Published in final edited form as:

Peptides. 2010 January ; 31(1): 27. doi:10.1016/j.peptides.2009.10.007.

Combining *in silico* transcriptome mining and biological mass spectrometry for neuropeptide discovery in the Pacific white shrimp *Litopenaeus vannamei*

Mingming Ma¹, Ashley L. Gard², Feng Xiang¹, Junhua Wang¹, Naveed Davoodian², Petra H. Lenz³, Spencer R. Malecha⁴, Andrew E. Christie^{2,3,*}, and Lingjun Li^{1,5,*}

¹ School of Pharmacy, University of Wisconsin, 777 Highland Avenue, Madison, Wisconsin 53705-2222 USA

² Center for Marine Functional Genomics, Mount Desert Island Biological Laboratory, P.O. Box 35, Old Bar Harbor Road, Salisbury Cove, Maine 04672 USA

³ Békésy Laboratory of Neurobiology, Pacific Biosciences Research Center, University of Hawaii at Manoa, 1993 East-West Road, Honolulu, Hawaii 96822 USA

⁴ Department of Human Nutrition, Food and Animal Science, College of Tropical Agriculture and Human Resources, University of Hawaii at Manoa, 1955 East West Road, Honolulu, Hawaii 96822 USA

⁵ Department of Chemistry, University of Wisconsin, 1101 University Avenue, Madison, Wisconsin 53706-1396 USA

Abstract

The shrimp *Litopenaeus vannamei* is arguably the most important aquacultured crustacean, being the subject of a multi-billion dollar industry worldwide. To extend our knowledge of peptidergic control in this species, we conducted an investigation combining transcriptomics and mass spectrometry to identify its neuropeptides. Specifically, *in silico* searches of the *L. vannamei* EST database were conducted to identify putative prepro-hormone-encoding transcripts, with the mature peptides contained within the deduced precursors predicted via online software programs and homology to known isoforms. MALDI-FT mass spectrometry was used to screen tissue fragments and extracts via accurate mass measurements for the predicted peptides, as well as for known ones from other species. ESI-Q-TOF tandem mass spectrometry was used to *de novo* sequence peptides from tissue extracts. In total 120 peptides were characterized using this combined approach, including 5 identified both by transcriptomics and by mass spectrometry (*e.g.* pQTFQYSRGWTamide, Arg⁷-corazonin, and pQDLDHVFLRFamide, a myosuppressin), 49 predicted via transcriptomics only (*e.g.* pQIRYHQCYFNPISCF and pQIRYHQCYFIPVSCF, two C-type allatostatins, and RYLPT, authentic proctolin), and 66 identified solely by mass spectrometry (*e.g.* the orcokinin NFDEIDRAGMGFA). While some of the characterized peptides were known *L. vannamei* isoforms (*e.g.* the pyrokinins DFAFSPRLamide and ADFAFNPRamide), most were novel, either for this

*Correspondence to either: Dr. Andrew E. Christie, Center for Marine Functional Genomics, Mount Desert Island Biological Laboratory, P.O. Box 35, Old Bar Harbor Road, Salisbury Cove, ME 04672. Phone: 207-288-9880 ext. 284; FAX: 207-288-2130; achristi@mdibl.org., Dr. Lingjun Li, School of Pharmacy, University of Wisconsin, 777 Highland Avenue, Madison, Wisconsin 53705-2222 USA; Phone: 608-265-8491; Fax: 608-262-5345; lli@pharmacy.wisc.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

species (*e.g.* pEGFYSQRYamide, an RYamide) or in general (*e.g.* the tachykinin-related peptides APAGFLGMRamide, APSGFNGMRamide and APSGFLDMRamide). Collectively, our data not only expand greatly the number of known *L. vannamei* neuropeptides, but also provide a foundation for future investigations of the physiological roles played by them in this commercially important species.

Keywords

functional genomics; expressed sequence tag (EST); matrix-assisted laser desorption/ionization Fourier transform mass spectrometry (MALDI-FTMS); electrospray ionization quadrupole time-of-flight tandem mass spectrometry (ESI-Q-TOF MS/MS)

1. Introduction

Over the past fifty years, aquaculture has become an increasingly important source of marine and freshwater species for human consumption. With respect to decapod crustaceans, large-scale commercial aquaculture is limited primarily to the penaeid shrimp, which now constitute a multi-billion dollar industry worldwide [42]. To bring shrimp to market size requires rearing them through multiple developmental stages under environmental conditions controlled through management [25]. However, in recent years, shrimp production per unit area has leveled off, due largely to the exhaustion of new management improvement options, which tend to contribute to increased yields only during the first few decades that an undomesticated species is maintained under culture; shrimp aquaculture began in the 1940s. To reverse this trend the animals themselves must be changed, *i.e.* domesticated, and/or their physiology manipulated, as has been done in terrestrial animal agriculture and in the aquaculture of some fish, *e.g.* the Atlantic salmon *Salmo salar* [35]. In this regard, management improvement of penaeid shrimp must be based on knowledge of the basic physiological processes that control important components of their aquaculture. Currently, the management of these components in penaeid species, such as reproductive maturation, mating, spawning, larval and adult growth, and disease resistance, is based on a very limited knowledge of the underlying endocrinological mechanisms that control them [21,36]. The purpose of the study presented here is to extend our knowledge of peptidergic control in one of the most important cultured penaeid shrimp, *Litopenaeus vannamei*, in order to contribute to improvement in its aquaculture, as well as improvement in the rearing of other cultured penaeid species.

In decapod crustaceans, as in most animals, locally-released paracrine and circulating hormones contribute critically to physiological control systems. While a variety of substances can function as paracrine/hormones in any given species, peptides constitute by far the largest single class of signaling agents present in most multicellular organisms [39]. In higher animals, including the decapods, the nervous system is a major source of locally-released peptide paracrine and circulating peptide hormones [39].

While much work has focused on elucidating the neuropeptide complement of several decapod crustaceans, *e.g.* the American lobster *Homarus americanus* [48] and the green crab *Carcinus maenas* [47], no large-scale study has yet been conducted on any penaeid species. Here, we have undertaken such an investigation, focusing on elucidating the neuropeptidome of the Pacific white shrimp *L. vannamei*, the single most important farmed penaeid worldwide [42], using a combination of transcriptomics and mass spectrometry. Specifically, the publicly accessible expressed sequence tags (ESTs) for *L. vannamei* were searched for orthologs of known neuropeptide-encoding precursors using previously identified arthropod sequences as queries; the mature peptides encoded within the deduced proteins were predicted via a combination of on-line software programs and homology to known peptide isoforms. Matrix-

assisted laser desorption/ionization Fourier transform mass spectrometry (MALDI-FTMS)-based high resolution mass profiling was used subsequently to screen tissue fragments or tissue extracts via accurate mass measurements for the predicted peptides, as well as other known ones, while nanoscale biochemical separation/derivatization coupled to electrospray ionization quadrupole time-of-flight tandem mass spectrometry (ESI-Q-TOF MS/MS) was used to *de novo* sequence both known and novel peptides from tissue extracts. In total, 120 peptides were characterized using this combined approach (49 identified by transcriptomics only, 66 identified by mass spectrometry only, and 5 identified by both methodologies), with the vast majority (approximately 98%) being new to this species. Collectively, our data not only expand greatly the catalog of known *L. vannamei* peptide paracrine/hormones, but also provide a foundation for future functional studies to improve commercial rearing technologies.

2. Materials and methods

2.1. Animals

Pacific white shrimp, *L. vannamei*, were purchased from Island Aquaculture (Kaneohe, HI) and were maintained in aerated tanks of 24 ppt seawater at a temperature of approximately 22 °C.

2.2. Peptide prediction via in silico analyses

2.2.1. Database searches—Database searches were conducted using methods modified from several recent publications [7,8,9,15,¹⁸,33,47,48,72]. Specifically, the online program tblastn (National Center for Biotechnology Information [NCBI], Bethesda, MD; <http://www.ncbi.nlm.nih.gov/BLAST/>) was used to mine for ESTs encoding putative *L. vannamei* peptide precursors via queries using known arthropod prepro-hormone sequences. For all searches, the default settings of the program were used, with the exceptions that the database searched was set to non-human, non-mouse ESTs (*i.e.* EST_others) and was restricted to *L. vannamei* transcripts (*i.e.* taxid:6685). All hits were fully translated (see Section 2.2.2) and checked manually for homology to the target query, as well as for typical peptide precursor features, including start and stop codons (*i.e.* a full-length prepro-hormone), the presence of a signal sequence and pro-hormone convertase processing sites. For each of the putative neuropeptide-encoding transcripts identified, the BLAST score and BLAST-generated E-value for significant alignment are provided in Table 1.

2.2.2. Prediction of mature peptide structures—Prediction of the structures of the peptides encoded by the transcripts identified in Section 2.2.1 was accomplished via previously established procedures [7,8,9,15,¹⁸,33,47,48,72]. Specifically, translation of the nucleotide sequences of ESTs was performed using the Translate tool of ExPASy (Swiss Institute of Bioinformatics, Basel, Switzerland; <http://www.expasy.ch/tools/dna.html>). Signal peptide prediction was done via the online program SignalP 3.0, using both the Neural Networks and the Hidden Markov Models algorithms (Center for Biological Sequence Analysis, Technical University of Denmark, Lyngby, Denmark; <http://www.cbs.dtu.dk/services/SignalP/>) [2]. Pro-hormone convertase cleavage sites were predicted based on the information presented in Veenstra [84], as well as on homology to known pro-hormone processing schemes. Prediction of the sulfation state of Tyr residues was done using the online program Sulfinator (Swiss Institute of Bioinformatics; <http://www.expasy.org/tools/sulfinator/>) [54]. Where applicable, other post-translational modifications, *e.g.* cyclization of amino (N)-terminal Gln/Glu residues, disulfide bridging between Cys residues, and carboxyl (C)-terminal amidation at Gly residues, were predicted by homology to known peptide isoforms.

2.3. Mass spectral characterization of peptide complement

2.3.1. Tissue collection—The major regions of the *L. vannamei* CNS (*i.e.* the eyestalk ganglia [including the sinus gland], the supraoesophageal ganglion [brain] and the ventral nerve cord) were isolated by manual micro-dissection and immediately placed in acidified methanol (90% methanol [Fisher Scientific, Pittsburgh, PA]; 9% glacial acetic acid [Fisher]; 1% deionized water) and stored at -80°C until utilized for peptide extraction or direct tissue mass spectral analysis (see Section 2.3.2). Most of the collected tissue was pooled by CNS region for peptide extraction; some individual tissues were also obtained and stored as single samples for direct tissue mass spectral analysis.

2.3.2. Mass spectral analyses

2.3.2.1. Tissue extraction and HPLC fractionation: Pooled tissues were homogenized and extracted with acidified methanol (see Section 2.3.1). Extracts were dried in a Savant SC 110 SpeedVac concentrator (Thermo Electron Corporation, West Palm Beach, FL) and re-suspended in approximately 100 μl of 0.1% formic acid. The re-suspended extracts were then vortexed and briefly centrifuged, with the resulting supernatants subsequently fractionated via high performance liquid chromatography (HPLC).

HPLC separations were performed using a Rainin Dynamax HPLC system, which was equipped with a Dynamax UV-D II absorbance detector (Rainin Instrument Inc., Woburn, MA). The mobile phases used for chromatographic separation were: deionized water containing 0.1% formic acid (mobile phase A), and acetonitrile (HPLC grade, Fisher Scientific) containing 0.1% formic acid (mobile phase B). For each separation run, 20 μl of extract was injected onto a Macrosphere C18 column (2.1 mm i.d. $\times$ 250 mm length, 5 μm particle size; Alltech Assoc. Inc., Deerfield, IL). The separation consisted of a 120-minute gradient of 5%–95% mobile phase B with fractions automatically collected every two minutes using a Rainin Dynamax FC-4 fraction collector.

2.3.2.2. MALDI-FTMS: MALDI-FTMS experiments were performed on an IonSpec ProMALDI Fourier transform mass spectrometer (Lake Forest, CA) equipped with a 7.0 Tesla actively-shielded superconducting magnet. This FTMS instrument contains a high pressure MALDI source where the ions from multiple laser shots can be accumulated in the external hexapole storage trap before being transferred to the ICR cell via a quadrupole ion guide. A 337 nm nitrogen laser (Laser Science, Inc., Franklin, MA) was used for ionization/desorption. The ions were excited prior to detection with a radio frequency sweep beginning at 7050 ms with a width of 4 ms and amplitude of 150 V base to peak. The filament and quadrupole trapping plates were initialized to 15 V, and both were ramped to 1 V from 6500 to 7000 ms to reduce baseline distortion of peaks. Detection was performed in broadband mode from m/z 108.00 to 4500.00.

For direct tissue analysis, tissue fragments were desalted by briefly rinsing in a solution of DHB prepared in deionized water (10 mg/ml). The tissue was then placed onto the MALDI sample plate along with 0.3 μl of saturated DHB matrix (prepared as described in Section 2.3.2.2), which was subsequently allowed to crystallize at room temperature (approximately 22°C).

Off-line analysis of HPLC fractions (prepared as described in Section 2.3.2.1) was performed by spotting 0.3 μl of saturated DHB on the MALDI sample plate and adding 0.3 μl of the HPLC fraction of interest. The resulting mixture was allowed to crystallize at room temperature, with subsequent MALDI-FTMS analysis performed as described above.

2.3.2.3. Capillary LC-ESI-Q-TOF MS/MS: Nanoscale liquid chromatography (LC)-ESI-Q-TOF MS/MS was performed using a Waters capillary LC system coupled to a Q-TOF Micro mass spectrometer (Waters Corp., Milford, MA). Chromatographic separations were performed on a C18 reverse phase capillary column (75 μ m internal diameter $\times$ 150 mm length, 3 μ m particle size; Micro-Tech Scientific Inc., Vista, CA). The mobile phases used were: deionized water with 5% acetonitrile and 0.1% formic acid (mobile phase A), acetonitrile with 5% deionized water and 0.1% formic acid (mobile phase B), and deionized water with 0.1% formic acid (mobile phase C). An aliquot of 6.0 μ l of an HPLC fraction (see Section 2.3.2.1) was injected and loaded onto the trap column (PepMapTM C18; 300 μ m column internal diameter $\times$ 1 mm, 5 μ m particle size; LC Packings, Sunnyvale, CA) using mobile phase C at a flow rate of 30 μ l/min for 3 minutes. Following injection, the stream select module was switched to a position at which the trap column became in line with the analytical capillary column, and a linear gradient of mobile phases A and B was initiated. A splitter was added between the mobile phase mixer and the stream select module to reduce the flow rate from 15 μ l/min to 200 nl/min.

