Stemmler, E.A., Christie, A.E., 2009a. Identification of SYWKQCAFNAVSCFamide: a
 656 broadly conserved crustacean C-type allatostatin-like peptide with both neuromodulatory and
 657 cardioactive properties. *J. Exp. Biol.* 212, 1140-1152.
 658
 659 Dierick, H.A., Greenspan, R.J., 2007. Serotonin and neuropeptide F have opposite modulatory
 660 effects on fly aggression. *Nat. Genet.* 39, 678-682.

661

662 Dircksen, H., Neupert, S., Predel, R., Verleyen, P., Huybrechts, J., Strauss, J., Hauser, F.,
663 Stafflinger, E., Schneider, M., Pauwels, K., Schoofs, L., Grimmelikhuijzen, C.J., 2011.
664 Genomics, transcriptomics, and peptidomics of *Daphnia pulex* neuropeptides and protein
665 hormones. J. Proteome Res. 10, 4478-4504.

666

667 Elekonich, M.M., Horodyski, F.M., 2003. Insect allatotropins belong to a family of structurally-
668 related myoactive peptides present in several invertebrate phyla. Peptides 24,1623-1632.

669

670 Elliott, K.L., Stay, B., 2008. Changes in juvenile hormone synthesis in the termite
671 *Reticulitermes flavipes* during development of soldiers and neotenic reproductives from groups
672 of isolated workers. J. Insect Physiol. 54, 492-500.

673

674 Ferrè, F., Clote, P., 2005. DiANNA: a web server for disulfide connectivity prediction. Nucleic
675 Acids Res. 33,W230-W232.

676

677 Gard, A.L., Lenz, P.H., Shaw, J.R., Christie, A.E., 2009. Identification of putative peptide
678 paracrines/hormones in the water flea *Daphnia pulex* (Crustacea; Branchiopoda; Cladocera)
679 using transcriptomics and immunohistochemistry. Gen. Comp. Endocrinol. 160, 271-287.

680

681 Hansen, K.K., Hauser, F., Williamson, M., Weber, S.B., Grimmelikhuijzen, C.J., 2011. The
682 *Drosophila* genes CG14593 and CG30106 code for G-protein-coupled receptors specifically

683 activated by the neuropeptides CCHamide-1 and CCHamide-2. *Biochem. Biophys. Res.*
 684 *Commun.* 404, 184-189.
 685
 686 Hansen, K.K., Stafflinger, E., Schneider, M., Hauser, F., Cazzamali, G., Williamson, M.,
 687 Kollmann, M., Schachtner, J., Grimmelikhuijzen, C.J., 2010. Discovery of a novel insect
 688 neuropeptide signaling system closely related to the insect adipokinetic hormone and corazonin
 689 hormonal systems. *J. Biol. Chem.* 285, 10736-10747.
 690
 691 Hauser, F., Grimmelikhuijzen, C.J., 2014. Evolution of the AKH/corazonin/ACP/GnRH
 692 receptor superfamily and their ligands in the Protostomia. *Gen. Comp. Endocrinol.* 209, 35-49.
 693
 694 Hummon, A.B., Richmond, T.A., Verleyen, P., Baggerman, G., Huybrechts, J., Ewing, M.A.,
 695 Vierstraete, E., Rodriguez-Zas, S.L., Schoofs, L., Robinson, G.E., Sweedler, J.V., 2006. From
 696 the genome to the proteome: uncovering peptides in the *Apis* brain. *Science* 314, 647-649.
 697
 698 Inward, D.J., Vogler, A.P., Eggleton, P., 2007. A comprehensive phylogenetic analysis of
 699 termites (Isoptera) illuminates key aspects of their evolutionary biology. *Mol. Phylogenet. Evol.*
 700 44, 953-967.
 701
 702 Janssen, T., Claeys, I., Simonet, G., De Loof, A., Girardie, J., Vanden Broeck, J., 2001. cDNA
 703 cloning and transcript distribution of two different neuroparsin precursors in the desert locust,
 704 *Schistocerca gregaria*. *Insect Mol. Biol.* 10, 183-189.
 705

706 Kastin, A.J., 2006. Handbook of Biologically Active Peptides. Academic Press. Burlington.
 707
 708 Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7:
 709 improvements in performance and usability. Mol. Biol. Evol. 30, 772-780.
 710
 711 Kay, I., Patel, M., Coast, G.M., Totty, N.F., Mallet, A.I., Goldsworthy, G.J., 1992. Isolation,
 712 characterization and biological activity of a CRF-related diuretic peptide from *Periplaneta*
 713 *americana* L. Regul. Pept. 42, 111-122.
 714
 715 Liebrich, W., Kellner, R., Gäde, G., 1995. Isolation and primary structures of neuropeptides of
 716 the AKH/RPCH family from various termite species. Peptides 16, 559-564.
 717
 718 Ma, M., Bors, E.K., Dickinson, E.S., Kwiatkowski, M.A., Sousa, G.L., Henry, R.P., Smith,
 719 C.M., Towle, D.W., Christie, A.E., Li, L., 2009. Characterization of the *Carcinus maenas*
 720 neuropeptidome by mass spectrometry and functional genomics. Gen. Comp. Endocrinol. 161,
 721 320-334.
 722
 723 Ma, M., Gard, A.L., Xiang, F., Wang, J., Davoodian, N., Lenz, P.H., Malecha, S.R., Christie,
 724 A.E., Li, L., 2010. Combining *in silico* transcriptome mining and biological mass spectrometry
 725 for neuropeptide discovery in the Pacific white shrimp *Litopenaeus vannamei*. Peptides 31, 27-
 726 43.
 727

