Published in final edited form as:
J Proteomics. 2013 October 8; 91: . doi:10.1016/j.jprot.2013.06.021.

A Multi-scale Strategy for Discovery of Novel Endogenous Neuropeptides in the Crustacean Nervous System

Chenxi Jia, **Christopher B. Lietz**, **Hui Ye**, **Limei Hui**, **Qing Yu**, **Sujin Yoo**, and **Lingjun Li***
School of Pharmacy and Department of Chemistry, University of Wisconsin–Madison, WI

Abstract

The conventional mass spectrometry (MS)-based strategy is often inadequate for the comprehensive characterization of various size neuropeptides without assistance of genomic information. This study evaluated sequence coverage of different size neuropeptides in two crustacean species, blue crab *Callinectes sapidus* and Jonah crab *Cancer borealis* using conventional MS methodologies and revealed limitations to mid- and large-size peptide analysis. Herein we attempt to establish a multi-scale strategy for simultaneous and confident sequence elucidation of various sizes of peptides in the crustacean nervous system. Nine novel neuropeptides spanning a wide range of molecular weights (0.9-8.2 kDa) were fully sequenced from a major neuroendocrine organ, the sinus gland of the spiny lobster *Panulirus interruptus*. These novel neuropeptides included seven allatostatin (A- and B-type) peptides, one crustacean hyperglycemic hormone precursor-related peptide, and one crustacean hyperglycemic hormone. Highly accurate multi-scale characterization of a collection of varied size neuropeptides was achieved by integrating traditional data-dependent tandem MS, improved bottom-up sequencing, multiple fragmentation technique-enabled top-down sequencing, chemical derivatization, and *in silico* homology search. Collectively, the ability to characterize a neuropeptidome with vastly differing molecule sizes from a neural tissue extract could find great utility in unraveling complex signaling peptide mixtures employed by other biological systems.

Keywords

Neuropeptide; peptidomics; *de novo* sequencing; mass spectrometry; crustacean

INTRODUCTION

Neuropeptides, including endogenous peptide neuromodulators and hormones, mediate or modulate neuronal communication by acting on cell surface receptors and are involved in a broad range of physiological and behavioral processes [1-4]. Mass spectrometry (MS)-based neuropeptidomics aims to completely characterize the neuropeptides in a target organism as an important first step toward a better understanding of the structure and function of these complex signaling molecules [5-17]. A significant challenge to this goal is that many of these endogenous neuropeptides display large diversity at the molecular and cellular level, such as various molecular sizes [15], extensive and multiple post-translational modifications

***Correspondence:** Professor Lingjun Li, School of Pharmacy and Department of Chemistry, University of Wisconsin–Madison, 777 Highland Avenue, Madison, Wisconsin 53705-2222, lli@pharmacy.wisc.edu, **Fax:** +1-608-262-5345 .

(PTMs) [18], different hydrophobicities [12], and a wide dynamic range of concentrations [19]. Because of this, a uniform approach for comprehensive neuropeptide characterization is difficult to engineer.

A neuropeptidome usually contains peptides of various sizes [1]. Although liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) with data-dependent acquisition is a powerful tool in peptidome research [5, 6, 11-13, 17], it lacks the capability for *de novo* sequencing of mid-size and large peptides due to inefficient fragmentation of peptides larger than 4 kDa [20]. Innovation of MS-based instrumentation has greatly advanced our capability to analyze complex peptide mixtures, including those large peptides and intact proteins with extensive modifications [21]. For example, emergence of ultra-high resolution Fourier transform mass spectrometers greatly facilitates the bioinformatics-assisted peptidome research, as mass accuracy is a critical factor in scoring algorithm for peptide sequence assignment and protein identification [12, 17]. The observed mass values of peptide precursors and their resulting fragment ions are matched to the theoretical values arising from a genome/cDNA sequence, and thus partial peptide sequence coverage from interpretation of MS/MS spectra can usually produce confident hits [10]. However, there are many valuable animal models whose genomes have not been sequenced yet, thus no genomic database or large cDNA database is available for database searching strategy commonly used [1, 3]. Peptide discovery in these model organisms would need to rely on obtaining full sequence coverage, including enhanced local identification confidence on each amino acid residue.

The California spiny lobster *Panulirus interruptus* has long served as an important animal model for many areas of research in endocrinology and neurobiology [22-24], but its genome has not been sequenced yet and there is no protein/cDNA database. In particular, its stomatogastric ganglion has been utilized as a powerful model system to understand the cellular mechanisms of rhythmic pattern generation in neuronal networks [22]. Many studies reported that neurotransmitters and neuropeptides regulate the functional output of these well-defined neuronal circuits [25-28]. Therefore, it is highly desirable to obtain accurate molecular information on neuropeptides in this species.

In a previous study [29], we established a strategy by combining bottom-up, off-line top-down, and on-line top-down MS methods for confident *de novo* sequencing of crustacean hyperglycemic hormone (CHH)-family neuropeptides with molecular weight (MW) around 8.4-9.2 kDa. The current work aims to discover and confidently identify various sizes of novel neuropeptides (MW 0.9-8.2 kDa) in the crustacean nervous system. We evaluate the possibility for improvement of current approaches on sequencing of small-, middle- and large-molecular sizes of neuropeptides. A multi-scale strategy is established by rational optimization of methodology and further validated by sequencing of nine novel neuropeptides in *P. interruptus* sinus gland, a major neuroendocrine structure that secretes peptide hormones to regulate many essential functions of the animal.

MATERIALS AND METHODS

Chemicals

Methanol, glacial acetic acid, borane pyridine and formaldehyde were obtained from Sigma-Aldrich (St. Louis, MO). Optima grade formic acid, acetonitrile (ACN), water, and methanol were purchased from Fisher Scientific (Pittsburgh, PA).

Animals and Tissue Dissection

Blue crabs *Callinectes sapidus* and Jonah crabs *Cancer borealis* were shipped from the Fresh Lobster Company (Gloucester, MA); and the California spiny lobsters were purchased from

Catalina Offshore Products (San Diego, CA). Tissue dissection was performed according to our previous report [30]. Briefly, ten animals were anesthetized in ice, and the sinus glands were dissected and collected in chilled acidified methanol and stored in −80 °C freezer prior to further sample processing.

Tissue Extraction and LC Fractionation

The tissues were homogenized and extracted with 100 μL of acidified methanol (methanol:H_2O:acetic acid, 90:9:1, v:v:v) [30, 31]. After centrifugation, supernatant fractions were combined and the extraction procedure was repeated three times. After dryness in a Savant SC 110 SpeedVac concentrator (Thermo Electron Corporation, CA), the sample was re-suspended in 100 μL of deionized water for further analysis. High-performance liquid chromatography (HPLC) separations were performed with a Waters Alliance HPLC system (Milford, MA) according to our previous report [31]. The mobile phases included solution A (water containing 0.1% formic acid) and solution B (ACN containing 0.1% formic acid). Approximately 50 μL of extract was injected onto a Phenomenex Gemini C18 column (2.1 mm i.d., 150 mm length, 5 μm particle size; Torrance, CA). The separations consisted of a 120 min gradient of 5–95% solution B. The flow rate was 0.2 mL/min. Fractions were automatically collected every 2 min, followed by drying in SpeedVac, re-suspended in 40 μL water, and stored in −80 °C.

Formaldehyde Labeling, and Peptide Digestion

For formaldehyde labeling [14], 1 μL of peptide fraction was mixed with 9 μL of ACN with 0.1% formic acid, followed by the addition of 1 μL of formaldehyde (4% in H_2O) and 1 μL of borane pyridine (120 mM in methanol). The labeling reaction was allowed to take place for 10 min with mixing on Vortex. Excess formaldehyde was quenched via the addition of 1 μL of ammonium bicarbonate buffer (0.2 M). The resulting solution was stored at −20 °C before further analysis.