The nanoflow ESI source conditions were set as follows: capillary voltage 3200 V, sample cone voltage 35 V, extraction cone voltage 1 V, source temperature 120°C, cone gas (N₂) 10 l/hr. A data-dependent acquisition was employed for the MS survey scan and the selection of precursor ions and subsequent MS/MS of the selected parent ions. The MS scan range was from m/z 300–2000 and the MS/MS scan was from m/z 50–1800. The MS/MS *de novo* sequencing was performed with a combination of manual sequencing and automatic sequencing by PepSeq software (Waters Corp.).

2.4. Figure production

For all figures illustrating prepro-hormone alignments (Figs. 1–3), the amino acid sequences of the precursors were aligned using the online program MAFFT version 6 (<http://align.bmr.kyushu-u.ak.jp/mafft/online/server/>), with the resulting alignments copied into and colored using Microsoft Word 2004 (Microsoft Corporation, Redmond, WA). All MS/MS figures (Figs. 4A and 5–7) were produced using a combination of Fireworks MX 2004 and Microsoft Windows paint tool. The MALDI-FTMS figure (Fig. 4B) was produced by converting the spectrum obtained using IonSpec version 7.0 (IonSpec Corp.) into a bitmap image using Boston University Data Analysis (BUDA) software (version 1.4; Boston University, Boston, MA). The BUDA file was then pasted into Fireworks MX 2004 (Macromedia, Inc., San Francisco, CA) and resampled to improve its resolution.

3. Results

To identify the maximum number of *L. vannamei* peptides possible, we employed an approach combining *in silico* transcriptome mining, with subsequent bioinformatic prediction of the peptides encoded within the deduced proteins, and mass spectrometry. For our transcriptome mining, the sequences of known insect and crustacean peptide precursors were used to query the NCBI *L. vannamei* EST database (on or before October 2, 2008) for putative peptide-encoding transcripts. Twenty-nine known arthropod peptide families/subfamilies were queried for, with ESTs putatively encoding members of 10 of the 29 target groups identified (Figs. 1–3 and Tables 1–2). In the interest of space, only those searches that identified putative precursors are described here. For mass spectral elucidation of the neuropeptides present in the *L. vannamei* CNS (Table 3), we used a strategy combining MALDI-FTMS-based high resolution mass profiling, both direct tissue and off-line HPLC fraction analyses (Fig. 4B), and nanoscale biochemical separation coupled to ESI-Q-TOF MS/MS *de novo* sequencing (Figs. 4A and 5–7). Whenever possible, we have grouped the identified peptides into families of related isoforms (Tables 2–3), and these are presented below in alphabetical order based on

family name. Unless otherwise noted, all of the peptides described here are novel, either for *L. vannamei* or in a general sense.

3.1. A-type allatostatin

Members of the A-type allatostatin (A-AST) family are characterized by the presence of the C-terminal motif –YXFGLamide, where X is a variable amino acid [71]. Prior to our study, a large number of A-ASTs had been identified from a variety of decapod crustaceans [20,22,23,29,³⁷,47,48,92,93]. No A-AST precursors were identified via transcriptome mining. However, in our mass spectral analyses, four A-type peptides, HGSYAFGLamide, ANQYAFGLamide, DRLYAFGLamide and SSKPYAFGLamide, were sequenced via ESI-Q-TOF MS/MS from the brain of *L. vannamei* (Table 3). Of these isoforms, ANQYAFGLamide and DRLYAFGLamide were described previously from the shrimp *Penaeus monodon* [23].

3.2. B-type allatostatin

Members of the B-type allatostatin (B-AST) family exhibit the characteristic C-terminal motif –WX₆Wamide, X₆ indicating six variable residues [71]. Recently, numerous B-AST isoforms have been identified from members of the Decapoda [7,29,31,47,48]. While no B-type precursors were identified via transcriptome mining, eight isoforms of B-AST were identified via a combination of MALDI-FTMS and ESI-Q-TOF MS/MS (Table 3). Six of the eight peptides, KWAAGRSWamide, RWSKFQGSWamide, ADWNKFQGSWamide, LTWNKFQGSWamide, SADWNSLRGTWamide and STNWSNLRGTWamide, were sequenced/detected only from the ventral nerve cord (Table 3), while one isoform, VPNDWAHFRGSWamide, was identified only from the brain (Table 3). One peptide, NWNKFQGSWamide, was sequenced/detected from the eyestalk ganglia, brain and ventral nerve cord (Table 3). Of the identified B-type peptides, NWNKFQGSWamide and VPNDWAHFRGSWamide are known *Cancer* crabs B-ASTs [29,31] and ADWNKFQGSWamide was previously predicted from the shrimp *Marsupenaeus japonicus* [7].

3.3. C-type allatostatin

The C-type allatostatins (C-ASTs) are a family of pentadecapeptides characterized by a pyroglutamine blocked N-terminus, an unamidated –PISCF C-terminus, and a disulfide bridge between two internal Cys residues [71]. In our study, four *L. vannamei* ESTs (Table 1) were identified as encoding putative C-AST precursors via a query using the sequence of a fruit fly *Drosophila melanogaster* prepro-C-AST (accession no. [AAK40100](#)) [89]. These ESTs were identified as encoding putative C-type precursors in two previous studies [49,72]; for the ease of later discussion, they are re-described here. Translation of ESTs **FE182974** and **FE175093** revealed each to encode a 139 amino acid, putative full-length prepro-hormone, which differed only in a Lys vs Phe residue at position 106 (Fig. 1A). Translation of **FE182975** and **FE180026** revealed similar, though not identical, putative C-terminal partial pro-hormones of 26 and 18 amino acids, respectively. These two partial pro-hormones differed from the full-length precursor predicted from **FE182974** at position 114 (Asp vs Gly), and positions 131 (Ile vs Asn) and 133 (Val vs Ile), respectively (numbering based on the sequence of the full-length prepro-hormone; Fig. 1A). Bioinformatic prediction of the peptides encoded by the deduced precursor proteins suggests that a maximum of six peptides are produced via post-translational processing (Table 2), including one isoform each of C-AST: pQIRYHQCYFNPISCF from **FE182974**, **FE175093** and **FE182975** or pQIRYHQCYFIPVSCF from **FE180026** (disulfide bridging between the two Cys residues in both peptides).

Neither of the putative C-ASTs nor any of the C-AST precursor-related peptides were identified via mass spectral analyses.

3.4. Bursicon α

The melanization and sclerotisation of the cuticle in newly ecdysed insects is controlled by bursicon, a heterodimeric cysteine knot protein comprised of α and β (see Section 3.5) subunit peptides [46,52]. In our study, five ESTs (Table 1) were identified as encoding putative bursicon α precursors via a query using the sequence of a green crab *C. maenas* prepro-bursicon α (accession no. [ABX55995](#)) [87]. Translation of these transcripts revealed [FE187805](#), [FE175634](#) and [FE173462](#) to each encode a 142 amino acid, putative full-length prepro-hormone, with [FE173463](#) and [FE173025](#) encoding 47 and 57 amino acid, putative partial C- and N-terminal precursors, respectively. Comparisons of the deduced protein sequences showed that the prepro-hormones derived from [FE187805](#) and [FE175634](#) varied at a single residue, Thr⁸³ vs Ala⁸³ (Fig. 1B). Likewise, the precursor deduced from [FE173462](#) differed from that of [FE187805](#) at one residue, an Asp for Gly substitution at position 36 (Fig. 1B). The partial protein deduced from [FE173463](#) was similar to the corresponding C-termini of the full-length precursors, with the exceptions of Phe for Leu and Pro for Ala substitutions at positions 100 and 126, respectively (Fig. 1B). Similarly, the partial sequence of [FE173025](#) showed extensive conservation with the corresponding N-terminus of the full-length precursors deduced from [FE187805](#) and [FE175634](#), differing only in Arg for Leu, Gln for Glu, Cys for Gly, and Gly for Cys substitutions at positions 8, 22, 49 and 51, respectively (Fig. 1B). Putative post-translational processing of the deduced precursors is predicted to produce a single bursicon α isoform from each protein (Table 2):

DECSLTPVIHILSYPGCNSKPIPSFACQGRCTSYVQVSGSKIWQTERSCMCCQESGER
EASVTLNCPKARPGEPRMRKILTRAPIDCMCRPCTDVEEGTVLAQEIANFIEDSPME
NVPFL from [FE187805](#),
DECSLTPVIHILSYPGCNSKPIPSFACQGRCTSYVQVSGSKIWQTERSCMCCQESGER
EASVALNCPKARPGEPRMRKILTRAPIDCMCRPCTDVEEGTVLAQEIANFIEDSPME
NVPFL from [FE175634](#), and
DECSLTPVIHILSYPDCNSKPIPSFACQGRCTSYVQVSGSKIWQTERSCMCCQESGER
EASVTLNCPKARPGEPRMRKILTRAPIDCMCRPCTDVEEGTVLAQEIANFIEDSPME
NVPFL from [FE173462](#). The partial C- and N-terminal peptides -
MRKIFTRAPIDCMCRPCTDVEEGTVLAQEIPNFIEDSPMENVPFL and
DQCSLTPVIHILSYPGCNSKPIPSFACQCRGTSYVQV- are predicted from [FE173463](#) and
[FE173025](#), respectively.

None of the putative bursicon α isoforms were identified via mass spectral analyses, though this is not surprising, given that the peptides are too large to be fully sequenced using the mass spectral methods used here.

3.5. Bursicon β

In addition to the transcripts encoding bursicon α isoforms, three ESTs (Table 1) were identified as encoding putative bursicon β precursors via a query using the sequence of a *C. maenas* prepro-bursicon β (accession no. [ABX55996](#)) [87]. Translation of these transcripts revealed each to encode a putative partial precursor (Fig. 1C), with that derived from [FE178442](#) being a 136 amino N-terminal partial protein (a start but not a stop codon present), and those deduced from [FE184710](#) and [FE184711](#) being 58 and 42 amino acid, C-terminal partial peptides, respectively (stop but not start codons evident). Comparisons of the overlapping portion of the three sequences revealed a number of variable residues (Fig. 1C). Putative post-translational processing of the deduced proteins suggests that a single bursicon β isoform is produced from each of the proteins, all identified here being partial sequences (Table 2):

GPSRAHTYGSECETLPSTMHVAKEEFDDAGRLVRTCEEDLAVNKCEGACVSKVQPS

VNTPSGFLKDCRCCRETHLRARDVVLTHCCDGDGNRITGDNGKLTVKLREPAELP–
from [FE178442](#), –
FLKDCRCCRETHLRARDVVLTHCYDGDGNRITGDNGKLTVKLREPADCQCFKCGN
SI from [FE184710](#), and –DVGLTPCYDGDGNRIPGDNGKLPVKLREPGDCQCFKCGNSI
from [FE184711](#).

Like the predicted bursicon α subunits, none of the putative bursicon β isoforms were identified via mass spectral analyses, though, again, this is not surprising, given that the peptides are large, and thus could not be fully sequenced using the mass spectral platforms employed in our study.

3.6. Corazonin

The peptide pQTFQYSRGWTSamide was originally identified from the cockroach *Periplaneta americana*, where it was found to be cardioactive [82]. Given its bioactivity, this peptide was named corazonin [82]. Subsequently, several structurally-related peptides were identified from members of other arthropods, though the Arg⁷ variant appears to be the most broadly conserved of the corazonin isoforms [7,43,47,48,61]. In crustaceans, Arg⁷-corazonin has been characterized via mass spectrometry from the crabs *C. borealis* and *C. maenas*, as well as from the lobster *H. americanus* [43,47,48]. Here, two ESTs (Table 1) were identified as encoding putative corazonin precursors via a query using the sequence of a *D. melanogaster* prepro-corazonin (accession no. [Q26377](#)) [78]. Translation of these transcripts revealed each to encode a 112 amino acid, putative full-length prepro-hormone (Fig. 1D). The two deduced precursors were identical with the exception of residue 57, which in [FE154856](#) was Leu and in [FE154857](#) was Phe (Fig. 1D). Predicted processing of the deduced prepro-hormones suggests that two peptides are produced from each protein (Table 2), including pQTFQYSRGWTSamide, which is identical in structure to Arg⁷-corazonin.

Via ESI-Q-TOF MS/MS, Arg⁷-corazonin was sequenced from the brain of *L. vannamei* (Table 3). This peptide was also detected via MALDI-FTMS in both the brain and ventral nerve cord (Table 3).

3.7. Crustacean cardioactive peptide

The peptide PFCNAFTGCamide (disulfide bridging between the two Cys residues) was originally isolated from the crab *C. maenas*, where it was shown to be a potent modulator of the heart, and thus named crustacean cardioactive peptide or CCAP [68]. This peptide has subsequently been identified from a variety of other decapod species [e.g. 13]. In our study, 13 ESTs (Table 1) were identified as encoding putative CCAP precursors via a query using the sequence of a *C. maenas* prepro-CCAP (accession no. [ABB46291](#)) [13]. Translation of ESTs [FE187476](#), [FE174994](#), [FE179552](#), [FE190948](#), [FE177552](#) and [FE189608](#) revealed each to encode a putative full-length precursor of 139 amino acids, with ESTs [FE173081](#), [FE175049](#) and [FE187520](#) each encoding an N-terminal partial protein of 142, 132 and 71 amino acids, respectively, and ESTs [FE177553](#) and [FE180576](#) encoding C-terminal partial sequences of 118 and 65 amino acids, respectively (Fig. 2A). Translation of ESTs [FE189609](#) and [FE175050](#) revealed each to encode a putative internal fragment of a precursor, with that derived from [FE189609](#) being 76 amino acids in length and that of [FE175050](#) being 59 amino acids long (Fig. 2A). As shown in Figure 2A, comparisons of the overlapping portion of the thirteen sequences revealed numerous variable residues. Putative post-translational processing of the deduced precursors suggests that a maximum of six peptides are liberated from each of the proteins (Table 2), including the well-known [e.g.,¹³,17,33,48,68] CCAP isoform PFCNAFTGCamide (disulfide bridging predicted between the two Cys) from all the full-length and N-terminal partial precursors except [FE177553](#) and [FE189609](#); PFCNAFTPGCamide is predicted from [FE177553](#) and DIADLLDGKDKSPFCKAFPGFamide

from **FE189609**, disulfide bridging is predicted between the two Cys residues in the former peptide.