728 McCallum, M.L., Rao, K.R., Riehm, J.P., Mohrherr, C.J., Morgan, W.T., 1991. Primary
 729 structure and relative potency of an analog of beta-PDH (pigment-dispersing hormone) from the
 730 crayfish *Procambarus clarkii*. *Pigment Cell Res.* 4, 201-208.
 731
 732 Misof, B., Liu, S., Meusemann, K., Peters, R.S., Donath, A., Mayer, C., Frandsen, P.B., Ware, J.,
 733 Flouri, T., Beutel, R.G., Niehuis, O., Petersen, M., Izquierdo-Carrasco, F., Wappler, T., Rust, J.,
 734 Aberer, A.J., Aspöck, U., Aspöck, H., Bartel, D., Blanke, A., Berger, S., Böhm, A., Buckley,
 735 T.R., Calcott, B., Chen, J., Friedrich, F., Fukui, M., Fujita, M., Greve, C., Grobe, P., Gu, S.,
 736 Huang, Y., Jermin, L.S., Kawahara, A.Y., Krogmann, L., Kubiak, M., Lanfear, R., Letsch, H.,
 737 Li, Y., Li, Z., Li, J., Lu, H., Machida, R., Mashimo, Y., Kapli, P., McKenna, D.D., Meng, G.,
 738 Nakagaki, Y., Navarrete-Heredia, J.L., Ott, M., Ou, Y., Pass, G., Podsiadlowski, L., Pohl, H.,
 739 von Reumont, B.M., Schütte, K., Sekiya, K., Shimizu, S., Slipinski, A., Stamatakis, A., Song,
 740 W., Su, X., Szucsich, N.U., Tan, M., Tan, X., Tang, M., Tang, J., Timelthaler, G., Tomizuka, S.,
 741 Trautwein, M., Tong, X., Uchifune, T., Walz, M.G., Wiegmann, B.M., Wilbrandt, J., Wipfler,
 742 B., Wong, T.K., Wu, Q., Wu, G., Xie, Y., Yang, S., Yang, Q., Yeates, D.K., Yoshizawa, K.,
 743 Zhang, Q., Zhang, R., Zhang, W., Zhang, Y., Zhao, J., Zhou, C., Zhou, L., Ziesmann, T., Zou, S.,
 744 Li, Y., Xu, X., Zhang, Y., Yang, H., Wang, J., Wang, J., Kjer, K.M., Zhou, X., 2014.
 745 Phylogenomics resolves the timing and pattern of insect evolution. *Science* 346, 763-767.
 746
 747 Monigatti, F., Gasteiger, E., Bairoch, A., Jung, E., 2002. The Sulfinator: predicting tyrosine
 748 sulfation sites in protein sequences. *Bioinformatics* 18, 769-770.
 749

Nässel, D.R., Winther, A.M., 2010. *Drosophila* neuropeptides in regulation of physiology and behavior. Prog. Neurobiol. 92, 42-104.

Petersen, T.N., Brunak, S., von Heijne, G., Nielsen, H., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods 8,785-786.

Plowes, N., 2010. An Introduction to Eusociality. Nature Education Knowledge 3:7.

Predel, R., Wegener, C., 2006. Biology of the CAPA peptides in insects. Cell. Mol. Life Sci. 63, 2477-2490.

Roth, S., Fromm, B., Gäde, G., Predel, R., 2009. A proteomic approach for studying insect phylogeny: CAPA peptides of ancient insect taxa (Dictyoptera, Blattoptera) as a test case. BMC Evol. Biol. 9, 50.

Shellman-Reeve, J.S., 1997. The spectrum of eusociality in termites. In: Choe, J.C., Crespi, B.J. (eds). Social competition and cooperation in insects and arachnids. Cambridge University Press, Cambridge, UK, pp 52-93.

Stafflinger, E., Hansen, K.K., Hauser, F., Schneider, M., Cazzamali, G., Williamson, M., Grimmelikhuijzen, C.J., 2008. Cloning and identification of an oxytocin/vasopressin-like receptor and its ligand from insects. Proc. Natl. Acad. Sci. USA 105, 3262-3267.

773 Taylor, P.A. 3rd, Bhatt, T.R., Horodyski, F.M., 1996. Molecular characterization and expression
 774 analysis of *Manduca sexta* allatotropin. Eur. J. Biochem. 239, 588-596.
 775
 776 Terrapon, N., Li, C., Robertson, H.M., Ji, L., Meng, X., Booth, W., Chen, Z., Childers, C.P.,
 777 Glastad, K.M., Gokhale, K., Gowin, J., Gronenberg, W., Hermansen, R.A., Hu, H., Hunt, B.G.,
 778 Huylmans, A.K., Khalil, S.M., Mitchell, R.D., Munoz-Torres, M.C., Mustard, J.A., Pan, H.,
 779 Reese, J.T., Scharf, M.E., Sun, F., Vogel, H., Xiao, J., Yang, W., Yang, Z., Yang, Z., Zhou, J.,
 780 Zhu, J., Brent, C.S., Elsik, C.G., Goodisman, M.A., Liberles, D.A., Roe, R.M., Vargo, E.L.,
 781 Vilcinskas, A., Wang, J., Bornberg-Bauer, E., Korb, J., Zhang, G., Liebig, J., 2014. Molecular
 782 traces of alternative social organization in a termite genome. Nat. Commun. 5, 3636.
 783
 784 Vázquez-Acevedo, N., Rivera, N.M., Torres-González, A.M., Rullan-Matheu, Y., Ruíz-
 785 Rodríguez, E.A., Sosa, M.A., 2009. GYRKPPFNGSIFamide (Gly-SIFamide) modulates
 786 aggression in the freshwater prawn *Macrobrachium rosenbergii*. Biol. Bull. 217, 313-326.
 787
 788 Veenstra, J.A., 2000. Mono- and dibasic proteolytic cleavage sites in insect neuroendocrine
 789 peptide precursors. Arch. Insect. Biochem. Physiol. 43, 49-63.
 790
 791 Veenstra, J.A., 2009. Allatostatin C and its paralog allatostatin double C: the arthropod
 792 somatostatins. Insect Biochem. Mol. Biol. 39, 161-170.
 793
 794 von Reumont, B.M., Jenner, R.A., Wills, M.A., Dell'ampio, E., Pass, G., Ebersberger, I., Meyer,
 795 B., Koenemann, S., Iliffe, T.M., Stamatakis, A., Niehuis, O., Meusemann, K., Misof, B., 2012.

796 Pancrustacean phylogeny in the light of new phylogenomic data: support for Remipedia as the
 797 possible sister group of Hexapoda. *Mol. Biol. Evol.* 29, 1031-1045.
 798
 799 Watson, J.A.L., Metcalf, E.C., Sewell, J.J., 1977. A re-examination of the development of castes
 800 in *Mastotermes darwiniensis* Froggatt (Isoptera). *Aust. J. Zool.* 25, 25-42.
 801
 802 Wilson, E.O., 1971. *The Insect Societies*. Cambridge, MA: Belknap Press of Harvard University
 803 Press.
 804
 805 Wilson, E.O., Hölldobler, B., 2005. Eusociality: origin and consequences. *Proc. Natl. Acad.*
 806 *Sci. USA.* 102, 13367-13371.
 807
 808 White, B.H., Ewer, J., 2014. Neural and hormonal control of postecdysial behaviors in insects.
 809 *Annu. Rev. Entomol.* 59, 363-381.