For trypsin digestion of CHHs, 1 μL of peptide fraction was reduced and alkylated by incubation in 2.5 mM Dithiothreitol (DTT) for 1 h at 37 °C followed by incubation in 7 mM iodoacetamide (IAA) in the dark at room temperature for 1 h. A 1 μL of peptide was digested at 37 °C overnight after addition of 50 mM ammonium bicarbonate buffer with 0.5 μg of trypsin (Promega, Madison, WI). For Lys-C digestion of CPRPs, 1 μL of peptide fraction was digested overnight at 37 °C after re-suspending in 10μL of 50 mM Tris-HCl, pH 8.0, and addition of 0.5 μg of Lys-C (Princeton Separations, Adelphia, NJ). Finally, each digest was quenched by the addition of formic acid to a final concentration of 0.5%, desalted on a C18 ZipTip (Millipore, Bedford, MA), and the eluent was further dried in SpeedVac.

Nano-LC-Electrospray (ESI)-Linear Trap Quadrupole (LTQ)-Orbitrap Elite

A 1 μL of crude tissue extract was reduced by incubation in 2.5 mM DTT for 1 h at 37 °C and desalted by C18 ZipTip and resuspended in 10 μL of water containing 0.1% formic acid. A 0.5 μL of peptide sample was injected onto an Ultimate 3000 RSLCnano system coupled to an Orbitrap Elite mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) for data-dependent MS/MS analysis with electron-transfer dissociation (ETD) and higher-energy collisional dissociation (HCD) according to our previous report [29].

NanoLC-ESI-Quadrupole Time of Flight (QTOF)-MS/MS

Nanoscale LC-ESI-Q-TOF MS/MS was performed on a Waters nanoAcquity ultra performance LC system coupled to a Synapt G2 high-definition mass spectrometer (Waters Corp., Milford, MA). Chromatographic separations were performed on a Waters BEH 130Å C18 reversed phase capillary column (150 mm × 75 μm, 1.7 μm). The mobile phases used

were: 0.1% formic acid in deionized water (A); 0.1% formic acid in ACN (B). An aliquot of 2.0 μL of HPLC fraction was injected and loaded onto the Waters NanoEase trap column using 95% mobile phase A and 5% mobile phase B at a flow rate of 10 μL/min for 3 min. For neuropeptides, the linear gradient was from 5 to 45% buffer B over 75 min. A data dependent acquisition was employed for the MS survey scan and the selection of three precursor ions and subsequent MS/MS of the selected parent ions. The MS scan range was from m/z 400–2000 and the MS/MS scan was from *m/z* 50–2000.

Top-down MS/MS on ESI-LTQ- Fourier Transform Ion Cyclotron Resonance (FTICR) mass spectrometer

A 0.5 μL of crustacean hyperglycemic hormone (CHH) peptide fraction was reduced by incubation in 2.5 mM DTT for 1 h at 37 °C and desalted by C18 ZipTip and resuspended in 10 μL of 50% ACN containing 2% formic acid. The DTT-reduced CHH sample was directly infused into a 7T LTQ-FTICR Ultra hybrid mass spectrometer (Thermo Scientific Inc., Bremen, Germany) equipped with an automated chip-based nano-ESI source (Triversa NanoMate; Advion BioSciences, Ithaca, NY) as described previously [29]. For collision-induced dissociation (CID) and electron-capture dissociation (ECD) fragmentation, individual charge states of peptide molecular ions were first isolated and then dissociated using 22–28% of normalized collision energy for CID or 4% electron energy for ECD with a 60 ms duration with no delay. Typically, 1000 transients were averaged to ensure high quality MS/MS spectra. All FTICR spectra were processed with Xtract Software (Xcalibur 2.1, Thermo Scientific Inc., Bremen, Germany) using a S/N threshold of 1.5 and fit factor of 40% and validated manually. The resulting mass lists were further assigned using the in-house developed “Ion Assignment” software [32]. The assigned ions were manually validated to ensure the quality and accuracy of assignments. The analysis of CHH-precursor related peptides (CPRPs) employed the same protocol of CHHs, except for no DTT reduction of disulfide bonds.

In Silico Homology Search

For small neuropeptides, the resulting sequence was put in NCBI tBLASTn search engine against nucleotide collection database (no. of sequence 16341740). Crustacean (taxid: 6657) was chosen as organism [16].

For mid-size neuropeptides, the raw data from ESI-LTQ-FTICR were deisotoped with Xtract v2.1 (ThermoFisher, Bremen, Germany) and uploaded in a custom 168-core ProSightPC 3.0 cluster [17, 20]. The data was searched in a “Sequence Tag Search” mode against a home-built crustacean neuropeptide database, in which minimum tag score was set at 0.01, and compiler tolerance was 10 ppm. A minimum tag size of 4 residues was required. The database was created with 693 previously identified crustacean neuropeptides.

For large neuropeptides, the LC-ESI-QTOF-MS/MS raw data were converted to peak list (.pkl) files using ProteinLynx software 2.4 (Waters). Peptides were identified by searching against an NCBInr protein database 20120519 (18132328 sequences; 6219145704 residues) using the Mascot v2.4 search engine. Trypsin was selected as enzyme allowing up to 2 missed cleavages. Carboxylmethyl cysteine was specified as fixed modifications, and methionine oxidation and pyro-Glu as variable modifications. Precursor and MS/MS tolerances were set within 0.5 Da and 0.5 Da for monoisotopic mass, respectively. Peptide charge states include 1+, 2+ and 3+ charged peptides.

RESULTS AND DISCUSSION

Attempts for Rational Improvement of Methodology

In the past decade, there has been a great success with mass spectrometry (MS)-based neuropeptidomics [1, 5-11, 15, 33, 34]. However, the majority of these published peptide lists show excellent coverage rates on small neuropeptides (< 2.5 kDa), but few identifications in the higher mass range because the traditional strategies are inefficient for *de novo* sequencing of mid-size (2.5-5 kDa) and large (>5 kDa) neuropeptides. The first aim of this study is to evaluate the performance of traditional strategies on their sequence coverage of crustacean neuropeptides from previously studied model organisms. The newly released Thermo LTQ-Orbitrap Elite instrument [35] has demonstrated powerful performance in both bottom-up and top-down proteomics. This instrument platform provides ultra-high resolving power and allows multiple complementary fragmentation methods (ETD, HCD, and CID) to be performed on liquid chromatographic time scales [35]. Therefore, it is tempting to employ such a state-of-the-art analytical platform for a comprehensive sequencing analysis of various sizes of neuropeptides using data-dependent LC-MS/MS with multiple fragmentation techniques. Here, we evaluated two neuropeptide extracts from the sinus glands of *Callinectes sapidus* [31] and *Cancer borealis* [15]. The former extract contained $^{Cas-}$Orc (*Callinectes sapidus, Cas;* orcokinin, Orc; MW 1474.64 Da), $^{Cas-}$CPRP (3836.99 Da), and $^{Cas-}$CHH (8472.80 Da); and the latter sample yielded $^{Cab-}$Orc (*Cancer borealis, Cab*; MW 1473.66 Da), $^{Cab-}$CPRP (3976.06 Da), and $^{Cab-}$CHH (8539.88 Da). Their sequences obtained on this new instrument platform agreed well with previously published sequences [15, 29-31, 36]. The two model samples were analyzed by LC-MS/MS with HCD and ETD, respectively. As shown in Figure 1A, 69% sequence coverage was obtained by HCD for the two small peptides orcokinins. However, for the two mid-size peptides, 64% and 37% were obtained by HCD, and 83% and 55% by ETD, respectively. For the large peptides, less than 32% of sequence was determined by either HCD or ETD. Because the mid- and large-sized peptide analysis yielded much lower sequence coverage, complete sequence elucidation or identification cannot be obtained in the absence of a genome or a cDNA database.

Compared with on-line LC-MS/MS, the off-line MS/MS method requires highly purified samples and lengthy data acquisition times, which lowers the analytical throughput. However, it can generate high quality MS/MS spectra for sequence interpretation [32]. Therefore, we evaluated whether or not off-line top-down MS could achieve complete sequencing of mid- and large-size peptides. The LC-purified $^{Cas-}$CPRP, $^{Cas-}$CHH, $^{Cab-}$CPRP, and $^{Cab-}$CHH were directly infused into an ESI-LTQ-FTICR mass spectrometer for top-down fragmentation by CID and ECD, respectively. Compared with the on-line LC-MS/MS results, this off-line strategy offers 16% increase of total sequence coverage for mid-size peptides and 50% increase for large peptides (Figure 1B). However, complete sequence coverage was not obtained by this off-line strategy, indicating its limitation. It should be noted that the off-line top-down MS can also be carried out on LTQ-Orbitrap using ETD and HCD with comparable results. In addition, because the CHHs contain disulfide bonds, the samples containing CHHs were reduced with DTT before analysis to obtain efficient MS fragmentation [29].

