Neuropeptide discovery in *Symphylella vulgaris* (Myriapoda, Symphyla): *in silico* prediction of the first myriapod peptidome

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; pseudocentipede

Abstract

Arthropods have contributed greatly to our understanding of peptidergic control of physiology and behavior, and being the largest and most diverse animal phylum, represent a model for investigating peptide hormone evolution. Surprisingly, one arthropod subphylum, the Myriapoda, is uninvestigated in terms of its peptide hormones. The public deposition of a transcriptome for *Symphylella vulgaris*, a pseudoscorpion, provides a means for peptide discovery in myriapods. Here, *in silico* transcriptome mining was used to identify 47 *S. vulgaris* neuropeptide-encoding transcripts within this dataset. The identified transcripts allowed for the deduction of 31 unique pre/preprohormone sequences, with 97 distinct mature peptides predicted from the deduced proteins. The predicted *S. vulgaris* peptidome includes members of the adipokinetic hormone/red pigment concentrating hormone, adipokinetic hormone-corazonin-like peptide, allatostatin A, allatostatin C (AST-C), allatotropin, CCHamide, crustacean cardioactive peptide, GSEFLamide, insulin-like peptide, intocin, proctolin, pyrokinin, short neuropeptide F, SIFamide and sulfakinin families. This is the first, and thus far only, peptidome predicted for a myriapod. Of particular note were a modified AST-C, TYWKQCAFNAVSRFamide, that lacks one of two cysteine residues (*i.e.* one at position 13) stereotypically present in members of this peptide family (and hence is missing the disulfide bridge that spans these residues) and a SIFamide, PPFNGSIFamide, that is truncated due to a lysine for arginine substitution in the dibasic residue pair commonly located at positions 3 and 4 of stereotypical full-length isoforms (*e.g.* the crustacean peptide GYRKPPFNGSIFamide). The peptides predicted here represent the only extant resource for initiating investigations of native peptidergic signaling in the Myriapoda.

1. Introduction

Arthropods have a long history of use in the field of neuroendocrinology, and in particular have contributed greatly to our understanding of the basic principles that underlie peptidergic control of physiology and behavior (*e.g.* Altstein and Nässel, 2010; Bendena, 2010; Christie, 2011; Christie et al., 2010a; Clynen et al., 2010; Nässel and Winther, 2010; Van Wielendaele et al., 2013). Of the extant arthropod subphyla, the Hexapoda and the Crustacea are by far the most studied in terms of their native neuropeptides (*e.g.* Audsley et al., 2015; Christie, 2008a, 2014a, 2014b, 2014c, 2014d, 2014e, 2014f, 2015a, 2015b; Christie and Chi, 2015a, 2015b; Christie et al., 2008a, 2010b, 2011a, 2013; Dircksen et al., 2011; Fu et al., 2005a, 2005b; Gard et al., 2009; Hui et al., 2013; Hummon et al., 2006; Huybrechts et al., 2010; Li et al., 2003; Ma et al., 2008; 2009, 2010; Rahman et al., 2013; Suwansa-Ard et al., 2015; Toullec et al., 2013; Ventura et al., 2014; Xu et al., 2015; Yan et al., 2012), with some structural data also available for members of the Chelicerata (*e.g.* Christie, 2008b, 2015c; Christie and Chi, 2015c; Christie et al., 2011b; Donohue et al., 2010; Egekwu et al., 2014; Neupert et al., 2009; Veenstra et al., 2012). Interestingly, the Myriapoda (which includes millipedes, centipedes, symphylans [pseudocentipedes] and pauropods) remains uninvestigated in terms of its native peptide hormones. In fact, to the best of this author's knowledge, no native neuropeptides have been described from any myriapod species.

Interest in understanding the relatedness between and within arthropod subphyla has led to ever growing transcriptomic resources for the Arthropoda (*e.g.* Misof et al., 2014; Rehm et al., 2014; von Reumont et al., 2012). When a whole organism (or the nervous system) is used as the source for RNA, the resulting transcriptome can be used to identify neuropeptide-encoding transcripts by homology to known precursor protein sequences (*e.g.* Christie et al., 2010a). Via

translation of the identified transcripts, and subsequent processing of the deduced proteins through a simple peptide prediction workflow, mature neuropeptide structures can easily be predicted (*e.g.* Christie et al., 2010a). Using this strategy, large peptidomes for a wide variety of arthropod species, including ones from little studied taxa, have recently been predicted (*e.g.* Christie, 2008a, 2008b, 2014a, 2014b, 2014c, 2014d, 2014e, 2014f, 2015a, 2015b; 2015c; Christie and Chi, 2015a, 2015b, 2015c; Christie et al., 2008a, 2010b, 2011b, 2011c, 2013; Gard et al., 2009; Ma et al., 2009, 2010).

One species for which a relatively large (~70,000 sequences) transcriptome (**BioProject Accession No. PRJNA208708**; Meusemann and Misof, unpublished direct submission) has recently been publicly released is the pseudoscorpion *Symphylella vulgaris* (Rehm et al., 2014). This transcriptome was derived from a whole individual (see the description of the BioSample [**Accession No. SAMN01907470**] associated with **BioProject PRJNA208708** at: http://www.ncbi.nlm.nih.gov/biosample?LinkName=bioproject_biosample_all&from_uid=208708), and thus represents a powerful new resource for *in silico* neuropeptide discovery in a member of the Myriapoda.

In the study presented here, the *S. vulgaris* transcriptome shotgun assembly (TSA) dataset was mined for putative neuropeptide-encoding sequences. Specifically, tblastn searches of the extant *S. vulgaris* TSA collection were conducted using known arthropod proteins as the input queries; these searches identified 47 transcripts as encoding putative neuropeptide precursors. Translation of these sequences allowed for the deduction of 31 unique pre/preprohormones, with 97 distinct mature peptides predicted from the deduced proteins using a simple bioinformatics workflow. The identified peptides included members of the adipokinetic hormone (AKH)/red pigment concentrating hormone (RPCH), adipokinetic hormone-corazonin-like peptide (ACP),

allatostatin A (AST-A), allatostatin C (AST-C), allatotropin, CCHamide, crustacean cardioactive peptide (CCAP), GSEFLamide, insulin-like peptide (ILP), intocin, proctolin, pyrokinin, short neuropeptide F (sNPF), SIFamide and sulfakinin families, as well as a large number of linker/precursor-related peptides. All of the peptides predicted here are novel for *S. vulgaris*, and for that matter the Myriapoda, with the majority being novel discoveries in a general sense. Being the first peptides predicted from any myriapod, the *S. vulgaris* peptidome described here provides the first resource for initiating studies of peptidergic signaling in members of the Myriapoda.

2. Materials and methods

2.1. Database searches

Database searches were conducted on or before May 15, 2015, using methods modified from a well-vetted protocol (*e.g.* Christie, 2008a, 2008b, 2014a, 2014b, 2014c, 2014d, 2014e, 2014f, 2015a, 2015b; 2015c; Christie and Chi, 2015a, 2015b, 2015c; Christie et al., 2008a, 2010b, 2011b, 2011c; 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. vulgaris* “(taxid:1288507)”. Known arthropod peptide precursors were input into tblastn as the query sequences, and all hits returned by a given search were fully translated using the “Translate” tool of ExPASy (<http://web.expasy.org/translate/>) and then checked manually for homology to the target query. The complete list of peptide families searched for in this study, as well as the specific queries used, 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 workflow (*e.g.* Christie, 2008a, 2008b, 2014a, 2014b, 2014c, 2014d, 2014e, 2014f, 2015a, 2015b; 2015c; Christie and Chi, 2015a, 2015b, 2015c; Christie et al., 2008a, 2010b, 2011a, 2011b; 2011c, 2013; Gard et al., 2009; Ma et al., 2009, 2010). Specifically, each of the deduced precursor proteins was assessed for the presence of a signal peptide using the online program SignalP 4.1 (<http://www.cbs.dtu.dk/services/SignalP/>; Petersen et al., 2011); the D-cutoff values for the program were set to “Sensitive”. Prohormone cleavage sites were identified based on the information presented in Veenstra (2000) and/or by homology to known arthropod pre/preprohormone processing schemes. When present, prediction of the sulfation state of tyrosine residues was conducted using the online program “Sulfinator” (<http://www.expasy.org/tools/sulfinator/>; Monigatti et al., 2002). Disulfide bonding between cysteine residues was predicted by homology to known peptide isoforms and/or by using the online program “DiANNA” (<http://clavius.bc.edu/~clotelab/DiANNA/>; Ferrè and Clote, 2005). Other post-translational modifications, *e.g.* cyclization of amino (N)-terminal glutamine/glutamic acid residues and carboxyl (C)-terminal amidation at glycine residues, were predicted by homology to known arthropod peptide isoforms. 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/>;
Kato and Standley, 2013).

3. Results

Thirty-six distinct peptide families/subfamilies were searched for within the extant *S. vulgaris* 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 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/red pigment concentrating hormone

Using a prepro-RPCH from the crayfish *Cherax quadricarinatus* (**Accession No. AAV80404**; Martinez-Perez et al., 2007) as the tblastn query sequence, four *S. vulgaris* transcripts were identified as encoding putative AKH/RPCH precursor proteins (Table 1). Translation of two of the transcripts revealed identical, 70 amino acid, full-length preprohormones (Symvu-prepro-AKH I; Fig. 1A), with a 56 amino acid, full-length precursor deduced from a third sequence (Symvu-prepro-AKH II; Fig. 2A). Translation of the fourth transcript revealed a 64 amino acid C-terminal partial protein that is identical to the C-terminus of Symvu-prepro-AKH I. Three distinct peptides were predicted from Symvu-prepro-AKH I (Table 2 and Fig. 1A), one of which, pQVNFSPGWamide, is identical in structure to an isoform

of AKH described previously from an unidentified member of the hexapod family Mantophasmatidae (Gäde et al., 2005). Two peptides were predicted from Symvu-prepro-AKH II (Table 2 and Fig. 2A), pQVNFSPGWamide, which is the same AKH isoform predicted from Symvu-prepro-AKH I, and the linker/precursor related peptide SESPCYHLQGDSDDY_(SO₃H)LCSHSE (disulfide bridge predicted by DiANNA between the two cysteine residues), which is distinct in structure from those predicted from Symvu-prepro-AKH I.

3.2. Adipokinetic hormone-corazonin-like peptide

Eight *S. vulgaris* transcripts were identified as encoding putative ACP precursors using the sequence of a partial prepro-ACP from the remipede *Speleonectes cf. tulumensis* (Christie, 2014a) as the tblastn query protein (Table 1). Translation of these sequences revealed each to encode an 82 amino acid, full-length precursor. Four distinct preprohormones were identified by alignment of the eight proteins (alignment not shown): Symvu-prepro-ACP I (Fig. 2B1; encoded by GAKX01010333, GAKX01032230, GAKX01040717 and GAKX01055497), Symvu-prepro-ACP II (Fig. 2B2; encoded by GAKX01018684 and GAKX01009943), Symvu-prepro-ACP III (Fig. 2B3; encoded by GAKX01062638) and Symvu-prepro-ACP IV (Fig. 2B4; encoded by GAKX01023780). All variation seen between Symvu-prepro-ACP I-III is restricted to the signal peptide, with Symvu-prepro-ACP II differing from Symvu-prepro-ACP I at one residue and Symvu-prepro-ACP III differing from Symvu-prepro-ACP I at two residues; Symvu-prepro-ACP II and III share one of the substituted positions. Three distinct peptides were predicted from the common prohormone of Symvu-prepro-ACP I-III (Table 2 and Figs. 2B1-2B3), one of which, pQVTFSKDWKAamide, differs at just two residues (both conservative

amino acid substitutions) from a broadly conserved insect/crustacean isoform of ACP, *i.e.* pQVTFSRDWNAamide (Christie, 2014a; Hansen et al., 2010). Three peptides were also predicted from Symvu-prepro-ACP IV (Table 2 and Fig. 2B4), pQVTFSKDWKAamide and LQRIVKMETERLEECGGSLANILDEIPRIE, both of which are shared with the other ACP precursors, and SQDGCSASLFALL, which differs at a single residue (a leucine for serine substitution at position 9) from its counterpart in the other three proteins.

