A MASS SPECTROMETRY-BASED NEUROPEPTIDE DISCOVERY PIPELINE: SEQUENCE, STRUCTURE AND FUNCTIONALITY

By

Chenxi Jia

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Pharmaceutical Sciences)

at the

UNIVERSITY OF WISCONSIN-MADISON

2013

Date of final oral examination: 11/26/13

The dissertation is approved by the following members of the Final Oral Committee:
Lingjun Li, Professor, Pharmaceutical Sciences and Chemistry
Tim S. Bugni, Assistant Professor, Pharmaceutical Sciences
Richard P. Hsung, Professor, Pharmaceutical Sciences and Chemistry
Joshua J. Coon, Professor, Chemistry and Biomolecular Chemistry
Jean-Michel Ané, Associate Professor, Agronomy

Acknowledgements

Over the past four years I have received support and encouragement from professors, colleagues, friends and families. First of all, I would like to express my sincere gratitude to my advisor Prof. Lingjun Li, without whom none of the work in this thesis may be finished. In the earlier days of my PhD study, I was struggling with no progress and tons of ideas did not work. Her patience, motivation, enthusiasm, and immense knowledge continuously support me to pursue my research goals and dreams in science. Because of her guidance and enlightening, my thesis research has turned into a successful and rewarding journey. Prof. Li offers me the freedom and cultivates my curiosity in scientific world. She provided me tremendous training opportunities to become an independent researcher. I am proud of being Prof. Lingjun Li's student and working in her world-class laboratory.

Besides my advisor, I would like to thank the rest of my thesis committee: Prof. Richard P. Hsung, Prof. Joshua J. Coon, Prof. Tim S. Bugni and Prof. Jean-Michel Ané. Their encouragement, constructive criticism and insightful comments substantially improve the academic caliber of this work.

My sincere thanks also go to two members of the Li Lab, outstanding researchers and my important colleagues: Christopher B. Lietz and Qing Yu. Chris is a mentor who guides me into the ion mobility mass spectrometry field, although he joined Li Lab later than me. Qing provides excellent research assistance and tremendous contributions to this thesis work. I always cherish the time we were working together to pursue our scientific ideas. I sincerely wish them to be successful and to make great achievements in academia.

In addition to Chris and Qing, several other people have contributed directly to this work.

This includes the current lab members, Jingxin Wang, Zhidan Liang, Chuanzi Ouyang and Fengfei Ma as well as the former lab members Dr. Limei Hui, Dr. Ruibing Chen, Dr. Claire Schmerberg, Dr. Xiaoyue Jiang, Dr. Hui Ye, Dr. Weifeng Cao. I also would like to thank other labmates Dr. Zichuan Zhang, Dr. Nicole Woodards, Dr. Junhua Wang, Dr. Mingming Ma, Dr. Robert Cunningham, Dr. Robert Sturm, Dr. Yuzhuo Zhang, Dr. Feng Xiang, Dr. Jiang Zhang, Dr. Xin Wei, Dr. Yan Liu, Shan Jiang, Dr. Xuefei Zhong, Chenxi Yang, Erin Gemperline, Tyler Creer, Dustin Frost, Bingming Chen and Ling Hao for inspiring discussions and day-to-day interactions.

I would like to express my sincere gratitude to Dr. Cameron O. Scarlett in the Analytical Instrumentation Center at the School of Pharmacy, where I worked as a Project Assistant for four years. Dr. Scarlett's immense knowledge about mass spectrometry and instrumentation support me helped me to become a better mass spectrometrist.

I also would like to thank Dr. Lu Bai and Prof. Jonathan V. Sweedler at the University of Illinois at Urbana-Champaign for inspiring me on the project of D-amino acid-containing peptide characterization. I would like to acknowledge my collaborators, Prof. Neil Kelleher, Dr. Paul Martin Thomas, and Adam Catherman at Northwestern University and Prof. Ying Ge at the University of Wisconsin-Madison for aiding in the top-down mass spectrometry experiments; Dr. Zhe Wu and Prof. Qiang Cui at University of Wisconsin-Madison for performing theoretical calculations.

Last but not least, I would like to thank my parents Jianjun Jia and Xianping Zhang, and my grandmother Anxiang Zhao for their support and encouragement for me to pursue a career in science.

Table of Contents

A Mass Spectrometry-Based Neuropeptide Discovery Pipeline: Sequence, Structure and Functionality

Chenxi Jia

Under the supervision of Professor Lingjun Li

At the University of Wisconsin-Madison

ABSTRACT

Neuropeptides represent the most complex and diverse group of naturally occurring endogenous molecules in the nervous system that play important roles in the regulation of physiological process. Elucidation of these signaling molecules is a crucial step towards understanding the underlying mechanism of neuromodulation. However, the conventional mass spectrometry (MS)-based strategy is often inadequate for comprehensive characterization of neuropeptides. Herein, my thesis research focuses on development of an efficient peptide discovery pipeline that accelerates neuropeptidomic analysis in crustacean models, toward filling the technical gaps in sequence analysis and structural elucidation and better understanding of the functional roles of these signaling molecules.

De novo sequencing of large neuropeptides is challenging due to the inefficient fragmentation of peptides larger than 4 kDa. We established a high-definition MS approach for identification and characterization of large crustacean hyperglycemic hormone (CHH)-family neuropeptides, which was achieved by combining bottom-up, off-line top-down, and on-line top-down tandem MS methods. We further refined and incorporated this high-definition approach into a multi-scale strategy for simultaneous and confident sequence elucidation of various sizes of peptides in the crustacean nervous system. A wide range of neuropeptides with various molecular weights (0.9-8.2 kDa) were fully sequenced.

The well-characterized primary sequences of neuropeptides provide opportunities for elucidation of their structural information. Generally, the D-amino acid-containing peptides (DAACPs) exhibit dramatically higher bioactivities than their all-L counterparts. To rapidly and precisely localize D-amino acids in DAACP candidates, we developed a novel site-specific strategy by ion mobility spectrometry (IMS) analysis of MS-generated epimeric fragment ions. Our results indicate that the isomerization of L- to D-Phe occurs in the American lobster CHHs. In addition to characterization of conformational isomers of neuropeptides, our method allows elucidating structural isomers in gas phase. We report on a novel strategy for qualitative and quantitative analysis of b-type ion isomers by combining electron transfer dissociation, IMS and formaldehyde labeling. The analysis results provide evidence to support the proposed fragmentation mechanism of peptide sequence scrambling in gas phase.

The final goal of our study is to understand the roles of neuropeptides play in neural circuits. Feeding behavior is critical for animal survival, and is also a fundamental aspect of energy homeostasis. To investigate if CHHs are involved in regulation of food intake, a new isotopic labeling-assisted top-down MS strategy was developed to directly monitor the abundance changes of endogenous large neuropeptides. Comparative analysis in unfed and fed crabs revealed that the CHH abundance in the sinus glands was significantly increased after food intake. In addition, label-free quantitative analysis reveals that the orcokinin and RFamide peptides were down-regulated after food intake, suggesting that the release of these neuropeptides might be altered by feeding behavior.

Collectively, the methods developed in this study are transferable to other biological systems and multiple areas of research. The detailed molecular information of identified neuropeptides

provides a framework for future investigations of their physiological roles. The quantitative studies offer preliminary evidence to support the peptidergic regulation of feeding behavior.

Chapter 1

Introduction: Brief Background and Research Summary

Abstract

My thesis research focuses on development of a neuropeptide discovery pipeline to facilitate peptidomic studies of the nervous system of crustacean models. In terms of research objective, the overall work can be divided into three aspects: sequence analysis (Chapter 2 and 3), structural elucidation (Chapter 4 and 5) and functional study (Chapter 6 and 7). For *de novo* sequencing of neuropeptides, one of the most challenging tasks is to characterize the amino acid sequences of large neuropeptides without assistance of genome information. In Chapter 2, I present a high-definition MS approach for identification and characterization of large crustacean hyperglycemic hormone (CHH)-family neuropeptides, a class of peptide hormones that play central roles in the regulation of many important physiological processes of crustaceans. Furthermore, by incorporating this high-definition approach, a multi-scale strategy was established in Chapter 3, which enables simultaneous and confident characterization of various sizes of peptides in the crustacean nervous system. The well-characterized sequence information of neuropeptides offers opportunities for the structural elucidation of these signaling molecules. Isomerization of an L- to D-amino acid causes structural variation of peptides and thus significantly changes their bioactivities. Therefore, Chapter 4 introduces a novel site-specific strategy to rapidly and precisely localize D-amino acids in peptides by ion mobility spectrometry (IMS). Our results indicate that the isomerization of L- to D-Phe occurs at the third residue of the American lobster crustacean hyperglycemic hormones (CHHs). In addition to the conformational isomerization of peptides, the structural isomerization of peptide fragment ions during MS gas phase analysis may cause re-arrangement of peptide chains, thus leading to sequence scrambling and misidentification. Chapter 5 reports on a strategy for qualitative and quantitative analysis of b-type fragment ion isomers by combining electron transfer dissociation, IMS and chemical

labeling and theoretical calculation using molecular dynamic simulation. The ultimate goal of my study is to improve our understanding of the functionalities of these neuropeptides. In Chapter 6, a new isotopic labeling-assisted top-down MS strategy was developed, which enabled direct monitoring of the abundance changes of endogenous large neuropeptides. Comparative analysis of CHHs in unfed and fed crabs revealed that the peptide abundance in the sinus glands was significantly increased after food intake, suggesting that the release of CHHs might be altered by feeding behavior. In Chapter 7, label-free quantitative analysis reveals that the abundances of orcokinin and RFamide peptides were decreased after food intake.

1.1 Brief Background

1.1.1 Neuropeptides

Neuropeptides acts as neurohormones, neuromodulators, or neurotransmitters in neural circuits. Many of essential physiological behaviors are regulated by this complex suite of singling molecules [1]. The pre-propeptide is synthesized in the rough endoplasmic reticulum where the signal peptide is removed. The propeptide is then transported to the Golgi apparatus and packaged into vesicles. It is further cleaved and modified by proteases to generate the final bioactive peptides [2-5].

To clearly understand the neural circuitry, an initial step is to elucidate the full cast of molecular players and their mode of action on each of the elements of the circuit. However, neuropeptides feature high chemical complexity and wide distribution. The precise molecular mechanisms at the cellular and network levels remain ambiguous, which is partially due to a lack of analytical capabilities to measure and identify these low abundance endogenous signaling molecules in a complex microenvironment [4]. Clearly, the development of efficient analytical tools for identification and quantitation of neuropeptides is in great demand. Herein, my thesis research focuses on development of a neuropeptide discovery pipeline toward probing the sequence, structure and function of these signaling molecules.

1.1.2 Crustacean Model System

Compared with complex mammalian nervous systems, the crustacean stomatogastric nervous system (STNS) is highly simplified [1, 6-10]. It is an ideal model for studying neuromodulation in a well-defined neural network and for understanding the neural basis of rhythmic behavior. For example, overlapping subsets of these neurons contribute to the circuits responsible for generating the rhythmic motor patterns for chewing (gastric mill circuit) and the filtering of

chewed food (pyloric circuit) [7, 11]. Such neural activity patterns are dependent on the modulatory input from peptidergic projection neurons located in other components of the STNS, including the oesophageal ganglion and the paired commissural ganglia [8, 10]. Research in this model system led to the notion that different modulators produce completely different outputs for the same neural network [10]. Additionally, neuroendocrine organs and structures such as the pericardial organs (PO) and sinus glands (SG) contain a wide array of neuropeptides that can be released into circulating system and regulate the functions of distant organs [8]. In this study, the crustacean model was used for development of analytical methodology and elucidation of these intriguing bioactive molecules present in the neural circuits.

1.1.3 Peptidomics

The peptidomic approach aims at simultaneous visualization and identification of the entire complement of peptides in a cell or tissue [5, 12-17]. While some of the proteomics methodology may seem readily applicable to peptidomic studies, several aspects of peptidomics must be distinguished from those of proteomics. For example, it is important to identify the native and endogenous forms of neuropeptides with post-translational modifications (PTMs) in peptidomics studies, while most proteomics studies identify proteins by detecting the peptide fragments produced by enzymatic digestion [18]. In proteomics studies, identification of a protein can be achieved by detection of a small number of tryptic fragments. In contrast, because each neuropeptide typically has a distinct function, it is important to identify all the neuropeptides in peptidomics studies, even if they arise from the same prepropeptide. Another major difference, and a challenging aspect of peptidomics, is the almost complete lack of genomic information on the organism of interest [19-21]. This will make the database searching strategy difficult for neuropeptide identification. Even for neuropeptide analysis in organisms with genomic sequence

information available, database searching is often complicated by the lack of knowledge of the proteases involved. Therefore, there is a great need to develop new and improved bioinformatics and analytical strategies that are specifically tailored for the global analysis of neuropeptides.

1.2 Research Summary

1.2.1 Sequence Analysis of Neuropeptides

1.2.1.1 High-definition *De Novo* Sequencing of Crustacean Hyperglycemic Hormone (CHH)-Family Neuropeptides (Chapter 2)

A complete understanding of the biological functions of large signaling peptides (>4 kDa) require comprehensive characterization of their amino acid sequences and PTMs, which presents significant analytical challenges [4]. In the past decade, there has been great success with mass spectrometry (MS)-based *de novo* sequencing of small neuropeptides [12, 19-30]. However, these approaches are less applicable to larger neuropeptides due to the inefficient fragmentation of peptides larger than 4 kDa and their lower endogenous abundance. The conventional proteomics approach focuses on large-scale determination of protein identities via database searching, lacking the ability for in-depth elucidation of individual amino acid residues. Here, I present a multifaceted MS approach for identification and characterization of large CHH-family neuropeptides [31-33], a class of peptide hormones that play central roles in the regulation of many important physiological processes of crustaceans. Six crustacean CHH-family neuropeptides (8-9.5 kDa), including two novel peptides with extensive disulfide linkages and PTMs, were fully sequenced without reference to genomic databases. High-definition *de novo* sequencing was achieved by a combination of bottom-up, off-line top-down, and on-line top-down tandem MS methods. Statistical evaluation indicated that these methods provided complementary information for sequence interpretation and increased the local identification

confidence of each amino acid and PTM. Further investigations by MALDI imaging MS mapped the spatial distribution and colocalization patterns of various CHH-family neuropeptides in the neuroendocrine organs, revealing that two CHH-subfamilies are involved in distinct signaling pathways. Conformational changes in CHH peptides caused by pyro-Glu modification were also observed in ion mobility MS.

1.2.1.2 A Multi-scale Strategy for Discovery of Novel Endogenous Neuropeptides in the Crustacean Nervous System (Chapter 3)

The conventional MS-based strategy is often inadequate for the comprehensive characterization of various size neuropeptides without assistance of genomic information [23, 24]. 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 sillico* homology search. Collectively, the ability to simultaneously address 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.

1.2.2 Structural Elucidation of Neuropeptides

1.2.2.1 Site-Specific Characterization of D-Amino Acid-Containing Peptide Epimers by Ion Mobility Spectrometry (Chapter 4)

Traditionally, the D-amino acid-containing peptide (DAACP) candidate can be discovered by observing the differences of biological activity and chromatographic retention time between the synthetic peptides and naturally occurring peptides [34-39]. However, it is difficult to determine the exact position of D-amino acid in the DAACP candidates. Herein, we developed a novel site-specific strategy to rapidly and precisely localize D-amino acids in peptides by ion mobility spectrometry (IMS) [40, 41] analysis of tandem MS-generated epimeric fragment ions. Briefly, the D/L-peptide epimers were separated by on-line reversed-phase liquid chromatography and fragmented by collision induced dissociation (CID), followed by IMS analysis. The epimeric fragment ions resulting from D/L-peptide epimers exhibit conformational differences, thus showing different mobility in IMS. The drift time shift between the epimeric fragment ions was used as criteria to localize the D-amino acid substitution. The utility of this strategy was demonstrated by analysis of peptide epimers with different molecular sizes, [D-Trp]-Melanocyte stimulating hormone, [D-Ala]-Deltorphin, [D-Phe]-Achatin-I and their all-L forms. Furthermore, the 8.5 kDa CHHs were isolated from American lobster and identified by integration of MS-based bottom-up and top-down sequencing approaches. The IMS data acquired with use of our novel site-specific strategy indicates that the isomerization of L- to D-Phe occurs at the third residue of the CHHs from the N-terminus. Collectively, this study highlights a new route for discovery of DAACPs using IMS technique centered on localization of D-amino acid residues.

1.2.2.2 Gas-phase Ion Isomer Analysis Reveals the Mechanism of Peptide Sequence Scrambling (Chapter 5)

Peptide sequence scrambling during mass spectrometry-based gas-phase fragmentation analysis causes misidentification of peptides and proteins [42-46]. Thus, there is a need to develop an efficient approach to probing the gas-phase fragment ion isomers related to sequence scrambling and the underlying fragmentation mechanism, which will facilitate the development of bioinformatics algorithm for proteomics research. Herein, we report on the first use of electron transfer dissociation [47-49] (ETD)-produced diagnostic fragment ions to probe the components of gas-phase peptide fragment ion isomers. In combination with ion mobility spectrometry (IMS) and formaldehyde labeling [50], this novel strategy enables qualitative and quantitative analysis of b-type fragment ion isomers. ETD fragmentation produced diagnostic fragment ions indicative of the precursor ion isomer components, and subsequent IMS analysis of b ion isomers provided their quantitative and structural information. The isomer components of three representative b ions (b_9, b_{10}, and b_{33} from three different peptides) were accurately profiled by this method. IMS analysis of the b_9 ion isomers exhibited dynamic conversion among these structures. Furthermore, molecular dynamics simulation predicted theoretical drift time values which were in good agreement with experimentally measured values. Our results strongly support the mechanism of peptide sequence scrambling via b ion cyclization, and provide the first experimental evidence to support that the conversion from molecular precursor ion to cyclic b ion ($M \rightarrow {}^{c}b$) pathway is less energetically (or kinetically) favored.

1.2.3 Functional Aspects of Neuropeptides Related to Feeding Behavior

1.2.3.1 Qualitative and Quantitative Top-down Mass Spectral Analysis of Crustacean Hyperglycemic Hormones in Response to Feeding (Chapter 6)

An efficient pipeline for peptide discovery accelerates peptidomic analysis and facilitates a better understanding of the functional roles of neuropeptides. However, qualitative and quantitative analysis of large neuropeptides are challenging due to the bigger molecular sizes, multiple post-translational modifications, and interference by homologous isoforms. Herein, we refined two methodologies in the pipeline for highly confident and efficient MS-based peptide discovery. For the qualitative analysis, the so-called "high resolution/accurate mass" measurement on Orbitrap mass spectrometers was integrated with computer-assisted homology search, which was successfully applied to decipher the substituted amino acid residues in large neuropeptides by referring to homologous sequences. For the quantitative analysis, a new isotopic labeling-assisted top-down MS strategy was developed, which enabled direct monitoring of the abundance changes of endogenous large neuropeptides. By using the refined peptide discovery pipeline, one novel CHH from the Dungeness crab sinus glands was confidently identified and *de novo* sequenced, and its relative abundance was quantified. Comparative analysis of CHHs in unfed and fed crabs revealed that the peptide abundance in the sinus glands was significantly increased after food intake, suggesting that the release of CHHs might be altered by feeding behavior.

1.2.3.2 Rapid and Sensitive Characterization of Neuropeptidome in Dungeness Crabs on a Q-Exactive High-resolution Mass Spectrometer and Peptide Alterations in Response to Feeding (Chapter 7)

The Dungeness crab *C. magister* is a favored animal model for neurobiological studies due to their relatively simple and well-defined nervous system. However, the neuropeptidome is not fully discovered in this species. In this study, we employed high resolution/accurate mass measurement and home-built database searching to elucidate the neuropeptidome in the nervous

system of Dungeness crab. In total, 127 peptides were identified with high confidence, including 11 novel ones. In addition, one large mandibular organ-inhibiting hormone was completely *de novo* sequenced by bottom-up strategy. Label-free quantitative analysis of the sinus glands reveals that the abundances of orcokinin and RFamide peptides were decreased after food intake.

1.3 Conclusions

In summary, this research work contributes to three aspects: peptide discovery, methodology development, and study of neuroendocrine signaling. Specifically, this study (1) results in discovery of novel neuropeptides and comprehensive characterization of neuropeptidomes in several crustacean species; (2) generates a set of improved MS tools to allow high-definition sequencing of large neuropeptides, multi-scale characterization of peptidome, localization of D-amino acid in peptides, analysis of ion isomers in gas phase, and top-down quantification; and (3) provides experimental evidence and practical insights into an improved understanding of neuropeptides involved in regulation of feeding behavior. The perspective is to establish a simplified neuroendocrine model of energy regulation using crustaceans, which provides transferable insights to extremely complicated neuroendocrine network of higher level vertebrates and even human, facilitating pathological and pharmacological studies.

1.4 References

[1] Marder E. Neuromodulation of neuronal circuits: back to the future. Neuron. 2012;76:1-11.
[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.
[3] Christie AE, Stemmler EA, Dickinson PS. Crustacean neuropeptides. Cell Mol Life Sci. 2010;67:4135-69.
[4] Li L, Sweedler JV. Peptides in the brain: mass spectrometry-based measurement approaches and challenges. Annu Rev Anal Chem. 2008;1:451-83.

[5] 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.

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

[7] Nusbaum MP, Beenhakker MP. A small-systems approach to motor pattern generation. Nature. 2002;417:343-50.

[8] Kilman VL, Marder E. Ultrastructure of the stomatogastric ganglion neuropil of the crab, Cancer borealis. J Comp Neurol. 1996;374:362-75.

[9] Dickinson PS, Mecsas C, Marder E. Neuropeptide fusion of two motor-pattern generator circuits. Nature. 1990;344:155-8.

[10] Skiebe P. Neuropeptides are ubiquitous chemical mediators: Using the stomatogastric nervous system as a model system. J Exp Biol. 2001;204:2035-48.

[11] Marder E, Calabrese RL. Principles of rhythmic motor pattern generation. Physiological reviews. 1996;76:687-717.

[12] 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.

[13] Fricker LD. Neuropeptidomics to study peptide processing in animal models of obesity. Endocrinology. 2007;148:4185-90.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] Angel TE, Aryal UK, Hengel SM, Baker ES, Kelly RT, Robinson EW, et al. Mass spectrometry-based proteomics: existing capabilities and future directions. Chem Soc Rev. 2012;41:3912-28.

[19] 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.

[20] Ma M, Szabo TM, Jia C, Marder E, Li L. Mass spectrometric characterization and physiological actions of novel crustacean C-type allatostatins. Peptides. 2009;30:1660-8.

[21] Ma M, Sturm RM, Kutz-Naber KK, Fu Q, Li L. Immunoaffinity-based mass spectrometric characterization of the FMRFamide-related peptide family in the pericardial organ of Cancer borealis. Biochem Biophys Res Commun. 2009;390:325-30.

[22] 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.

[23] Hui L, Zhang Y, Wang J, Cook A, Ye H, Nusbaum MP, et al. Discovery and functional study of a novel crustacean tachykinin neuropeptide. ACS Chem Neurosci. 2011;2:711-22.

[24] 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.

[25] Chen R, Jiang X, Conaway MC, Mohtashemi I, Hui L, Viner R, et al. Mass spectral analysis of neuropeptide expression and distribution in the nervous system of the lobster Homarus americanus. J Proteome Res. 2011;9:818-32.

[26] Fricker LD. Analysis of mouse brain peptides using mass spectrometry-based peptidomics: implications for novel functions ranging from non-classical neuropeptides to microproteins. Mol Biosyst. 2010;6:1355-65.

[27] 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.

[28] 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.

[29] 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;77:7783-95.

[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.

[31] 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.

[32] Aquiloni L, Giulianini PG, Mosco A, Guarnaccia C, Ferrero E, Gherardi F. Crustacean hyperglycemic hormone (cHH) as a modulator of aggression in crustacean decapods. PLoS One. 2012;7:e50047.

[33] Chung JS, Zmora N, Katayama H, Tsutsui N. Crustacean hyperglycemic hormone (CHH) neuropeptidesfamily: Functions, titer, and binding to target tissues. Gen Comp Endocrinol. 2010;166:447-54.

[34] 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.

[35] Tao Y, Quebbemann NR, Julian RR. Discriminating D-amino acid-containing peptide epimers by radical-directed dissociation mass spectrometry. Anal Chem. 2012;84:6814-20.

[36] Soyez D, Toullec JY, Montagne N, Ollivaux C. Experimental strategies for the analysis of D-amino acid containing peptides in crustaceans: a review. J Chromatogr B Analyt Technol Biomed Life Sci. 2011;879:3102-7.

[37] Ewing MA, Wang J, Sheeley SA, Sweedler JV. Detecting D-amino acid-containing neuropeptides using selective enzymatic digestion. Anal Chem. 2008;80:2874-80.

[38] Sheeley SA, Miao H, Ewing MA, Rubakhin SS, Sweedler JV. Measuring D-amino acid-containing neuropeptides with capillary electrophoresis. Analyst. 2005;130:1198-203.

[39] Buczek O, Yoshikami D, Bulaj G, Jimenez EC, Olivera BM. Post-translational amino acid isomerization: a functionally important D-amino acid in an excitatory peptide. J Biol Chem. 2005;280:4247-53.

[40] Kanu AB, Dwivedi P, Tam M, Matz L, Hill HH, Jr. Ion mobility-mass spectrometry. J Mass Spectrom. 2008;43:1-22.

[41] Bohrer BC, Merenbloom SI, Koeniger SL, Hilderbrand AE, Clemmer DE. Biomolecule analysis by ion mobility spectrometry. Annu Rev Anal Chem (Palo Alto Calif). 2008;1:293-327.

[42] Saminathan IS, Wang XS, Guo Y, Krakovska O, Voisin S, Hopkinson AC, et al. The extent and effects of peptide sequence scrambling via formation of macrocyclic B ions in model proteins. J Am Soc Mass Spectrom. 2010;21:2085-94.

[43] Molesworth S, Osburn S, Van Stipdonk M. Influence of amino acid side chains on apparent selective opening of cyclic b5 ions. J Am Soc Mass Spectrom. 2010;21:1028-36.

[44] Bythell BJ, Knapp-Mohammady M, Paizs B, Harrison AG. Effect of the His residue on the cyclization of b ions. J Am Soc Mass Spectrom. 2010;21:1352-63.

[45] Bleiholder C, Osburn S, Williams TD, Suhai S, Van Stipdonk M, Harrison AG, et al. Sequence-scrambling fragmentation pathways of protonated peptides. J Am Chem Soc. 2008;130:17774-89.

[46] Harrison AG, Young AB, Bleiholder C, Suhai S, Paizs B. Scrambling of sequence information in collision-induced dissociation of peptides. J Am Chem Soc. 2006;128:10364-5.

[47] Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Proc Natl Acad Sci U S A. 2004;101:9528-33.

[48] Coon JJ, Ueberheide B, Syka JE, Dryhurst DD, Ausio J, Shabanowitz J, et al. Protein identification using sequential ion/ion reactions and tandem mass spectrometry. Proc Natl Acad Sci U S A. 2005;102:9463-8.

[49] Coon JJ. Collisions or electrons? Protein sequence analysis in the 21st century. Anal Chem. 2009;81:3208-15.

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

Chapter 2

High-definition *De Novo* Sequencing of Crustacean Hyperglycemic Hormone (CHH)-family Neuropeptides

Adapted from: Chenxi Jia, Limei Hui, Weifeng Cao, Christopher B. Lietz, Xiaoyue Jiang, Ruibing Chen, Adam D. Catherman, Paul M. Thomas, Ying Ge, Neil L. Kelleher and Lingjun Li, High-definition *De Novo* Sequencing of Crustacean Hyperglycemic Hormone (CHH)-family Neuropeptides. *Molecular and Cellular Proteomics*, 2012, 11:1951-64.

Abstract

A complete understanding of the biological functions of large signaling peptides (>4 kDa) require comprehensive characterization of their amino acid sequences and post-translational modifications (PTMs), which presents significant analytical challenges. In the past decade, there has been great success with mass spectrometry (MS)-based *de novo* sequencing of small neuropeptides. However, these approaches are less applicable to larger neuropeptides due to the inefficient fragmentation of peptides larger than 4 kDa and their lower endogenous abundance. The conventional proteomics approach focuses on large-scale determination of protein identities via database searching, lacking the ability for in-depth elucidation of individual amino acid residues. Here, we present a multifaceted MS approach for identification and characterization of large crustacean hyperglycemic hormone (CHH)-family neuropeptides, a class of peptide hormones that play central roles in the regulation of many important physiological processes of crustaceans. Six crustacean CHH-family neuropeptides (8-9.5 kDa), including two novel peptides with extensive disulfide linkages and PTMs, were fully sequenced without reference to genomic databases. High-definition *de novo* sequencing was achieved by a combination of bottom-up, off-line top-down, and on-line top-down tandem MS methods. Statistical evaluation indicated that these methods provided complementary information for sequence interpretation and increased the local identification confidence of each amino acid and PTM. Further investigations by MALDI imaging MS mapped the spatial distribution and colocalization patterns of various CHH-family neuropeptides in the neuroendocrine organs, revealing that two CHH-subfamilies are involved in distinct signaling pathways. Conformational changes in CHH peptides caused by pyro-Glu modification were also observed in ion mobility MS.

2.1 Introduction

Neuropeptides and hormones comprise a diverse class of signaling molecules involved in numerous essential physiological processes, including analgesia, reward, food intake, learning and memory [1]. Disorders of the neurosecretory and neuroendocrine systems influence many pathological processes. For example, obesity results from failure of energy homeostasis in association with endocrine alterations [2, 3]. Previous work from our lab used crustaceans as model organisms found that multiple neuropeptides were implicated in control of food intake, including RFamides, tachykinin related peptides, RYamides, and pyrokinins [4, 5].

Crustacean hyperglycemic hormone (CHH) family neuropeptides play a central role in energy homeostasis of crustaceans [6-16]. Hyperglycemic response of the CHHs was first reported after injection of crude eyestalk extract in crustaceans. Based on their preprohormone organization, the CHH family can be grouped into two sub-families: subfamily-I containing CHH, and subfamily-II containing molt-inhibiting hormone (MIH) and mandibular organ-inhibiting hormone (MOIH). The preprohormones of the subfamily-I have a CHH precursor related peptide (CPRP) that is cleaved off during processing; and preprohormones of the subfamily-II lack the CPRP [8]. Uncovering their physiological functions will provide new insights into neuroendocrine regulation of energy homeostasis.

Characterization of CHH-family neuropeptides is challenging. They are comprised of more than 70 amino acids and often contain multiple post-translational modifications (PTMs) and complex disulfide bridge connections [6]. In addition, physiological concentrations of these peptide hormones are typically below picomolar level, and most crustacean species do not have available genome and proteome databases to assist MS-based sequencing.

MS-based neuropeptidomics provides a powerful tool for rapid discovery and analysis of a large number of endogenous peptides from the brain and the central nervous system. Our group and others have greatly expanded the peptidomes of many model organisms [3, 17-32]. For example, we have discovered more than 200 neuropeptides with several neuropeptide families consisting of as many as 20-40 members in a simple crustacean model system [5, 24-30, 33]. However, a majority of these neuropeptides are small peptides with 5-15 amino acid residues long, leaving a gap of identifying larger signaling peptides from organisms without sequenced genome. The observed lack of larger size peptide hormones can be attributed to the lack of effective *de novo* sequencing strategies for neuropeptides larger than 4 kDa, which are inherently more difficult to fragment using conventional techniques [33-36]. Although classical proteomics studies examine larger proteins, these tools are limited to identification based on database searching with one or more peptides matching without complete amino acid sequence coverage [35, 37].

Large populations of neuropeptides from 4-10 kDa exist in the nervous systems of both vertebrates and invertebrates [8, 30, 38]. Understanding their functional roles requires sufficient molecular knowledge and a unique analytical approach. Therefore, developing effective and reliable methods for *de novo* sequencing of large neuropeptides at the individual amino acid residue level is an urgent gap to fill in neurobiology. In this study, we present a multifaceted MS strategy aimed at high-definition *de novo* sequencing and comprehensive characterization of the CHH-family neuropeptides in crustacean central nervous system. The high-definition *de novo* sequencing was achieved by a combination of three methods: (1) enzymatic digestion and LC-tandem mass spectrometry (MS/MS) bottom-up analysis to generate detailed sequences of proteolytic peptides; (2) off-line LC fractionation and subsequent top-down MS/MS to obtain

high-quality fragmentation map of intact peptides; and (3) on-line LC coupled to top-down MS/MS to allow rapid sequence analysis of low abundance peptides. Combining the three methods overcomes the limitations of each, and thus offers complementary and high-confidence determination of amino acid residues and complete characterization of site-specific PTMs. We report the complete sequence analysis of six CHH-family neuropeptides including the discovery of two novel peptides. With the accurate molecular information, MALDI imaging and ion mobility MS were conducted for the first time to explore their anatomical distribution and biochemical properties.

2.2 Experimental Procedures

2.2.1 Materials and Chemicals

All chemical reagents were obtained from Sigma-Aldrich (St. Louis, MO) unless otherwise noted. Optima grade formic acid, ACN, water, and methanol were purchased from Fisher Scientific (Pittsburgh, PA).

2.2.2 Animals, Tissue Dissection and Extraction

Blue crabs *Callinectes sapidus* and Jonah crabs *Cancer borealis* were shipped from the Fresh Lobster Company (Gloucester, MA), and then maintained in artificial seawater. The animals were anesthetized in ice, and the sinus glands (SGs) and pericardial organs (POs) were dissected and collected in acidified methanol. The tissue was then homogenized and extracted with acidified methanol. After centrifugation, supernatant fractions were combined and concentrated to dryness. The sample was re-suspended in 100 μL of water for further analysis [5]. The detailed protocol is described in Supplementary Materials.

2.2.3 HPLC Fractionation

HPLC separations were performed with a Waters Alliance HPLC system (Milford, MA). 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 with a Rainin Dynamax FC-4 fraction collector, followed by lyophilized, re-suspended in 20 μL water, and stored in −80°C.

2.2.4 MALDI-TOF/TOF Analysis

A model 4800 MALDI-TOF/TOF analyzer (Applied Biosystems, Framingham, MA) equipped with a 200 Hz, 355 nm Nd:YAG laser was used. Acquisitions were performed in positive ion reflectron mode. Instrument parameters were set using the 4000 Series Explorer software (Applied Biosystems). Mass spectra were obtained by averaging 1000 laser shots covering mass range *m/z* 500–4000 in reflectron mode and *m/z* 2000–10000 in linear mode. MS/MS was achieved by 1 kV CID. For sample analysis, 0.4 μL of sample was spotted on MALDI plate first and allowed to dry followed by the addition of 0.4 μL 2,5-dihydroxybenzoic acid (DHB) matrix [4].

2.2.5 Bottom-up MS on Nano-LC-ESI-QTOF

An aliquot of 3 μL peptide fraction was reduced and alkylated by DTT and iodoacetamide (IAA), followed by digestion with trypsin, Glu-C and Lys-C [39] (see Supplementary Materials for details). Nano-LC-ESI-QTOF MS/MS was performed using a Waters nanoAcquity UPLC system coupled to a QTOF Micro mass spectrometer (Waters, Milford, MA) as described previously [5]. The MS/MS raw data were converted to peak list (.pkl) files using ProteinLynx software 2.4 (Waters) [5]. Peptides were identified by searching against a NCBI protein database

using the Mascot search engine. Database searching was restricted to tryptic peptides with two missed cleavages allowed. Carboxylmethyl cysteine was specified as fixed modifications. Precursor and MS/MS tolerances were set within 300 ppm and 0.6 Da for monoisotopic mass, respectively. Peptide charge states include 1+, 2+ and 3+ charged peptides.

2.2.6 Off-line Top-down MS on ESI-LTQ-FTICR

A 0.5 µL of 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 sample was analyzed using a 7T linear trap quadrupole (LTQ)/Fourier transform ion cyclotron resonance (FTICR) (LTQ-FT 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 [40]. For CID and 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 [40]. All FTICR spectra were processed with Xtract Software (Xcalibur 2.0.5, 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. The assigned ions were manually validated to ensure the quality of assignments [40].

2.2.7 On-line Top-down MS on Nano-LC-ESI-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 0.2% formic acid. On-line top-down MS was carried out on an Ultimate 3000 RSLCnano system coupled to an

Orbitrap Elite mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). A 0.5 μL of peptide sample was injected onto a 2 cm, 150 μm i.d. PLRP-S (dp 5 μm, pore size 1000Å) trap column. A 10 cm, 75 μm i.d. PLRP-S column was used for separation. The gradient was delivered at 300 nL/min starting at 5% B (95% acetonitrile and 0.2% formic acid) and rose to 10% B at 7 min, 50% B at 50 min, and 85% B at 58 min. The mass spectrometer was operated in the data-dependent mode to switch automatically between full-MS (scan 1), higher-energy collision dissociation (HCD)-MS^2 (scan 2), and electron transfer dissociation (ETD)-MS^2 (scan 3). The isolation width was set at 10 Da [35]. The data processing method is the same as off-line top-down method described above.

2.2.8 MALDI Imaging

Immediately following dissection, the eyestalk was embedded in gelatin (100 mg/mL aqueous) and snap-frozen. Sectioning into 12 μm slices at −25 °C was performed on a cryostat (Microtom HM505E, Waldorf, Germany), and the slices were thaw-mounted onto a glass slide (Bruker Daltonics). An airbrush was used to spray coat the tissues with DHB. The airbrush was held perpendicular to the glass slide at a distance of 35 cm. Five coats of matrix were applied by spraying each sample for 30 s with 1 min dry time between each application [5].

Mass spectrometric analyses were performed in the linear, positive mode at +20 kV accelerating potential on a time-of-flight mass spectrometer (Bruker Autoflex III TOF-TOF; Bruker Daltonics, Bremen, Germany), which was equipped with a Smartbeam laser capable of operating at a repetition rate of 200 Hz with optimized delayed extraction time and laser beam size was set to medium. Laser energy was optimized for signal-to-noise in each preparation. Using Bruker Protein Standard 1 (Bruker Daltonics, Bremen, Germany), a linear external calibration was applied to the instrument prior to data collection. Mass spectral data sets were

acquired over a whole eyestalk section using FlexImaging software (Bruker Daltonics, Bremen, Germany) in the mass range of *m/z* 3,000 to 10,000, with a raster step size of 50 μm and 500 laser shots per spectrum. After data acquisition, molecular images were reconstituted using the FlexImaging software. Data was normalized using FlexImaging software, and each *m/z* signal plotted +/− 10 mass units [5].

2.2.9 Ion Mobility MS

All ESI-ion mobility (IM)-MS experiments were performed using a Synapt G2 HDMS mass spectrometer equipped with a nano-ESI ion source and MassLynx data processor (Waters, Milford, MA, USA). Instrument acquisition parameters used were as follows: an inlet capillary voltage of 3.0 kV, a sampling cone setting of 35 V, and a source temperature of 120 °C. The argon pressure in the traveling wave ion guide trap and the traveling wave ion guide transfer were 2.44×10^{-2} and 2.61×10^{-2} mbar, respectively. The wave height, the wave velocity, and the nitrogen pressure in the traveling wave IM drift cell were 32.0 V, 850 m/s, and 2.96 mbar, respectively. Samples were directly infused into the mass spectrometer at a rate of 0.5–0.8 μL/min. All IM-MS acquisitions were acquired for 5 min. Data processing was conducted using Waters MassLynx 4.1 and DriftScope 2.1. The experimental procedure for calibration of CCS was described in Supplementary Materials.

2.3 Results

2.3.1 Establishment and Validation of the High-definition *De Novo* Sequencing Strategy

Our MS strategy for identification of CHH-family neuropeptides from the crustacean nervous system (Figure 1) involves three steps: peptide candidate scanning, *in silico* homology searching, and *de novo* sequencing. Although the cDNA sequence of *C. sapidus* $^{Cas\text{-}SG\text{-}}$CHH (*Cas C. sapidus*, SG sinus gland) preprohormone has been obtained by PCR-based cloning strategy,

its amino acid sequence has not been identified by MS or Edman degradation [15, 16]. Here, the multifaceted strategy is established and validated by performing MS-sequencing of the *Cas*-SG-CHH peptide.

Peptide Candidate Scanning. Two unique features of the CHH-family neuropeptides, a molecular weight (MW) ranging from 8-10 kDa and the presence of three disulfide bonds [11], were used as screening criteria for candidate identification. The sinus gland organs were pooled from ten animals, homogenized, lyophilized and subjected to reversed phase (RP)-HPLC fractionation (Figure 2A), followed by off-line direct infusion on an ESI-LTQ-FTICR mass spectrometer. Multiply charged CHH-family peptide ions were detected in the high-resolution spectrum of Fraction #17 (Figure 2B), and the accurate MW of this large peptide was determined as 8472.948 Da.

To screen for disulfide bonds, the peptide candidate was treated with DTT and IAA respectively to reduce the disulfide bonds and then alkylate the free thiol groups followed by analysis via MALDI-MS. MALDI can tolerate higher levels of salt and mostly produce singly charged ions facilitating mass comparison. Figure 2C and 2D show the MALDI mass spectra of the original and derivatized peptides, respectively. The ions at *m/z* 8728, 8777, and 8824 correspond to the reduced peptide alkylated with attachment of 4, 5 and 6 carbamidomethyl groups, respectively. Incomplete alkylation might be caused by the large peptide size. These results suggest the presence of three disulfide bonds in the peptide *Cas*-SG-CHH.

In Silico Sequence Homology Searching. Sequence homology searching can aid in *de novo* sequencing of small neuropeptides. Here, we designed a bottom-up method to extend the utility of the protein database for sequencing of large neuropeptides. The candidate peptide was digested by trypsin and analyzed by LC-MS/MS, followed by Mascot searching against NCBI

database. The first two hits were the CHH preprohormones of *C. sapidus* (52% sequence match) and *Portunus trituberculatus* (29% sequence match). The goal of our method is to analyze neuropeptides from species without proteome and genome databases. So, we assumed that there were no such knowledge for this target peptide *Cas*-SG-CHH, and the homologous CHH preprohormone from *P. trituberculatus* (Figure S1) was used as a reference sequence for *de novo* sequencing. CHH family members share the characteristic feature of six cysteines located in the identical or similar positions, i.e. C^{7}, C^{23}/C^{24}, C^{26}/C^{27}, C^{39}/C^{40}, C^{43}/C^{44}, C^{52}/C^{53}. Therefore, the sequence AA^{64}-AA^{139} of *P. trituberculatus* CHH preprohormone (Figure S1) was mined as a reference peptide for the following *de novo* sequencing.

Bottom-up De Novo Sequencing. The tryptic fragments of *Cas*-SG-CHH were analyzed by LC-QTOF-MS/MS and MALDI-TOF/TOF-MS/MS. Figure 3A is the MALDI-TOF/TOF mass spectrum of tryptic peptides of the Fraction #17. Figure 3B displays a representative MS/MS spectrum of the tryptic peptide. Using the preprohormone sequence of *P. trituberculatus* CHH as a reference, the MS/MS spectra were carefully analyzed by software PepSeq [5] to assign fragment ions and determine peptide sequences. Table S1 lists all the sequenced tryptic peptides arising from *Cas*-SG-CHH. By this bottom-up sequencing method, 81% of the sequence was determined.

Off-line Top-down De Novo Sequencing. To determine the rest of the sequence, the intact *Cas*-SG-CHH was fragmented with top-down MS using CID on an ESI-ultra-high resolution (UHR)-QTOF maXis mass spectrometer via direct infusion (Figure 4A). Two sets of b_{55}-b_{65} and y_{7}-y_{15} ions, generated by cleavage of amide bonds closed to the peptide C-terminus, were detected with intense signals. However, only a few of fragment ions from middle region of the peptide chain were observed. Although only 14% of b and y ions were assigned, a sequence tag

^{56}LLIMDNFEEY^{65} (in this study, the isobaric I/L residues were assigned using homologous sequences) was confidently determined from the intense C-terminal fragmentation (Figure 4A). Similarly, this intact peptide was fragmented on a different instrument, an ESI-LTQ-FTICR mass spectrometer, with CID and ECD, resulting in poor fragmentation as well (data not shown). A possible explanation is that fragmentation of this peptide is hindered by its tertiary structure resulting from three disulfide bonds which crosslink the residues between Cys^{7}-Cys^{52}. Therefore, denaturation of peptides by breaking disulfide bonds with DTT and IAA could facilitate peptide fragmentation. However, the large peptide size causes incomplete alkylation as shown previously (Figure 2D).

Here, we adopted a different strategy for peptide denaturation, where the peptide was treated with DTT in urea solution for complete reduction of disulfide bonds, and then stored in 50% ACN containing 2% formic acid. The acidic environment prevented the disulfide bond from reforming although the free thiol groups were not blocked by protective groups. After storing the sample either at 5 ^{o}C for 5 hours or -20 ^{o}C for 4 days, no disulfide bond formation was observed (data not shown). The subsequent fragmentation of the reduced peptide on the LTQ-FTICR with CID and ECD produced 42% of b- and y-type ions and 39% of c- and z-type ions, respectively. Figure 4B is the ECD spectrum, and Figure 5A shows the fragmentation map of ECD and CID. As a result, 75% of sequence and the location of two PTMs, N-terminal pyroGlu and C-terminal amidation, were identified by this off-line top-down strategy.

On-line Top-down De Novo Sequencing. Top-down LC-MS/MS was carried out on an LTQ-Orbitrap Elite mass spectrometer. Here, the crude tissue extract treated with DTT was directly analyzed with top-down LC-MS/MS under data-dependent mode switching between ETD and

HCD. Figure 4C shows the HCD spectrum, and Figure S2 shows the fragmentation map of ETD and HCD for the *Cas*-SG-CHH peptide.

Sequence Assembling. Figure 5C summarizes the sequence coverage identified by the three *de novo* sequencing methods (bottom-up, off-line top-down and on-line top-down). These methods are complementary and enable maximizing the sequence identification. The identified sequence percentages are 81%, 75%, and 50%, respectively. Only two residues AA^{32} and AA^{35} were not identified, which can be further determined by so-called Dipeptide Mass Deduction and Peptide Mass Deduction.

Manual De Novo Sequencing. Dipeptide Mass Deduction: The mass of AA^{31}+AA^{32} can be calculated by c_{32} minus c_{30} (3826.720 - 3556.569 = 270.151 Da), only matching to dipeptide RN (Dipeptide mass table in ProteinProspector, http://prospector.ucsf.edu/prospector/html/misc/dipep.htm). So, AA^{32} is an asparagine because AA^{31} was already determined as an arginine (Figure 5C). Peptide Mass Deduction: The residue AA^{35} can be determined by peptide MW minus the rest of the sequence (8478.878 – 8379.802 = 99.076 Da, mass value of a valine residue). Finally, the complete sequence was determined.

In summary, this multifaceted MS strategy offers a systematic approach to elucidate amino acid sequences and characterize PTMs of the CHH-family neuropeptides, which is applicable to characterization of large intact peptide hormones with or without known cDNA sequences.

2.3.2 CHH-family Neuropeptidome in *C. sapidus* and *C. borealis*.

With the strategies described above, we have discovered and identified six CHH-family neuropeptides and two modified isoforms from *C. sapidus* and *C. borealis* (Table 1). For each peptide, the MWs of both intact and "reduced" peptides were measured to improve the identification confidence ("reduced" refers to the peptides after DTT reduction), where the

theoretical and experimental MWs match within 17.5 ppm. The sequence alignments among homologous species are shown in Figure 6. This high-level confidence for complete sequence coverage arises from a combination of the three *de novo* sequencing methods.

2.3.3 Identification of Two Novel CHH-family Neuropeptides in *C. borealis*.

The Novel MIH. Although *C. borealis* is a widely used animal model in neurobiological studies, there is no available genomic database, making peptide identification in this model organism challenging [24, 26]. With the multifaceted MS strategy, we found one novel MIH in the sinus gland. The accurate MWs of intact and "reduced" forms were determined as 8932.302 and 8938.230 Da, respectively. From the 6-Da mass difference, we inferred that this peptide contains three disulfide bonds and thus placing this peptide as a putative candidate for the CHH-family neuropeptides. The homology search using bottom-up proteomic method resulted in a hit of the MIH from a homologous species, *Cancer pagurus* [41], with an 89% of matched peptide coverage (Figure S3). In the subsequent bottom-up *de novo* sequencing, multiple proteases trypsin, Glu-C and Lys-C were employed to generate proteolytic peptides from different cleavage site. Followed by LC-MS/MS analysis, computer-assisted sequencing resulted in identification of proteolytic peptides listed in Table 2. In addition, these results indicated that Mascot search caused false positive identification of the tryptic peptide ^{68}QWVGILGAGRE^{78} (Figure S3). By overlapping these proteolytic peptides and referring the homologous sequence of *Cancer pagurus* $^{Cap\text{-}SG\text{-}}$MIH, we identified the sequence AA^{1}-AA^{59} of the new $^{Cab\text{-}SG\text{-}}$MIH in *C. borealis*. The sequence AA^{60}-AA^{77} remained unknown.

Previous work on sequencing of the *C. sapidus* $^{Cas\text{-}SG\text{-}}$CHH showed that CID fragmentation of intact (non-reduced) CHH peptide scan result in ample cleavage of C-terminal amide bonds, facilitating identification of the C-terminal residues. Similarly, we fragmented this putative $^{Cab\text{-}SG\text{-}}$

MIH peptide on ESI-LTQ-FTICR using CID, and observed one set of b ions with intense signals arising from sequential cleavage of the C-terminal amide bonds. Accordingly, AA^{68}-$AA^{77}$was determined as ^{68}EWVGILGAGS^{77} with a C-terminal amidation. The remaining task was to identify AA^{60}-AA^{67}.