811 **Figure legends**

812 Figure 1. Three examples of the *in silico* workflow used for the prediction of putative mature
 813 *Mastotermes darwiniensis* peptide structures. (A) Predicted processing scheme for prepro-
 814 adipokinetic hormone (AKH). The structure of the putative mature AKH isoform is shown in
 815 red, with that of a putative mature linker/precursor-related peptide shown in blue. In this panel,
 816 “pQ” indicates the pyroglutamic acid residue present in the putative mature AKH; an inverted
 817 blue bracket indicates a disulfide bridge between two cysteine residues in the putative mature
 818 linker/precursor-related peptide. (B) Predicted processing scheme for prepro-allatostatin C

(AST-C). The structure of the putative mature AST-C isoform is shown in red, with that of a putative mature linker/precursor-related peptide shown in blue. In this panel, the inverted red bracket indicates a predicted disulfide bond between the two cysteine residues in the putative mature AST-C. (C) Predicted processing scheme for prepro-sulfakinin. The structures of two putative mature sulfakinin isoforms are shown in red, with that of a putative mature linker/precursor-related peptide shown in blue. In this panel, “Y_(SO₃H)” indicates the presence of sulfated tyrosine residues in the two putative mature sulfakinins, while “pQ” or “pE” indicates the presence of pyroglutamic acid in the sulfakinins.

Figure 2. Putative *Mastotermes darwiniensis* pre/preprohormones deduced from transcriptome shotgun assembly sequence data. (A) The amino (N)-terminal portion of prepro-adipokinetic hormone-corazonin-like peptide. (B) Prepro-allatostatin A. (C) Prepro-allatotropin (D) Pre-bursicon β. (E) Prepro-CCHamide. (F) Prepro-corazonin. (G) Prepro-crustacean cardioactive peptide. (H) Prepro-crustacean hyperglycemic hormone/ion transport peptide. (I1) Prepro-diuretic hormone 31 I. (I2) The N-terminal portion of prepro-diuretic hormone 31 II. (J) Prepro-diuretic hormone 44. (K) The N-terminal portion of prepro-FMRFamide-like peptide. (L1) Prepro-insulin-like peptide I. (L2) Prepro-insulin-like peptide II. (L3) Prepro-insulin-like peptide III. (M1) An internal protein fragment of Prepro-leucokinin I. (M2) The carboxyl (C)-terminal portion of prepro-leucokinin II. (N) Prepro-myosuppressin. (O) Pre-neuroparsin. (P1) Prepro-neuropeptide F I. (P2) Prepro-neuropeptide F II. (Q) Prepro-orcokinin. (R) The C-terminal portion of prepro-pigment dispersing hormone. (S) Prepro-pyrokinin. (T) The C-terminal portion of prepro-RYamide. (U) Prepro-short neuropeptide F. (V1) Prepro-SIFamide I. (V2) Prepro-SIFamide II. (W) Prepro-tachykinin-related peptide. In this figure, signal peptides

842 are shown in gray, while all mono/dibasic cleavage loci are shown in black. For each sequence,
843 the isoform(s) of the peptide for which the precursor is named is/are shown in red, with all
844 linker/precursor related peptides shown in blue. The “+” symbol indicate the presence of
845 additional, unknown, amino acid residues at the N- and/or C-terminus.

Table 1

Table 1. Putative <i>Mastotermes darwiniensis</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	GAZE01042474	1015	70	F	49.3	2e-07
Adipokinetic hormone-corazonin-like peptide	GAZE01369298	253	68	N	47.0	2e-07
Allatostatin A	GAZE01597045	3296	385	F	56.2	8e-09
Allatostatin B						
Allatostatin C	GAZE01596866	2795	96	F	95.9	9e-23
Allatotropin	GAZE01594779	1526	122	F	67.0	4e-12
Bursicon α						
Bursicon β	GAZE01044152	1338	143	F	188	6e-57
CCHamide	GAZE01587523	868	132	F	30.8	2.3
Corazonin	GAZE01569676	558	136	F	37.4	0.009
Crustacean cardioactive peptide	GAZE01589689	968	167	F	64.7	5e-12
Crustacean hyperglycemic hormone/ion transport peptide	GAZE01566786	534	115	F	92.4	4e-23
DENamide						
Diuretic hormone 31	GAZE01584950	784	115	F	87.4	9e-21
	GAZE01414042	257	82	N	46.6	5e-07
Diuretic hormone 44	GAZE01044942	1689	198	F	70.1	3e-12
DXXRLamide						
Ecdysis-triggering hormone						
Eclosion hormone						
FMRFamide-like peptide	GAZE01569095	552	170	N	60.5	5e-10
FXGGXamide						
GSEFLamide						
Insulin-like peptide	GAZE01012208	1996	147	F	39.3	0.002
	GAZE01012206	2378	147	F	39.3	0.002
	GAZE01016992	1262	132	F	35.8	0.030
	GAZE01572726	585	135	F	30.8	1.2
Intocin						
Leucokinin	GAZE01541820	424	141	I	41.2	5e-04
	GAZE01590193	996	291	C	32.3	1.1
Myosuppressin	GAZE01595045	1597	98	F	65.5	1e-12
Neuroparsin	GAZE01588475	907	106	F	78.6	9e-18
Neuropeptide F	GAZE01583053	737	91	F	42.7	4e-05
	GAZE01593796	1331	124	F	31.6	0.39
Orcokinin	GAZE01031957	2105	164	F	55.1	4e-08
	GAZE01031955	8602	164	F	55.1	5e-08

Pigment dispersing hormone	GAZE01394118	255	41	C	40.0	9e-05
Proctolin						
Pyrokinin	GAZE01583360	744	185	F	29.3	5.0
RYamide	GAZE01568334	546	115	C	37.0	0.008
Short neuropeptide F	GAZE01595464	1725	101	F	39.3	0.011
SIFamide	GAZE01594511	1465	72	F	68.2	7e-14
	GAZE01510888	354	80	F	37.0	8e-04
Sulfakinin	GAZE01588629	914	124	F	55.5	5e-09
Tachykinin-related peptide	GAZE01594670	1503	337	F	59.3	4e-09