In this study, we established a multi-scale strategy for complete amino acid residue sequencing of various sizes of neuropeptides (Figure 2) based on the instrument platforms of LC-ESI-QTOF and ESI-LTQ-FTICR. We chose the *P. interruptus* sinus gland as a target neuroendocrine organ to explore the multi-scale strategy and peptide discovery. The sinus glands located in crustacean eyestalk synthesizes and secretes numerous peptide hormones regulating multiple physiological activities, such as molting, hemolymph glucose levels, osmoregulation, and integument color changes [37]. Therefore, the elucidation of novel

neuropeptide content will facilitate their functional studies. Table 1 lists nine such novel neuropeptides including one large, one mid-size and seven small peptides from this understudied model organism.

Small Neuropeptides

For *de novo* sequencing of small neuropeptides, we followed the conventional strategy [30]. The LC fractions #7, 8, and 9 (Figure 2) of *P. interruptus* sinus gland crude extract were analyzed by LC-MS/MS on an ESI-QTOF instrument. The resulting MS/MS spectra were processed by a software called Pepseq [31] for sequence interpretation, leading to identification of seven novel allatostatin (AST)-family neuropeptides listed in Table 1, $^{Pai-}$AST-An (*Pai, Panulirus interruptus*; A-type; n=1, 2, 3) and $^{Pai-}$AST-Bn (B-type; n=1, 2, 3, 4). A-type AST [15] is a peptide family possessing the C-terminal motif –YXFGLamide (X stands for a variable amino acid, and commonly either A or S), and the B-type AST-family peptides [38] possess the C-terminal motif −WX_6Wamide (X_6 are six variable amino acids). Figure 3 displays three representative *de novo* sequencing MS/MS spectra (others are shown in Figure S-3). To differentiate the isobaric (i.e., equal mass) residues (lysine vs. glutamine) in the four discovered B-type AST peptides, formaldehyde labeling [14, 39-41] reactions were performed to dimethylate the N-termini of the peptide chain and a putative lysine side chain. As shown in Figure 3D and 3E, there has 2×28 Da of mass increase after dimethylation, indicating that the fifth residue from N-terminus is lysine. Furthermore, the resulting peptide sequences were validated by BLAST homology search, with a match to autologous or homologous AST preprohomone genome (Table S-1). Interestingly, the $^{Pai-}$AST-A1 and $^{Pai-}$AST-A2 peptide were validated by the autologous AST preprohomone from *P. interruptus* with complete sequence match, confirming the sequence assignment and peptide identification. In addition, it should be noted that the MS approach used here could not differentiate leucine from isoleucine, so these residues were assigned according to homologous sequences by BLAST in this study.

The AST neuropeptide was first discovered and named based on their function of inhibiting juvenile hormone production in the corpora allata of insects [42, 43]. In crustaceans, ASTs regulate a range of important processes and can act as inhibitors of endocrine function, as neuromodulators, on muscle, and directly on metabolic pathways [43]. The A- and B-type ASTs possess similar physiological functions but exhibit distinct sequence diversity, suggesting that their functions have undergone convergent evolution, highlighting the importance of allatostatic substances in crustaceans.

Mid-size Neuropeptides

CPRP is produced concurrent with the cleavage of CHH from its preprohormone [30, 31, 44]. CPRP is located between the signal peptide and the CHH sequence and is separated from the CHH by a dibasic cleavage site. There are currently no reports on the physiological roles played by CPRPs in crustaceans.

In LC fraction #13 (Figure 2), a peptide candidate with MW 3468.68 was directly infused into ESI-LTQ-FTICR mass spectrometer for top-down ECD and CID fragmentation. Sequence Tag Search in ProSight is a powerful tool for homology search of mid-size peptide [20]. Specifically, sequence tags were obtained from continuous fragment ions in top-down MS/MS spectra, followed by subsequent matching against the homologous sequences in peptide database. Here, the ECD data of this putative peptide was searched against a home-built crustacean neuropeptide database, and resulted matches to several CPRP peptides from other crustacean species (Figure S-4), suggesting that this peptide belongs to CPRP family.

Deducing the peptide sequence solely based on the top-down ECD data may cause misidentification, as shown in the evaluation data acquired on LTQ-Orbitrap as described above. Because CPRP-family peptides contain multiple arginine and lysine, trypsin digestion produces too short tryptic peptides to be detected by MS, and also it is challenging to assemble multiple tryptic peptides from a mixture. Bonet-Costa et al. [45] reported an integrated top-down and bottom-up strategy to characterize histone proteins, where similar challenges exist in that the target proteins contain multiple basic amino acid residues. In our study, to sequence the mid-size CPRP peptides with multiple basic amino acid residues, we introduced a modified bottom-up *de novo* sequencing method. The $^{Pai-}$CPRP candidate peptide was allowed to react with formaldehyde [39], yielding a mass increase of 2×28 Da due to dimethylation on both the N-terminus and lysine side chain (Figure 4A and B), which indicated that this $^{Pai-}$CPRP contained one lysine. Consequently, after Lys-C digestion, the peptide is broken into two parts, and the proteolytic peptides can be identified by assembling peptide pair which yields total mass of 3467.08 + 18.01 = 3485.09 Da. As we expected, two proteolytic peptides were observed with MW 1230.75 and 2254.24 Da (sum = 3484.99 Da). Interpretation of their MS/MS spectra by Pepseq generated amino acid sequences as ^{1}RSSSGLVRLEK^{11} (Figure 5A) and ^{12}LLSSRSSSSTPLGLLSADHNVN^{33} (Figure 5B). Accordingly, the top-down ECD spectrum in Figure 5C was confidently annotated, resulting in assembly of this $^{Pai-}$CPRP as shown in Figure 5D. Figure 8A illustrates the fragmentation maps by both top-down ECD and CID showing 91% of sequence coverage, which further confirms the sequence resulting from the modified bottom-up method. Subsequently, this novel $^{Pai-}$CPRP was aligned with other four CPRPs [15, 30, 31] from related crustacean species showing a high degree of homology across multiple species (Figure S-5).

Large Neuropeptides

CHH family is a group of structurally related peptide hormones that play multifunctional roles during the development and throughout the entire life cycle of crustaceans [44]. They regulate carbohydrate metabolism and also exert an inhibitory effect on molting, reproduction, and on osmoregulatory functions [37]. Characterization of CHHs family neuropeptides is challenging, as they contain 70-80 amino acids and complex disulfide bridge connections.

In LC fraction #18 (Figure 2), accurate mass measurement showed a peptide candidate with MW 8271.12 Da, and further DTT reduction of this peptide produced 6 Da of mass increase (Figure 6). These observations matched the two unique features of the CHH neuropeptide family, i.e., MW ranging from 8-10 kDa and PTMs containing thee disulfide bonds, so this peptide was tentatively assigned as the CHH candidate in *P. interruptus* (termed as $^{Pai-}$CHH).

Different from the mid-size peptides, tryptic digestion of large CHH can produce tryptic segments matching to the homologous sequences. Here, the homology search of CHH-family neuropeptides was performed by bottom-up proteomic method. The tryptic digest of the $^{Pai-}$CHH candidate was analyzed by LC-MS/MS on an ESI-QTOF mass spectrometer, followed by Mascot searching against a NCBI protein database. The first hit was the intact $^{Jal-}$CHH peptide from *Jasus lalandii* [46] with 45% homology match (Figure S-6), which was further used as reference sequence to conduct the bottom-up *de novo* sequencing.