In the homologous $^{Cap\text{-}SG\text{-}}$MIH (Figure S3), AA^{67} is an arginine and thus AA^{60}-AA^{67} forms a tryptic peptide by trypsin digestion. So, it is possible that in the putative $^{Cab\text{-}SG\text{-}}$MIH the residue AA^{60}-AA^{67} also form a tryptic peptide. Based on this hypothesis, the MW of the tryptic peptide can be calculated by Peptide Mass Deduction, i.e., the measured $^{Cab\text{-}SG\text{-}}$MIH peptide MW minus the calculated remaining sequence mass (8938.230 – 8003.781 = 934.449 Da). By carefully searching the bottom-up *de novo* sequencing data, one tryptic peptide was found matching the MW 934.449 Da (Figure 7B). Analysis of the MS/MS spectrum using PepSeq determined its sequence as ^{60}TAEMSQLR^{67} (Figure 7C). With the addition of this fragment, the complete sequence of this novel $^{Cab\text{-}SG\text{-}}$MIH was determined. To confirm it, off-line and on-line top-down MS/MS were carried out. In the resulting ECD spectrum (Figure 7D), observation of the series of $z^{\cdot}$ ions confirms the sequence ^{63}MSQL^{66} which was determined by the above bottom-up tryptic peptide in Figure 7C. The fragmentation map resulting from ECD and CID (Figure 5B) shows a 58% of identified sequence coverage, and that of ETD and HCD gives 50% of sequence coverage (Figure S4). Figure 7E illustrates the sequencing process step by step. This approach offers a confident and accurate sequence analysis for the discovery of novel large neuropeptides without genomic information.

The Novel MOIH. In *C. borealis* sinus gland, we found one putative $^{Cab\text{-}SG\text{-}}$MOIH whose MWs of intact and "reduced" forms were observed as 9252.346 and 9258.278 Da, respectively. Once again, 6-Da mass difference confirmed the candidate identity. Homology search hit the *Cancer*

pagurus $^{Cap\text{-}SG}$MOIH (Figure S5) with an 89% of matched sequence coverage [42]. With bottom-up *de novo* sequencing method using multiple protease digestion, twelve proteolytic peptides corresponding to 97% of sequence coverage (Table 3) were identified to generate a putative sequence

RRINNDCQNFIGNRAMYEKVDWICKDCANIFRQDGLLNNCRSNCFYNTEFLWCIDATEN

TR**XX**EQLEQWAAILGAGWN (**X** stands for unknown amino acid residue). The mass of the unknown residues AA^{62}-AA^{63} was calculated by Peptide Mass Deduction: measured MW of the intact peptide minus theoretical mass of the rest of residues 9258.278 – 8993.151 = 265.127 Da, which matches the only dipeptide HK in Dipeptide mass table of ProteinProspector. Compared with the homologous $^{Cap\text{-}SG}$MOIH sequence, the sequence order should be HK rather than KH. By the off-line top-down method, this putative $^{Cab\text{-}SG}$MOIH could not be selected for fragmentation due to low abundance and suppressed ionization (data not shown). Here, the on-line top-down method shows the advantage that this putative $^{Cab\text{-}SG}$MOIH was fragmented using ETD and HCD by on-line precursor ion selection, which produced a fragmentation map with 52% of sequence coverage (Figure S6).

2.3.4 MS-based Distribution Map and Conformation Analysis Reveal Biological Significance

Distribution Map by MALDI Imaging. The functional roles that various compounds play in the organism are closely related to their locations. The X-organ/sinus gland complex located in eyestalks represents a major neurodocrine structure in decapod crustaceans [9]. Previous studies using immunohistochemical techniques for peptide profiling indicate that the subfamily-I peptides (CHH) rarely overlap with the subfamily-II peptides (MIH and MOIH) [8, 13, 14, 43]. However, the immunohistochemical method suffers from antibody cross-reaction problem as the

epitope peptides share a high degree of sequence homology (see Figure 6) [13]. To overcome this limitation, we demonstrated the first use of MALDI MS imaging technique to map the endogenous CHH-family peptides in crustacean neurosecretory structures.

Figure 8A shows direct tissue MALDI MS analysis of Jonah crab *C. borealis* sinus gland, which contains *Cab*-SG-CPRP dimer [5], *Cab*-SG-CHH-I, *Cab*-SG-CHH-II, and *Cab*-SG-MOIH (MWs and PTMs are listed in Table 1). The MALDI imaging results of entire eyestalk reveal that all these peptides are located in the sinus gland (Figure 8B). Interestingly, the zoom-in images (Figures 8C-G) of the sinus gland illustrate that *Cab*-SG-CPRP and subfamily-I peptides, *Cab*-SG-CHH-I and *Cab*-SG-CHH-II are co-localized; and subfamily-II peptides, *Cab*-SG-MIH and *Cab*-SG-MOIH are also co-localized. Moreover, the *Cab*-SG-CPRP, *Cab*-SG-CHH-I and *Cab*-SG-CHH-II have almost no overlap with *Cab*-SG-MIH and *Cab*-SG-MOIH, exhibiting distinct distribution patterns for the two CHH subfamily peptides. Furthermore, the similar distribution pattern was observed from MALDI imaging of the blue crab *C. sapidus* sinus gland, which shows that the *Cas*-SG-MIH and *Cas*-SG-CHH are differentially distributed, and *Cas*-SG-CPRP and *Cas*-SG-CHH are co-localized (data not shown). These results are consistent with the previous studies using immunohistochemical method.

Conformation Analysis by Ion Mobility MS. The N-terminal glutamine of neuropeptides can be modified by cyclization of the glutamine via condensation of the α-amino group with the side-chain carboxyl group (pyro-Glu modification). Modified peptides show variation of half-life and biological activity related to conformational changes [44]. The *Cab*-SG-CHH-I and *Cab*-SG-CHH-II in Figure 8A are the modified and unmodified isoforms, respectively. The release sites of the two isoforms are co-localized in the sinus gland (Figure 8C and 8D), suggesting that they exhibit similar tissue-specific distribution patterns. Thus, it is interesting to investigate the potential

peptide conformation change induced by this modification. Figure 8H shows the calibrated collision cross section (CCS) [45] of the *Cab*-SG-CHH-I and *Cab*-SG-CHH-II with 0.4 nm^2 of CCS difference measured by ion mobility spectrometry. This suggests that the N-terminal pyro-Glu may alter the peptide's gas-phase conformation, and possibly the solution-phase conformation as well.

2.3.5 Statistical Analysis Evaluates Performance Characteristics of the Three *De Novo* Sequencing Strategies

Complementary Characteristics for Complete Sequence Interpretation. Implementation of three *de novo* sequencing methods provides more confident and effective sequence elucidation of large peptides. Here, we use *Cas*-SG-CHH, *Cas*-SG-MIH, *Cab*-SG-CHH and *Cab*-SG-MIH as model peptides to evaluate the three figures of merit: number of identified residues, number of residues in sequence tags, and number of cleavage sites.

The ultimate goal of *de novo* sequencing is to determine the complete amino acid sequence of a given peptide or protein. The Venn diagram in Figure 9A shows the contributions of the three methods to the number of identified residues. While the four model peptides contain a combined total of 274 residues, only 85 residues are commonly identified by all of the three methods. Each of the three methods enables determination of numerous uniquely identified residues (off-line top-down method yields 20 residues, on-line top-down method shows 11 and bottom-up technique generates 35 residues), and a large number of residues are in the overlap in two methods. The bottom-up method yielded the highest identification percentage, showing11% and 22% higher than the off-line and on-line top-down methods, respectively.

Interpretation of consecutive fragment ions can generate sequence tags, which is critical for peptide sequencing in the absence of a sequenced genome. The Venn diagram in Figure 9B

shows the contributions of the three methods to the number of residues that sequence tags contain. The bottom-up method shows the most significant contribution of 65 residues, indicating that bottom-up MS/MS of tryptic peptides exhibit preference to generate consecutive fragment ions for sequence tag interpretation in comparison with top-down MS/MS of intact large peptides.

The numbers of cleavage sites between two adjacent residues were evaluated. Cleavage site includes both amide bond cleavage and α-carbon bond cleavage primarily produced by proton-driven (CID and HCD) and electron-driven (ECD and ETD) dissociations, respectively. The Venn diagram in Figure 9C illustrates that of the 302 cleavage sites observed, 113 are shared by the three methods.

Figure 9D summarizes the impact of using multiple *de novo* sequencing methods. Combined use of the three methods leads to 60%, 100% and 36% of boosts in the values of the three evaluated aspects, respectively. These results indicate that the three *de novo* sequencing methods have complementary effects on improving sequence coverage.

Local Identification Confidence for Individual Amino Acid Residue. The combination of off-line and on-line top-down methods can also improve the local identification confidence for each amino acid residue. The *de novo* sequencing algorithm PEAKS has a scoring function for evaluation of the local confidence [46]. Similarly, we established a simpler evaluation scheme as shown in Figure 10A and legend. Figure 10B plots the confidence scores of each residue of $^{Cas\text{-}SG\text{-}}$CHH. After combination of the two methods, most of scores for individual residues were increased. Figure 10C summarizes the average scores of off-line top-down, on-line top-down and combination for the four model peptides, $^{Cas\text{-}SG\text{-}}$CHH, $^{Cas\text{-}SG\text{-}}$MIH, $^{Cab\text{-}SG\text{-}}$CHH and $^{Cab\text{-}SG\text{-}}$MIH. All the confidence scores are increased with combination of the two methods.

The use of multiple fragmentation techniques, including ECD, ETD, CID, and HCD, produces fragment ions with broader mass ranges in top-down fragmentation of large peptides. Each fragmentation technique has a unique and complementary mass range preference. Figure S7 plots the correlation between length of peptide fragment ions and fragmentation techniques. CID gives most fragment ions located in low and high mass ranges, containing 20 and 70 residues. HCD and ETD give predominantly fragment ions covering the low mass range. ECD produces the fragment ions spanning a broad mass range.

2.4 Discussion

2.4.1 Advantages of the High-definition *De Novo* Sequencing Approach

We have described a multifaceted strategy for identification and characterization of the CHH-family neuropeptides. High-definition *de novo* sequencing not only offers complementary sequencing information but also improves confidence for elucidation of individual amino acid residues and PTMs. Specifically, the bottom-up *de novo* sequencing method, employing the multiple protease digestion and LC-MS/MS technique, enables deep amino acid sequencing by analysis of the overlapped proteolytic peptides [36, 47]. However, complications arise from incomplete characterization of alternative splice forms, labile PTMs and truncated isoforms, leading to misidentification of the native intact peptide isoforms [34, 48, 49]. Combining bottom-up and top-down methods can overcome this issue. Off-line and on-line top-down *de novo* sequencing methods combined use of multiple fragmentation techniques CID, ECD, HCD and ETD [50], offering comprehensive cleavage of intact large peptide molecules while retaining site-specific PTMs. This allows us to obtain the precise molecular details of these important peptide hormones.

Off-line Top-down Strategy. One challenge for top-down MS/MS of large peptides and proteins is the limited capability to detect low-abundance fragment ions. To overcome this problem, our off-line top-down experiment was performed on a platform established by Ge and coworkers [40], where a 7T LTQ-FTICR is coupled with a chip-based nano-ESI, enabling femtomole level of sample consumption and consistent acquisition of spectrum. Thousands of MS/MS scans were averaged to generate a high-quality top-down MS/MS spectrum which facilitated the detection of numerous low-abundance fragment ions. For example, the ECD spectrum in Figure 4B resulted from averaging of 1000 scans with enhanced S/N ratio, which allows us to detect the low-abundance ions and obtain a 75% of sequence coverage (Figure 5A and 5C).

On-line Top-down Strategy. Another challenge of top-down MS/MS is the incompatibility of long duty cycle of FT transient with fast LC-MS time scale where analytes of interest are (co)eluted in a short time window [35]. Here, we carried out the on-line top-down experiment on a platform constructed by Kelleher and co-workers [35], where a newly developed ultrahigh resolution LTQ-Orbitrap Elite FTMS system [51] was coupled with nano-LC for high throughput LC-MS/MS analysis. The optimized ion transfer optics facilitates high speed top-down MS/MS and fast switching between CID, HCD and ETD, enabling sensitive and rapid identification of the eluted peptide ions. For instance, the $^{Cab\text{-}SG\text{-}}$MOIH was co-eluted with $^{Cab\text{-}SG\text{-}}$CHH from LC fractionation. The major component $^{Cab\text{-}SG\text{-}}$CHH suppressed the detection of $^{Cab\text{-}SG\text{-}}$MOIH and thus the off-line top-down fragmentation of $^{Cab\text{-}SG\text{-}}$MOIH cannot be performed. In contrast, using on-line top-down system, HCD and ETD spectra were collected with one scan resulting in a 52% of sequence coverage (Figure S6).

Sample Preparation Strategy. In addition, we employed a simple strategy to dramatically improve the fragmentation efficiency of large peptides containing extensive disulfide bonds. The

peptides are reduced by DTT and stored in acidic solvent followed by MS analysis, instead of IAA alkylation which usually causes incomplete reaction due to large peptide sizes and thus decrease the target ion intensities. Compared with our previous study [33] using IAA alkylation on CHH, this method resulted in ~3 fold increase of observed fragment ions and identified sequence coverage (comparison of fragmentation maps shown in Figure S8). Furthermore, this method can potentially be applied to top-down proteomics to enhance fragmentation efficiency and sequence coverage.

Homology Search Strategy. In this study, the CHH-family peptides share high degree of sequence homology (Figure 6). Therefore, homology search using tryptic peptides as query in Mascot [52] ensures to find the desired homologous preprohormone. Alternatively, other algorithms with homology search function, such as PLGS [53] and PEAKS [46], have been examined to obtain similar results. For utility of the on-line top-down LC-MS/MS data (fragmentation of entire large peptides), ProSightPC [35] enables searching against both a manually curated peptide database and a database of homologous preprohormones. In addition, manual homology search using NCBI Blast (http://blast.ncbi.nlm.nih.gov/Blast.cgi) with the sequence tag LLIMDNFEEY (Figure 4A) yielded the same preprohormones as using Mascot.

2.4.2 Functional Aspects of the Identified CHH-family Neuropeptides

Previous studies in crustacean endocrinology highlighted the central role of the CHH-family neuropeptides in signaling system of energy homeostasis, and proposed a model that employs neurotransmitters to control secretion and release of CHHs, followed by triggering the downstream energy metabolism by the second messengers [6, 7]. However, the precise molecular mechanisms underlying the interactions between neurotransmitters and CHHs at the cellular and network levels remain elusive. This is, in part, due to a lack of analytical capabilities

to identify and characterize these low abundance endogenous signaling molecules in a complex microenvironment.

Our current large neuropeptidome analysis by use of high-definition *de novo* sequencing allowed precise characterization of six CHH-family peptides with PTMs, including the novel MIH and MOIH. Their detailed molecular information can contribute to further functional and physiological studies to explore the modulation mechanism in controlling the animal metabolism, osmoregulation, molting and reproduction, among other essential biological processes [8]. One of the applications is to study their regulatory roles on energy homeostasis, with a goal to establish a simplified neuroendocrine model of energy regulation using crustaceans [7]. A critical element of studying the complex interactions between multiple molecular players is the ability to map their spatial distributions and colocalization patterns. Towards this end, we employed MALDI imaging mass spectrometry and ion mobility MS to analyze the spatial distribution and molecular conformation of several CHH-family neuropeptides.

Distribution Mapping. Generally, each CHH-family peptide plays distinct functional role. For example, CHHs regulate the blood glucose metabolism [15]; MIHs suppress synthesis of ecdysteroids delaying molting[11]; and MOIHs inhibit synthesis of methyl farnesoate and thus control somatic and gonadal growth [41]. Nevertheless, many previous studies have suggested that these large peptides are multifunctional hormones and often display overlapping biological activities [7, 8]. For instance, the osmoregulatory function of CHH is related with crustacean molt cycles [54]. Thus, visualization of the CHH-family peptide hormone release sites can provide insights into an improved understanding of the neurosecretion and signal transduction pathways as well as complex hormonal integration of these processes. Many efforts have been

dedicated to the use of immunohistochemical techniques for distribution analysis of the CHH-family peptides in the neurosecretory system [8, 13, 14, 43]. In general, the immunoreactivities of subfamily-I peptides rarely overlap with those of subfamily-II peptides, but co-localization of individual isoforms among subfamily-II peptides was previously observed. Our MALDI imaging results of both *C. sapidus* and *C. borealis* sinus glands (Figure 8) provide the first direct biochemical evidence to confirm this distribution pattern. The MS-based imaging method provides unambiguous identification and simultaneous measurement and mapping of multiple CHH-family peptides including the intact CHH and CPRP as well as MIH and MOIH using a single eyestalk tissue section. This new method overcomes several limitations of traditional immunohistochemical techniques such as cross-reactivity and limited throughput by offering accurate and simultaneous mapping of numerous endogenous CHH-family peptide isoforms in a single experiment. Nonetheless, the MALDI MS imaging technique does not rival the cellular spatial resolution offered by antibody-based staining technique. Here, we used 50x50 μm^2 pixels for MS image acquisition and visualization, enabling the generation of spatial distribution of various CHH-family neuropeptides within a ~1x1 mm^2 tissue area of this important neuroendocrine organ.

Conformation Analysis. Ion-mobility MS is an analytical technique used to separate and identify ionized molecules in the gas phase based on mass, charge, and shape, with latter being measured as collision cross section (CCS) in an inert carrier buffer gas [45, 55]. When carefully measured, the drift behaviors of peptides reflect information for structural and conformational assessment. It has been shown that some peptides must overcome significant energy barriers to change conformation when transferring from solution to gas phase, indicating that gas-phase measurements made by ion IM-MS may reflect the conformation in solution. Pyroglutamate is

commonly observed in endogenous signaling peptides that contain glutamate or glutamine residues at their N-termini. It makes the peptides more resistant to amino peptidases and has many functional roles. The results of the IM-MS investigations indicated that an N-terminal pyroGlu has the ability to alter the CHH conformation in the gas-phase and may also change the conformation of this peptide hormone in solution. Molecular modeling based on the experimental CCS measurements will be required in future studies to determine the exact gas-phase structures of the modified and unmodified peptides, and more accurate methods to predict any changes that may occur during ESI desolvation are being developed.

2.5 Conclusions

The overall work described here represents a new route to discovery and characterization of large neuropeptides. This multifaceted MS-based strategy involves a comprehensive and systematic implementation of peptide candidate scanning, *in silico* homology searching, *de novo* sequencing, distribution mapping, and conformation analysis. The accurate sequence, PTMs, spatial distribution pattern, and conformational analysis of the CHH-family neuropeptides were elucidated with a combination of MS tools. The high-definition *de novo* sequencing strategy combines well-established bottom-up, on-line top-down and off-line top-down methods, offering complementary sequence information and in-depth characterization of PTMs at residue level. Because CHH-family peptides represent the typical molecular features of large neuropeptides, this multifaceted strategy is applicable to other biological systems.

With the knowledge of the precise molecular details of these CHH-family neuropeptides, the following work focuses on studying their functional roles and modulation mechanism. A novel quantitative top-down MS method described in Chapter 6 enables the monitoring of CHH-family

peptide release and deciphering the regulatory pathways in energy homeostasis and feeding behavior.

2.6 Acknowledgements

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). YG acknowledges support from the Wisconsin Partnership Fund for a Healthy Future and NIH R01HL096971. NLK thanks support from NIH grants R01 GM067193 and P30 DA018310 and the Searle Funds at the Chicago Community Trust (to Chicago Biomedical Consortium). LL acknowledges an H. I. Romnes Faculty Research Fellowship. CJ thanks an Oversea Training Fellowship and UW Vilas Conference Presentation Funds. The authors thank Bruker Daltonics for graciously loaning the Autoflex III MALDI TOF/TOF mass spectrometer. We are also grateful to the UW School of Pharmacy Analytical Instrumentation Center for access to UHR-TOF maXis, and to Lisa Xu and Huseyin Guner for experimental assistance with LTQ-FTICR.

2.7 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] Morton GJ, Cummings DE, Baskin DG, Barsh GS, Schwartz MW. Central nervous system control of food intake and body weight. Nature. 2006;443:289-95.

[3] Fricker LD. Neuropeptidomics to study peptide processing in animal models of obesity. Endocrinology. 2007;148:4185-90.

[4] 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.

[5] 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.

[6] Chung JS, Zmora N, Katayama H, Tsutsui N. Crustacean hyperglycemic hormone (CHH) neuropeptidesfamily: Functions, titer, and binding to target tissues. Gen Comp Endocrinol. 2009;166:447-54.

[7] 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.

[8] 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.

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

[10] 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.

[11] Nakatsuji T, Lee CY, Watson RD. Crustacean molt-inhibiting hormone: structure, function, and cellular mode of action. Comp Biochem Physiol A Mol Integr Physiol. 2009;152:139-48.

[12] Covi JA, Chang ES, Mykles DL. Conserved role of cyclic nucleotides in the regulation of ecdysteroidogenesis by the crustacean molting gland. Comp Biochem Physiol A Mol Integr Physiol. 2009;152:470-7.

[13] Hsu YW, Messinger DI, Chung JS, Webster SG, de la Iglesia HO, Christie AE. Members of the crustacean hyperglycemic hormone (CHH) peptide family are differentially distributed both between and within the neuroendocrine organs of Cancer crabs: implications for differential release and pleiotropic function. J Exp Biol. 2006;209:3241-56.

[14] Chung JS, Webster SG. Expression and release patterns of neuropeptides during embryonic development and hatching of the green shore crab, Carcinus maenas. Development. 2004;131:4751-61.

[15] Chung JS, Zmora N. Functional studies of crustacean hyperglycemic hormones (CHHs) of the blue crab, Callinectes sapidus - the expression and release of CHH in eyestalk and pericardial organ in response to environmental stress. FEBS J. 2008;275:693-704.

[16] Zheng J, Chen HY, Choi CY, Roer RD, Watson RD. Molecular cloning of a putative crustacean hyperglycemic hormone (CHH) isoform from extra-eyestalk tissue of the blue crab

(Callinectes sapidus), and determination of temporal and spatial patterns of CHH gene expression. Gen Comp Endocrinol. 2010;169:174-81.

[17] 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.

[18] 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. In press.

[19] Colgrave ML, Xi L, Lehnert SA, Flatscher-Bader T, Wadensten H, Nilsson A, et al. Neuropeptide profiling of the bovine hypothalamus: thermal stabilization is an effective tool in inhibiting post-mortem degradation. Proteomics. 2011;11:1264-76.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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;77:7783-95.

[28] Chen R, Jiang X, Conaway MC, Mohtashemi I, Hui L, Viner R, et al. Mass spectral analysis of neuropeptide expression and distribution in the nervous system of the lobster Homarus americanus. J Proteome Res. 2011;9:818-32.

[29] Ma M, Sturm RM, Kutz-Naber KK, Fu Q, Li L. Immunoaffinity-based mass spectrometric characterization of the FMRFamide-related peptide family in the pericardial organ of Cancer borealis. Biochem Biophys Res Commun. 2009;390:325-30.

[30] Mita M, Yoshikuni M, Ohno K, Shibata Y, Paul-Prasanth B, Pitchayawasin S, et al. A relaxin-like peptide purified from radial nerves induces oocyte maturation and ovulation in the starfish, Asterina pectinifera. Proc Natl Acad Sci U S A. 2009;106:9507-12.

[31] 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.

[32] 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.

[33] 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.

[34] Resemann A, Wunderlich D, Rothbauer U, Warscheid B, Leonhardt H, Fuchser J, et al. Top-down de Novo protein sequencing of a 13.6 kDa camelid single heavy chain antibody by matrix-assisted laser desorption ionization-time-of-flight/time-of-flight mass spectrometry. Anal Chem. 2010;82:3283-92.

[35] 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.

[36] Chen ZW, Fuchs K, Sieghart W, Townsend RR, Evers AS. Deep amino acid sequencing of native brain GABAA receptors using high-resolution mass spectrometry. Mol Cell Proteomics. 2012;11:M111 011445.

[37] Fang L, Kaake RM, Patel V, Yang Y, Baldi P, Huang L. Mapping the protein interaction network of the human COP9 signalosome (CSN) complex using a label-free QTAX strategy. Mol Cell Proteomics. in press.

[38] Nassel DR, Wegener C. A comparative review of short and long neuropeptide F signaling in invertebrates: Any similarities to vertebrate neuropeptide Y signaling? Peptides. 2011;32:1335-55.

[39] Swaney DL, Wenger CD, Coon JJ. Value of using multiple proteases for large-scale mass spectrometry-based proteomics. J Proteome Res. 2010;9:1323-9.

[40] 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.

[41] Wainwright G, Webster SG, Wilkinson MC, Chung JS, Rees HH. Structure and significance of mandibular organ-inhibiting hormone in the crab, Cancer pagurus. Involvement in multihormonal regulation of growth and reproduction. J Biol Chem. 1996;271:12749-54.

[42] Tang C, Lu W, Wainwright G, Webster SG, Rees HH, Turner PC. Molecular characterization and expression of mandibular organ-inhibiting hormone, a recently discovered neuropeptide involved in the regulation of growth and reproduction in the crab Cancer pagurus. Biochem J. 1999;343:355-60.

[43] Ollivaux C, Gallois D, Amiche M, Boscameric M, Soyez D. Molecular and cellular specificity of post-translational aminoacyl isomerization in the crustacean hyperglycaemic hormone family. FEBS J. 2009;276:4790-802.

[44] Vitorino R, Alves R, Barros A, Caseiro A, Ferreira R, Lobo MC, et al. Finding new posttranslational modifications in salivary proline-rich proteins. Proteomics. 2010;10:3732-42.

[45] Bush MF, Hall Z, Giles K, Hoyes J, Robinson CV, Ruotolo BT. Collision cross sections of proteins and their complexes: a calibration framework and database for gas-phase structural biology. Anal Chem. 2010;82:9557-65.

[46] Zhang J, Xin L, Shan B, Chen W, Xie M, Yuen D, et al. PEAKS DB: De Novo sequencing assisted database search for sensitive and accurate peptide identification. Mol Cell Proteomics. In press.

[47] Bandeira N, Pham V, Pevzner P, Arnott D, Lill JR. Automated de novo protein sequencing of monoclonal antibodies. Nat Biotechnol. 2008;26:1336-8.

[48] Wu SL, Jardine I, Hancock WS, Karger BL. A new and sensitive on-line liquid chromatography/mass spectrometric approach for top-down protein analysis: the comprehensive analysis of human growth hormone in an E. coli lysate using a hybrid linear ion trap/Fourier transform ion cyclotron resonance mass spectrometer. Rapid Commun Mass Spectrom. 2004;18:2201-7.

[49] Eliuk SM, Maltby D, Panning B, Burlingame AL. High resolution electron transfer dissociation studies of unfractionated intact histones from murine embryonic stem cells using on-line capillary LC separation: determination of abundant histone isoforms and post-translational modifications. Mol Cell Proteomics. 2010;9:824-37.

[50] 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.

[51] Michalski A, Damoc E, Lange O, Denisov E, Nolting D, Mueller 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. In press.

[52] Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis. 1999;20:3551-67.

[53] Silva JC, Gorenstein MV, Li GZ, Vissers JP, Geromanos SJ. Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. Mol Cell Proteomics. 2006;5:144-56.

[54] Chang ES, Chang SA, Beltz BS, Kravitz EA. Crustacean hyperglycemic hormone in the lobster nervous system: localization and release from cells in the subesophageal ganglion and thoracic second roots. J Comp Neurol. 1999;414:50-6.

[55] Uetrecht C, Rose RJ, van Duijn E, Lorenzen K, Heck AJ. Ion mobility mass spectrometry of proteins and protein assemblies. Chem Soc Rev. 2010;39:1633-55.

Table 1. CHH-family neuropeptides identified in this study

Species	Peptide name	Calcd. MW[a]	Exptl. MW[a]	Error/ppm	PTM[b]	Organ
C. sapidus	*Cas*-SG-CHH	8472.800/8478.847	8472.948/8478.878	17.5/3.7	*p*Q,C-NH_2	SG
	Cas-SG-MIH	9065.364/9071.435	9065.391/9071.424	3.0/1.2	-	SG
	Cas-PO-CHH	8395.644/8401.691	8395.643/ -	0.1/ -	*p*Q,3Mox	PO
C. borealis	*Cab*-SG-CHH-I	8539.876/8545.923	8539.892/8545.934	1.9/1.3	*p*Q, C-NH_2	SG
	Cab-SG-CHH-II	8556.903/8562.950	8556.913/8562.958	1.2/0.9	C-NH_2	SG
	Cab-SG-MIH	8932.189/8938.235	8932.302/8938.230	12.7/0.6	C-NH_2	SG
	Cab-SG-MOIH	9252.258/9258.305	9252.346/9258.278	9.5/2.9	-	SG

[a] MW of intact peptide / MW of "reduced" peptide.
[b] *p*Q, pyro-Glu; C-NH_2, C-terminal amidation; Mox, methionine oxidation.

Table 2. Proteolytic peptides of $^{Cab\text{-}SG}$-MIH identified by bottom-up *de novo* sequencing

Exptl. MW	Calcd. MW	Sequence[a]	Proteolytic Peptide	Protease
1654.96	1654.82	RVINDDCPNLIGNR	MIH [1-14]	trypsin
1498.79	1498.71	VINDDCPNLIGNR	MIH [2-14]	trypsin
2018.20	2017.99	VINDDCPNLIGNRDLYK	MIH [2-18]	trypsin
1854.98	1854.84	KVEWICEDCSNIFR	MIH [19-32]	trypsin
1726.94	1726.74	VEWICEDCSNIFR	MIH [20-32]	trypsin
1022.59	1022.46	NTGMATLCR	MIH [33-41]	trypsin
2398.30	2398.05	KNCFFNEDFLWCVYATER	MIH [42-59]	trypsin
2270.21	2269.96	NCFFNEDFLWCVYATER	MIH [43-59]	trypsin
1769.98	1769.85	RVINDDCPNLIGNRD	MIH [1-15]	Glu-C
1366.80	1366.70	LYKKVEWICE	MIH [16-25]	Glu-C

[a]Fixed carbamidomethyl is on cysteine.

Table 3. Proteolytic peptides of $^{Cab\text{-}SG}$-MOIH identified by bottom-up *de novo* sequencing

Exptl. MW	Calcd. MW	Sequence[a]	Proteolytic Peptide	Protease
1619.83	1619.76	RINNDCQNFIGNR	MOIH [2-14]	trypsin
1463.73	1463.66	INNDCQNFIGNR	MOIH [3-14]	trypsin
1695.86	1695.79	VDWICKDCANIFR	MOIH [20-32]	trypsin
894.45	894.40	DCANIFR	MOIH [26-32]	trypsin
1964.98	1964.89	DCANIFRQDGLLNNCR	MOIH [26-41]	trypsin
1088.56	1088.50	QDGLLNNCR	MOIH [33-41]	trypsin
2540.19	2540.07	SNCFYNTEFLWCIDATENTR	MOIH [42-61]	trypsin
1684.91	1684.82	EQLEQWAAILGAGWN	MOIH [64-78]	trypsin
2270.04	2270.04	RRINNDCQNFIGNRAMYE	MOIH [1-18]	Glu-C
2286.03	2286.04	RRINNDCQNFIGNRAMoxYE	MOIH [1-18]	Glu-C
2865.31	2865.27	CANIFRKDGLLNNCRSNCFYNTE	MOIH [27-49]	Glu-C
1185.65	1185.59	QWAAILGAGWN	MOIH [68-78]	Glu-C

[a]Fixed carbamidomethyl is on cysteine; Mox, methionine oxidation.

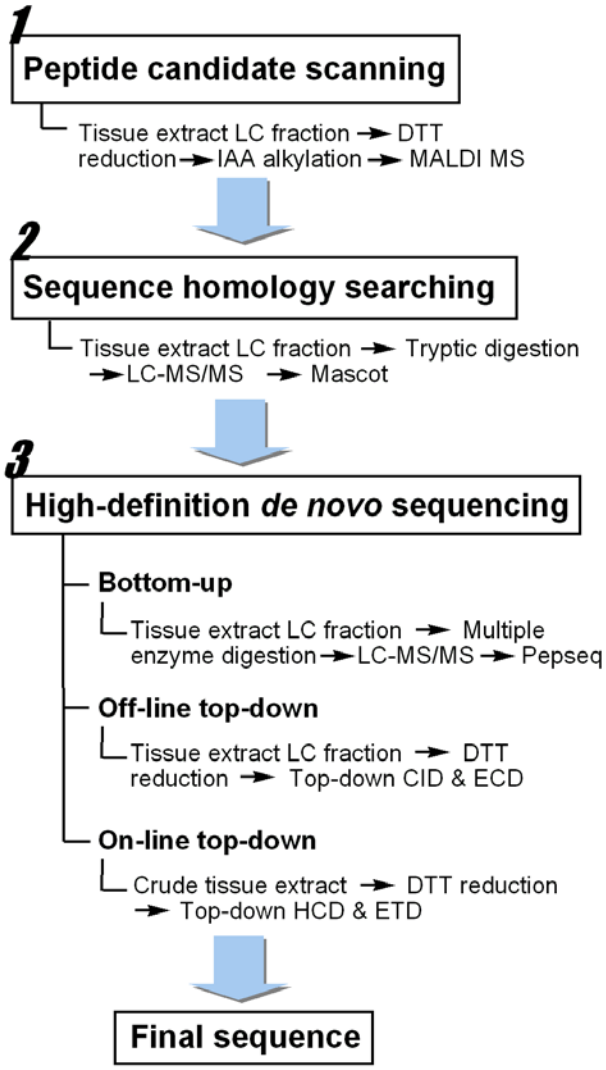


Figure 1.Workflow of the multifaceted strategy for identification and characterization of the CHH-family neuropeptides.

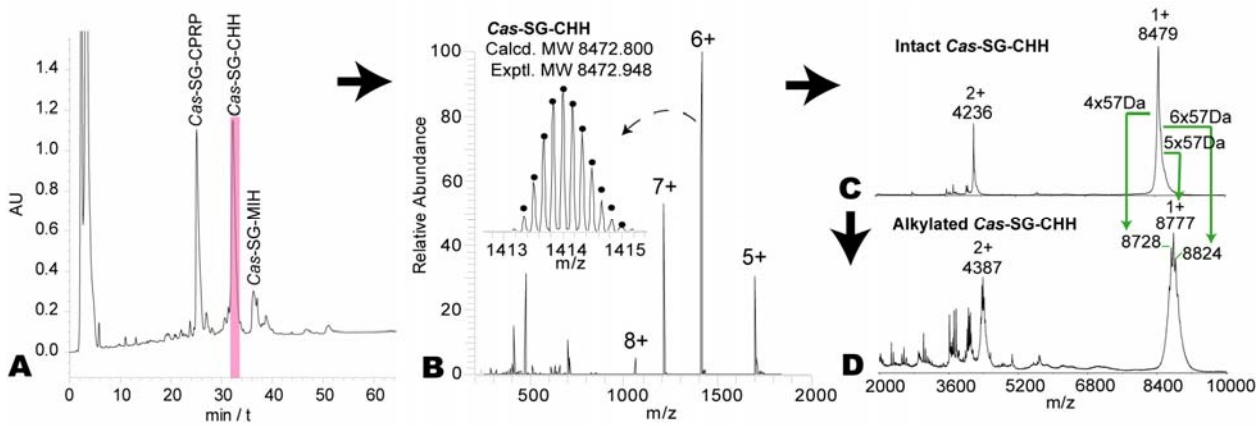


Figure 2. CHH-family neuropeptide candidate scanning in *C. sapidus* sinus gland. (A) Reversed phase HPLC fractionation of *C. sapidus* sinus gland tissue extract showing the presence of putative CPRP, CHH, and MIH peptide peaks. (B) ESI-LTQ-FTICR spectra of LC Fraction #17. (C) MALDI-TOF/TOF mass spectra of original, intact $^{Cas\text{-}SG\text{-}}$CHH and (D) DTT-reduced and IAA-treated LC Fraction #17 showing multiple alkylated CHH peptide forms.

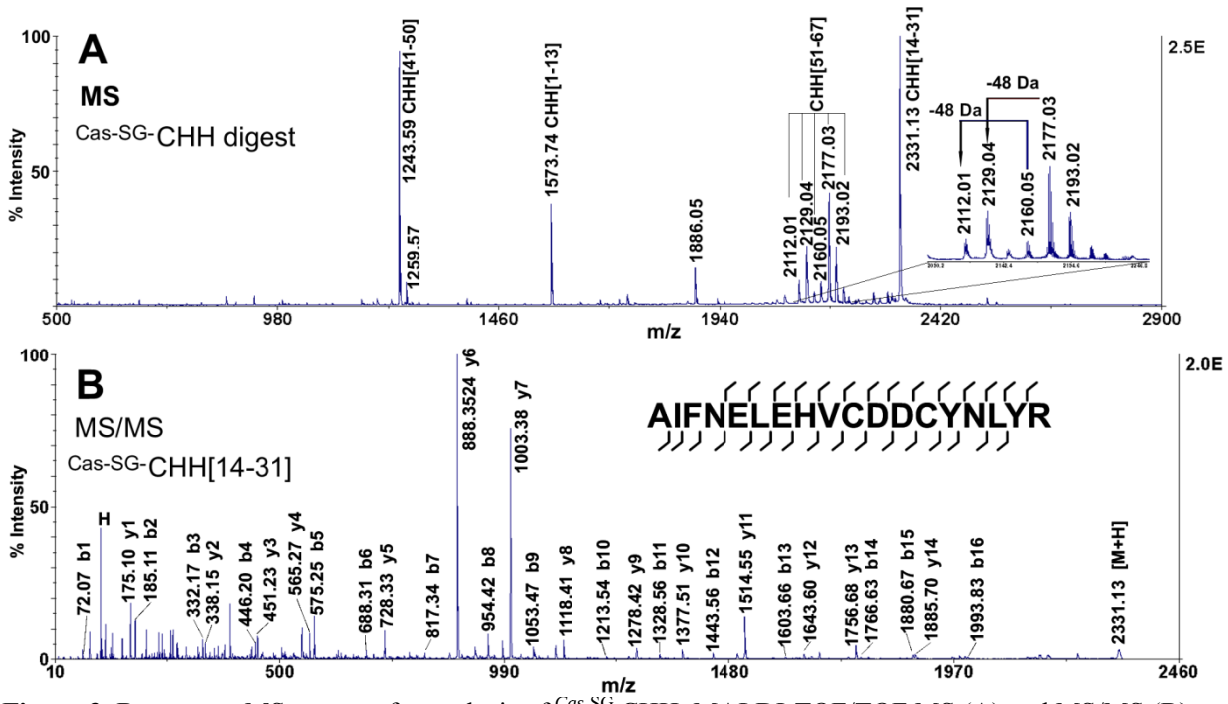


Figure 3. Bottom-up MS strategy for analysis of *Cas*-SG-CHH. MALDI-TOF/TOF MS (A) and MS/MS (B) spectra of tryptic digest of LC Fraction #17 from *C. sapidus* sinus gland tissue extract.

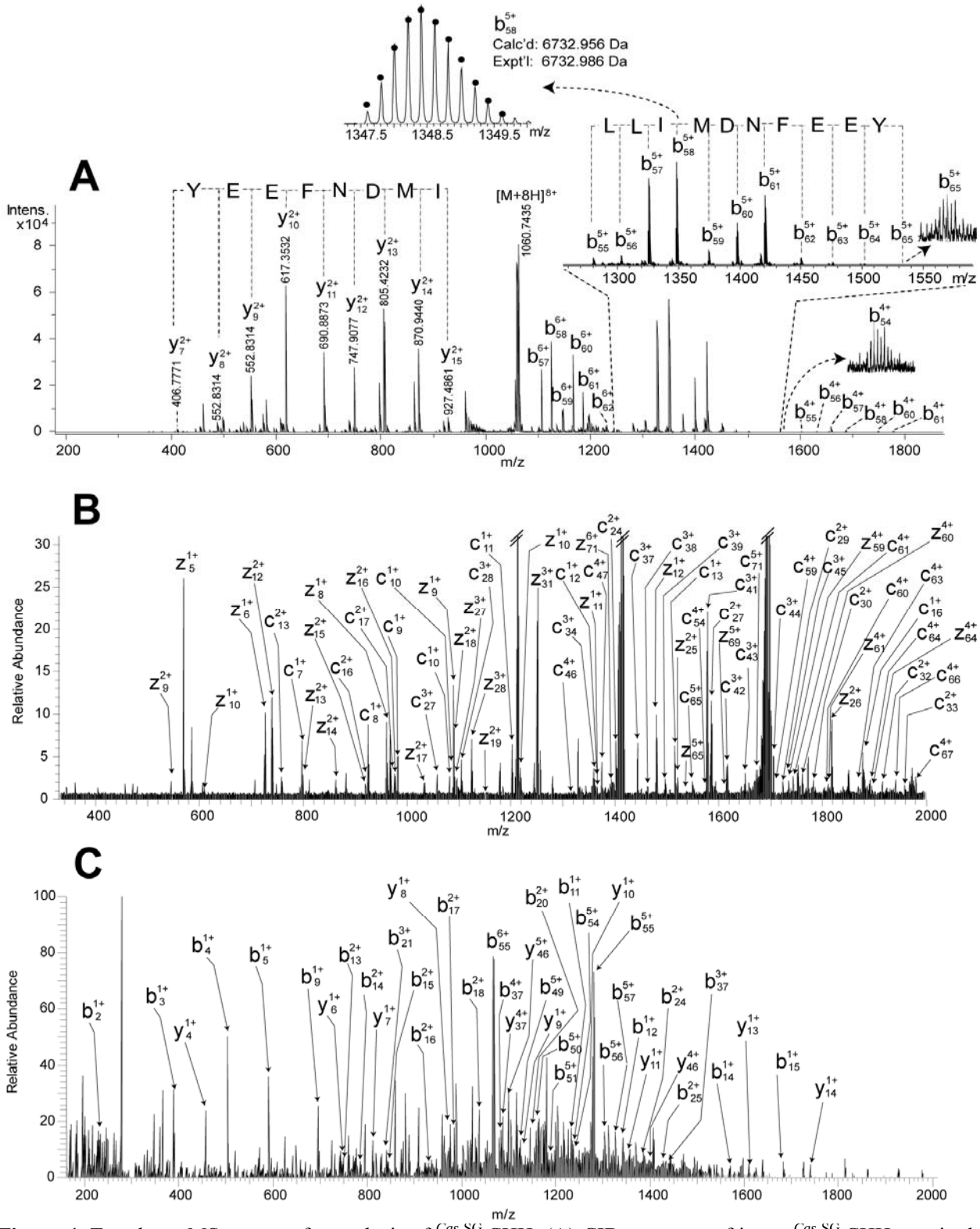


Figure 4. Top-down MS strategy for analysis of $^{Cas\text{-}SG\text{-}}$CHH. (A) CID spectrum of intact $^{Cas\text{-}SG\text{-}}$CHH acquired on ESI-UHR-QTOF maXis. (B) ECD spectrum of DTT-reduced $^{Cas\text{-}SG\text{-}}$CHH acquired on ESI-LTQ-FTICR. (C) HCD spectrum of DTT-reduced $^{Cas\text{-}SG\text{-}}$CHH acquired on ESI-LTQ-Orbitrap Elite.

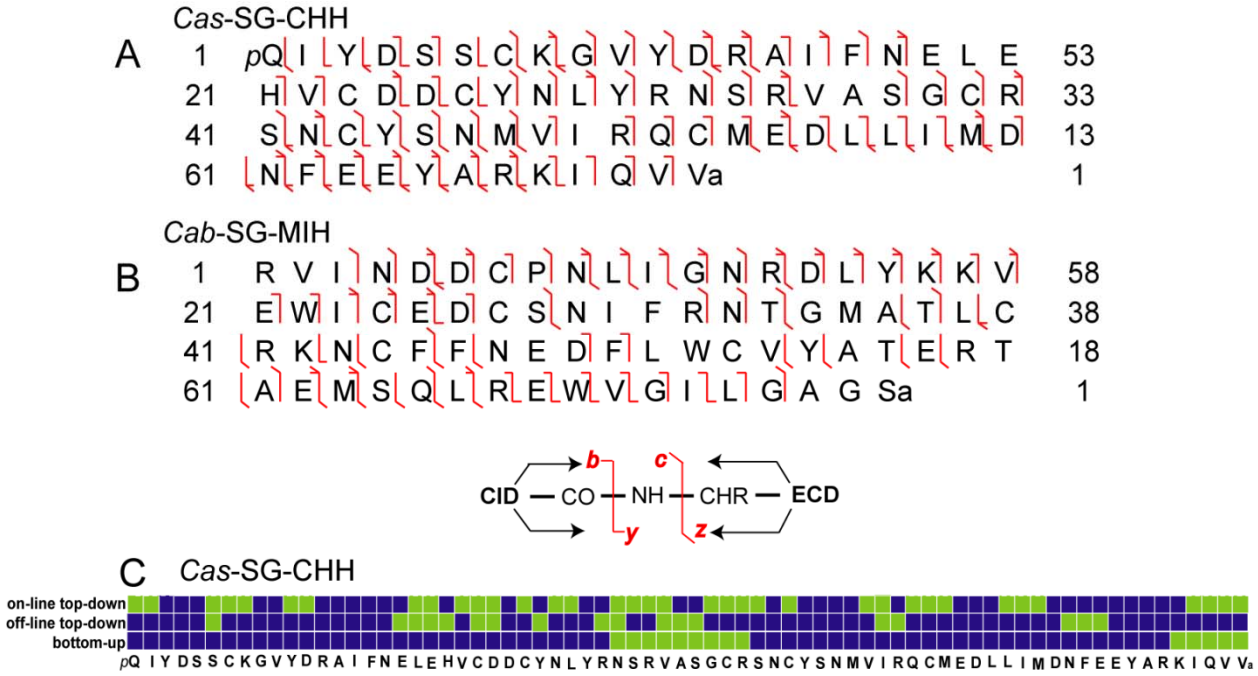


Figure 5. Peptide fragmentation maps and sequence coverage. Fragmentation maps of (A) *Cas*-SG-CHH and (B) *Cab*-SG-MIH. (C) Sequence coverage map of *Cas*-SG-CHH. Green square, unidentified; blue square, identified.

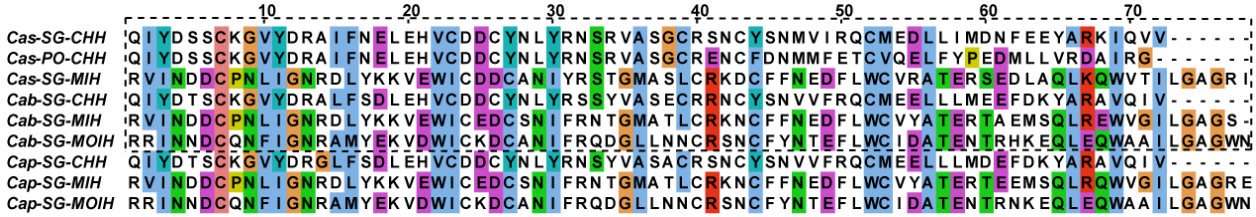


Figure 6. Sequence alignment of the CHH-family neuropeptides. *Cas, Callinectes sapidus; Cab, Cancer borealis; Cap, Cancer pagurus.*

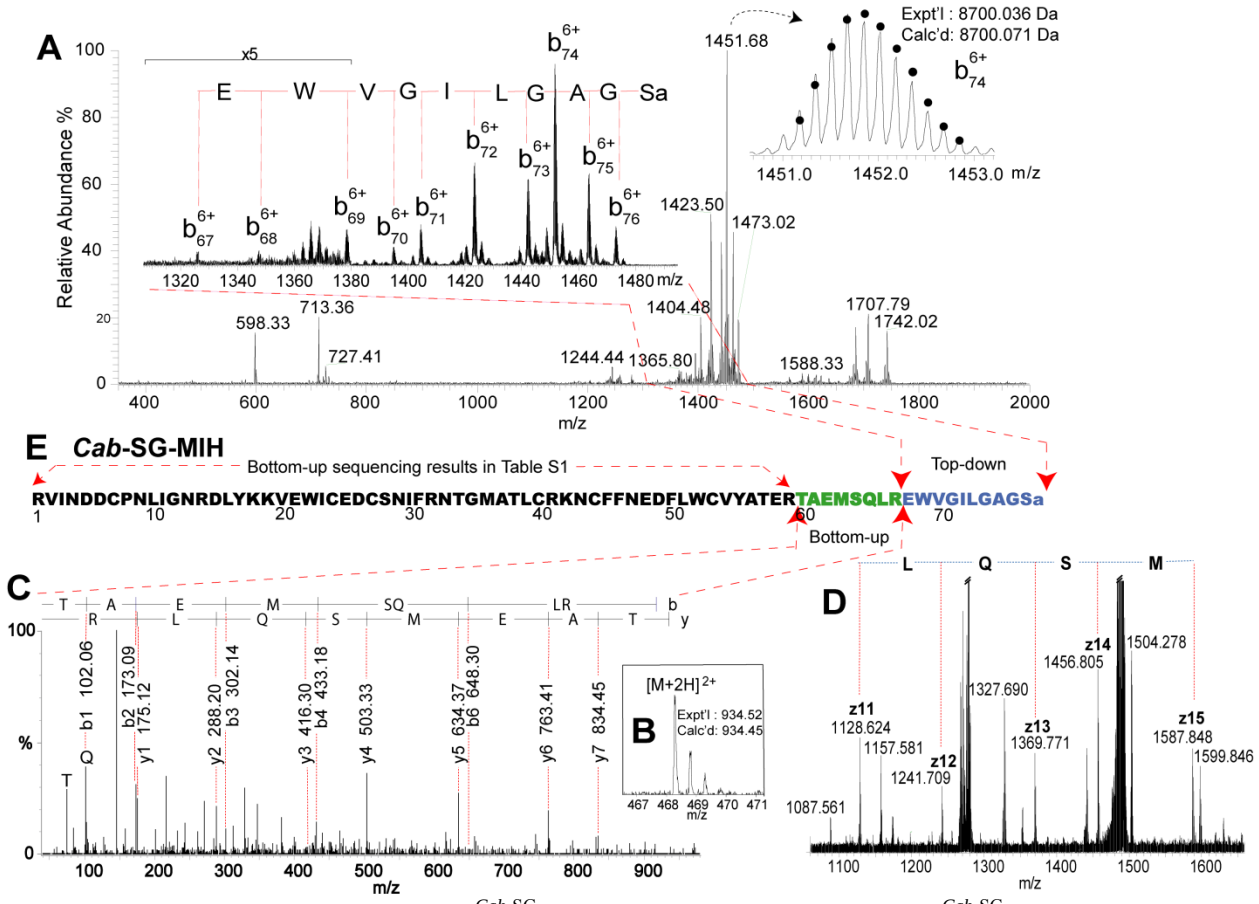


Figure 7. *De novo* sequencing of novel $^{Cab\text{-}SG\text{-}}$MIH. (A) CID spectrum of intact $^{Cab\text{-}SG\text{-}}$MIH (*m/z* 1278.055, charge state +7). (B) MS and (C) MS/MS spectrum of tryptic peptide TAEMSQLR from $^{Cab\text{-}SG\text{-}}$MIH. (D) Zoom-in ECD spectrum of DTT-reduced $^{Cab\text{-}SG\text{-}}$MIH. (E) Sequence assembling of $^{Cab\text{-}SG\text{-}}$MIH.