*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 (AKH), *Drosophila melanogaster* prepro-AKH (**Accession No. [AAF47846](#)**; Adams et al., 2000); adipokinetic hormone-corazonin-like peptide (ACP), *Tribolium castaneum* prepro-ACP (**Accession No. [ADF28807](#)**; Hansen et al., 2010); allatostatin A (AST-A), *D. melanogaster* prepro-AST-A (**Accession No. [AAF56331](#)**; Adams et al., 2000); allatostatin B (AST-B), *D. melanogaster* prepro-AST-B (**Accession No. [AAF49354](#)**; Adams et al., 2000); allatostatin C (AST-C), *Homarus americanus* prepro-AST-C (Dickinson et al., 2009a); allatotropin, *Manduca sexta* prepro-allatotropin (**Accession No. [AAB08759](#)**; Taylor et al., 1996); bursicon α , *D. melanogaster* pre-bursicon (**Accession No. [AAF55915](#)**; Adams et al., 2000); bursicon β , *D. melanogaster* pre-partner of bursicon (**Accession No. [AAF53396](#)**; Adams et al., 2000); CCHamide, *D. melanogaster* prepro-CCHamide 1 (**Accession No. [AAF55014](#)**; Adams et al., 2000); corazonin, *D. melanogaster* prepro-corazonin (**Accession No. [AAF55046](#)**; Adams et al., 2000); crustacean cardioactive peptide (CCAP), *D. melanogaster* prepro-CCAP (**Accession No. [AAF56045](#)**; Adams et al., 2000); crustacean hyperglycemic hormone (CHH)/ion transport peptide (ITP), *D. melanogaster* prepro-ITP (**Accession No. [ABI31114](#)**; Adams et al., 2000); DENamide, *Daphnia pulex* prepro-DENamide (Dirksen et al., 2011); diuretic hormone 31 (DH31), *D. melanogaster* prepro-DH31 (**Accession No. [AAF52685](#)**; Adams et al., 2000); diuretic hormone 44 (DH44), *D. melanogaster* prepro-DH44 (**Accession No. [AAF54421](#)**; Adams et al., 2000); DXXRLamide, *Tigriopus californicus* prepro-DXXRLamide Ia (Christie, 2014b); ecdysis-triggering hormone (ETH), *D. melanogaster* prepro-ETH (**Accession No. [AAF47275](#)**; Adams et al., 2000); eclosion hormone (EH), *D. melanogaster* pre-EH (**Accession No. [AAF55423](#)**; Adams et al., 2000); FMRamide-like peptide (FLP); *D. melanogaster* prepro-FLP (**Accession No. [AAF58874](#)**; Adams et al., 2000); FXGGXamide, *T. californicus* prepro-FXGGXamide Ia (Christie, 2014b); GSEFLamide, *T. californicus* prepro-GSEFLamide Ia (Christie, 2014b); insulin-like peptide (ILP), *D. melanogaster* prepro-ILP 2 (**Accession No. [AAF50204](#)**; Adams et al., 2000); intocin, *T. castaneum* prepro-vasopressin-like peptide (**Accession No. [ABX52000](#)**; Stafflinger et al., 2008); leucokinin, *Eucyclops serrulatus* prepro-leucokinin (Christie, 2015b); myosuppressin, *D. melanogaster* prepro-myosuppressin (**Accession No. [AAF56283](#)**; Adams et al., 2000); neuroparsin, *Schistocerca gregaria* pre-neuroparsin 1 (**Accession No. [CAC38869](#)**; Janssen et al., 2001); neuropeptide F (NPF), *D. melanogaster* prepro-NPF (**Accession No. [AAF55339](#)**; Adams et al., 2000); orckinin, *H. americanus* prepro-orckinin I (**Accession No. [ACB41787](#)**; Dickinson et al., 2009b); pigment dispersing hormone, *D. melanogaster* prepro-pigment dispersing factor (**Accession No. [AAF56593](#)**; Adams et al., 2000); proctolin, *D. melanogaster* prepro-proctolin (**Accession No. [AAF52572](#)**; Adams et al., 2000); pyrokinin, *D. melanogaster* prepro-capability (**Accession No. [AAF56969](#)**; Adams et al., 2000); RYamide, *Procambarus clarkii* prepro-RYamide (Christie and Chi, 2015b); short neuropeptide F (sNPF), *D. melanogaster* prepro-sNPF (**Accession No. [AAN11060](#)**; Adams et al., 2000); SIFamide, *D. melanogaster* prepro-SIFamide (**Accession No. [AFH08249](#)**; Adams et al., 2000); sulfakinin, *D. melanogaster* prepro-drosulfakinin (**Accession No. [AAF52173](#)**; Adams et al., 2000); tachykinin-related peptide (TRP), *D. melanogaster* prepro-TRP (**Accession No. [AAF54735](#)**; Adams et al., 2000).

Table 2

Table 2. Predicted peptidome of <i>Mastotermes darwiniensis</i>	
Peptide group	Predicted peptide structure
Adipokinetic hormone	pQVNFSPNwa
Adipokinetic hormone linker/precursor-related peptide	SGLQDGPKTSTD ^{SLMYIYKLIQSEAQKLVDCEKFGAN}
Adipokinetic hormone-corazonin-like peptide	pQVTFSRDWNaa
Adipokinetic hormone-corazonin-like peptide linker/precursor-related peptide	SPADLQCN ^{AVRSADEICKAFLDEF} RHLAA+
Allatostatin A	LPVYKFGLa SKEYGFGLa AGADGGRFYSFGLa VRPYSFGLa SPSSGIQRLYGFGVa GRSLYSFGLa YSFGMa SLHYAFGLa PFKFGLa
Allatostatin A linker/precursor-related peptide	TESATTGTHVAPEEPSPASGGNVGLVPHHIESSAADNSELDFV LYDSGFa AYSVSEY EDDDDY _(8/19H) VQEEEDENQINGNDELEDSDVDTMD ERLYSFDLE SDGRLYPFGLE PASSGRQSGSRFNFGLL SYDTNLEEFEDEME ^{EEEA} SPQa EVAPSELQAVRNEERVREMDNKEESRNNGTAEGYH ^{HS} AERV AYDLYSSLDGEDDEVGDEEFTRLV IPIYDFGTEKLSE
Allatostatin C	SYWKQCAFNAVSCFa
Allatostatin C linker/precursor-related peptide	KNLGQPDDKDR ^{LISELDLVDDDGSIETALINYLFAKQVVSRLRSQMDVSDLO}
Allatotropin	GFKNVALSTARGFa
Allatotropin linker/precursor-related peptide	GSSYQ ^{TRNKPTI} DGALTYLAGNANTAAESTPESLPVEWFVEELRTNP ^{ELA} IIIVHKFVDADQDGELSAEELLRP ^{MY}
Bursicon b	pQPEDDMT ^{CETIPSEIHI} IKEEFDDLGR ^{LQRTC} SSEVS ^{VNKCE} GACASQVQPSVITPTGFIKE ^{CYC} CRESFLRERII ^{SLTHCYDPDGVRLTQEGQASLEV} KIREPVD ^{CKCFKCGDFS}
CCHamide	GCSAFGHSCFGGHa
CCHamide linker/precursor-related peptide	VDEDVLLLP ^{GS} SDSQQPKLVFPPPAARPVGEDDVTRQAGEYGVRS ^{PSSLP} SDGIVSLITSSHP ^{SHYNLS} PFRLQWLQSY STGDI
Corazonin	pQTFQYSRGWTNa
Corazonin linker/precursor-related peptide	GGPAPPMFIPSSASGAVLQNLDESSVIN ^{PCSQLQRI} RFLLGARNPQ ^{QFYYP} CD ^{SWRDI} SETPSEEV ^{SERF} APQDSAEGNNT ^{EEN}