Bottom-up *de novo* sequencing was carried out using Pepseq to interpret the tryptic peptide sequences. However, it is challenging to *de novo* sequence the homologous tryptic peptides which contain different residues from the reference sequence $^{Jal-}$CHH. We tried Spider Homology Search in PEAKS [47], where the bottom-up MS/MS data was searched against a home-built database containing $^{Jal-}$CHH. However, no hit was found. Alternatively, we manually sequenced the homologous tryptic peptides. The tryptic peptides were selected

according to two criteria, i.e., MW < reference tryptic peptide ±129 Da (based on the assumption that a single amino acid substitution may occur from the reference sequence and the largest amino acid residue difference between tryptophan and glycine equals 129 Da) and ion intensity > 200 counts (based on typical precursor ion threshold for obtaining decent MS/MS spectra), and then submitted to Pepseq for sequence interpretation. Five new tryptic peptides were determined: CHH[14-21], CHH[14-32], CHH[32-50], CHH[33-50], and CHH [41-50] (sequences shown in Table 2), which contains residues different from the reference sequence $^{Jal-}$CHH. By this bottom-up sequencing method, the first 50 amino acid portion AA^{1}-AA^{50} was determined. The sequence assembly is shown in Figure 7E, and all the tryptic peptides are listed in Table 2.

Top-down MS-based sequencing was then performed to determine the remaining residues AA^{51}-AA^{72}. The $^{Pai-}$CHH peptide was treated with DTT to reduce disulfide bonds according to our previous report [29], and then was directly infused into ESI-LTQ-FTICR for CID and ECD fragmentation. In Figure 7A, a set of continuous b ions, b50 – b71, was clearly detected, leading to confident identification of the remaining residues as ^{51}QCLDDLLLVDVVDEYVASVQSV^{72}. Figure 8B are the ECD and CID fragmentation maps of $^{Pai-}$CHH, which further confirms the bottom-up MS results described above, especially on the different residues from the reference sequence. For example, the Gly^{9} residue was confirmed by observation of c8, c9 and z63, z64 ion pairs. The sequence assembly is illustrated in Figure 7E, and further sequence alignment of CHHs from homologous species was shown in Figure S-8. A combination of the bottom-up and top-down methods not only provided complementary sequence interpretation, but also increased the local identification confidence on each amino acid residue of this large peptide hormone.

CONCLUSIONS

This study systematically evaluated limitations and several improvements of current approaches to the discovery of neuropeptides of various sizes, which facilitates the rational design of methodology for comprehensive characterization of neuropeptides in the crustacean nervous system. A multi-scale strategy was established enabling accurate identification of nine novel neuropeptides spanning a wide range of molecular sizes in *P. interruptus* sinus gland. The results provide a foundation for future functional and mechanistic studies of these novel neuropeptides.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors wish to thank Prof. Neil L. Kelleher's group at Northwestern University for access to an LTQ-Orbitrap Elite system and Drs. Adam D. Catherman and Paul M. Thomas from the Kelleher group for instrument assistance. We would also like to thank Prof. Deborah Baro from the Georgia State University for providing some of the spiny lobsters used in this study. We are also grateful to Prof. Ying Ge at UW-Madison for helpful suggestions on top-down MS analysis and Matt Lawrence at the UW Human Proteomics Program for experimental assistance with the LTQ-FTICR instrument. This work is supported in part by the National Institutes of Health grant (R01DK071801 to LL) and the National Science Foundation grant (CHE-0967784 to LL). LL acknowledges an H. I. Romnes Faculty Research Fellowship. CJ thanks an Oversea Training Fellowship and UW Vilas Conference Presentation Funds. CBL acknowledges an NIH-supported Chemistry Biology Interface Training Program Predoctoral Fellowship (grant number T32-GM008505) and an NSF predoctoral graduate fellowship.

References

[1]. Li L, Sweedler JV. Peptides in the brain: mass spectrometry-based measurement approaches and challenges. Annu Rev Anal Chem. 2008; 1:451–83.

[2]. Mykles DL, Adams ME, Gade G, Lange AB, Marco HG, Orchard I. Neuropeptide action in insects and crustaceans. Physiol Biochem Zool. 2010; 83:836–46. [PubMed: 20550437]

[3]. Christie AE, Stemmler EA, Dickinson PS. Crustacean neuropeptides. Cell Mol Life Sci. 2010; 67:4135–69. [PubMed: 20725764]

[4]. Hummon AB, Richmond TA, Verleyen P, Baggerman G, Huybrechts J, Ewing MA, et al. From the genome to the proteome: uncovering peptides in the *Apis* brain. Science. 2006; 314:647–9. [PubMed: 17068263]

[5]. Che FY, Yuan Q, Kalinina E, Fricker LD. Peptidomics of Cpe fat/fat mouse hypothalamus: effect of food deprivation and exercise on peptide levels. J Biol Chem. 2005; 280:4451–61. [PubMed: 15572367]

[6]. Dowell JA, Heyden WV, Li L. Rat neuropeptidomics by LC-MS/MS and MALDI-FTMS: Enhanced dissection and extraction techniques coupled with 2D RP-RP HPLC. J Proteome Res. 2006; 5:3368–75. [PubMed: 17137338]

[7]. Fricker LD, Lim J, Pan H, Che FY. Peptidomics: identification and quantification of endogenous peptides in neuroendocrine tissues. Mass Spectrom Rev. 2006; 25:327–44. [PubMed: 16404746]

[8]. Jimenez CR, Spijker S, de Schipper S, Lodder JC, Janse CK, Geraerts WP, et al. Peptidomics of a single identified neuron reveals diversity of multiple neuropeptides with convergent actions on cellular excitability. J Neurosci. 2006; 26:518–29. [PubMed: 16407549]

[9]. Fricker LD. Neuropeptidomics to study peptide processing in animal models of obesity. Endocrinology. 2007; 148:4185–90. [PubMed: 17584973]

[10]. Falth M, Skold K, Svensson M, Nilsson A, Fenyo D, Andren PE. Neuropeptidomics strategies for specific and sensitive identification of endogenous peptides. Mol Cell Proteomics. 2007; 6:1188–97. [PubMed: 17401030]

[11]. Hatcher NG, Atkins N Jr. Annangudi SP, Forbes AJ, Kelleher NL, Gillette MU, et al. Mass spectrometry-based discovery of circadian peptides. Proc Natl Acad Sci U S A. 2008; 105:12527–32. [PubMed: 18719122]

[12]. Zhang X, Petruzziello F, Zani F, Fouillen L, Andren PE, Solinas G, et al. High Identification Rates of Endogenous Neuropeptides from Mouse Brain. J Proteome Res. 2012; 11:2819–27. [PubMed: 22424378]

[13]. Shen Y, Tolic N, Xie F, Zhao R, Purvine SO, Schepmoes AA, et al. Effectiveness of CID, HCD, and ETD with FT MS/MS for degradomic-peptidomic analysis: comparison of peptide identification methods. J Proteome Res. 2011; 10:3929–43. [PubMed: 21678914]

[14]. Chen R, Hui L, Cape SS, Wang J, Li L. Comparative Neuropeptidomic Analysis of Food Intake via a Multi-faceted Mass Spectrometric Approach. ACS Chem Neurosci. 2010; 1:204–14. [PubMed: 20368756]

[15]. Ma M, Wang J, Chen R, Li L. Expanding the Crustacean neuropeptidome using a multifaceted mass spectrometric approach. J Proteome Res. 2009; 8:2426–37. [PubMed: 19222238]

[16]. Ma M, Bors EK, Dickinson ES, Kwiatkowski MA, Sousa GL, Henry RP, et al. Characterization of the *Carcinus maenas* neuropeptidome by mass spectrometry and functional genomics. Gen Comp Endocrinol. 2009; 161:320–34. [PubMed: 19523386]

[17]. Lee JE, Atkins N Jr. Hatcher NG, Zamdborg L, Gillette MU, Sweedler JV, et al. Endogenous peptide discovery of the rat circadian clock: a focused study of the suprachiasmatic nucleus by ultrahigh performance tandem mass spectrometry. Mol Cell Proteomics. 2010; 9:285–97. [PubMed: 19955084]

[18]. Bai L, Romanova EV, Sweedler JV. Distinguishing endogenous D-amino acid-containing neuropeptides in individual neurons using tandem mass spectrometry. Anal Chem. 2011; 83:2794–800. [PubMed: 21388150]

[19]. Chen R, Ma M, Hui L, Zhang J, Li L. Measurement of neuropeptides in crustacean hemolymph via MALDI mass spectrometry. J Am Soc Mass Spectrom. 2009; 20:708–18. [PubMed: 19185513]

[20]. Tran JC, Zamdborg L, Ahlf DR, Lee JE, Catherman AD, Durbin KR, et al. Mapping intact protein isoforms in discovery mode using top-down proteomics. Nature. 2011; 480:254–8. [PubMed: 22037311]

[21]. McLafferty FW. A century of progress in molecular mass spectrometry. Annu Rev Anal Chem. 2012; 4:1–22.