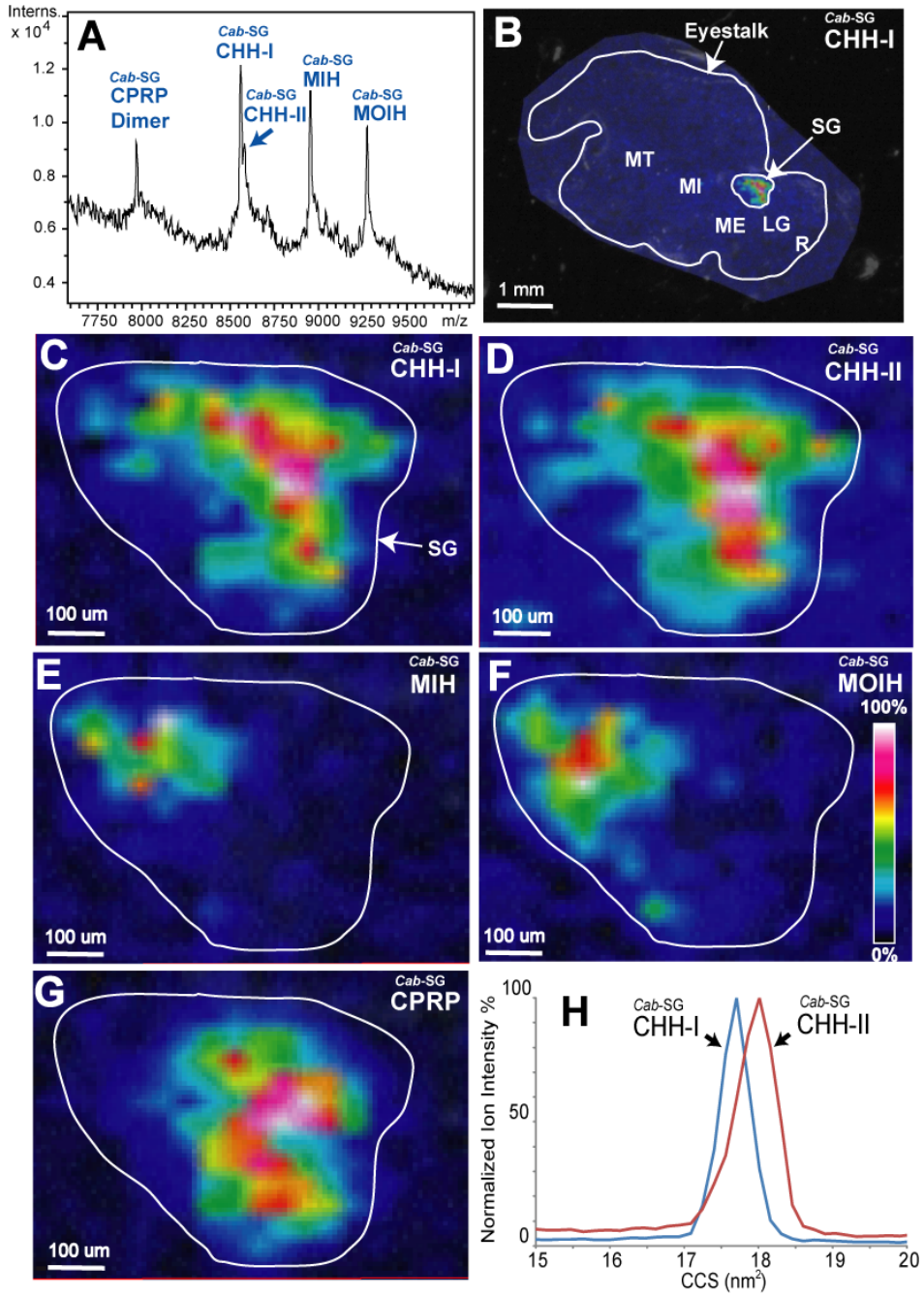


Figure 8. Spatial distribution mapping and conformational analysis of the CHH-family neuropeptides in *C. borealis* sinus gland. (A) MALDI mass spectrum of direct tissue analysis of *C. borealis* sinus gland. (B) MALDI-MSI image of $^{Cab\text{-}SG\text{-}}$CHH-I on *C. borealis* eyestalk. Zoom-in MALDI images of $^{Cab\text{-}SG\text{-}}$CHH-I (C), $^{Cab\text{-}SG\text{-}}$CHH-II (D), $^{Cab\text{-}SG\text{-}}$MIH (E), $^{Cab\text{-}SG\text{-}}$MOIH (F), and $^{Cab\text{-}SG\text{-}}$CPRP (G) on the sinus gland of *C. borealis* eyestalk. (H) Plot of normalized ion intensity *vs.* collision cross section of $^{Cab\text{-}SG\text{-}}$CHH-I and $^{Cab\text{-}SG\text{-}}$CHH-II from ion mobility MS analysis.

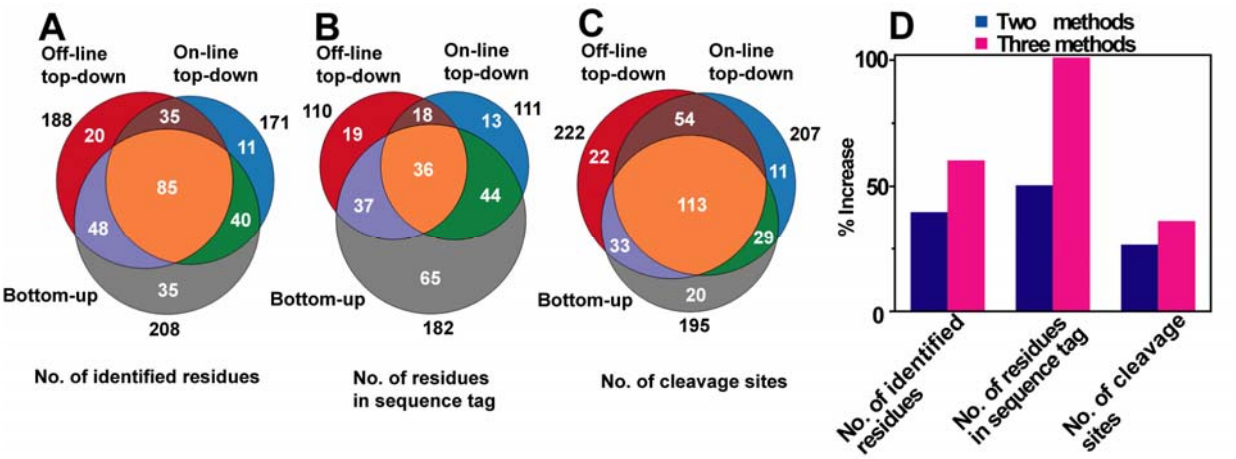


Figure 9. Statistical evaluation of bottom-up, off-line top-down and on-line top-down *de novo* sequencing strategies. Contributions of the three *de novo* sequencing methods to number of identified residues (A), number of residues in sequence tags (B), and number of cleavage sites (C). (D) Percent increase of the three numbers by combination of two and three methods *vs.* one method. One method: on-line top-down; two methods: on-line top-down and off-line top-down; three methods: on-line top-down, off-line top-down and bottom up.

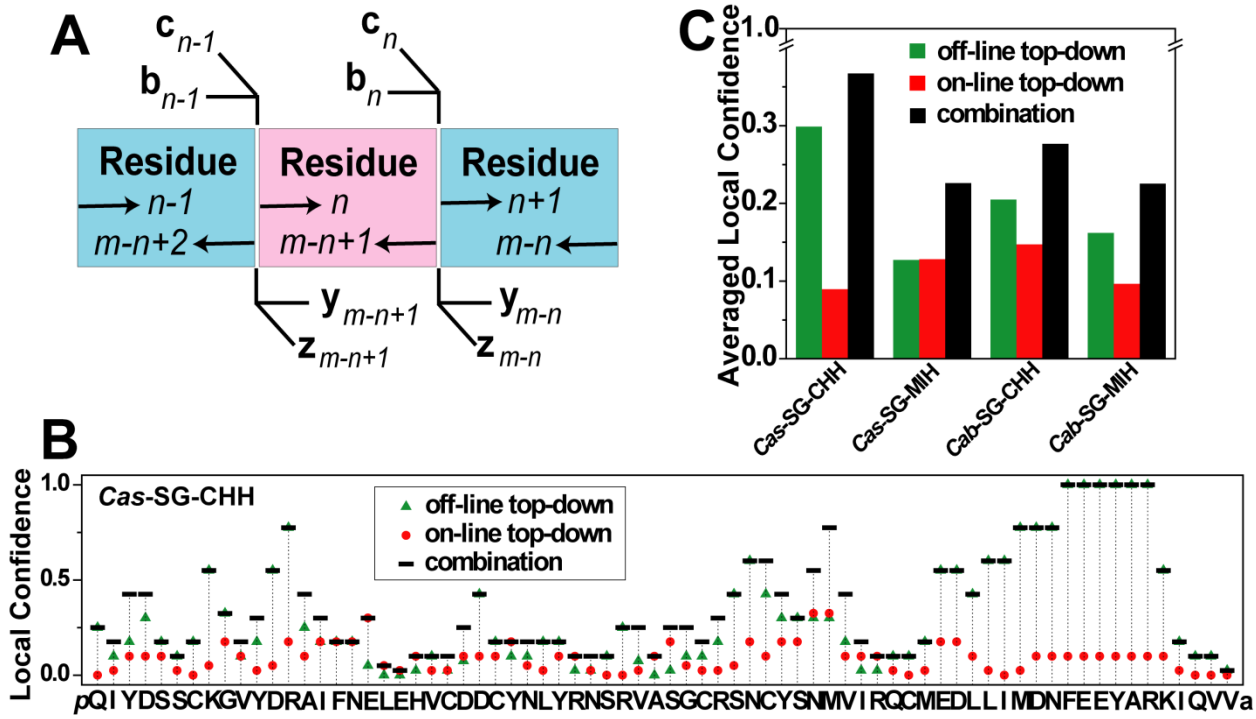


Figure 10. Statistical evaluation of off-line top-down and on-line top-down *de novo* sequencing strategies. (A) Demonstration of ion pairs (same type of two consecutive ions, e.g. b_{n-1}/b_n or y_{m-n+1}/y_{m-n}, etc) and fragment ions for the calculation of local confidence score for the n^{th} residue (counting from N-terminus) of a peptide with a length of *m* residues. (B) Distribution of local confidence scores for *Cas*-SG-CHH. (C) Averaged local confidence scores of *Cas*-SG-CHH, *Cas*-SG-MIH, *Cab*-SG-CHH and *Cab*-SG-MIH. We assume that an ideal identification of the n^{th} residue requires observation of at least eight fragment ions under proton-driven and electron-driven fragmentation, *i.e.* b_{n-1}, b_n, c_{n-1}, c_n, y_{m-n+1}, y_{m-n}, z_{m-n+1}, z_{m-n}. However, in practice, not all these fragments can be detected. Therefore, we define a local confidence score to evaluate the degree of identification for a residue. The local confidence score for a given residue (e.g. the n^{th} residue) is calculated by the following point-based method. 2 points are assigned to an ion pair and 1 point is given to a fragment ion. The local confidence score of the n^{th} residue is calculated by the equation $S_n = (X_{obsd} \cdot 2 + Y_{obsd} \cdot 1)/(X_{thel} \cdot 2 + Y_{thel} \cdot 1)$, where X_{obsd} stands for the total number of observed ion pairs, Y_{obsd} is the total number of observed fragment ions, X_{thel} is the total number of theoretical ion pairs, and Y_{thel} represents the total number of theoretical ions. The averaged local confidence score $\bar{S}$was calculated by the equation $\bar{S} = \sum_1^m S_n/m$, where *m* stands for the total residue number of a peptide.

Supplementary Materials for Chapter 2

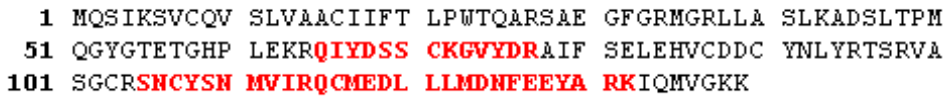


Figure S1. Homologous preprohormone of *P. trituberculatus* CHH hit by Mascot search. Matched peptides are shown in bold red.

Reference: Zhu D.F., Shen J.M., Yang J.F., Su Q. Cloning and characterization of hyperglycemic hormone from the crab *Portunus trituberculatus*. Submitted (JAN-2008) to the EMBL/GenBank/DDBJ databases.

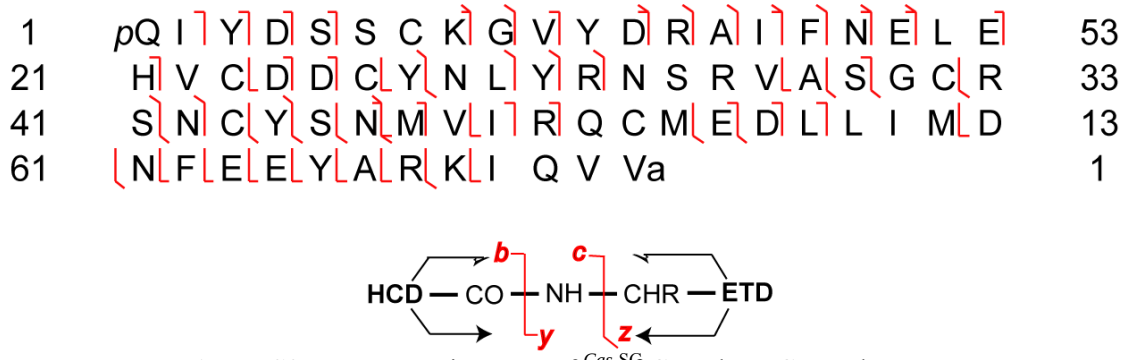


Figure S2. Fragmentation map of *Cas*-SG-CHH by HCD and ETD.

```
  1 RVINDDCPNL IGNRDLYKKV EWICEDCSNI FRNTGMATLC RKNCFFNEDF
 51 LWCVYATERT EEMSQLRQWV GILGAGRE
```

Figure S3. Homologous *Cap*-SG-MIH hit by Mascot search. Matched peptides are shown in bold red.

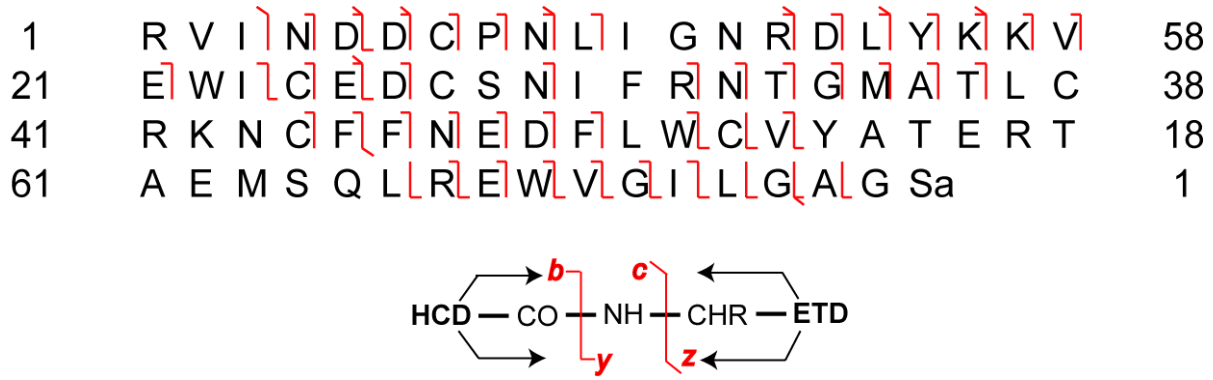


Figure S4. Fragmentation map of $^{Cab\text{-}SG}$MIH by HCD and ETD.

Figure S5. Homologous *Cap*-SG-MOIH hit by Mascot search. Matched peptides are shown in bold red.

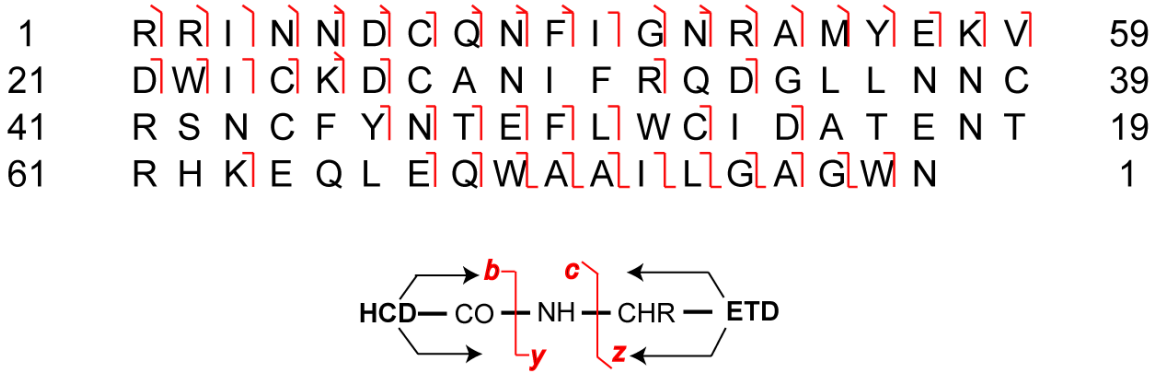


Figure S6. Fragmentation maps of *Cab*-SG-MOIH by HCD and ETD.

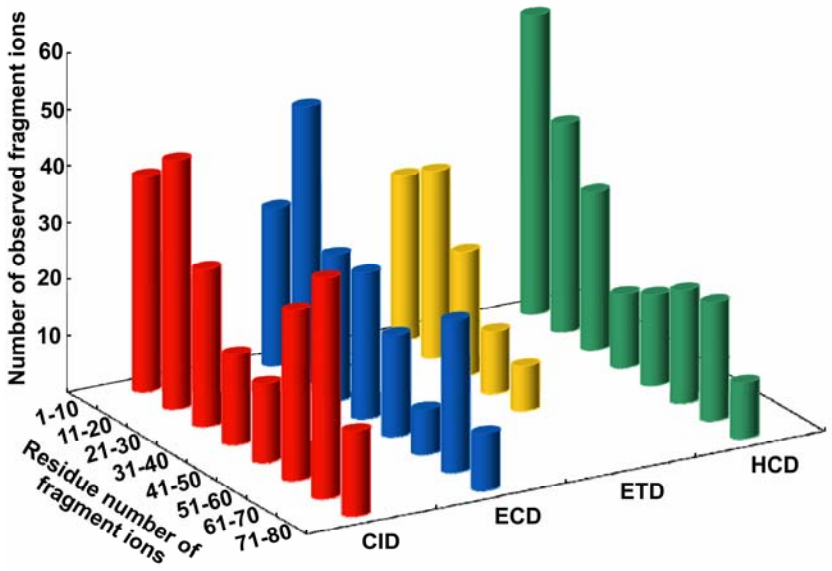


Figure S7. Correlation between length of peptide fragment ions and fragmentation techniques.

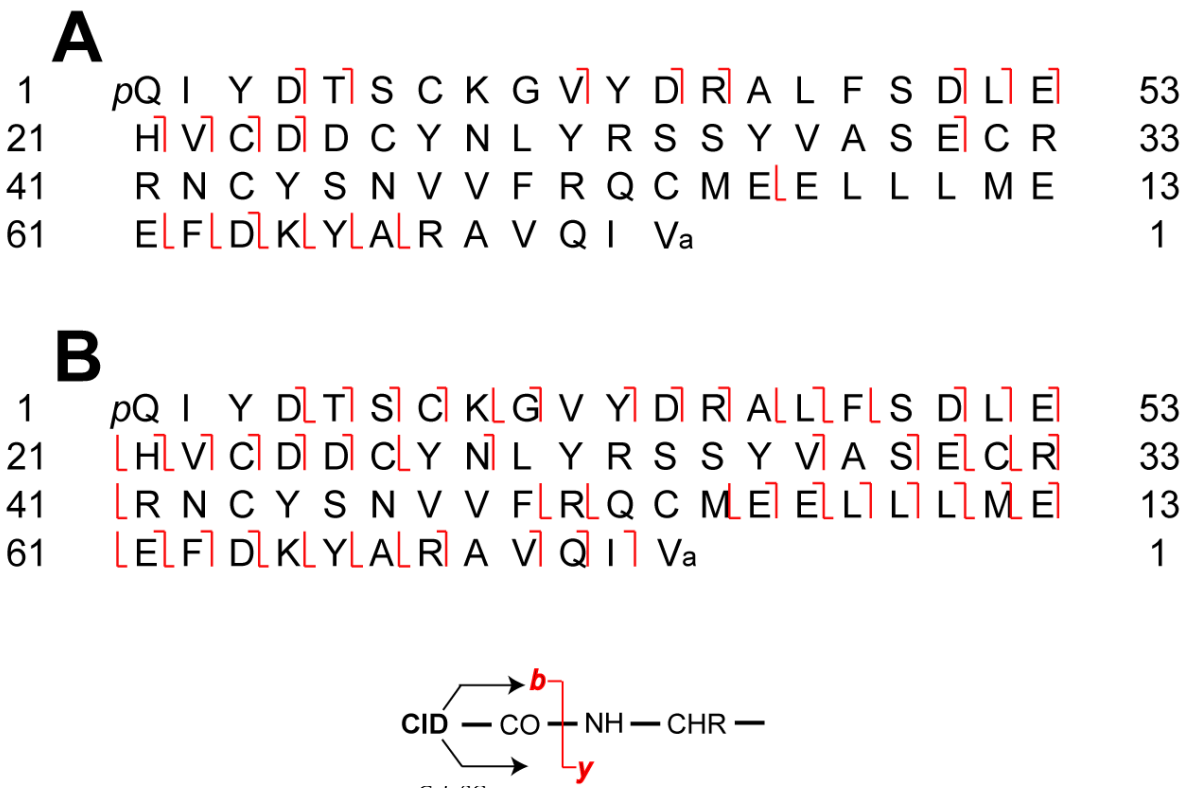


Figure S8. Fragmentation maps of *Cab*-SG-CHH. (A) Our previous study using DTT reduction followed by IAA alkylation. (B) This study using DTT reduction and acidic solvent storage. Compared with our previous study, the current strategy resulted in ~3 fold increase of observed fragment ions and identified sequence coverage.

Table S1.Tryptic peptides of *Cas*-SG-CHH identified by bottom-up *de novo* sequencing

Exptl MW	Calcd MW	Sequence[a]	Tryptic Peptide
1572.58	1572.68	*p*QIYDSSCKGVYDR	CHH[1-13]
2329.85	2330.00	AIFNELEHVCDDCYNLYR	CHH[14-31]
1242.33	1242.55	SNCYSNMVIR	CHH[41-50]
1258.57	1258.55	SNCYSNMoxVIR	CHH[41-50]
2158.75	2158.90	*p*QCMEDLLIMDNFEEYAR	CHH[51-67]
2175.68	2175.92	QCMEDLLIMDNFEEYAR	CHH[51-67]
2191.70	2191.92	QCMEDLLIMoxDNFEEYAR	CHH[51-67]

[a]Fixed carbamidomethyl is on cysteine; *p*Q, pyro-Glu; Mox, methionine oxidation.

EXPERIMENTAL PROCEDURES (Supplemental Materials)

Animals, Tissue Dissection and Extraction

C. sapidus and *C. borealis* were shipped from the Fresh Lobster Company (Gloucester, MA), and then maintained without food in artificial seawater at 10–12 °C. For tissue collection, animals were anesthetized by packing them in ice for 30–60 minutes. After anesthesia, the eyestalks and dorsolateral portions of the pericardial chamber were removed. The sinus glands (SGs) and pericardial organs (POs) were subsequently dissected from these structures, respectively, in chilled (approximately 4°C) physiological saline (composition: 440 mM NaCl; 11 mM KCl; 13 mM $CaCl_2$; 26 mM $MgCl_2$; 10 mM HEPES acid, pH 7.4, adjusted with NaOH). After dissection, tissue was either pinned in a Sylgard 184 (KR Anderson, Santa Clara, CA)-lined Petri dish and subsequently processed for anatomical studies or immediately placed in acidified methanol (90% methanol:9% glacial acetic acid:1% water) and stored at −80°C.

The tissue was then homogenized, after which the extraction liquid was transferred to a 1.5 ml microcentrifuge tube (Fisher Scientific) and centrifuged at 16,100 rcf for 10 min in an Eppendorf 5415 D microcentrifuge (Brinkmann Instruments). The supernatant fraction was retrieved, and the pellet was re-extracted with acidified methanol and re-centrifuged. Supernatant fractions were combined and concentrated to dryness using a Savant SC 110 SpeedVac concentrator (Thermo Electron). Finally, 100 µl of Millipore water was added to the extract. This resuspended extract was then vortexed and centrifuged. After centrifugation, the clear solution at the upper layer of the extract was used for further analysis.

Multiple Protease Digestion

Three μL of peptide fraction was reduced and alkylated by incubation in 2.5 mM 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 50mM ammonium bicarbonate with 0.5 μg of trypsin (Promega, Madison, WI). A 1 μL of peptide was digested overnight at 37 °C after addition of 50 mM Tris-HCl, pH 8.0, and 0.5 μg of Lys-C (Princeton Separations, Adelphia, NJ). One μL of peptide was digested in 2 M urea, 50 mM Tris, 25 mM ammonium bicarbonate at room temperature overnight with 0.5 μg of Glu-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 lyophilized.

Off-line Top-down MS on ESI-UHR-QTOF maXis

Peptide sample was directly infused on a hybrid UHR-QTOF maXis (Bruker Daltonics, Bremen, Germany) by a 25μL Hamilton syringe using infusion pump (KD Scientific, MA, USA). Further settings: flow rate 1.0 μL/min, nebulizer 1.6 psi, dry gas rate 4 L/min, dry gas temperature 200 °C, funnel rf 400 Vpp, multipole rf 400 Vpp, ion energy 5 V, low mass cutoff 300, collision energy 10 V, collision rf 1200 Vpp, ion cooler rf 400 Vpp, transfer time 120 μs, prepulse storage time 10 μs. For top-down MS/MS, the parent ion was selected with a 10 Da isolation window and fragmented by 40% of normalized collision energy. Spectra were acquired in profile mode with a scan speed of 1.0 Hz. DataAnalysis software (Bruker Daltonics) was used for data processing.

Calibration of Collision Cross Section on Ion Mobility MS

The calibration of the SYNAPT G2 travelling wave (TW) IM mass spectrometer to measure gas-phase CCS mainly followed the protocols outlined by Smith et al.[i] The CCS (Ω) of an ion

can be calculated as a function of charge (z), ion mass (m_I), drift gas mass (m_N), temperature (T), pressure (P), drift gas number density (N), and drift time (t_D):

$$\Omega = \left[\frac{18\pi}{K_b T}\left(\frac{1}{m_I} + \frac{1}{m_N}\right)\right]^{1/2} \frac{ze}{16}\frac{760}{P}\frac{T}{273.2}\frac{1}{N} A {t_D}^B \qquad (1)$$

The constants in Equation 1 include the Boltzmann constant (K_b), elementary charge (e), a correction factor for the pulsed electric field parameters (A), and an exponent to compensate for the non-linear Ω-t_D relationship in a TW-IMS instrument (B). To simplify the calculation, equation 1 can be divided by ze and reduced mass to yield a mass and charge independent CCS (Ω'). Then, all constants are combined with A to yield a new constant (A') and produce a new equation:

$$\Omega' = A' {t_D}^B \quad (2)$$

A set of standards were analyzed with known CCS values[ii] to obtain values for A' and B. The Ω' of an ion can be determined by plugging the drift time into Equation 2, and Ω can be calculated as the product of Ω', the absolute charge of ion, and the square-root of the inverse reduced mass.

Table S2. Ions used for CCS calibration.

Standard	z	mass (Da)	Ω (nm^2)
Ubiquitin	7	8561	19.1
Ubiquitin	8	8561	19.9
Ubiquitin	9	8561	20.9
Ubiquitin	10	8561	22.0
Ubiquitin	11	8561	23.4
Ubiquitin	12	8561	24.8
Ubiquitin	13	8561	26.0
Myoglobin	15	16951	40.6
Myoglobin	16	16951	41.8
Myoglobin	17	16951	43.1
Myoglobin	18	16951	44.4
Myoglobin	19	16951	45.7
Myoglobin	20	16951	47.0
Myoglobin	21	16951	48.2
Myoglobin	22	16951	49.2
Myoglobin	23	16951	50.1
Myoglobin	24	16951	50.9
Cytochrome C	13	12327	30.8
Cytochrome C	14	12327	32.0
Cytochrome C	15	12327	33.3
Cytochrome C	16	12327	34.5
Cytochrome C	17	12327	36.0
GRGDS	1	490.21	2.06
GRGDS	2	490.21	2.15
Angiotensin Fragment	2	898.47	3.32
RASG-1	2	1000.49	3.32
Angiotensin II	2	1045.53	3.74
Bradykinin	2	1059.56	3.67
Angiotensin I	3	1295.68	3.12
Renin Substrate	3	1757.93	3.88
Enolase T35	3	1871.96	4.16

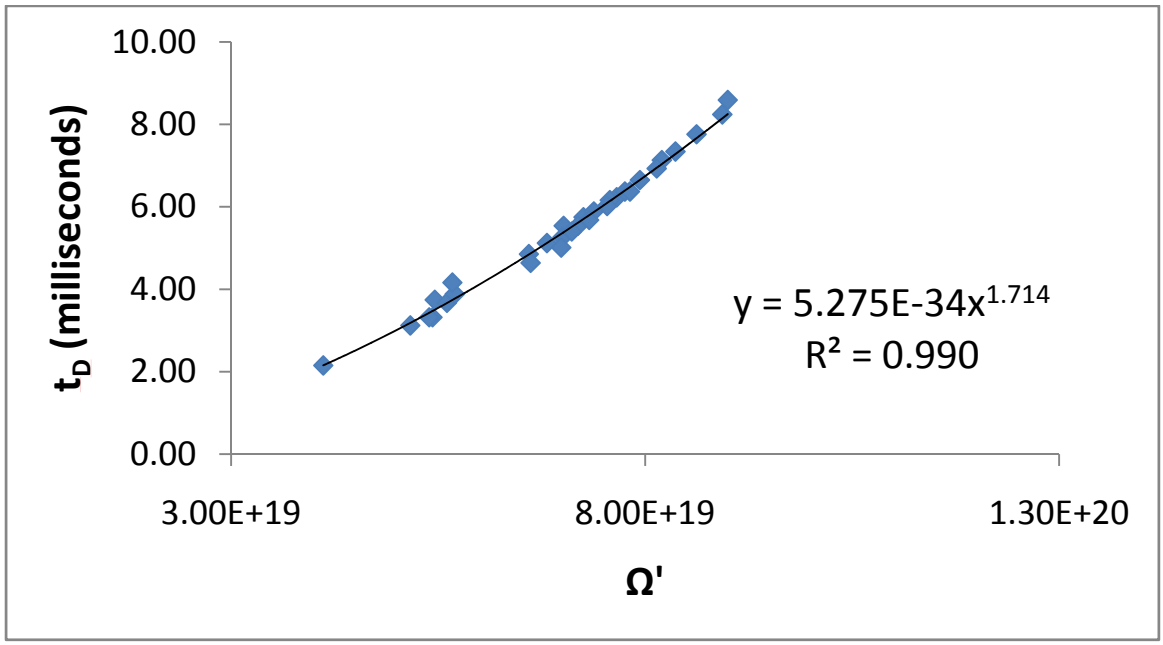


Figure S9. Calibration curve of Ω' *vs.* t_D.

[i] Smith, D. P.; Knapman, T. W.; Campuzano, I.; Malham, R. W.; Berryman, J. T.; Radford, S. E.; Ashcroft, A. E. Deciphering drift time measurements from travelling wave ion mobility spectrometry-mass spectrometry studies. *Eur. J. Mass Spectrom.***2009**, *15*, 113-130.

[ii] Bush, M.F.; Hall, Z.; Giles, K.; Hoyes, J.; Robinson, C. V.; Ruotolo, B. T. Collision cross sections of proteins and their complexes: a calibration framework and database for gas-phase structural biology. *Anal. Chem.***2010**, *82*, 9557-9565.

Chapter 3

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

Adapted from: Chenxi Jia, Christopher B. Lietz, Hui Ye, Limei Hui, Qing Yu, Sujin Yoo, and Lingjun Li, A Multi-scale Strategy for Discovery of Novel Endogenous Neuropeptides in the Crustacean Nervous System. *Journal of Proteomics*, 2013, 91:1-12.

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 sillico* homology search. Collectively, the ability to simultaneously address 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.

3.1 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 (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.

3.2 Materials and Methods

3.2.1 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).

3.2.2 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.

3.2.3 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.

3.2.4 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.

3.2.5 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]

3.2.6 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 X 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.

3.2.7 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.

3.2.8 *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.

3.3 Results and Discussion

3.3.1 Attempts for Rational Improvement of Methodology

In the past decade, there has been a great success with mass spectrometry (MS)-based neuropeptidomics [1, 5-12, 15, 33]. However, the majority of these published peptide lists show excellent coverage rates on small neuropeptides (< 3 kDa), but few identifications in the higher mass range because the traditional strategies are inefficient for *de novo* sequencing of mid-size (3-5 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 [34] exhibits 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 [34]. 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, 35]. 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 orcokinin. 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 [36]. 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.

3.3.2 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 the B-type AST-family peptides [37] 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, 38-40] 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 2x28 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 [41, 42]. 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 [42]. 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.

3.3.3 Mid-size Neuropeptides.

CPRP is produced concurrent with the cleavage of CHH from its preprohormone [30, 31, 43]. 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. [44] reported that characterization of hisotone also encountered the same issue, as the target proteins contain multiple basic amino acid residues. Therefore, a strategy in combination with top-down and bottom-up was employed. In our study, to sequence this mid-size CPRP peptide, we introduced a modified bottom-up *de novo* sequencing method. The $^{Pai-}$CPRP candidate peptide was treated with formaldehyde [38], which leads to 2x28 Da of mass increase due to dimethylation on N-terminus and lysine side chain (Figure 4A and B), indicating that this $^{Pai-}$CPRP contains 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 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).

3.3.4 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 [43]. They regulate carbohydrate metabolism and also exert an inhibitory effect on molting, reproduction, and on osmoregulatory functions [36]. 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 match 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* [45] 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 $^{Jal\text{-}}$CHH. We tried Spider Homology Search in PEAKS [46], where the bottom-up MS/MS data was searched against a home-built database containing $^{Jal\text{-}}$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 $^{Jal\text{-}}$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 $^{Pai\text{-}}$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 $^{Pai\text{-}}$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.

3.4 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.

3.5 Acknowledgements

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).

3.6 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.

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

[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.

[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.

[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.

[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.

[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.

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

[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.

[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.

[12] 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.

[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.

[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.

[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.

[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.

[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.

[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.

[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.

[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.

[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.

[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.

[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.

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

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

[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.

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

[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.

[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.

[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.

[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.

[33] 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.

[34] 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.

[35] 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.

[36] 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.

[37] 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.

[38] 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.

[39] 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.

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

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

[42] Stay B, Tobe SS. The role of allatostatins in juvenile hormone synthesis in insects and crustaceans. Annu Rev Entomol. 2007;52:277-99.

[43] Chung JS, Zmora N, Katayama H, Tsutsui N. Crustacean hyperglycemic hormone (CHH) neuropeptidesfamily: Functions, titer, and binding to target tissues. Gen Comp Endocrinol. 2009;166:447-54.

[44] 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.

[45] 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.

[46] 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.

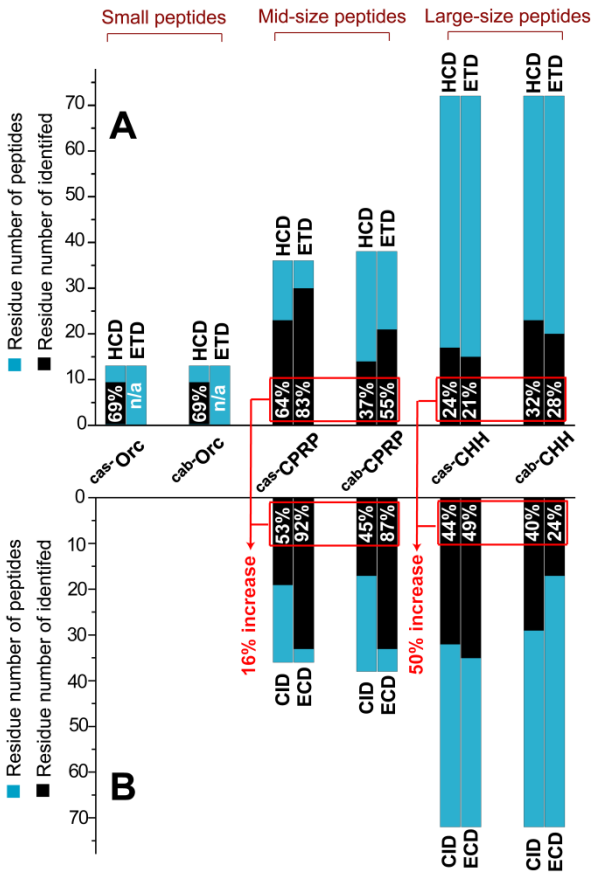


Figure 1. Evaluation results for sequencing of various sizes of neuropeptides. (A) On-line LC-MS/MS; (B) off-line top-down MS/MS. The percentage of peptide sequence coverage is labeled in the black bar of every peptide. The percentage of increase in red is calculated by ∑identified residue number/∑peptide residue number. The MS/MS fragmentation maps of these model peptides are shown in Figure S-1 and S-2.

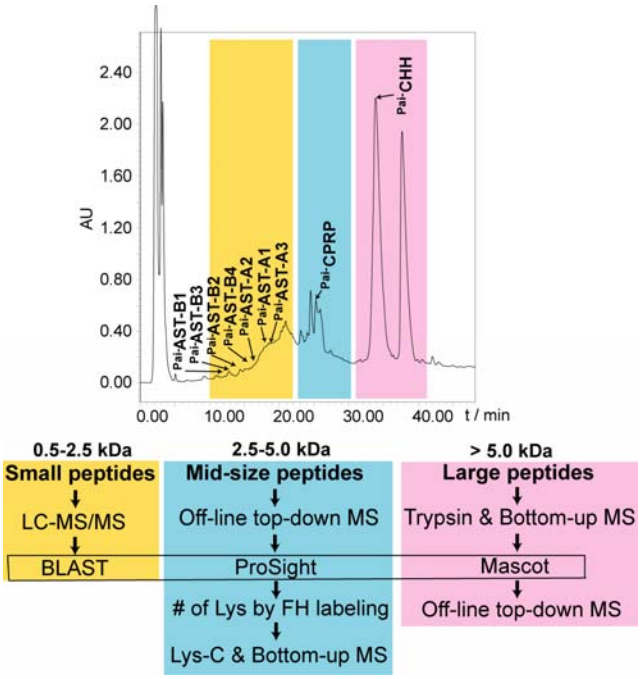


Figure 2. Workflow of the multi-scale strategy. The top panel shows reversed phase HPLC fractionation of *P. interruptus* sinus gland tissue extract showing the presence of various sizes of novel neuropeptides. The bottom panel outlines each specific strategy for discovery and identification of small, mid-size and large peptides. It should be noted that the retention behavior of peptides in reverse phase HPLC is not only based on molecular size. Here, we show three different color regions in the LC profile for the purpose of illustration.

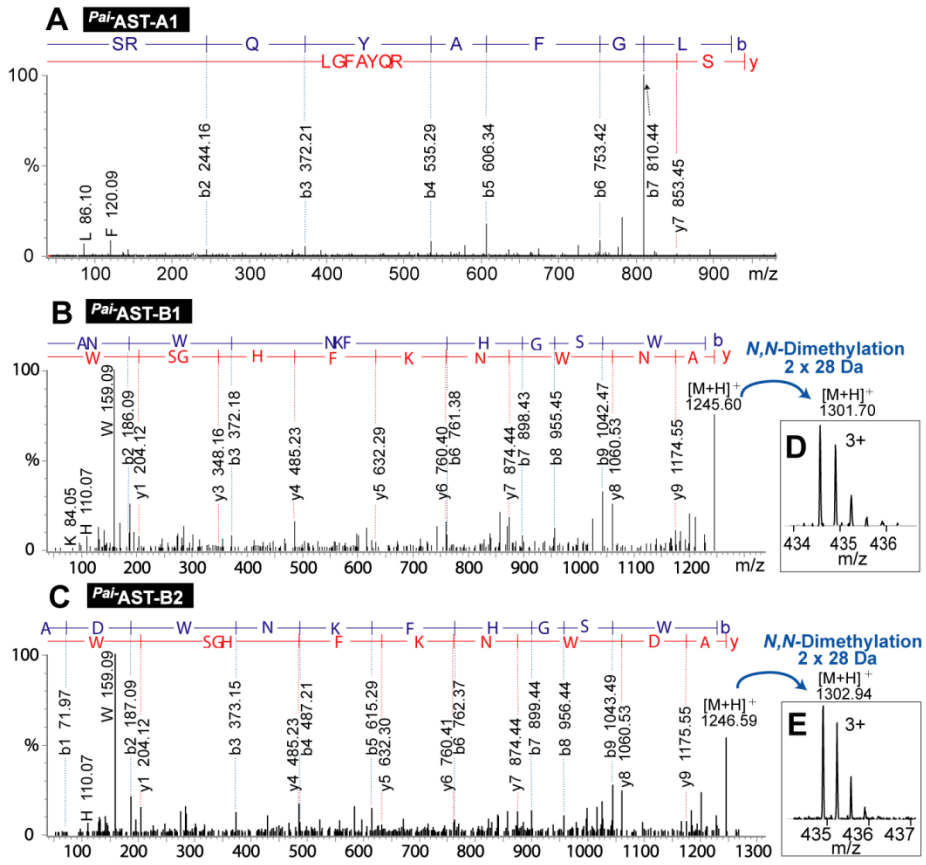


Figure 3. *De novo* sequencing of AST-family neuropeptides. MS/MS spectra of $^{Pai-}$AST-A1 (A), $^{Pai-}$AST-B1 (B), and $^{Pai-}$AST-B2 (C). MS spectra of dimethylated $^{Pai-}$AST-B1 (D), and $^{Pai-}$AST-B2 (E). Panels A, B and C are deconvoluted spectra exported from Pepseq, and Panels D and E are original zoom-in mass spectra.

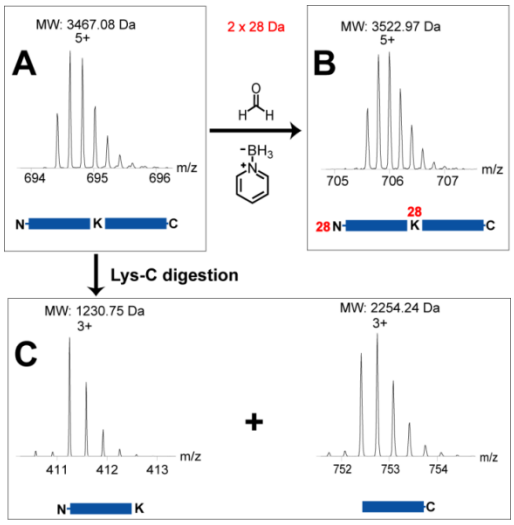


Figure 4. Strategy for *de novo* sequencing of a mid-sized CPRP neuropeptide. Zoom-in MS spectra of (A) intact $^{Pai-}$CPRP, (B) dimethylated $^{Pai-}$CPRP, and (C) proteolytic peptides of $^{Pai-}$CPRP after Lys-C digestion. The 28 Da of mass increase of peptides arises from *N*,*N*-dimethylation.

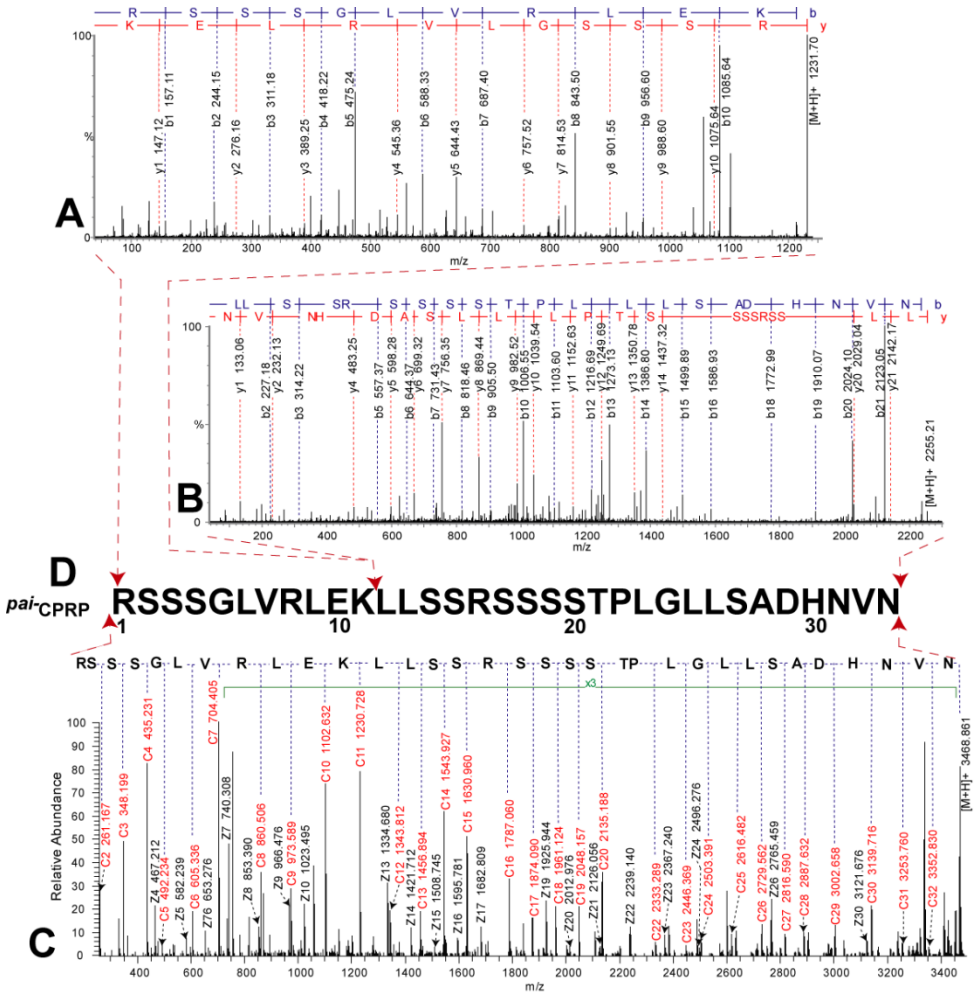


Figure 5. *De novo* sequencing MS/MS spectra of ^{Pai}CPRP. Bottom-up MS/MS spectra of (A) ^{Pai}CPRP [1-11] and (B) ^{Pai}CPRP [12-33]. (C) Top-down ECD spectra of intact ^{Pai}CPRP exported from Xtract after charge deconvolution. (D) Sequence assembly of ^{Pai}CPRP.

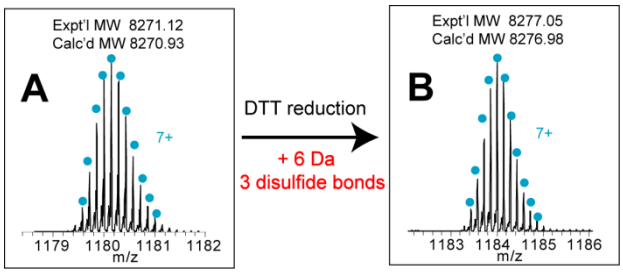


Figure 6. Determination of disulfide bond numbers in $^{Pai-}$CHH. High resolution zoom-in spectra of (A) intact $^{Pai-}$CHH and (B) DTT-reduced $^{Pai-}$CHH.