Using ESI-Q-TOF MS/MS, PFCNAFTGCamide was sequenced from the brain of *L. vannamei* (Table 3). This peptide, including a disulfide bridge present between the two Cys residues, was also detected via MALDI-FTMS in both the brain and the ventral nerve cord (Table 3). In addition, we also sequenced the predicted CCAP precursor-related peptide DIADLLDGKD (see Table 2) from both the brain and the ventral nerve cord via ESI-Q-TOF MS/MS (Table 3).

3.8. Crustacean hyperglycemic hormone

The crustacean hyperglycemic hormone (CHH) family is a group of structurally-related, large peptides, which, among other functions, plays a role in the regulation of hemolymph glucose levels [24]. Members of the CHH superfamily can be divided into two subfamilies based on the presence or absence of a precursor-related peptide within their prepro-hormone [24]; the CHH subfamily possesses a precursor-related peptide, whereas it is absent in members of the moult-inhibiting hormone (MIH) subgroup [24]. Here, two ESTs (Table 1) were identified as encoding putative CHH/ion transport peptide (ITP) precursors via a query using the sequence of a *L. vannamei* prepro-ITP (accession no. **ABN11282**) [77]. Translation of these transcripts revealed each to encode a 142 amino acid, putative full-length prepro-hormone (Fig. 2B).

Comparison of the two deduced sequences revealed them to be nearly identical, differing only at position 20 (located within the signal peptide; Fig. 2B), which was Val in **FE101547** and Ala in **FE057303** (Fig. 2B). The predicted post-translational processing of the deduced proteins suggests that two peptides are cleaved from each precursor (Table 2):

RSVDGVGRLEKLLSSSSSSSGSSSPLDALGGDHSV and
DTFDHSCCKGIYDRELFRKLDRCEDCYNLYRKPYVATECKSNKYANFVFKQCLDD
LLMVDAIDEYVNTVQLVamide, the former a putative isoform of CHH precursor-related peptide (CPRP) and the latter a putative CHH/ITP isoform.

The predicted CHH/ITP isoform is identical in sequence to that of *L. vannamei* MIH 2 (accession no. **AA86057**) identified previously by Lago-Lestón and colleagues [41], with the exception of the C-terminal post-translational modification predicted here, *i.e.*

carboxypeptidase cleavage of the C-terminal Lys and α -amidation at a subsequently exposed Gly residue. Likewise, the CPRP isoform predicted in our study is identical to that of a C-terminal partial peptide encoded with *L. vannamei* MIH 2 [41], the latter missing residues 1–9 in our predicted peptide.

As with the bursicon α and β isoforms discussed earlier, neither the predicted CHH nor the predicted CPRP peptides were identified via mass spectral analyses. For the isoform of CHH this is expected, as the peptide is too large to be fully sequenced using the mass spectral platforms employed here. It should be noted that full sequence derivation of putative CHH peptides can be achieved by tandem MS following trypsin cleavage of a full-length isoform. However, this procedure requires a substantially larger amount of starting material for the purification of the peptide than was available for our study, thus it was not pursued here.

3.9. FMRamide-related peptide

The FMRamide-related peptides (FaRPs) are a large and diverse family of peptides found in both invertebrates and vertebrates [94]. In arthropods, a number of distinct subfamilies have been identified, including the myosuppressins, neuropeptide Fs, short neuropeptide Fs (sNPFs), and sulfakinins [*e.g.* 4,32,56]. In our study, members of the myosuppressin and sNPF subfamilies were identified, as were several FaRPs possessing –FLRFamide, –YLRFamide or –FVRamide C-termini; these peptide subgroups are discussed in turn below.

3.9.1. Myosuppressin—Members of the myosuppressin subfamily of the FaRPs are characterized by the presence of the C-terminal motif –HVFLRFamide [86]. In decapod crustaceans, the myosuppressin isoform pQDL DHVFLRFamide has been shown to be broadly, and perhaps ubiquitously, conserved [74]. In addition, the putative precursor of pQDL DHVFLRFamide, *i.e.* QDL DHVFLRFamide, has been characterized by mass spectrometry from several species [47,48]. In our study, one EST (Table 1) was identified as encoding a putative myosuppressin precursor via a query using the sequence of a *D. melanogaster* prepro-myosuppressin (accession no. **P61849**). Translation of this EST, **FE188748**, revealed it to encode a 108 amino acid, putative C-terminal partial precursor protein (a stop, but no start codon present; Fig. 3A). Given the lack of a start codon, it is impossible to predict where the signal peptide cleavage locus is located within this sequence, if, in fact, it is present in the portion of the precursor deduced from the transcript. However, post-translational processing of this partial protein suggests that at least three peptides are cleaved from prepro-hormone (Table 2), including pQDL DHVFLRFamide, which has a predicted structure identical to that of the known, mature decapod isoform of myosuppressin [74].

Via ESI-Q-TOF MS/MS both pQDL DHVFLRFamide and QDL DHVFLRFamide were sequenced from the brain and ventral nerve cord of *L. vannamei* (Table 3). In addition, both peptides were detected by MALDI-FTMS from the ventral nerve cord (Table 3).

3.9.2. Short neuropeptide F—FaRPs possessing the C-terminal motif –RXRFamide, where X represents a variable residue, most commonly Leu, are classified as members of the sNPF subfamily [32]. A number of sNPF isoforms have been identified from decapods [37, 47,48,64]. While, no sNPF-encoding transcripts were identified by transcriptome mining, five peptides possessing the C-terminal motif –RLRFamide were characterized via ESI-Q-TOF MS/MS and/or MALDI-FTMS (Table 3). Specifically, SMPSLRLRFamide, PSLRLRFamide, SM(O)PSLRLRFamide, M(O) representing an oxidized Met residue, and DGRTPALRLRFamide were identified in the ventral nerve cord, while PSMRLRFamide was sequenced/detected from the eyestalk ganglia, brain and ventral nerve cord; each of the *L. vannamei* sNPFs has been identified previously from at least one other decapod species [37, 47,48,64].

3.9.3. Other FaRPs—In addition to the subfamilies named above, a number of additional FaRPs have been described, including many isoforms possessing either –FLRFamide or –YLRamide C-termini [*e.g.* ^{14,29,47,48,53,80}]. While no *L. vannamei* transcripts encoding members of these FaRP groups were identified, a combination of ESI-Q-TOF MS/MS and MALDI-FTMS did result in the characterization of four N-terminally extended FLRFamides (Table 3): NRNFLRFamide, DGRNFLRFamide and APERNFLRFamide from both the brain and ventral nerve cord and SENRNFLRFamide from both the eyestalk ganglia and brain. Of these peptides, NRNFLRFamide and SENRNFLRFamide are previously known –FLRFamide isoforms from other decapods [29,30,47]. In addition, a single FaRP of the –YLRamide subtype, GAHKNYLRFamide, was sequenced/detected from the eyestalk ganglia and ventral nerve cord of *L. vannamei* (Table 3); this peptide is known from several decapod species [14,47,48]. Surprisingly, two FaRPs possessing novel –FVRFamide C-termini, GYSNKNFVRFamide and GYSNKDFVRFamide, were also sequenced via ESI-Q-TOF MS/MS from the brain of *L. vannamei* (Table 3). The former peptide was also detected via MALDI-FTMS in the brain and ventral nerve cord (Table 3).

3.10. Neuroparsin

The neuroparsins are a family of pleiotropic neuropeptides originally described from insects; these peptide possess a characteristic arrangement of 12 Cys residues that allow for the formation of six disulfide bridges, which is considered a hallmark of the neuroparsins [1]. In

decapod crustaceans a neuroparsin-like peptide was recently predicted from the crab, *C. maenas* [47]. Here, four ESTs (Table 1) were identified as encoding putative neuroparsin precursors via a query using the sequence of a locust *Schistocerca gregaria* prepro-neuroparsin 1 (accession no. **CAC38869**) [38]. Translation of these transcripts revealed each to encode a 101 amino acid, putative full-length precursor protein (Fig. 3B), with those deduced from **FE068047**, **FE060966** and **FE056058** being identical in sequence (Fig. 3B). The precursor predicted from **FE056059** differed from that of **FE068047**, **FE060966** and **FE056058** at five residues, *i.e.* Phe vs Leu at positions 17, 19 and 21, Leu vs Phe at position 45, and Pro vs Thr at position 92 (Fig. 3B). Bioinformatics conducted on the deduced proteins suggests a single neuroparsin-like peptide is produced from each precursor (Table 2):
 TPVCPGTRDPPQDLSKCKFGVVKDWCRNTVCAKGPRETCGGRWLEHGRCGLGMY
 CRCGHCAGCTSTLECVLGRFC from **FE068047**, **FE060966** and **FE056058** and
 TPVCPGTRDPPQDLSKCKLGVVKDWCRNTVCAKGPRETCGGRWLEHGRCGLGMY
 CRCGHCAGCTSPLECVLGRFC from **FE056059**.

Neither of the putative *L. vannamei* neuroparsin-like peptides was identified in our mass spectral experiments.

3.11. Orcokinin

In decapod crustaceans, members of the orcokinin family are characterized by the presence of the N-terminal motif NFDEIDR-, as well as exhibiting an overall length of 13 amino acids [6]. As the family name implies, the original member of this group of peptides was isolated and characterized from the crayfish *Orconectes limosus* [69]; isoforms of orcokinin have subsequently been identified from a wide variety of decapod species [*e.g.*,^{5,6,7,19,44,65,66,69,90}]. In our study, no orcokinin-encoding transcripts were identified. However, a large collection of orcokinin and orcokinin-related peptides was characterized via ESI-Q-TOF MS/MS and/or MALDI-FTMS (Table 3), including seven full-length isoforms, *i.e.* NFDEIDRAGMGFA (from eyestalk ganglia, brain and ventral nerve cord) and its Met oxidized form NFDEIDRAGM(O)GFA (from the same tissues), NFDEIDRSGFGFA (from brain), NFDEIDRAGFGFN (from brain and ventral nerve cord), NFDEIDRTGFGFH (from brain), NFDEIDRSGFGFN (from brain and ventral nerve cord) and NFDEIDRAGFGFL (from ventral nerve cord), nine putative truncations (all from brain unless otherwise noted), *i.e.* NFDEIDRAGM(O)GF, FDEIDRAGM(O)GFA, NFDEIDRSGFG, NFDEIDRSGFA, DEIDRAGM(O)GFA, FDEIDRAGMG, EIDRSGFGFA, NFDEIDRAG and NFDEIDRA (from brain and ventral nerve cord), and three amidated variants (each from brain), *i.e.* NFDEIDRAGFamide, NFDEIDRSGFamide and DFDEIDRAGFamide. Of the full-length orcokinins, NFDEIDRAGMGFA, NFDEIDRSGFGFA, NFDEIDRTGFGFH and NFDEIDRSGFGFN are peptides previously identified from other decapods [*e.g.*,^{5,6,7,19,44,47,48,65,66,69,90}]. Similarly, a subset of the truncated forms, *i.e.* NFDEIDRSGFG, NFDEIDRSGFA, EIDRSGFGFA, and NFDEIDRA, are known from other species [*e.g.*,^{6,15,44,47,65}]. Likewise, the amidated truncation NFDEIDRSGFamide is a previously known variant [48].

3.12. Orcomyotropin

The myotropic peptide FDAFTTGfamide was originally isolated and characterized from the crayfish *O. limosus* [19]. Given its source and bioactivity, this peptide was named orcomyotropin [19]. C-terminally extended variants of this peptide, *e.g.* FDAFTTGFGHN and FDAFTTGFGHS, possible precursors of FDAFTTGfamide, have been identified from a variety of decapod species [74], and recent molecular studies show that an extended variant is encoded with multiple copies of orcokinin on a common precursor protein [15,90]. In our study, FDAFTTGFGHS was sequenced via ESI-Q-TOF MS/MS from the eyestalk ganglia, brain and

ventral nerve cord of *L. vannamei* (Table 3). This peptide was also detected via MALDI-FTMS in the brain (Table 3).

3.13. Proctolin

The peptide RYLPT, originally identified from an insect and commonly referred to as proctolin [3,70], is a well-known crustacean neuropeptide [*e.g.* 29,43,47,48]. Here, one EST (Table 1) was identified as encoding a putative proctolin precursor via a query using the sequence of a *D. melanogaster* prepro-proctolin (accession no. **CAD30643**) [76]. Translation of this EST, **FE183480**, revealed it to encode a 123 amino acid, putative full-length precursor protein. Putative post-translational processing of the deduced precursor predicts the liberations of as many as six peptides (Table 2), including authentic proctolin.

Neither proctolin nor any of the other predicted proctolin precursor-related peptides were identified by mass spectrometry.

3.14. Pyrokinin

The pyrokinin/pheromone biosynthesis activating neuropeptide (PBAN) family of peptides is characterized by the C-terminal motif –FXPRLamide (where X is a variable amino acid). Members of this peptide family have been characterized from a number of decapod species, including *L. vannamei* [47,48,62,79]. While no pyrokinin-encoding transcripts were identified in our study, nine peptides possessing –FXPRLamide C-termini were characterized in *L. vannamei* via a combination of MALDI-FTMS and/or ESI-Q-TOF MS/MS (Table 3). Six of the pyrokinins, DFAFNPRamide, DFSFNPRamide, GDFAFSPRLamide, ADFAFSPRLamide, GDFAFNPRamide and SGGFAFSPRLamide, were sequenced via ESI-Q-TOF MS/MS from the brain (SGGFAFSPRLamide was also detected via MALDI-FTMS from this tissue), while the three remaining peptides, YSFLPRamide, DFAFSPRLamide and ADFAFNPRamide, were sequenced from both the brain and the ventral nerve cord. Both DFAFSPRLamide and ADFAFNPRamide are previously known *L. vannamei* peptides [79]. SGGFAFSPRLamide has been identified previously from other decapod species [62].