3.3. Allatostatin A

Five *S. vulgaris* transcripts were identified as encoding putative prepro-AST-As using the sequence of an AST-A precursor protein from the crayfish *Procambarus clarkii* (**Accession No. BAE45266**; Yasuda-Kamatani and Yasuda, 2006) as the tblastn query (Table 1). Translation of one of these sequences revealed a 273 amino acid, C-terminal partial prehormone (Symvu-prepro-AST-A I; Fig. 2C1); another transcript encodes a 138 amino acid, C-terminal partial protein that is identical in sequence to Symvu-prepro-AST-A I. A third transcript also encodes a 138 amino acid, C-terminal partial precursor (Symvu-prepro-AST-A II; Fig. 2C2), which differs from the corresponding portion of Symvu-prepro-AST-A I at one position (alignment not shown); a 128 amino acid portion of Symvu-prepro-AST-A II is encoded by another transcript. Translation of the final sequence revealed a 136 amino acid, C-terminal partial protein (Symvu-prepro-AST-A III; Fig. 2C3); this partial precursor differs from the extant portions of Symvu-prepro-AST-A I and II at 3 and 2 positions, respectively (alignments not shown). Twenty-three peptides (22 full-length and one partial) were predicted from Symvu-prepro-AST-A I (Fig. 2C1). Of these peptides, 12 possess the C-terminal motif –YXFGLamide, where *X* represents a variable residue, which is the hallmark of the AST-A family (Christie et al., 2010a); the partial peptide

predicted from Symvu-prepro-AST-A I is also likely an AST-A family member. Of the twelve full-length AST-As derived from Symvu-prepro-AST-A I, six are present as single copies within the extant portion of this precursor, with six copies of one isoform, *i.e.* AAKSYAFGLamide, present in the protein (Table 2). All of the linker/precursor-related peptides derived from Symvu-prepro-AST-A I possess distinct structures (Table 2). Ten peptides were predicted from the extant portion of Symvu-prepro-AST-A II (Fig. 2C2), nine of which (four AST-As and five linker/precursor-related peptides) are shared with Symvu-prepro-AST-A I; the tenth peptide is a novel isoform of AST-A, *i.e.* SASDRMYSFGLamide (Table 2), which possesses a serine for alanine substitution at position 8 relative to its counterpart in Symvu-prepro-AST-A I. The extant portion of Symvu-prepro-AST-A III is also predicted to liberate ten peptides (Fig. 2C3), and here eight (three AST-As and five linker/precursor-related peptides) are shared with Symvu-prepro-AST-A I. Like Symvu-prepro-AST-A II, Symvu-prepro-AST-A III is also predicted to produce the AST-A isoform SASDRMYSFGLamide, with an additional novel AST-A also likely liberated from this partial protein, *i.e.* VVKSYAFGLamide (Table 2), which possesses valine residues at positions 1 and 2 rather than the alanines seen in its counterpart in Symvu-prepro-AST-A I and II.

3.4. Allatostatin C

A single *S. vulgaris* transcript was identified as encoding a putative AST-C precursor using the sequence of *Homarus americanus* prepro-AST-C-like peptide (Dickinson et al., 2009a) as the tblastn query protein (Table 1). Translation of this sequence revealed a 106 amino acid, full-length prehormone (Symvu-prepro-AST-C; Fig. 2D), from which three distinct peptides were predicted (Table 2). One of the predicted peptides, TYWKQCAFNAVSRFamide, differs

from the known and highly conserved arthropod AST-C-like peptide SYWKQCAFNAVSCFamide (*e.g.* Dickinson et al., 2009a; Hummon et al., 2006; Veenstra, 2009) at just two residues, namely a threonine for serine substitution at position 1 and an arginine for cysteine substitution at position 13. The latter substitution is of particular note as it predicted to result in TYWKQCAFNAVSRFamide lacking the stereotypical disulfide bridge that exists between the cysteine typically located at position 13 and one located at position 6, *i.e.* the bridge that exists between the two cysteines in SYWKQCAFNAVSCFamide.

3.5. Allatotropin

Two *S. vulgaris* transcripts were identified as encoding putative allatotropin precursors using the sequence of a copepod *Tigriopus californicus* prepro-allatotropin (Christie, 2014b) as the tblastn query protein (Table 1). Translation of one of these sequences revealed a 113 amino acid, full-length preprohormone (Symvu-prepro-allatotropin I; Fig. 2E1), while translation of the other revealed a 102 amino acid, C-terminal partial protein (Symvu-prepro-allatotropin II; Fig. 2E2). Three distinct peptides were predicted from Symvu-prepro-allatotropin I (Table 2 and Fig. 2E1), one of which, GFKNPALSTARGFamide, possess the C-terminal motif –TARGFamide, a feature considered the hallmark of the allatotropin family (Elekonich and Horodyski, 2003). Three peptides were also predicted from Symvu-prepro-allatotropin II (Table 2 and Fig. 2E3), the allatotropin GFKNPALSTARGFamide (which is shared with Symvu-prepro-allatotropin I) and two linker/precursor-related peptides that are similar to, but distinct from, their counterparts in Symvu-prepro-allatotropin I.

3.6. CCHamide

Two *S. vulgaris* transcripts were identified as encoding putative CCHamide precursors using the sequence of *Litopenaeus vannamei* prepro-CCHamide (Christie, 2014f) as the tblastn query protein (Table 1). Translation of one of the two transcripts revealed 107 amino acid, full-length preprohormone (Symvu-prepro-CCHamide I; Fig. 1B), while translation of the other sequence revealed a 104 amino acid, full-length precursor (Symvu-prepro-CCHamide II; Fig. 2F). Four distinct peptides were predicted from Symvu-prepro-CCHamide I (Table 2 and Fig. 1B), including LKGCSSFGHSCFGAHamide (disulfide bridging predicted between the two cysteine residues), which appears to be an N-terminally extended CCHamide, as it possesses the structural motif XCXXW/Y/FGXXCXGXHamide (disulfide bridging between the position 2 and 9 cysteines), which is the hallmark of CCHamide family members (Hansen et al., 2011). Three peptides were predicted from Symvu-prepro-CCHamide II (Table 2 and Fig. 2F): the same N-terminally extended CCHamide predicted from Symvu-prepro-CCHamide I and two novel linker/precursor-related sequences.

3.7. Crustacean cardioactive peptide

Seven *S. vulgaris* transcripts were identified as encoding putative CCAP precursors using *Homarus gammarus* prepro-CCAP (Accession No. ABB46292; Chung et al., 2006) as the tblastn query sequence (Table 1). Translation of five of the transcripts revealed each to encode a 129 amino acid, full-length protein. Three different proteins were identified from this collection CCAP I, (Fig. 2G1; encoded by GAKX01056315, GAKX01073013 and GAKX01005072) Symvu-prepro-CCAP II (Fig. 2G2; encoded by GAKX01035926) and Symvu-prepro-CCAP III (Fig. 2G3; encoded by GAKX01019483); these proteins differ from one another by one or two amino acids. Translation of the sixth transcript revealed a 126 amino acid, full-length

preprohormone (Symvu-prepro-CCAP IV; Fig. 2G4) that, while similar to Symvu-prepro-CCAP I-III, is considerably more distinct in sequence. Translation of the final transcript revealed a 98 amino acid, full-length protein (Symvu-prepro-CCAP V; Fig. 2G5), which possesses a sequence distinct from all of the other *S. vulgaris* CCAP precursors. As all of the amino acid variation seen between Symvu-prepro-CCAP I-III lies in the signal sequence, a common set of six distinct peptides was predicted from each protein (Table 2 and Figs 2G1-2G3). One of the predicted peptides PFCNAFTGCamide (disulfide bridging between the two cysteine residues) possesses a structure identical to that of authentic CCAP (Christie et al., 2010a). Five peptides were predicted from Symvu-prepro-CCAP IV (Table 2 and Fig. 2G4). Three of these peptides, authentic CCAP and the linker/precursor-related peptides ESERTEEFSGSDEPLMP and SDTNYVDREDDV, were also predicted from Symvu-prepro-CCAP I-III, with the remaining two being novel linker/precursor-related sequences. Three peptides were from Symvu-prepro-CCAP V (Table 2 and Fig. 2G5): authentic CCAP, ESERTEEFSGSDEPLMP and a novel linker/precursor-related peptide.

3.8. GSEFLamide

A single transcript was identified as encoding a putative GSEFLamide precursor using *T. californicus* prepro-GSEFLamide Ia (Christie, 2014b) as the tblastn query protein. Translation of this sequence revealed a 169 amino acid, N-terminal partial preprohormone (Symvu-prepro-GSEFLamide; Fig. 2H). Eleven peptides (ten full-length and one partial) were predicted from the extant portion of Symvu-prepro-GSEFLamide (Table 2 and Fig. 2H). Of these peptides, six possess a –GSEFLamide C-terminus, or a close approximation thereof, which is the defining feature of the GSEFLamide family (Christie, 2014b). The GSEFLamide variants

VGSEFIamide, LGSEFIamide and LGSEFVamide are present as single copies in the extant portion of Symvu-prepro-GSEFIamide, where as LGSEFIamide is encoded three times in the protein (Table 2 and Fig. 2H). The five linker/precursor-related peptides (four full-length and one partial) all possess distinct structures (Table 2).

3.9. Insulin-like peptide

A single *S. vulgaris* transcript was identified as encoding a putative ILP precursor using *Daphnia pulex* prepro-ILP 2 (**Accession No. EFX70023**; Colbourne et al., 2011) as the tblastn protein query. Translation of this sequence revealed a 142 amino acid, full-length preprohormone (Symvu-prepro-ILP; Fig. 2I). Four distinct peptides were predicted from Symvu-prepro-ILP (Table 2 and Fig. 2I). Three of the predicted peptides show sequence homology to known arthropod A-, B- and C-chain ILPs, *e.g.* those derived from *D. pulex* prepro-ILP 2 (Dirksen et al., 2011), and based on homology to known isoforms, disulfide bridging is predicted to be present within the *S. vulgaris* A-chain peptide DVGFYRTLRLDGVHDECCKKSCSYQELLQYCATPDNTIIVD (the first and third cysteines bridged; Table 2) and between it and the predicted B-chain isoform STGFKFCGARLVGILSMLCDSNYYDSTS (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; the cysteines involved in these inter-peptide bridges are not highlighted in the peptide structures shown in Table 2).

3.10. Intocin

A single *S. vulgaris* transcript was identified as encoding a putative intocin precursor using *L. vannamei* prepro-intocin (Christie, 2014f) as the tblastn query sequence. Translation of this transcript revealed a 149 amino acid, full-length protein (Symvu-prepro-intocin; Fig. 2J). Two distinct peptides were predicted from Symvu-prepro-intocin (Table 2 and Fig. 2J), one of which, CFITNCPIGamide (disulfide bridging between the position 1 and 6 cysteines), possesses the structural hallmarks of the intocin family, *i.e.* an overall length of nine amino acids, the structure CXXXNCPXGamide, and a disulfide bridge present between the position 1 and 6 cysteine residues (Stafflinger et al., 2008). DiANNA analysis of the linker/precursor-related protein derived from Symvu-prepro-intocin predicts the presence of disulfide bonds between the first and ninth, second and fourth, third and fourteenth, fifth and eighth, sixth and tenth, seventh and eleventh and twelfth and thirteenth cysteines in this peptide (Table 2).

3.11. Proctolin

Three *S. vulgaris* transcripts were identified as encoding putative proctolin precursors using the sequence of *Penaeus monodon* prepro-proctolin (Christie, 2014f) as the tblastn query protein (Table 1). Translation of these sequences revealed all to encode identical, 80 amino acid, full-length preprohormones (Symvu-prepro-proctolin; Fig. 2K). Four distinct peptides were predicted from Symvu-prepro-proctolin (Table 2 and Fig. 2K), one of which, RYLPT, possesses a structure identical to authentic proctolin (Christie et al., 2010a).