Crustacean cardioactive peptide	PFCNAFTGCa
Crustacean cardioactive peptide linker/precursor-related peptide	VVIQ VDSGEVGRLLDP SDDSMGTLVELNSEPAVEDLSRQILSEAKLWEAIQEAKEVELL QEQQQLQQQSQYATAMERPVPPLPIAGY AATAEGQAPAAHSATSEPQEQTWTWS
Crustacean hyperglycemic hormone/ion transport peptide	SFFEIQCKGVYDKSIFARLDRICEDCYNLFREPQLHSLCRSKCFSSDYFKGCLEALLLKQEEQKFNQMEYVa
Crustacean hyperglycemic hormone/ion transport peptide linker/precursor-related peptide	SPTSGILLGHSLN
Diuretic hormone 31	GLDLGLSRGFSGSQAAKHLMGLAAANYAGGPa
Diuretic hormone 31 linker/precursor-related peptide	APLSSHRNNFISDQDSEPDSEY _(SN3H) VLEMLARLGQSII ANDLENS SDESS FNVDAAPQSYWSQFDDDEDPEALMEIIARLGHSIMRNPELENT
Diuretic hormone 44	TGAGPSLSIVNPLDVLQRLLLEIARRRMRSQDQIQANREILQSIa
Diuretic hormone 44 linker/precursor-related peptide	YYEPPLLQALAAPSADHETTSYLLPRLSAKFRPQGDWESAPDPRFYVLTETERESNQAA DVNQSLQRNNEQDDDDDEIDSEFMISSTDEKPGNADPENNRPPETPDWSAASSSRWNNEYTSQHHS
FMRFamide-like peptide	CSSRVFVRLa GSSAGPASELESSGEKDTGLLRLa GRTSDDNFIRFa GGRDNFIRLa DRSDNFIRFa GRTDNFIRFa GKSDNFIRFa
FMRFamide-like peptide linker/precursor-related peptide	YPTDIPITEPPNIVLTAPDDVENSIDNIATLDDCETEKVMSFPVST NEDEDEPPV
Insulin-like peptide	DATHRYCGSNLVTTLRFLCRSQYNGIDDE NAEPIVKKSLQVSDPDNSLWTQLQTVDELPETEFPFRSRSIANTFANRVF ERSDGGIVDECC _(SN3H) IYKGCTVNELTEY _(SN3H) CSS NTNKYCGRNLANMLRYVCNGNYYPMF ASQDVEDTNDSGIWILSPPTEEPQLQFPFRSRSNAASLVPGSF HTRGVYDECC _(SN3H) CRKSC _(SN3H) TIQEMASYCa pEAAFRYCGPNLANILRLLCNGSYHTDEHW pQTRDVGIVECC _(SN3H) MFKGCT _(SN3H) LNELTEY _(SN3H) CAD
Insulin-like peptide linker/precursor-related peptide	LPENPSAVLQLV QSDLFQLGD ESHLSL NVATIQ QAPDMDGSLWIQPAAAEEAFPPFRPWWMANQLTNRVF
Leucokinin	NYAFSSWGa pEPAFSSWGa APSFSSWGa

Leucokinin linker/precursor-related peptide	APNFHSWGa PVFSSWGa +ILGNKMDSWGa DPAFYSWGa GAQFSSWGa DFSNWGa SYTSVSERPREGKALFSPWGa WGSSPAGAVFSSWGa GADFYSWGa
	+QPE SWEGSMDRLSSEGRPGFTSVGS DIFASGISGSKTEPAFWILNPKPDTETNTLGAISDPAFRIMGNHDYAPFGILKSY+ DAEIPNWEHKYEST DFHGDENE NYICDITKEDAEETa EAHEENSARNGTTNCKEEQVEERKEAY LLEELDSKTKLGSGNSKEFMGNSEETEACNTDTCQPDSDNELDEGDVNTGVPGAIL SHESPSLFSILGSMHTWHDPELLSDLF GSAHETPGF SDNLTGRNKHKTPPQGTGRQL
Myosuppressin	pQVDVHVFLRFa
Myosuppressin linker/precursor-related peptide	VPPPQCNPNFQQEDIPLRVRKVCAALSTFYELNAMEAY ^(503H) LDDKVARDNNNPLVDITGV
Neuroparsin	SHICKPCVGSECNLEPEFCDHGIERDYCGRKVCAKGPGEFCGGPSDVRGKCGEGMHCACGKCNGCSLSTLDCYFGLDQLQCL
Neuropeptide F	KPTDPEQLAAMADTLKYLQELDRYYSQVARPRFa
Neuropeptide F linker/precursor-related peptide	PKVFTSPDQLRVYLQELGNYYAIEGRPRFa AEMRPIPEQDAAPDDSNDRLW FAS DPLPAGAEIASRPT LATPTYRSSNNVVGLPLSAGLSSAVTPSNYLRYPAQSAIRNDLYQLLFYDE
Orcokinin	NFDEIDRSGFDGFV
Orcokinin linker/precursor-related peptide	NFDEIDRADFGSFV QSDTLRQTQTAF DYGGDSAYEDRGA HFDPIGGSHVLRVDGLTHFP TRSGLDSLSGVTFGWN LDSLSGITFGNQ NAPLLLA YDKEN
Pigment dispersing hormone	NSEIINSILGLPKNLNEAa
Pigment dispersing hormone linker/precursor-related peptide	+LPLVKGLSEFFQRNAEVH
Pyrokinin	SDPEISGMWFGPRLa SADEMQDDLEVVEVLRETPWALVPVKGKKHHSGGFI PRLa DSNQDEDPDIVDLRSPPFAPRLa LVPFRPRMa DHQPHDVYSPRLa