3.15. Red pigment concentrating hormone

The peptide pELNFSPGWamide was originally isolated and characterized from the shrimp *Pandalus borealis*, and due to its ability to concentrate pigment in erythrophores, was named red pigment concentrating hormone or RPCH [26]. Since its initial description, this peptide has been identified in authentic form from many decapod species [*e.g.* ^{17,40,43,45,47,48,51}]. While no RPCH-encoding transcripts were identified by transcriptome mining, a peptide identical to authentic RPCH was sequenced from the eyestalk ganglia of *L. vannamei* via ESI-Q-TOF MS/MS (Table 3).

3.16. RYamide

A family of peptides possessing the C-terminal motif –RYamide, was recently identified from the pericardial organ of the crab *C. borealis* [43]. Members of this peptide family have subsequently been identified from a number of other decapod species [29,47,73]. Though no –RYamide-encoding transcripts were identified via transcriptome mining, two –RYamide isoforms, pEGFYSQRYamide and SGFYANRYamide [29,43,47,73], were characterized in the brain and ventral nerve cord of *L. vannamei* via a combination of ESI-Q-TOF MS/MS and MALDI-FTMS (Table 3).

3.17. SIFamide

Members of the SIFamide family are characterized by the C-terminal motif –SIFamide [85]. Here, one EST (Table 1) was identified as encoding a putative SIFamide precursor via a query using the sequence of an American lobster *H. americanus* prepro-Val¹-SIFamide (accession no. **ABV21807**) [16]. Translation of this EST, **FE187321**, revealed it to encode a 76 amino acid, putative full-length precursor protein (Fig. 3D). Putative post-translational processing of the deduced prepro-hormone suggests that two peptides are cleaved from it, including an isoform of SIFamide (Table 2). As the signal peptide predicted by SignalP differed between the Neural Networks and the Hidden Markov Models algorithms, the encoded SIFamide isoform would be either EPVSAGYRKPPFNGSIFamide or GYRKPPFNGSIFamide (depending on which signal sequence is used; Fig. 3D). GYRKPPFNGSIFamide, *e.g.* Gly¹-SIFamide, is a known and broadly conserved decapod variant [64,74].

Using ESI-Q-TOF MS/MS, GYRKPPFNGSIFamide was sequenced from the eyestalk ganglia, brain and ventral nerve cord; it was also identified via MALDI-FTMS from all three tissues as well (Table 3). While we did not detect EPVSAGYRKPPFNGSIFamide in our study, a second variant, RKPPFNGSIFamide, a peptide previously described from the lobster *H. americanus* [48], was sequenced via ESI-Q-TOF MS/MS from the brain and ventral nerve cord (Table 3).

3.18. Tachykinin-related peptide

A number of peptides that share sequence similarity to the vertebrate tachykinins have been identified in invertebrates. Members of this peptide family, often referred to as the tachykinin-related peptides (TRPs), typically possess the C-terminal motif –FX₁GX₂Ramide, where X₁ and X₂ represent variable residues [55]. While no TRP-encoding ESTs were identified from *L. vannamei*, seven family members were characterized via a combination of ESI-Q-TOF MS/MS and/or MALDI-FTMS (Table 3): APSGFLGMRamide (from brain and ventral nerve cord), its Met-oxidized form APSGFLGM(O)Ramide (from eyestalk ganglia, brain and ventral nerve cord) and its putative precursor APSGFLGMRG (from eyestalk ganglia and ventral nerve cord), APAGFLGMRamide (from brain and ventral nerve cord) and its Met-oxidized form APAGFLGM(O)Ramide (from eyestalk ganglion and brain), APSGFNGM(O)Ramide (from brain only) and APSGFLDM(O)Ramide (from brain only). Of these TRPs, APSGFLGMRamide is a well-known and broadly conserved decapod isoform [12,74,90], including being identified previously from *L. vannamei* [57]. The Met-oxidized form of APSGFLGMRamide and the putative immature form the peptide are also known decapod variants [47,48,75].

3.19. Other peptides

In addition to peptides with sequences that place them into known peptide families, a novel peptide, L/IPEPDPMAGHEL/I, was sequenced from the eyestalk ganglia via ESI-Q-TOF MS/MS (Table 3). Due to the methodology used for its identification, we cannot differentiate between Leu and Ile in this peptide's sequence.

4. Discussion

4.1. Combining transcriptomics and mass spectrometry for peptide discovery in *Litopenaeus vannamei*

Growth and reproduction are among the many physiological processes that are controlled, at least in part, by circulating hormones and/or locally-released paracrine. Given the importance of these processes for the aquaculture of *L. vannamei*, a thorough cataloging of the peptide hormones/paracrine present in this species is an important first step in understanding endocrine control in it. To this end, we have conducted a study to elucidate the neuropeptidome

of *L. vannamei*. To achieve the greatest coverage of peptides, we surveyed multiple tissues using a strategy combining *in silico* transcriptome mining and mass spectrometry. For our mass spectral analyses, two distinct platforms were used: MALDI-FTMS for high resolution mass profiling of known/predicted peptides and ESI-Q-TOF MS/MS for sequencing both known and novel ones. In total, 120 peptides were identified using this combined approach (Tables 2–3), with approximately 98% of them novel, either for the species or in a general sense.

Of the peptides characterized in our study, five were identified by both transcriptomics and at least one of the mass spectral platforms (Tables 2–3), such as Arg⁷-corazonin, which was identified in our *in silico* analyses and by both MALDI-FTMS and ESI-Q-TOF MS/MS. This dual identification provides a high degree of confidence in the structural assignment of these peptides, with the transcriptomics allowing for the unambiguous assignment of amino acids that can not be differentiated via mass spectral means, for example the isobaric amino acids Leu and Ile, and the mass spectrometry confirming post-translational modifications predicted by the bioinformatics, *e.g.* N-terminal cyclization of Gln or Glu, C-terminal amidation and disulfide bridging between Cys residues. Other peptides were identified by only one of the two methods (compare Tables 2–3), *i.e.* members of the C-type AST family only by transcriptomics vs. multiple isoforms of the A-type and the B-type ASTs solely by mass spectrometry. For peptides present in small quantities in the nervous system, or ones that possess structures that are not readily ionizable, a lack of detection by mass spectrometry is not surprising. Likewise, the ESTs currently extant represent only a portion of the *L. vannamei* transcriptome, and therefore many peptide-encoding transcripts are undoubtedly absent in the collection that is currently available publicly. Thus, the combined approach used here to identify the neuropeptides has capitalized on the strengths, and minimized the weaknesses, of the two techniques employed.

Despite the combined approach to neuropeptide discovery employed here, and the large number of peptides identified using it, it is important to note that our results clearly represent only a portion of the total peptidome present in the *L. vannamei* nervous system, as a number of peptides/peptide families previously described from this species were not re-identified in our study. For example, no members of the pigment dispersing hormone (PDH) family were found, though several PDH isoforms are known from *L. vannamei*, *i.e.*

NSELINSLLGIPKVMNDAamide and NSELINSLGLPKVMNDAamide [58]. Similarly, while one CHH peptide was identified here (see Section 3.8), additional family members are known that were not re-identified in our study [*e.g.* 77,78,88]. As stated earlier, it is likely that some of the missing peptides are present in low abundance, which could render them below the limit of detection by mass spectrometry, and/or possess structures that are not readily ionizable using the mass spectral techniques employed here. Similarly some peptides, *e.g.* the bursicons, CHHs and the neuroparsins, are too large to be fully-sequenced via the mass spectral methods used in our study. Clearly as more ESTs are produced for *L. vannamei*, it will be interesting to see what additional peptides are found via database mining, and whether or not these peptides fill in the known gaps currently extant in our dataset.

4.2. Identification of crustacean C-type allatostatins

Among the peptide-encoding transcripts identified here, and in two other studies [49,72], were four containing an isoform of C-AST, either pQIRYHQCYFNPISCF (in three of the four ESTs) or pQIRYHQCYFIPVSCF (in the remaining transcript), disulfide bridging predicted between the two Cys residues in each peptide. The former peptide possesses a pyroglutamine blocked N-terminus, an unamidated C-terminal motif –PISCF, and a disulfide bridge between the Cys residues located at positions 7 and 14, which are considered as the hallmarks of the C-AST family [71]. With the exception of a Val for Ile substitution in the C-terminal motif, pQIRYHQCYFIPVSCF too is predicted to possess all of the hallmarks of an authentic C-type

AST peptide. Recently, the peptide pQIRYHQCYFNPISCF was identified and sequenced by mass spectrometry in Jonah crab *Cancer borealis* and was found to exert state-dependent inhibitory effect on the pyloric rhythm [49]. pQIRYHQCYFNPISCF was also identified in 27 other decapods by accurate mass measurements via MALDI FTMS [72]. These findings suggest that members of the C-type allatostatin family are not restricted to holometabolous insects, but may well be broadly distributed within the Arthropoda, certainly at least within the Pancrustacea. Interestingly, while *in silico* analysis of crustacean ESTs identified several transcripts encoding PISCF-type (C-type) allatostatin precursors, none of the predicted C-type AST peptides, including pQIRYHQCYFNPISCF, was detected by mass spectrometry in *L. vannamei*. Given the fact that at least pQIRYHQCYFNPISCF is detectable via the mass spectral techniques employed in our study [49], our lack of detection of pQIRYHQCYFNPISCF and/or pQIRYHQCYFIPVSCF suggests that these peptides are present in low quantities within the nervous system, at least in the areas surveyed here.

4.3. Molecular confirmation of crustacean corazonin

pQTFQYSRGWTamide, often referred to as Arg⁷-corazonin, is a well known insect peptide hormone [61]. Based on accurate mass measurements, it also appears to be present in the neural tissues of a number of decapod crustaceans [43,47,48], including *L. vannamei*. Molecular confirmation of this peptide's existence in decapod species, however, has remained elusive. Here, we also identified *L. vannamei* transcripts encoding precursors of pQTFQYSRGWTamide, therein providing the first molecular confirmation of the peptide's presence in authentic form in crustaceans. Interestingly, the source of the ESTs from which Arg⁷-corazonin was predicted was the lymphoid organ, a hemolymph filtering structure which has been proposed to play a role in innate immunity [59,81]. The expression of corazonin-encoding transcripts in this structure suggests that tissues in addition to neural ones may produce this peptide in *L. vannamei*, and potentially other species as well. Evidence for "neuropeptides" being produced and secreted by non-neural tissues is growing in decapod crustaceans, for example from the midgut [10,11,75], and certainly the lymphoid organ's proximity to the hemolymph makes it a logical candidate for endocrine release. As additional studies are conducted, it will be interesting to see if this structure does in fact produce and release corazonin. Likewise, it will be important to determine if the lymphoid organ synthesizes and secretes other peptide hormones, particularly other known "neuropeptides", as well as to determine what factors influence peptide release from this structure, and what functional roles lymphoid organ-derived hormones/paracrine play in *L. vannamei*.

4.4. Identification of the first crustacean proctolin-encoding transcript

Like corazonin, RYLPT, commonly known as proctolin, is a peptide originally described from insects [3,70] that has subsequently been found broadly conserved in crustaceans [28,29,43,47,48,63,67]. Despite its biochemical/mass spectral detection in numerous decapods, no proctolin-encoding transcript had been identified in any crustacean species. Here, we have identified a transcript from *L. vannamei* that encodes RYLPT, confirming, for the first time at the transcript level, proctolin's presence in a crustacean. The identification of this and the other peptide encoding transcripts described in our study are of note, as these transcripts now serve as templates for gene discovery in other decapods, and can be used for gene-based manipulation of their respective peptidergic systems *in vivo*, e.g. RNAi knockdown studies, thereby providing a new avenue for pursuing the functional roles played by neuropeptides in members of the Decapoda.

4.5. In contrast to other decapods, *L. vannamei* appears to contain a large collection of tachykinin-related peptides

Over a quarter of a century ago, immunohistochemistry suggested the presence of tachykinins, or related peptides, in members of the Decapoda [e.g. 27,34,50]. It was nearly two decades later, however, before the first TRP, APSGFLGMRamide, was isolated and characterized from crustacean tissues [12]. In contrast to insects, where multiple TRP isoforms are commonly found in any given species [55], it was long held that decapods possessed only APSGFLGMRamide [74,90]. Recently, this dogma was challenged by the finding of a second TRP, TPSGFLGMRamide, in some members of the Decapoda [10,47,75]. Still, a maximum of two isoforms in any given species was far fewer than that found in most insects, whereas over a dozen TRPs have been characterized from several species [55], e.g. 13 TRPs each in the cockroaches *P. americana* and *Leucophaea maderae* [60]. Here, using mass spectral techniques, we have identified the TRP APSGFLGMRamide, as well as three additional novel full-length isoforms, APAGFLGMRamide, APSGFNGMRamide and APSGFLDMRamide, from *L. vannamei*. This collection of four peptides represents by far the largest number of TRP isoforms identified from any decapod, and in fact rivals the complements present in many insects [55]. Why *L. vannamei* possess this extensive complement of TRPs, relative to other decapods, is unknown. Interestingly, this species is a member of the suborder Dendrobranchiata, the most basal of the decapod taxa, and thus is potentially more closely related to the insect lineage than are the other decapods. Clearly it will be interesting to see if other penaeids also possess multiple TRPs, as well as to examine shrimp species from the Pleocyemata, i.e. stenopodid and caridean shrimp, for their TRP complements, as they too are considered basal relative to most of the other decapod taxa, and have, like the penaeids, been poorly studied in terms of their neuropeptide complements.