[22]. Thuma JB, White WE, Hobbs KH, Hooper SL. Pyloric neuron morphology in the stomatogastric ganglion of the lobster, *Panulirus interruptus*. Brain Behav Evol. 2009; 73:26–42. [PubMed: 19223685]

[23]. Thuma JB, Harness PI, Koehnle TJ, Morris LG, Hooper SL. Muscle anatomy is a primary determinant of muscle relaxation dynamics in the lobster (*Panulirus interruptus*) stomatogastric system. J Comp Physiol A Neuroethol Sens Neural Behav Physiol. 2007; 193:1101–13. [PubMed: 17710408]

[24]. Ouyang Q, Goeritz M, Harris-Warrick RM. *Panulirus interruptus* Ih-channel gene PIIH: modification of channel properties by alternative splicing and role in rhythmic activity. J Neurophysiol. 2007; 97:3880–92. [PubMed: 17409170]

[25]. Hopkins PM. The eyes have it: A brief history of crustacean neuroendocrinology. Gen Comp Endocrinol. 2012; 175:357–66. [PubMed: 22197211]

[26]. Marder E. Neuromodulation of neuronal circuits: back to the future. Neuron. 2012; 76:1–11. [PubMed: 23040802]

[27]. Selverston AI, Szucs A, Huerta R, Pinto R, Reyes M. Neural mechanisms underlying the generation of the lobster gastric mill motor pattern. Front Neural Circuits. 2009; 3:1–12. [PubMed: 19225575]

[28]. Marder E, Bucher D. Understanding circuit dynamics using the stomatogastric nervous system of lobsters and crabs. Annu Rev Physiol. 2007; 69:291–316. [PubMed: 17009928]

[29]. Jia C, Hui L, Cao W, Lietz CB, Jiang X, Chen R, et al. High-definition *De Novo* Sequencing of Crustacean Hyperglycemic Hormone (CHH)-family Neuropeptides. Mol Cell Proteomics. 2012; 11:1951–64. [PubMed: 23028060]

[30]. 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. 2005; 337:765–78. [PubMed: 16214114]

[31]. Hui L, Cunningham R, Zhang Z, Cao W, Jia C, Li L. Discovery and characterization of the Crustacean hyperglycemic hormone precursor related peptides (CPRP) and orcokinin neuropeptides in the sinus glands of the blue crab *Callinectes sapidus* using multiple tandem mass spectrometry techniques. J Proteome Res. 2011; 10:4219–29. [PubMed: 21740068]

[32]. Ge Y, Rybakova IN, Xu Q, Moss RL. Top-down high-resolution mass spectrometry of cardiac myosin binding protein C revealed that truncation alters protein phosphorylation state. Proc Natl Acad Sci U S A. 2009; 106:12658–63. [PubMed: 19541641]

[33]. Petruzziello F, Fouillen L, Wadensten H, Kretz R, Andren PE, Rainer G, et al. Extensive characterization of *Tupaia belangeri* neuropeptidome using an integrated mass spectrometric approach. J Proteome Res. 2012; 11:886–96. [PubMed: 22070463]

[34]. Clynen E, Baggerman G, Huybrechts J, Vanden Bosch L, De Loof A, Schoofs L. Peptidomics of the locust *corpora allata*: identification of novel pyrokinins (-FXPRLamides). Peptides. 2003; 24:1493–500. [PubMed: 14706528]

[35]. Michalski A, Damoc E, Lange O, Denisov E, Nolting D, Muller M, et al. Ultra High Resolution Linear Ion Trap Orbitrap Mass Spectrometer (Orbitrap Elite) Facilitates Top Down LC MS/MS and Versatile Peptide Fragmentation Modes. Mol Cell Proteomics. 2012; 11:1–11.

[36]. Ma M, Chen R, Ge Y, He H, Marshall AG, Li L. Combining bottom-up and top-down mass spectrometric strategies for *de novo* sequencing of the crustacean hyperglycemic hormone from Cancer borealis. Anal Chem. 2009; 81:240–7. [PubMed: 19046072]

[37]. Webster SG, Keller R, Dircksen H. The CHH-superfamily of multifunctional peptide hormones controlling crustacean metabolism, osmoregulation, moulting, and reproduction. Gen Comp Endocrinol. 2012; 175:217–33. [PubMed: 22146796]

[38]. Szabo TM, Chen R, Goeritz ML, Maloney RT, Tang LS, Li L, et al. Distribution and physiological effects of B-type allatostatins (myoinhibitory peptides, MIPs) in the stomatogastric

nervous system of the crab *Cancer borealis*. J Comp Neurol. 2011; 519:2658–76. [PubMed: 21491432]

[39]. Boersema PJ, Raijmakers R, Lemeer S, Mohammed S, Heck AJ. Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. Nat Protoc. 2009; 4:484–94. [PubMed: 19300442]

[40]. Yang SJ, Nie AY, Zhang L, Yan GQ, Yao J, Xie LQ, et al. A novel quantitative proteomics workflow by isobaric terminal labeling. J Proteomics. 2012; 75:5797–806. [PubMed: 22813877]

[41]. Hsu JL, Huang SY, Chow NH, Chen SH. Stable-isotope dimethyl labeling for quantitative proteomics. Anal Chem. 2003; 75:6843–52. [PubMed: 14670044]

[42]. Audsley N, Matthews HJ, Price NR, Weaver RJ. Allatoregulatory peptides in Lepidoptera, structures, distribution and functions. J Insect Physiol. 2008; 54:969–80. [PubMed: 18377924]

[43]. 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]

[44]. Chung JS, Zmora N, Katayama H, Tsutsui N. Crustacean hyperglycemic hormone (CHH) neuropeptides family: Functions, titer, and binding to target tissues. Gen Comp Endocrinol. 2009; 166:447–54. [PubMed: 20026335]

[45]. Bonet-Costa C, Vilaseca M, Diema C, Vujatovic O, Vaquero A, Omenaca N, et al. Combined bottom-up and top-down mass spectrometry analyses of the pattern of post-translational modifications of *Drosophila melanogaster* linker histone H1. J Proteomics. 2012; 75:4124–38. [PubMed: 22647927]

[46]. Marco HG, Hansen IA, Scheller K, Gade G. Molecular cloning and localization of a cDNA encoding a crustacean hyperglycemic hormone from the South African spiny lobster, *Jasus lalandii*. Peptides. 2003; 24:845–51. [PubMed: 12948836]

[47]. Zhang J, Xin L, Shan BZ, Chen WW, Xie MJ, Yuen D, et al. PEAKS DB: *De Novo* Sequencing Assisted Database Search for Sensitive and Accurate Peptide Identification. Mol Cell Proteomics. 2012; 11:1–8.

Highlights

- Sequence coverage of neuropeptide analysis revealed limitations to mid- and large-size peptide analysis.
- A multi-scale strategy was established for elucidation of various sizes of peptides.
- Nine novel neuropeptides (MW 0.9-8.2 kDa) were fully sequenced in the spiny lobster *P. interruptus*.

SIGNIFICANCE

Mass spectrometry (MS)-based neuropeptidomics aims to completely characterize the neuropeptides in a target organism as an important first step toward a better understanding of the structure and function of these complex signaling molecules. Although liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) with data-dependent acquisition is a powerful tool in peptidomic research, it often lacks the capability for *de novo* sequencing of mid-size and large peptides due to inefficient fragmentation of peptides larger than 4 kDa. This study describes a multi-scale strategy for complete and confident sequence elucidation of various sizes of neuropeptides in the crustacean nervous system. The aim is to fill a technical gap where the conventional strategy is inefficient for comprehensive characterization of a complex neuropeptidome without assistance of genomic information. Nine novel neuropeptides in a wide range of molecular weights (0.9-8.2 kDa) were fully sequenced from a major neuroendocrine organ of the spiny lobster, *Panulirus interruptus*. The resulting molecular information extracted from such multi-scale peptidomic analysis will greatly accelerate functional studies of these novel neuropeptides.