Pyrokinin linker/precursor-related peptide	CNVSGLRVSTDPLEDAILLGLDGTGDDTLMV SVNPSLPHE QPPLH
RYamide	GDLSQSRSMFWSGSRYa FFIGSRYa
RYamide linker/precursor-related peptide	NSGTSQVKGHQRGNSPVEVAVRND SEEP TTTTDETVGFGDMLLAEDDNSQVTC _{LYTGV} TNLYRCYY DNSSSEDASIEH
Short neuropeptide F	ANRSPSLRLRFa
Short neuropeptide F linker/precursor-related peptide	APSY _(SO₃H) SDY _(SO₃H) ESVRDLYQLLLQKEALENRLQQGGQHEIV ADPLLTGSPFSEHSSVESSIAEN
SIFamide	TYRKPPFNGSIFa LPFNGSIYa
SIFamide linker/precursor-related peptide	GNTADY _(SO₃H) DGTGKALSAMCEIASEACSAWFPQSDN NPNPGPQF AASALPMDYDSSSKSLSSLC _{ETVAEVCAASFPQQTEN}
Sulfakinin	pQSEDY _(SO₃H) GHMRFa pEQFDDY _(SO₃H) GHMRFa
Sulfakinin linker/precursor-related peptide	APPTSDVVSAGGSNLEGAGHRGRGRSFLQTPPRTSQYVRARLVPLEPAADVLSDFVIDDEIMDFN SLD
Tachykinin-related peptide	APSGFLGVRa GPSMGFQGVra GPSMGFHGMra APTMGFMGMra APNVGFHGMra VPAAGFFGMra VPASGFFGMra GPSVGFFAMra APSPGFLGMra APSGFLGMra
Tachykinin-related peptide linker/precursor-related peptide	APESP DSGLASQDTYNDVMD DDQELAYN DRDSQQDILD DDAMDFDY _(SO₃H) FDAGAASLGFORMREKEQWEENSCTL DGDLEGDDY _(SO₃H) FELMADGDIWGNQEEGLQDAEELS DSERIEDLDTLLSYLEAAYQHGRN NGERAPGS DWPGEQV _{ASSQGT} EPENHAPQLPDSE
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.	

A. Prepro-adipokinetic hormone (from GAZE01042474)

MVCMATKLFVVVALVLVLC**EAQ**VNFSPNWGKRSGLQDGPCKTSTD**SLMYIYKLIQSEAQ**KLVDCEKFGAN

↓ Signal peptidase (cleavage locus highlighted above)

QVNFSPNWGK**RS**GLQDGPCKTSTD**SLMYIYKLIQSEAQ**KLVDCEKFGAN

↓ Prohormone convertase (cleavage locus highlighted above)

QVNFSPNWG**KR** SGLQDGPCKTSTD**SLMYIYKLIQSEAQ**KLVDCEKFGAN

↓ Carboxypeptidase (cleavage locus highlighted above)

QVNFSPNW**G** SGLQDGPCKTSTD**SLMYIYKLIQSEAQ**KLVDCEKFGAN

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

QVNFSPNWamide SGLQDGPCKTSTD**SLMYIYKLIQSEAQ**KLVDCEKFGAN

↓ Glutaminyl cyclase (cyclization locus highlighted above)

pQVNFSPNWamide SGLQDGP**C**KTSTD**SLMYIYKLIQSEAQ**KLVD**C**EKFGAN

↓ Enzymatic disulfide bond formation (bridged cysteines highlighted above)

SGLQDGPCKTSTD**SLMYIYKLIQSEAQ****KLVDCEKFGAN**

B. Prepro-allatostatin C (from GAZE01596866)

MSTITATKLMLLMLISMLT**LSWALA**KNLGQPD**DKDRLISE**LDLVDDDDGSIETALIN**YLF**AKQVVSRLRSQMDVSD**LQ**RKRSYWKQCAFNAVSCFGK

↓ Signal peptidase (cleavage locus highlighted above)

KNLGQPD**DKDRLISE**LDLVDDDDGSIETALIN**YLF**AKQVVSRLRSQMDVSD**LQ**R**RS**YWKQCAFNAVSCFGK

↓ Prohormone convertase (cleavage locus highlighted above)

KNLGQPD**DKDRLISE**LDLVDDDDGSIETALIN**YLF**AKQVVSRLRSQMDVSD**LQ**R**KR** SYWKQCAFNAVSCFG**K**

↓ Carboxypeptidase (cleavage loci highlighted above)

KNLGQPD**DKDRLISE**LDLVDDDDGSIETALIN**YLF**AKQVVSRLRSQMDVSD**LQ** SYWKQCAFNAVSCF**G**

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

SYWKQCAFNAV**C**Famide

↓ Enzymatic disulfide bond formation (bridged cysteines highlighted above)

SYWKQCAFNAVSC**Famide**

C. Prepro-sulfakinin (from GAZE01588629)

MTSRVVATLLLT**LG**VYLI**LQ**HHYAAV**QA**APPTSDVVSAGGSNLEGAGHRGRGRSFLQTPPRTSQYV**RLV**PLEPAADVLSDFVIDDEIMDFNKRQS
EDYGHMRFGKRE**Q**FDDYGHMRFG**RS**LD

↓ Signal peptidase (cleavage locus highlighted above)

APPTSDVVSAGGSNLEGAGHRGRGRSFLQTPPRTSQYV**RLV**PLEPAADVLSDFVIDDEIMDFN**KR**Q**SE**DYGHMRFGK**RE**QFDDYGHMRFG**RS**LD

↓ Prohormone convertase (cleavage loci highlighted above)

APPTSDVVSAGGSNLEGAGHRGRGRSFLQTPPRTSQYV**RLV**PLEPAADVLSDFVIDDEIMDFN**KR** Q**SE**DYGHMRFG**KR** EQFDDYGHMRFG**R**
SLD

↓ Carboxypeptidase (cleavage loci highlighted above)