3.12. Pyrokinin

A single *S. vulgaris* transcript was identified as encoding a putative pyrokinin precursor using the sequence of a partial *L. vannamei* pyrokinin preprohormone (Christie, 2014f) as the

tblastn query protein (Table 1). Translation of this sequence revealed a 198 amino acid, C-terminal partial protein (Symvu-prepro-pyrokinin; Fig. 2L). Eleven peptides were predicted from the extant portion of Symvu-prepro-pyrokinin (Table 2 and Fig. 2L), including six that possess the C-terminal motif –FXPRLamide, or a close approximation thereof, which is the defining feature of members of the pheromone biosynthesis activating neuropeptide/diapause hormone/pyrokinin superfamily (Rafaeli, 2009).

3.13. Short neuropeptide F

Five *S. vulgaris* transcripts were identified as encoding putative sNPF precursors using the sequence of *D. pulex* prepro-sNPF (**Accession No. EFX90018**; Colbourne et al., 2011) as the tblastn protein query (Table 1). Translation of three of these sequences revealed 98 amino acid, full-length precursors. Two of these deduced proteins are identical in sequence (Symvu-prepro-sNPF I; Fig. 2M1 [encoded by **GAKX01076964** and **GAKX01011699**]), with the third differing in sequence at just a single residue (Symvu-prepro-sNPF II; Fig. 2M2 [encoded by **GAKX01076393**]). Translation of the remaining two transcripts revealed an identical 96 amino acid, full-length preprohormone (Symvu-prepro-sNPF III; Fig. 2M3); with the exception of a two amino acid deletion, this protein is identical in sequence to Symvu-prepro-sNPF I. Three peptides were predicted from each of the three *S. vulgaris* prepro-sNPFs (Table 2 and Figs. 2M1-2M3). All three proteins are predicted to give rise to AGRDPSLRLRFamide, which possesses the hallmarks of the sNPF family, *i.e.* the C-terminal motif –RXRFamide and an overall length of ten or so amino acids (Christie et al., 2010a), as well as the linker/precursor-related peptide SDPLWQQYGPEETQSEGPASN (Table 2 and Figs. 2M1-2M3); each prepro-sNPF is also

predicted to give rise to a third linker/precursor-related peptide that is distinct in structure (Table 2 and Figs. 2M1-2M3).

3.14. SIFamide

Four *S. vulgaris* transcripts were identified as encoding putative SIFamide precursors using the sequence of *H. americanus* prepro-valine¹-SIFamide (**Accession No. ABV21807**; Dickinson et al., 2008) as the tblastn protein query (Table 1). Translation of three of these sequences revealed identical, 71 amino acid, full-length precursors (Symvu-prepro-SIFamide I: Fig. 2N1), while translation of the remaining transcript revealed a 69 amino acid, full-length preprohormone (Symvu-prepro-SIFamide II: Fig. 2N2). Both proteins are predicted to give rise to PPFNGSIFamide, a peptide possessing the C-terminal motif –SIFamide, the defining feature of SIFamides (Christie et al., 2010a), and the dipeptide GY (Table 2 and Figs. 2N1-2N2). The second linker/precursor-related peptide in each preprohormone is distinct in structure (Table 2 and Figs. 2N1-2N2), *i.e.* APGELATDKLYAVCEIAVETCSQFFQGFESK (from Symvu-prepro-SIFamide I) and APGELATDKLYAVCEIAVETCYPIFPRL (from Symvu-prepro-SIFamide II), and possesses a disulfide bridge between its two cysteine residues.

3.15. Sulfakinin

Two *S. vulgaris* transcripts were identified as encoding putative sulfakinin precursors using the sequence of *H. americanus* prepro-sulfakinin (**Accession No. ABQ95346**; Dickinson et al., 2007) as the tblastn protein query (Table 1). Translation of these sequences revealed each to encode a C-terminal partial protein, with one partial preprohormone being 112 amino acids in length (Symvu-prepro-sulfakinin I; Fig. 1C) and the other, 97 amino acid long (Symvu-prepro-

sulfakinin II; Fig. 2O). Six distinct peptides were predicted from Symvu-prepro-sulfakinin I (Table 2 and Fig. 1C), two of which, pQSDDY_(SO₃H)GHLRFamide and DFDDY_(SO₃H)GHMRFamide, possess the structural signature of members of the sulfakinin family, *i.e.* the C-terminal motif –Y_(SO₃H)GHM/LRFamide (Christie et al., 2010a). With the exception of one linker/precursor-related peptide, *i.e.* KPAPTNGDNS (KPAPTNGNS from Symvu-prepro-sulfakinin I), an identical set of peptides was predicted from Symvu-prepro-sulfakinin II (Table 2 and Fig. 2O).

4. Discussion

4.1. Transcriptomic resources for the Myriapoda

With decreasing costs for high-throughput nucleotide sequencing projects, large transcriptomic data sets have recently been generated for many arthropod species, including members of a number of taxa that have previously had few, or no, molecular resources available for them, *e.g.* the crustacean class Remipedia (von Reumont et al., 2012) and hexapod class Entognatha (Misof et al., 2014). Interestingly, essentially no large-scale transcriptomic sequencing has been done for any myriapod, this despite the fact that the Myriapoda is one of the four extant subphyla of the Arthropoda. The recent public deposition of a 70,000⁺-sequence transcriptome for the pseudoscorpion *S. vulgaris* has begun to fill this void in sequence data for the myriapods; the *S. vulgaris* transcriptome constitutes nearly 50% of the total nucleotide sequences publicly accessible for the Myriapoda (171,445 total nucleotide sequences were present in NCBI on June 26, 2015). Thus, the transcriptome of *S. vulgaris* is currently the only large-scale molecular resource available for *in silico* protein discovery in myriapods.

4.2. Neuropeptide discovery in *Symphylella vulgaris* using *in silico* transcriptome mining

Classically, the neuropeptidome of an organism was characterized by identifying individual neuropeptides in a one-by-one fashion, typically via biochemical purification and sequencing and/or molecular cloning, until the full, or at least a large, catalog of peptide structures was assembled; because of the time and tissue need for neuropeptide discovery in a one-by one fashion (*e.g.* the 3,500 central nervous systems needed to isolate and characterize the two native sulfakinins from the shrimp *L. vannamei* [Torfs et al., 2002]), large peptidomes were rarely achieved via this approach. More recently, mass spectrometry, in particular high-resolution mass profiling and tandem mass spectrometric sequencing, has been used to rapidly identify large peptidomes based on accurate mass matching and/or *de novo* sequence analyses (*e.g.* Fu et al., 2005a, 2005b; Hui et al., 2013; Li et al., 2003; Ma et al., 2008; 2009, 2010). However, when the organism in question is microscopic, rare and/or geographically inaccessible, it is often difficult to obtain enough starting material for the elucidation of a full peptidome via mass spectral means. In these cases, *in silico* genome/transcriptome mining offers an alternative approach to peptide discovery. For this strategy, more limited starting material is typically needed to generate a nucleotide sequence dataset; in some instances tissue from just a single individual is sufficient. Moreover, once a genome or transcriptome is constructed, it provides a permanent resource for mining transcripts encoding essentially any protein of interest. Given *S. vulgaris*' small size and habitat (leaf litter), transcriptome mining provides perhaps the most efficient means of large-scale neuropeptide discovery in this species.

Via *in silico* mining of the extant *S. vulgaris* TSA dataset, 47 transcripts encoding putative neuropeptide were identified. Translation of these sequences revealed them to encode 31 unique pre/preprohormones, from which the mature structures of 97 distinct peptides were predicted.

These peptides include members of 15 different peptide families: AKH/RPCH, ACP, AST-A, AST-C, allatotropin, CCHamide, CCAP, GSEFLamide, ILP, intocin, proctolin, pyrokinin, sNPF, SIFamide and sulfakinin; a wide array of linker/precursor-related peptides whose structures do not place them into any described peptide family were also predicted. The peptidome described here for *S. vulgaris* is the first from any member of the Myriapoda.

While *S. vulgaris* isoforms of many well-known arthropod peptide families were predicted in this study, no transcripts were identified for other peptide groups that appear broadly conserved in the other arthropod subphyla. For example, no transcripts encoding members of the allatostatin B, bursicon α , bursicon β , corazonin, crustacean hyperglycemic hormone/ion transport peptide, diuretic hormone 31, diuretic hormone 44, ecdysis-triggering hormone, eclosion hormone, FMRFamide-like peptide, leucokinin, myosuppressin, neuroparsin, neuropeptide F, orcokinin, pigment dispersing hormone/factor, RYamide, and tachykinin-related peptide families were found. Given the absence of peptide sequence data from any other myriapod, it is at present impossible to determine if the lack of detection of transcripts for members of these peptide families in the *S. vulgaris* transcriptome truly represents a loss of these peptide groups in the Myriapoda. Given the relatively small number of transcripts in the *S. vulgaris* transcriptome, it seems likely that their lack of detection simply reflects low transcriptome coverage, and that as more sequence data is added for this species, that transcripts encoding these families will be identified.

4.3. The native *Symphylella vulgaris* allatostatin C and SIFamide possess unusual structures

While all of the peptides predicted here are novel discoveries for *S. vulgaris*, two are of particular interest, as they possess structures that are atypical for full-length members of their

451 respective families. These peptides are the AST-C isoform TYWKQCAFNAVSRFamide and
452 the SIFamide PPFNGSIFamide. With respect to the AST-Cs, stereotypical family members
453 exhibit an overall length of 15 amino acids, and possess a pyroglutamic acid blocked N-terminus,
454 an unamidated –PISCF C-terminus, and a disulfide bridge between the cysteine residues located
455 at positions 7 and 14, *e.g.* pQIRYHQCYFNPISCF (disulfide bridging between the two cysteine
456 residues), a conserved decapod crustacean isoform AST-C (Stemmler et al., 2010). A large
457 number of variant AST-C have also been identified, including SYWKQCAFNAVSCFamide
458 (disulfide bonding between the cysteines at positions 6 and 13), a peptide that is broadly
459 conserved in arthropods (*e.g.* Dickinson et al., 2009a; Hummon et al., 2006; Veenstra, 2009);
460 this AST-C is 14 amino acids long, possess an unblocked N-terminus, and an amidated C-
461 terminus, but possesses disulfide bridging between its cysteine residues, which are similarly
462 positioned to those present in stereotypical family members. TYWKQCAFNAVSRFamide, the
463 AST-C predicted here from the *S. vulgaris* transcriptome, differs from
464 SYWKQCAFNAVSCFamide at just two residues, a conservative threonine for serine
465 substitution at position 1 and a non-conservative arginine for cysteine substitution at position 13.
466 While substitutions similar to the former have been noted in other species (*e.g.* Veenstra, 2009),
467 the latter substitution is of note, as it removes one of two cysteine residues need for the disulfide
468 bridging that is ubiquitous in AST-C family members. It is important to remember
469 TYWKQCAFNAVSRFamide, like all of the peptides predicted here, is a theoretical construct,
470 having yet to be confirmed to actually exist in *S. vulgaris* biochemically. Given this, it is
471 possible that position 13 substitution is the result of a sequencing error; the codon that rise to the
472 arginine in the peptide is *cgt* and a miscalled nucleotide at the first position, could convert a
473 codon for cysteine, *tgt*, to that for the arginine present in the transcript in question. Similarly it

could be that the *S. vulgaris* used for RNA isolation had an individual-specific mutation that gave rise to the variant sequence, and that TYWKQCAFNAVSCFamide (with disulfide bridging present between its cysteines) is the isoform present in other members of this species. However, if TYWKQCAFNAVSRFamide is the native AST-C present in *S. vulgaris*, it will be interesting to see what, if any, effect the missing disulfide bridging has on bioactivity relative to a bridged analog.