Acknowledgments

The University of Wisconsin School of Pharmacy Analytical Instrumentation Center is thanked for providing us access to the MALDI-FTMS instrument. Dr. Peter O'Connor (Boston University) is thanked for providing BUDA software. L.L. acknowledges financial support from the University of Wisconsin School of Pharmacy, Wisconsin Alumni Research Foundation, National Science Foundation (CAREER Award CHE-0449991), National Institutes of Health (1R01DK071801) and a research fellowship from the Alfred P. Sloan Foundation. A.E.C. acknowledges financial support from the National Center for Research Resources' Maine INBRE Program (NIH P20 RR-016463; to Mount Desert Island Biological Laboratory [MDIBL]; Dr. Patricia Hand, Principle Investigator), a MDIBL New Investigator Award (from the Salisbury Cove Research Fund provided through the Thomas H. Maren Foundation), MDIBL institutional funds, and funds provided by the Cades Foundation of Honolulu, Hawaii.

References

1. Badisco L, Claeys I, Van Loy T, Van Hiel M, Franssens V, Simonet G, Vanden Broeck J. Neuroparsins, a family of conserved arthropod neuropeptides. *Gen Comp Endocrinol* 2007;153:64–71. [PubMed: 17475261]
2. Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 2004;340:783–95. [PubMed: 15223320]
3. Brown BE. Proctolin: a peptide transmitter candidate in insects. *Life Sci* 1975;17:1241–52. [PubMed: 575]
4. Brown MR, Crim JW, Arata RC, Cai HN, Chun C, Shen P. Identification of a *Drosophila* brain-gut peptide related to the neuropeptide Y family. *Peptides* 1999;20:1035–42. [PubMed: 10499420]
5. Bungart D, Hilbich D, Dirksen H, Keller R. Occurrence of analogues of the myotropic neuropeptide orcokinin in the nervous system of the shore crab, *Carcinus maenas*: evidence for a novel neuropeptide family. *Peptides* 1995;16:67–72. [PubMed: 7716076]
6. Cashman, CR. Thesis for the Biochemistry Program. Bowdoin College: 2007. The identification and comparative analysis of orcokinin family neuropeptides in decapod crustaceans using matrix assisted laser desorption/ionization-Fourier transform mass spectrometry.

7. Christie AE. Neuropeptide discovery in Ixodoidea: an *in silico* investigation using publicly accessible expressed sequence tags. *Gen Comp Endocrinol* 2008a;157:174–85. [PubMed: 18495123]
8. Christie AE. *In silico* analyses of peptide paracrine/hormones in Aphidoidea. *Gen Comp Endocrinol* 2008b;159:67–79. [PubMed: 18725225]
9. Christie AE, Cashman CR, Brennan HR, Ma M, Sousa GL, Li L, Stemmler EA, Dickinson PS. Identification of putative crustacean neuropeptides using *in silico* analyses of publicly accessible expressed sequence tags. *Gen Comp Endocrinol* 2008c;156:246–64. [PubMed: 18321503]
10. Christie AE, Cashman CR, Stevens JS, Smith CM, Beale KM, Stemmler EA, Greenwood SJ, Towle DW, Dickinson PS. Identification and cardiotropic actions of brain/gut-derived tachykinin-related peptides (TRPs) from the American lobster *Homarus americanus*. *Peptides* 2008b;29:1909–1918. [PubMed: 18706463]
11. Christie AE, Kutz-Naber KK, Stemmler EA, Klein A, Messinger DI, Goiney CC, Conterato AJ, Bruns EA, Hsu YW, Li L, Dickinson PS. Midgut epithelial endocrine cells are a rich source of the neuropeptides APSGFLGMRamide (*Cancer borealis* tachykinin-related peptide Ia) and GYRKPPFNGSIFamide (Gly¹-SIFamide) in the crabs *Cancer borealis*, *Cancer magister* and *Cancer productus*. *J Exp Biol* 2007;210:699–714. [PubMed: 17267655]
12. Christie AE, Lundquist CT, Nässel DR, Nusbaum MP. Two novel tachykinin-related peptides from the nervous system of the crab *Cancer borealis*. *J Exp Biol* 1997;200:2279–94. [PubMed: 9316266]
13. Chung JS, Wilcockson DC, Zmora N, Zohar Y, Dirksen H, Webster SG. Identification and developmental expression of mRNAs encoding crustacean cardioactive peptide (CCAP) in decapod crustaceans. *J Exp Biol* 2006;209:3862–72. [PubMed: 16985202]
14. Cruz-Bermúdez ND, Fu Q, Kutz-Naber KK, Christie AE, Li L, Marder E. Mass spectrometric characterization and physiological actions of GAHKNYLRFamide, a novel FMRFamide-like peptide from crabs of the genus *Cancer*. *J Neurochem* 2006;97:784–99. [PubMed: 16515542]
15. Dickinson PS, Stemmler EA, Barton EE, Cashman CR, Gardner NP, Rus S, Brennan HR, McClintock TS, Christie AE. Molecular, mass spectral, and physiological analyses of orckinins and orckinin precursor-related peptides in the lobster *Homarus americanus* and the crayfish *Procambarus clarkii*. *Peptides* 2009a;30:297–317. [PubMed: 19007832]
16. Dickinson PS, Stemmler EA, Cashman CR, Brennan HR, Dennison B, Huber KE, Peguero B, Rabacal W, Goiney CC, Smith CM, Towle DW, Christie AE. SIFamide peptides in clawed lobsters and freshwater crayfish (Crustacea, Decapoda, Astacidea): a combined molecular, mass spectrometric and electrophysiological investigation. *Gen Comp Endocrinol* 2008b;156:347–60. [PubMed: 18308319]
17. Dickinson PS, Stemmler EA, Christie AE. The pyloric neural circuit of the herbivorous crab *Pugettia producta* shows limited sensitivity to several neuromodulators that elicit robust effects in more opportunistically feeding decapods. *J Exp Biol* 2008a;211:1434–47. [PubMed: 18424677]
18. Dickinson PS, Wiwatpanit T, Gabranski ER, Ackerman RJ, Stevens JS, Cashman CR, Stemmler EA, Christie AE. Identification of SYWKQCAFNAVSCFamide: a broadly conserved crustacean C-type allatostatin-like peptide with both neuromodulatory and cardioactive properties. *J Exp Biol* 2009b;212:1140–52. [PubMed: 19423507]
19. Dirksen H, Burdick S, Sauter A, Keller R. Two orckinins and the novel octapeptide orcomyotropin in the hindgut of the crayfish *Orconectes limosus*: identified myostimulatory neuropeptides originating together in neurons of the terminal abdominal ganglion. *J Exp Biol* 2000;203:2807–18. [PubMed: 10952880]
20. Dirksen H, Skiebe P, Abel B, Agricola H, Buchner K, Muren JE, Nässel DR. Structure, distribution, and biological activity of novel members of the allatostatin family in the crayfish *Orconectes limosus*. *Peptides* 1999;20:695–712. [PubMed: 10477125]
21. Diwan AD. Current progress in shrimp endocrinology. A review. *Indian J Endocrinol* 2005;43:209–243.
22. Duve H, Johnsen AH, Maestro JL, Scott AG, Jaros PP, Thorpe A. Isolation and identification of multiple neuropeptides of the allatostatin superfamily in the shore crab *Carcinus maenas*. *Eur J Biochem* 1997;250:727–34. [PubMed: 9461295]
23. Duve H, Johnsen AH, Scott AG, Thorpe A. Allatostatins of the tiger prawn, *Penaeus monodon* (Crustacea: Penaeidea). *Peptides* 2002;23:1039–51. [PubMed: 12126730]

24. Fanjul-Moles ML. Biochemical and functional aspects of crustacean hyperglycemic hormone in decapod crustaceans: review and update. *Comp Biochem Physiol C Toxicol Pharmacol* 2006;142:390–400. [PubMed: 16403679]
25. Fast, AW.; Lester, LJ. Marine shrimp culture: principles and practices. Amsterdam: Elsevier; 1992.
26. Fernlund P. Structure of the red-pigment-concentrating hormone of the shrimp, *Pandalus borealis*. *Biochim Biophys Acta* 1974;371:304–11. [PubMed: 4433569]
27. Fingerman M, Hanumante MM, Kulkarni GK, Ikeda R, Vacca LL. Localization of substance P-like, leucine-enkephalin-like, methionine-enkephalin-like, and FMRFamide-like immunoreactivity in the eyestalk of the fiddler crab, *Uca pugnator*. *Cell Tissue Res* 1985;241:473–7. [PubMed: 2411411]
28. Fu Q, Goy MF, Li L. Identification of neuropeptides from the decapod crustacean sinus glands using nanoscale liquid chromatography tandem mass spectrometry. *Biochem Biophys Res Commun* 2005a; 337:765–78. [PubMed: 16214114]
29. Fu Q, Kutz KK, Schmidt JJ, Hsu YW, Messinger DI, Cain SD, de la Iglesia HO, Christie AE, Li L. Hormone complement of the *Cancer productus* sinus gland and pericardial organ: an anatomical and mass spectrometric investigation. *J Comp Neurol* 2005b;493:607–26. [PubMed: 16304631]
30. Fu Q, Li L. De novo sequencing of neuropeptides using reductive isotopic methylation and investigation of ESI QTOF MS/MS fragmentation pattern of neuropeptides with N-terminal dimethylation. *Anal Chem* 2005;77:7783–7795. [PubMed: 16316189]
31. Fu Q, Tang LS, Marder E, Li L. Mass spectrometric characterization and physiological actions of VPNDWAHFRGSWamide, a novel B type allatostatin in the crab, *Cancer borealis*. *J Neurochem* 2007;101:1099–107. [PubMed: 17394556]
32. Garczynski SF, Brown MR, Crim JW. Structural studies of *Drosophila* short neuropeptide F: occurrence and receptor binding activity. *Peptides* 2006;27:575–82. [PubMed: 16330127]
33. Gard AL, Lenz PH, Shaw JR, Christie AE. Identification of putative peptide paracrine/hormones in the water flea *Daphnia pulex* (Crustacea; Branchiopoda; Cladocera) using transcriptomics and immunohistochemistry. *Gen Comp Endocrinol* 2009;160:271–87. [PubMed: 19135444]
34. Goldberg D, Nusbaum MP, Marder E. Substance P-like immunoreactivity in the stomatogastric nervous systems of the crab *Cancer borealis* and the lobsters *Panulirus interruptus* and *Homarus americanus*. *Cell Tissue Res* 1988;252:515–22. [PubMed: 2456155]
35. Gross MR. One species with two biologies: Atlantic salmon (*Salmo salar*) in the wild and in aquaculture. *Can J Fish Aquat Sci* 1998;55(S1):131–44.
36. Huberman A. Shrimp endocrinology. A review. *Aquaculture* 2000;191:191–208.
37. Huybrechts J, Nusbaum MP, Bosch LV, Baggerman G, De Loof A, Schoofs L. Neuropeptidomic analysis of the brain and thoracic ganglion from the Jonah crab, *Cancer borealis*. *Biochem Biophys Res Commun* 2003;308:535–44. [PubMed: 12914784]
38. Janssen T, Claeys I, Simonet G, De Loof A, Girardie J, Vanden Broeck J. cDNA cloning and transcript distribution of two different neuropeptide precursors in the desert locust, *Schistocerca gregaria*. *Insect Mol Biol* 2001;10:183–9. [PubMed: 11422514]
39. Kastin, AJ. Handbook of biologically active peptides. Academic Press, Elsevier Inc; Burlington, MA: 2006. p. 201–6. p. 21–8.
40. Klein JM, Mohrher CJ, Sleutels F, Jaenecke N, Riehm JP, Rao KR. A highly conserved red pigment-concentrating hormone precursor in the blue crab *Callinectes sapidus*. *Biochem Biophys Res Commun* 1995;212:151–58. [PubMed: 7611999]
41. Lago-Lestón MA, Ponce E, Muñoz ME. Cloning and expression of hyperglycemic (CHH) and molt-inhibiting (MIH) hormones mRNAs from the eyestalk of shrimps of *Litopenaeus vannamei* grown in different temperature and salinity conditions. *Aquaculture* 2007;270:343–57.
42. Lem, A. An overview of global shrimp markets and trade. In: Leung, P.; Engle, C., editors. Shrimp culture: economics, market and trade. 1. Ames: Blackwell Publishing; 2006. p. 3–10.
43. Li L, Kelley WP, Billimoria CP, Christie AE, Pulver SR, Sweedler JV, Marder E. Mass spectrometric investigation of the neuropeptide complement and release in the pericardial organs of the crab, *Cancer borealis*. *J Neurochem* 2003;87:642–56. [PubMed: 14535947]
44. Li L, Pulver SR, Kelley WP, Thirumalai V, Sweedler JV, Marder E. Orcokinin peptides in developing and adult crustacean stomatogastric nervous systems and pericardial organs. *J Comp Neurol* 2002;444:227–244. [PubMed: 11840477]