APPTSDVVSAGGSNLEGAGHRGRGRSFLQTPPRTSQYV**RLV**PLEPAADVLSDFVIDDEIMDFN Q**SE**DYGHMRF**G** EQFDDYGHMRF**G**

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

Q**SE**DY**G**HMRFamide EQFDDY**G**HMRFamide

↓ Tyrosylprotein sulfotransferase (tyrosine targets highlighted above)

QSEDY_(S03H1)G**HMR**Famide **E**QFDDY_(S03H1)G**HMR**Famide

↓ Glutaminyl cyclase (cyclization loci highlighted above)

pQSEDY_(S03H1)G**HMR**Famide **pEQ**FDDY_(S03H1)G**HMR**Famide

A. Prepro-adipokinetic hormone-corazonin-like peptide (N-terminal partial protein from GAZE01369298)

MTHRALCQRALLFVGVLSCLYSRTSGQVTFSRDWNAGKRSPADLQCNVRSAD EICKAFLDEFRLHAA+

B. Prepro-allatostatin A (from GAZE01597045)

MLGLKPCFGSFHTAILCVLLMHLSTPTLGTESATTGTHVAPEEPSPASGGNVGLVPHHIESSAADNSELDFVKRLYDSGFGKRAYSYVSEYKRLPV
YKFGGLKRSKEYGFGGLKAGADGGRFYSGFLKREDDDDYVQEEEDENQINGNDELESDVDTMDKRELYSFDLEKVRPYSGFLKRSPPSSGIQ
RLYGFGVGKRGSRYSFGFLKGRSDGRLYPFGLKRPASSGRQSGSRNFGLLKKSYDNL EEFEDEMEEEAKRSPQGRYSGFMGKREVPASELOAV
RNEERVREMDNKEESRNNGTAEGYHSAERVKRSLHYAFGLGKRAYDLYSSLDGEDDEVGDEEFTRLVRRPFKFGGLKRIPIYDFGTEKLSEK

C. Prepro-allatotropin (from GAZE01594779)

MRASLSVYSIIAVTALVLLALCGCASSGSSYQTTNRNKPRTIRGFKNVALSTARGFGKRDGALTYLAGNANTAAESTPESLPVEWFVEELRTNPELAR
IIVHKFVDADQDGELSAEELLRPMY

D. Prepro- Pre-bursicon β (from GAZE01044152)

MVSEAAWRLRNLFVFIILFFMVSSSLAQPEDDMTCETIPSEIHIIKEEFDDLGRQLQRTCSSESVNKEGACASQVQPSVITPTGFIKECYCCRESFL
RERIISLTHCYDPDGVRILTQEGQASLEVKIREPVDCKCFKCGDFSR

E. Prepro-CCHamide (from GAZE01587523)

MTLYRRASVLTVMMAVLLMAEIDQSFAKRGCSAFGHSCFGGHGKRVDEVDVLLPGSDSDQQPKLVFPFPAARFVGEDDVTRQAGEYGVRSPPSLPS
DGIVSLITSSHPSHYNLSFFLRQWLQSYRRSTGDI

F. Prepro-corazonin (from GAZE01569676)

MHFHSTATASGCRSPRLPALLLILCCLTGVLAAQTFFQYSGRWTNGRKRKGGPAPPMFIPSSASGAVLQNLDESSVISNPCSQQLQIRIFLLGARNPQQF
YYPCDSWRDISETPSEEVSERFRRKAPQDSAEGNNTEN

G. Prepro-crustacean cardioactive peptide (from GAZE01589689)

MMCHVVIIGCTLAALLVILGLPRLSCDAVVIQKRVDSGEVGRLLDPKRRKPFCAFTGCGKKRSDDSMGTLVELNSEPAVEDLSRQILSEAKLWEAI
QEAKVELLRRRQEQQLQQQSQYATAMERPVPPLPIAGYRRKRAATAEGQAPAAHSATSEPQEQFTKTWSR

H. Prepro-crustacean hyperglycemic hormone/ion transport peptide (from GAZE01566786)

MQQPQLTRVLACVLLVSMVITSLAASPTSGLLGHSLNKRSSF EIQCKGVYDKSIFARLDRI CEDCYNLFREPQLHSLCRSKCFSSDYFKGCLEALL
LKQEEQKFNQMV EYVGKK

I1. Prepro-diuretic hormone 31 I (from GAZE01584950)

MNSCLILLTSTLLVGVIIMLSVAHAVESAPLSSHNNFTISDQSEPDSEYVLEMLARLGQSIIRANDLENSKRGDLGLSRGFSGSQAAKHLMGLAA
ANYAGGPGRRRRSSDESS

I2. Prepro-diuretic hormone 31 II (N-terminal partial protein from GAZE01414042)

MQQRVVLSCVLLAFVVFVAVSSFNVDAAPOSYWSQFDDDEDPEALMEI IARLGHSIMRNPELENTKRGDLGLSRGFSGSQAA+

J. Prepro-diuretic hormone 44 (from GAZE01044942)

MMLAVIPTLLLVALMCSMAYYEPPLLQALAAPADHETTSYLLPRLSAKFRPQGDWESAPDPRFYVLTETERESNQAAARRVKRTGAGPSLSIVNPL
DVLQRLLLEIARRRMRSQSDQIQANREILQSIGKRDVNQSLQRNNEQDDDEIDSEFMISSTDEKPGNADPENNRPPETPDWSAASSSRWNNEYTS
QHHS

K. Prepro-FMRamide-like peptide (N-terminal partial protein from GAZE01569095)

MKRVLVLVCAVAASYP TDIPTTEPPNIVLTAPDDVENSIDNIATLDDCETEKVMSVSTKRNED EDEPPVRRRCSSRVFVRLGRSSAGPASELE
SSGEKDTGLLRGCRGRTSDDNFIRFGRGGRNDNFIRLGRDRSDNFIRFGRGRTDNFIRFGRGKSDNFIRFGRG+

L1. Prepro-insulin-like peptide I (from GAZE01012208 and GAZE01012206)

MRLCLHVMLIGVICVCA LPENPSAVLQLVRKRDATHRYCGSNLVTTLRFLCRSQYNGIDDERKRNAEPIVKKSLQVSDPDNSLWTQLQTVDEL PET
EFPFRSRSIANTFANRVFRRRERSDGGIVDECCIIYKGC TVNELTEYCSSR