The vast majority of the SIFamide isoforms thus far identified possess the structure XYRKPPFNGSIFamide, where X, represents a variable residue, most commonly alanine in insects and glycine in crustaceans (*e.g.* Stemmler et al., 2007; Verleyen et al., 2009). Analysis of prohormone cleavage loci suggest that dibasic residue pair “RK” is rarely cleaved in insects (Veenstra, 2000), and mass spectrometric data have shown that this is the case for this residue pair in members of the SIFamide family (*e.g.* Stemmler et al., 2007; Verleyen et al., 2009). The *S. vulgaris* SIFamide precursor contains the sequence GYKKPPFNGSIFG, which possess “KK” rather than “RK”, a dibasic residue pair that does appear to be commonly cleaved. Thus, the native *S. vulgaris* SIFamide was predicted to be PPFNGSIFamide, rather than GYKKPPFNGSIFamide. Like the situation discussed earlier for the predicted *S. vulgaris* AST-C, the position three lysine for arginine substitution could result from a sequencing error or an individual-specific mutation, and if so, the native peptide in *S. vulgaris* could well be GYRKPPFNGSIFamide, which is highly conserved in crustaceans (*e.g.* Stemmler et al., 2007). If, however, PPFNGSIFamide is the correct structure of the native *S. vulgaris* SIFamide, it will be interesting to see what, if any, effect truncation has on SIFamide bioactivity in this species.

4.4. Summary

Prior to this study, no native neuropeptide structures were known from any member of the Myriapoda. Here, a recently deposited transcriptome for the pseudoscorpion *S. vulgaris* was probed for sequences encoding putative peptide precursor proteins. Forty-seven transcripts were identified as encoding putative neuropeptide pre/preprohormones, and using the proteins deduced from them, a total of 97 distinct mature peptides were predicted. These peptides include members of the AKH/RPCH, ACP, AST-A, AST-C, allatotropin, CCHamide, CCAP, GSEFLamide, ILP, intocin, proctolin, pyrokinin, sNPF, SIFamide and sulfakinin families. The discovery of these peptide groups in a member of the Myriapoda provides the first evidence that at least these peptide families are ubiquitously conserved among the extant subphyla of the Arthropoda, as isoforms of each family have now been identified from at least one member of each of the extant arthropod subphyla. In addition to being the first, and thus far only, peptidome predicted for any myriapod, the data presented here also serve as a template for future peptide discovery in other members of the Myriapoda, and as a resource for initiating investigations of peptide function in members of this little studied arthropod taxon.

Acknowledgements

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

References

Altstein, M., Nässel, D.R., 2010. Neuropeptide signaling in insects. *Adv. Exp. Med. Biol.* 692, 155-165.

520 Audsley, N., Down, R.E., Isaac, R.E., 2015. Genomic and peptidomic analyses of the
 521 neuropeptides from the emerging pest, *Drosophila suzukii*. *Peptides* 68, 33-42.
 522
 523 Bendena, W.G., 2010. Neuropeptide physiology in insects. *Adv. Exp. Med. Biol.* 692, 166-191.
 524
 525 Christie, A.E., 2008a. *In silico* analyses of peptide paracrines/hormones in Aphidoidea. *Gen.*
 526 *Comp. Endocrinol.* 159, 67-79.
 527
 528 Christie, A.E., 2008b. Neuropeptide discovery in Ixodoidea: an *in silico* investigation using
 529 publicly accessible expressed sequence tags. *Gen. Comp. Endocrinol.* 157, 174-185.
 530
 531 Christie, A.E., 2011. Crustacean neuroendocrine systems and their signaling agents. *Cell Tissue*
 532 *Res.* 345, 41-67.
 533
 534 Christie, A.E., 2014a. Prediction of the first neuropeptides from a member of the Remipedia
 535 (Arthropoda, Crustacea). *Gen. Comp. Endocrinol.* 201, 74-86.
 536
 537 Christie, A.E., 2014b. Prediction of the peptidomes of *Tigriopus californicus* and
 538 *Lepeophtheirus salmonis* (Copepoda, Crustacea). *Gen. Comp. Endocrinol.* 201, 87-106.
 539
 540 Christie, A.E., 2014c. Peptide discovery in the ectoparasitic crustacean *Argulus siamensis*:
 541 identification of the first neuropeptides from a member of the Branchiura. *Gen. Comp.*
 542 *Endocrinol.* 204, 114-125.

543

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

546

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

549

550 Christie, A.E., 2014f. Expansion of the *Litopenaeus vannamei* and *Penaeus monodon*
551 peptidomes using transcriptome shotgun assembly sequence data. Gen. Comp. Endocrinol. 206,
552 235-254.

553

554 Christie, A.E., 2015a. Neuropeptide discovery in *Eucyclops serrulatus* (Crustacea, Copepoda): in
555 silico prediction of the first peptidome for a member of the Cyclopoida. Gen. Comp. Endocrinol.
556 211, 92-105.

557

558 Christie, A.E., 2015b. *In silico* prediction of a neuropeptidome for the eusocial insect
559 *Mastotermes darwiniensis*. Gen. Comp. Endocrinol. In press.

560

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

563

564 Christie, A.E., Cashman, C.R., Brennan, H.R., Ma, M., Sousa, G.L., Li, L., Stemmler, E.A.,
 565 Dickinson, P.S., 2008a. Identification of putative crustacean neuropeptides using *in silico*
 566 analyses of publicly accessible expressed sequence tags. *Gen. Comp. Endocrinol.* 156, 246-64.
 567
 568 Christie, A.E., Cashman, C.R., Stevens, J.S., Smith, C.M., Beale, K.M., Stemmler, E.A.,
 569 Greenwood, S.J., Towle, D.W., Dickinson, P.S., 2008b. Identification and cardiotropic actions
 570 of brain/gut-derived tachykinin-related peptides (TRPs) from the American lobster *Homarus*
 571 *americanus*. *Peptides* 29, 1909-1918.
 572
 573 Christie, A.E., Chapline, M.C., Jackson, J.M., Dowda, J.K., Hartline, N., Malecha, S.R., Lenz,
 574 P.H., 2011d. Identification, tissue distribution and orexigenic activity of neuropeptide F (NPF)
 575 in penaeid shrimp. *J. Exp. Biol.* 214, 1386-1396.
 576
 577 Christie, A.E., Chi, M., 2015a. Identification of the first neuropeptides from the enigmatic
 578 hexapod order Protura. *Gen. Comp. Endocrinol.* In press.
 579
 580 Christie, A.E., Chi, M., 2015b. Prediction of the neuropeptidomes of members of the Astacidea
 581 (Crustacea, Decapoda) using publicly accessible transcriptome shotgun assembly (TSA)
 582 sequence data. *Gen. Comp. Endocrinol.* In press.
 583
 584 Christie, A.E., Chi, M., 2015c. Neuropeptide discovery in the Araneae (Arthropoda, Chelicerata,
 585 Arachnida): elucidation of true spider peptidomes using that of the Western black widow as a
 586 reference. *Gen. Comp. Endocrinol.* 213, 90-109.

587

588 Christie, A.E., Durkin, C.S., Hartline, N., Ohno, P., Lenz, P.H., 2010b. Bioinformatic analyses
589 of the publicly accessible crustacean expressed sequence tags (ESTs) reveal numerous novel
590 neuropeptide-encoding precursor proteins, including ones from members of several little studied
591 taxa. *Gen. Comp. Endocrinol.* 167, 164-178.

592

593 Christie, A.E., McCoole, M.D., Harmon, S.M., Baer, K.N., Lenz, P.H., 2011a. Genomic
594 analyses of the *Daphnia pulex* peptidome. *Gen. Comp. Endocrinol.* 171, 131-150.

595

596 Christie, A.E., Nolan, D.H., Garcia, Z.A., McCoole, M.D., Harmon, S.M., Congdon-Jones, B.,
597 Ohno, P., Hartline, N., Congdon, C.B., Baer, K.N., Lenz, P.H., 2011c. Bioinformatic prediction
598 of arthropod/nematode-like peptides in non-arthropod, non-nematode members of the
599 Ecdysozoa. *Gen. Comp. Endocrinol.* 170, 480-486.

600

601 Christie, A.E., Nolan, D.H., Ohno, P., Hartline, N., Lenz, P.H., 2011b. Identification of
602 chelicerate neuropeptides using bioinformatics of publicly accessible expressed sequence tags.
603 *Gen. Comp. Endocrinol.* 170, 144-155.

604

605 Christie, A.E., Roncalli, V., Wu, L.S., Ganote, C.L., Doak, T., Lenz, P.H., 2013. Peptidergic
606 signaling in *Calanus finmarchicus* (Crustacea, Copepoda): *in silico* identification of putative
607 peptide hormones and their receptors using a *de novo* assembled transcriptome. *Gen. Comp.*
608 *Endocrinol.* 187:117-35.

609

610 Christie, A.E., Stemmler, E.A., Dickinson, P.S., 2010a. Crustacean neuropeptides. Cell. Mol.
 611 Life Sci. 67, 4135-4169.
 612
 613 Christie, A.E., Stevens, J.S., Bowers, M.R., Chapline, M.C., Jensen, D.A., Schegg, K.M.,
 614 Goldwaser, J., Kwiatkowski, M.A., Pleasant, T.K. Jr., Shoenfeld, L., Tempest, L.K., Williams,
 615 C.R., Wiwatpanit, T., Smith, C.M., Beale, K.M., Towle, D.W., Schooley, D.A., Dickinson, P.S.,
 616 2010c. Identification of a calcitonin-like diuretic hormone that functions as an intrinsic
 617 modulator of the American lobster, *Homarus americanus*, cardiac neuromuscular system. J.
 618 Exp. Biol. 213, 118-127.
 619
 620 Chung, J.S., Wilcockson, D.C., Zmora, N., Zohar, Y., Dirksen, H., Webster, S.G., 2006.
 621 Identification and developmental expression of mRNAs encoding crustacean cardioactive
 622 peptide (CCAP) in decapod crustaceans. J. Exp. Biol. 209, 3862-3872.
 623
 624 Clynen, E., Reumer, A., Baggerman, G., Mertens, I., Schoofs, L., 2010. Neuropeptide biology in
 625 *Drosophila*. Adv. Exp. Med. Biol. 692, 192-210.
 626
 627 Colbourne, J.K., Pfrender, M.E., Gilbert, D., Thomas, W.K., Tucker, A., Oakley, T.H.,
 628 Tokishita, S., Aerts, A., Arnold, G.J., Basu, M.K., Bauer, D.J., Cáceres, C.E., Carmel, L., Casola,
 629 C., Choi, J.H., Detter, J.C., Dong, Q., Dusheyko, S., Eads, B.D., Fröhlich, T., Geiler-Samerotte,
 630 K.A., Gerlach, D., Hatcher, P., Jogdeo, S., Krijgsveld, J., Kriventseva, E.V., Kültz, D., Laforsch,
 631 C., Lindquist, E., Lopez, J., Manak, J.R., Muller, J., Pangilinan, J., Patwardhan, R.P., Pitluck, S.,
 632 Pritham, E.J., Rechtsteiner, A., Rho, M., Rogozin, I.B., Sakarya, O., Salamov, A., Schaack, S.,

633 Shapiro, H., Shiga, Y., Skalitzky, C., Smith, Z., Souvorov, A., Sung, W., Tang, Z., Tsuchiya, D.,
 634 Tu, H., Vos, H., Wang, M., Wolf, Y.I., Yamagata, H., Yamada, T., Ye, Y., Shaw, J.R., Andrews,
 635 J., Crease, T.J., Tang, H., Lucas, S.M., Robertson, H.M., Bork, P., Koonin, E.V., Zdobnov, E.M.,
 636 Grigoriev, I.V., Lynch, M., Boore, J.L., 2011. The ecoresponsive genome of *Daphnia pulex*.
 637 Science 331, 555-561.
 638
 639 Dickinson, P.S., Stemmler, E.A., Barton, E.E., Cashman, C.R., Gardner, N.P., Rus, S., Brennan,
 640 H.R., McClintock, T.S., Christie, A.E., 2009b. Molecular, mass spectral, and physiological
 641 analyses of orcokinins and orcokinin precursor-related peptides in the lobster *Homarus*
 642 *americanus* and the crayfish *Procambarus clarkii*. Peptides 30, 297-317.
 643
 644 Dickinson, P.S., Stemmler, E.A., Cashman, C.R., Brennan, H.R., Dennison, B., Huber, K.E.,
 645 Peguero, B., Rabacal, W., Goiney, C.C., Smith, C.M., Towle, D.W., Christie, A.E., 2008.
 646 SIFamide peptides in clawed lobsters and freshwater crayfish (Crustacea, Decapoda, Astacidea):
 647 a combined molecular, mass spectrometric and electrophysiological investigation. Gen. Comp.
 648 Endocrinol. 156, 347-360.
 649
 650 Dickinson, P.S., Stevens, J.S., Rus, S., Brennan, H.R., Goiney, C.C., Smith, C.M., Li, L., Towle,
 651 D.W., Christie, A.E., 2007. Identification and cardiotropic actions of sulfakinin peptides in the
 652 American lobster *Homarus americanus*. J. Exp. Biol. 210, 2278-2289.
 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 Dircksen, H., Neupert, S., Predel, R., Verleyen, P., Huybrechts, J., Strauss, J., Hauser, F.,
660 Stafflinger, E., Schneider, M., Pauwels, K., Schoofs, L., Grimmelikhuijzen, C.J., 2011.
661 Genomics, transcriptomics, and peptidomics of *Daphnia pulex* neuropeptides and protein
662 hormones. J. Proteome Res. 10, 4478-4504.