45. Linck B, Klein JM, Mangerich S, Keller R, Weidemann WM. Molecular cloning of crustacean red pigment concentrating hormone precursor. *Biochem Biophys Res Commun* 1993;195:807–13. [PubMed: 8373416]
46. Luo CW, Dewey EM, Sudo S, Ewer J, Hsu SY, Honegger HW, Hsueh AJ. Bursicon, the insect cuticle-hardening hormone, is a heterodimeric cystine knot protein that activates G protein-coupled receptor LGR2. *Proc Natl Acad Sci U S A* 2005;102(8):2820–5. [PubMed: 15703293]
47. Ma M, Bors EK, Dickinson ES, Kwiatkowski MA, Sousa GL, Henry RP, Smith CM, Towle DW, Christie AE, Li L. Characterization of the *Carcinus maenas* neuropeptide by mass spectrometry and functional genomics. *Gen Comp Endocrinol* 2009;161:320–34. [PubMed: 19523386]
48. Ma M, Chen R, Sousa GL, Bors EK, Kwiatkowski MA, Goiney CC, Goy MF, Christie AE, Li L. Mass spectral characterization of peptide transmitters/hormones in the nervous system and neuroendocrine organs of the American lobster *Homarus americanus*. *Gen Comp Endocrinol* 2008a;156:395–409. [PubMed: 18304551]
49. Ma M, Szabo MT, Jia C, Marder E, Li L. Mass spectrometric characterization and physiological actions of novel crustacean C-type allatostatins. *Peptides* 2009;30:1660–8. [PubMed: 19505516]
50. Mancillas JR, McGinty JF, Selverston AI, Karten H, Bloom FE. Immunocytochemical localization of enkephalin and substance P in retina and eyestalk neurones of lobster. *Nature* 1981;293:576–8. [PubMed: 6169995]
51. Martinez-Perez F, Zinker S, Aguilar G, Valdes J, Arechiga A. Circadian oscillations of RPCH gene expression in the eyestalk of the crayfish *Cherax quadricarinatus*. *Peptides* 2005;26:2434–44. [PubMed: 15992960]
52. Mendive FM, Van Loy T, Claeysen S, Poels J, Williamson M, Hauser F, Grimmelikhuijzen CJ, Vassart G, Vanden Broeck J. *Drosophila* molting neurohormone bursicon is a heterodimer and the natural agonist of the orphan receptor DLGR2. *FEBS Lett* 2005;579:2171–2176. [PubMed: 15811337]
53. Mercier AJ, Orchard I, TeBrugge V, Skerrett M. Isolation of two FMRFamide-related peptides from crayfish pericardial organs. *Peptides* 1993;14:137–43. [PubMed: 8387183]
54. Monigatti F, Gasteiger E, Bairoch A, Jung E. The Sulfinator: predicting tyrosine sulfation sites in protein sequences. *Bioinformatics* 2002;18:769–70. [PubMed: 12050077]
55. Nässel DR. Tachykinin-related peptides in invertebrates: a review. *Peptides* 1999;20:141–58. [PubMed: 10098635]
56. Nichols R, Bendena WG, Tobe SS. Myotropic peptides in *Drosophila melanogaster* and the genes that encode them. *J Neurogenet* 2002;16:1–28. [PubMed: 12420787]
57. Nieto J, Veelaert D, Derua R, Waelkens E, Cerstiaens A, Coast G, Devreese B, Van Beeumen J, Calderon J, De Loof A, Schoofs L. Identification of one tachykinin- and two kinin-related peptides in the brain of the white shrimp, *Penaeus vannamei*. *Biochem Biophys Res Commun* 1998;248:406–11. [PubMed: 9675150]
58. Ohira T, Tsutsui N, Kawazoe I, Wilder MN. Isolation and characterization of two pigment-dispersing hormones from the whiteleg shrimp, *Litopenaeus vannamei*. *Zoology Sci* 2006;23:601–6.
59. Pongsomboon S, Wongpanya R, Tang S, Chalorsrikul A, Tassanakajon A. Abundantly expressed transcripts in the lymphoid organ of the black tiger shrimp, *Penaeus monodon*, and their implication in immune function. *Fish Shellfish Immunol* 2008;25:485–93. [PubMed: 18692576]
60. Predel R, Neupert S, Roth S, Derst C, Nässel DR. Tachykinin-related peptide precursors in two cockroach species. *FEBS J* 2005;272:3365–3375. [PubMed: 15978042]
61. Predel R, Neupert S, Russell WK, Scheibner O, Nachman RJ. Corazonin in insects. *Peptides* 2007;28:3–10. [PubMed: 17140699]
62. Saideman SR, Ma M, Kutz-Naber KK, Cook A, Torfs P, Schoofs L, Li L, Nusbaum MP. Modulation of rhythmic motor activity by pyrokinin peptides. *J Neurophysiol* 2007;97:579–95. [PubMed: 17065249]
63. Schwarz TL, Lee GM, Siwicki KK, Standaert DG, Kravitz EA. Proctolin in the lobster: the distribution, release, and chemical characterization of a likely neurohormone. *J Neurosci* 1984;4:1300–11. [PubMed: 6144736]
64. Sithigorngul P, Pupuem J, Krungkasem C, Longyant S, Chaivisuthangkura P, Sithigorngul W, Petsom A. Seven novel FMRFamide-like neuropeptide sequences from the eyestalk of the giant tiger prawn

- Penaeus monodon*. Comp Biochem Physiol B Biochem Mol Biol 2002;131:325–37. [PubMed: 11959015]
65. Skiebe P, Dreger M, Boerner J, Meseke M, Weckwerth W. Immunocytochemical and molecular data guide peptide identification by mass spectrometry: Orcokinin and orcomyotropin-related peptides in the stomatogastric nervous system of several crustacean species. Cell Mol Biol (Noisy-le-grand) 2003;49:851–71. [PubMed: 14528921]
 66. Skiebe P, Dreger M, Meseke M, Evers JF, Hucho F. Identification of orcokinins in single neurons in the stomatogastric nervous system of the crayfish *Cherax destructor*. J Comp Neurol 2002;444:245–59. [PubMed: 11840478]
 67. Stangier J, Dirksen H, Keller R. Identification and immunocytochemical localization of proctolin in pericardial organs of the shore crab, *Carcinus maenas*. Peptides 1986;7:67–72. [PubMed: 2872661]
 68. Stangier J, Hilbich C, Beyreuther K, Keller R. Unusual cardioactive peptide (CCAP) from pericardial organs of the shore crab *Carcinus maenas*. Proc Natl Acad Sci USA 1987;84:575–9. [PubMed: 16593803]
 69. Stangier J, Hilbich C, Burdzik S, Keller R. Orcokinin: a novel myotropic peptide from the nervous system of the crayfish, *Orconectes limosus*. Peptides 1992;13:859–64. [PubMed: 1480511]
 70. Starratt AN, Brown BE. Structure of the pentapeptide proctolin, a proposed neurotransmitter in insects. Life Sci 1975;17:1253–6. [PubMed: 576]
 71. Stay B, Tobe SS. The role of allatostatins in juvenile hormone synthesis in insects and crustaceans. Annu Rev Entomol 2007;52:277–99. [PubMed: 16968202]
 72. Stemmler EA, Bruns EA, Cashman CR, Dickinson PS, Christie AE. Molecular and mass spectral identification of the broadly conserved decapod crustacean neuropeptide pQIRYHQCYFNPISCF: the first PISCF-allatostatin (*Manduca sexta*- or C-type allatostatin) from a non-insect. Gen Comp Endocrinol. 2009 In press.
 73. Stemmler EA, Bruns EA, Gardner NP, Dickinson PS, Christie AE. Mass spectrometric identification of pEGFYSQRYamide: a crustacean peptide hormone possessing a vertebrate neuropeptide Y (NPY)-like carboxy-terminus. Gen Comp Endocrinol 2007c;152:1–7. [PubMed: 17420018]
 74. Stemmler EA, Cashman CR, Messinger DI, Gardner NP, Dickinson PS, Christie AE. High-mass-resolution direct-tissue MALDI-FTMS reveals broad conservation of three neuropeptides (APSGFLGMRamide, GYRKPPFNGSIFamide and pQDLHDHVFLRFamide) across members of seven decapod crustacean infraorders. Peptides 2007a;28:2104–15. [PubMed: 17928104]
 75. Stemmler EA, Peguero B, Bruns EA, Dickinson PS, Christie AE. Identification, physiological actions, and distribution of TPSGFLGMRamide: a novel tachykinin-related peptide from the midgut and stomatogastric nervous system of *Cancer* crabs. J Neurochem 2007b;101:1351–66. [PubMed: 17437551]
 76. Taylor CA, Winther AM, Siviter RJ, Shirras AD, Isaac RE, Nässel DR. Identification of a proctolin prohormone gene (Proct) of *Drosophila melanogaster*: expression and predicted prohormone processing. J Neurobiol 2004;58:379–91. [PubMed: 14750150]
 77. Tiu SH, He JG, Chan SM. The LvCHH-ITP gene of the shrimp (*Litopenaeus vannamei*) produces a widely expressed putative ion transport peptide (LvITP) for osmo-regulation. Gene 2007;396:226–35. [PubMed: 17466469]
 78. Tsutsui N, Ohira T, Kawazoe I, Takahashi A, Wilder MN. Purification of sinus gland peptides having vitellogenesis-inhibiting activity from the whiteleg shrimp *Litopenaeus vannamei*. Mar Biotechnol 2007;9:360–9. [PubMed: 17357858]
 79. Torfs P, Nieto J, Cerstiaens A, Boon D, Baggerman G, Poulos C, Waelkens E, Derua R, Calderón J, De Loof A, Schoofs L. Pyrokinin neuropeptides in a crustacean. Isolation and identification in the white shrimp *Penaeus vannamei*. Eur J Biochem 2001;268:149–54. [PubMed: 11121115]
 80. Trimmer BA, Kobierski LA, Kravitz EA. Purification and characterization of FMRFamide-like immunoreactive substances from the lobster nervous system: isolation and sequence analysis of two closely related peptides. J Comp Neurol 1987;266:16–26. [PubMed: 3429714]
 81. van de Braak CB, Botterblom MH, Taverne N, van Muiswinkel WB, Rombout JH, van der Knaap WP. The roles of haemocytes and the lymphoid organ in the clearance of injected *Vibrio* bacteria in *Penaeus monodon* shrimp. Fish Shellfish Immunol 2002;13:293–309. [PubMed: 12443012]

82. Veenstra JA. Isolation and structure of corazonin, a cardioactive peptide from the American cockroach. *FEBS Lett* 1989;250:231–4. [PubMed: 2753132]
83. Veenstra JA. Isolation and structure of the *Drosophila* corazonin gene. *Biochem Biophys Res Commun* 1994;204:292–6. [PubMed: 7945373]
84. Veenstra JA. Mono- and dibasic proteolytic cleavage sites in insect neuroendocrine peptide precursors. *Arch Insect Biochem Physiol* 2000;43:49–63. [PubMed: 10644969]
85. Verleyen P, Huybrechts J, Schoofs L. SIFamide illustrates the rapid evolution in Arthropod neuropeptide research. *Gen Comp Endocrinol* 2009;162:27–35. [PubMed: 19014945]
86. Vilaplana L, Castresana J, Bellés X. The cDNA for leucomyosuppressin in *Blattella germanica* and molecular evolution of insect myosuppressins. *Peptides* 2004;25:1883–9. [PubMed: 15501519]
87. Wilcockson DC, Webster SG. Identification and developmental expression of mRNAs encoding putative insect cuticle hardening hormone, bursicon in the green shore crab *Carcinus maenas*. *Gen Comp Endocrinol* 2008;156:113–25. [PubMed: 18221939]
88. Wang YJ, Hayes TK, Holman GM, Chavez AR, Keeley LL. Primary structure of CHH/MIH/GIH-like peptides in sinus gland extracts from *Penaeus vannamei*. *Peptides* 2000;21:477–84. [PubMed: 10822102]
89. Williamson M, Lenz C, Winther AM, Nässel DR, Grimmekhuijzen CJ. Molecular cloning, genomic organization, and expression of a C-type (*Manduca sexta*-type) allatostatin preprohormone from *Drosophila melanogaster*. *Biochem Biophys Res Commun* 2001;282:124–30. [PubMed: 11263981]
90. Yasuda-Kamatani Y, Yasuda A. Identification of orcokinin gene-related peptides in the brain of the crayfish *Procambarus clarkii* by the combination of MALDI-TOF and online capillary HPLC/Q-TOF mass spectrometries and molecular cloning. *Gen Comp Endocrinol* 2000;118:161–72. [PubMed: 10753578]
91. Yasuda-Kamatani Y, Yasuda A. APSGFLGMRamide is a unique tachykinin-related peptide in crustaceans. *Eur J Biochem* 2004;271:1546–56. [PubMed: 15066180]
92. Yasuda-Kamatani Y, Yasuda A. 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* 2006;496:135–47. [PubMed: 16528723]
93. Yin GL, Yang JS, Cao JX, Yang WJ. Molecular cloning and characterization of FGLamide allatostatin gene from the prawn, *Macrobrachium rosenbergii*. *Peptides* 2006;27:1241–50. [PubMed: 16376458]
94. Zajac JM, Mollereau C. RFamide peptides. *Introduction Peptides* 2006;27:941–42.

(A) Prepro-C-type allatostatin

```

(FE182974)  MTPTKVHVLAAALLLALALA SPAPQDKPLEAAPNQPHVHLQKRAAPQDSSPEQDLAAF
(FE175093)  MTPTKVHVLAAALLLALALA SPAPQDKPLEAAPNQPHVHLQKRAAPQDSSPEQDLAAF
(FE182975)  -----
(FE180026)  -----

(FE182974)  KDLIVAQVAELRRSWQEPF SALRRTLVDDEGLEEEQAAQSKKKRM LAPLSGLP CELPTIK
(FE175093)  KDLIVAQVAELRRSWQEPF SALRRTLVDDEGLEEEQAAQSKKKRM LAPLSGLP CELPTIK
(FE182975)  -----
(FE180026)  -----

(FE182974)  RQIRYHCYFNP I SCFRRK*
(FE175093)  RQIRYHCYFNP I SCFRRK*
(FE182975)  RQIRYHCYFNP I SCFRRK*
(FE180026)  -QIRYHCYFIPV SCFRRK*

```

(B) Prepro-bursicon α

```

(FE187805)  MNGEFRT LLLALVFAQTQAD E CSLTPVIHILSYPC NSKPIPSFACQCRGTSYVQVSGS
(FE175634)  MNGEFRT LLLALVFAQTQAD E CSLTPVIHILSYPC NSKPIPSFACQCRGTSYVQVSGS
(FE173462)  MNGEFRT LLLALVFAQTQAD E CSLTPVIHILSYPC NSKPIPSFACQCRGTSYVQVSGS
(FE173463)  -----
(FE173025)  MNGEFRT LLLALVFAQTQAD E CSLTPVIHILSYPC NSKPIPSFACQCRGTSYVQV---

(FE187805)  KIWQTERSCMCQESGEREASV T LNC PKARPGEPMRKI L TRAPIDCMCRPCTDVEEGTV
(FE175634)  KIWQTERSCMCQESGEREASV T LNC PKARPGEPMRKI L TRAPIDCMCRPCTDVEEGTV
(FE173462)  KIWQTERSCMCQESGEREASV T LNC PKARPGEPMRKI L TRAPIDCMCRPCTDVEEGTV
(FE173463)  -----
(FE173025)  -----

(FE187805)  LAQEIANFIEDSPMENVPFLKK*
(FE175634)  LAQEIANFIEDSPMENVPFLKK*
(FE173462)  LAQEIANFIEDSPMENVPFLKK*
(FE173463)  LAQEIANFIEDSPMENVPFLKK*
(FE173025)  -----

```

(C) Prepro-bursicon β

```

(FE178442)  MMPHCTVGACVVITVA AVVVGGLA GPSRAHTYGECE TLPSTMHVAK EEFDDAGRLVRT
(FE184710)  -----
(FE184711)  -----

(FE178442)  CEEDLAVNKCEGACVSKVQPSVNTPSGFLKDCRCRETHLRARDV V LTHCCDGDGNRITG
(FE184710)  -----
(FE184711)  -----

(FE178442)  DNGKLT VKLREP AELF -----
(FE184710)  DNGKLT VKLREP ADCQCFKCGNSIR*
(FE184711)  DNGKLT VKLREP GDCQCFKCGNSIR*

```

(D) Prepro-corazonin

```

(FE154856)  MVRSFQQTVTVMAIALCLSLATA QTFQYSRGWTNG RKRSDMSMGVARTVNPLSFPLTRV
(FE154857)  MVRSFQQTVTVMAIALCLSLATA QTFQYSRGWTNG RKRSDMSMGVARTVNPLSFFLTRV

(FE154856)  LPVSLHQNAAYNFPVKTV EERLRALEAGFAA VIRASSANLSSGGDDEYYSEN*
(FE154857)  LPVSLHQNAAYNFPVKTV EERLRALEAGFAA VIRASSANLSSGGDDEYYSEN*

```

Figure 1.