L2. Prepro-insulin-like peptide II (from GAZE01016992)

MWRLCLRLVAIAAVCLCTLAQAQSDLFQLGDKRNTNKYCGRNLANMLRYVCNGNYIPMFKKASQDVEDTNDSGIWILSPPTTEPQLQFPFRSRSNAA
SLVPGSFRRHTRGVYDECCRKSTIQEMASYCGR

L3. Prepro-insulin-like peptide III (from GAZE01572726)

MRNILMRVTVIGMICSWALPESHLSLRKREAAFRYCGPNLANILRLLCNGSYHTDEHWKRN VATIQKKQAPMDGSLWIQPA AEEAFPRPWWMAN
QLTNRVFRRQTRDVGIVECCMFKGCTLNELTEYCADR

M1. Prepro-leucokinin I (internal protein fragment from GAZE01541820)

+QPEKKNYAFSSWGGKREPAFSSWGGKRAPSFSSWGGKRAPNFHSSWGGKRRSWEGSGMDRLSSEGRPGFTSVGSKRRPVFSSWGGKRDIFASGISGSK
TEPAFWILNPKPD TETNTLGAISDPAFRIMGNHDYAPFGILGKSY+

M2. Prepro-leucokinin II (C-terminal partial protein from GAZE01590193)

+ILGNKMDSWGGKRDPAFYSWGGRDAEIPNWEHKYESTKRG AQFSSWGGKRDHGDENEKRD FSNWGGKRN YICDITKDAEETGKREAHEENSAR
NGTTNCKEEQVEERKEAYRRLLEELDSKTKLGSGNSKEFMGNSEETEAE CNTDTCQPDSDNELDEGDVNTGVP GAILKRSYTSV SERPREGKALFSP
WGGKRSHESPSLFSILGSMHTWHD PPLLSDLFKKRGSAHETPGFKK WGSSPAGAVFSSWGGKRS DNLTGRNKHKTPPQGTGRQLRRGADFYSWGGR
R

N. Prepro-myosuppressin (from GAZE01595045)

MKHACAIVICVLAVLLACSPREVSASVPPPPQCNPNFQQEDIPLRVKVCALSTFYELSNAYLDDKVARDDNNPLVDITGVKRQDVHDVFLRFGRKR

O. Pre-neuroparsin (from GAZE01588475)

MKFCAVFTLVSAVLLLLLESCEVAGSHICKPCVGSECNLEPEFCDHGIERYDCGRKVCAGKPGGEFCGGPSDVRGKCGEGMHCACGKCNCGCSLSTLDCYFGLDQLQCL

P1. Prepro-neuropeptide F I (from GAZE01583053)

MQSSLCWILLVGCALILVFNLVFNVWAKPTDPEQLAAMADTLKYLQELDRYYSQVARPRFGKRAEMRPIPEQDAAPDSDNDRLWRRFASRR

P2. Prepro-neuropeptide F II (from GAZE01593796)

MQNLANLVLLACLCCVVISMAMPWCSDPLPAGAEIASRPTRPKVFTSPDQLRVYLQELGNYAIEGRPRFGKRLATPTYRSSNNVVGLPLSAGLSSAVTPSNYLRYPAQSAIRNDLYQLLFYDE

Q. Prepro-orcokinin (from GAZE01031957 and GAZE01031955)

MMTAPVAVLRSGMKVLPFLVVMVAVLSVPSRATSAQSDTLRQTQTAFRDYGGDSAYEDRGARHFDPIGGSHVLRVDGLTHFPRTTRSGLDLSLGSVTFGWNKRRLDSLSGITFGNQKRNFDEIDRSGFDGFFVKRNFDEIDRADFGSFVKRNAPLLLARRYDKENR

R. Prepro-pigment dispersing hormone (C-terminal partial protein from GAZE01394118)

+LPLVKGLSEFFQRNAEVHKRNSEIINSILGLPKNLNEAGRR

S. Prepro-pyrokinin (from GAZE01583360)

MRTGFSTTQRLIQSTVLLFLATVLTSCNVSGLRVSTDPLEDAILLGLDGTGDDTLMVKKSDPEISGMWFGPRIGRRDRRSADEMQDDLEVVEVLRETPTWALVPVKGKKHSGGGFIPRLGRDSNQDEDDPIDVLDLRSPPFAPRLGRRLVPPFRPRMGRDHQPHDVYS PRLGRSVNPSLPHEKKQPPLH

T. Prepro-RYamide (C-terminal partial protein from GAZE01568334)

+YGKRGDLSQSRSMFWSGSRYGRNSGTSQVKQHQRGNSPVEVAVRNDRFFIGSRYGKRSEEPTTTTDETVGFGDMLLAEEDDNSQVTCLYTGVTNLRYCYKRRKDNSSSEDASIEHKK

U. Prepro-short neuropeptide F (from GAZE01595464)

MHCFTVKCCTVALCFLIVLAEFVTSAPSYSDYESVRDLYQLLLQKEALENRLQQQGQHEIVRKANRSPSLRLRFGRRADPLLTGSPFSEHSSVESSTIAEN

V1. Prepro-SIFamide I (from GAZE01594511)

MQKRGIATCLLLLAVLLLAEFATA TYRKPPFNNGSIFGKRGNATADYDGTGKALSAMCEIASEACSAWFPQSDN

V2. Prepro-SIFamide II (from GAZE01510888)

MKFPCS SVFVLVTLILVLLVDCNPNPGPQFRRLPFNNGSIYGKRAASALPMDYDSSSKSLSSLCETVAEVCAASFQQTEN

W. Prepro-tachykinin-related peptide (from GAZE01594670)

MVLPCRCRCRAGALLVLAISLIAVVLCAPEESFKRAPSGFLGVRGKKDSGLASQDTYNDVMDKRGPSMGFQGVRGKKDDQELAYNKRGPSMGFHHMRGKKDRDSQQDILDKRAPTGMGFMGRGKKDDAMDFDYFDAGAASLGFRMREKEQWEENSGTLKRAPNVGFHGMGRGKKDGDLEGDDYFELMADGDIWGNQEEGLQDAEELSKRVPAAGFFGMRGKKVPASGFFGMRGKKGPSVGFFAMRGKKAPSPGFLGMRGKKDSEIEDLDITLLSYLEAAYQHGRNKRNGERAPGSKKAPSGFLGMRGKKDWPGEQVVASSQGTPEHNHAPQLPDSE