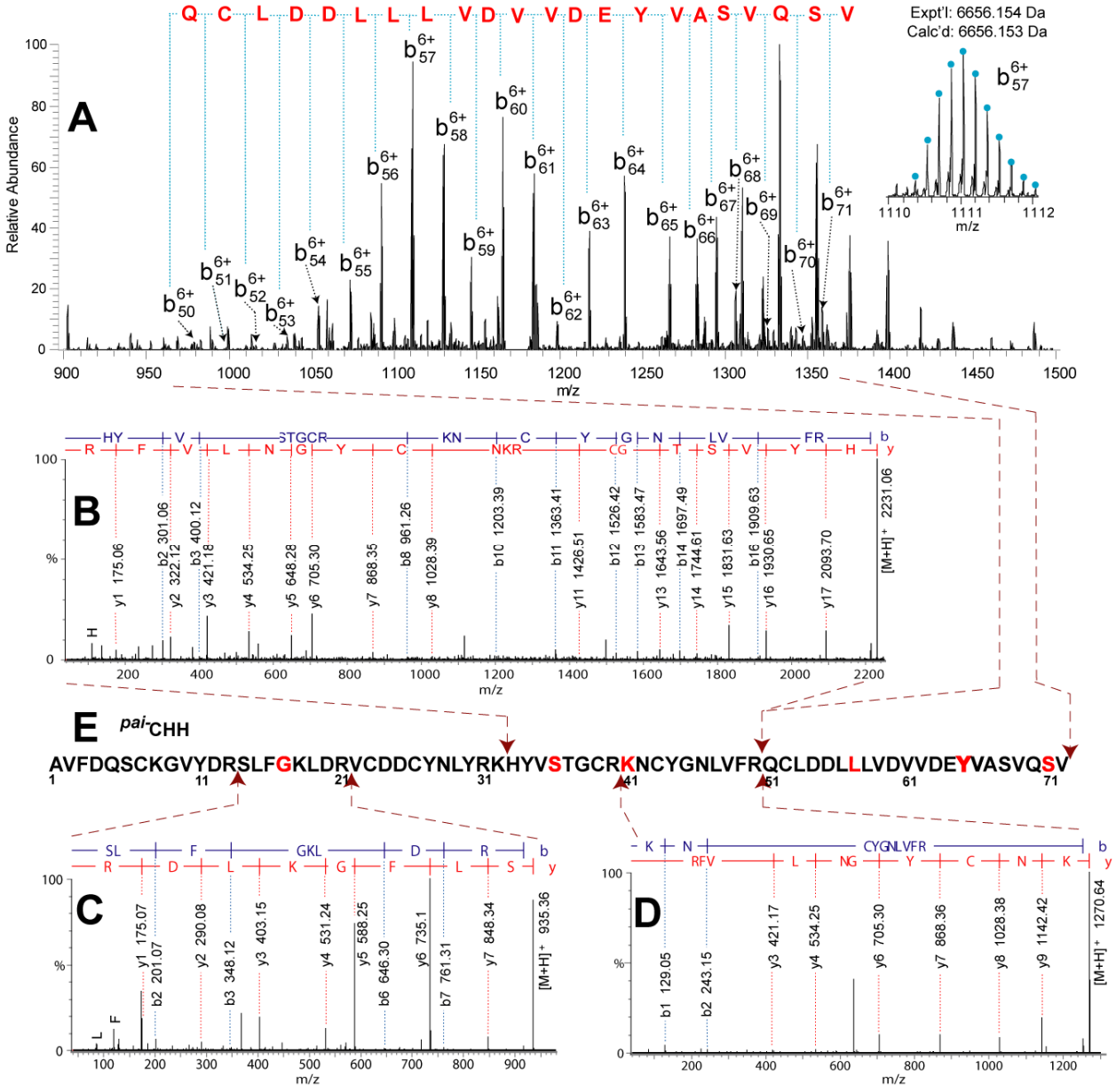


Figure 7. *De novo* sequencing MS/MS spectra of ^{Pai}CHH. (A) Top-down CID spectra of DTT-reduced ^{Pai}CHH. Bottom-up MS/MS spectra of (B) ^{Pai}CHH[33-50], (C) ^{Pai}CHH[14-21] and (D) ^{Pai}CHH[41-50]. (E) Sequence assembly of ^{Pai}CHH. The dipeptide segments ^{38}GC^{39}, ^{45}GN^{46}, and ^{49}FR^{50} were assigned according to the homologous sequence ^{Jal}CHH.

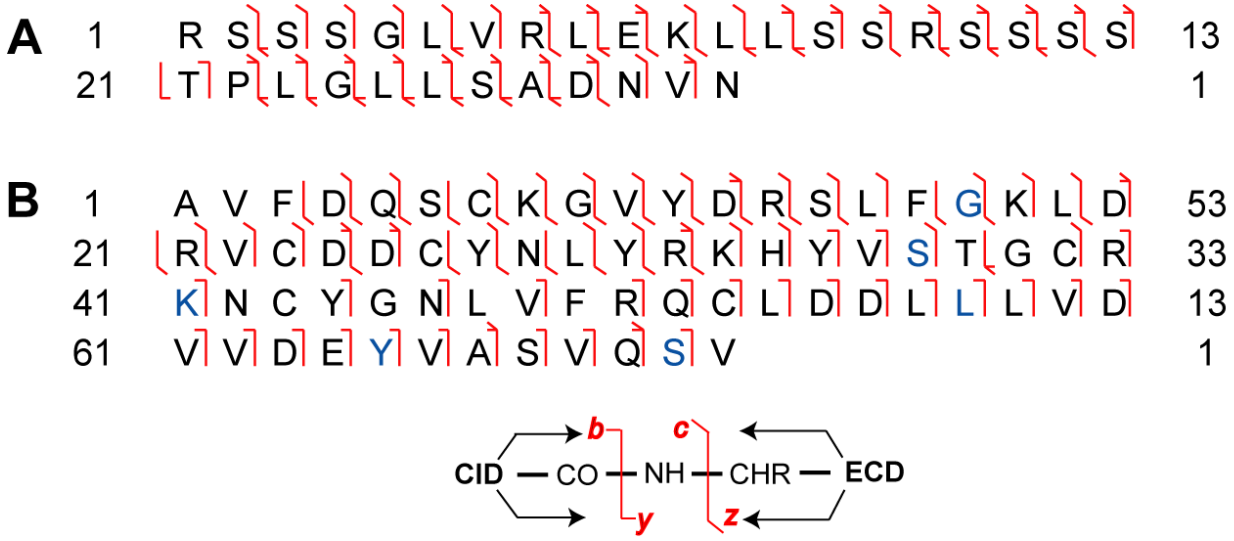


Figure 8. Fragmentation maps of ^{Pai}-CPRP (A) and ^{Pai}-CHH (B) from off-line top-down MS/MS. In Panel B, the residues different from the reference sequence ^{Jal}-CHH are highlighted in blue.

Table 1. Novel neuropeptides identified in this study.

Peptide	Sequence	Exptl MW [a]	Mass error (Da) [a]	Spectrum
Pai-AST-A1	SRQYAFGLamide	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	RSSSGLVRLEKLLSSRSSSSTPLGLLS ADHNVN	3466.86	0.00	Fig. 5
Pai-CHH	AVFDQSCKGVYDRSLFGKLDRVCD DCYNLYRKHYVSTGCRKNCYGNLV FRQCLDDLLLVDVVDEYVASVQSV [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.

Supplementary Materials for Chapter 3

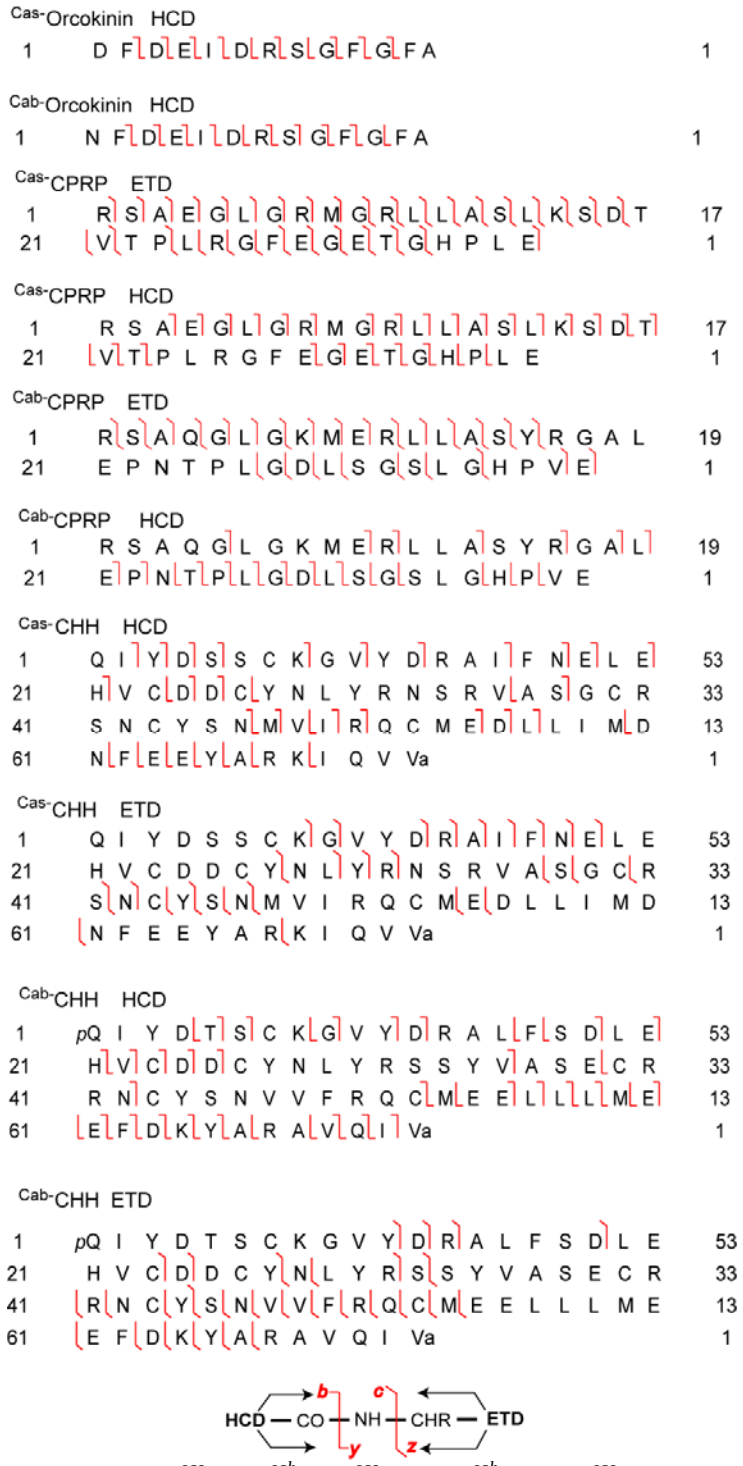


Figure S-1. Fragmentation maps of ^{cas}Orc, ^{cab}Orc, ^{cas}CPRP, ^{cab}CPRP, ^{cas}CHH, and ^{cab}CHH resulting from LC-MS/MS on LTQ-Orbritrap with HCD and ETD.

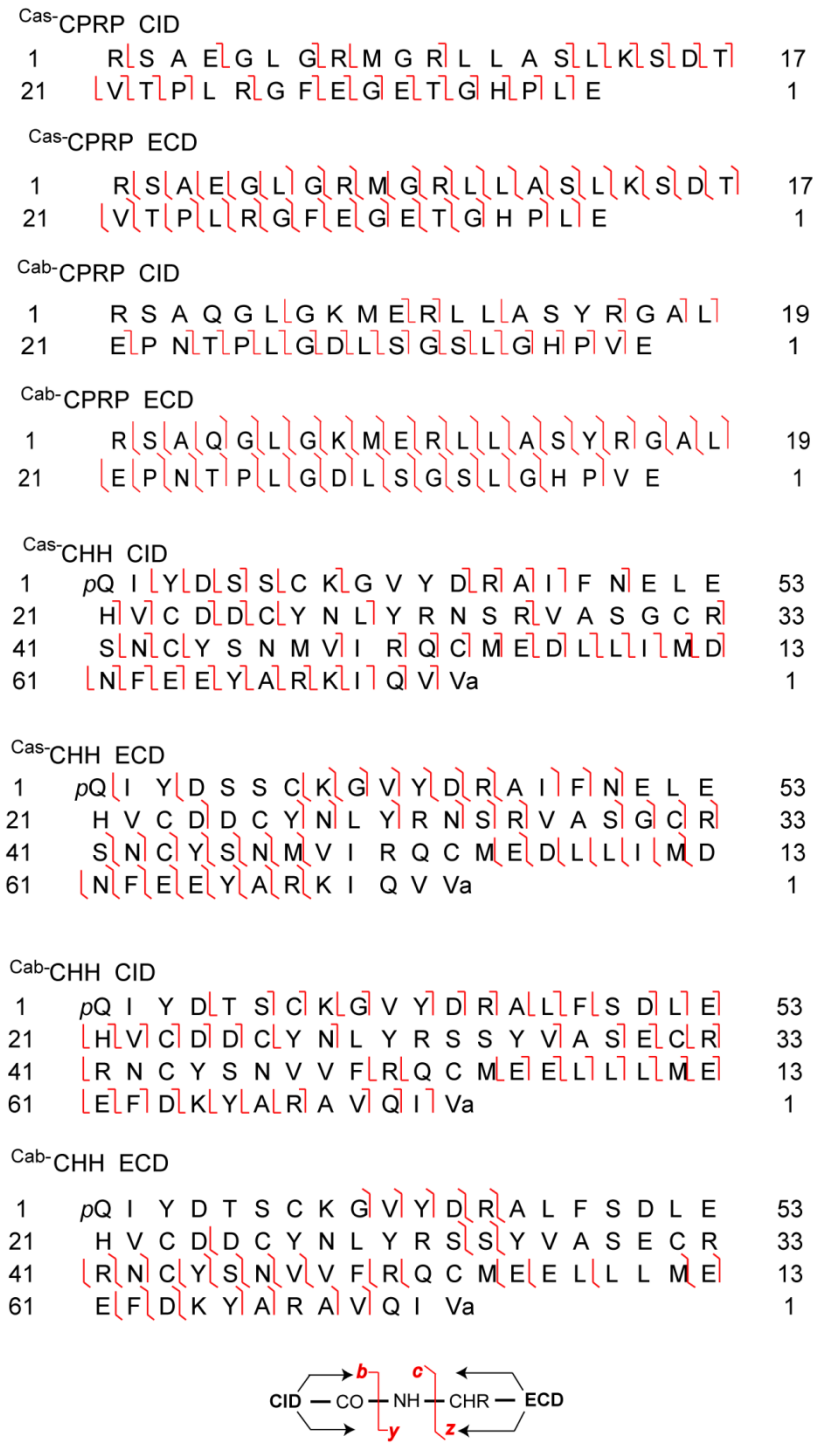


Figure S-2. Fragmentation maps of *cas*-CPRP, *cab*-CPRP, *cas*-CHH, and *cab*-CHH resulting from off-line top-down strategy on LTQ-FTICR with CID and ECD.

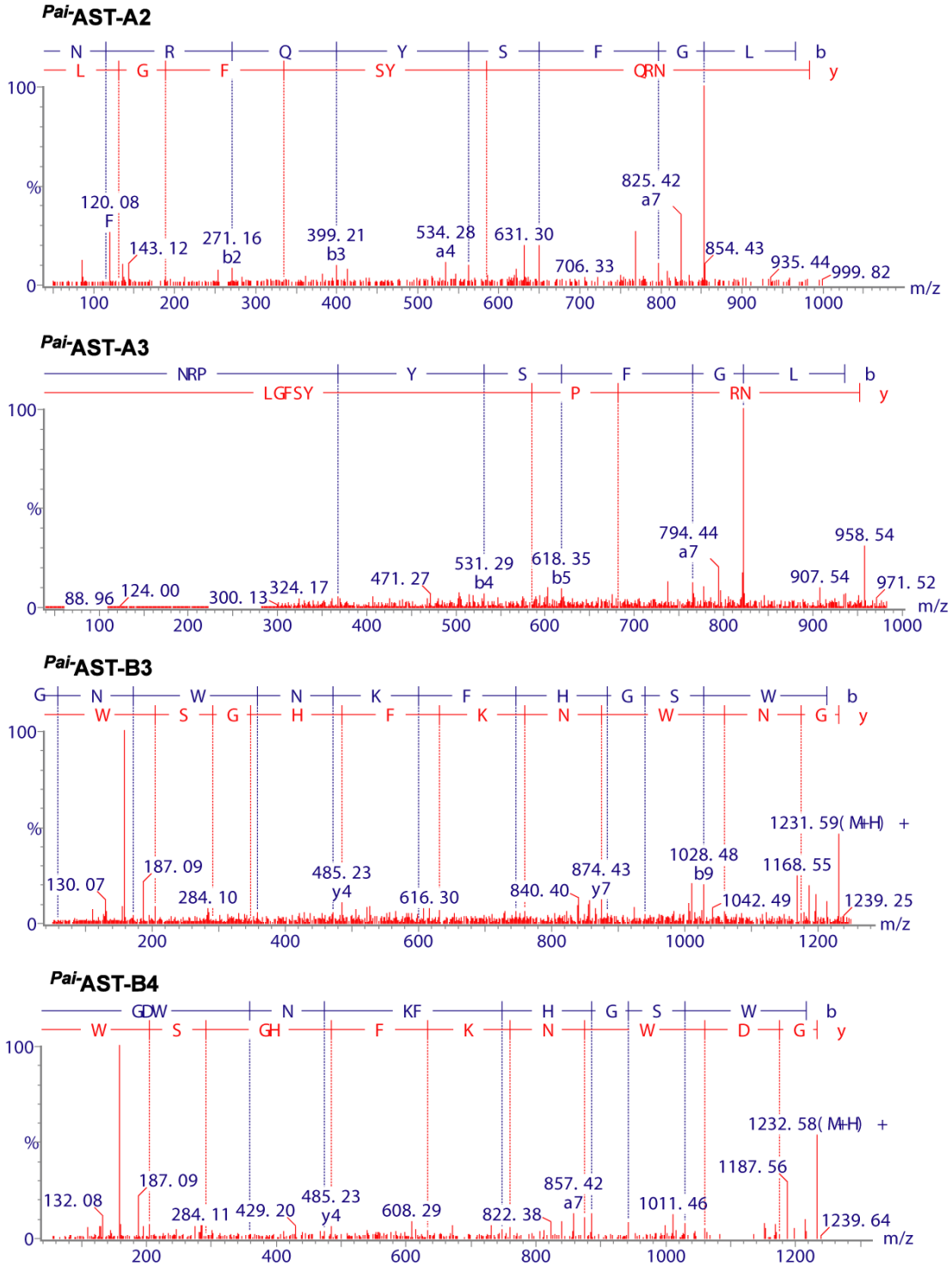


Figure S-3. *De novo* sequencing MS/MS spectra of *Pai*-AST-A2, *Pai*-AST-A3, *Pai*-AST-B3, and *Pai*-AST-B4.

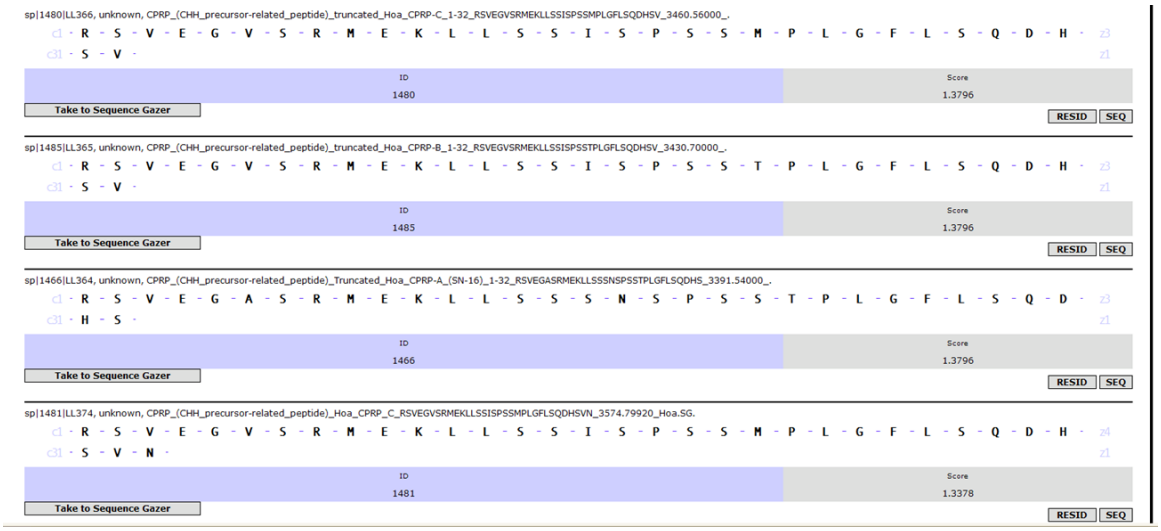


Figure S-4. Homology search result of ^{Pai}-CPRP by ProSight showing hits of several homologous CPRP peptides.

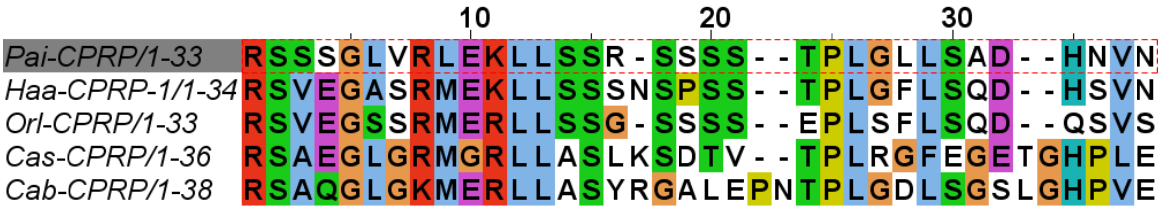


Figure S-5. Sequence alignment of five CPRPs from different crustacean species. *Pai, Panulirus interruptus; Haa, Homarus americanus; Orl, Orconectes limosus; Cas, Callinectes sapidus; Cab, Cancer borealis.*

1 **AVFDQSCKGV YDR**SLFSK**LD RVCDDCYNLY RK**HYVATGCR R**NCYGNLVFR**
51 QCLDDLMLVD VVDEYYVASV QMV

Figure S-6. Homology search result of ^{Pai}-CHH by Mascot showing match of homologous sequence of ^{Jal}-CHH. The residues in red are matched to the tryptic peptides of ^{Pai}-CHH.

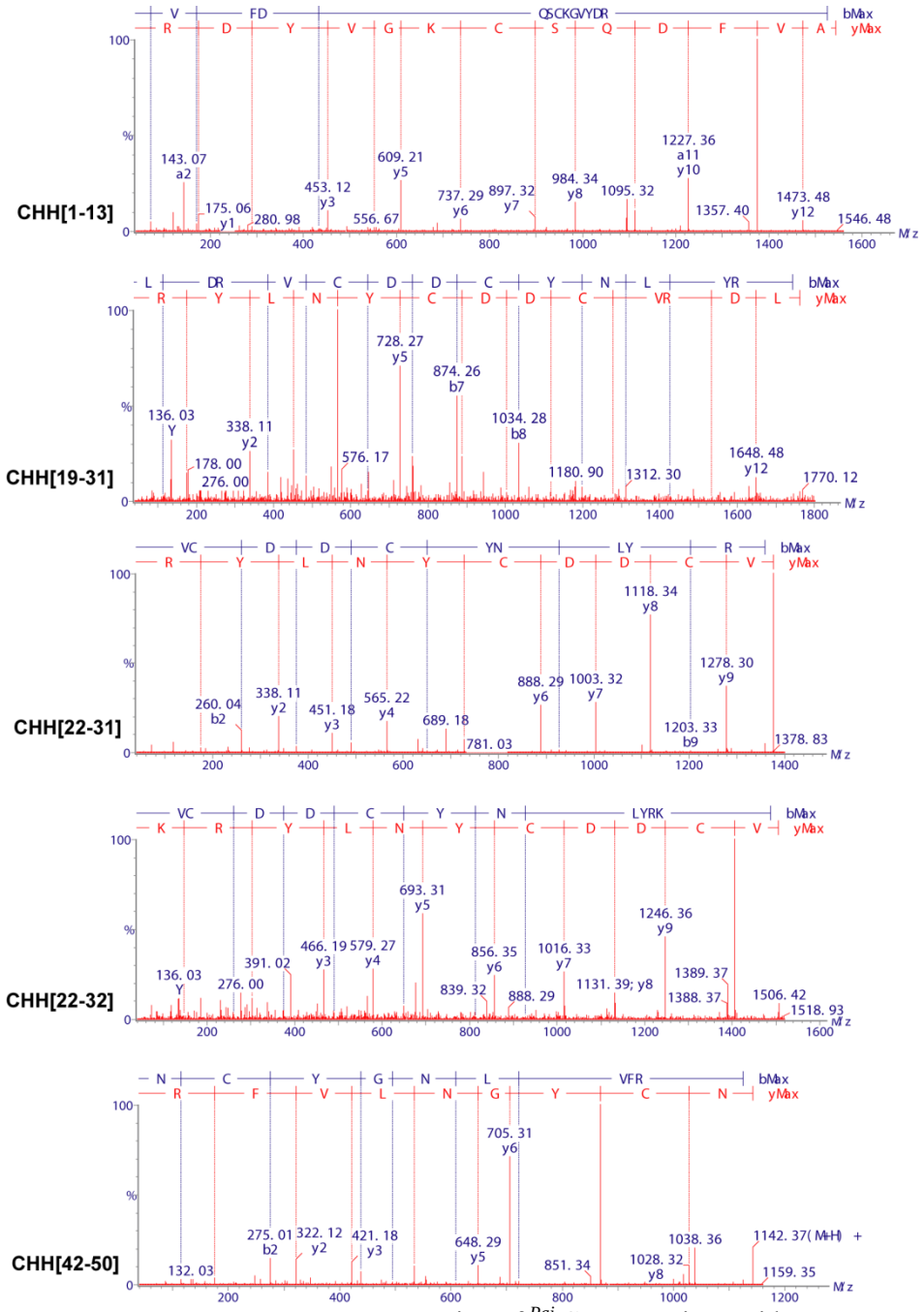


Figure S-7. *De novo* sequencing of $^{Pai-}$CHH tryptic peptides.

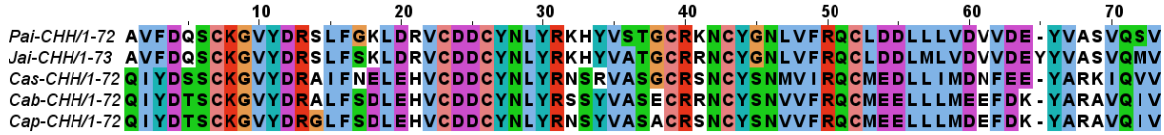


Figure S-8. Sequence alignment of five CHHs from different crustacean species. *Pai, Panulirus interruptus; Jal, Jasus lalandii; Cas, Callinectes sapidus; Cab, Cancer borealis; Cap, Cancer pagurus.*

Table S-1. BLAST homology search results of novel AST-family neuropeptides.

Peptide name	Peptide Sequence	Homo. species	Accession no.	Homo. gene name	Homo. sequence	E-value	Max Score
*Pai-*AST-A1	SRQYAFGLamide	*Panulirus interruptus* [a]	AB245091.1 [a]	prepro-AST mRNA [a]	SRQYAFGL [a]	0.003 [a]	20.4 [a]
*Pai-*AST-A2	NRQYSFGLamide	*Panulirus interruptus* [a]	AB245091.1 [a]	prepro-AST mRNA [a]	NRQYSFGL [a]	0.002 [a]	20.8 [a]
*Pai-*AST-A3	NRPYSFGLamide	*Panulirus interruptus* [a]	AB245091.1 [a]	prepro-AST mRNA [a]	PYSFGL [a]	0.037 [a]	16.9 [a]
*Pai-*AST-B1	ANWNKFHGSWamide	*Balanus amphitrite*	JQ864192.1	B-type prepro-AST	AKWNGFAGSW	0.066	18.1
*Pai-*AST-B2	ADWNKFHGSWamide	*Balanus amphitrite*	JQ864192.1	B-type prepro-AST mRNA	DWNALHGNW	0.032	18.9
*Pai-*AST-B3	GNWNKFHGSWamide	*Balanus amphitrite*	JQ864192.1	B-type prepro-AST mRNA	DWNALHGNW	0.22	16.9
*Pai-*AST-B4	GDWNKFHGSWamide	*Balanus amphitrite*	JQ864192.1	B-type prepro-AST mRNA	DWNALHGNW	0.056	18.9

[a] This gene is from the same species (*P. interruptus*), instead of homologous species.

Table S-2 Sequential b ions of ^{pai}-CHH produced by top-down CID fragmentation.

b ion	Exptl. MW	Calcd. MW	Error/ppm
b50	5855.77	5855.78	1.1
b51	5983.83	5983.84	2.2
b52	6086.84	6086.85	1.6
b53	6199.93	6199.93	0.8
b54	6314.96	6314.96	0.5
b55	6429.98	6429.99	0.3
b56	6543.05	6543.07	2.2
b57	6656.15	6656.15	0.1
b58	6769.24	6769.24	0.0
b59	6868.29	6868.31	2.1
b60	6983.33	6983.33	0.0
b61	7082.41	7082.40	0.7
b62	7181.47	7181.47	0.6
b63	7296.50	7296.49	0.3
b64	7425.54	7425.54	0.4
b65	7588.60	7588.60	0.1
b66	7687.67	7687.67	0.1
b67	7758.71	7758.71	0.1
b68	7845.73	7845.74	1.3
b69	7944.81	7944.81	0.2
b70	8072.86	8072.87	1.4
b71	8159.90	8159.90	0.0

Chapter 4

Site-Specific Characterization of D-Amino Acid-Containing Peptide Epimers by Ion Mobility Spectrometry

Adapted from: Chenxi Jia, Christopher B. Lietz, Qing Yu, and Lingjun Li, Site-Specific Characterization of D-Amino Acid-Containing Peptide Epimers by Ion Mobility Spectrometry. *Anal Chem*, submitted.

Abstract

Traditionally, the D-amino acid-containing peptide (DAACP) candidate can be discovered by observing the differences of biological activity and chromatographic retention time between the synthetic peptides and naturally occurring peptides. However, it is difficult to determine the exact position of D-amino acid in the DAACP candidates. Herein, we developed a novel site-specific strategy to rapidly and precisely localize D-amino acids in peptides by ion mobility spectrometry (IMS) analysis of mass spectrometry (MS)-generated epimeric fragment ions. Briefly, the D/L-peptide epimers were separated by on-line reversed-phase liquid chromatography and fragmented by collision induced dissociation (CID), followed by IMS analysis. The epimeric fragment ions resulting from D/L-peptide epimers exhibit conformational differences, thus showing different mobility in IMS. The drift time shift between the epimeric fragment ions was used as criteria to localize the D-amino acid substitution. The utility of this strategy was demonstrated by analysis of peptide epimers with different molecular sizes, [D-Trp]-Melanocyte stimulating hormone, [D-Ala]-Deltorphin, [D-Phe]-Achatin-I and their all-L forms. Furthermore, the crustacean hyperglycemia hormones (CHHs, 8.5 kDa) were isolated from American lobster and identified by integration of MS-based bottom-up and top-down sequencing approaches. The IMS data acquired with use of our novel site-specific strategy indicates that the isomerization of L- to D-Phe occurs at the third residue of the CHHs from the N-terminus. Collectively, this study highlights a new route for discovery of DAACPs using IMS technique centered on localization of D-amino acid residues.

4.1 Introduction

The isomerization of an L- to D-amino acid is a remarkable post-translational modification of peptides in RNA-based protein synthesis and has been documented in amphibians, invertebrates and mammals [1-4]. In many cases, the D-amino acid-containing peptides (DAACPs) exhibit dramatically higher affinity and selectivity for receptor binding than their all-L counterparts and thus are essential for biological function [4]. Generally, the targeted approaches for discovery of endogenous DAACPs include two steps: screening DAACP candidates in biological samples and then localizing D-amino acid residues [2, 5]. Many new DAACPs were found by observing the differences in biological activity or chromatographic retention time between synthetic peptides and naturally occurring peptides [2-5]. In addition, immunoassays based on conformational antibodies have been successfully used to screen DAACPs at the tissue and cellular level [2, 6]. For localization of D-amino acids in DAACP candidates, the most popular approach relies on matching chromatographic retention time of the naturally occurring peptide with a panel of synthetic peptides [2]. For example, validation of a deca-DAACP presumably requires testing ten synthetic peptides, each of which contains a D-amino acid at a various position, leading to high cost and limited analytical throughput. Other techniques utilize Edman degradation [7] or acid hydrolysis [2, 8] to release free amino acids, followed by chromatographic analysis of the free or derivatized amino acids. However, cleavage of amide bond by chemical methods induces a 3-15% level of racemization [9]. Therefore, there is a great demand for development of a simple and low-cost method to localize D-amino acids in a wide range of DAACP candidates.

In the last decade, mass spectrometry (MS) has become a powerful tool for peptidomics studies. However, differentiation of D/L-peptide epimers by MS represents a major analytical

challenge because the peptide epimers share identical masses and primary structures [10, 11]. Consequently, L- to D-amino acid isomerization is largely overlooked in MS-based peptide discovery. Recently, a variety of fragmentation techniques in tandem mass spectrometry (MS/MS), such as collision-induced dissociation (CID) [9, 12-14], metastable decomposition [14], electron capture dissociation [9, 15], and radical-directed dissociation [16], have been successfully applied in discriminating D/L-peptide epimers through the comparison of fragment ion intensities. Although excellent differentiation and quantitation between D/L-peptide epimers can be accomplished by these strategies, localization of D-amino acid in peptides is still difficult, as measurement of fragment ion intensities cannot provide accurate positional information of D-amino acids. To address this problem, this paper introduces a novel ion mobility spectrometry-mass spectrometry (IMS-MS)-based strategy enabling site-specific characterization of DAACP epimers to localize D-amino acids.

The IMS-MS technique has been widely used to probe the gas-phase conformations of biomolecules by measuring their mobility in a buffer gas [17-21] and has shown very broad applicability in the separation and identification of isomeric peptides [18, 22, 23]. Previous studies reported that intact D/L-peptide epimers displayed different mobility in IMS due to conformational differences attributed by substitution of an L- to D-amino acid [24, 25]. Presumably, under CID fragmentation the two peptide epimers may produce epimeric fragment ions which contain the same amino acid sequences but differ by substitution of an L- to D-amino acid. This raises a question of whether the epimeric fragment ions derived from D/L-peptide epimers exhibit different mobility in IMS due to conformational differences, and whether or not those differences can pinpoint the D-amino acid. For example, the peptide epimers PEP_DTIDE and PEP_LTIDE may respectively produce the epimeric y_6 ions,

EP_DTIDE^+ and EP_LTIDE^+, which show different mobility in IMS analysis due to substitution of D/L-Thr. In contrast, the y_3 ions IDE^+ derived from the two peptide epimers contain the same all-L amino acid residues and thus exhibit the same mobility. Therefore, sequential mobility analysis of epimeric fragment ions may be able to accurately identify D-amino acids in DAACP candidates. Based on this concept, we developed a novel LC-MS/MS-IMS strategy which allows site-specific characterization of peptide epimers. The practical utility was demonstrated by analysis of peptide standards, [D-Trp]-Melanocyte-stimulating hormones (MSH), [D-Ala]-Deltorphin (DTP), [D-Phe]-Achatin-I and their all-L forms. The strategy was then applied to determine the isomerization of an L- to D-Phe in crustacean hyperglycemia hormones (CHHs) isolated from sinus gland of American lobster.

4.2 Experimentals

4.2.1 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). Peptide standards, [D-Trp]-MSH and its all-L form were purchased from American Peptide Company; and [D-Ala]-Deltorphin and [D-Phe]-Achatin-I as well as their all-L forms were synthesized in Biotechnology Center, University of Wisconsin-Madison.

4.2.2 Animals, Tissue Dissection and Extraction.

American lobster *Homarus americanus* were purchased from Maine Lobster Direct Web site (http://www.mainelobsterdirect.com). All animals were kept in a circulating artificial seawater tank at 10−15 °C. Tissue dissection and extraction was performed according to our previous reports.[26, 27] Briefly, 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. The tissues were homogenized and extracted with 100 μL of acidified methanol (methanol:H_2O:acetic acid, 90:9:1, v:v:v) for three times.

4.2.3 LC-MS/MS Coupled to Ion Mobility Spectrometry

The LC-MS/MS-IMS experiments were performed on a Waters nanoAcquity ultra performance LC system coupled to a Synapt G2 high-definition mass spectrometer. Chromatographic separations were performed on a Waters BEH 300Å C18 reversed phase capillary column (150 mm X 75 μm, 1.7 μm). The mobile phases used were: 0.1% formic acid in water (A); 0.1% formic acid in ACN (B). The peptide sample was injected and loaded onto the Waters Symmetry C18 trap column (180 μm X 20 mm, 5 μm) using 97% mobile phase A and 3% mobile phase B at a flow rate of 5 μL/min for 3 min. The gradient started from 3 to 10% B during the first 5 min, increased to 45% B in the next 65 min, then was kept at 90% B for 20 min. A fixed MS/MS survey was employed to select the peptide molecular ions in a traveling-wave (T-Wave) trap cell for CID fragmentation with adjusted collision energy 22-30 eV. The resulting fragment ions were on-line submitted to T-Wave drift tube and time of flight analyzer to measure the drift time. Instrument acquisition parameters used were as follows: an inlet capillary voltage of 2.8 kV, a sampling cone setting of 35 V, and a source temperature of 70 °C. The argon gas pressure in the traveling wave ion guide trap and the traveling wave ion guide transfer cell were 2.44×10^{-2} and 2.61×10^{-2} mbar, respectively. The wave height, the wave velocity, and the nitrogen pressure in the traveling wave IM drift cell were 32.0 V, 800 m/s, and 2.96 mbar, respectively.

Data processing was conducted using Waters MassLynx 4.1 and DriftScope 2.1. The LC-MS/MS-IMS .raw data was opened in DriftScope and the Selection Tool is used to respectively

select the two peptide epimer LC peaks and export the two corresponding Masslynx .raw data by retaining drift time functions. The drift time distributions of interested ions were exported from Masslynx.

4.2.4 Calibration of Collision Cross Sections on Ion Mobility Spectrometry

The T-Wave N_2 drift tube of Synapt G2 was calibrated for collision cross section (CCS) measurements using a slightly modified version of the polyalanine method outlined by Bush et al [28]. and the calculation method by Ruotolo et al. [29] A 100 μg mL^{-1} solution (49:49:2, water:acetonitrile:acetic acid) of polyalanine was directly infused and acquired with the same instrument and method settings used to acquire the DAACP analyte. A peaklist containing *m*/*z*-corrected mobility and drift times was then exported from DriftScope to a csv for CCS calculation. Further details of the CCS calculation and a representative calibration curve can be found in Supporting Information. It should be noted that this method allows measurement of CCS_{He} in spite of using nitrogen as drift gas in our experiment [28]. Thus, the CCSs shown in this paper correspond to the helium drift gas.

4.2.5 Tryptic Digestion and Bottom-up sequencing of CHHs on Synapt G2 HDMS

For trypsin digestion of CHHs, 1 μL of tissue extract 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, and then digested at 37 °C overnight after addition of 50 mM ammonium bicarbonate buffer with 0.5 μg of trypsin (Promega, Madison, WI). The tryptic digest was injected into a Waters nanoAcquity UPLC system coupled to a Synapt G2 HDMS. Chromatographic conditions are the same as described above. 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.

4.2.6 Top-down MS/MS Fragmentation of CHHs on Q Exactive

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 0.2% formic acid. On-line top-down MS was carried out on Waters nanoAcquity ultra performance LC system coupled to a Q Exactive quadrupole orbitrap mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). A 0.5 µL of peptide sample was injected onto The peptide sample was injected and loaded onto the Waters Symmetry C18 trap column (180 µm x 20 mm, 5 µm) using 97% mobile phase A and 3% mobile phase B at a flow rate of 5 µL/min for 3 min. A Waters BEH 300Å C18 reversed phase capillary column (150 mm x 75 µm, 1.7 µm) was used for separation. The gradient started from 3 to 10% B during the first 5 min, increased to 55% B in the next 65 min, then was kept at 90% B for 20 min. Typical mass spectrometric conditions were as follows: spray voltage, 2.8 kV; no sheath and auxiliary gas flow; heated capillary temperature, 275 °C; normalized HCD collision energy 30%. The Q Exactive instrument was operated in targeted MS/MS mode with an inclusion list containing the targeted mass of the CHHs. The settings are: resolution 70, 000; automatic gain control $2e^5$; maximum ion injection time, 100 ms; isolation window, 6 m/z; fixed first mass, 100 m/z. All MS/MS spectra were processed with Xtract CI-3.0 Software (Thermo Scientific Inc., Bremen, Germany) using an 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 with of 10 ppm of mass error tolerance. The assigned ions were manually validated to ensure the quality of assignments.

4.3 Results and Discussion

4.3.1 Workflow and Rationale of the Proposed Site-Specific Strategy for localization of D-Amino Acid in Peptide Epimers.

Figure 1 illustrates the workflow for the proposed strategy. The analysis can be completed in one LC-MS/MS-IMS run. The peptide epimers are separated by reversed-phase LC and on-line submitted to CID fragmentation. The resulting peptide fragment ions are then subjected to IMS for measurement of drift time. The epimeric ions of y_6, y_5 and y_4 derived from the two peptide epimers respectively contain the L- or D-Thr, which possibly leads to conformational differences between each epimeric y ion pair, resulting in drift time shift during IMS analysis. In contrast, the two peptide epimers produce the same y_3, y_2 and y_1 ions containing all-L amino acids, since the D- or L-Thr has been removed from peptide chain by CID fragmentation. Thus, these y ion pairs show identical drift times. By determining at which the drift time shift suddenly appears, the D-amino acid can be confidently localized at the threonine. Sometimes, a LC-MS-IMS run without fragmentation is needed to measure the drift times of peptide molecular ions when the D-amino acid is at the N-terminus. To validate the proposed strategy, we analyzed three pairs of peptide epimers with various molecular sizes: [D-Trp]-MSH (MW 1569.73 Da), [D-Ala]-DTP (768.38 Da), [D-Phe]-Achatin-I (407.18 Da) and their all-L forms.

4.3.2 Site-specific Characterization of MSH Peptide Epimers for Localization of D-amino Acids.

The MSHs are produced by cells in the intermediate lobe of the pituitary gland, which stimulates the production and release of melanin and effects on appetite and sexual arousal [30]. It was reported that the D-Trp substituted isoform was the most selective analogue for the melanocortin receptors [30]. In our study, the peptide epimers D/L-MSH

(YVMGHFR$_D$WDRFG and YVMGHFRWDRFG, 1:1, concentration ratio) were analyzed by LC-MS-IMS on a Waters Synapt G2 HDMS mass spectrometer coupled to a nanoACQUITY UPLC system. Figure 2A shows the extracted ion chromatography of the two peptides with RPLC baseline separation. The $[M+3H]^{3+}$ ions of the two epimers exhibit different drift time at 2.32 and 2.39 ms (Figure 2B and C). Subsequently, the peptide mixture was analyzed by LC-MS/MS-IMS, where the two peptide epimers were fragmented by CID and then the resulting fragment ions were submitted to IMS for drift time measurement. In their CID spectra (Figure S1), a set of y ions, y_3~y_{11} are observed, and the corresponding drift time distributions are illustrated in Figure 3D. The D/L-epimeric ions of y_5, y_6, y_9 and y_{11} exhibit drift time shifts, whereas the paired y ions of y_3 and y_4 derived from the two peptide epimers have the identical drift time distributions. More importantly, the drift time shift between the epimeric y ions starts from y_5, which is the first y ion that contains tryptophan by counting from y_3 to y_{11}. The results indicate that the D-amino acid is localized at the tryptophan (YVMGHFR$_D$WDRFG) as annotated in Figure 2E. Although the epimeric ions of y_{10}, y_8 and y_{11} ions contain D/L-Trp, they do not show drift time shifts (Figure 2D). These exceptions may attribute to the possibilities that the conformational differences are too small to be resolvable due to limited revolving power of our IMS instrument, or the conformational differences of D/L-Trp are comprised by other factors that may affect structure changes.

Previous studies reported that D/L-peptide epimers exhibit conformational differences, leading to distinguished mobility in IMS [24, 25]. The orientation of the epimeric amino acid residues could produce unique intramolecular interactions that lead to a more extended or compact overall shape. In this study, we hypothesize that the same phenomenon occurs with the CID-produced epimeric fragments. If the fragments are epimers, the different amino acid

orientation could still produce unique intramolecular interactions and unique conformations. However, if the fragments are identical, then the intramolecular interactions and resulting conformations may also be identical. To support this assumption, we measured the CCSs of the precursor and fragment ions arising from D/L-MSH [28, 29]. The results are listed in Table S1 and their CCS differences (ΔCCSs) are illustrated in Figure 2F. The epimeric $[M+3H]^{3+}$ ions of D-MSH and L-MSH show a 6.9 $Å^2$ of ΔCCS, which reveals their conformational differences attributed to the substitution of D/L-Trp. The epimeric fragment ions of y_5~y_{11} containing D/L-Trp show an averaged ΔCCS of 1.16 $Å^2$, while the y_3 and y_4 ions without D/L-Trp have an averaged ΔCCS of 0.19 $Å^2$. More importantly, the epimeric y_5 ions with D/L-Trp at the N-terminus exhibit a remarkable ΔCCS of 4.30 $Å^2$, which clearly indicate that the D-amino acid is localized at the tryptophan. In addition, the ΔCCSs of epimeric y_{10}, y_8 and y_7 ions are 0.20, 0.21 and 0.75, respectively. These small differences indicate a possible reason why their drift time shifts are not resolved in Figure 2D. To validate the assumption, more evidences should be obtained and several fundamental questions should be answered, which will be investigated in our follow-up work. This paper focuses on the analytical methodology for discrimination of peptide epimers. Based on the concept described above, two peptide standards with smaller sizes are further investigated to validate the proposed site-specific strategy.

4.3.3 Site-specific Characterization of DTP and Achatin-I Peptide Epimers for Localization of D-amino Acids.

Deltophins are a family of naturally occurring peptides found in skin extracts of frogs. They have high affinity and selectivity for delta opioid binding sites [31]. Here, we choose [D-Ala]-DTP ($Y_DAFDVVG\text{-}NH_2$) and its all-L form ($YAFDVVG\text{-}NH_2$) to further validate our

proposed strategy. The two peptide epimers are baseline separated by LC and eluted at 32.77 and 37.29 min (Figure 3A and B), respectively. The drift time measurement indicates a shift of 0.16 ms between the two epimeric $[M+H]^+$ ions (Figure 3C). In the subsequent LC-MS/MS-IMS experiment, the epimeric ions of y_4, b_4, b_5 and b_6 exhibit different mobility, leading to drift time shift as shown in Figure 3D. In contrast, the paired y ions of y_2~y_5 derived from the two peptide epimers display nearly identical drift times. Figure 3E summarizes the results of drift time shifts of these fragment ions, clearly indicating that the D-amino acid is localized at the alanine ($Y_DAFDVVG\text{-}NH_2$).

Achatin-I (G_DFAD) is a neuroexcitatory peptide in *Achatina fulica* ganglia [32]. On our nanoRPLC system, the Achatin-I and its all-L form (GFAD) were not baseline separated due to a short retention (data not shown). Regarding the issue of poor chromatographic behavior of some peptides in RPLC, other separation techniques, such as chiral chromatograph [33] or capillary electrophoresis [34], could be incorporated into our strategy. To exam the proof-of-principal, we assume that the two peptide epimers had been off-line separated, and the individual peptide epimer was directly infused into mass spectrometer for MS-IMS and MS/MS-IMS analysis. The $[M+H]^+$ ions of the two peptide epimers show the same drift time at 5.20 ms. The immeasurable drift time shift may attribute to the limited resolution of our IMS instrument. Interestingly, the epimeric y_3 ions derived from the peptide epimers exhibit a 0.08 ms of drift time shift as shown in Figure 3F. In contrast, the y_2 ions from the two peptide epimers have the identical drift time at 2.79 ms. These results provide precise positional information for localization of D-amino acid at the phenylalanine of G_DFAD (Figure 3F).

4.3.4 Site-Specific Characterization of D-amino Acids in CHHs Isolated from American

Lobster.

To apply the site-specific strategy to identification of DAACPs in real biological samples, we sequenced the CHHs in American lobster and determined the D-amino acid position in these peptides. CHHs are a family of neurohormone released from crustacean sinus glands, which regulates the glycemia through the classical mechanisms of glycogen mobilization [35]. Soyez et al. [5, 36] reported that the sinus gland of American lobster secrets two peptide hormones, CHH_{-A} and CHH_{-B}, and both of CHH_{-A} and CHH_{-B} occur as two isoforms with D/L-Phe at the third residue from the N-terminus. The biological activities of the D/L-isoforms differ in the kinetics of their hyperglycemic effect [5].

The first step is to identify the CHHs from American lobster. The identification is challenging, as CHHs contain more than 70 residues and multiple PTMs, including three disulfide bonds [26, 35]. In this work, we employed an on-line top-down approach. The intact peptide and dithiothreitol (DTT)-reduced peptide were analyzed on a high-resolution Q Exactive Orbitrap mass spectrometer by targeted MS/MS to acquire high-quality top-down MS and MS/MS spectra. Figure 4A shows the isotopic distributions of the CHH_{-A} before and after reduction of disulfide bonds with an error less than 2 ppm. The mass increase of 6 Da resulting from DTT reduction means that the peptides contain three disulfide bonds, which is an important criterion to discriminate CHHs from other peptide families [35]. Figure 4B is the top-down HCD MS/MS spectrum of the reduced CHH_{-A}. Compared with HCD fragmentation of the intact CHH_{-A} (data not shown), the efficiency of peptide fragmentation is dramatically improved due to elimination of disulfide linkage, where 46% of sequence coverage and 62% of amide bond cleavage are achieved. The fragmentation maps of CHH_{-A} and CHH_{-B} are shown in Figure 4C. These results confirm the identities of the two CHHs in

American lobster. However, baseline separation of D- and L-CHHs by nanoRPLC is difficult (Figure S2) because of the negligible differences of the hydrophobicity and sequence-dependent effects of the D/L-isoforms attributed to the third D/L-Phe. Therefore, we adopted a bottom-up approach in the following experiment.

The second step is to find the target tryptic peptide epimers that contain D/L-amino acids. After treatment with DTT and iodoacetamide (IAA), the crude extract was digested by trypsin and analyzed by RPLC-MS/MS in a data dependent mode and processed using software PEAKS [37] against a database containing the sequences of CHH_{-A} and CHH_{-B}. Table S2 lists all the tryptic peptides and their corresponding LC retention times. The tryptic peptide CHH[1-18] is eluted at 21.42 and 26.69 min, and CHH[1-17] at 33.21 and 35.70 min. The elution pattern of split speaks means that the two peptides are D/L-epimer candidates. In contrast, the rest of tryptic peptides are eluted at single time points. Figure 5A shows the representative extracted ion chromatograph of tryptic peptide *p*QVFDQAC*K (*p*Q, pyro-Gln; *, carbamindomethyl), where the D/L-peptide epimer candidates are baseline separated. It should be noted that CHH_{-A} and CHH_{-B} share the same N-terminal sequence from ^{1}pyro-Gln to ^{19}Leu, and the residue D/L-Phe is at the third position from N-terminus [5, 36]. To obtain the precise D-amino acid information of each peptide, the two CHHs should be purified and separately analyzed. The goal of our study is to examine the practical utility of the site-specific strategy in real biological sample. Herein, we used a simplified experimental procedure by directly investigating the entire tryptic digest of the tissue extracts.