Deduced amino acid sequences of *L. vannamei* C-type allatostatin (C-AST)-, bursicon α -, bursicon β - and corazonin-encoding prepro-hormones. (A) C-AST precursors. (B) Bursicon α precursors. (C) Bursicon β precursors. (D) Corazonin precursors. Accession nos. of the ESTs from which the prepro-hormones were predicted are shown to the left, with the deduced amino acid sequences of the precursor proteins shown on the right. The predicted signal peptides, when present, are shown in gray, with all predicted prohormone convertase cleavage sites shown in black. Isoform(s) of the named peptide family, *i.e.* (A) C-AST, (B) bursicon α , (C) bursicon β or (D) corazonin, are shown in red, while all other precursor-related peptides are shown in blue. Asterisks indicate the presence of a stop codon. Amino acid residues that vary between the precursors of a given family are highlighted in yellow.

(A) Prepro-crustacean cardioactive peptide

```

(FE173081) MSTMNWCGRAGVIAAASLLLLL LAAHSHAGPVAKRDIADLLDGKDKRPFCAFTGCGKKR
(FE187476) MSTMNWCGRAGVIAAASLLLLL LAAHSHAGPVAKRDIADLLDGKDKRPFCAFTGCGKKR
(FE174994) MSTMNWCGRAGVIAAASLLLLL LAAHSHAGPVAKRDIADLLDGKDKRPFCAFTGCGKKR
(FE179552) MSTMNWCGRAGVIAAASLLLLL LAAHSHAGPVAKRDIADLLDGKDKRPFCAFTGCGKKR
(FE190948) MSTMNWCGRAGVIAAASLLLLL LAAHSHAGPVAKRDIADLLDGKDKRPFCAFTGCGKKR
(FE177552) MSTMNWCGRAGVIAAASLLLLL LAAHSHAGPVAKRDIADLLDGKDKRPFCAFTGCGKKR
(FE189608) MSTMNWCGRAGVIAAASLLLLL LAAHSHAGPVAKRDIADLLDGKDKRPFCAFTGCGKKR
(FE175049) MSTMNWCGRAGVIAAASLLLLL LAAHSHAGPVAKRDIADLLDGKDKRPFCAFTGCGKKR
(FE177553) ----- LAAHSHAGPVAKRDIADLLDGKDKRPFCAFTGCGKKR
(FE189609) ----- FLAAPSAGPVAKRDIADLLDGKDKRPFCAFTGCGKKR
(FE187520) MSTMNWCGRAGVIAAASLLLLL LAAHSHAGPVAKRDIADLLDGKDKRPFCAFTGCGKKR
(FE180576) ----- LAAHSHAGPVAKRDIADLLDGKDKRPFCAFTGCGKKR
(FE175050) -----

(FE173081) SEPKVEPLSSGAELDALAKHVVAEAKLWEQLONKMEVMRSLAARMEHPIFRRKRSPAPA
(FE187476) SEPKVEPLSSGAELDALAKHVVAEAKLWEQLONKMEVMRSLAARMEHPIFRRKRSPAPA
(FE174994) SEPKVEPLSSGAELDALAKHVVAEAKLWEQLONKMEVMRSLAARMEHPIFRRKRSPAPA
(FE179552) SEPKVEPLSSGAELDALAKHVVAEAKLWEQLONKMEVMRSLAARMEHPIFRRKRSPAPA
(FE190948) SEPKVEPLSSGAELDALAKHVVAEAKLWEQLONKMEVMRSLAARMEHPIFRRKRSPAPA
(FE177552) SEPKVEPLSSGAELDALAKHVVAEAKLWEQLONKMEVMRSLAARMEHPIFRRKRSPAPA
(FE189608) SEPKVEPLSSGAELDALAKHVVAEAKLWEQLONKMEVMRSLAARMEHPIFRRKRSPAPA
(FE175049) SEPKVEPLSSGAELDALAKHVVAEAKLWEQLONKMEVMRSLAARMEHPIFRRKRSPAPA
(FE177553) SEPKVEPLSSGGGLDALPKHVVAEAKLWEQLONKMEVMRSLAARMEHPIFRRKRSPAPA
(FE189609) SEPKMGPPSSGAELDAPPKHVVAEAKLWEQLONKMGV-----
(FE187520) SEPKVEPLSSG-----
(FE180576) ----- GPLPKFVVPETQLWEQLPNKMEVMRSLAARMEHPIFRRKRSPAPA
(FE175050) ----- AKLWEQLPNKMEVMRSLAARMEHPIFRRKRSPAPA

(FE173081) HAPQAAAPSCNQOMDTSSEED-
(FE187476) HAPQAAAPSCNQOMDSDKQ*
(FE174994) HAPQAAAPSCNQOMDSDKQ*
(FE179552) HAPQAAAPSCNQOMDSDKQ*
(FE190948) HAPQAAAPSCNQOMDSDKQ*
(FE177552) HAPQAAAPSCNQOMDSDKQ*
(FE189608) HAPQAAALSCNQOMDSDKQ*
(FE175049) HAPQAAAPLPKE-
(FE177553) HAPQAAAPSCNQOMDSDKQ*
(FE189609) -----
(FE187520) -----
(FE180576) HAPQAAAPSCNQOMDTAKQ*
(FE175050) HAPQAAAPSCNQOMDSDKQWGRM-

```

(B) Prepro-crustacean hyperglycemic hormone

```

(FE101547) MVSFLSLRMVCSAALVSL LALSSRSFAFARSVDGVRLEKLLSSSSSSSGSSSPDLALG
(FE057303) MVSFLSLRMVCSAALVSL LALSSRSFAFARSVDGVRLEKLLSSSSSSSGSSSPDLALG

(FE101547) GDHSVNRDFTFDHSCGKIYDRELFRKLDRCVCEDCYNLYRKPVYATECKSNCFANFVFRQC
(FE057303) GDHSVNRDFTFDHSCGKIYDRELFRKLDRCVCEDCYNLYRKPVYATECKSNCFANFVFRQC

(FE101547) LDDLMLVDAIDEYVNTVQLVGK*
(FE057303) LDDLMLVDAIDEYVNTVQLVGK*

```

Figure 2.

Deduced amino acid sequences of *L. vannamei* crustacean cardioactive peptide (CCAP)- and crustacean hyperglycemic hormone (CHH)-encoding prepro-hormones. (A) CCAP precursors. (B) CHH precursors. Accession nos. of the ESTs from which the prepro-hormones were predicted are shown to the left, with the deduced amino acid sequences of the precursor proteins shown on the right. The predicted signal peptides, when present, are shown in gray, with all predicted prohormone convertase cleavage sites shown in black. Isoform(s) of the named peptide family, *i.e.* (A) CCAP or (B) CHH, are shown in red, while all other precursor-related peptides are shown in blue. Asterisks indicate the presence of a stop codon. Amino acid residues that vary between the precursors of a given family are highlighted in yellow.

(A) Prepro-myosuppressin

(FE188748) -TSPQNTTTQWCSASSSWFSLLVGVVVVLGVFAGVCEAMPVPVCTDQKLPLSPYAQKLCA
ALINIADFSRAMEDYLDKAVIRNSMFINPEV**KRQDL**DHVFLRFGRSQQ*

(B) Prepro-neuroparsin

(FE056059) MGISTSAVACLTACLT**LFHF**QAAQGT**TPVCPGTRDPPQDL**SKCKLG**VVKD**WCRNTVCAKG
(FE068047) MGISTSAVACLTACLTLLHLQAAQGT**TPVCPGTRDPPQDL**SKCKFG**VVKD**WCRNTVCAKG
(FE060966) MGISTSAVACLTACLTLLHLQAAQGT**TPVCPGTRDPPQDL**SKCKFG**VVKD**WCRNTVCAKG
(FE056058) MGISTSAVACLTACLTLLHLQAAQGT**TPVCPGTRDPPQDL**SKCKFG**VVKD**WCRNTVCAKG

(FE056059) **PRETCGGRWLEHGRCGLMYCRGHCAGCTSTLECVLGRFC***
(FE068047) **PRETCGGRWLEHGRCGLMYCRGHCAGCTSTLECVLGRFC***
(FE060966) **PRETCGGRWLEHGRCGLMYCRGHCAGCTSTLECVLGRFC***
(FE056058) **PRETCGGRWLEHGRCGLMYCRGHCAGCTSTLECVLGRFC***

(C) Prepro-proctolin

(FE183480) MSQKSHCTSRANTDLTTPYKMCRSSVFIVLGLVVMVAASQA**RYLPT**RADDSRLDEIREL
IRELLERTAE~~GGNNRMMANS~~GYEK**RYL**PK**RS**AVEGSSNNNNNLSNLGGEVATPLIN
LPQ*

(D) Prepro-SIFamide

(FE187321) MSVQARVVLAVLVVLAVFT**TEPV**SA**GYRKP**PFNGS**IFG**K**RS**GGDAVYEPKALASACQ
IAVEACAGWFP**GP**EKK*

Figure 3.

Deduced amino acid sequences of *L. vannamei* myosuppressin-, neuroparsin-, proctolin- and SIFamide-encoding prepro-hormones. (A) Myosuppressin precursor. (B) Neuroparsin precursors. (C) Proctolin precursor. (D) SIFamide precursor. Accession nos. of the ESTs from which the prepro-hormones were predicted are shown to the left, with the deduced amino acid sequences of the precursor proteins shown on the right. The predicted signal peptides, when present, are shown in gray (and pink in the case of one signal sequence prediction for the SIFamide precursor; see Results), with all predicted prohormone convertase cleavage sites shown in black. Isoform(s) of the named peptide family, *i.e.* (A) myosuppressin, (B) neuroparsin, (C) proctolin or SIFamide, are shown in red (and the pink sequence in the case of one signal sequence prediction for the SIFamide prepro-hormone; see Results), while all other precursor-related peptides are shown in blue. Asterisks indicate the presence of a stop codon. Amino acid residues that vary between the precursors of a given family are highlighted in yellow.

Figure 4.

The combined use of ESI QTOF MS/MS and MALDI FTMS for peptide identification in *L. vannamei*. (A) Collision-induced dissociation spectrum of a novel B-type allatostatin peptide STNWSNLRGTWamide from LC-MS/MS of ventral nerve cord extract via ESI QTOF. (B) Accurate mass profiling of an orckinin peptide NFDEIDRAGMGFA from an HPLC fraction of the brain extract by MALDI FTMS. Inset shows tandem MS fragmentation spectrum supporting the identification of peptide by accurate mass measurement.

Figure 5. Collision-induced dissociation spectra of two *de novo* sequenced peptides. ESI-Q-TOF MS/MS sequencing of two FaRPs: (A) GYSNKNFVRamide (615.82²⁺) and (B) APERNFLRFamide (574.71²⁺).

Figure 6. Collision-induced dissociation spectra of three *de novo* sequenced peptides. ESI-Q-TOF MS/MS sequencing of three pyrokinins: (A) DFSFNPRamide (497.76²⁺), (B) ADFAFSPRamide (511.75²⁺) and (C) SYFIPRamide (447.76²⁺).

Figure 7. Collision-induced dissociation spectra of a *de novo* sequenced tachykinin-related peptide APAGFLGMRamide (459.75^{2+}).

Table 1Bioinformatics of putative *Litopenaeus vannamei* peptide encoding expressed sequence tags (ESTs)

Queried peptide family (subfamily)	ESTs identified	Accession no.	Blast score *	E-value *
A-type allatostatin	–			
B-type allatostatin	–			
C-type allatostatin	+	FE182974	47.0	5e-06
		FE175093	47.0	6e-06
		FE182975	45.8	1e-05
		FE180026	33.5	0.067
Allatotropin	–			
Bursicon (α subunit)	+	FE187805	239	7e-64
		FE175634	237	2e-63
		FE173462	236	3e-63
		FE173463	84.7	2e-17
		FE173025	73.6	5e-14
Bursicon (β subunit)	+	FE178442	176	5e-45
		FE184710	98.2	2e-21
		FE184711	62.4	1e-10
Corazonin	+	FE154856	38.5	0.002
		FE154857	38.5	0.002
Crustacean cardioactive peptide	+	FE173081	146	5e-36
		FE187476	146	6e-36
		FE174994	146	6e-36
		FE179552	146	6e-36
		FE190948	146	6e-36
		FE177552	146	6e-36
		FE189608	146	6e-36
		FE175049	145	1e-35
		FE177553	132	1e-31
		FE189609	100	5e-22
		FE187520	80.1	6e-16
		FE180576	52.4	1e-07
		FE175050	47.4	4e-06

Queried peptide family (subfamily)	ESTs identified	Accession no.	Blast score *	E-value *
Crustacean hyperglycemic hormone/ion transport peptide (CHH-like)	+	FE101547	179	7e-46
		FE057303	177	2e-45
Crustacean hyperglycemic hormone/ion transport peptide (MIH-like)	-			
Diuretic hormone (calcitonin-like)	-			
Diuretic hormone (corticotropin-releasing factor-like)	-			
Ecdysis-triggering hormone	-			
Eclosion hormone	-			
FMRFamide-related peptide (F/YLRFamide)	-			
FMRFamide-related peptide (myosuppressin)	+	FE188748	45.1	2e-05
FMRFamide-related peptide (neuropeptide F)	-			
FMRFamide-related peptide (short neuropeptide F)	-			
FMRFamide-related peptide (sulfakinin)	-			
Insect kinin	-			
Intocin	-			
Neuroparsin	+	FE056059	65.1	2e-11
		FE068047	64.3	3e-11
		FE060966	64.3	3e-11
		FE056058	64.3	3e-11
Orcokinin	-			
Pigment dispersing hormone	-			
Proctolin	+	FE183480	38.5	0.002
Pyrokinin	-			

Queried peptide family (subfamily)	ESTs identified	Accession no.	Blast score *	E-value *
Red pigment concentrating hormone	–			
SIFamide	+	FE187321	96.3	7e-21
Tachykinin-related peptide	–			

* It should be noted that many of the bioactive peptides search for in this study are small (<20 amino acids) and homology/identity between the prepro-hormones containing them is often limited to the sequences of these small peptides. This combination of factors automatically leads to low BLAST scores/high E-values, and in several cases, the scores reported here are low and high, respectively, however this should not be considered unusual for neuropeptide transcripts.