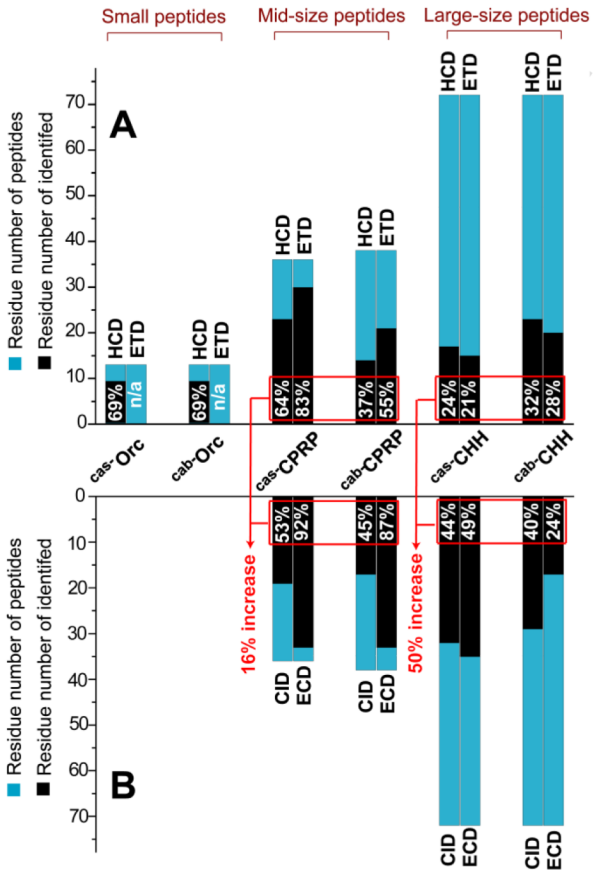

Figure 1. Evaluation results for sequencing of various sizes of neuropeptides

(A) On-line LC-MS/MS; (B) Off-line top-down MS/MS. The percentage of peptide sequence coverage is labeled in the black bar of every peptide. The percentage of increase in red is calculated by the following expression Σ identified residue number/ Σ peptide residue number. The MS/MS fragmentation maps of these model peptides are shown in Figure S-1 and S-2.

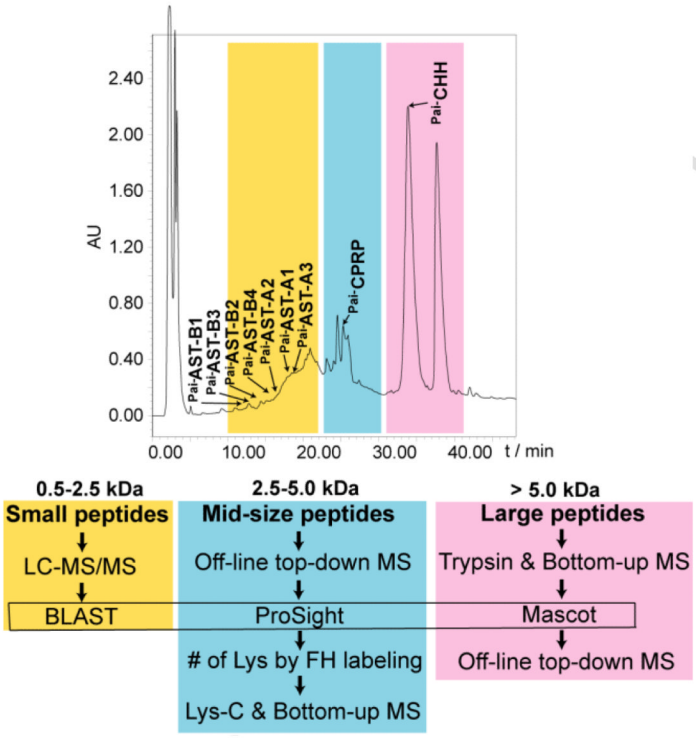

Figure 2. Workflow of the multi-scale strategy

The top panel shows reversed phase HPLC fractionation of *P. interruptus* sinus gland tissue extract showing the presence of various sizes of novel neuropeptides. The bottom panel outlines each specific strategy for discovery and identification of small, mid-size and large peptides. It should be noted that the retention behavior of peptides in the reversed phase HPLC is not only based on molecular size, but also the hydrophobicity and other sequence-dependent effects. Here, we show three different color regions in the LC profile for the purpose of illustration.

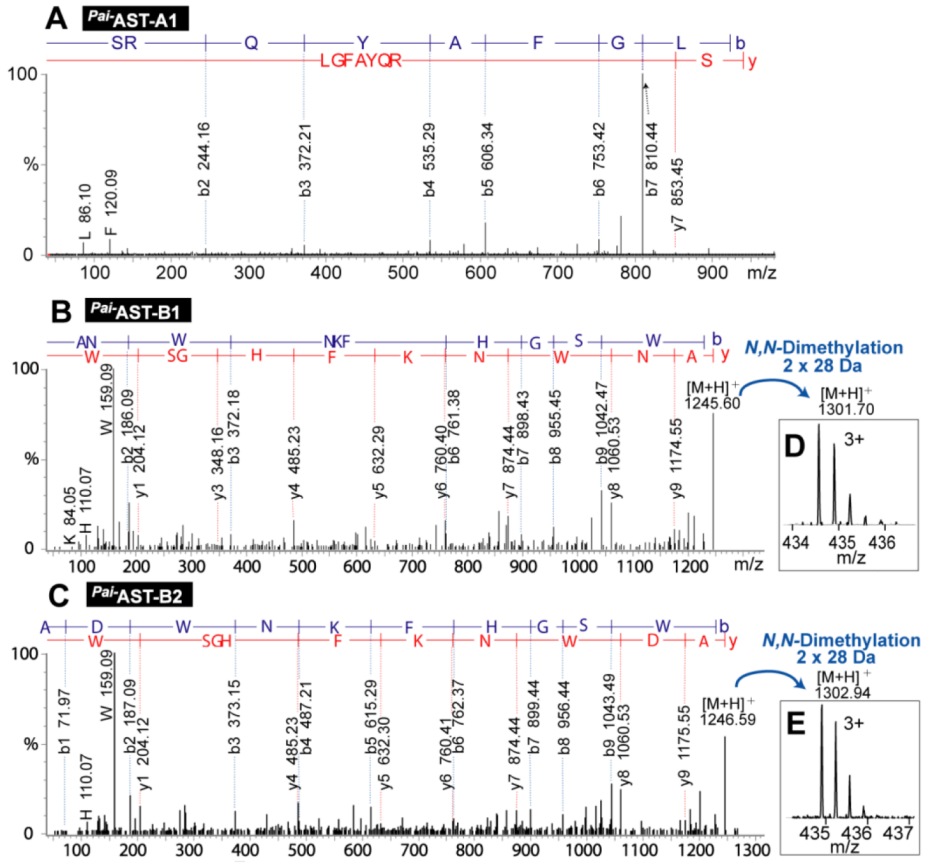

Figure 3. *De novo* sequencing of AST-family neuropeptides

MS/MS spectra of $^{Pai-}$AST-A1 (A), $^{Pai-}$AST-B1 (B), and $^{Pai-}$AST-B2 (C). MS spectra of dimethylated $^{Pai-}$AST-B1 (D), and $^{Pai-}$AST-B2 (E). Panels A, B and C are deconvoluted spectra exported from Pepseq, and Panels D and E are original zoom-in mass spectra.