The third step is to localize the D-amino acid in the target tryptic peptides, *p*QVFDQAC*K and *p*QVFDQAC*KGVYDRNLFK. The entire tryptic digest of the tissue extracts was analyzed by LC-MS-IMS and LC-MS/MS-IMS. Figure 5B and C show the MS/MS spectra of

peptide epmiers, *p*QV$_{D/L}$FDQAC*K, and the corresponding drift time distributions of the fragment ions. As summarized in Figure 4D, the paired y ions of y_2~y_5 derived from the two peptide epimers have the same drift time, whereas those of b_4, y_6, $[M+H]^+$ and $[M+2H]^{2+}$ display drift time shifts. Similarly, LC-MS/MS-IMS analysis of peptide epimers *p*QVFDQAC*KGVYDRNLFK (Figure 6) indicates that the paired y ions of y_5~y_{14} show the identical drift time distributions, while those of y_{15}, b_4, and $[M+H]^+$ ions present drift time shifts. These results suggest that the D-amino acid is localized at the phenylalanine of the third residue of the CHHs from the N-terminus.

4.3.5 Issues for Practical Utility of the Site-Specific Strategy.

In this work, we demonstrated the broad application and practical utility of the site-specific strategy for characterization of DAACPs and their all-L counterparts. By employing LC-MS/MS-IMS, the position of D-amino acid in the peptide epimers can be rapidly and precisely determined. Specifically, the LC-MS/MS experiment serves for screening D/L-peptide epimers in biological samples, so the initial candidates of D/L-peptide epimers must display differential LC retention times. An LC-MS-IMS experiment can be carried out to measure the drift time of the peptide molecular ions to obtain complementary IMS evidence to suggest epimers. Lastly, the site-specific strategy based on LC-MS/MS-IMS experiment is utilized to localize the D-amino acid.

The co-existence of DAACP and all-L forms is widely found in biological systems due to incompletely enzymatic isomerization of an L- to D-amino acid. In contrast, some DAACPs are found without their all-L counterparts [2, 4]. Characterization of them does not require the on-line separation prior to MS/MS-IMS analysis. The purified DAACP candidate and the synthetic all-L counterpart are respectively infused into mass spectrometer for IMS analysis,

and the D-amino acid can be localized by the site-specific strategy. In addition, our site-specific strategy can be incorporated into the non-targeted strategies for DAACP discovery. For example, Ewing et al [38]. employed microsomal alanyl aminopeptidase to selectively degrade peptides lacking a D-amino acid in the second position from the N-terminus, so the DAACPs can be identified from a complex mixture. Our site-specific strategy can serve as a downstream tool for validation of these DAACP candidates. However, we cannot exclude the possibility that some peptide epimers and their CID-produced fragments show the same shape (maybe different conformation) in spite of containing D/L-amino acids, or their shape differences are too slight to be resolvable by IMS, such as the epimeric y_7, y_8 and y_{10} ions of D/L-MSH (Figure 2D).

In this study, we measured the drift time distributions of y, b and molecular ions of peptide epimers, indicative of the D-amino acid position. It is expected that other peptide fragment ions, such as a, c, z, neutral loss ions, etc, could be used as indicators for site-specific characterization because of being sensitive to chirality. In addition, the peptide epimers are eluted at different time points by LC solvents containing different percentage of organic phase. For example, the D-MSH and L-MSH were eluted in 24% and 25% of acetonitrile, respectively. To rule out the possibility of drift time shift caused by various organic phase percentages in solvents, we analyzed the L-MSH in 24% and 30% of acetonitrile by direct infusion into MS/MS-IMS and found that the drift time distributions of peptide precursor ions and fragment ions were not changed (data not shown). This result means that the measurement conditions between the two peptide epimers in one LC-MS/MS-IMS run have a negligible effect on the changes of their drift time distributions.

4.4 Conclusions

In this study, a novel strategy was developed for site-specific characterization of peptide epimers, which allows rapid and precise localization of D-amino acids in DAACP candidates. The analysis can be finished in one LC-MS/MS-IMS run, followed by data processing in a simple and straightforward manner. The efficiency and utility of the strategy were demonstrated by analysis of a set of peptides with various molecular sizes, MSH, DTP and Achatin-I and their all-L forms. In the CHHs isolated from American lobster, our data indicates that the peptides contain an isomerization site of L- to D-Phe at the third residue from the N-terminus. This study represents a new route to obtain positional information of amino acid isomerizations in peptides by elucidating the IMS data of peptide fragment ions. By coupling with efficient screening approaches, the developed strategy is potentially applicable to large-scale discovery of DAACPs; and the proposed concept is transferable to characterization of other post-translational isomerizations in large biological molecules.

4.5 Acknolegements

We are grateful to Prof. Jonathan V. Sweedler, Dr. Lu Bai, and Itamar Livnat (UIUC) for helpful discussions and insightful suggestions on this project. This work is supported in part by the National Institutes of Health (NIH) grant (R01DK071801 to LL) and the National Science Foundation grant (CHE-0957784 to LL). We would like to acknowledge NIH shared instrument program for funding the instrument purchase (S10 RR029531). LL acknowledges an H. I. Romnes Faculty Research Fellowship. We thank the UW School of Pharmacy Analytical Instrumentation Center for access to Q Exactive Orbitrap mass spectrometers.

4.6 References

[1] Iida T, Santa T, Toriba A, Imai K. Amino acid sequence and D/L-configuration determination methods for D-amino acid-containing peptides in living organisms. Biomed Chromatogr. 2001;15:319-27.

[2] Soyez D, Toullec JY, Montagne N, Ollivaux C. Experimental strategies for the analysis of D-amino acid containing peptides in crustaceans: a review. J Chromatogr B Analyt Technol Biomed Life Sci. 2011;879:3102-7.

[3] Buczek O, Yoshikami D, Bulaj G, Jimenez EC, Olivera BM. Post-translational amino acid isomerization: a functionally important D-amino acid in an excitatory peptide. J Biol Chem. 2005;280:4247-53.

[4] Bai L, Sheeley S, Sweedler JV. Analysis of Endogenous D-Amino Acid-Containing Peptides in Metazoa. Bioanal Rev. 2009;1:7-24.

[5] Soyez D, Van Herp F, Rossier J, Le Caer JP, Tensen CP, Lafont R. Evidence for a conformational polymorphism of invertebrate neurohormones. D-amino acid residue in crustacean hyperglycemic peptides. J Biol Chem. 1994;269:18295-8.

[6] Serrano L, Grousset E, Charmantier G, Spanings-Pierrot C. Occurrence of L- and D-crustacean hyperglycemic hormone isoforms in the eyestalk X-organ/sinus gland complex during the ontogeny of the crayfish Astacus leptodactylus. J Histochem Cytochem. 2004;52:1129-40.

[7] Iida T, Matsunaga H, Santa T, Fukushima T, Homma H, Imai K. Amino acid sequence and D/L-configuration determination of peptides utilizing liberated N-terminus phenylthiohydantoin amino acids. J Chromatogr A. 1998;813:267-75.

[8] Scaloni A, Simmaco M, Bossa F. Determination of the chirality of amino acid residues in the course of subtractive Edman degradation of peptides. Anal Biochem. 1991;197:305-10.

[9] Adams CM, Zubarev RA. Distinguishing and quantifying peptides and proteins containing D-amino acids by tandem mass spectrometry. Anal Chem. 2005;77:4571-80.

[10] Hurtado PP, O'Connor PB. Differentiation of isomeric amino acid residues in proteins and peptides using mass spectrometry. Mass Spectrom Rev. 2012;31:609-25.

[11] Tao WA, Cooks RG. Parallel Reactions for Enantiomeric Quantification of Peptides by Mass Spectrometry This work was supported by the U.S. Department of Energy, Office of Energy Research. W.A.T. acknowledges fellowship support from Triangle Pharmaceuticals. Angew Chem Int Ed Engl. 2001;40:757-60.

[12] 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.

[13] Serafin SV, Maranan R, Zhang K, Morton TH. Mass spectrometric differentiation of linear peptides composed of L-amino acids from isomers containing one D-amino acid residue. Anal Chem. 2005;77:5480-7.

[14] Sachon E, Clodic G, Galanth C, Amiche M, Ollivaux C, Soyez D, et al. D-amino acid detection in peptides by MALDI-TOF-TOF. Anal Chem. 2009;81:4389-96.

[15] Adams CM, Kjeldsen F, Zubarev RA, Budnik BA, Haselmann KF. Electron capture dissociation distinguishes a single D-amino acid in a protein and probes the tertiary structure. J Am Soc Mass Spectrom. 2004;15:1087-98.

[16] Tao Y, Quebbemann NR, Julian RR. Discriminating D-amino acid-containing peptide epimers by radical-directed dissociation mass spectrometry. Anal Chem. 2012;84:6814-20.

[17] Bohrer BC, Merenbloom SI, Koeniger SL, Hilderbrand AE, Clemmer DE. Biomolecule analysis by ion mobility spectrometry. Annu Rev Anal Chem (Palo Alto Calif). 2008;1:293-327.

[18] Kanu AB, Dwivedi P, Tam M, Matz L, Hill HH, Jr. Ion mobility-mass spectrometry. J Mass Spectrom. 2008;43:1-22.

[19] Enders JR, McLean JA. Chiral and structural analysis of biomolecules using mass spectrometry and ion mobility-mass spectrometry. Chirality. 2009;21 Suppl 1:E253-64.

[20] Verbeck GF, Ruotolo BT, Sawyer HA, Gillig KJ, Russell DH. A fundamental introduction to ion mobility mass spectrometry applied to the analysis of biomolecules. J Biomol Tech. 2002;13:56-61.

[21] Shvartsburg AA, Tang K, Smith RD. Two-dimensional ion mobility analyses of proteins and peptides. Methods Mol Biol. 2009;492:417-45.

[22] Shvartsburg AA, Creese AJ, Smith RD, Cooper HJ. Separation of a set of peptide sequence isomers using differential ion mobility spectrometry. Anal Chem. 2011;83:6918-23.

[23] Ibrahim YM, Shvartsburg AA, Smith RD, Belov ME. Ultrasensitive identification of localization variants of modified peptides using ion mobility spectrometry. Anal Chem. 2011;83:5617-23.

[24] de Magalhaes MT, Barbosa EA, Prates MV, Verly RM, Munhoz VH, de Araujo IE, et al. Conformational and functional effects induced by D- and L-amino acid epimerization on a single gene encoded peptide from the skin secretion of Hypsiboas punctatus. PLoS One. 2013;8:e59255.

[25] Wu C, Siems WF, Klasmeier J, Hill HH, Jr. Separation of isomeric peptides using electrospray ionization/high-resolution ion mobility spectrometry. Anal Chem. 2000;72:391-5.

[26] 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.

[27] Jia C, Lietz CB, Ye H, Hui L, Yu Q, Yoo S, et al. A multi-scale strategy for discovery of novel endogenous neuropeptides in the crustacean nervous system. J Proteomics. 2013;91:1-12.

[28] Bush MF, Campuzano ID, Robinson CV. Ion mobility mass spectrometry of peptide ions: effects of drift gas and calibration strategies. Anal Chem. 2012;84:7124-30.

[29] Ruotolo BT, Benesch JL, Sandercock AM, Hyung SJ, Robinson CV. Ion mobility-mass spectrometry analysis of large protein complexes. Nat Protoc. 2008;3:1139-52.

[30] Grieco P, Balse PM, Weinberg D, MacNeil T, Hruby VJ. D-Amino acid scan of gamma-melanocyte-stimulating hormone: importance of Trp(8) on human MC3 receptor selectivity. J Med Chem. 2000;43:4998-5002.

[31] Erspamer V, Melchiorri P, Falconieri-Erspamer G, Negri L, Corsi R, Severini C, et al. Deltorphins: a family of naturally occurring peptides with high affinity and selectivity for delta opioid binding sites. Proc Natl Acad Sci U S A. 1989;86:5188-92.

[32] Takeuchi H, Kim KH, Liu GJ, Yasuda-Kamatani Y, Minakata H, Nomoto K. Achatin-I, an excitatory neurotransmitter having a D-phenylalanine residue of Achatina giant neurones. Acta Biol Hung. 1992;43:147-58.

[33] Li T. Peptide and peptidomimetic chiral selectors in liquid chromatography. J Sep Sci. 2005;28:1927-31.

[34] Sheeley SA, Miao H, Ewing MA, Rubakhin SS, Sweedler JV. Measuring D-amino acid-containing neuropeptides with capillary electrophoresis. Analyst. 2005;130:1198-203.

[35] Chung JS, Zmora N, Katayama H, Tsutsui N. Crustacean hyperglycemic hormone (CHH) neuropeptidesfamily: Functions, titer, and binding to target tissues. Gen Comp Endocrinol. 2010;166:447-54.

[36] Ollivaux C, Gallois D, Amiche M, Boscameric M, Soyez D. Molecular and cellular specificity of post-translational aminoacyl isomerization in the crustacean hyperglycaemic hormone family. FEBS J. 2009;276:4790-802.

[37] Zhang J, Xin L, Shan B, Chen W, Xie M, Yuen D, et al. PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. Mol Cell Proteomics. 2012;11:M111 010587.

[38] Ewing MA, Wang J, Sheeley SA, Sweedler JV. Detecting D-amino acid-containing neuropeptides using selective enzymatic digestion. Anal Chem. 2008;80:2874-80.

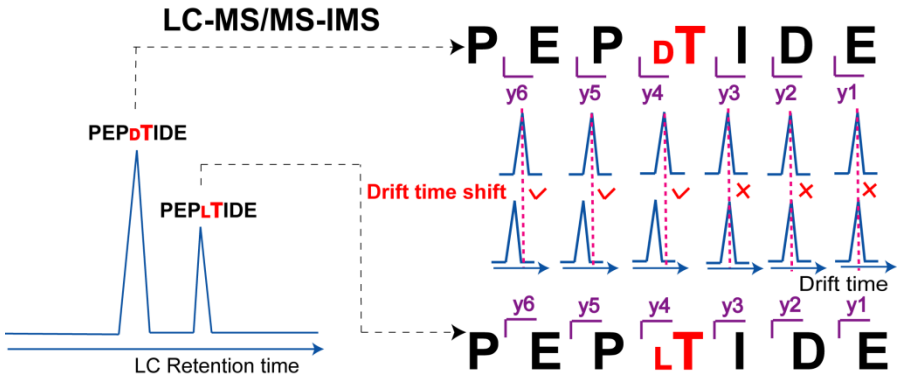


Figure 1. Workflow of the proposed strategy for localization of D-amino acids in peptides. The analysis can be performed in one LC-MS/MS-IMS run. The two peptide epimers are seperated by RPLC and respectively fragmented by CID. Their fragment ions are then submitted to IMS for drift time meausurement. By comparing the drift time distributions between the two sets of fragment ions, the position of D-amino acid can be determined. √, Drift time shift. ×, No shift. For illustration purpose, only y ions are listed in this workflow. Note that other fragment ions can also act as indicators for localization of D-amino acids.

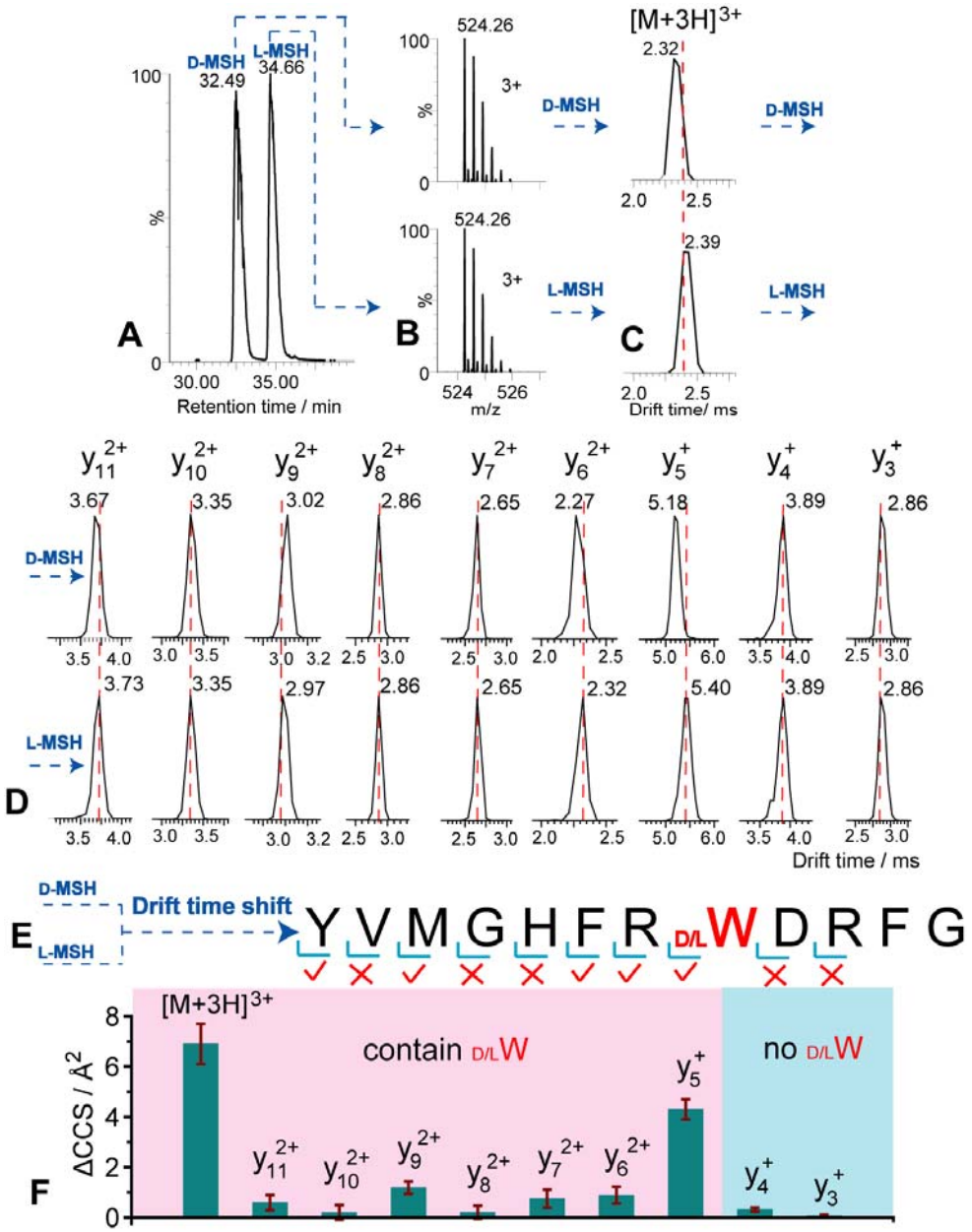


Figure 2. Site-specific characterization of D/L-MSH peptide epimers. (A) Extracted ion chromatagram of LC-MS analysis of D/L-MSH peptides. (B) Molecular ions and (C) the corresponding IMS distributions of D/L-MSH peptides. (D) IMS distributions of fragment ions of D/L-MSH peptides. (E) Localization of D-amino acid residue postion by comparison of drift time shift. √, Drift time shift. ×, No shift. (F) CCS differences (ΔCCS, absolute values) of peptide precusor and fragment ions. Error bars stand for standard deviations.

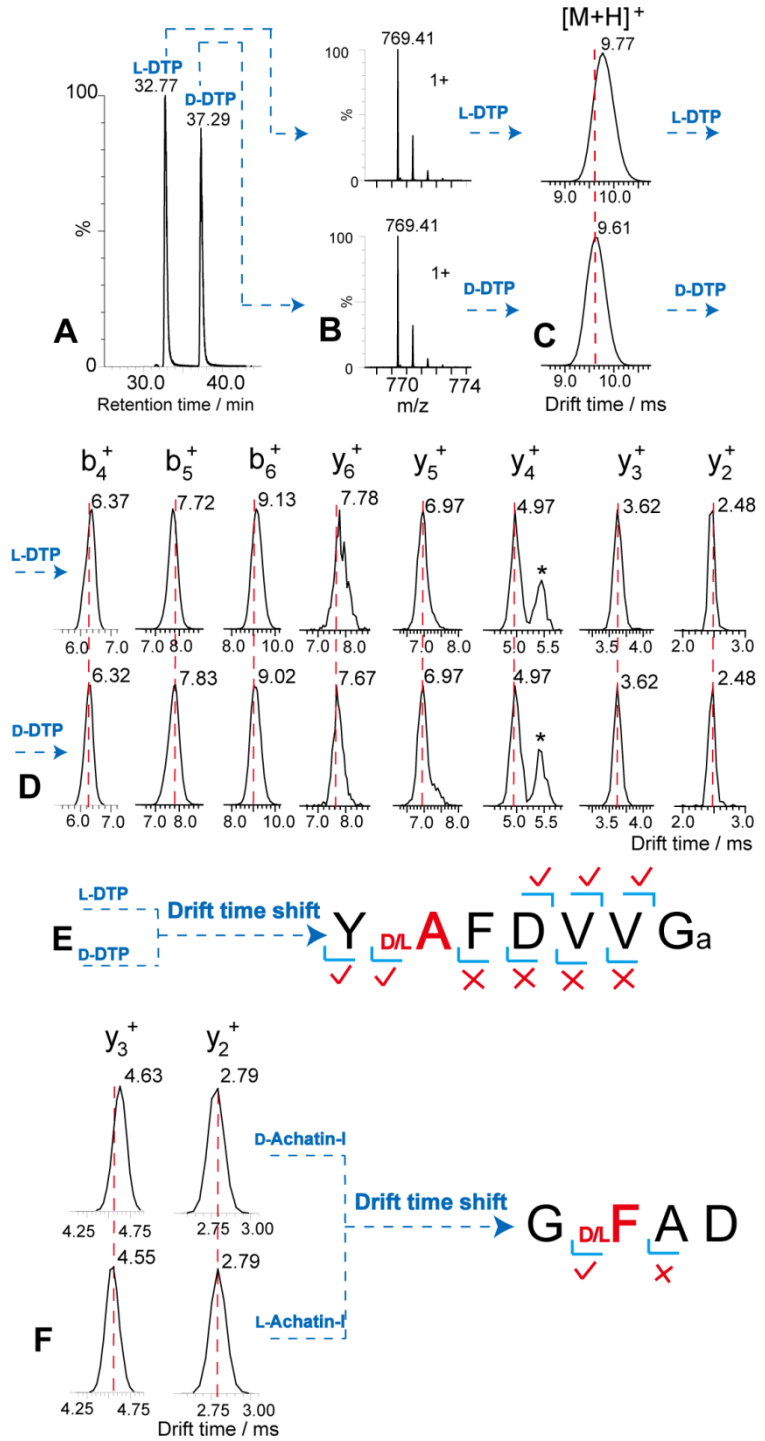


Figure 3. Site-specific characterization of $_{D/L}$-DTP and $_{D/L}$-Achatin-I peptide epimers. (A) Extracted ion chromatagram of LC-MS analysis of $_{D/L}$-DTP peptides. (B) Molecular ions and (C) the corresponding IMS distributions of $_{D/L}$-DTP peptides. (D) IMS distributions of fragment ions of $_{D/L}$-DTP peptides. (E) Localization of D-amino acid residue postion by comparison of drift time shift. (F) IMS analysis of $_{D/L}$-Achatin-I and localization of D-amino acid position. √, Drift time shift; ×, No shift; * Interference ions.

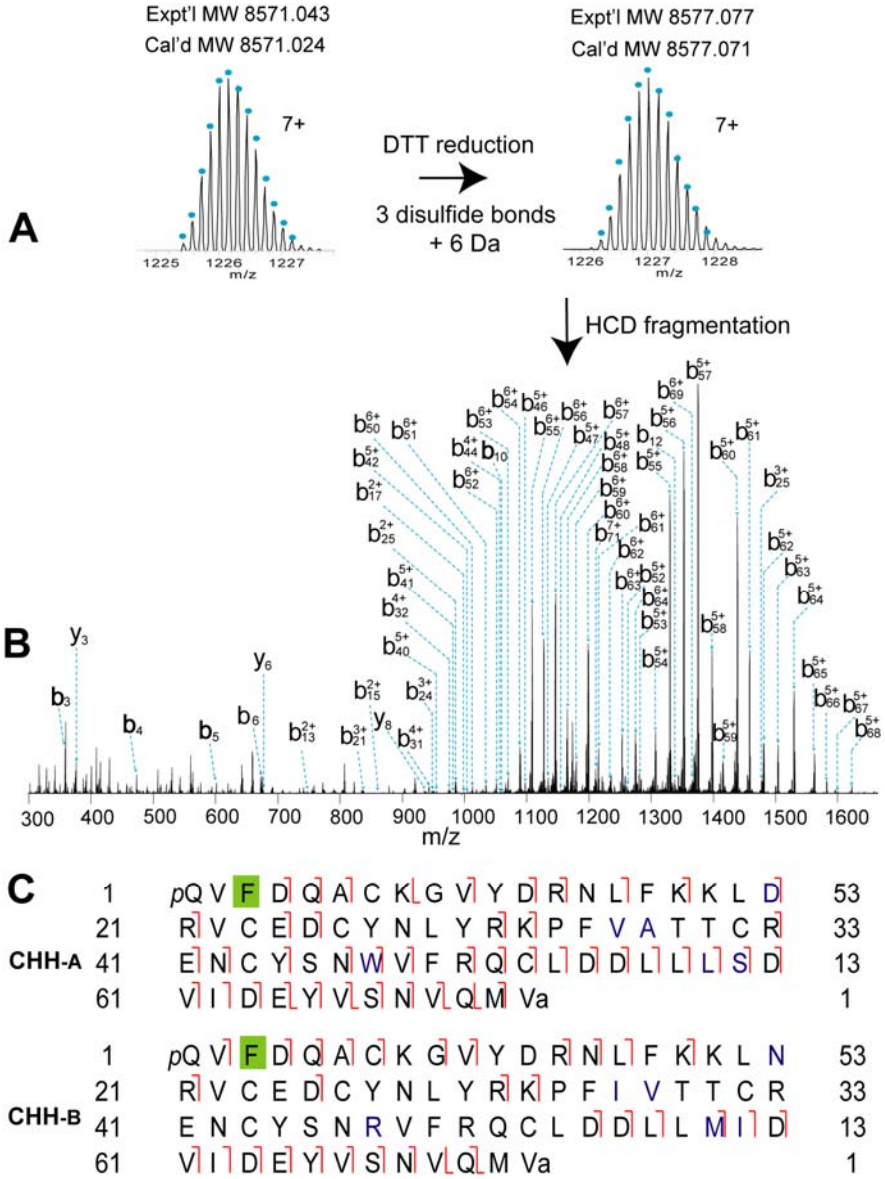


Figure 4. Identification of CHH$_{-A}$ and CHH$_{-B}$ peptides by top-down MS/MS. (A) Isotopic distributions of intact and DTT-reduced CHH$_{-A}$. (B) HCD MS/MS spectrum of DTT-reduced CHH$_{-A}$. (C) Top-down fragmentation maps of CHH$_{-A}$ and CHH$_{-B}$. ┓, b ions, ┗, y ions. The different residues between the two CHH peptides are highlighted in blue. The D-Phe residue is localized in the third position from the N-terminus.

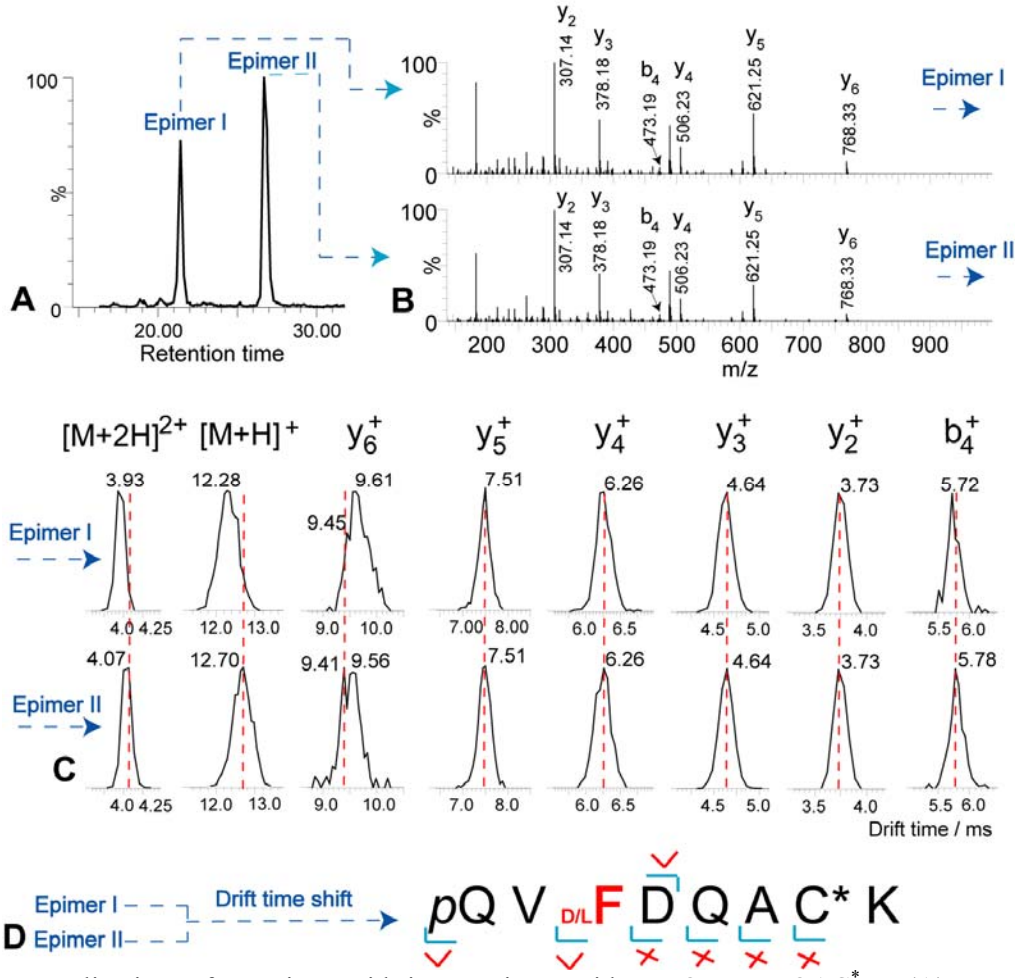


Figure 5. Localization of D-amino acid in tryptic peptides, *p*QV$_{D/L}$FDQAC*K. (A) Extracted ion chromatograph and (B) MS/MS of the tryptic peptide epimer I and II. (C) IMS distributions of fragment ions from tryptic peptide epimer I and II. (D) Localization of D-amino acid residue by comparison of drift time shift. √, Drift time shift. ×, No shift. It should be noted that the elution order of the two D/L-peptide epmiers cannot be determined by our current method, so we use Epimer I and II for annotation.

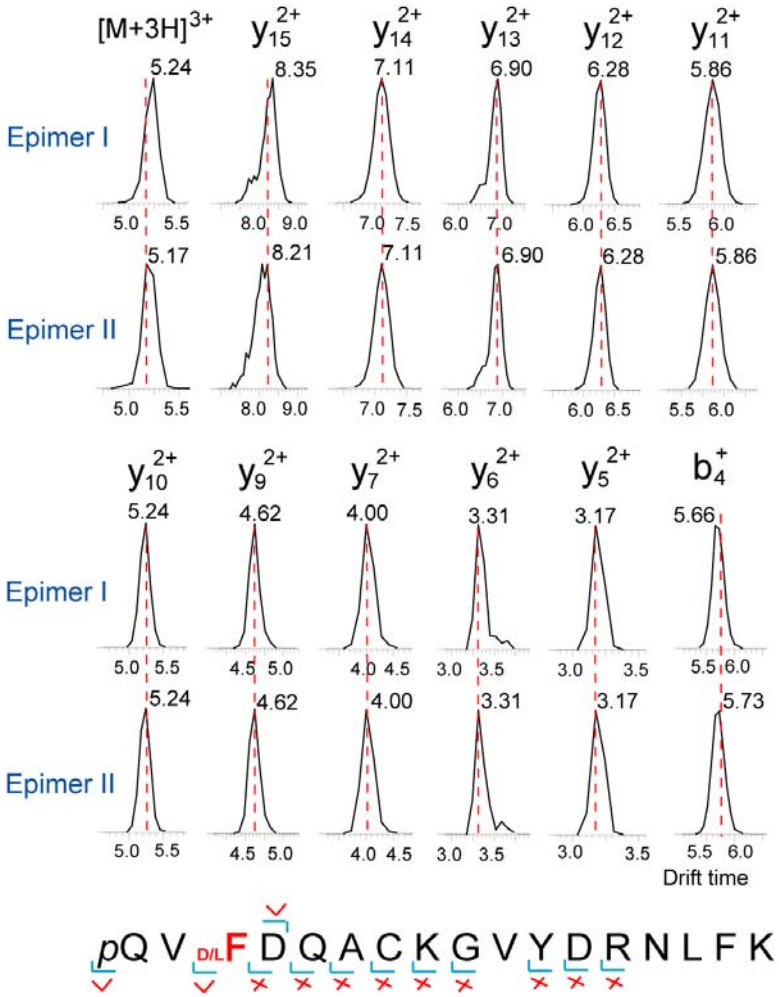


Figure 6. IMS distributions of fragment ions from tryptic peptide epimers, *p*QV$_{D/L}$FDQAC*KGVYDRNFLK. Tryptic peptide epimers I and II were eluted from RPLC at 33.21 and 35.70 min, respectively. √, Drift time shift. ×, No shift.

Supplementary Materials for Chapter 4

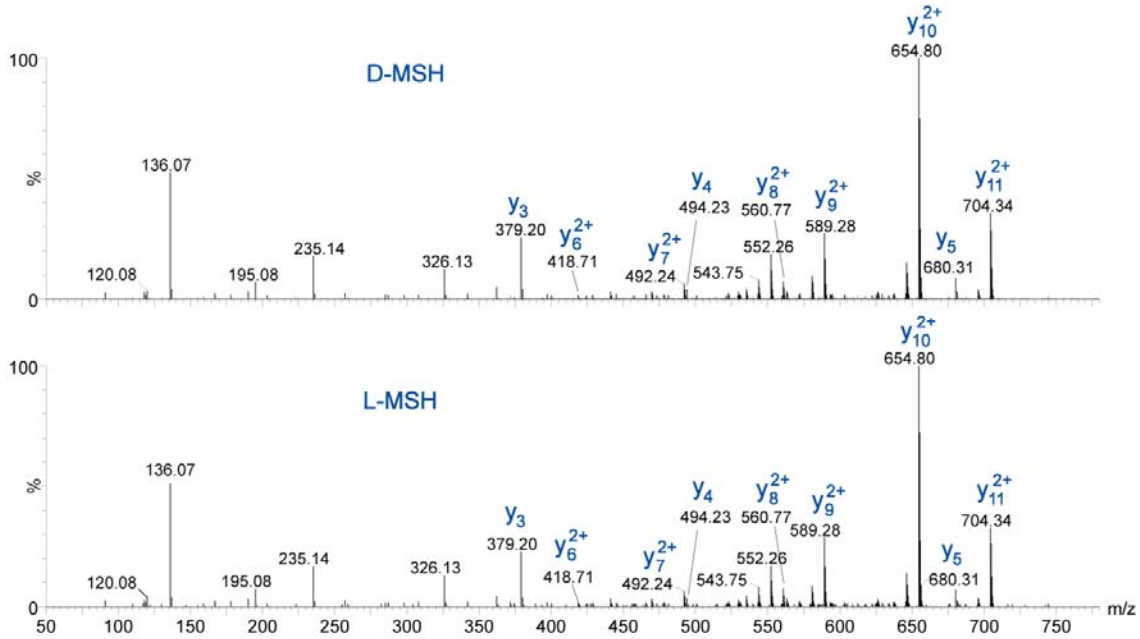


Figure S-1. CID MS/MS spectra of the D- and L-MSH peptides acquired from LC-MS/MS.

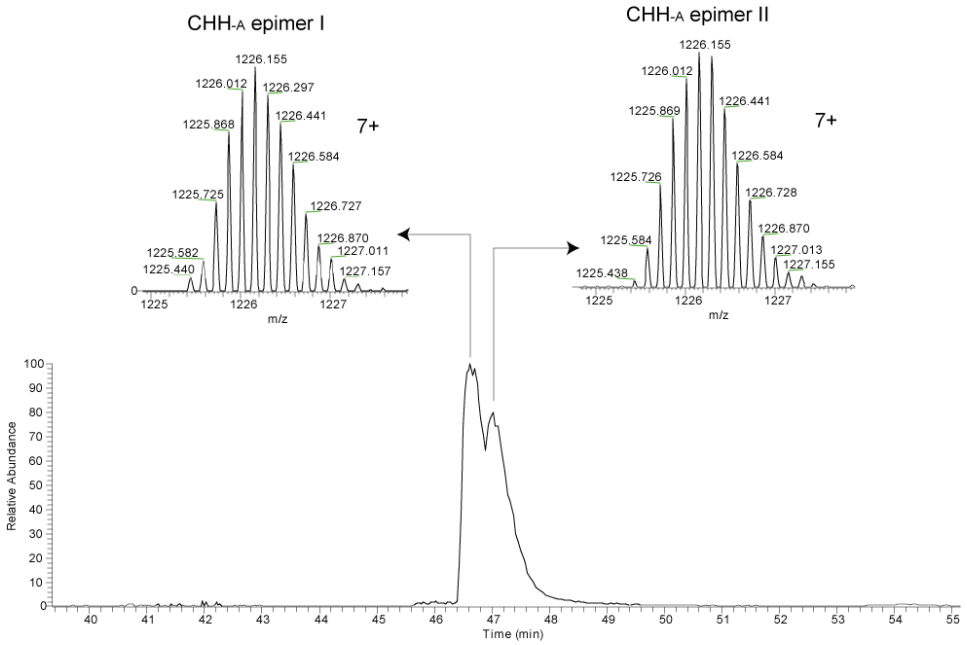


Figure S-2. Extract ion chromatograph of the LC-MS analysis of CHH_{-A} epimers I and II. The two inset panels show the isotopic distributions of the two epimers. They show the same mass, but are eluted at different retention times.

Table S1. Measured CCS_{He} of precursor and fragment ions of D/L-MSH.

	D-MSH	L-MSH
$[M+3H]^{3+}$	340.3 ± 0.4	347.2 ± 0.4
y_{11}^{2+}	294.1 ± 0.2	294.7 ± 0.1
y_{10}^{2+}	277.4 ± 0.2	277.2 ± 0.1
y_{9}^{2+}	260.2 ± 0.2	259.01± 0.05
y_{8}^{2+}	252.7 ± 0.2	252.91 ± 0.06
y_{7}^{2+}	236.75 ±0.06	237.5 ± 0.3
y_{6}^{2+}	216.7 ± 0.3	217.59± 0.03
y_{5}^{+}	173.5 ± 0.1	177.8 ± 0.3
y_{4}^{+}	143.70 ± 0.04	143.38 ±0.04
y_{3}^{+}	118.85 ± 0.03	118.79 ±0.03

Table S2. Tryptic peptides of CHHs identified by bottom-up sequencing approaches.[a]

CHH_{-A}

Tryptic peptides	Position	-10logP	mass	m/z	ppm	Elution time /min
*p*QVFDQAC*K	CHH_{-A}[1-8]	84.06	977.43	489.71	-23.7	21.42, 26.69
*p*QVFDQAC*KGVYDRNLFK	CHH_{-A}[1-17]	108.94	2069.99	690.99	-22.3	33.21,35.70
GVYDRNLFK	CHH_{-A}[9-17]	86.12	1110.58	556.28	-24.9	21.82
KLDRVC*EDC*YNLYR	CHH_{-A}[18-31]	124.97	1902.87	635.28	-27.2	20.12
VC*EDC*YNLYR	CHH_{-A}[22-31]	115.62	1390.56	696.27	-25.8	20.99
VC*EDC*YNLYRKPFVATTC*R	CHH_{-A}[22-40]	103.51	2451.11	613.77	-28.1	24.77
KPFVATTC*RENC*YSNWVFR	CHH_{-A}[32-50]	124.26	2434.13	609.52	-27.2	29.83
ENC*YSNWVFR	CHH_{-A} [41-50]	134.61	1373.58	687.78	-22.0	32.14

CHH_{-B}

Tryptic peptides	Position	-10logP	mass	m/z	ppm	Elution time /min
*p*QVFDQAC*K	CHH_{-B}[1-8]	84.06	977.43	489.71	-23.7	21.42, 26.69
*p*QVFDQAC*KGVYDRNLFK	CHH_{-A}[1-17]	108.94	2069.99	690.99	-22.3	33.21, 35.70
GVYDRNLFK	CHH_{-B} [9-17]	86.12	1110.58	556.28	-24.9	21.82
LNRVC*EDC*YNLYR	CHH_{-B}[19-31]	135.91	1773.79	592.25	-27.0	20.99
VC*EDC*YNLYRKPFIVTTC*R	CHH_{-B}[22-40]	115.93	2493.16	624.28	-26.0	27.94
KPFIVTTC*R	CHH_{-B}[32-40]	111.98	1120.60	561.29	-23.6	17.61
KPFIVTTC*RENC*YSNRVFR	CHH_{-B}[32-50]	35.36	2446.19	612.54	-29.8	22.48
ENC*YSNRVFR	CHH_{-B} [41-50]	45.18	1343.60	448.86	-27.8	16.78

[a] The two CHHs share parts of sequences, so some tryptic peptides are the same in the two tables.

Table S3. The ions used for collision cross section calibration.

Polyalanine Standards

n	mass (Da)	charge	m/z	Ω_{He} (Å^2)
11	799.42	2	400.72	197
12	870.46	2	436.24	208
13	941.49	2	471.75	220
14	1012.53	2	507.27	232
15	1083.57	2	542.79	243
16	1154.60	2	578.31	255
17	1225.64	2	613.83	265
18	1296.68	2	649.35	276
19	1367.72	2	684.87	287
20	1438.75	2	720.38	297
21	1509.79	2	755.90	308
22	1580.83	2	791.42	317
23	1651.86	2	826.94	327
24	1722.90	2	862.46	337
25	1793.94	2	897.98	348
26	1864.98	2	933.50	358
21*	1509.79	3	504.27	361
22	1580.83	3	527.95	373
23	1651.86	3	551.63	386
24	1722.90	3	575.31	399
25	1793.94	3	598.99	412
26	1864.98	3	622.67	425
27	1936.01	3	646.35	438
28	2007.05	3	670.02	452
29	2078.09	3	693.70	465
30	2149.12	3	717.38	479
31	2220.16	3	741.06	490
32	2291.20	3	764.74	502
33	2362.24	3	788.42	516

* The ion was detected at all wave velocities.

EXPERIMENTALS (Supplementary Materials)

Collision Cross Section Measurement

Gas-phase helium collision cross section values (CCS_{He}) for all of the ions described in the manuscript were measured on the Synapt G2 travelling-wave ion mobility mass spectrometer with nitrogen buffer gas. CCS_{He} calibration was performed using combined methods from Bush et al.[1] and Ruotolo et al.[2] to obtain accurate measurements. Polyalanine peptides (Sigma Aldrich) were dissolved in 49.5/49.5/1 water/ACN/formic acid at a concentration of 10 µg/mL. A separate calibration spectrum was acquired for each of the following wave velocities (m/s): 600, 700, and 800. The wave height was kept constant at 40 V. Following acquisitions, the log of the drift times (t_D) were plotted against the log of the reduced CCS todetermine the constants needed to calculate unknown Ω_{He}. The equations in the following paragraph were obtained from previously published sources.[2, 3] In travelling wave ion mobility, t_D and CCS_{He} (Ω_{He}) are non-linearly related by Equation 1:

$$\Omega_{He} = \frac{ze}{16}\left[\frac{18\pi}{k_b T}\left(\frac{1}{m}+\frac{1}{M_{He}}\right)\right]^{1/2}\frac{760}{P}\frac{T}{273.2}\frac{1}{NL}A t_D^{\,B} \qquad (1)$$

The variables z and e make up the charge of the analyte, T is the temperature of the drift gas, M_{He} is the mass of the helium drift gas, m is the mass of the analyte, P is the pressure inside the drift cell, N is the number density of drift gas molecules, L is the length of the drift cell, and k_b is Boltzmann's constant. A and B are constants that arise from the non-uniformity of the travelling-wave electric field and must be empirically determined by calibration.

The t_D for each calibration standard was converted to corrected drift time (t_D') by Equation 2 to account for the *m*/*z*-dependent travel time through the Synapt G2's ion optics:

$$t_D' = t_D - \left(\frac{C\sqrt{m/z}}{1000}\right) \qquad (2)$$

C is the delay constant set by the MS control software. The reduced collision cross section (Ω') was normalized for mass and charge contributions and calculated by Equation 3:

$$\Omega' = \frac{\Omega_{He}}{z\left(\frac{1}{m}+\frac{1}{M_{He}}\right)^{1/2}} \qquad (3)$$

Plotting the natural log of Ω' versus the natural log of t_D' yielded a linear best-fit line, the slope of which is B from Equation 1. From here, the doubly corrected drift time (t_D'') was calculated by Equation 4:

$$t_D'' = z\left(t_D'\right)^B\left(\frac{1}{m}+\frac{1}{M_{He}}\right)^{1/2} \qquad (4)$$

A final plot was constructed with t_D'' on the x-axis and Ω_{He} on the y-axis. The equation of the best-fit line was then used to calculate the Ω_{He} of the unknown peptides and peptide fragments.

References for Supplementary Materials

[1] Bush MF, Campuzano ID, Robinson CV. Ion mobility mass spectrometry of peptide ions: effects of drift gas and calibration strategies. Anal Chem. 2012;84:7124-30.

[2] Ruotolo BT, Benesch JL, Sandercock AM, Hyung SJ, Robinson CV. Ion mobility-mass spectrometry analysis of large protein complexes. Nat Protoc. 2008;3:1139-52.

[3] Smith DP, Knapman TW, Campuzano I, Malham RW, Berryman JT, Radford SE, et al. Deciphering drift time measurements from travelling wave ion mobility spectrometry-mass spectrometry studies. Eur J Mass Spectrom (Chichester, Eng). 2009;15:113-30.

Chapter 5

Gas-phase Ion Isomer Analysis reveals the Mechanism of Peptide Sequence Scrambling

Adapted from: Chenxi Jia, Zhe Wu, Christopher B. Lietz, Zhidan Liang, Qiang Cui, and Lingjun Li, Gas-phase Ion Isomer Analysis reveals the Mechanism of Peptide Sequence Scrambling. *Anal Chem*, under revision.

Abstract

Peptide sequence scrambling during mass spectrometry-based gas-phase fragmentation analysis causes misidentification of peptides and proteins. Thus, there is a need to develop an efficient approach to probing the gas-phase fragment ion isomers related to sequence scrambling and the underlying fragmentation mechanism, which will facilitate the development of bioinformatics algorithm for proteomics research. Herein, we report on the first use of electron transfer dissociation (ETD)-produced diagnostic fragment ions to probe the components of gas-phase peptide fragment ion isomers. In combination with ion mobility spectrometry (IMS) and formaldehyde labeling, this novel strategy enables qualitative and quantitative analysis of b-type fragment ion isomers. ETD fragmentation produced diagnostic fragment ions indicative of the precursor ion isomer components, and subsequent IMS analysis of b ion isomers provided their quantitative and structural information. The isomer components of three representative b ions (b_9, b_{10}, and b_{33} from three different peptides) were accurately profiled by this method. IMS analysis of the b_9 ion isomers exhibited dynamic conversion among these structures. Furthermore, molecular dynamics simulation predicted theoretical drift time values which were in good agreement with experimentally measured values. Our results strongly support the mechanism of peptide sequence scrambling via b ion cyclization, and provide the first experimental evidence to support that the conversion from molecular precursor ion to cyclic b ion ($M \rightarrow {}^{c}b$) pathway is less energetically (or kinetically) favored.

5.1 Introduction

Mass spectrometry (MS)–based proteomics has become an attractive technology for global analysis of protein composition, modifications, and dynamics [1, 2]. Tandem mass spectrometry (MS/MS) analysis of peptides is an essential tool for proteomic studies [2, 3]. Although bioinformatics algorithms provide necessary support to process the enormous quantity of MS/MS spectra generated from large scale analysis of biological samples, false positive identifications of peptides and proteins may exist [2]. One of the reasons causing such misidentifications stems from peptide sequence scrambling that could occur during tandem MS analysis of peptides via collision-induced dissociation (CID) [4].

Typically, CID causes the peptide bonds to be sequentially cleaved along the backbone to produce so-called "direct sequence ions" [4]. However, if a peptide chain is cyclized and rearranged in the gas phase, these resulting ion isomers will undergo further dissociation to produce "non-direct sequence ions" (Figure S1) leading to sequence scrambling [4, 5]. Interpreting these fragment ions without an efficient bioinformatics algorithm may lead to misidentification of peptides and proteins.

Recent studies [4] reported that head-to-tail cyclization of peptide b-type fragment ions under CID fragmentation caused sequence scrambling. Two fragmentation pathways (Figure S2) were proposed and studied [4-29]. In M→^{c}b pathway (M, molecular ion of peptides; ^{c}b, cyclic b ion; see Table S1 for special ion symbols), the primary amine at the N-terminus of the peptide chain reacts with the carbonyl carbon near the C-terminus of the peptide chain via nucleophilic attack to produce ^{c}b ion. The ^{c}b ion can also be produced via ^{ol}b→^{c}b pathway by head-to-tail cyclization of ^{ol}b ion (^{ol}b, original linear b ions). However, for analysis of polypeptides from complicated biological samples, there are many factors influencing the propensity for various peptides to

form ^{c}b ions thus leading to sequence scrambling. These factors include peptide length,[9, 26] residue acidity [12], proline and histidine effects [13, 21], and residue side chains [14], etc. Therefore, it is important to establish a method for accurate and fast identification of ion isomers, which allows for large-scale analysis of b ions to improve our understanding of the underlying fragmentation chemistry and facilitate development of bioinformatics algorithm.