Table 2

Predicted structures of mature *Litopenaeus vannamei* peptide paracrine/hormones identified via transcriptomics

Table 2. Predicted structures of mature <i>Litopenaeus vannamei</i> peptide paracrine/hormones identified via transcriptomics	
Family (subfamily)	Predicted mature peptide structure
C-AST	pQIRYHQCFNPISCF pQIRYHQCFIPVSCF
C-AST-PRP	SPAFQDKPLEAANQFQHPVHLQ AAPQDSSFEQDLAAFKDLDVAQVAEL SAQEPFSA TLVDEDEGLEEQAAQS MLAPLSGLPGELPTI MFAPLSGLPGELPTI -DELPTI
Bursicon α	DECSLTPVIHILSYPGNSKPIPSFACQGRCTSYVQVSGSKIWQTERSCMCQESGEREASVTLNCPKARPGEPMRKILTRAPIDCMCRPCTDVEEGTVLAQEIANTFEDSPMENVPFL DECSLTPVIHILSYPGNSKPIPSFACQGRCTSYVQVSGSKIWQTERSCMCQESGEREASVTLNCPKARPGEPMRKILTRAPIDCMCRPCTDVEEGTVLAQEIANTFEDSPMENVPFL DECSLTPVIHILSYPGNSKPIPSFACQGRCTSYVQVSGSKIWQTERSCMCQESGEREASVTLNCPKARPGEPMRKILTRAPIDCMCRPCTDVEEGTVLAQEIANTFEDSPMENVPFL -MRKILTRAPIDCMCRPCTDVEEGTVLAQEIANTFEDSPMENVPFL
Bursicon β	DQCSLTPVIHILSYPGNSKPIPSFACQGRCTSYVQV- GFSRAHTYGSECETLPSTMHVAREEFDAGRLVRTCEDLAVNCKEGACVSKVQPSVNTPSGFLKDCRCRETHLRARDVVLTHCCDGNRITGDNGKLTVKLRPAELP- -FLKDCRCRETHLRARDVVLTHCYDGNRITGDNGKLTVKLRPAELPDCQCFKCGNSI -DVGLTPCYDGNRITGDNGKLTVKLRPAELPDCQCFKCGNSI
Corazonin	pQTFCISRGWTNA
Corazonin-PRP	RSDMSGVARTVNPLSPFLTRVLPVSLHQAAYNPPVKTVEERLRALEAGFAVIRASSANLSSGGDEYYSEN RSDMSGVARTVNPLSPFPTRVLPVSLHQAAYNPPVKTVEERLRALEAGFAVIRASSANLSSGGDEYYSEN
CCAP	PFCAFTGCa PFCAFTGCa DIADLLDGKDKSPFCAPFGFa
CCAP-PRP	GVA DIADLLDGKD RSEPKVEPLSSGAEALDALAKHVVAEAKLWEQLQNKMEVMRSLAA RSEPKVEPLSSGGGLDALPKHVVAEAKLWEQLQNKMEVMRSLAA RSEPKMGFPSSGAEALDALPKHVVAEAKLWEQLQNKMGV- -GFLPKPVVETQLWEQLQNKMEVMRSLAP -AKLWEQLQNKMEVMRSLAA MEHPIF MEHPIF SPAPAHAPQAAAPSONQMDTDQ SPAPAHAPQAAAPSONQMDTDQ SPAPAHAPQAAAPSONQMDTAKQ SPAPAHAPQAAAPSONQMDTSSSEED- SPAPAHAPQAAAPSONQMDTDQWGRM- SPAPAHAPQAAAPLKKP-
CHH	DTFDHSCGIYDRELFRKLDVCEDCYNLYRKPYVATECKSNYANFVKQLDLDLLMVDAIDEVYNTVQLVa
CHH-PRP	RSVDGVRLEKLLSSSSSSSSSSPLDALGGDHSV
FMRFamide (MS)	pQDLHVFRLFa
MS-PRP	-TSPQNTTTQWCSASSWSFLVGVVVLGVFAGVGEAMPFPVCTDQKLPLSPYAQLCAALINIADFSRADEY ₁₀₀ LDKAVIRNSMPINEPEV SQQ
Neuroparsin	TPVCFGTDRPQDLCKKLGVVVDWCRNTVCAGPRETCGGRWLEHGRCLGMYCRGHCAGCTSTLECVLGRFC TPVCFGTDRPQDLCKKLGVVVDWCRNTVCAGPRETCGGRWLEHGRCLGMYCRGHCAGCTSTLECVLGRFC
Proctolin	RYLPT
Proctolin-PRP	ADDSRLDEI ELIRELLE TAEGGNRMWANSQYE YLF SAVPEGSSNNNNNNLSNLGGEVATPLINLPQ EPVSAGYRPPFNGSIFa GYRPPFNGSIFa
SIFamide	
SIFamide-PRP	SGGDVAYEPSKALASACQIAYEACAGWTFGPe

Abbreviations: a, amide; C-AST, C-type allatostatin; CCAP, crustacean cardioactive peptide; CHH, crustacean hyperglycemic hormone; MS, myosuppressin; PRP, precursor-related peptide.

– indicates the presence of a putative partial peptide with the dash indicating the end of the peptide likely missing an unknown number of amino acids.

Peptides denoted in red are peptides predicted here for the first time.

Peptides denoted in blue are ones previously identified in other species, but are described here for *L. vannamei* for the first time.

Peptides denoted in bold font were also identified via mass spectrometry (see Table 3).

Table 3

Litopenaeus vannamei neuropeptides identified via mass spectrometry

Family	m/z	Sequence	MALDI FTMS			QTOF MS/MS		
			EG	Br	VNC	EG	Br	VNC
A-type AST	850.42	HGSYAFGLa	-	-	-	-	+	-
	882.45	ANQYAFGLa	-	-	-	-	+	-
	953.52	DRLYAFGLa	-	-	-	-	+	-
	968.52	SSKPYAFGLa	-	-	-	-	+	-
B-type AST	1031.55	KWAAAGRSaW _a	-	-	-	-	-	+
	1165.55	NWNKFQGSW _a	-	+	+	+	-	-
	1180.60	RWSKFQGSW _a	-	-	-	-	-	+
	1237.57	ADWNKFQGSW _a	-	-	-	-	-	+
	1265.64	LTWNKFQGSW _a	-	-	+	-	-	+
	1291.62	SADWNSLRGTW _a	-	-	+	-	-	+
	1320.64	STNWSNLRGTW _a	-	-	+	-	-	+
	1470.70	VPNDWAHFRGSW _a	-	-	-	-	+	-
Corazonin	1369.64	pQTFQYSRGWTNa	-	+	+	-	+	-
CCAP	956.37	PFCNAFTGCa	-	+	+	-	+	-
CCAP-PRP	1074.53	DIADLLDGKD	-	-	-	-	+	+
FaRP	887.56	PSLRLRFa	-	-	-	-	-	+
	905.51	PSMRLRFa	-	+	-	+	-	+
	965.54	NRNFLRFa	-	+	-	-	-	+
	1023.55	DGRNFLRFa	-	-	+	-	+	+
	1104.61	GAHKNYLRFa	+	-	+	+	-	-
	1105.63	SMPSLRLRFa	-	-	+	-	-	+

Family	m/z	Sequence	MALDI FTMS			QTOF MS/MS		
			EG	Br	VNC	EG	Br	VNC
Orcokinin	1121.66	SM(O)PSLR ¹ LR ² Fa	-	-	+	-	-	-
	1148.63	APER ¹ NFLR ² Fa	-	+	+	-	+	+
	1181.62	SEN ¹ RNFLR ² Fa	-	+	-	+	-	-
	1230.64	GYS ¹ NKNFVR ² Fa	-	+	+	-	+	-
	1231.62	GYS ¹ NKDFVR ² Fa	-	-	-	-	+	-
	1271.67	pQDL ¹ DHVF ² LR ³ Fa	-	-	+	-	+	+
	1288.68	QDL ¹ DHVF ² LR ³ Fa	-	-	+	-	+	+
	1300.76	DGR ¹ TPAL ² LR ³ Fa	-	-	+	-	-	-
	979.45	NFDEIDRA	-	+	-	-	+	+
	1036.47	NFDEIDRAG	-	-	-	-	+	-
	1098.52	EIDRS ¹ GGF ² GFA	-	+	-	-	-	-
	1110.49	FDEIDRAG ¹ MG	-	-	-	-	+	-
	1182.55	NFDEIDRAG ¹ Fa	-	-	-	-	+	-
	1183.54	DFDEIDRAG ¹ Fa	-	-	-	-	+	-
	1197.52	DEIDRA ¹ GM(O) ² GFA	-	-	-	-	+	-
	1198.55	NFDEIDRS ¹ GFa	-	+	-	-	-	-
	1256.55	NFDEIDRS ¹ GF ² G	-	+	-	-	-	-
	1270.57	NFDEIDRS ¹ GFA	-	+	-	-	-	-
	1344.59	FDEIDRA ¹ GM(O) ² GFA	-	-	-	-	+	-
	1387.49	NFDEIDRAG ¹ (O) ² GF	-	-	-	-	+	-
	1442.64	NFDEIDRAG ¹ MGFA	-	+	+	+	+	+
	1458.63	NFDEIDRA ¹ GM(O) ² GFA	-	+	+	+	+	+
	1474.66	NFDEIDRS ¹ GGF ² GFA	-	-	-	-	+	-
	1500.71	NFDEIDRAG ¹ GF ² FL	-	-	+	-	-	+
	1501.67	NFDEIDRAG ¹ GF ² FN	-	-	+	-	+	-
	1517.67	NFDEIDRS ¹ GGF ² FN	-	+	+	-	+	-
	1554.70	NFDEIDRT ¹ GF ² GF ³ H	-	-	-	-	+	-

Family	<i>m/z</i>	Sequence	MALDI FTMS			QTOF MS/MS		
			EG	Br	VNC	EG	Br	VNC
Oromyotropin	1186.52	FDAFTTGFGHS	-	+	-	+	+	+
Pyrokinin	894.52	YSFLPRL Amide	-	-	-	-	+	+
	951.50	DFAFSPRLAmide	-	-	-	-	+	+
	978.52	DFAFNPR LAmide	-	-	-	-	+	-
	994.51	DFSFNPR LAmide	-	-	-	-	+	-
	1008.53	GDAFSPRL Amide	-	-	-	-	+	-
	1022.54	ADFAFSPRL Amide	-	-	-	-	+	-
	1035.54	GDAFNPR LAmide	-	-	-	-	+	-
	1037.55	SGGFAFSPRL Amide	-	+	-	-	+	-
	1049.55	ADFAFNPR LAmide	-	-	-	-	+	+
RPCH	930.46	pELNFS PGW ^a	-	-	-	+	-	-
RYamide	976.46	SGFYANRY ^a	-	+	+	-	-	+
	1030.47	pEGFYSQRY ^a	-	+	+	-	+	-
SIFamide	1161.65	RKPPENG SIF ^a	-	-	-	-	+	+
	1381.74	GYRKPPENG SIF ^a	+	+	+	+	+	+
TRP	918.50	APAGFLGM R ^a	-	-	+	-	+	+
	934.49	APAGFLGM(O) R ^a	+	-	-	+	+	-
	934.49	APSGFLGM ^a R ^a	-	+	+	-	+	+
	950.49	APSGFLGM(O) R ^a	-	-	-	+	+	+
	951.45	APSGFNGM(O) R ^a	-	-	-	-	+	-
	992.50	APSGFLGM IRG	-	-	-	+	-	+
	1008.49	APSGFLDM(O) R ^a	-	-	-	-	+	-

			MALDI FTMS				QTOF MS/MS			
Family	<i>m/z</i>	Sequence	EG	Br	VNC	EG	Br	VNC		
Other	1505.69	L/IPEPDPMAEAGHEL/I	-	-	-	+	-	-	-	

Abbreviations: MALDI FTMS, matrix-assisted laser desorption/ionization Fourier transform mass spectrometry; QTOF MS/MS, electrospray ionization quadrupole time-of-flight tandem mass spectrometry; EG, eyestalk ganglia (including the sinus gland); Br, brain; VNC, ventral nerve cord (including both the thoracic and abdominal ganglia); a, amide; AST, allatostatin; CCAP, crustacean cardioactive peptide; FaRP, FMRFamide-related peptide; PRP, precursor-related peptide; RPCH, red pigment concentrating hormone; TRP, tachykinin-related peptide.

Peptides denoted in red are peptides identified here for the first time.

Peptides denoted in blue are ones previously identified in other species, but are described here for the first time in *L. vannamei*.

Peptides denoted in black are previously known *L. vannamei* peptides.

Peptides denoted in bold font were also predicted via transcriptome mining and bioinformatics (see Table 2).

+ indicates presence and – indicates absence.