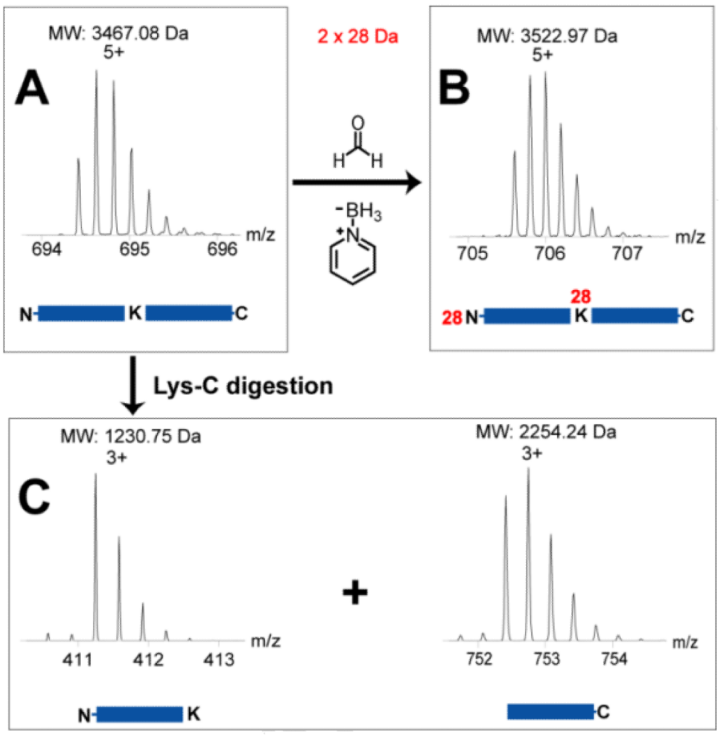

Figure 4. Strategy for *de novo* sequencing of a mid-sized CPRP neuropeptide

Zoom-in MS spectra of (A) intact $^{Pai-}$CPRP, (B) dimethylated $^{Pai-}$CPRP, and (C) proteolytic peptides of $^{Pai-}$CPRP after Lys-C digestion. The 28 Da of mass increase of peptides arises from *N,N*-dimethylation.

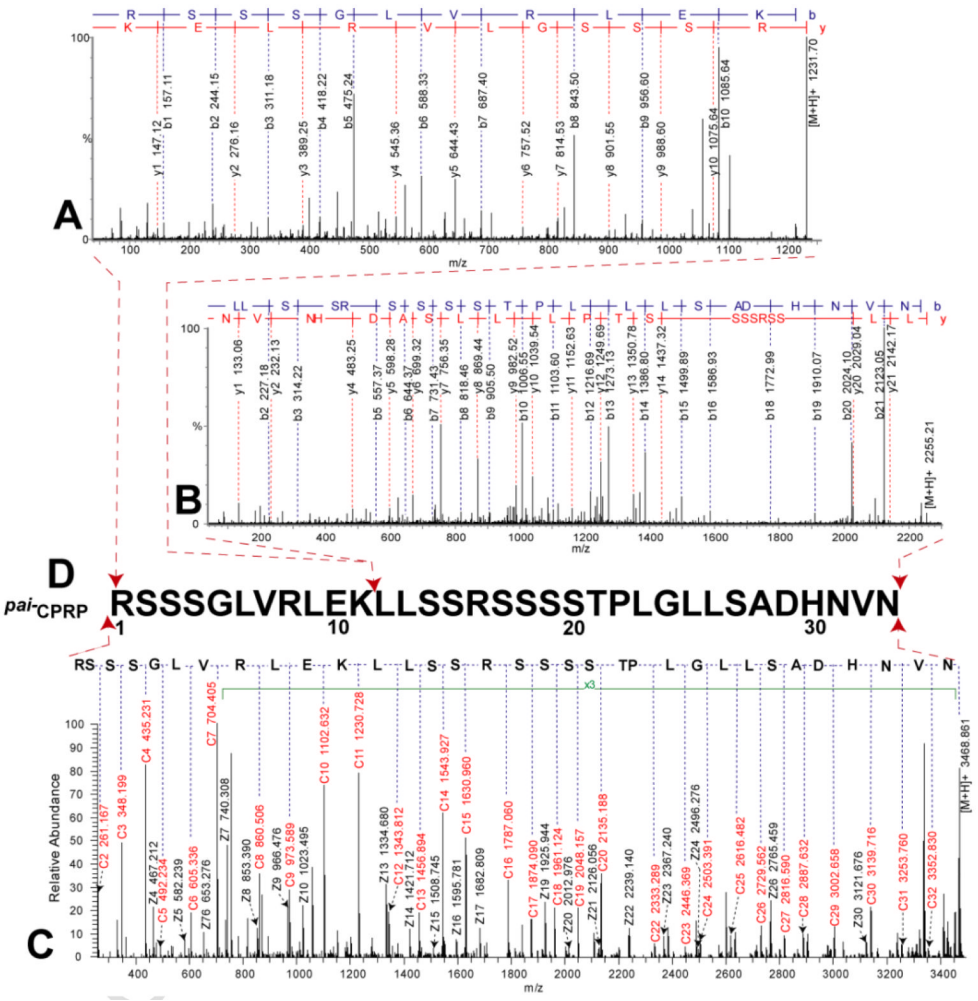

Figure 5. ***De novo*** **sequencing MS/MS spectra of** $^{Pai-}$**CPRP**

Bottom-up MS/MS spectra of (A) $^{Pai-}$CPRP [1-11] and (B) $^{Pai-}$CPRP [12-33]. (C) Top-down ECD spectra of intact $^{Pai-}$CPRP exported from Xtract after charge deconvolution. (D) Sequence assembly of $^{Pai-}$CPRP.

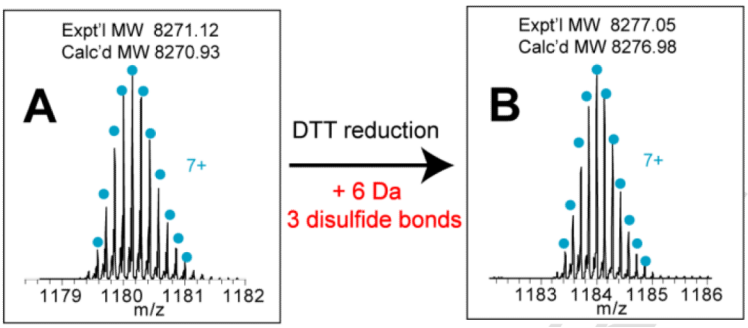

Figure 6. Determination of disulfide bond numbers in $^{Pai-}$CHH

High resolution zoom-in spectra of (A) intact $^{Pai-}$CHH and (B) DTT-reduced $^{Pai-}$CHH.