Structural elucidation and profiling of ion isomers are challenging, because these gas-phase ion isomers share identical *m/z* values in MS measurement and they may undergo dynamic conversion which could be difficult to characterize. Previous studies have established a great variety of strategies for analysis of ion isomers, such as infrared spectroscopy [9, 15, 27], gas-phase H/D exchange [15, 24], guest-host chemistry [10], etc, which provided useful insights into underlying mechanism of peptide sequence scrambling. However, these strategies lack the capability to elucidate the precise components of various ion isomers in a complex mixture. To fill in this gap, herein we employ novel use of electron transfer dissociation (ETD), an electron-based fragmentation technique, to probe the gas-phase composition and structures of fragment ion isomers. In combination with ion mobility spectrometry (IMS) and formaldehyde (FH) labeling, this novel strategy enables simultaneous qualitative and quantitative analysis of various b ion isomers including linear and cyclic b ions such as ^{ol}b, ^{c}b, and ^{rl}b ions (^{rl}b, rearranged linear b ions).

5.2 Experimentals

5.2.1 Materials and Chemicals

All chemical reagents were obtained from Sigma-Aldrich (St. Louis, MO) unless otherwise noted. Optima grade formic acid, acetonitrile (ACN) and water were purchased from Fisher Scientific (Pittsburgh, PA). Peptide standards, neurokinin and Substance P were purchased from

American Peptide Co. (Sunnyvale, CA). The crustacean hyperglycemic hormone precursor related peptide (CPRP) was isolated and purified from sinus glands of blue crabs *Callinectes sapidus* as described in our previous report [30].

5.2.2 Formaldehyde Labeling of Peptides

One microgram of peptide sample was labeled in 10 μL of water solution by adding 1 μL of borane pyridine (C_5H_8BN, 120 mM in 10% methanol) and then mixing with formaldehyde (15% in H_2O, 1 μL). The reaction mixture was then vortexed at room temperature for 15 min and quenched with ammonium bicarbonate solution (1 μL, 0.2 M). After drying down in Speedvac, the sample was desalted by Ziptip for direct infusion analysis on mass spectrometers.

5.2.3 Fragmentation of b Ions by ETD

The ETD experiments were performed on an amaZon ETD ion trap mass spectrometer (Bruker Daltonics, Bremen, Germany) equipped with a CaptiveSpray ESI source. Optimization of the CaptiveSpray source resulted in dry gas temperature, 130 °C, dry gas, 6.0 L/min, capillary voltage, −1250 V, end plate offset, −500 V. The peptide sample was dissolved in ACN-water-formic acid (50:50:0.1) at a concentration of 100 ng/mL and was directly infused into mass spectrometer at 300 nL/min. The peptide molecular ions were selected with a 4 Da isolation window in MS^2 and then fragmented by collision induced dissociation (CID) with an adjusted fragmentation amplitude (0.7-1.2) that allows the parent ions to be completely fragmented. In MS^3 the CID-produced b ions were further selected with a 2 Da isolation window and fragmented with ETD. The ion charge control (ICC) target was set to 200 000, maximum accumulation time, 120 ms, 10 spectral averages, 5 rolling averaging (a filter that operates on the time series of mass spectra that are generated by the ion trap), and acquisition range of *m/z* 100–1800. The ETD reagent parameters were set to 400 000 ICC target for the fluoranthene radical

anion and the ETD duration time was 100 ms. Data were acquired in enhanced resolution mode with ~3000 of resolving power for 1 hour and processed using software DataAnalysis to generate a high quality MS^3 ETD spectrum. The fragment ions were assigned within 0.1 Da of mass error.

5.2.4 Ion Mobility Analysis

The IMS experiments were performed using a Synapt G2 HDMS mass spectrometer equipped with a nano-ESI ion source and MassLynx data processor (Waters, Milford, MA, USA). Instrument acquisition parameters used were as follows: an inlet capillary voltage of 3.0 kV, a sampling cone setting of 35 V, and a source temperature of 120 °C. The argon gas pressure in the traveling wave ion guide trap and the traveling wave ion guide transfer cell were 2.44×10^{-2} and 2.61×10^{-2} mbar, respectively. The wave height, the wave velocity, and the nitrogen pressure in the traveling wave ion mobility drift cell were 32.0 V, 800 m/s, and 2.96 mbar, respectively. Samples were directly infused into the mass spectrometer at a rate of 0.5–0.8 μL/min. The peptide molecular ions were selected and fragmented in Traveling-Wave trap cell by CID with an adjusted collision energy 25-30 eV that allows the parent ions to be completely fragmented. All the fragment ions were submitted into drift tube with their drift time measured. The spectra were acquired for 5 min. Data processing was conducted using Waters MassLynx 4.1 and DriftScope 2.1. The experimental procedures for collision cross section measurement are described in Supporting Information.

The dynamics study was conducted as shown in the schematic of Figure 3A. The peptides were fragmented by in-source dissociation (ISD) in nozzle-skimmer region. To result in ISD, the adjusted source parameters were used: an inlet capillary voltage of 3.5 kV, a sampling cone setting of 55 V, and a source temperature of 120 °C. The target b ions were selected by quadrupole and accumulated in trap cell, where various collision energies 0-22 eV were used to

activate the b ions. Subsequently, the b ions were submitted into drift tube with the same parameters as described above.

5.2.5 High-resolution CID MS/MS of CPRP Neuropeptide

The high-resolution CID MS/MS of CPRP peptide was carried out on a 7T linear trap quadrupole (LTQ)/Fourier transform ion cyclotron resonance (FTICR) (LTQ-FT Ultra) hybrid mass spectrometer (Thermo Scientific Inc., Bremen, Germany) as previous report.[31] The experimental details are described in Supporting Information.

5.2.6 Molecular Dynamics (MD) Simulation

MD simulations were performed using the program CHARMM (c37) with the CHARMM22 force field for proteins [32]. Parameters were generated with the CHARMM General Force field (CGenFF) program [33] for the oxazolone structure. MD simulations were performed in vacuum at 450 K, which mimics the low gas density environment in the drift tube with enhanced molecular vibrations due to N_2 collision. Standard non-bonded cut-off scheme and temperature-coupling protocol in CHARMM were adopted. For each molecule, a 30 ns MD trajectory was calculated and the last 15 ns was used to generate structures for the collision cross section (Ω_{He}) calculations. For each b ion isomer in Figures S3 and S4 (protonation sites listed in Table S2), 1500 structures were collected to compute the Ω_{He} values using the Sigma program [34-36] with a scaled Leonard-Jones projection protocol. Finally, the Ω_{He} values were converted into drift times to generate distributions of drift times for the b ion isomers. Nonlinear fitting was carried out using Gaussian functions. It should be noted that the Synapt G2 HDMS used 0.069 ms of bin size as default setting. To be consistent, the same bin size was used in theoretical simulation.

5.3 Results and Discussion

Figure 1 describes the workflow of the proposed strategy for fragment ion isomer analysis. The CID-produced b ion isomers are subjected to ETD fragmentation and IMS analysis, respectively. The formaldehyde (FH)-labeled peptides are fragmented by CID to generate FH-labeled linear b ion, followed by ETD fragmentation and IMS analysis, respectively. On the left panel of Figure 1, under ETD fragmentation each b ion isomer generates unique diagnostic fragment ions. The FH-labeled b ion (dimethylation on N-terminus and lysine side chain) [37] is used for alignment. On the right panel, each b ion isomer is separated in IMS, followed by alignment with FH-labeled b ions.

ETD fragmentation is based on gas-phase ion/ion reactions, where an anion transfers one electron to the protonated peptide cation to initiate radical-driven backbone cleavage for peptide fragmentation [38, 39]. Unlike CID which relies on "slow heating" threshold fragmentation, ETD occurs in a much shorter time scale than that is required for internal energy distribution, thus resulting in rapid fragmentation and reducing the propensity of b ion cyclization [38]. Previous studies [40, 41] have demonstrated that ETD or electron capture dissociation (ECD) of cyclic peptides can induce a free radical reaction cascade generating [z-c] and [z-a] ions. Because the ^{c}b ion has the same structure as a protonated cyclic peptide, ETD fragmentation of ^{c}b ion produces the same diagnostic ions [z-c] and [z-a] as well. In contrast, ETD fragmentation of ^{l}b ions generates diagnostic [z-b] and c ions. This distinction provides the key molecular signatures to differentiate ^{c}b from ^{l}b ions. Additionally, ^{ol}b and ^{rl}b ions can be distinguished by assigning sequential c and [z-b] ions. Figure S5 illustrates the structures of these diagnostic ions. Recently, Li et al. [42] reported observation of these diagnostic fragment ions from ^{c}b and ^{l}b ions using ECD. Herein, we extend this concept to the development of a novel diagnostic analysis of ion isomers using ETD, as ETD may minimize internal energy distribution and thus reduce

secondary fragmentation [38]. IMS analysis offers accurate profiling of these b ions by measuring drift time differences of isobaric isomer species due to their different shapes and conformations in the gas phase. Alignment of unlabeled b ion with $_{FH}$-b (formaldehyde-labeled b ion) provides complementary information to ETD and IMS analyses. To demonstrate this strategy, we analyzed three representative b ions from three peptides with diverse isomer components, i.e., neurokinin b_9 ion, substance P b_{10} ion, and CPRP b_{33} ion.

5.3.1 Neurokinin b_9 ion

The ETD experiments were performed on an ion trap Bruker amaZon with electrospray ionization (ESI). The $[M+2H]^{2+}$ ion of neurokinin (HKTDSFVGLM-NH_2) was selected in MS^2 and fragmented using CID to generate the b_9^{2+} ion, followed by ETD fragmentation in MS^3. In the resulting ETD spectrum (Figure 2A), fragment ions from four different b ion isomers were detected, i.e., c_2~c_7 from $^{ol}b_9^{2+}$ (HKTDSFVGL^{2+}), $[z\text{-}b]_{A3}$~$[z\text{-}b]_{A8}$ from $^{rl}b_{A9}^{2+}$ (TDSFVGLHK^{2+}), $[z\text{-}b]_{B5}$~$[z\text{-}b]_{B8}$ from $^{rl}b_{B9}^{2+}$ (SFVGLHKTD^{2+}), as well as $[z\text{-}c]_3$~$[z\text{-}c]_7$ and $[z\text{-}a]_3$~$[z\text{-}a]_7$ from $^{c}b_9^{2+}$ (cyclo-HKTDSFVGL^{2+}) (See Table S1 for special ion symbols). In contrast, the N-terminus and lysine side chain were dimethylated with formaldehyde [37], which blocks cyclization reaction of b ions (mechanisms shown in Figure S6). Thus, the $_{FH}$-b_9^{2+} ion produced by CID should maintain linear structure. As expected, the ETD spectrum of $_{FH}$-b_9 in MS^3 (Figure 2C) only contains c and [z-b] ions from $^{l}b_9$, confirming that FH labeling blocks the formation of cyclic b ion. Subsequently, the IMS experiments were performed on an ESI-QTOF Waters Synapt G2 HDMS. The CID-produced b_9^{2+} ion of neurokinin was submitted into IMS drift tube with drift time measured. The resulting IMS distribution (Figure 2B) shows four peaks at 4.02, 4.29, 4.43 and 4.78 ms respectively corresponding to four ion isomers. Their collision cross sections Ω_{He} were 229, 237, 244 and 256 Å^2, which were converted from Ω_{N2} according to the

method reported by Bush et al. [43] (Table S3 lists all Ω_{He} determined in this study). This was in agreement with the four isomers determined by ETD method (Figure 2A). It should be noted that the isomer peak at 4.29 ms was further resolved by the subsequent dynamic study. In contrast, IMS analysis of $_{FH}$-b_9^{2+} (Figure 2D) displayed only one peak at 4.55 ms (248 Å^2), which was also consistent with the ETD data (Figure 2C). Here, we tentatively assign the peak at 4.02 ms in Figure 2B as $^{c}b_9^{2+}$, since cyclic structure is more compact than the linear ones and has the highest mobility and thus shortest drift time. The peak at 4.43 ms is tentatively assigned as the $^{ol}b_9^{2+}$ ion, as the drift time of $_{FH}$-b_9^{2+} at 4.55 ms is the closest to the peak at 4.43 ms. This argument can be supported by the fact that the $[M+2H]^{2+}$ and $_{FH}$-$[M+2H]^{2+}$ of neurokinin (Figure S7) also show similar drift times under the same IMS conditions. In addition, IMS can resolve conformational isomers as well as structural isomers [44]. The IMS distribution of intact peptide of neurokinin (Figure S7A) showed one major peak corresponding to one IMS-resolved conformation, so more possibly the four peaks in IMS distribution of b_9^{2+} (Figure 2B) were different structural isomers instead of different conformational isomers.

To study the dynamics of structural conversion among these isomers under CID activation, we designed the MS^3-IMS experiment (Figure 3A) on Synapt G2. The b_9^{2+} ion isomers were generated by in-source dissociation (ISD) of neurokinin $[M+2H]^{2+}$ ion in nozzle-skimmer region (MS^2), isolated by quadrupole, and then was further activated by collision energy in the Triwave® trap cell (MS^3). After that, all the ion isomers were submitted into drift tube to monitor their conversion dynamics. The benefit of this design is that tandem MS in the trap cell can be optimized to directly activate the target fragment ions, instead of precursor ions, thus improving the capability to monitor the conversion dynamics of these ion isomers. Figure 3B shows the IMS distributions of the b_9^{2+} isomers after activation by various collision energy (CE) levels.

With elevation of CE from 0 to 22 eV, the relative peak abundances at 4.02 and 4.43 ms were reduced; and those at 4.29 and 4.78 ms exhibited consistent elevation. This trend revealed the conversion dynamics of these ion isomers (depicted in Figure S3), with the peak at 4.02 ms assigned to $^{c}b_9^{2+}$; 4.43 ms attributed to $^{ol}b_9^{2+}$; and 4.29 and 4.78 ms assigned to the two $^{rl}b_9^{2+}$ ions.

To further confirm the above assignments, we conducted the MS^3-IMS-MS^4 experiment (Figure 3A). All the ions from the IMS drift tube were fragmented in the Triwave® transfer cell (MS^4) by CID with varied CE in the trap cell. This experiment allowed the mobility-separated $^{c}b_9^{2+}$ and $^{ol}b_9^{2+}$ ions to be fragmented at successive time points. Ideally, CID of $^{c}b_9^{2+}$ should produce higher abundance of rearranged fragment ions than that of $^{ol}b_9^{2+}$. As expected, we observed different intensity ratio of b_8 and $^{r}b_8$ (1:3.7 and 1:2) in the two resulting CID spectra (Figure 3C and D). The b_8 ion arose directly from $^{ol}b_9^{2+}$, and the rearranged fragment ion $^{r}b_8$ was produced from $^{c}b_9^{2+}$ via pathway $^{c}b_9^{2+} \rightarrow {}^{r}b_9^{2+} \rightarrow {}^{r}b_8$ (details shown in Figure S8). This result provided the third piece of evidence to confirm the assignment of $^{ol}b_9^{2+}$ and $^{c}b_9^{2+}$ in Figure 3B. Although the results from ETD and the conversion dynamics studies indicate that the peaks at 4.29 and 4.78 ms were the two rearranged ions $^{rl}b_{A9}^{2+}$ and $^{rl}b_{B9}^{2+}$, the order of their mobility in the IMS drift tube could not be determined.

To further assign the four ion isomers, we calculated their Ω_{He} distributions by molecular dynamics simulation [45, 46]. In practice, the ISD-produced b ion isomers were calibrated by polyalanine under the MS^3-IMS experimental condition. Figure S9 displays the measured Ω_{He} distributions. It should be noted that the individual drift times of the four isomers remained consistent under CE 0~22 eV, assuring that the experimental CE in the trap cell used here would not cause remarkable drift time shifts. For theoretical simulation [47], the structures (Figure S3

and Table S2) of the four b ion isomers were sampled by molecular dynamics, and then the Ω_{He} values were calculated using the Sigma program [35] with a scaled Leonard-Jones projection approximation procedure (see Supporting Information). To make a direct comparison with Figure 3B, the theoretical Ω_{He} values were converted to drift times, followed by statistical processing to generate simulated distributions in Figure 3E. The theoretical and experimental drift time values agree with each other, supporting the assignment of the four ion isomers. Collectively, the agreement of the results from ETD fragmentation, IMS measurement, and theoretical simulation support the proposed mechanism of b ion cyclization outlined in Figure S3. In addition to the standard ^{rl}b ions containing oxazolone structures at the C-terminus, two non-standard ^{rl}b ions could be produced respectively via two alternative ring opening pathways due to nucleophilic effects of the side chains of Lys and Asp (Figure S4) [48]. The two non-standard ^{rl}b ions, termed as $^{rl}b_{A9,}^{2+}$ and $^{rl}b_{B9,}^{2+}$, respectively contain the same amino acid sequences as $^{rl}b_{A9}^{2+}$ and $^{rl}b_{B9}^{2+}$, but different C-terminal structures (Figure S4). Figure S10 shows the theoretical drift time distributions of $^{rl}b_{A9,}^{2+}$ and $^{rl}b_{B9,}^{2+}$ centered at 4.32 and 4.64 ms, which agrees with the measured drift times of $^{rl}b_{A9}^{2+}$ and $^{rl}b_{B9}^{2+}$ at 4.29 and 4.78 ms (Figure 3B). Although our experimental results and theoretical calculations support the assignment of the four ion isomers in Figure 3B, further studies using other technologies, such as infrared photodissociation spectroscopy [49], gas-phase hydrogen/deuterium exchange [15], etc could make the ion assignments more confident. In addition, it was previously reported that the primary amine on the lysine side chain in a peptide may attack the C-terminus of the peptide chain via nucleophilic reaction to produce a macrocycle structure [50]. In the ETD spectrum (Figure 2A) of b_9^{2+}, we did not observe any fragment ions from this macrocycle structure, indicating that this side reaction may occur via a minor pathway (or did not happen under the experimental conditions used in this study).

5.3.2 Substance P b_{10} ion

Figure 2E is the ETD fragmentation of the b_{10}^{2+} ion from substance P (RPKPQQFFGLM-NH_2). Interestingly, one set of [z-b] ions was clearly detected, which arose from the $^{rl}b_{10}^{2+}$ ion ($PQQFFGLRPK^{2+}$). Consistently, only one peak at 4.90 ms (260 $Å^2$) was present in the corresponding IMS distribution of this b_{10}^{2+} ion (Figure 2F). Thus, this IMS peak can be assigned as $^{rl}b_{10}^{2+}$. Furthermore, ETD fragmentation of the $_{FH}$-b_{10}^{2+} ion (Figure 2G) produced a set of c ions with original sequence; and accordingly the IMS distribution (Figure 2H) showed one peak at 5.74 ms (278 $Å^2$). Aligning the two IMS peaks of $^{rl}b_{10}^{2+}$ and $_{FH}$-b_{10}^{2+} revealed a large drift time shift of 0.84 ms, in contrast to small deviation often observed for original linear b-ion and their formaldehyde labeled pair. Therefore, we conclude that the peak at 4.90 ms can be assigned as the rearranged linear b-ion $^{rl}b_{10}^{2+}$.

5.3.3 CPRP b_{33} ion

Tirado and Polfer [9] recently developed an infrared multiple photon dissociation-based strategy which allows probing the components of b ions containing up to 12 amino acids. To the best of our knowledge, the b_{12} ion is the largest b ion whose component is accurately analyzed so far. Thus, examination of a b-type ion with 33 amino acid residues using our strategy may provide useful insight into peptide scrambling of large b ions. We investigated the b_{33}^{4+} ion of CPRP (RSAEGLGRMGRLLASLKSDTVTPLRGFEGETGHPLE), a neuropeptide isolated from the sinus gland of blue crab *Callinectes sapidus* [30]. Figure 4A shows the CID spectrum (MS^2) of $[M+5H]^{5+}$ from CPRP on a high resolution instrument LTQ-FTICR, giving rise to abundant b_{33}^{4+} ion. CID fragmentation of b_{33}^{4+} in MS^3 (Figure 4B) produced intense ^{r}b ions (ion nomenclature proposed by Chawner et al. [51]) leading to sequence scrambling. The proposed pathway of b_{33}

cyclization and re-arrangement is shown in Figure S12. Interestingly, two ^{r}b ions, $[b_{33}19]b_{29}^{3+}$ and $[b_{33}19]b_{14}^{+}$ with very low intensities were observed in the MS^2 spectrum. This suggests that peptide sequence scrambling may be a minor problem in top-down MS^2, while CID fragmentation of large b ions in MS^3 can cause significant sequence scrambling, leading to misidentification of large peptides.

To investigate the isomer component of this b_{33}^{4+} ion, ETD, FH labeling and IMS experiments were performed. Surprisingly, ETD fragmentation patterns (Figures 2I and 2K) of b_{33}^{4+} and $_{FH}$-b_{33}^{4+} were remarkably consistent, where two sets of c and [z-b] ions from the original sequence were observed in each ETD spectrum. IMS distributions (Figures 2J and 2L) of b_{33}^{4+} and $_{FH}$-b_{33}^{4+} agree with each other as well, where the two peaks (6.14 and 6.00 ms) has 0.14 ms of shift due to N-terminal dimethylation. These results clearly indicate that the IMS peak at 6.14 ms is $^{ol}b_{33}^{4+}$. In other words, CID-produced b_{33}^{4+} in MS^2 maintains the original linear structure without rearrangement.

Previous studies [4, 5] proposed two pathways regarding formation of ^{c}b ions, M→^{c}b and M→^{ol}b→^{c}b (Figure S2). Harrison, Paizs and co-workers [4] suggested that the former pathway is less energetically favored via theoretical calculation. So far, there has been no experimental evidence to support this theoretical prediction, because all of the b ions under study undergo cyclization in MS^2. Therefore, a critical experiment is to find the b ions in MS^2 displaying original linear structures as the major molecular species. Interestingly, our results indicated that b_{33}^{4+} maintains the original linear sequence via CID fragmentation of $[M+5H]^{5+}$, suggesting that majority of b_{33}^{4+} ion did not undergo cyclization in MS^2. This finding provided the first experimental evidence to support the calculation result [4] that the M→^{c}b pathway is indeed less energetically favored.

Our strategy enables qualitative and quantitative analysis of b ion isomers. Table 1 summarizes the analysis results of the three representative b ions, including the ion isomer components and their corresponding ratios in relative abundances. The goal of this work is to improve our understanding of the underlying fragmentation chemistry of peptide sequence scrambling and to faciliate the development of bioinformatics algrithm for proteomics research. Although our strategy is demonstrated in the study of multiply charged b ions, the general fragmentation mechanism and identification rules are applicable to b ions at all charge states. To apply our strategy to large-scale analysis of b ion components, we may choose an alternative enzyme such as Lys-N [52] to digest protein mixture and generate multiply charged peptide ions containing basic amino acid residues at the N-terminus, which produces high abundance of multiply charged b ions. These b ions will be subsequently analyzed by ETD and IMS, followed by computer-assisted data processing. The qualitative and quantitative information of b ion isomer composition can be extracted from the resulting ETD spectra and IMS distributions, respectively. In this study, the IMS and ETD experiments were respectively performed on two different mass spectrometers, since unfortunately the Synapt G2 mass spectrometer does not offer the capability of ETD fragmentation. If the Synapt G2 is upgraded by installation of ETD source [53], the peptide can be fragmented by in-source dissociation and the resulting b ions can be transferred to Triwave trap cell for ETD fragmentation. This experiment design allows performing IMS and ETD in the same instrument.

In addition, Li et al. [42] previously applied ECD to fragment b ions and study cyclization of b ions. However, ECD-produced fragment ions were not utilized as a diagnostic tool for b ion isomer component analysis. Due to relatively higher energy level applied to these ions,

ECD may cause secondary fragmentation, which may not provide accurate information to allow characterizing the components of b ions. Moreover, compared with ECD technique realized on FTICR instruments, mass spectrometers equipped with ETD technique often provide much faster scan rate, which may improve the throughput of analysis [38]. In addition, from the ETD spectra of the three $_{FH}$-b ions (Figure 2C, G and K), only the fragment ions from linear structures were observed; and IMS distrbutions of the three $_{FH}$-b ions (Figure 2D, H and L) only displayed as single peaks. These experimental evidence supports that *N*,*N*-dimethylation with formaldehyde can fully prevent the cyclization of b ions, therefore producing greatly simplified fragmentation due to elimination of sequence scrambling. Fu and Li [54] previously reported that the CID spectra of FH-labeled peptides were simplified due to the formation of enhanced N-terminal fragment ions and suppressed internal fragments. Our current work provided strong experimental support that such fragmentation simplification and improved *de novo* sequencing is mainly due to the blockage of cyclic b-ion formation via *N*,*N*-dimethylation.

5.4 Conclusions

In comparison with previous studies [9, 10, 15], our method has several advantages: (1) all the isomers (^{ol}b, ^{c}b, and ^{rl}b) can be simultaneously probed; (2) the experiments can be conducted using commercially available mass spectrometers; and (3) the method is applicable for large-scale screening. In this study, ion isomer analysis of the three representative b ions has demonstrated the utility of a novel strategy combining experimental analysis via ETD, IMS and chemical labeling and theoretical calculation using molecular dynamic simulation. The detailed analyses of gas-phase ion isomer components and their

dynamic conversions reveal the proposed fragmentation mechanism of b ion cyclization and their potential impact on large-scale proteomic studies.

5.5 Acknowlegements

This work was funded by the National Institutes of Health (R01DK071801 to LL) and National Science Foundation (CHE-0957784 to LL, CHE-0957285 to QC, and CHE-0840494 to phoenix cluster). LL acknowledges an HI Romnes Faculty Research Fellowship, and CBL thanks an NIH-CBIT fellowship (T32-GM008505) and a NSF graduate research fellowship. We thank Prof. Cheng Lin for his helpful discussions about fragmentation chemistry. We are also grateful to the UW SoP AIC and UW HPP for access to mass spectrometers.

5.6 References

[1] Walther TC, Mann M. Mass spectrometry-based proteomics in cell biology. J Cell Biol. 2010;190:491-500.

[2] Aebersold R, Mann M. Mass spectrometry-based proteomics. Nature. 2003;422:198-207.

[3] Wysocki VH, Resing KA, Zhang Q, Cheng G. Mass spectrometry of peptides and proteins. Methods. 2005;35:211-22.

[4] Harrison AG, Young AB, Bleiholder C, Suhai S, Paizs B. Scrambling of sequence information in collision-induced dissociation of peptides. J Am Chem Soc. 2006;128:10364-5.

[5] Yague J, Paradela A, Ramos M, Ogueta S, Marina A, Barahona F, et al. Peptide rearrangement during quadrupole ion trap fragmentation: added complexity to MS/MS spectra. Anal Chem. 2003;75:1524-35.

[6] Jia C, Qi W, He Z. Cyclization reaction of peptide fragment ions during multistage collisionally activated decomposition: an inducement to lose internal amino-acid residues. J Am Soc Mass Spectrom. 2007;18:663-78.

[7] Bleiholder C, Osburn S, Williams TD, Suhai S, Van Stipdonk M, Harrison AG, et al. Sequence-scrambling fragmentation pathways of protonated peptides. J Am Chem Soc. 2008;130:17774-89.

[8] Harrison AG. To b or not to b: the ongoing saga of peptide b ions. Mass Spectrom Rev. 2009;28:640-54.

[9] Tirado M, Polfer NC. Defying entropy: forming large head-to-tail macrocycles in the gas phase. Angew Chem Int Ed Engl. 2012;51:6436-8.

[10] Somogyi A, Harrison AG, Paizs B. Using Gas-Phase Guest-Host Chemistry to Probe the Structures of b Ions of Peptides. J Am Soc Mass Spectrom. 2012;23:2055-8.

[11] Polfer NC, Bohrer BC, Plasencia MD, Paizs B, Clemmer DE. On the dynamics of fragment isomerization in collision-induced dissociation of peptides. J Phys Chem A. 2008;112:1286-93.

[12] Atik AE, Yalcin T. A systematic study of acidic peptides for b-type sequence scrambling. J Am Soc Mass Spectrom. 2011;22:38-48.

[13] Bythell BJ, Knapp-Mohammady M, Paizs B, Harrison AG. Effect of the His residue on the cyclization of b ions. J Am Soc Mass Spectrom. 2010;21:1352-63.

[14] Molesworth S, Osburn S, Van Stipdonk M. Influence of amino acid side chains on apparent selective opening of cyclic b5 ions. J Am Soc Mass Spectrom. 2010;21:1028-36.

[15] Chen X, Yu L, Steill JD, Oomens J, Polfer NC. Effect of peptide fragment size on the propensity of cyclization in collision-induced dissociation: oligoglycine b(2)-b(8). J Am Chem Soc. 2009;131:18272-82.

[16] Saminathan IS, Wang XS, Guo Y, Krakovska O, Voisin S, Hopkinson AC, et al. The extent and effects of peptide sequence scrambling via formation of macrocyclic B ions in model proteins. J Am Soc Mass Spectrom. 2010;21:2085-94.

[17] Goloborodko AA, Gorshkov MV, Good DM, Zubarev RA. Sequence scrambling in shotgun proteomics is negligible. J Am Soc Mass Spectrom. 2011;22:1121-4.

[18] Durand S, Rossa M, Hernandez O, Paizs B, Maitre P. IR Spectroscopy of b4 Fragment Ions of Protonated Pentapeptides in the X-H (X = C, N, O) Region. J Phys Chem A. 2013;117:2508-16.

[19] Dong NP, Liang YZ, Yi LZ. Investigation of scrambled ions in tandem mass spectra. Part 1. Statistical characterization. J Am Soc Mass Spectrom. 2012;23:1209-20.

[20] Tasoglu C, Gorgulu G, Yalcin T. Investigation of peptide size, residue position, neighbor amino acid and side chain effect on macrocyclization of b(n) (n=5-7) ions. Int J Mass Spectrom. 2012;316:108-16.

[21] Miladi M, Zekavat B, Munisamy SM, Solouki T. A systematic study on the effect of histidine position and fragment ion size on the formation of b(n) ions. Int J Mass Spectrom. 2012;316:164-73.

[22] Chen X, Tirado M, Steill JD, Oomens J, Polfer NC. Cyclic peptide as reference system for b ion structural analysis in the gas phase. J Mass Spectrom. 2011;46:1011-5.

[23] Molesworth SP, Van Stipdonk MJ. Apparent inhibition by arginine of macrocyclic b ion formation from singly charged protonated peptides. J Am Soc Mass Spectrom. 2010;21:1322-8.

[24] Fattahi A, Zekavat B, Solouki T. H/D exchange kinetics: experimental evidence for formation of different b fragment ion conformers/isomers during the gas-phase peptide sequencing. J Am Soc Mass Spectrom. 2010;21:358-69.

[25] Harrison AG. Cyclization of peptide b9 ions. J Am Soc Mass Spectrom. 2009;20:2248-53.

[26] Molesworth S, Osburn S, Van Stipdonk M. Influence of size on apparent scrambling of sequence during CID of b-type ions. J Am Soc Mass Spectrom. 2009;20:2174-81.

[27] Erlekam U, Bythell BJ, Scuderi D, Van Stipdonk M, Paizs B, Maitre P. Infrared spectroscopy of fragments of protonated peptides: direct evidence for macrocyclic structures of b5 ions. J Am Chem Soc. 2009;131:11503-8.

[28] Harrison AG. Peptide sequence scrambling through cyclization of b(5) ions. J Am Soc Mass Spectrom. 2008;19:1776-80.

[29] Riba Garcia I, Giles K, Bateman RH, Gaskell SJ. Studies of peptide a- and b-type fragment ions using stable isotope labeling and integrated ion mobility/tandem mass spectrometry. J Am Soc Mass Spectrom. 2008;19:1781-7.

[30] 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.

[31] 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.

[32] MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. The journal of physical chemistry B. 1998;102:3586-616.

[33] Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, et al. CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. J Comput Chem. 2010;31:671-90.

[34] Mesleh MF, Hunter JM, Shvartsburg AA, Schatz GC, Jarrold MF. Structural information from ion mobility measurements: Effects of the long-range potential. J Phys Chem. 1996;100:16082-6.

[35] Wyttenbach T, Witt M, Bowers MT. On the stability of amino acid zwitterions in the gas phase: The influence of derivatization, proton affinity, and alkali ion addition. J Am Chem Soc. 2000;122:3458-64.

[36] Vonhelden G, Hsu MT, Kemper PR, Bowers MT. Structural Information from Ion Mobility Measurements: Effects of the Long-Range Potential. J Chem Phys. 1991;95:3835-7.

[37] Boersema PJ, Raijmakers R, Lemeer S, Mohammed S, Heck AJ. Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. Nature protocols. 2009;4:484-94.

[38] Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Proc Natl Acad Sci U S A. 2004;101:9528-33.

[39] Cole SR, Ma X, Zhang X, Xia Y. Electron transfer dissociation (ETD) of peptides containing intrachain disulfide bonds. J Am Soc Mass Spectrom. 2012;23:310-20.

[40] Guan F, Uboh CE, Soma LR, Rudy J. Sequence elucidation of an unknown cyclic peptide of high doping potential by ETD and CID tandem mass spectrometry. J Am Soc Mass Spectrom. 2011;22:718-30.

[41] Leymarie N, Costello CE, O'Connor PB. Electron capture dissociation initiates a free radical reaction cascade. J Am Chem Soc. 2003;125:8949-58.

[42] Li X, Huang Y, O'Connor PB, Lin C. Structural heterogeneity of doubly-charged peptide b-ions. J Am Soc Mass Spectrom. 2011;22:245-54.

[43] Bush MF, Campuzano ID, Robinson CV. Ion mobility mass spectrometry of peptide ions: effects of drift gas and calibration strategies. Anal Chem. 2012;84:7124-30.

[44] Pierson NA, Chen L, Valentine SJ, Russell DH, Clemmer DE. Number of solution states of bradykinin from ion mobility and mass spectrometry measurements. J Am Chem Soc. 2011;133:13810-3.

[45] McLean JR, McLean JA, Wu Z, Becker C, Perez LM, Pace CN, et al. Factors that influence helical preferences for singly charged gas-phase peptide ions: the effects of multiple potential charge-carrying sites. The journal of physical chemistry B. 2010;114:809-16.

[46] Chen L, Shao Q, Gao YQ, Russell DH. Molecular dynamics and ion mobility spectrometry study of model beta-hairpin peptide, trpzip1. J Phys Chem A. 2011;115:4427-35.

[47] Chirot F, Calvo F, Albrieux F, Lemoine J, Tsybin YO, Dugourd P. Statistical analysis of ion mobility spectrometry. I. Unbiased and guided replica-exchange molecular dynamics. J Am Soc Mass Spectrom. 2012;23:386-96.

[48] Paizs B, Suhai S. Fragmentation pathways of protonated peptides. Mass Spectrom Rev. 2005;24:508-48.

[49] Tirado M, Rutters J, Chen X, Yeung A, van Maarseveen J, Eyler JR, et al. Disfavoring macrocycle b fragments by constraining torsional freedom: the "twisted" case of QWFGLM b6. J Am Soc Mass Spectrom. 2012;23:475-82.

[50] Tang XJ, Thibault P, Boyd RK. Fragmentation reactions of multiply-protonated peptides and implications for sequencing by tandem mass spectrometry with low-energy collision-induced dissociation. Anal Chem. 1993;65:2824-34.

[51] Chawner R, Gaskell SJ, Eyers CE. Proposal for a common nomenclature for peptide fragment ions generated following sequence scrambling during collision-induced dissociation. Rapid Commun Mass Spectrom. 2012;26:205-6.

[52] Taouatas N, Drugan MM, Heck AJ, Mohammed S. Straightforward ladder sequencing of peptides using a Lys-N metalloendopeptidase. Nat Methods. 2008;5:405-7.

[53] Rand KD, Pringle SD, Morris M, Engen JR, Brown JM. ETD in a traveling wave ion guide at tuned Z-spray ion source conditions allows for site-specific hydrogen/deuterium exchange measurements. J Am Soc Mass Spectrom. 2011;22:1784-93.

[54] 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;77:7783-95.

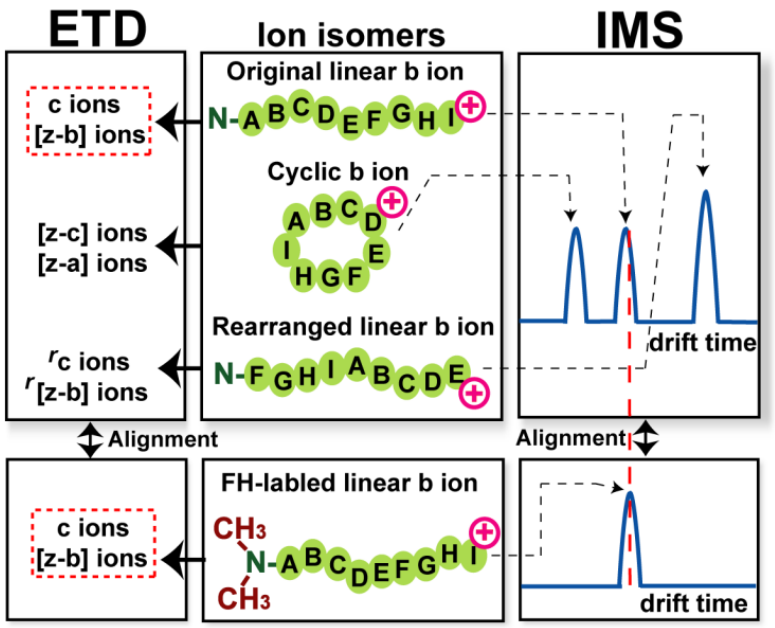


Figure 1. Workflow of the developed strategy for gas-phase ion isomer analysis. On the left panel, under ETD fragmentation each b ion isomer generates certain diagnostic fragment ions and the FH-labeled b ion is used for alignment. On the right panel, each isomer is separated by IMS, followed by comparison with the drift time of the FH-labeled b ions. Although ETD fragments multiply charged ions, we show singly charged ions here for simplicity in illustration.

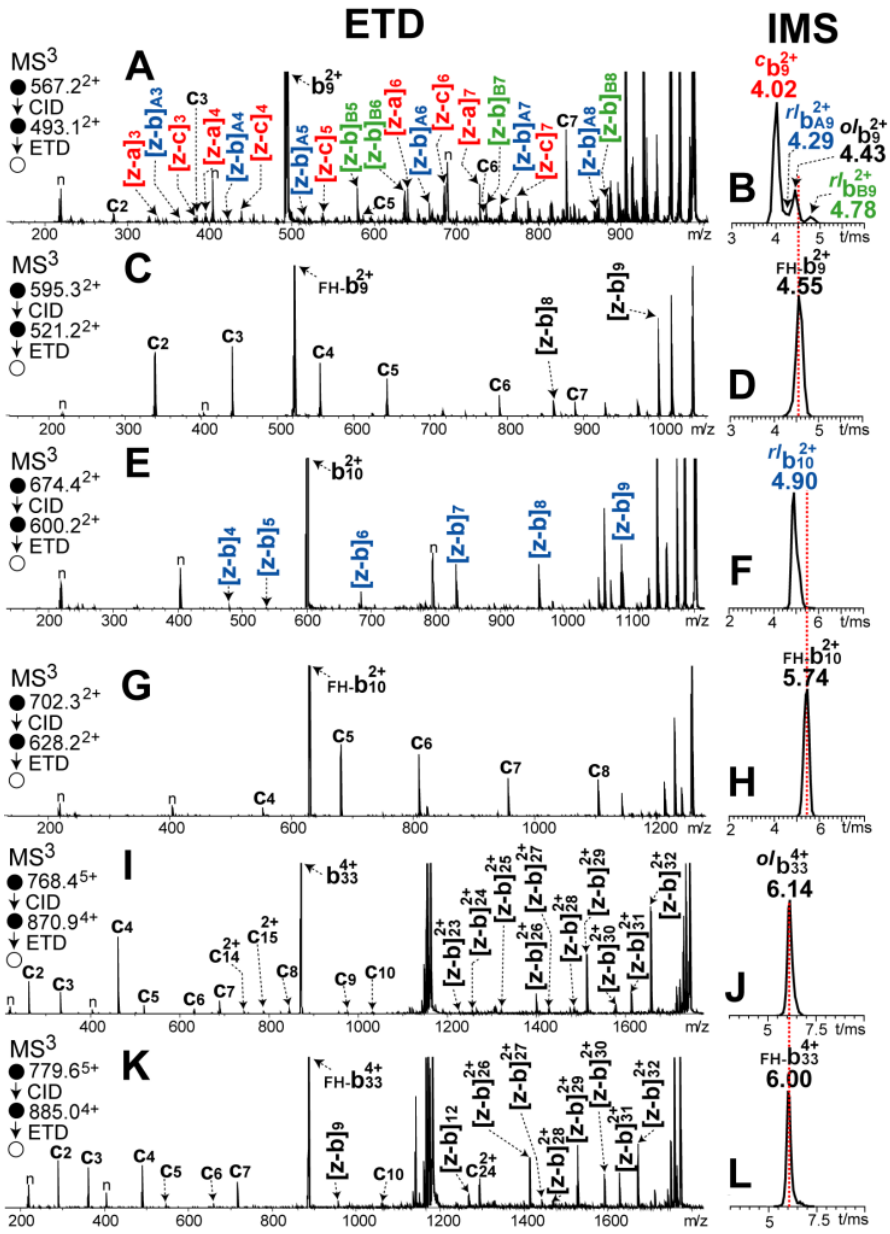


Figure 2. ETD fragmentation spectra and IMS distributions of three b ions. (A,B) b_9^{2+} ion of neurokinin. (C,D) $_{FH}$-b_9^{2+} ion of neurokinin. (E,F) b_{10}^{2+} ion of substance P. (G,H) $_{FH}$-b_{10}^{2+} ion of substance P. (I,J) b_{33}^{4+} ion of CPRP. (K,L) $_{FH}$-b_{33}^{4+} ion of CPRP. The letter n denotes system noise. The details of ion assignment are listed in Tables S4-9.

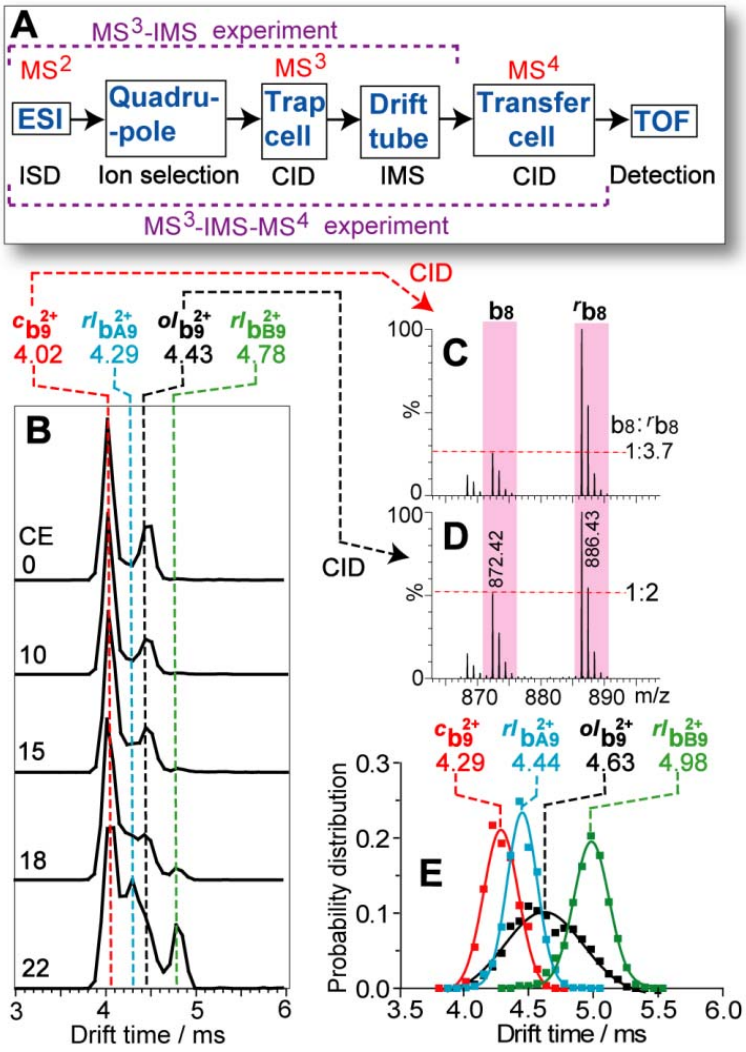


Figure 3. IMS dynamics study of neurokinin b_9. (A) Workflow of the experimental design on ESI-QTOF-IMS. (B) Drift time distributions of b_9^{2+} under various collision energy values (CE, eV; bin, 0.069 ms). (C,D) CID spectra of b_9^{2+} from MS^3-IMS-MS^4 experiment. (E) Simulated drift time distributions of the four b_9^{2+} isomers (bin, 0.069 ms). Note that the integrated distributions from simulations are the same for the four isomers because 1500 structures are used for the simulation of each isomer. By contrast, the populations of the four isomers are different in experiment.

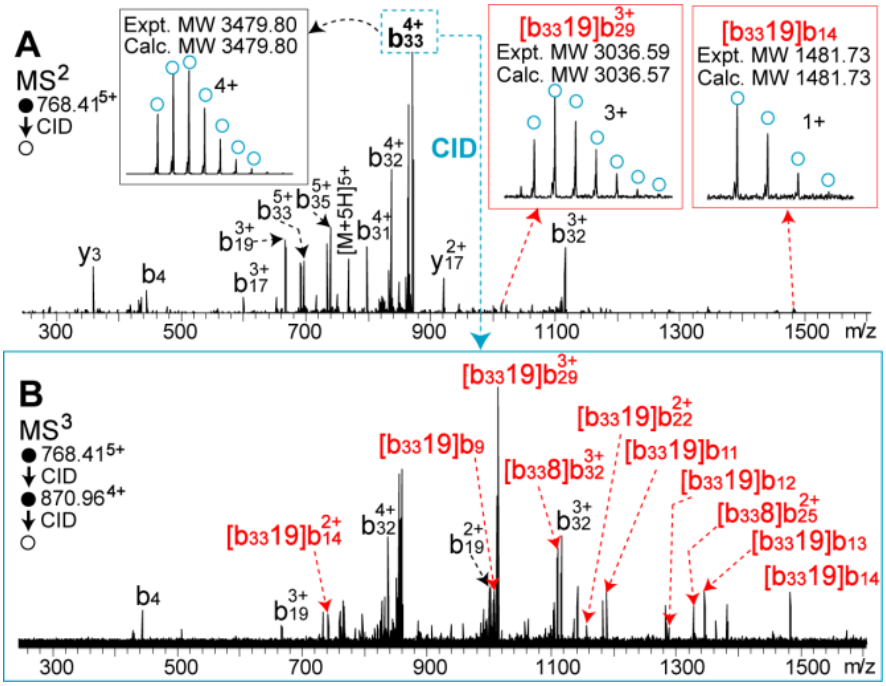


Figure 4. MS^2 (CID) of CPRP peptide (A) and MS^3 (CID) of b_{33} ion (B) acquired on LTQ-FTICR. The ^{r}b ions were highlighted with red and assigned according to ion nomenclature reported in Ref. 48. The details of ion assignment are listed in Table S10. Note that the $[b_{33}19]b_9$~$[b_{33}19]b_{14}$ and $[b_{33}8]b_{25}$ ions could also be internal fragments produced by backbone cleavage of $^{ol}b_{33}$.

Table 1. Summary of fragment ion isomer analysis of the three representative b ions.

The b ions under investigation [a]	Ion isomers	Ratio [b]
Neurokinin b_9^{2+} HKTDSFVGL^{2+}	$^{c}\mathbf{b_9}^{2+}$ Cyclo-(HKTDSFVGL)$^{2+}$	80%
	$^{ol}\mathbf{b_9}^{2+}$ HKTDSFVGL^{2+}	16%
	$^{rl}\mathbf{b_{A9}}^{2+}$ TDSFVGLHK^{2+}	_ [c]
	$^{rl}\mathbf{b_{B9}}^{2+}$ SFVGLHKTD^{2+}	4%
Substance P b_{10}^{2+} RPKPQQFFGL^{2+}	$^{rl}\mathbf{b_{10}}^{2+}$ PQQFFGLRPK^{2+}	100%
CPRP b_{33}^{4+} RSAEGLGRMGRLLASLKS DTVTPLRGFEGETGH^{4+}	$^{ol}\mathbf{b_{33}}^{4+}$ RSAEGLGRMGRLLASLKS DTVTPLRGFEGETGH^{4+}	100%

[a] The sequences of original linear structures are listed.
[b] The ratios are obtained from IMS distributions in Figure 2B, F and J, respectively.
[c] The $^{rl}b_{A9}^{2+}$ ion is not fully resolved between $^{c}b_9^{2+}$ and $^{ol}b_9^{2+}$ (Figure 2B). Therefore, the ratios of $^{c}b_9^{2+}$ and $^{ol}b_9^{2+}$ include small amount of $^{rl}b_{A9}^{2+}$.