663

664 Donohue, K.V., Khalil, S.M., Ross, E., Grozinger, C.M., Sonenshine, D.E., Roe R.M., 2010.
665 Neuropeptide signaling sequences identified by pyrosequencing of the American dog tick
666 synganglion transcriptome during blood feeding and reproduction. Insect Biochem. Mol. Biol.
667 40, 79-90.

668

669 Egekwu, N., Sonenshine, D.E., Bissinger, B.W., Roe, R.M., 2014. Transcriptome of the female
670 synganglion of the black-legged tick *Ixodes scapularis* (Acari: Ixodidae) with comparison
671 between Illumina and 454 systems. PLoS One 9, e102667.

672

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

675

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

678

679 Fu, Q., Christie, A.E., Li, L., 2005a. Mass spectrometric characterization of crustacean
 680 hyperglycemic hormone precursor-related peptides (CPRPs) from the sinus gland of the crab,
 681 *Cancer productus*. Peptides 26, 2137-2150.
 682
 683 Fu, Q., Kutz, K.K., Schmidt, J.J., Hsu, Y.W., Messinger, D.I., Cain, S.D., de la Iglesia, H.O.,
 684 Christie, A.E., Li, L., 2005b. Hormone complement of the *Cancer productus* sinus gland and
 685 pericardial organ: an anatomical and mass spectrometric investigation. J. Comp. Neurol. 493,
 686 607-626.
 687
 688 Gäde, G., Marco, H.G., Simek, P., Marais, E., 2005. The newly discovered insect order
 689 Mantophasmatodea contains a novel member of the adipokinetic hormone family of peptides.
 690 Biochem. Biophys. Res. Commun. 330, 598-603.
 691
 692 Gard, A.L., Lenz, P.H., Shaw, J.R., Christie, A.E., 2009. Identification of putative peptide
 693 paracrines/hormones in the water flea *Daphnia pulex* (Crustacea; Branchiopoda; Cladocera)
 694 using transcriptomics and immunohistochemistry. Gen. Comp. Endocrinol. 160, 271-287.
 695
 696 Hansen, K.K., Hauser, F., Williamson, M., Weber, S.B., Grimmlikhuijzen, C.J., 2011. The
 697 *Drosophila* genes CG14593 and CG30106 code for G-protein-coupled receptors specifically
 698 activated by the neuropeptides CCHamide-1 and CCHamide-2. Biochem. Biophys. Res.
 699 Commun. 404, 184-189.
 700

701 Hansen, K.K., Stafflinger, E., Schneider, M., Hauser, F., Cazzamali, G., Williamson, M.,
 702 Kollmann, M., Schachtner, J., Grimmelikhuijzen, C.J., 2010. Discovery of a novel insect
 703 neuropeptide signaling system closely related to the insect adipokinetic hormone and corazonin
 704 hormonal systems. *J. Biol. Chem.* 285, 10736-10747.
 705
 706 Hsu, Y.W., Stemmler, E.A., Messinger, D.I., Dickinson, P.S., Christie, A.E., de la Iglesia, H.O.,
 707 2008b. Cloning and differential expression of two β -pigment-dispersing hormone (β -PDH)
 708 isoforms in the crab *Cancer productus*: evidence for authentic β -PDH as a local neurotransmitter
 709 and β -PDH II as a humoral factor. *J. Comp. Neurol.* 508, 197-211.
 710
 711 Hsu, Y.W., Weller, J.R., Christie, A.E., de la Iglesia, H.O., 2008a. Molecular cloning of four
 712 cDNAs encoding prepro-crustacean hyperglycemic hormone (CHH) from the eyestalk of the red
 713 rock crab *Cancer productus*: identification of two genetically encoded CHH isoforms and two
 714 putative post-translationally derived CHH variants. *Gen. Comp. Endocrinol.* 155, 517-525.
 715
 716 Hui, L., D'Andrea, B.T., Jia, C., Liang, Z., Christie, A.E., Li, L., 2013. Mass spectrometric
 717 characterization of the neuropeptidome of the ghost crab *Ocypode ceratophthalma* (Brachyura,
 718 Ocypodidae). *Gen. Comp. Endocrinol.* 184, 22-34.
 719
 720 Hummon, A.B., Richmond, T.A., Verleyen, P., Baggerman, G., Huybrechts, J., Ewing, M.A.,
 721 Vierstraete, E., Rodriguez-Zas, S.L., Schoofs, L., Robinson, G.E., Sweedler, J.V., 2006. From
 722 the genome to the proteome: uncovering peptides in the *Apis* brain. *Science* 314, 647-649.
 723

724 Huybrechts, J., Bonhomme, J., Minoli, S., Prunier-Leterme, N., Dombrovsky, A., Abdel-Latief,
 725 M., Robichon, A., Veenstra, J.A., Tagu, D., 2010. Neuropeptide and neurohormone precursors
 726 in the pea aphid, *Acyrtosiphon pisum*. *Insect Mol. Biol.* 19, 87-95.
 727
 728 Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7:
 729 improvements in performance and usability. *Mol. Biol. Evol.* 30, 772-780.
 730
 731 Li, L., Kelley, W.P., Billimoria, C.P., Christie, A.E., Pulver, S.R., Sweedler, J.V., Marder, E.,
 732 2003. Mass spectrometric investigation of the neuropeptide complement and release in the
 733 pericardial organs of the crab, *Cancer borealis*. *J. Neurochem.* 87, 642-656.
 734
 735 Ma, M., Bors, E.K., Dickinson, E.S., Kwiatkowski, M.A., Sousa, G.L., Henry, R.P., Smith,
 736 C.M., Towle, D.W., Christie, A.E., Li, L., 2009. Characterization of the *Carcinus maenas*
 737 neuropeptidome by mass spectrometry and functional genomics. *Gen. Comp. Endocrinol.* 161,
 738 320-334.
 739
 740 Ma, M., Chen, R., Sousa, G.L., Bors, E.K., Kwiatkowski, M.A., Goiney, C.C., Goy, M.F.,
 741 Christie, A.E., Li, L., 2008. Mass spectral characterization of peptide transmitters/hormones in
 742 the nervous system and neuroendocrine organs of the American lobster *Homarus americanus*.
 743 *Gen. Comp. Endocrinol.* 156, 395-409.
 744
 745 Ma, M., Gard, A.L., Xiang, F., Wang, J., Davoodian, N., Lenz, P.H., Malecha, S.R., Christie,
 746 A.E., Li, L., 2010. Combining *in silico* transcriptome mining and biological mass spectrometry

747 for neuropeptide discovery in the Pacific white shrimp *Litopenaeus vannamei*. Peptides 31, 27-
748 43.

749

750 Marco, H.G., Anders, L., Gäde, G., 2014. cDNA cloning and transcript distribution of two novel
751 members of the neuroparsin peptide family in a hemipteran insect (*Nezara viridula*) and a
752 decapod crustacean (*Jasus lalandii*). Peptides 53, 97-105.

753

754 Martínez-Pérez, F., Durán-Gutiérrez, D., Delaye, L., Becerra, A., Aguilar, G., Zinker, S., 2007.
755 Loss of DNA: a plausible molecular level explanation for crustacean neuropeptide gene
756 evolution. Peptides 28, 76-82.

757

758 Misof, B., Liu, S., Meusemann, K., Peters, R.S., Donath, A., Mayer, C., Frandsen, P.B., Ware, J.,
759 Flouri, T., Beutel, R.G., Niehuis, O., Petersen, M., Izquierdo-Carrasco, F., Wappler, T., Rust, J.,
760 Aberer, A.J., Aspöck, U., Aspöck, H., Bartel, D., Blanke, A., Berger, S., Böhm, A., Buckley,
761 T.R., Calcott, B., Chen, J., Friedrich, F., Fukui, M., Fujita, M., Greve, C., Grobe, P., Gu, S.,
762 Huang, Y., Jermin, L.S., Kawahara, A.Y., Krogmann, L., Kubiak, M., Lanfear, R., Letsch, H.,
763 Li, Y., Li, Z., Li, J., Lu, H., Machida, R., Mashimo, Y., Kapli, P., McKenna, D.D., Meng, G.,
764 Nakagaki, Y., Navarrete-Heredia, J.L., Ott, M., Ou, Y., Pass, G., Podsiadlowski, L., Pohl, H.,
765 von Reumont, B.M., Schütte, K., Sekiya, K., Shimizu, S., Slipinski, A., Stamatakis, A., Song,
766 W., Su, X., Szucsich, N.U., Tan, M., Tan, X., Tang, M., Tang, J., Timelthaler, G., Tomizuka, S.,
767 Trautwein, M., Tong, X., Uchifune, T., Walz, M.G., Wiegmann, B.M., Wilbrandt, J., Wipfler,
768 B., Wong, T.K., Wu, Q., Wu, G., Xie, Y., Yang, S., Yang, Q., Yeates, D.K., Yoshizawa, K.,
769 Zhang, Q., Zhang, R., Zhang, W., Zhang, Y., Zhao, J., Zhou, C., Zhou, L., Ziesmann, T., Zou, S.,

770 Li, Y., Xu, X., Zhang, Y., Yang, H., Wang, J., Wang, J., Kjer, K.M., Zhou, X., 2014.
 771 Phylogenomics resolves the timing and pattern of insect evolution. *Science* 346, 763-767.
 772
 773 Monigatti, F., Gasteiger, E., Bairoch, A., Jung, E., 2002. The Sulfinator: predicting tyrosine
 774 sulfation sites in protein sequences. *Bioinformatics* 18, 769-770.
 775
 776 Nässel, D.R., Winther, A.M., 2010. *Drosophila* neuropeptides in regulation of physiology and
 777 behavior. *Prog. Neurobiol.* 92, 42-104.
 778
 779 Neupert, S., Russell, W.K., Predel, R., Russell, D.H., Strey, O.F., Teel, P.D., Nachman, R.J.,
 780 2009. The neuropeptidomics of *Ixodes scapularis* synganglion. *J. Proteomics* 72, 1040-1045.
 781
 782 Petersen, T.N., Brunak, S., von Heijne, G., Nielsen, H., 2011. SignalP 4.0: discriminating signal
 783 peptides from transmembrane regions. *Nat. Methods* 8, 785-786.
 784
 785 Rafaeli, A., 2009. Pheromone biosynthesis activating neuropeptide (PBAN): regulatory role and
 786 mode of action. *Gen. Comp. Endocrinol.* 162, 69-78.
 787
 788 Rahman, M.M., Neupert, S., Predel, R., 2013. Neuropeptidomics of the Australian sheep
 789 blowfly *Lucilia cuprina* (Wiedemann) and related Diptera. *Peptides* 41, 31-37.
 790
 791 Rehm, P., Meusemann, K., Borner, J., Misof, B., Burmester, T., 2014. Phylogenetic position of
 792 Myriapoda revealed by 454 transcriptome sequencing. *Mol. Phylogenet. Evol.* 77, 25-33.