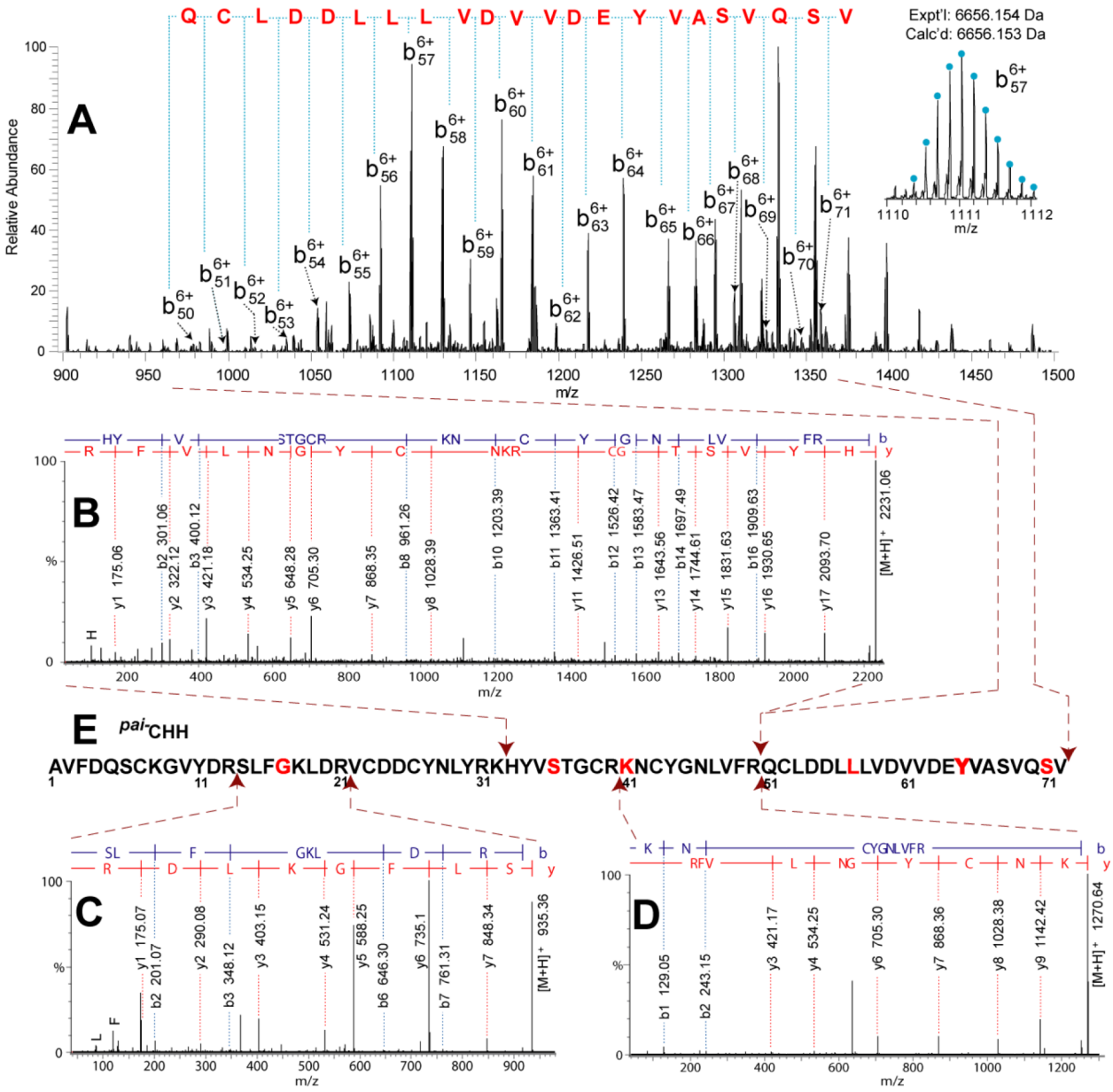

Figure 7. *De novo* sequencing MS/MS spectra of $^{Pai-}$CHH

(A) Top-down CID spectra of DTT-reduced $^{Pai-}$CHH. Bottom-up MS/MS spectra of (B) $^{Pai-}$CHH[33-50], (C) $^{Pai-}$CHH[14-21] and (D) $^{Pai-}$CHH[41-50]. (E) Sequence assembly of $^{Pai-}$CHH. The dipeptide segments ^{38}GC^{39}, ^{45}GN^{46}, and ^{49}FR^{50} were assigned according to the homologous sequence $^{Jal-}$CHH.

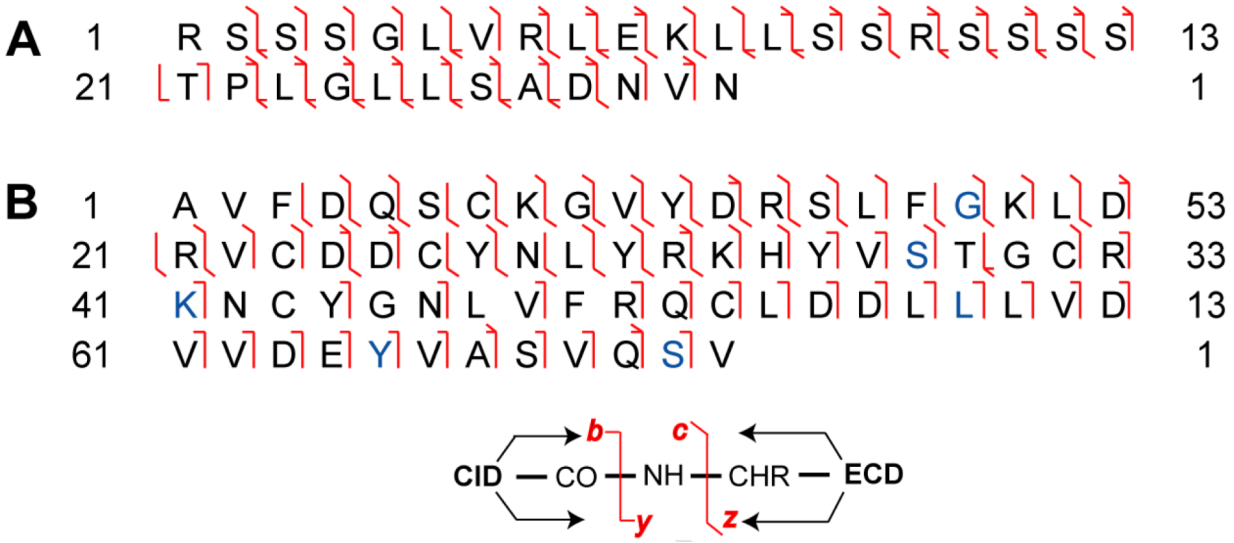

Figure 8. Fragmentation maps of $^{Pai-}$CPRP (A) and $^{Pai-}$CHH (B) from off-line top-down MS/MS

In Panel B, the residues different from the reference sequence $^{Jal-}$CHH are highlighted in blue.

Table 1

Novel neuropeptides identified in this study.

Peptide	Sequence	Exptl MW [a]	Mass error (Da) [a]	Spectrum
Pai–AST-A1	SRQYAF GLamide	940.55	0.06	Fig. 3A
Pai–AST-A2	NRQYSFGLamide	982.50	0.02	Fig. S3
Pai–AST-A3	NRPYSFGLamide	951.55	0.05	Fig. S3
Pai–AST-B1	ANWNKFHGSWamide	1244.59	0.00	Fig. 3B
Pai–AST-B2	ADWNKFHGSWamide	1245.60	0.02	Fig. 3C
Pai–AST-B3	GNWNKFHGSWamide	1230.59	0.01	Fig. S3
Pai–AST-B4	GDWNKFHGSWamide	1231.57	0.01	Fig. S3
Pai–CPRP	RSS SGLVRLEKLL SSRSS SSTPLGLLS ADHNVN	3466.86	0.00	Fig. 5
Pai–CHH	AVFDQSCKGVYDRSLFGKLDRVCDD CYNLYRKHYVSTGCRKNCYGNLVFR QCLDDLLLVDVVDEYVASVQSV [b]	8271.12	0.19	Fig. 7

[a] Small peptides were measured on ESI-QTOF; mid-size and large peptides were analyzed on an LTQ-FTICR mass spectrometer.

[b] Pai–CHH contains three disulfide bonds. The disulfide linkages are the same as the homologous CHH peptide(Ref. 46).

Table 2

Tryptic peptides of CHH identified by bottom-up *de novo* sequencing method.

Exptl MW	Calcd MW	Sequence [a]	Tryptic peptide	MS/MS spectrum [b]
1543.50	1543.71	AVFDQSCKGVYDR	CHH[1-13]	Fig. S7
2075.66	2076.01	AVFDQSCKGVYDRSLFGK	CHH[1-18]	–
934.36	934.52	SLFGKLDR	CHH[14-21]	Fig. 7C
2420.79	2421.15	SLFGKLDRVCDDCYNLYRK	CHH[14-32]	–
1760.49	1760.76	LDRVCDDCYNLYR	CHH[19-31]	Fig. S7
1888.53	1888.85	LDRVCDDCYNLYRK	CHH[19-32]	–
1376.35	1376.55	VCDDCYNLYR	CHH[22-31]	Fig. S7
1504.39	1504.64	VCDDCYNLYRK	CHH[22-32]	Fig. S7
2358.16	2358.15	KHYVSTGCRKNCYGNLVFR	CHH[32-50]	–
2230.06	2230.05	HYVSTGCRKNCYGNLVFR	CHH[33-50]	Fig. 7B
1269.64	1269.63	KNCYGNLVFR	CHH[41-50]	Fig. 7D
1141.35	1141.53	NCYGNLVFR	CHH[42-50]	Fig. S7

[a] Fixed carbamidomethyl modification is on cysteine. The residues different from the homologous sequence are highlighted with underline.

[b] Data is selectively shown for covering the sequence CHH[1-50]. CHH[51-71] was determined by top-down *de novo* sequencing.