Supporting Information for Chapter 5

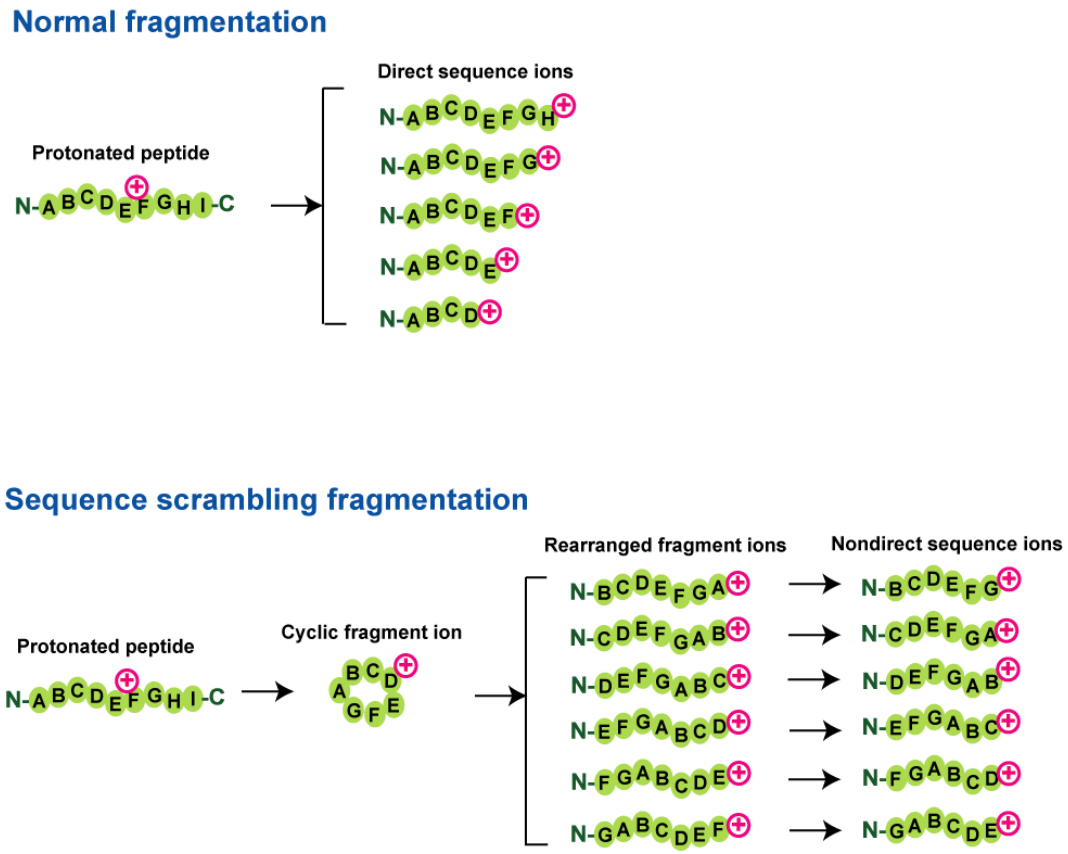


Figure S1. Graphical representation for normal fragmentation and sequence scrambling fragmentation of protonated peptides.

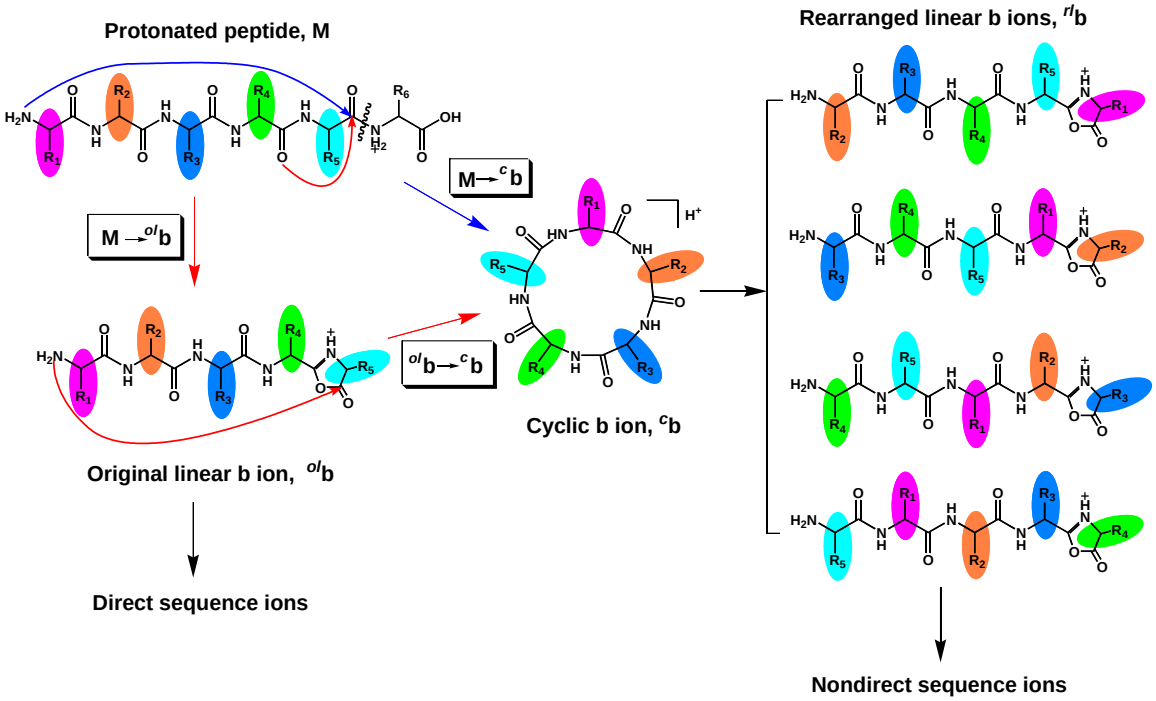


Figure S2. Proposed fragmentation pathways for cyclization and rearrangement of peptide b ions. The cyclic b ion (^{c}b) can be formed via pathway M→^{c}b or M→^{ol}b→^{c}b, and then the ring of ^{c}b ion opens at different amide bonds to produce rearranged linear b ions (^{rl}b). These ^{rl}b ions further dissociate to generate nondirect sequence ions, causing peptide sequence scrambling.

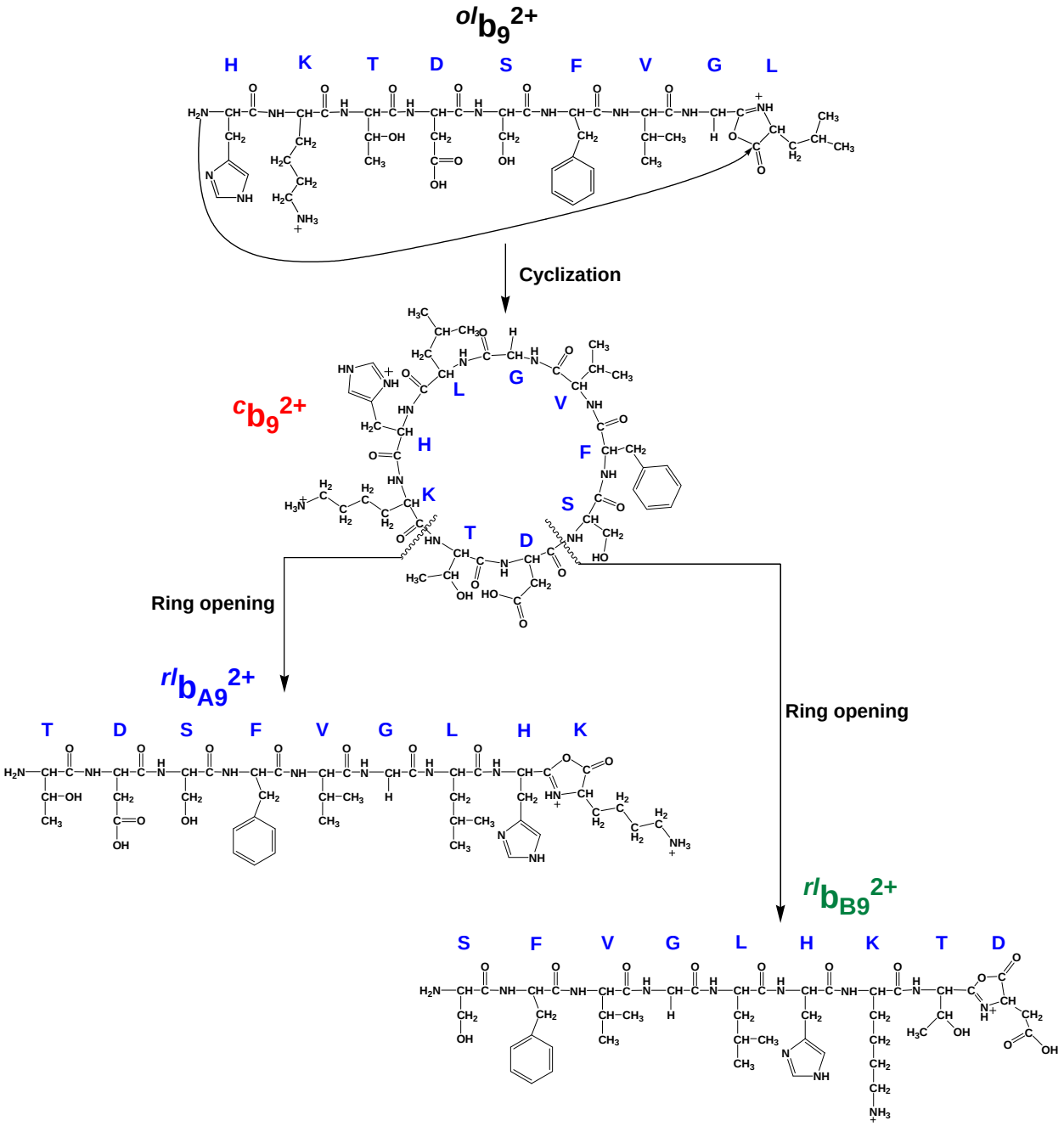


Figure S3. Proposed fragmentation pathway of neurokinin b_9 ion. These structures were used for molecular dynamics simulation.

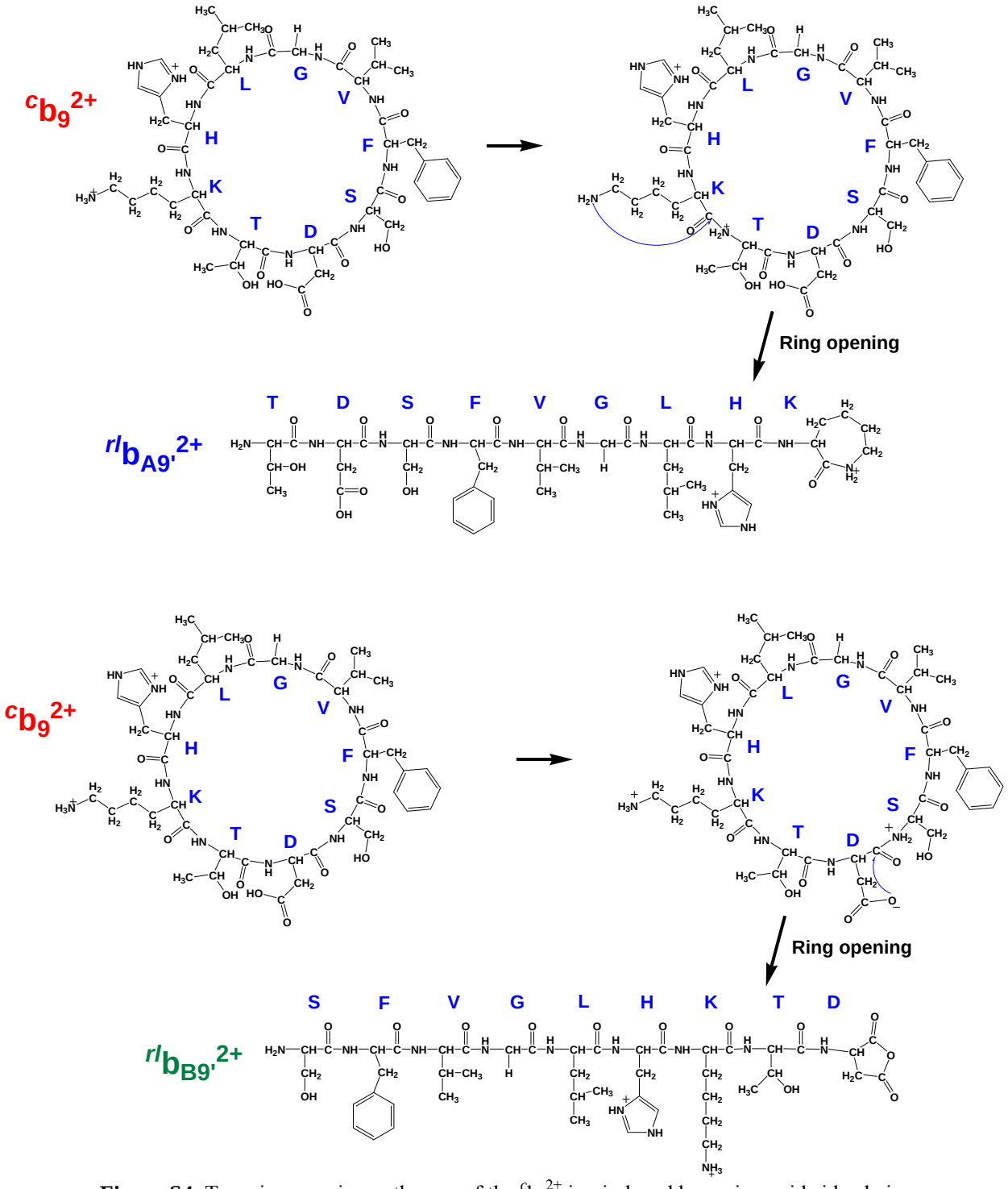


Figure S4. Two ring opening pathways of the $^{c}b_{9}^{2+}$ ion induced by amino acid side chain.

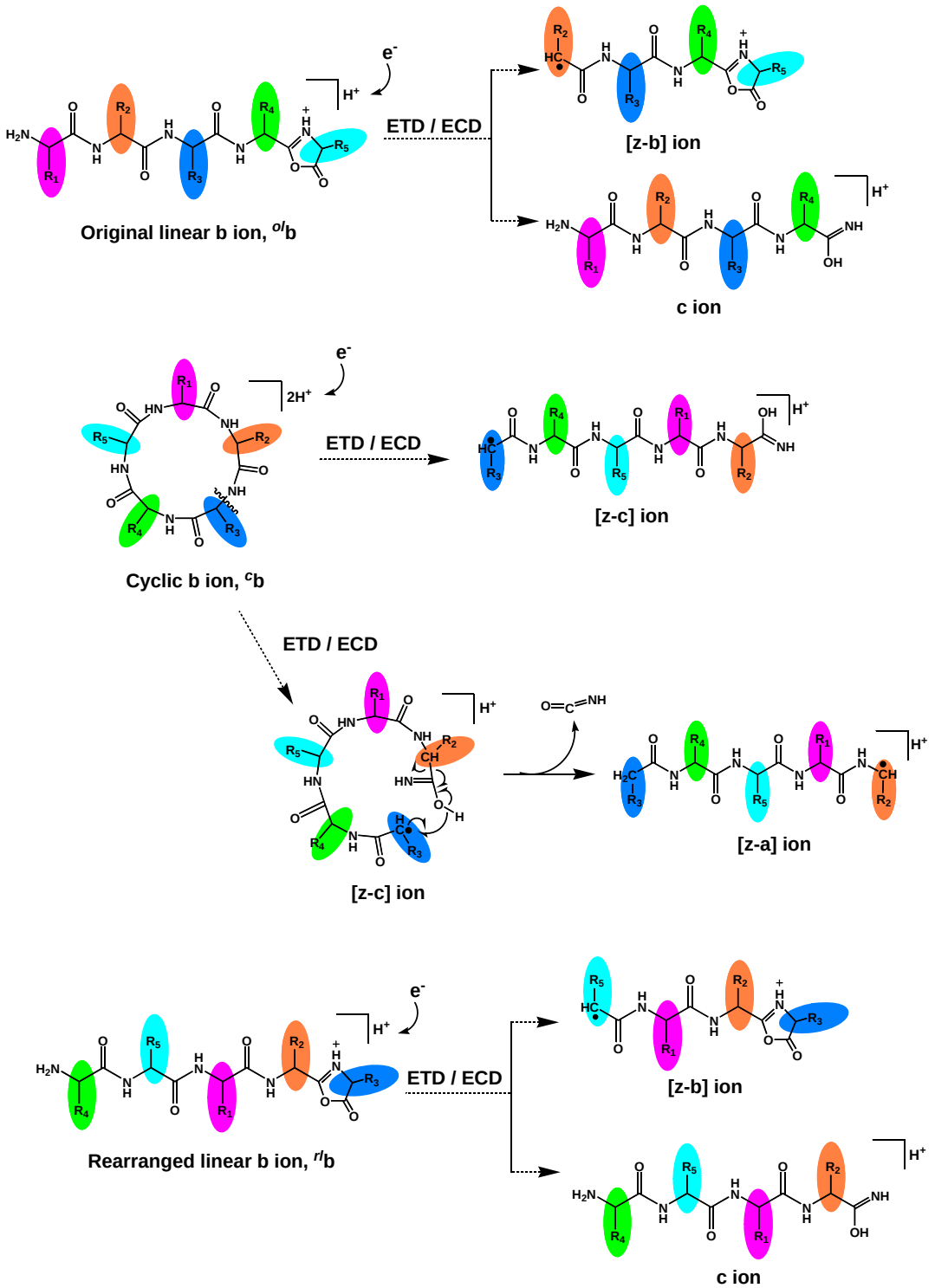


Figure S5. The proposed structures of diagnostic fragment ions produced by ETD or ECD fragmentation of ^{ol}b, ^{c}b and ^{rl}b ions. Previous studies[1-3] reported that the [z-c] ions were formed directly from protonated cyclic peptides, which underwent further dissociation to produce [z-a] ions. The ^{c}b ions have the same structures as the protonated cyclic peptides[4, 5], therefore they share identical fragmentation mechanisms under ETD or ECD conditions.

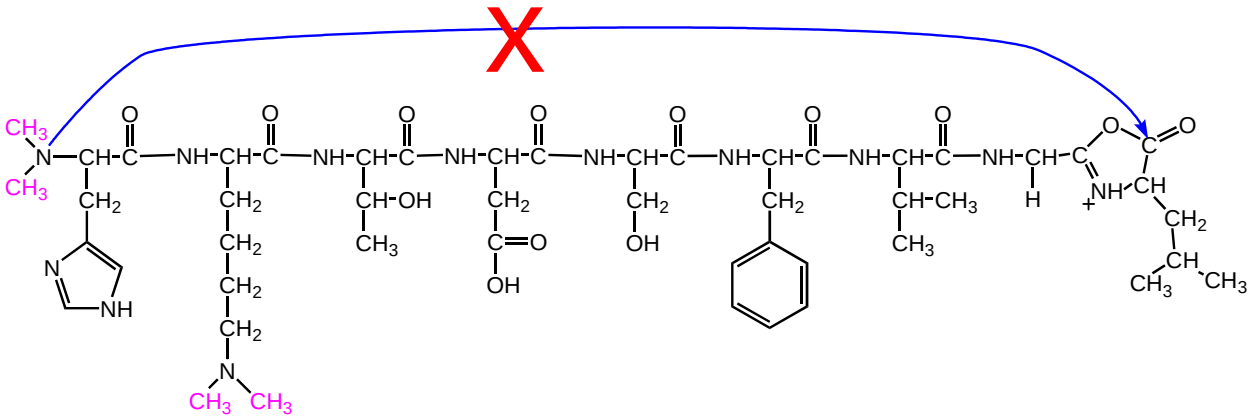


Figure S6. Cyclization reaction of b ions is prevented by (formaldehyde labeling) *N*,*N*-dimethylation of peptides due to blocking of nucleophilic attack by primary amine on the carbonyl carbon via dimethylation at the N-terminus.

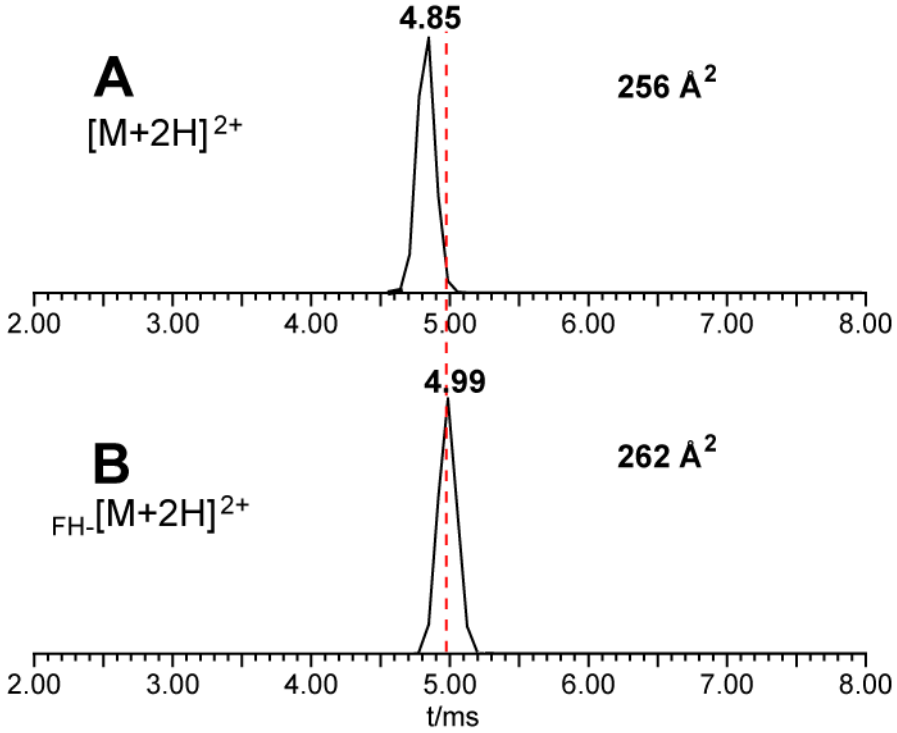


Figure S7. IMS distributions of $[M+2H]^{2+}$ (A) and $_{FH-}[M+2H]^{2+}$ (B) of neurokinin peptide.

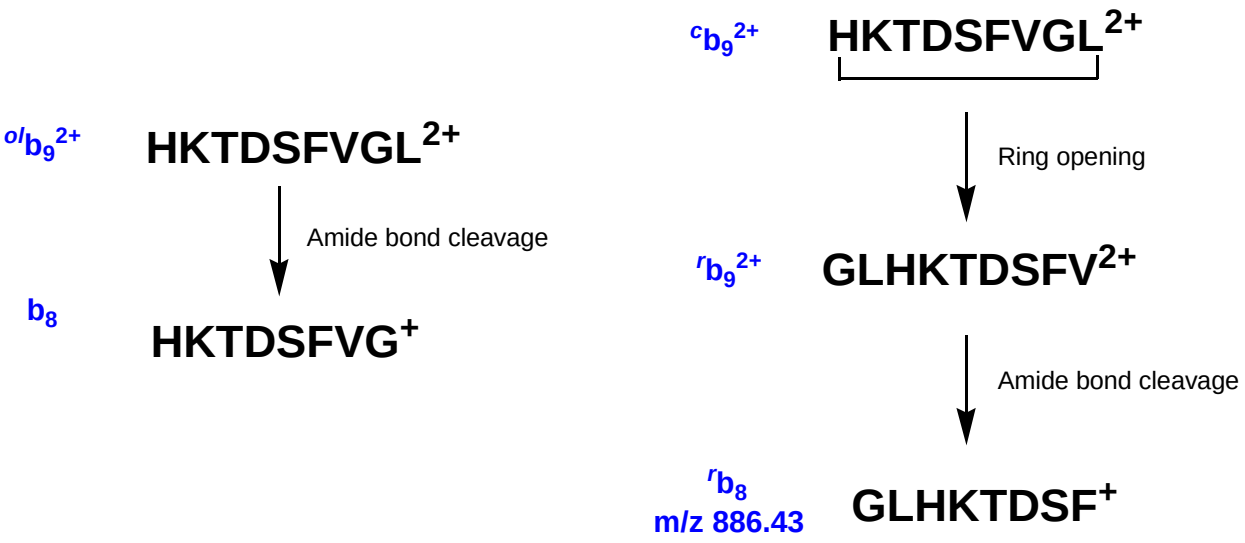


Figure S8. Fragment ions and proposed pathways for Figure 3C and D.

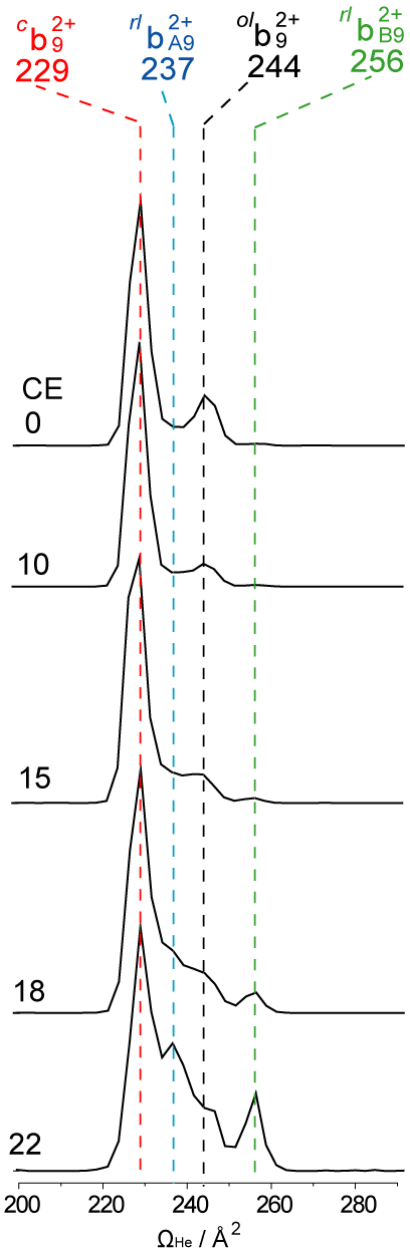


Figure S9. Collision cross section distributions of neurokinin b_9 obtained from dynamics study. Corresponding to Figure 3B. CE, collision energy.

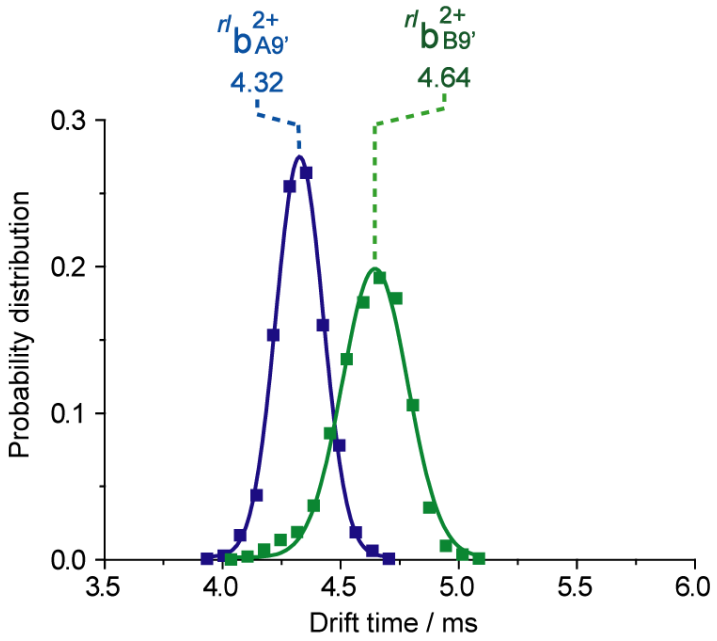


Figure S10. Theoretical drift time distributions of $^{rl}b_{A9'}^{2+}$ and $^{rl}b_{B9'}^{2+}$.

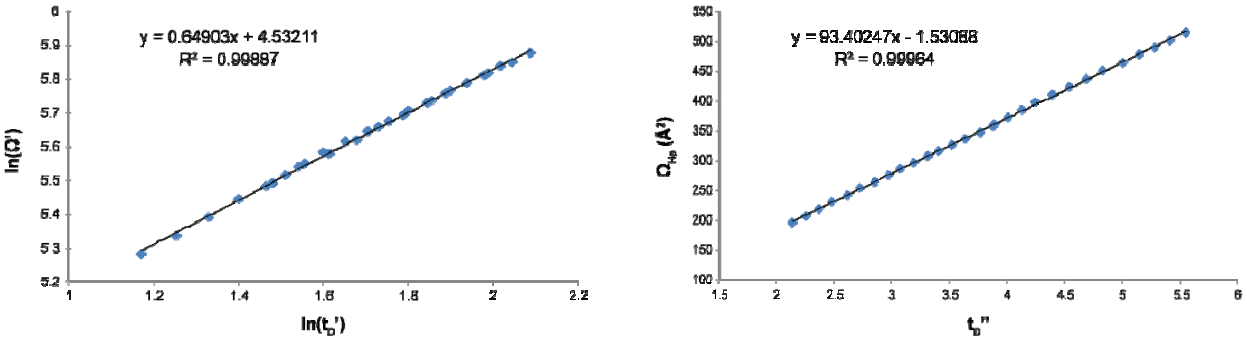


Figure S11. Sample collision cross section calibration curves acquired at a wave velocity of 800 ms^{-1} and a wave height of 40 V. These curves represent the lowest R^2 values obtained throughout the investigation.

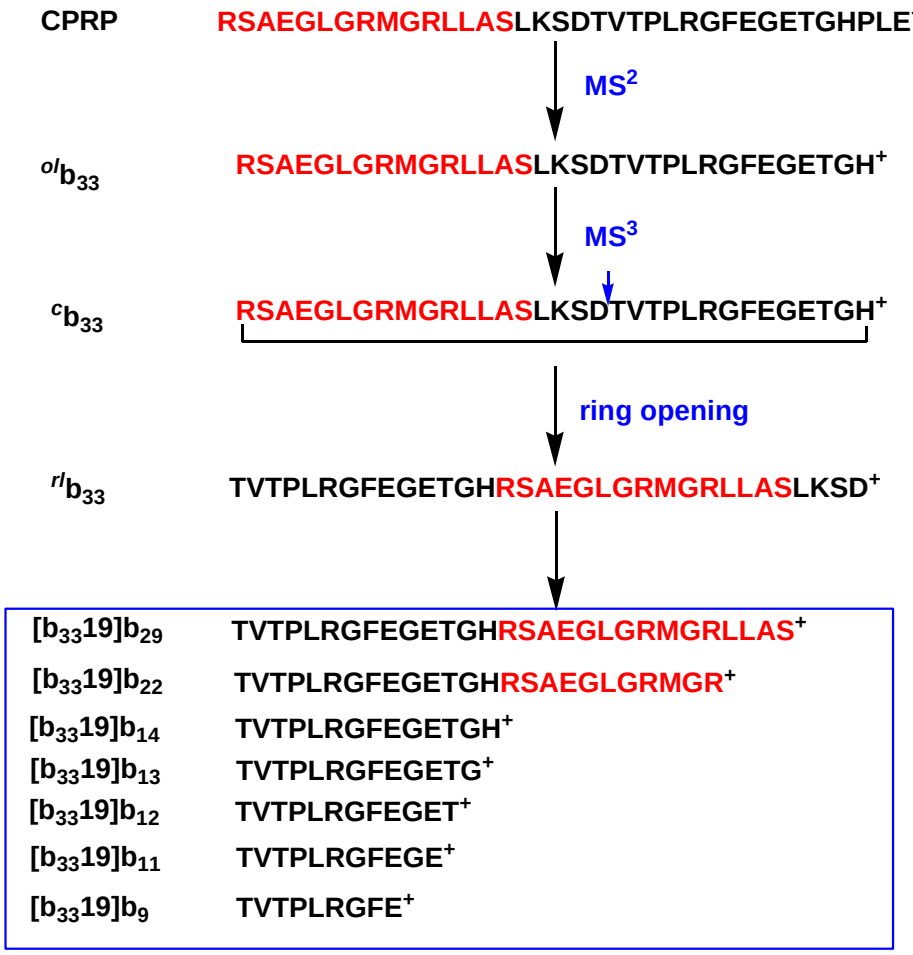


Figure S12. Proposed pathway of CPRP peptide to generate re-arranged b ions. The re-arranged sequences are highlighted in red, which moves from the N-terminus of the peptide precursor ion to the C-terminus of the $[b_{33}19]b_{29}$ ion, resulting in a sequential re-arranged b ions down to $[b_{33}19]b_9$. The mass accuracy of the observed re-arranged b ions is less than 8 ppm as shown in Table S10. For illustration purpose, only single charge is labeled. Note that the $[b_{33}19]b_9$~$[b_{33}19]b_{14}$ could also be internal fragments of $^{ol}b_{33}$.

Table S1. Nomenclature for special ions described in this study.

M	Peptide molecular ion
^{l}b	Linear b ion
^{ol}b	Original linear b ion
^{rl}b	Rearranged linear b ion
^{c}b	Cyclic b ion
^{r}[z-b]	Rearranged [z-b] ion
^{r}c	Rearranged c ion
$[z\text{-}b]_{A}$	The first set of rearranged [z-b]
$[z\text{-}b]_{B}$	The second set of rearranged [z-b]
$_{FH}$-b	Formaldehyde labeled b ion
$_{FH}$-$[M+H]^{+}$	Formaldehyde labeled $[M+H]^{+}$ ion

Table S2. Protonation sites of the b ion isomers for molecular dynamic simulation.

Fragment ion	**Protonation site**	**Structure**
$^{c}b_9^{2+}$	Side chains of histidine and lysine	Figure S5
$^{rl}b_{A9}^{2+}$	C-terminal Oxazolone structure and lysine side chain	Figure S5
$^{rl}b_{B9}^{2+}$	C-terminal Oxazolone structure and lysine side chain	Figure S5
$^{ol}b_9^{2+}$	C-terminal Oxazolone structure and lysine side chain	Figure S5
$^{rl}b_{A9'}^{2+}$	C-terminal seven-membered ring structure and histidine side chain	Figure S6
$^{rl}b_{B9'}^{2+}$	Side chains of histidine and lysine	Figure S6

Table S3. Measured collision cross sections (Ω) in this work.

Peptide	Ion	Measured Ω_{He} ($Å^2$) [a]
Neurokinin	$^{c}b_9^{2+}$	228.8±1.3
	$^{ol}b_9^{2+}$	243.6±1.2
	$^{rl}b_{A9}^{2+}$	237.2±1.2
	$^{rl}b_{B9}^{2+}$	255.7±0.9
	$_{FH\text{-}}b_9^{2+}$	247.6±0.6
	$[M+2H]^{2+}$	256.0±1.5
	$_{FH\text{-}}[M+2H]^{2+}$	262.3±0.6
Substance P	$^{rl}b_{10}^{2+}$	260.0±1.3
	$_{FH\text{-}}b_{10}^{2+}$	277.8±1.0

[a] In practice, the measurements were performed using N_2 gas. The Ω_{N2} was converted to Ω_{He} according to the method reported by Bush et al[6]. This is consistent with the calculated Ω_{He}, as software Sigma can only simulate Ω_{He} calculation.

Table S4. Fragment ion assignment in MS^3 (ETD) of neurokinin b_9. Corresponding to Figure 2A.

Obsd. *m/z*	Charge	Expt. $[M+H]^+$	Calc. $[M+H]^+$	Ion assignment
266.8	1	266.8	266.2	c_2
384.1	1	384.1	384.2	c_3
586.2	1	586.2	586.3	c_5
733.3	1	733.3	733.4	c_6
832.3	1	832.3	832.4	c_7
363.2	1	363.2	363.2	$[z\text{-}b]_{A3}$, LHK
420.2	1	420.2	420.2	$[z\text{-}b]_{A4}$, GLHK
519.3	1	519.3	519.3	$[z\text{-}b]_{A5}$, VGLHK
666.3	1	666.3	666.4	$[z\text{-}b]_{A6}$, FVGLHK
753.3	1	753.3	753.4	$[z\text{-}b]_{A7}$, SFVGLHK
868.3	1	868.3	868.4	$[z\text{-}b]_{A8}$, DSFVGLHK
579.2	1	579.2	579.3	$[z\text{-}b]_{B5}$, LHKTD
636.2	1	636.2	636.3	$[z\text{-}b]_{B6}$, GLHKTD
735.3	1	735.3	735.4	$[z\text{-}b]_{B7}$, VGLHKTD
882.3	1	882.3	882.5	$[z\text{-}b]_{B8}$, FVGLHKTD
381.2	1	381.3	381.3	$[z\text{-}c]_3$, LHK
438.2	1	438.2	438.3	$[z\text{-}c]_4$, GLHK
537.2	1	537.2	537.4	$[z\text{-}c]_5$, VGLHK
684.3	1	684.3	684.4	$[z\text{-}c]_6$, FVGLHK
771.4	1	771.4	771.5	$[z\text{-}c]_7$, SFVGLHK
337.1	1	337.1	337.2	$[z\text{-}a]_3$, LHK
394.2	1	394.2	394.3	$[z\text{-}a]_4$, GLHK
640.3	1	640.3	640.4	$[z\text{-}c]_6$, FVGLHK
727.3	1	727.3	727.4	$[z\text{-}c]_7$, SFVGLHK

Table S5. Fragment ion assignment in MS^3 (ETD) of neurokinin $_{FH}$-b_9. Corresponding to Figure 2C.

Obsd. m/z	Charge	Expt. $[M+H]^+$	Calc. $[M+H]^+$	Ion assignment
339.1	1	339.1	339.3	c_2
440.2	1	440.2	440.3	c_3
555.2	1	555.2	555.3	c_4
642.2	1	642.2	642.4	c_5
789.3	1	789.3	789.4	c_6
888.4	1	888.4	888.4	c_7
860.4	1	860.4	860.5	$[z\text{-}b]_8$, K*TDSFVGL
997.5	1	997.5	997.5	$[z\text{-}b]_9$, HK*TDSFVGL

* Dimethylation.

Table S6. Fragment ion assignment in MS^3 (ETD) of Substance P b_{10}. Corresponding to Figure 2E.

Obsd. m/z	Charge	Expt. $[M+H]^+$	Calc. $[M+H]^+$	Ion assignment
479.2	1	479.2	479.3	$[z\text{-}b]_4$, LRPK
536.2	1	536.2	536.3	$[z\text{-}b]_5$, GLRPK
683.3	1	683.3	683.4	$[z\text{-}b]_6$, FGLRPK
830.4	1	830.4	830.5	$[z\text{-}b]_7$, FFGLRPK
958.5	1	958.5	958.5	$[z\text{-}b]_8$, QFFGLRPK
1086.5	1	1086.5	1086.6	$[z\text{-}b]_9$, QQFFGLRPK

Table S7. Fragment ion assignment in MS^3 (ETD) of Substance P_{FH}-b_{10}. Corresponding to Figure 2G.

Obsd. m/z	Charge	Expt. $[M+H]^+$	Calc. $[M+H]^+$	Ion assignment
552.3	1	552.3	552.4	c_4
680.4	1	680.4	680.5	c_5
808.4	1	808.4	808.5	c_6
955.5	1	955.5	955.6	c_7
1102.6	1	1102.6	1102.7	c_8

Table S8. Fragment ion assignment in MS^3 (ETD) of CPRP b_{33}. Corresponding to Figure 2I.

Obsd. m/z	Charge	Expt. $[M+H]^+$	Calc. $[M+H]^+$	Ion assignment
261.1	1	261.1	261.2	c_2
332.1	1	332.1	332.2	c_3
461.2	1	461.2	461.2	c_4
518.2	1	518.2	518.3	c_5
631.3	1	631.3	631.4	c_6
688.3	1	688.3	688.4	c_7
844.4	1	844.4	844.5	c_8
975.5	1	975.5	975.5	c_9
1032.5	1	1032.5	1032.5	c_{10}
743.4	2	1485.8	1485.8	c_{14}^{2+}
786.9	2	1572.8	1572.9	c_{15}^{2+}
1225.6	2	2450.2	2450.3	$[z\text{-}b]_{23}^{2+}$
1254.1	2	2507.2	2507.3	$[z\text{-}b]_{24}^{2+}$
1319.6	2	2638.2	2638.3	$[z\text{-}b]_{25}^{2+}$
1397.7	2	2794.4	2794.4	$[z\text{-}b]_{26}^{2+}$
1426.2	2	2851.4	2851.5	$[z\text{-}b]_{27}^{2+}$
1482.7	2	2964.4	2964.5	$[z\text{-}b]_{28}^{2+}$
1511.2	2	3021.4	3021.6	$[z\text{-}b]_{29}^{2+}$
1575.7	2	3150.4	3150.6	$[z\text{-}b]_{30}^{2+}$
1611.3	2	3221.6	3221.7	$[z\text{-}b]_{31}^{2+}$
1654.8	2	3308.6	3308.7	$[z\text{-}b]_{32}^{2+}$

Table S9. Fragment ion assignment in MS^3 (ETD) of CPRP $_{FH}$-b_{33}. Corresponding to Figure 2K.

Obsd. m/z	Charge	Expt. $[M+H]^+$	Calc. $[M+H]^+$	Ion assignment
289.1	1	289.1	289.2	c_2
360.1	1	360.1	360.2	c_3
489.2	1	489.2	489.3	c_4
546.3	1	546.3	546.3	c_5
659.3	1	659.3	659.4	c_6
716.3	1	716.3	716.4	c_7
1060.6	1	1060.6	1060.6	c_{10}
1292.1	2	2583.2	2583.5	c_{24}^{2+}
955.3	1	955.3	955.4	$[z\text{-}b]_9$
1266.5	1	1266.5	1266.6	$[z\text{-}b]_{12}$
1411.6	2	2822.2	2822.5	$[z\text{-}b]_{26}^{2+}$
1440.2	2	2879.4	2879.5	$[z\text{-}b]_{27}^{2+}$
1496.7	2	2992.4	2992.6	$[z\text{-}b]_{28}^{2+}$
1525.2	2	3049.4	3049.6	$[z\text{-}b]_{29}^{2+}$
1589.7	2	3178.4	3178.6	$[z\text{-}b]_{30}^{2+}$
1625.2	2	3249.4	3249.6	$[z\text{-}b]_{31}^{2+}$
1668.7	2	3336.4	3336.7	$[z\text{-}b]_{32}^{2+}$

Table S10. Non-direct sequence ions detected from MS^3 (CID) of CPRP b_{33}. Corresponding to Figure 4B.

Ion assignment	Obsd. m/z	Charge	Expt. $[M+H]^+$	Calc. $[M+H]^+$	Error/ppm
$[b_{33}19]b_9$	1001.54	1	1001.53	1001.53	0
$[b_{33}19]b_{11}$	1187.61	1	1187.60	1187.60	0
$[b_{33}19]b_{12}$	1288.65	1	1288.64	1288.65	8
$[b_{33}19]b_{13}$	1345.67	1	1345.67	1345.67	0
$[b_{33}19]b_{14}$	1482.73	1	1482.72	1482.73	7
$[b_{33}19]b_{22}$	1155.09	2	2309.16	2309.17	4
$[b_{33}19]b_{29}^{3+}$	1013.20	3	3037.59	3037.57	7
$[b_{33}8]b_{25}^{2+}$	1327.69	2	2654.36	2654.36	0
$[b_{33}8]b_{32}^{3+}$	1108.91	3	3324.70	3324.70	0

Table S11. The ions used for collision cross section calibration [6].

Polyalanine Standards

n	mass (Da)	charge	m/z	Ω_{He} (Å^2)
11	799.42	2	400.72	197
12	870.46	2	436.24	208
13	941.49	2	471.75	220
14	1012.53	2	507.27	232
15	1083.57	2	542.79	243
16	1154.60	2	578.31	255
17	1225.64	2	613.83	265
18	1296.68	2	649.35	276
19	1367.72	2	684.87	287
20	1438.75	2	720.38	297
21	1509.79	2	755.90	308
22	1580.83	2	791.42	317
23	1651.86	2	826.94	327
24	1722.90	2	862.46	337
25	1793.94	2	897.98	348
26	1864.98	2	933.50	358
21*	1509.79	3	504.27	361
22	1580.83	3	527.95	373
23	1651.86	3	551.63	386
24	1722.90	3	575.31	399
25	1793.94	3	598.99	412
26	1864.98	3	622.67	425
27	1936.01	3	646.35	438
28	2007.05	3	670.02	452
29	2078.09	3	693.70	465
30	2149.12	3	717.38	479
31	2220.16	3	741.06	490
32	2291.20	3	764.74	502
33	2362.24	3	788.42	516

* The ion was detected at all wave velocities.

EXPERIMENTALS (Supplementary Materials)

Collision Cross Section Measurement

Gas-phase helium collision cross section values (Ω_{He}) for all of the ions described in the manuscript were measured on the Synapt G2 travelling-wave ion mobility mass spectrometer. Ω_{He} calibration was performed using combined methods from Bush et al.[6] and Ruotolo et al.[7] to obtain accurate measurements for doubly charged cations. Although N_2 was actually used as the IM buffer gas, Bush et al.[6] has shown that using the known Ω_{He} values for the calibration standards accounts for many of the systematic differences between He and N_2, and accurate Ω_{He} for unknowns can be obtained. Polyalanine peptides were purchased from Sigma Aldrich and dissolved in 49.5/49.5/1 water/ACN/formic acid at a concentration of 10 µg/mL. A separate calibration spectrum was acquired for each of the following wave velocities (m/s): 500, 600, 700, and 800. The wave height was kept constant at 40 V. Following acquisitions, the drift times (t_D) of the doubly and triply charged ions were used to create a best-fit power curve and determine the constants needed to calculate unknown Ω_{He}. Bush et al.[6] has shown that singly charged and multiply charged ions display distinct calibration distributions, thus we opted to omit singly charged polyalanines from our calibrations as they may introduce error when measuring doubly charged unknowns. Table S11 lists the ions of polyalanine used for collision cross section calibration.

The equations in the following paragraph were obtained from previously published sources.[7, 8] In travelling wave ion mobility, t_D and Ω_{He} are non-linearly related by Equation 1:

$$\Omega_{He} = \frac{ze}{16}\left[\frac{18\pi}{k_b T}\left(\frac{1}{m}+\frac{1}{M_{He}}\right)\right]^{1/2}\frac{760}{P}\frac{T}{273.2}\frac{1}{NL}Al_D^{B} \qquad (1)$$

The variables z and e make up the charge of the analyte, T is the temperature of the drift gas, M_{He} is the mass of the helium drift gas, m is the mass of the analyte, P is the pressure inside the drift cell, N is the number density of drift gas molecules, L is the length of the drift cell, and k_b is Boltzmann's constant. A and B are constants that arise from the non-uniformity of the travelling-wave electric field and must be empirically determined by calibration.

The t_D for each calibration standard was converted to corrected drift time (t_D') by Equation 2 to account for the *m*/*z*-dependent travel time through the Synapt G2's ion optics:

$$t_D' = t_D - \left(\frac{C\sqrt{m/z}}{1000}\right) \qquad (2)$$

C is the delay constant set by the MS control software. The reduced collision cross section (Ω') was normalized for mass and charge contributions and calculated by Equation 3:

$$\Omega' = \frac{\Omega_{He}}{z\left(\frac{1}{m}+\frac{1}{M_{He}}\right)^{1/2}} \qquad (3)$$

Plotting the natural log of Ω' versus the natural log of t_D' yielded a linear best-fit line, the slope of which is B from Equation 1. From here, the doubly corrected drift time (t_D'') was calculated by Equation 4:

$$t_D'' = z\left(t_D'\right)^B\left(\frac{1}{m}+\frac{1}{M_{He}}\right)^{1/2} \qquad (4)$$

A final plot was constructed with t_D'' on the x-axis and Ω_{He} on the y-axis. The equation of the best-fit line was then used to calculate the Ω_{He} of the unknown peptides and peptide fragments.

It should be noted that, to the knowledge of the authors, this is the first time collision cross sections have been measured for MS/MS fragments on Synapt G2. It was very important for us to investigate whether or not subjecting ions to the collision energies herein would cause shifts in t_D and introduce error into Ω_{He}. As shown in Figure 3B and Figure S8, the apex value

for t_D and Ω_{He} for any particular structure remains stable with increasing energy. We attributed the broadening of $^{c}b_9$ at 22 eV collision energy to insufficient resolution from the overlapping $^{rl}b_{A9}$. Additionally, increasing cone voltage from 30 V to 50 V did not cause a t_D shift. Therefore, we assert that the Ω_{He} calibrations and measurements are valid for MS/MS fragments.

High-resolution Tandem MS of CPRP Neuropeptide

The high-resolution tandem MS of CPRP peptide was carried out on a 7T linear trap quadrupole (LTQ)/Fourier transform ion cyclotron resonance (FTICR) (LTQ-FT Ultra) hybrid mass spectrometer (Thermo Scientific Inc., Bremen, Germany) equipped with an automated chip-based nano-ESI source (Triversa NanoMate; Advion BioSciences, Ithaca, NY). The spray voltage was 1.5 kV versus the inlet of the mass spectrometer, resulting in a flow of 100 nL/min. Ion transmission into the linear trap and subsequently into the FTICR cell was automatically optimized for the maximal ion signal. The peptide molecular ions were first isolated (MS^2) and then dissociated by CID using 25% of normalized collision energy. In MS^3, the b_{33} ion was selected and fragmented by CID with 16% of normalized collision energy. Typically, 1000 transients were averaged to ensure high quality MS/MS spectra. All FTICR spectra were processed with Xtract Software (Xcalibur 2.0.5, Thermo Scientific Inc., Bremen, Germany) using a S/N threshold of 1.5 and fit factor of 40% and validated manually.

References

[1] Leymarie N, Costello CE, O'Connor PB. Electron capture dissociation initiates a free radical reaction cascade. J Am Chem Soc. 2003;125:8949-58.

[2] Li X, Huang Y, O'Connor PB, Lin C. Structural heterogeneity of doubly-charged peptide b-ions. J Am Soc Mass Spectrom. 2011;22:245-54.

[3] Li X, Lin C, Han L, Costello CE, O'Connor PB. Charge remote fragmentation in electron capture and electron transfer dissociation. J Am Soc Mass Spectrom. 2010;21:646-56.

[4] Harrison AG, Young AB, Bleiholder C, Suhai S, Paizs B. Scrambling of sequence information in collision-induced dissociation of peptides. J Am Chem Soc. 2006;128:10364-5.

[5] Harrison AG. To b or not to b: the ongoing saga of peptide b ions. Mass Spectrom Rev. 2009;28:640-54.

[6] Bush MF, Campuzano ID, Robinson CV. Ion mobility mass spectrometry of peptide ions: effects of drift gas and calibration strategies. Anal Chem. 2012;84:7124-30.

[7] Ruotolo BT, Benesch JL, Sandercock AM, Hyung SJ, Robinson CV. Ion mobility-mass spectrometry analysis of large protein complexes. Nat Protoc. 2008;3:1139-52.