793

794 Sharp, J.H., Wilcockson, D.C., Webster, S.G., 2010. Identification and expression of mRNAs
 795 encoding bursicon in the plesiomorphic central nervous system of *Homarus gammarus*. Gen.
 796 Comp. Endocrinol. 169, 65-74.

797

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

801

802 Stemmler, E.A., Bruns, E.A., Cashman, C.R., Dickinson, P.S., Christie, A.E., 2010. Molecular
 803 and mass spectral identification of the broadly conserved decapod crustacean neuropeptide
 804 pQIRYHQCYFNPISCF: the first PISCF-allatostatin (*Manduca sexta*- or C-type allatostatin)
 805 from a non-insect. Gen. Comp. Endocrinol. 165, 1-10.

806

807 Stemmler, E.A., Cashman, C.R., Messinger, D.I., Gardner, N.P., Dickinson, P.S., Christie, A.E.,
 808 2007. High-mass-resolution direct-tissue MALDI-FTMS reveals broad conservation of three
 809 neuropeptides (APSGFLGMRamide, GYRKPPFNGSIFamide and pQDLDHVFLRFamide)
 810 across members of seven decapod crustacean infraorders. Peptides 28, 2104-2115.

811

812 Stevens, J.S., Cashman, C.R., Smith, C.M., Beale, K.M., Towle, D.W., Christie, A.E.,
 813 Dickinson, P.S., 2009. The peptide hormone pQDLDHVFLRFamide (crustacean
 814 myosuppressin) modulates the *Homarus americanus* cardiac neuromuscular system at multiple
 815 sites. J. Exp. Biol. 212, 3961-3976.

816

817 Suwansa-Ard, S., Thongbuakaew, T., Wang, T., Zhao, M., Elizur, A., Hanna, P.J., Sretarugsa, P.,
818 Cummins, S.F., Sobhon, P., 2015. *In silico* neuropeptidome of female *Macrobrachium*
819 *rosenbergii* based on transcriptome and peptide mining of eyestalk, central nervous system and
820 ovary. PLoS One 10, e0123848.

821

822 Torfs, P., Baggerman, G., Meeusen, T., Nieto, J., Nachman, R.J., Calderon, J., De Loof, A.,
823 Schoofs, L., 2002. Isolation, identification, and synthesis of a disulfated sulfakinin from the
824 central nervous system of an arthropods the white shrimp *Litopenaeus vannamei*. Biochem.
825 Biophys. Res. Commun. 299, 312-320.

826

827 Toullec, J.Y., Corre, E., Bernay, B., Thorne, M.A., Cascella, K., Ollivaux, C., Henry, J., Clark,
828 M.S., 2013. Transcriptome and peptidome characterisation of the main neuropeptides and
829 peptidic hormones of a euphausiid: the Ice Krill, *Euphausia crystallorophias*. PLoS One 8,
830 e71609.

831

832 Van Wielendaele, P., Badisco, L., Vanden Broeck, J., 2013. Neuropeptidergic regulation of
833 reproduction in insects. Gen. Comp. Endocrinol. 188, 23-34.

834

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

837

838 Veenstra, J.A., 2009. Allatostatin C and its paralog allatostatin double C: the arthropod
839 somatostatins. *Insect Biochem. Mol. Biol.* 39, 161-170.
840

841 Veenstra, J.A., Rombauts, S., Grbić, M., 2012. *In silico* cloning of genes encoding
842 neuropeptides, neurohormones and their putative G-protein coupled receptors in a spider mite.
843 *Insect Biochem. Mol. Biol.* 42, 277-295.
844

845 Ventura, T., Cummins, S.F., Fitzgibbon, Q., Battaglione, S., Elizur, A., 2014. Analysis of the
846 central nervous system transcriptome of the eastern rock lobster *Sagmariasus verreauxi* reveals
847 its putative neuropeptidome. *PLoS One* 9, e97323.
848

849 Verleyen, P., Huybrechts, J., Schoofs, L., 2009. SIFamide illustrates the rapid evolution in
850 arthropod neuropeptide research. *Gen. Comp. Endocrinol.* 162, 27-35.
851

852 von Reumont, B.M., Jenner, R.A., Wills, M.A., Dell'ampio, E., Pass, G., Ebersberger, I., Meyer,
853 B., Koenemann, S., Iliffe, T.M., Stamatakis, A., Niehuis, O., Meusemann, K., Misof, B., 2012.
854 Pancrustacean phylogeny in the light of new phylogenomic data: support for Remipedia as the
855 possible sister group of Hexapoda. *Mol. Biol. Evol.* 29, 1031-1045.
856

857 Xu, Z., Zhao, M., Li, X., Lu, Q., Li, Y., Ge, J., Pan, J., 2015. Transcriptome profiling of the
858 eyestalk of precocious juvenile Chinese mitten crab reveals putative neuropeptides and
859 differentially expressed genes. *Gene* In press.
860

Yan, X.C., Chen, Z.F., Sun, J., Matsumura, K., Wu, R.S., Qian, P.Y., 2012. Transcriptomic analysis of neuropeptides and peptide hormones in the barnacle *Balanus amphitrite*: evidence of roles in larval settlement. PLoS One 7, e46513.

Yasuda-Kamatani, Y., Yasuda, A., 2006. Characteristic expression patterns of allatostatin-like peptide, FMRFamide-related peptide, orcokinin, tachykinin-related peptide, and SIFamide in the olfactory system of crayfish *Procambarus clarkii*. J. Comp. Neurol. 496, 135-147.

Figure legends

Figure 1. Three examples of the *in silico* workflow used for the prediction of putative mature *Symphylella vulgaris* peptide structures. (A) Predicted processing scheme for prepro-adipokinetic hormone (AKH) I. The structure of the putative mature AKH isoform is shown in red, with those of two putative mature linker/precursor-related peptides shown in blue. In this panel, “pQ” indicates the pyroglutamic acid residue present in the putative mature AKH. (B) Predicted processing scheme for prepro-CCHamide I. The structure of the putative mature CCHamide isoform is shown in red, with those of three putative mature linker/precursor-related peptides shown in blue. In this panel, the inverted red bracket indicates a disulfide bond between the two cysteine residues in the putative mature CCHamide. (C) Predicted processing scheme for the extant portion of prepro-sulfakinin I. The structures of two putative mature sulfakinin isoforms are shown in red, with those of four putative mature linker/precursor-related peptides shown in blue. In this panel, “Y_(SO₃H)” indicates the presence of sulfated tyrosine residues in the two putative mature sulfakinins. “pQ” indicates the presence of pyroglutamic acid in one of the sulfakinins.

884

885 Figure 2. Putative *Symphylella vulgaris* pre/preprohormones deduced from transcriptome
886 shotgun assembly sequence data. **(A)** Prepro-adipokinetic hormone II. **(B1)** Prepro-adipokinetic
887 hormone-corazonin-like peptide (ACP) I. **(B2)** Prepro-ACP II. **(B3)** Prepro-ACP III. **(B4)**
888 Prepro-ACP IV. **(C1)** The carboxyl (C)-terminal portion of prepro-allatostatin A (AST-A) I.
889 **(C2)** The C-terminal portion of prepro-AST-A II. **(C3)** The C-terminal portion of prepro-AST-A
890 III. **(D)** Prepro-allatostatin C. **(E1)** Prepro-allatotropin I. **(E2)** The C-terminal portion of prepro-
891 allatotropin II. **(F)** Prepro-CCHamide II. **(G1)** Prepro-crustacean cardioactive peptide (CCAP)
892 I. **(G2)** Prepro-CCAP II. **(G3)** Prepro-CCAP III. **(G4)** Prepro-CCAP IV. **(G5)** Prepro-CCAP V.
893 **(H)** The amino (N)-terminal portion of prepro-GSEFLamide. **(I)** Prepro-insulin-like peptide. **(J)**
894 Prepro-intocin. **(K)** Prepro-proctolin. **(L)** The C-terminal portion of prepro-pyrokinin. **(M1)**
895 Prepro-short neuropeptide F (sNPF) I. **(M2)** Prepro-sNPF II. **(M3)** Prepro-sNPF III. **(N1)**
896 Prepro-SIFamide I. **(N2)** Prepro-SIFamide II. **(O)** The C-terminal portion of prepro-sulfakinin
897 II. In this figure, signal peptides are shown in gray, while all mono/dibasic cleavage loci are
898 shown in black. For each sequence, the isoform(s) of the peptide for which the precursor is
899 named is/are shown in red, with all linker/precursor related peptides shown in blue. The “+”
900 symbol indicate the presence of additional, unknown, amino acid residues at the N- or C-
901 terminus.

Table 1

Table 1. Putative <i>Symphylella vulgaris</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/red pigment concentrating hormone	GAKX01033834	539	70	F	38.9	9e-05
	GAKX01012891	579	70	F	38.9	1e-04
	GAKX01000634	451	64	C	38.9	7e-05
	GAKX01037082	348	56	F	29.3	0.19
Adipokinetic hormone-corazonin-like peptide	GAKX01010333	383	82	F	42.4	2e-06
	GAKX01032230	404	82	F	42.0	3e-06
	GAKX01040717	426	82	F	42.0	4e-06
	GAKX01055497	425	82	F	42.0	4e-06
	GAKX01018684	355	82	F	41.2	5e-06
	GAKX01009943	388	82	F	41.2	7e-06
	GAKX01062638	380	82	F	41.2	6e-06
	GAKX01023780	351	82	F	40.0	1e-05
Allatostatin A	GAKX01038863	1018	273	C	120	1e-29
	GAKX01008414	477	138	C	67.8	7e-13
	GAKX01019845	432	138	C	67.0	1e-12
	GAKX01063138	1214	128	C	55.1	4e-11
	GAKX01010441	468	136	C	67.0	1e-12
Allatostatin B						
Allatostatin C	GAKX01045392	501	106	F	94.0	1e-24
Allatotropin	GAKX01071027	552	113	F	32.7	0.052
	GAKX01071530	439	102	C	32.7	0.041
Bursicon α						
Bursicon β						
CCHamide	GAKX01004850	430	107	F	37.0	7e-04
	GAKX01071454	484	104	F	37.0	0.001
Corazonin						
Crustacean cardioactive peptide	GAKX01056315	486	129	F	42.7	1e-05
	GAKX01073013	479	129	F	42.4	1e-05
	GAKX01005072	546	129	F	42.7	2e-05
	GAKX01035926	539	129	F	42.4	2e-05
	GAKX01019483	593	129	F	42.7	2e-05
	GAKX01059057	459	126	F	42.7	9e-06
	GAKX01013544	450	98	F	42.4	1e-05
Crustacean hyperglycemic hormone/ion transport peptide						
DENamide						

Diuretic hormone 31						
Diuretic hormone 44						
DXXRLamide						
Ecdysis-triggering hormone						
Eclosion hormone						
FMRFamide-like peptide						
FXGGXamide						
GSEFLamide	GAKX01061676	649	169	N	57.4	5e-10
Insulin-like peptide	GAKX01045439	587	144	F	48.9	4e-07
Intocin	GAKX01066982	744	149	F	91.7	1e-22
Leucokinin						
Myosuppressin						
Neuroparsin						
Neuropeptide F						
Orcokinin						
Pigment dispersing hormone						
Proctolin	GAKX01012799	322	80	F	33.9	0.003
	GAKX01017321	354	80	F	33.9	0.004
	GAKX01043206	975	80	F	33.9	0.008
Pyrokinin	GAKX01015801	635	198	C	53.5	1e-09
RYamide						
Short neuropeptide F	GAKX01076964	432	98	F	48.9	3e-08
	GAKX01011699	533	98	F	47.8	9e-08
	GAKX01076393	488	98	F	50.4	7e-09
	GAKX01064368	487	96	F	48.9	3e-08
	GAKX01026236	540	96	F	48.1	8e-08
SIFamide	GAKX01063861	368	71	F	48.5	2e-08
	GAKX01043940	438	71	F	47.8	3e-08
	GAKX01054729	420	71	F	47.0	4e-08
	GAKX01035385	369	69	F	46.2	8e-08
Sulfakinin	GAKX01066370	376	112	C	53.5	1e-06
	GAKX01023815	568	97	C	52.0	5e-06
Tachykinin-related peptide						

*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/red pigment concentrating hormone (RPCH), *Cherax quadricarinatus* prepro-RPCH (**Accession No. AAV80404**;

Martínez-Pérez et al., 2007); adipokinetic hormone-corazonin-like peptide (ACP), *Speleonectes cf. tulumensis* (Christie, 2014a); allatostatin A (AST-A),

Procambarus clarkii prepro-AST-A (**Accession No. BAE45266**; Yasuda-Kamatani and Yasuda, 2006); allatostatin B (AST-B); *Litopenaeus vannamei* prepro-

AST-B (Christie, 2014f); allatostatin C (AST-C), *Homarus americanus* prepro-AST-C-like peptide (Dickinson et al., 2009a); allatotropin, *Tigriopus californicus* prepro-allatotropin (Christie, 2014b); bursicon α , *Homarus gammarus* pre-bursicon α (**Accession No. ADI86242**; Sharp et al., 2010); bursicon β , *H. americanus* pre-bursicon β (Christie et al., 2010b); CCHamide, *L. vannamei* prepro-CCHamide (Christie, 2014f); corazonin, *Daphnia pulex* prepro-corazonin (**Accession No. EFX86608**; Colbourne et al., 2011); crustacean cardioactive peptide (CCAP), *H. gammarus* prepro-CCAP (**Accession No. ABB46292**; Chung et al., 2006); crustacean hyperglycemic hormone (CHH)/ion transport peptide, *Cancer productus* prepro-CHH I (**Accession No. ABQ41269**; Hsu et al., 2008a); DENamide, *D. pulex* prepro-DENamide (Dirksen et al., 2011); diuretic hormone 31 (DH31), *H. americanus* prepro-calcitonin-like diuretic hormone (**Accession No. ACX46386**; Christie et al., 2010c); diuretic hormone 44 (DH44), *Argulus siamensis* prepro-DH44 (Christie, 2014c); DXXRLamide, *Tigriopus californicus* prepro-DXXRLamide Ia (Christie, 2014b); ecdysis-triggering hormone (ETH), *D. pulex* prepro-ETH (Dirksen et al., 2011); eclosion hormone (EH), *D. pulex* pre-EH (Dirksen et al., 2011); FMRFamide-like peptide; *P. clarkii* prepro-FLRFamide A (**Accession No. BAE06262**; Yasuda-Kamatani and Yasuda, 2006); FXGGXamide, *T. californicus* prepro-FXGGXamide Ia (Christie, 2014b); GSEFLamide, *T. californicus* prepro-GSEFLamide Ia (Christie, 2014b); insulin-like peptide (ILP), *D. pulex* prepro- Insulin/IGF/relaxin-like peptide 2 (**Accession No. EFX70023**; Colbourne et al., 2011); intocin, *L. vannamei* prepro-intocin (Christie, 2014f); leucokinin, *Eucyclops serrulatus* prepro-leucokinin (Christie, 2015a); myosuppressin, *H. americanus* prepro-myosuppressin (**Accession No. ACX46385**; Stevens et al., 2009); neuroparsin, *Jasus lalandii* pre-neuroparsin (**Accession No. AHG98659**; Marco et al., 2014); neuropeptide F (NPF), *L. vannamei* prepro-NPF I (**Accession No. AEC12204**; Christie et al., 2011d); orckinin, *H. americanus* prepro-orckinin I (**Accession No. ACB41787**; Dickinson et al., 2009b); pigment dispersing hormone (PDH), *C. productus* prepro- β -PDH I (**Accession No. ABV68725**; Hsu et al., 2008b); proctolin, *Penaeus monodon* prepro-proctolin (Christie, 2014f); pyrokinin, *L. vannamei* prepro-CAPA (Christie, 2014f); RYamide, *L. vannamei* prepro-RYamide (Christie, 2014f); short neuropeptide F (sNPF), *D. pulex* prepro-sNPF (**Accession No. EFX90018**; Colbourne et al., 2011); SIFamide, *H. americanus* prepro-Val¹-SIFamide (**Accession No. ABV21807**; Dickinson et al., 2008); sulfakinin, *H. americanus* prepro-sulfakinin (**Accession No. ABQ95346**; Dickinson et al., 2007); tachykinin-related peptide, *H. americanus* prepro-tachykinin (**Accession No. ACB41786**; Christie et al., 2008b).

Table 2

Table 2. Predicted peptidome of <i>Symphylla vulgaris</i>	
Peptide group	Predicted peptide structure
Adipokinetic hormone (AKH)/red pigment concentrating hormone (RPCH)	
AKH/RPCH linker/precursor-related peptide	pQVNFSPGwa SESPCSIYKETVMTIYALIQNEA ISDCNGNYS SESPCYHLQGDSDDY _(503u) LCSHSE
Adipokinetic hormone-corazonin-like peptide (ACP)	
ACP linker/precursor-related peptide	pQVTFSKDWKaa SQDGCSASSFALL LQRIVKMETERLEECGGS LANILDEIPRIE SQDGCSASLFALL
Allatostatin A (AST-A)	
AST-A linker/precursor-related peptide	+AFGLa AAKSYAFGLa PSNRVYSFGLa PNNRVYSFGLa AARAYNFGLa ARQNSYAFGLa VPMYSFGLa SASDRMYAFGLa SASDRMYSFGLa VVKSYAFGLa EMEDD EIDDMDED EIDDIEDD DNSDYSED EIEDD SNNEDSLTWDDDNFEDS AEDMYSDELNED TRSWDa EEGEES TSNETRTEAVSTTTSTATTKPTRS
Allatostatin C (AST-C)	
AST-C linker/precursor-related peptide	TYWKQCAFNAVSRFa SKGENERHGFVPEIGSDPVYQHQMIDDDGSLDTALINYL FARQMM LQTGADTLDLQ
Allatotropin	
Allatotropin linker/precursor-related peptide	GFKNPALSTARGFa RQQPNVPRVV TPLGTADELLTSELAARKPLDQDLVLCFYNTWCRLMLRQQGLMVPSYHNLTAGEPITSEERFWVFF RQQPNVPRAV TPLGTSDELLTSELAARKPLDQDLVLCFYNTWCRLMLRQQGLMVPSYHNLTAGEPITSEE
CCHamide	
CCHamide linker/precursor-related peptide	LKGCSFSGHSCFGAHa TDPSETDNSNNLLLSRLRSSLNQGSLSDSRIVGKEA PGLPYDLLSAWAASAAERAINSP IKPI TDPSETDNSNNLLLSRLRSSLNPA SLSDSRFAGKEP

PGLPYDLLSAWAASAAERSNQSQED	
Crustacean cardioactive peptide (CCAP)	
CCAP linker/precursor-related peptide	<p>PFCNAFTGCa</p> <p>ESERTEEFSGSDEFLMP</p> <p>SDTNY_(S03H)VDREDDV</p> <p>QRLFYPPQSERPRGWSDLQAFRTPTYFP</p> <p>DYLGTYGFSa</p> <p>STQHLEQQNQNRGSEQ</p> <p>QRLFYPPQSERPRGWSDLQAFRTPTYFPVLAVVPNSQYRDYLGTYGFSa</p> <p>GARST</p> <p>SDTNY_(S03H)VDREDDVDNGYSTLRNRNVQGDGPISRPEHLPTSRY</p>
GSEFLamide	
GSEFLamide linker/precursor-related peptide	<p>VGSEFIa</p> <p>LGSEFIa</p> <p>LGSEFLa</p> <p>LGSEFVa</p> <p>YDPPETRKLTSDHLPYSVYDLLRMRLRa</p> <p>SDAIDDLSESARSYLM</p> <p>NVDYYYGSPD</p> <p>SDSNEGPEEESSSWNES</p> <p>SVSDPKAAG+</p>
Insulin-like peptide (ILP)	
ILP linker/precursor-related peptide	<p>STGFKFCGARLVGILSMLCDSNYYDSTS</p> <p>SDSSTDSEWKHWYLNDKTPFTSYPAARAFLV</p> <p>DVGFYRTL RDGVHDECCKKSCSYQELLQYCATPDNTIIVD</p> <p>LRERDRTHPLF</p>
Intocin	
Intocin linker/precursor-related peptide	<p>CFITNCPIGa</p> <p>SGHSLSSSVALRQCSPCGPGGAGRCFGKSACCVPGMGCFLGPAHTATCRLEHSYPVPCDNGTPACGENGNRC AVEGLCCTGDGCFYEDTCRDSSLLTSSLHHS GSTLEDETTO</p>
Proctolin	
Proctolin linker/precursor-related peptide	<p>RYLPT</p> <p>GDNARKEQIKDLLRELLDMQPDSPPEME</p> <p>FLV</p> <p>EANLPGYSANAEDWN</p>
Pyrokinin	
Pyrokinin linker/precursor-related peptide	<p>pQGLIPFPRVa</p> <p>SNSFSPRLa</p> <p>AISAFSPRLa</p> <p>AMIPRLa</p> <p>AVAAFSPLa</p> <p>ANVAFSPRLa</p> <p>FHLYGEE</p> <p>SPYSHSSLSAADGSELDDNENEF LGYPLRYA</p> <p>SAENADDLEDISEVLEKSPWAILALKDSISGID</p> <p>SDEIAESAEMS DY_(S03H)Y</p> <p>AGGLQERSVAFSPRLGDRQ</p>
Short neuropeptide F (sNPF)	
sNPF linker/precursor-related peptide	<p>AGRDP SLRLRFa</p>

	<div>EFGTLAAPAYGDYEGKGIRDLYEMLLRNELFNERMGHQVV SDPLWQQYGPEETQSEGPASN EFGTLAAPAYGDYEGKGIRDLYEMLLRNELFNERMSHQVV EFGTLAAPAYGDYEGIRDLYEMLLRNELFNERMSHQVV</div>
SIFamide	<div>PPFNGSIFa</div>
SIFamide linker/precursor-related peptide	<div>GY APGELATDKLYAVCEIAVETCSQFFQGFESK APGELATDKLYAVCEIAVETCYPPIFPRL</div>
Sulfakinin	<div>pQSDDY_(SO3H)GHLRFa DFDDY_(SO3H)GHMRFa</div>
Sulfakinin linker/precursor-related peptide	<div>KPAPTNGNS QHRFTHLILPYEDQTQ SKLDLASLIALRESGMMRDGPGEDFEDADSGELREAS SAV KPAPTNGDNS</div>
<div>Abbreviations: a, amidated C-terminus; +, likely partial peptide with additional unknown amino acids; pQ, 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.</div>	

A. Prepro-adipokinetic hormone (from GAKX01033834, GAKX01012891 and GAKX01000634)

MRAVFGVGLFLGLMVILMTVNLAE**GQ**VNFSPGWGKRSESPCSIYKETVMTIYALIQNEARRISDCNGNYSR

↓ Signal peptidase (cleavage locus highlighted above)

QVNFSPGWGKR**SE**SPCSIYKETVMTIYALIQNEARR**I**SDCNGNYSR

↓ Prohormone convertase (cleavage loci highlighted above)

QVNFSPGWG**KR** SESPCSIYKETVMTIYALIQNEARR ISDCNGNYS**R**

↓ Carboxypeptidase (cleavage loci highlighted above)

QVNFSPGW**G** **SE**SPCSIYKETVMTIYALIQNEA ISDCNGNYS

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

QVNFSPGWamide

↓ Glutaminyl cyclase (cyclization locus highlighted above)

pQVNFSPGWamide

B. Prepro-CCHamide I (from GAKX01004850)

MKQCQIICVFYIILVVLFSATA**AL**KGCSSFHSCFGAHGKRTPSETDNSNNLLSRLRSSLNQGSLSDSRIVGKEARRPGLPYDLLSAWAASAAERA
INSPKKIKPI

↓ Signal peptidase (cleavage locus highlighted above)

LKGCSSFHSCFGAHG**KR**TPSETDNSNNLLSRLRSSLNQGSLSDSRIVGKEARR**P**GLPYDLLSAWAASAAERAINSP**KK**IKPI

↓ Prohormone convertase (cleavage loci highlighted above)

LKGCSSFHSCFGAHG**KR** TDPSETDNSNNLLSRLRSSLNQGSLSDSRIVGKEARR PGLPYDLLSAWAASAAERAINSP**KK** IKPI

↓ Carboxypeptidase (cleavage loci highlighted above)

LKGCSSFHSCFGAH**G** TDPSETDNSNNLLSRLRSSLNQGSLSDSRIVGKEA PGLPYDLLSAWAASAAERAINSP

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

LKG**C**SSFHSC**F**GHAHamide

↓ Enzymatic disulfide bond formation (bridged cysteines highlighted above)

LKGCSSFHSCFGAHamide

C. Prepro-sulfakinin I (C-terminal partial protein from GAKX01066370)

+LLVSVLLVVICGLKCL**LAK**PAPTNGNSKRQHRFTHLILPYEDQTQRRSKLDLASLIALRESGMMRDGPGEDFEDADSGELREASKRQSDDYGHLRFGK
RDFDDYGHMRFGRSAV

↓ Signal peptidase (cleavage locus highlighted above)

KPAPTNGNSKR**Q**HRFTHLILPYEDQT**QRR**SKLDLASLIALRESGMMRDGPGEDFEDADSGELREAS**RQ**SDDYGHLRFG**KRD**FDDYGHMRFG**RS**AV

↓ Prohormone convertase (cleavage loci highlighted above)

KPAPTNGNS**KR** QHRFTHLILPYEDQT**QRR** SKLDLASLIALRESGMMRDGPGEDFEDADSGELREAS**KR** QSDDYGHLRFG**KR** DFDDYGHMRFG**R**
SAV

↓ Carboxypeptidase (cleavage loci highlighted above)

KPAPTNGNS QHRFTHLILPYED**QTQ** SKLDLASLIALRESGMMRDGPGEDFEDADSGELREAS QSDDYGHLRFG DFDDYGHMRFG

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

QSDDY**Y**GHLRFamide DFDD**Y**GHMRFamide

↓ Tyrosylprotein sulfotransferase (tyrosine targets highlighted above)

QSDDY_(S03H)GHLRFamide **DFDDY**_(S03H)GHMRFamide

↓ Glutaminyl cyclase (cyclization locus highlighted above)

pQSDDY_(S03H)GHLRFamide

A. Prepro-adipokinetic hormone II (from GAKX01037082)

MRAVLVGLFLGLMVLMTVNLAEGQVNFSPGWGKRSESPCYHLQGDSDDYLCSSHSE

B1. Prepro-adipokinetic hormone-corazonin-like peptide (ACP) I (from GAKX01010333, GAKX01032230, GAKX01040717 and GAKX01055497)

MKNIVAILAVMIVLASFLPDRSNGQVTFSDKWKAGKRSQDGCSASSFALLRRLQRIVKMETERLEECGGSANILDEIPRIE

B2. Prepro-ACP II (from GAKX01018684 and GAKX01009943)

MKNIVAILAVMIVIASFLPDRSNGQVTFSDKWKAGKRSQDGCSASSFALLRRLQRIVKMETERLEECGGSANILDEIPRIE

B3. Prepro-ACP III (from GAKX01062638)

MKNLVAAILAVMIVIASFLPDRSNGQVTFSDKWKAGKRSQDGCSASSFALLRRLQRIVKMETERLEECGGSANILDEIPRIE

B4. Prepro-ACP IV (from GAKX01023780)

MKNLVAAILAVMIVIASFLPDRSNGQVTFSDKWKAGKRSQDGCSASSFALLRRLQRIVKMETERLEECGGSANILDEIPRIE

C1. Prepro-allatostatin A (AST-A) I (C-terminal partial protein from GAKX01038863 and GAKX01008414)

+AFGLGKREMEDDKRAAKSYAFGLGKREIDDMDEDKRAAKSYAFGLGKREIDDIEDDKRAAKSYAFGLGKRPSNRVYSFGLGKRDNDSYSEDKRAAKSYAFGLGKREIEDDKRAAKSYAFGLGKRPNNRVYSFGLGKRNNEDSLTWDDDNFEDSKRAARAYNFGGLGKRAEDMYSELDNEDKRAQNSYAFGLGKRTRSWDGKRVPMYSFGLGKRASDRMYAFGLGKREEGEESKRAAKSYAFGLGKRTRNETRTEAVSTTTSTATTKPETRS

C2. Prepro-AST-A II (C-terminal partial protein from GAKX01019845 and GAKX01063138)

+KRNNEDSLTWDDDNFEDSKRAARAYNFGGLGKRAEDMYSELDNEDKRAQNSYAFGLGKRTRSWDGKRVPMYSFGLGKRASDRMYAFGLGKREEGEESKRAAKSYAFGLGKRTRNETRTEAVSTTTSTATTKPETRS

C3. Prepro-AST-A III (C-terminal partial protein from GAKX01010441)

+SNNEDSLTWDDDNFEDSKRAARAYNFGGLGKRAEDMYSELDNEDKRAQNSYAFGLGKRTRSWDGKRVPMYSFGLGKRASDRMYAFGLGKREEGEESKRVVKSAYAFGLGKRTRNETRTEAVSTTTSTATTKPETRS

D. Prepro-allatostatin C (from GAKX01045392)

MVSSVVRISLLSLSLLLFCVVSRYRTEASKGENERHGFVPEIGSDPVYQHQMIDDDGSLDTALINYLFAQQMMKRLQTGADTLDLQKRRTYWKQCAFNAVSRFGKK

E1. Prepro-allatotropin I (from GAKX01071027)

MKLGVLVLCVLACLLVISESRQQPNVPRVVRGFKNPALSTARFGFKRTPLGTADELLTSELAARKPLDQDLVLCFYNTWCRLMLRQQGLMVPSYNHLNTAGEPITSEERFWVF

E2. Prepro-allatotropin II (C-terminal partial protein from GAKX01071530)

+VCVLACLLVISESRQQPNVPRVVRGFKNPALSTARFGFKRTPLGTSDELLTSELAARKPLDQDLVLCFYNTWCRLMLRQQGLMVPSYNHLNTAGEPITSEER

F. Prepro-CCHamide II (from GAKX01071454)

MKQCQIICVFYIILVLFSAATAALKGCSSFGHSCFGAHCRTDPSETDNSNNLLSRLRSSLPASLSDSRFAGKEPRRPGLPYDLLSAWAASAAERSNQSQED

G1. Prepro-crustacean cardioactive peptide (CCAP) I (from GAKX01056315, GAKX01073013 and GAKX01005072)

MYTISSTLISFTILCTILLYQVNAESERTEEFSGSDEPLMPKKRPFCNAFTGCGRKRSNTNYVDREDDVRQRLFYPPQSERPRGWSDLQAFRTPTYFPDYLGTYGFGSGRRKRSTQHLEQQNQNRGSEQ

G2. Prepro-CCAP II (from GAKX01035926)

MYTISSTLISFTILCTILLYQVNAESERTEEFSGSDEPLMPKKRPFCNAFTGCGRKRSNTNYVDREDDVRQRLFYPPQSERPRGWSDLQAFRTPTYFPDYLGTYGFGSGRRKRSTQHLEQQNQNRGSEQ

G3. Prepro-CCAP III (from GAKX01019483)

MYTISGTLISFTILCTILLYQVNAESERTEEFSGSDEPLMPKKRPFCNAFTGCGRKRSNTNYVDREDDVRQRLFYPPQSERPRGWSDLQAFRTPTYFPDYLGTYGFGSGRRKRSTQHLEQQNQNRGSEQ

G4. Prepro-CCAP IV (from GAKX01059057)

MYTISSTLISFTILCTILLYQVNAESERTEEFSGSDEPLMPKKRPFCNAFTGCGRKRSNTNYVDREDDVRQRLFYPPQSERPRGWSDLQAFRTPTYFPVLAVPNSQYRDYLGTYGFGSGRRKGARST

G5. Prepro-CCAP V (from GAKX01013544)

MYTISSTLISFTILCTILLYQVNAESERTEEFSGSDEPLMPKKRPFCNAFTGCGRKRSNTNYVDREDDVDNGYSTLRNRNVQGDGPISRPSHLPTSR

Y

H. Prepro-GSEFLamide (N-terminal partial protein from GAKX01061676)

MQRCGLLVLQWILGLMTLVMLSGPSEAYDPPETRKLTSDHLPYSVYDLLRMRLRGRRVGSEFIGKRLGSEFIGRKRSDAIDDLSESARSYLMKRLGSEFLGKRNVDDYYGSPDKKLGSEFVGKRLGSEFLGKRSDSNEGPEEESSSWNESKRLGSEFLGKRVSVDPAAG+

I. Prepro-insulin-like peptide (from GAKX01045439)

MRTCMSFVAIGVICTFALLIQQIGGLRERDRTHPLFRRSTGFKFCGARLVGILSMLCDSNYDYDSTSKRSDSSTDSEWEGKHWYLNDKTPFTSYPAARA
FLVKRDVGIFYRTL RDGVHDECCCKKSCSYQELLQYCATPDNTIIVD

J. Prepro-intocin (from GAKX01066982)

MSSGKSDITNAFVLLTFVSTVSA CFITNCPIGGKRSGHSLSSSVLRQCSPCGPGGAGRCFGKSACCVPGMGCF LGPAHTATCRLEHSYPVPCDNGT
PACGENGNRC AVEGLCCTGDGCFYEDTCRDSSLLTSSLHSGSTLEDETTQ

K. Prepro-proctolin (from GAKX01012799, GAKX01017321 and GAKX01043206)

MAVRNSFLLTVVVMVCFWVAQA RYLPTRGDNARKEQIKDLLRELLDMQPD SRPPEMEKRFLVKREANLPGYSANAEDWN

L. Prepro-pyrokinin (C-terminal partial protein from GAKX01015801)

+VCLSLVVLVVIGSCTGFHLYGEEKRQGLIPFPVRGSRSPYSHSSLSAADGSELDDNENEF LGYPLRYARSNSFS PRLGRKKR SAENADDLEDISEVL
EKSPWAILALKDSISGIDKKAISAFSPRLGRAMI PRLGKKR SDEIAESADEMSDYKRAVAAFSPRLGRANVAFSPRLGRAGGLQERSVAFSPRLGD
RQRRN

M1. Prepro-short neuropeptide F (sNPF) I (from GAKX01076964 and GAKX01011699)

MSSSILVKCCALTLCALLVLG EFGTLAAPAYGDYEGKGIRDLYEMLLRNELFNERM SHQVVRKAGRDP SLRLRFGRRS DPLWQQYGPEETQSEGPAS
N

M2. Prepro-sNPF II (from GAKX01076393)

MSSSILVKCCALTLCALLVLG EFGTLAAPAYGDYEGKGIRDLYEMLLRNELFNERM GHQVVRKAGRDP SLRLRFGRRS DPLWQQYGPEETQSEGPAS
N

M3. Prepro-sNPF III (from GAKX01064368 and GAKX01026236)

MSSSILVKCCALTLCALLVLG EFGTLAAPAYGDYEGGIRDLYEMLLRNELFNERM SHQVVRKAGRDP SLRLRFGRRS DPLWQQYGPEETQSEGPASN

N1. Prepro-SIFamide I (from GAKX01063861, GAKX01043940 and GAKX01054729)

MSCKTLLLVCLVTL SVMLLLETAEAGYKKPPFN GSIFGKRAPGELATDKLYAVCEIAVETCSQFFQGFESK

N2. Prepro-SIFamide II (from GAKX01035385)

MSCKTLLLVCLVTL SVMLLLKETAEAGYKKPPFN GSIFGKRAPGELATDKLYAVCEIAVETCYPIFPRL

O. Prepro-sulfakinin II (C-terminal partial protein from GAKX01023815)

+KPAPTNGDNSKRQHRFTHLILPYEDQTQRRSKLDLASLIALRESGMMRDGP GEDFEDADSGELREASKRQSDDYGH LRFGKRDFDDYGHMRFGRSA
V