[8] Smith DP, Knapman TW, Campuzano I, Malham RW, Berryman JT, Radford SE, et al. Deciphering drift time measurements from travelling wave ion mobility spectrometry-mass spectrometry studies. Eur J Mass Spectrom (Chichester, Eng). 2009;15:113-30.

Chapter 6

Qualitative and Quantitative Top-down Mass Spectral Analysis of Crustacean Hyperglycemic Hormones in Response to Feeding

Adapted from: Chenxi Jia, Qing Yu, Jingxin Wang, and Lingjun Li, Qualitative and Quantitative Top-down Mass Spectral Analysis of Crustacean Hyperglycemic Hormones in Response to Feeding. *Proteomics*, under revision.

Abstract

An efficient pipeline for peptide discovery accelerates peptidomic analysis and facilitates a better understanding of the functional roles of neuropeptides. However, qualitative and quantitative analysis of large neuropeptides are challenging due to the bigger molecular sizes, multiple post-translational modifications, and interference by homologous isoforms. Herein, we refined two methodologies in the pipeline for highly confident and efficient mass spectrometry (MS)-based peptide discovery. For the qualitative analysis, the so-called "high resolution/accurate mass" measurement on Orbitrap mass spectrometers was integrated with computer-assisted homology search, which was successfully applied to decipher the substituted amino acid residues in large neuropeptides by referring to homologous sequences. For the quantitative analysis, a new isotopic labeling-assisted top-down MS strategy was developed, which enabled direct monitoring of the abundance changes of endogenous large neuropeptides. By using the refined peptide discovery pipeline, one novel crustacean hyperglycemic hormone (CHH) from the Dungeness crab sinus glands was confidently identified and *de novo* sequenced, and its relative abundance was quantified. Comparative analysis of CHHs in unfed and fed crabs revealed that the peptide abundance in the sinus glands was significantly increased after food intake, suggesting that the release of CHHs might be altered by feeding behavior.

6.1 Introduction

Neuropeptides represents a complex and diverse group of naturally occurring endogenous molecules in nervous system that play important roles in regulation of physiological process, such as stress, memory, circadian rhythm, and energy homeostasis [1-4]. Among these important physiological studies, the investigation of functional roles of neuropeptides in food intake has attracted enormous interest due to potential link with energy homeostasis disorders leading to obesity and diabetes [2, 3]. Several neuropeptides in the mammalian hypothalamus have been reported to control food intake, including neuropeptide Y, proopiomelanocortin, melanin-concentrating hormone, neurotensin, cholecystokinin (CCK), leptin and ghrelin [2, 3, 5, 6]. Previously, we investigated crustacean neuropeptidome by comparative mass spectrometry (MS)-based approaches and observed the expression level changes of RY-amides, tachykinin-related peptides, RFamide-related peptides, and pyrokinins in response to feeding [6, 7].

Many of the feeding-related signaling molecules and pathways underlying complex behaviors are conserved across species [8, 9]. For example, neuropeptide F [10] and CCK-like peptides [8] as well as the related signaling pathways have been strongly implicated in the regulation of food intake in vertebrates and invertebrates. In crustaceans, the balanced secretion of crustacean hyperglycemia hormone (CHH) and insulin-like peptides is an essential part of the control of energy homeostasis [11]. CHH is able to initiate glycemia by increasing D-glucose levels in the hemolymph via mobilization of D-glucose from the hepatopancreas and muscle glycogen stores. In contrast, the insulin-related peptides act to cause a decrease in hemolymph D-glucose levels. To some extent, the balanced secretion mechanism of CHHs and insulin-like peptides in crustaceans is comparable with the well-

known opposite regulatory roles of glucagon and insulin in glucose metabolism of humans [12]. Regarding control of food intake in humans, insulin sensitizes the brain to short-term meal-generated satiety signals; and glucagon acts to reduce meal size [12]. However, the potential functional role of CHHs on control of crustacean food intake is still unknown so far. Before answering this complex biological question, the initial step is to obtain the physiologically relevant evidence that the expression level of CHHs can be altered by feeding behavior, which is the goal of this study.

Qualitative and quantitative analysis of CHHs at endogenous level is challenging because these large peptides contain multiple post-translational modifications and their molecular sizes are over 8 kDa [13]. In our previous study, we established a multi-faceted MS-based platform which allows confident *de novo* sequencing of CHH-family neuropeptides without assistance of genome database [14]. However, the substituted residues of the target CHHs need to be deciphered according to homologous sequences manually, thus limiting the overall throughput. In addition to qualitative analysis, accurate quantitation of large neuropeptides is essential for probing their functional dynamics. However, because the large neuropeptides exhibit broader peak elution patterns on reversed-phase liquid chromatography (RPLC) leading to poor reproducibility [15], the label-free method cannot offer reliable and accurate measurement. Alternatively, tryptic digest of large neuropeptides may be quantified by label-free or isotopic labeling approaches. However, the large peptide isoforms share homologous sequences [16], thus producing the same tryptic peptides and causing interference by each other.

In this study, we refined two major components of method units in the peptide discovery pipeline for qualitative and quantitative analysis of large neuropeptides. For qualitative

analysis, integration of "high-resolution/accurate mass (HR/AM)" measurement [17, 18] and algorithm-assisted homology search [19] enabled prompt and accurate sequence prediction for large neuropeptides. For quantitative analysis, we developed a stable isotopic labeling-assisted [20-22] top-down MS strategy that offered precise accurate and reliable quantitative results to monitor changes of endogenous abundance of large neuropeptides. By using the refined pipeline for peptide discovery, the novel ^{Cam}-CHH and ^{Cam}-CHH$_{pQ}$ were identified from the sinus glands of the Dungeness crab *Cancer magister*, and their relative abundances were altered upon feeding.

6.2 Materials and methods

6.2.1 Materials

Methanol, glacial acetic acid, borane pyridine, formaldehyde, formaldehyde-d_2, Dithiothreitol (DTT) and iodoacetamide (IAA) were obtained from Sigma-Aldrich (St. Louis, MO). Optima grade formic acid, acetonitrile (ACN), water, and methanol were purchased from Fisher Scientific (Pittsburgh, PA).

6.2.2 Animals, feeding experiments and tissue dissection

Dungeness crabs *C. magister* were purchased from The Fresh Lobster Company (Gloucester, MA) and kept without food in a circulating artificial seawater tank at 10−15 °C for 6 days before use. In the feeding experiment, one crab was fed with seafood until it stopped eating, which usually took about 20 min. After 40 min, the fed crab and one unfed crab (as control) were cold anesthetized by packing them in ice for 15 min [6]. The sinus glands were dissected in chilled physiological saline and collected in chilled acidified methanol. The tissues were homogenized and extracted with 100 μL of acidified methanol (methanol:H_2O:acetic acid, 90:9:1, v:v:v) for

three times [14]. After drying down in speedvac, the tissue extract was re-suspended in 100 μL of water and stored in -80 °C for further analysis.

6.2.3 Isotopic labeling of peptides by formaldehyde

Two 20 μL of aliquots of tissue extracts of fed and unfed crabs were dried down and re-suspend in 10 μL of solution (90% ACN in water). The two samples were added with borane pyridine (C_5H_8BN, 120 mM in 10% methanol) and then mixed with light and heavy formaldehyde (15% in water, 1 μL), respectively. The reaction solutions were kept at room temperature for 1 h. After quenched by adding 1 μL of ammonium bicarbonate solution (0.2 M), the two samples of fed and unfed crabs were mixed at equal volume. After desalting by Ziptip C_{18}, the sample was re-suspended in 20 μL of water for further MS analysis.

6.2.4 Top-down and bottom-up MS on Q Exactive Orbitrap mass spectrometer

For top-down analysis, 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 C_{18} ZipTip and resuspended in 10 μL of water containing 0.2% formic acid. On-line top-down MS was carried out on Waters nanoAcquity ultra performance LC system coupled to a Q Exactive quadrupole orbitrap mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). A 0.5 μL of peptide sample was injected and loaded onto the Waters Symmetry C_{18} trap column (180 μm x 20 mm, 5 μm) using 97% mobile phase A (0.1% of formic acid in water) and 3% mobile phase B (0.1% of formic acid in ACN) at a flow rate of 5 μL/min for 3 min. A Waters BEH 300Å C_{18} reversed phase capillary column (150 mm x 75 μm, 1.7 μm) was used for separation. The gradient started from 3 to 10% B during the first 5 min, increased to 55% B in the next 65 min, then was kept at 90% B for 20 min. Typical mass spectrometric conditions were as follows: spray voltage, 2.8 kV; no sheath and auxiliary gas flow; heated capillary temperature, 275 °C; normalized

high-energy collision dissociation (HCD) collision energy 30%. The Q Exactive instrument was operated in targeted MS/MS mode with an inclusion list containing the targeted mass of the CHHs. The settings are: resolution 70, 000; automatic gain control $2e^5$; maximum ion injection time, 100 ms; isolation window, 6 *m/z*; fixed first mass, 100 *m/z*. All MS/MS spectra were processed with Xtract CI-3.0 Software (Thermo Scientific Inc., Bremen, Germany) using an 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 with of 10 ppm of mass error tolerance. The assigned ions were manually validated to ensure the quality of assignments.

For bottom-up analysis, a 1 μL of crude tissue extract was treated with DTT and IAA followed by tryptic digestion according to previous reports [14, 23-25]. The tryptic digest was analzyed on the same LC-MS system as described above with different LC gradient and MS paratmeters. The LC gradient started from 3 to 10% B during the first 5 min, increased to 45% B in the next 65 min, then was kept at 90% B for 20 min. The Q Exactive mass spectrometer was operated in a data-dependent mode (top 10) with the following settings: Full MS, resolution 70,000, AGC $1xe^6$, maximum injection time 100 ms, scan range 200-2000 *m/z*; dd-MS^2, resolution 17,500, AGC $1xe^5$, maximum injection time 100 ms, loop count 10, isolation window 3.0 *m/z*, fixed fist mass 100 *m/z*, nomolized collision energy 30%.

6.2.5 Off-line top-down MS on Orbitrap Elite mass spectrometer

The purified peptide was re-suspended in 50% of ACN with 0.1% formic acid and directly infused into Orbitrap Elite mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) via syringe pump at a flow rate of 0.3 μL/ml. The peptide parent ions were selected with an isolation window of 6 Da and submitted to HCD fragmentation with 30% of normalized collision energy.

The data was acquired for 1000 scans to generate a high quality HCD MS/MS spectrum with a resolving power of 60,000. The data processing method was the same as off-line top-down method described above.

6.2.6 Database search and data interpretation by Mascot and PEAKS

Peptides were identified by searching against a NCBInr protein database using the Mascot v2.1 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 0.1 Da and 0.1 Da for monoisotopic mass, respectively. Peptide charge states include 1+, 2+ and 3+ charged peptides.

The data acquired by bottom-up strategy on Q Exactive were processed by Peaks Studio 5.2 (BSI, Canada) for spectral interpretation [19, 26, 27]. First, data was refined and subjected to Auto *De Novo* program for sequencing with the mass tolerance at 0.1 Da. Trypsin was specified for enzyme cleavage. Variable PTMs included C-terminal amidation, pyro-Glu and methionine oxidation, and fixed PTM is was carboxylmethyl on cysteine. Second, the homologous CHH sequences obtained from Mascot search were constructed into the small database I. The data from Auto *De Novo* was subsequently subjected to Spider Homology Search against the small database I with the same settings as above. Third, the predicted tryptic peptides were assembled to generate a predicted CHH sequence and the small database II. The data from Auto *De Novo* was searched against the small database II to match all the tryptic peptides in the sample. The settings were also the same as above.

6.3 Results and discussion

6.3.1 Identification of novel CHHs in Dungeness crab by a refined peptide discovery pipeline

The Dungeness crab *C. magister* is a favored animal model for neurobiologists due to their relativiely simple and well-organized nervous system [28]. However, its genome has not been sequenced yet and thus no protein/cDNA database is currently available. Although the presence of CHH of Dungeness crab sinus gland has been documented by immunocytochemistry using an antibody staining against a homologous CHH of *Cancer pagurus* [29], the exact amino acid sequence has not been determined. In this work, we refine our peptide discovery pipeline to determine the full amino acid sequence of the Dungeness crab CHHs and study its dynamic changes in response to food intake.

CHH-family neuropeptides have two unique features, namely the presence of three disulfide bonds and molecular weights ranging from 8 to 10 kDa [13, 16], which are criteria for screening peptide candidates. The tissue extract of Dungeness crab sinus glands was treated with DTT and kept in acidic solvent to prevent the free sulfhydryl from reforming disulfide bonds. The original and reduced tissue extract samples were respectively analyzed by LC-MS/MS under data-dependent mode on a Q Exactive hybrid quadrupole-Orbitrap mass spectrometer [17, 18, 30] for HR/AM measurement. By comparing the mass differences of the detected peptides between the two runs, two CHH candidates (MWs, 8453.824 and 8470.848 Da) were found by observing 6 Da of mass increase due to DTT reduction as shown in Figure 1. The 17 Da of mass difference between the two peptides could be attributed to N-terminal pyro-Gln modification. The two candidates were tentatively named as $^{Cam\text{-}}$CHH$_{pQ}$ and $^{Cam\text{-}}$CHH. We then performed *de novo* sequencing of $^{Cam\text{-}}$CHH$_{pQ}$.

To find homologous preprohormone and predict putative sequence, we performed

bottom-up sequencing on the CHH candidates. They were purified by RPLC and digested by trypsin, respectively. The tryptic digests were analyzed by LC-MS/MS on Q Exactive mass spectrometer, followed by Mascot searching against a NCBI database. The first hit was *Cancer productus* CHH-preprohormone [31] (Figure S1) with Mascot score of 1651 and homologous sequence coverage of 40%. Subsequently, we used this homologous sequence as a reference to search the bottom-up data by PEAKS Spider homology search [19, 27]. One set of tryptic peptides, $^{Cam\text{-}}$CHH$_{pQ}$[1-8], [32-40], [41-50], [32-50], and [41-67] in Table 1, were determined with significant Spider scores [19, 27], each of which contains contained one substituted residue in comparison to the homologous sequence. Figure 2 A, B and C show three representative MS/MS spectra of these tryptic peptides (others spectra shown in Figure S2-S8). In addition, there are other tryptic peptides, $^{Cam\text{-}}$CHH$_{pQ}$[14-31], [51-64], and [51-67] in Table 1, were also matched, which contained the same residues as the homologous sequence. Subsequently, the residues ^{6}Ser, ^{37}Ser, and ^{47}Val were substituted by the newly identified ones, ^{6}Thr, ^{37}Thr, and ^{47}Leu (Figure 2D) to generate a putative sequence. This sequence was used to search the bottom-up data again by means of PEAKS database search [19], and all the tryptic peptides were identified with significant -10logP scores as shown in Table 1 except for the unidentified residues ^{68}Xxx-^{72}Xxx. The sequence coverage of bottom-up sequencing was illustrated in Figure 3A. The remaining residues ^{68}Xxx-^{72}Xxx was further identified by top-down HCD fragmentation of the intact CHH peptide (Figure 2E), where the purified peptide was off-line injected into the Orbitrap Elite mass spectrometer. One set of large b ions, b_{54}~b_{71} deriving from enhanced cleavage of C-terminal amide bonds, were observed, indicative of the residues ^{55}Xxx-^{72}Xxx. Finally, the DTT-reduced tissue extract was submitted to Q Exactive by means of LC-targeted MS/MS to acquire a high quality of

top-down HCD spectrum, enabling validation of the newly identified sequence. Figure 3B shows the fragmentation map of this peptide, and with a 53% of sequence coverage and 65% of amide bond cleavage are being attained, confidently confirming the sequences of $^{Cam\text{-}}CHH_{pQ}$ as identified above with great confidence. It should be noted that the experiment in Figure 2E can also be performed on Q Exactive by either off-line or on-line top-down MS/MS. The reason of using Orbitrap Elite here is to demonstrate that our approach is applicable to diverse Orbitrap instruments. In addition, leucine and isoleucine in $^{Cam\text{-}}CHH_{pQ}$ are assigned consistent to the homologous sequence, since mass spectrometry lacks the capability to differentiate them. Similarly, bottom-up identification of the unmodified $^{Cam\text{-}}$CHH (MW 8470.848, no pyro-Gln) exported the same tryptic peptides as $^{Cam\text{-}}CHH_{pQ}$, except for observation of N-terminal tryptic segments without pyro-Gln. In addition, top-down HCD fragmentation of $^{Cam\text{-}}$CHH produced one set of large b ions without pyro-Gln at the N-terminus (data not shown). These results confirmed the identity of $^{Cam\text{-}}$CHH.

In our previous studies, we established a MS-based platform for confident *de novo* sequencing of large neuropeptides [14]. This multi-faceted pipeline, while effective, was time consuming due to the need for manual inspecting the MS/MS spectra of tryptic peptides to locate the substituted residues in the target peptides that were distinct from the homologous sequences. In this study, we take advantage of Orbitrap instruments with capability of HR/AM measurement and HCD technique [17], which greatly improves the quality of MS/MS spectra for both bottom-up and top-down analysis. By processing the data with PEAKS Spider homology search, the tryptic peptides containing substituted residues can be predicted in an automatic and confident manner. The modification of this aspect of the methodology greatly enhanced the throughput of the peptide discovery pipeline. Previously,

we also tested the data acquired by mass spectrometers with TOF analyzers to perform PEAKS Spider Homology Search [32], but nothing was obtained presumably due to lower mass accuracy of TOF analyzers and fragmentation efficiency of CID. Zhang and co-works [26, 27] have successfully applied this method to characterize mammalian neuropeptides with molecular weight less than 5 kDa. To the best of our knowledge, our study represents the first application of this method to sequence large neuropeptides.

6.3.2 Isotopic labeling-assisted quantitative top-down MS for monitoring the changes of CHHs at endogenous level

To monitor the relative abundance of the newly identified CHHs in neural tissue, we developed an isotopic labeling-assisted quantitative top-down strategy. The reductive amination using formaldehyde (FH) and cyanoboroydride has been reported as a fast and simple reaction that is highly selective to N-termini and lysine side chain [20, 21, 33, 34]. This reaction has been reported to label intact proteins resulting in high labeling yield [33, 35, 36]. The $^{Cam\text{-}}$CHH$_{pQ}$ (Figure 3) contains two lysines and the $^{Cam\text{-}}$CHH has an extra free N-terminus, which are active sites for dimethyl labeling. Two aliquots of crude tissue extracts were respectively treated with light and heavy FH and mixed at an equal ratio, followed by LC-MS analysis (Figure 3A). In the full scan spectrum (Figure 3B), the isotopic peaks of light- and heavy-FH labeled $^{Cam\text{-}}$CHH$_{pQ}$ were significantly overlapped, leading to inaccurate and ambiguous quantitative information. Herein, we employed a top-down MS-based strategy that both light- and heavy-FH labeled $^{Cam\text{-}}$CHH$_{pQ}$ were selected in C-trap [17] under LC-targeted MS/MS mode with a 15 Da of mass isolation window and subsequently fragmented in HCD cell [17], followed by parallel mass measurement in Orbitrap. In the resulting top-down MS/MS spectrum (Figure 3C), one set of FH-labeled y ions, $_{FH\text{-}}y_9$~ $_{FH\text{-}}y_{17}$,

were observed, which could be used as quantitative reporter ions indicative of relative abundance ratios of the two labeled samples. Figures 3D, E and F illustrate the zoom-in spectra of $_{FH}$-y_9 ions at light to heavy ratios of 1:1, 1:2 and 1:10, demonstrating a straightforward and accurate quantitation approach. Finally, $_{FH}$-y_9 and $_{FH}$-y_{13} were chosen as reporter ions because of no background interference around them. In addition, the accurate quantitation of large neuropeptides benefits from the paralleled detection [37] of Q Exactive instrument that allows all the fragment ions of light- and heavy-labeled large neuropeptides to be simultaneously measured in the Orbitrap cell, which cannot be realized by TOF instruments. Coon and co-workers [37] reported that this paralleled detection was able to enhance run-to-run reproducibility and measurement accuracy in targeted proteomics. Similarly, we take advantage of this unique feature from advancement of instrumentation and introduced the isotopic labeling-assisted quantitative top-down strategy into the domain of large neuropeptide analysis.

We also examined the potential deuterium effect [20] on RPLC retentions of labeled CHHs. Figure S9 shows the extract ion chromatograms of the light- and heavy-labeled ^{Cam}-CHH_{pQ} with the identical retention time at 47.19 min, suggesting that the deuterium effect is negligible in isotopic labeling of large peptides. To achieve a complete labeling by FH, we optimized the solvent component for labeling reaction and found that the use of 90% of ACN resulting in better than 95% of labeling efficiency presumably due to denaturation of large peptides in the high organic solvent.

6.3.3 Alteration of CHH expression levels in response to feeding

Feeding behavior is crucial for animal survival and is a fundamental aspect of energy homeostasis [3, 13]. Our previous study has observed expression level changes of multiple

small neuropeptides in brain and pericardial organ upon feeding [6, 7]. In this study, we examined the relative changes of large CHH peptide expression in the sinus gland from three groups of fasting and satiated crabs. The tissue samples were prepared according to the workflow shown in Figure 5A, followed by differential analysis with the newly developed quantitative top-down tandem MS approach. Significant increases of both ^{Cam}-CHH and ^{Cam}-CHH$_{pQ}$ abundances were observed (Figure 5B), showing 3 to 4 folds of elevation after food intake. In order to examine the reliability of our experiment, two unfed crabs from control groups were respectively labeled with light- and heavy-FH and measured by the top-down strategy. A 1:1 ratio of light- to heavy-labeled reporter ions was detected (data not shown).

CHH is a multifunctional hormone and plays important roles throughout the life cycle of crustaceans [38]. The primary role of CHH in decapods is to induce hyperglycemia and hyperlipidemia to meet the energy requirements of the organs and tissues [13]. It has been reported that at membrane level CHH is able to stimulate the GMP synthesis by activating guanylate cyclase [39]; and at the tissue level, CHH inhibits glycogen synthase and activates glycogen phosphorylase, thus improving glycogenolysis in muscle and midgut gland [40]. In addition to the downstream regulatory pathways, the secretion and release of CHH is subjected modulation by a feedback control system. D-glucose inhibits the release of CHH by activating a K^+ current that induces hyperpolarization of CHH-secreting cells [41].

Our quantitative data (Figure 5B) indicates that the CHH level in crustacean sinus gland is dramatically enhanced after feeding. Presumably, one of the reasons is due to the increasing glucose level after feeding that is commonly observed in most of organisms after eating [42]. The elevation of glucose level could initiate the feedback control system is initiated to inhibit the release of CHH, thus causing accumulation of CHH in the sinus glands.

Our results suggest that the tissue level of CHH from neuroendocrine organ can be altered and associated with feeding behavior. To probe the potential roles of CHH played in regulation of feeding, further experiments will need to be designed and conducted. In this study, the utility and efficiency of the newly developed quantitative top-down strategy has been demonstrated for monitoring expression changes of CHH at the tissue level. More systematic and in-depth investigation of CHH changes in both tissue expression and circulation in response to feeding coupled with physiological studies will provide more complete understanding of the functional roles of this important class of peptide hormones.

6.4 Concluding remarks

This study demonstrates the utility of an improved peptide discovery pipeline for qualitative and quantitative analysis of large neuropeptide hormones and related physiological study. First, the refinement of qualitative analysis targets at a significant bottleneck that hindered the overall throughput of homologous peptide discovery. By incorporating the HR/AM capability of Orbitrap instruments with PEAKS spider homology search, the substituted residues of target large peptides can be accurately and rapidly predicted according to homologous sequences. Second, because the traditional label-free and isotopic labeling methods cannot provide accurate quantitation for large neuropeptides, we developed a novel quantitative top-down tandem MS strategy with assistance of isotopic labeling, greatly improving the quantitation accuracy. Third, by applying the qualitative and quantitative approaches to peptide discovery in Dungeness crab, novel CHHs were confidently *de novo* sequenced without assistance of genome information, and quantitative monitoring of the abundance changes at the tissue level was achieved. Forth, the quantitative analysis of CHH peptides in response to food intake indicates alteration of CHH abundances

in the sinus glands upon feeding suggesting potential link of CHH expression and secretion with feeding behavior. Collectively, this peptide discovery pipeline highlights the ability for qualitative and quantitative analysis of large neuropeptides and is expected to aid in future research efforts to unravel large complex signaling peptides employed by other biological systems.

6.5 Acknowledgements

This work is supported in part by the National Institutes of Health (NIH) grant (R01DK071801 to LL) and the National Science Foundation grant (CHE-0957784 to LL). We would like to acknowledge NIH shared instrument program for funding the instrument purchase (S10 RR029531). LL acknowledges an H. I. Romnes Faculty Research Fellowship. We are also grateful to the UW School of Pharmacy Analytical Instrumentation Center for access to Q Exactive and Elite Orbitrap mass spectrometers.

6.6 References

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

[2] Frese CK, Boender AJ, Mohammed S, Heck AJ, Adan RA, Altelaar AF. Profiling of diet-induced neuropeptide changes in rat brain by quantitative mass spectrometry. Anal Chem. 2013;85:4594-604.

[3] Morton GJ, Cummings DE, Baskin DG, Barsh GS, Schwartz MW. Central nervous system control of food intake and body weight. Nature. 2006;443:289-95.

[4] Fricker LD. Analysis of mouse brain peptides using mass spectrometry-based peptidomics: implications for novel functions ranging from non-classical neuropeptides to microproteins. Molecular bioSystems. 2010;6:1355-65.

[5] Jing J, Vilim FS, Horn CC, Alexeeva V, Hatcher NG, Sasaki K, et al. From hunger to satiety: reconfiguration of a feeding network by Aplysia neuropeptide Y. J Neurosci. 2007;27:3490-502.

[6] 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.

[7] Hui L, Zhang Y, Wang J, Cook A, Ye H, Nusbaum MP, et al. Discovery and functional study of a novel crustacean tachykinin neuropeptide. ACS Chem Neurosci. 2011;2:711-22.

[8] Yu N, Nachman RJ, Smagghe G. Characterization of sulfakinin and sulfakinin receptor and their roles in food intake in the red flour beetle Tribolium castaneum. Gen Comp Endocrinol. 2013;188:196-203.

[9] Matsuda K, Sakashita A, Yokobori E, Azuma M. Neuroendocrine control of feeding behavior and psychomotor activity by neuropeptideY in fish. Neuropeptides. 2012;46:275-83.

[10] Lee KS, You KH, Choo JK, Han YM, Yu K. Drosophila short neuropeptide F regulates food intake and body size. J Biol Chem. 2004;279:50781-9.

[11] Verri T, Mandal A, Zilli L, Bossa D, Mandal PK, Ingrosso L, et al. D-glucose transport in decapod crustacean hepatopancreas. Comp Biochem Physiol A Mol Integr Physiol. 2001;130:585-606.

[12] Woods SC, Lutz TA, Geary N, Langhans W. Pancreatic signals controlling food intake; insulin, glucagon and amylin. Philos Trans R Soc Lond B Biol Sci. 2006;361:1219-35.

[13] 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.

[14] 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.

[15] 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.

[16] Chung JS, Zmora N, Katayama H, Tsutsui N. Crustacean hyperglycemic hormone (CHH) neuropeptidesfamily: Functions, titer, and binding to target tissues. Gen Comp Endocrinol. 2010;166:447-54.

[17] Michalski A, Damoc E, Hauschild JP, Lange O, Wieghaus A, Makarov A, et al. Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. Mol Cell Proteomics. 2011;10:M111 011015.

[18] Zubarev RA, Makarov A. Orbitrap mass spectrometry. Anal Chem. 2013;85:5288-96.
[19] 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.
[20] Hsu JL, Huang SY, Chow NH, Chen SH. Stable-isotope dimethyl labeling for quantitative proteomics. Anal Chem. 2003;75:6843-52.
[21] 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.
[22] Iliuk A, Galan J, Tao WA. Playing tag with quantitative proteomics. Anal Bioanal Chem. 2009;393:503-13.
[23] Hu S, Xie Y, Ramachandran P, Ogorzalek Loo RR, Li Y, Loo JA, et al. Large-scale identification of proteins in human salivary proteome by liquid chromatography/mass spectrometry and two-dimensional gel electrophoresis-mass spectrometry. Proteomics. 2005;5:1714-28.
[24] Fang L, Kaake RM, Patel VR, Yang Y, Baldi P, Huang L. Mapping the protein interaction network of the human COP9 signalosome complex using a label-free QTAX strategy. Mol Cell Proteomics. 2012;11:138-47.
[25] Yu Y, Leng T, Yun D, Liu N, Yao J, Dai Y, et al. Global analysis of the rat and human platelet proteome - the molecular blueprint for illustrating multi-functional platelets and cross-species function evolution. Proteomics. 2010;10:2444-57.
[26] Ranc V, Petruzziello F, Kretz R, Argandona EG, Zhang X, Rainer G. Broad characterization of endogenous peptides in the tree shrew visual system. J Proteomics. 2012;75:2526-35.
[27] 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.
[28] Verley DR, Doan V, Trieu Q, Messinger DI, Birmingham JT. Characteristic differences in modulation of stomatogastric musculature by a neuropeptide in three species of Cancer crabs. J Comp Physiol A Neuroethol Sens Neural Behav Physiol. 2008;194:879-86.
[29] Hsu YW, Messinger DI, Chung JS, Webster SG, de la Iglesia HO, Christie AE. Members of the crustacean hyperglycemic hormone (CHH) peptide family are differentially distributed both

between and within the neuroendocrine organs of Cancer crabs: implications for differential release and pleiotropic function. J Exp Biol. 2006;209:3241-56.

[30] Kharchenko A, Vladimirov G, Heeren RM, Nikolaev EN. Performance of Orbitrap mass analyzer at various space charge and non-ideal field conditions: simulation approach. J Am Soc Mass Spectrom. 2012;23:977-87.

[31] Hsu YW, Weller JR, Christie AE, de la Iglesia HO. Molecular cloning of four cDNAs encoding prepro-crustacean hyperglycemic hormone (CHH) from the eyestalk of the red rock crab Cancer productus: identification of two genetically encoded CHH isoforms and two putative post-translationally derived CHH variants. Gen Comp Endocrinol. 2008;155:517-25.

[32] Jia C, Lietz CB, Ye H, Hui L, Yu Q, Yoo S, et al. A multi-scale strategy for discovery of novel endogenous neuropeptides in the crustacean nervous system. J Proteomics. 2013:doi: 10.1016/j.jprot.2013.06.021.

[33] Kovanich D, Cappadona S, Raijmakers R, Mohammed S, Scholten A, Heck AJ. Applications of stable isotope dimethyl labeling in quantitative proteomics. Anal Bioanal Chem. 2012;404:991-1009.

[34] Sun Z, Qin H, Wang F, Cheng K, Dong M, Ye M, et al. Capture and dimethyl labeling of glycopeptides on hydrazide beads for quantitative glycoproteomics analysis. Anal Chem. 2012;84:8452-6.

[35] Dottavio-Martin D, Ravel JM. Radiolabeling of proteins by reductive alkylation with [14C]formaldehyde and sodium cyanoborohydride. Anal Biochem. 1978;87:562-5.

[36] Means GE, Feeney RE. Reductive alkylation of proteins. Anal Biochem. 1995;224:1-16.

[37] Peterson AC, Russell JD, Bailey DJ, Westphall MS, Coon JJ. Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. Mol Cell Proteomics. 2012;11:1475-88.

[38] Aquiloni L, Giulianini PG, Mosco A, Guarnaccia C, Ferrero E, Gherardi F. Crustacean hyperglycemic hormone (cHH) as a modulator of aggression in crustacean decapods. PLoS One. 2012;7:e50047.

[39] Goy MF. Activation of membrane guanylate cyclase by an invertebrate peptide hormone. J Biol Chem. 1990;265:20220-7.

[40] Sedlmeier D. The crustacean hyperglycemic hormone (CHH) releases amylase from the crayfish midgut gland. Regul Pept. 1988;20:91-8.

[41] Glowik RM, Golowasch J, Keller R, Marder E. D-glucose-sensitive neurosecretory cells of the crab Cancer borealis and negative feedback regulation of blood glucose level. J Exp Biol. 1997;200:1421-31.

[42] Sanchez-Paz A, Garcia-Carreno F, Muhlia-Almazan A, Peregrino-Uriarte AB, Hernandez-Lopez J, Yepiz-Plascencia G. Usage of energy reserves in crustaceans during starvation: status and future directions. Insect Biochem Mol Biol. 2006;36:241-9.

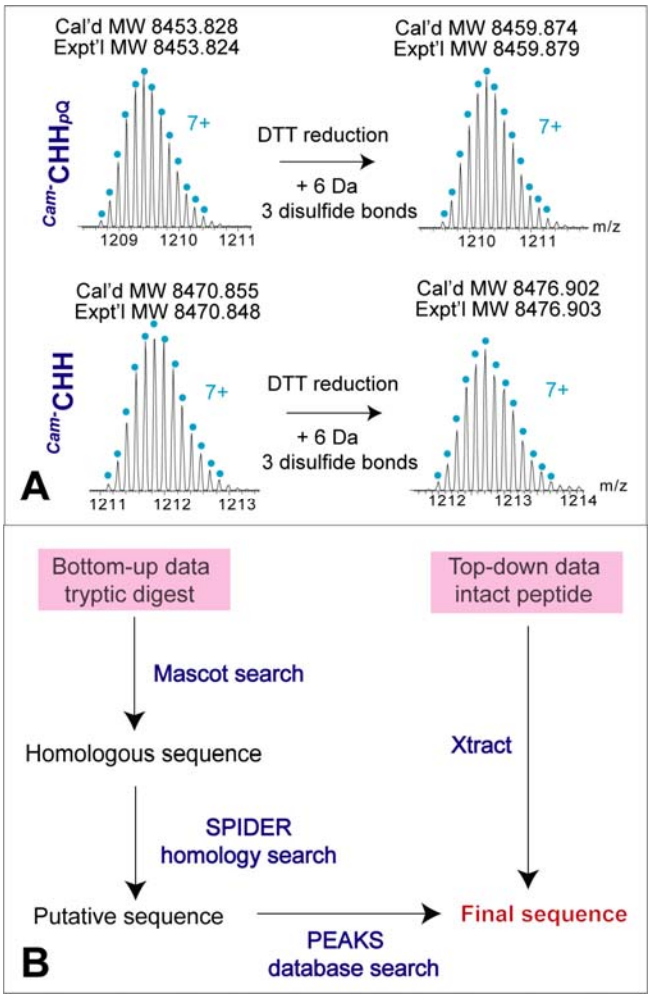


Figure 1. Discovery and identification of novel CHHs in the sinus gland of Dungeness crabs. (A) High resolution isotopic distributions of intact and DTT-reduced CHHs. (B) Refined workflow for *de novo* sequencing of CHHs.

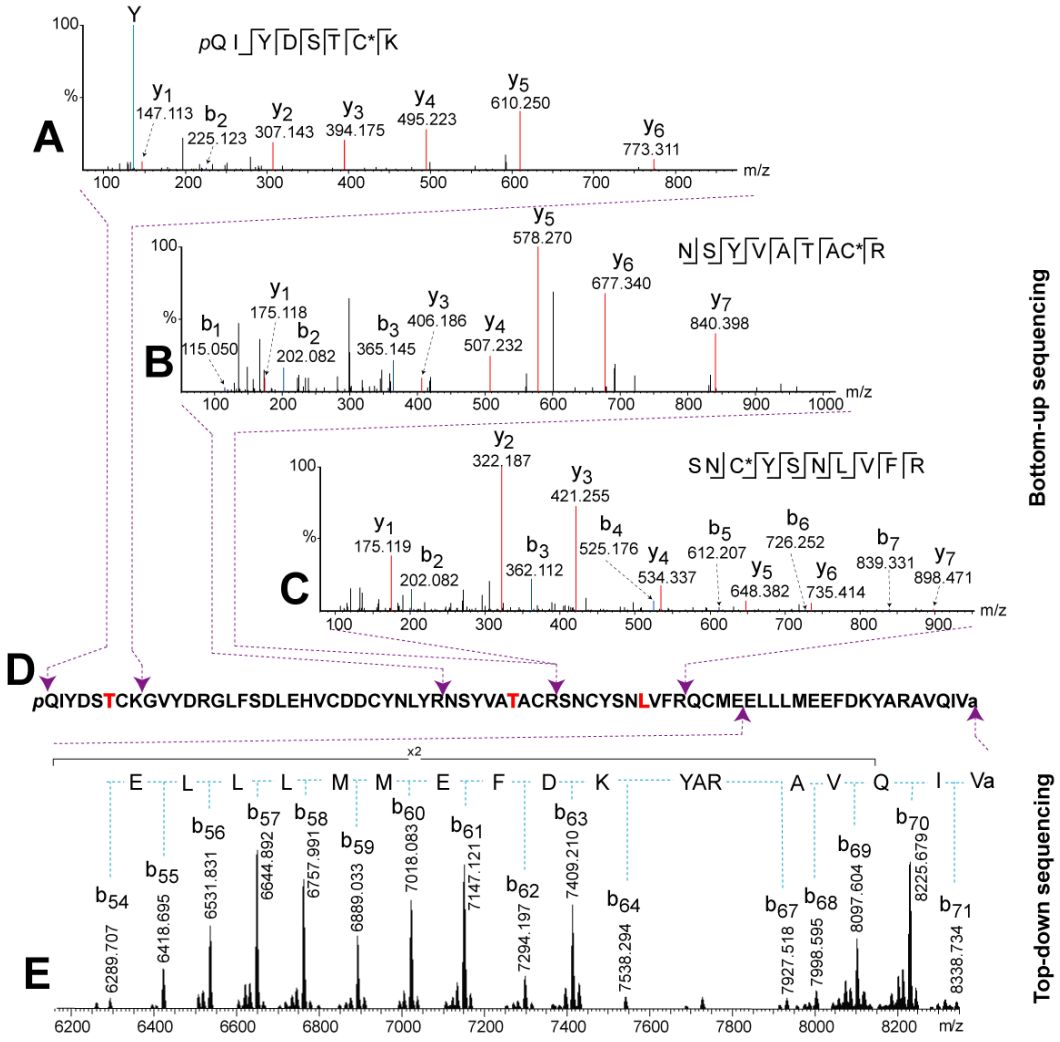


Figure 2. Mass spectral sequencing of ^{Cam}-CHH$_{pQ}$. MS/MS spectra of tryptic peptides [1-8] (A), [32-40] (B) and [41-50] (C). Sequence assembly of CHH (D). Top-down HCD MS/MS spectrum of intact CHH (E). The spectra in Panel A, B, and C are exported from PEAKS after deconvolution and deisotoping. The spectrum in Panel E is processed by Xtract for deconvolution. The substituted residues of ^{Cam}-CHH$_{pQ}$ are highlighted with red fond in panel D. *, carbamidomethyl; *p*Q, pyro-Gln; a, amidation.

Figure 3. Sequence coverage map (A) and fragmentation maps (B) of ^{Cam}-CHH$_{pQ}$ respectively obtained by bottom-up and top-down sequencing strategies. In panel A, the blue bars below sequence represent tryptic peptide sequences being detected and identified; and the bond letters are identified residues. *p*Q, pyro-Gln; a, amidation.

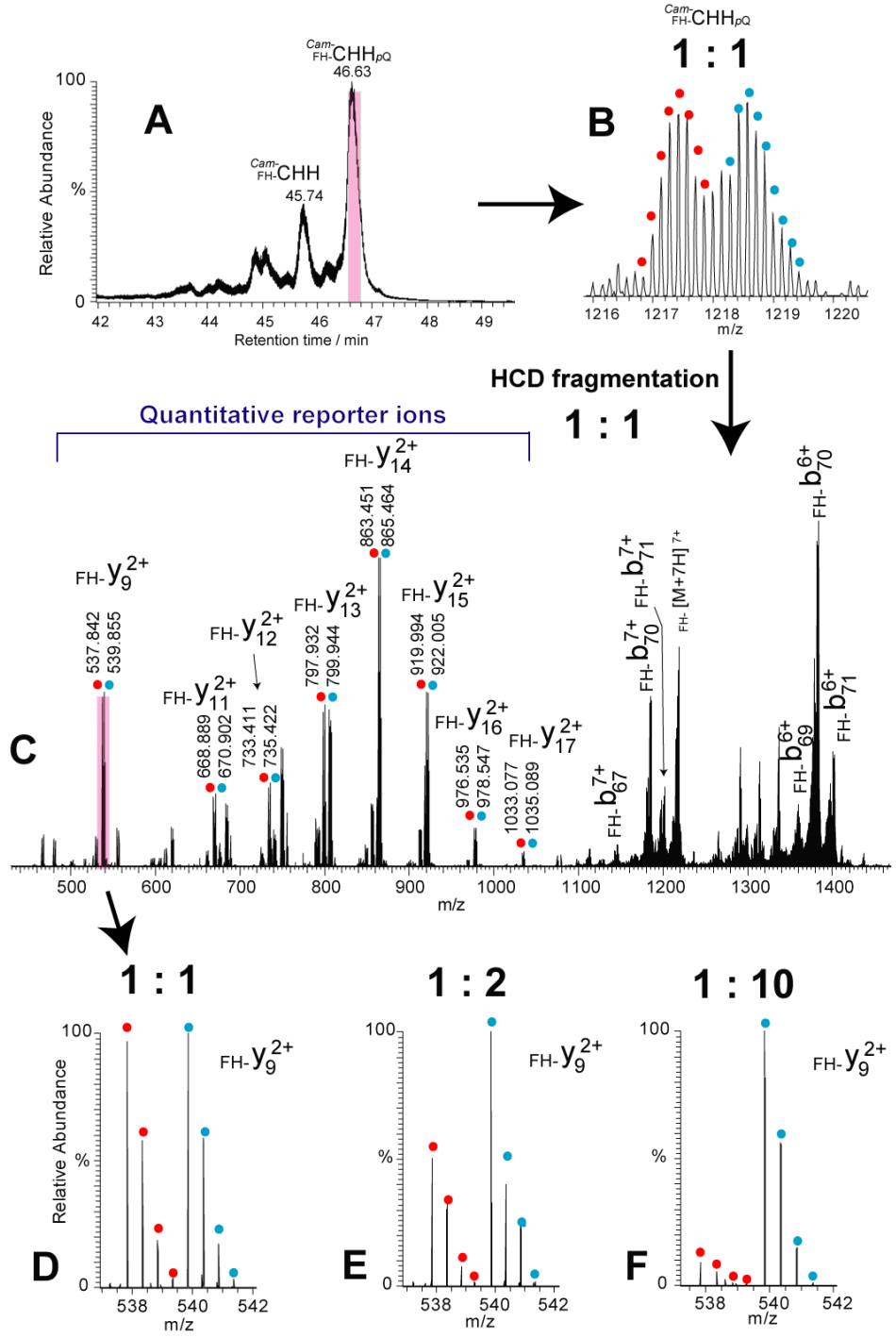


Figure 4. Isotopic labeling-assisted quantitative top-down MS/MS strategy. (A) Total ion chromatogram of sinus gland tissue extract treated with FH-labeling. (B) Isotopic distributions of light and heavy FH-labeled $^{Cam\text{-}}$CHH$_{pQ}$. (C) Top-down HCD MS/MS spectrum of light and heavy FH-labeled $^{Cam\text{-}}$CHH$_{pQ}$ represented by red and blue dots, respectively. Zoom-in spectra showing light and heavy FH-labeled y_9^{2+} ions at ratios of 1:1 (D), 1:2 (E), and 1:10 (F).

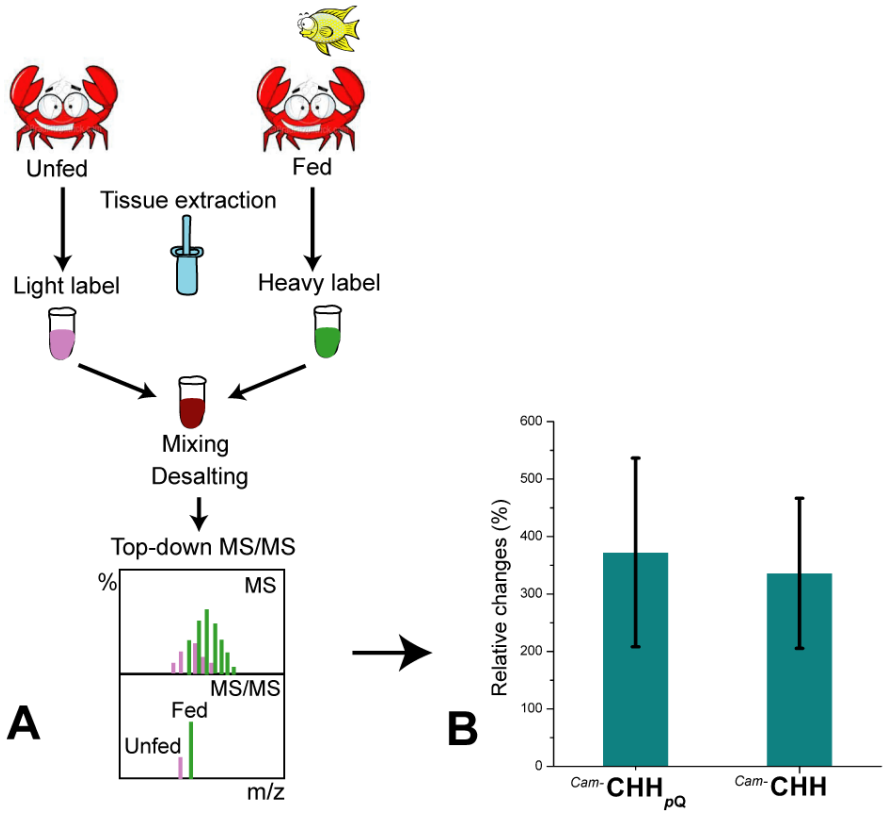


Figure 5. Monitoring of relative changes of CHH expression in the sinus glands of Dungeness crabs in response to food intake. (A) The workflow of feeding and quantitation experiments. (B) Significant changes of ^{Cam}-CHH and ^{Cam}-CHH$_{pQ}$ in the sinus gland upon feeding. Error bars indicate s.e.m. calculated from three feeding experiments (p-value < 0.05).

Table 1. Tryptic peptides identified in bottom-up sequencing of ^{Cam}-CHH$_{pQ}$.

Tryptic peptides [a]	Position	Spider score [b]	-10logP score [c]	mass	*m/z*	ppm	Fig.#
*p*QIYDS**T**C*K	CHH[1-8]	24.00	98.49	996.422	499.218	-1.8	Fig.2A
*p*QIYDS**T**C*KGVYDR	CHH[1-13]	–	132.98	1586.704	529.908	-0.3	Fig.S2
*p*QIYDS**T**C*KGVYDRGLF SDLEHVC*DDC*YNLYR	CHH[1-31]	–	128.95	3843.660	961.918	-4.8	Fig.S3
GLFSDLEHVC*DDC*YNL YR	CHH[14-31]	60.00	161.28	2274.967	1138.488	-2.4	Fig.S4
NSYVA**T**AC*R	CHH[32-40]	25.50	116.61	1040.471	521.240	-5.3	Fig.2B
SNC*YSN**L**VFR	CHH[41-50]	27.75	115.28	1258.576	420.532	-2.8	Fig.2C
NSYVA**T**AC*RSNC*YSNL VFR	CHH[32-50]	55.00	152.35	2281.037	761.351	-2.0	Fig.S5
SNC*YSN**L**VFRQC*MEEL LLMEEFDKYAR	CHH[41-67]	41.45	128.47	3444.561	862.146	-1.4	Fig.S6
QC*MEELLLMEEFDK	CHH[51-64]	44.50	125.56	1813.794	907.901	-2.6	Fig.S7
QC*MEELLLMEEFDKYA R	CHH[51-67]	54.00	154.11	2203.995	1103.003	-1.8	Fig.S8

[a] The substituted residues are underlined. *p*, pyro-Gln. * carbamidomethylation.
[b] The score is obtained from PEAKS Spider Homology Search. Some tryptic peptides are not identified in this run.
[c] The score is obtained from PEAKS database search.

Supplementary Materials for Chapter 6

1 MLTSRTLPTI ILGVLCIYLS TIPNAHARSA QGMGKMEHLL ASYRGALESN
51 TPTGDLPGGL VHPVEKR**QIY DSSCKGVYDR GLFSDLEHVC DDCYNLYR**NS
101 YVASACR**SNC YSNVVFRQCM EELLLMEEFD KYAR**AVQIVG KKK

Figure S1. Homologous preprohormone of *Cancer productus* CHH identified by Mascot search. Matched peptides are shown in bold fonts.

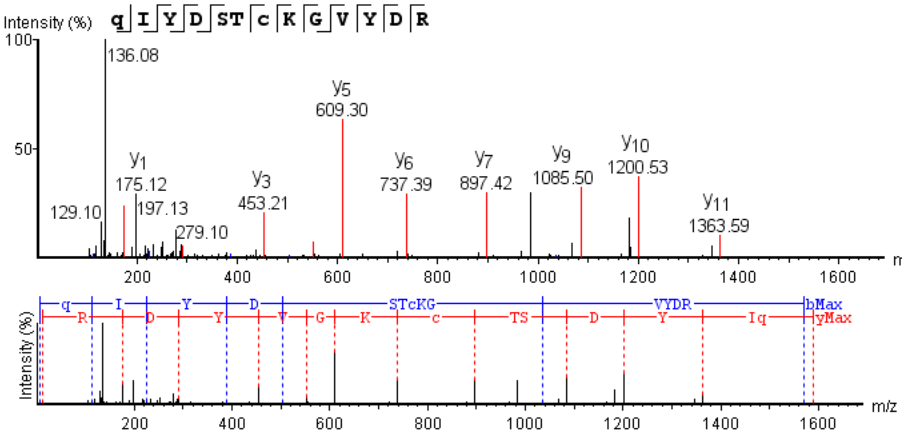


Figure S2. Annotated MS/MS spectrum of tryptic peptide CHH[1-13]. Exported from PEAKS. q, pyro-Gln; c, carbamidomethylated cysteine.

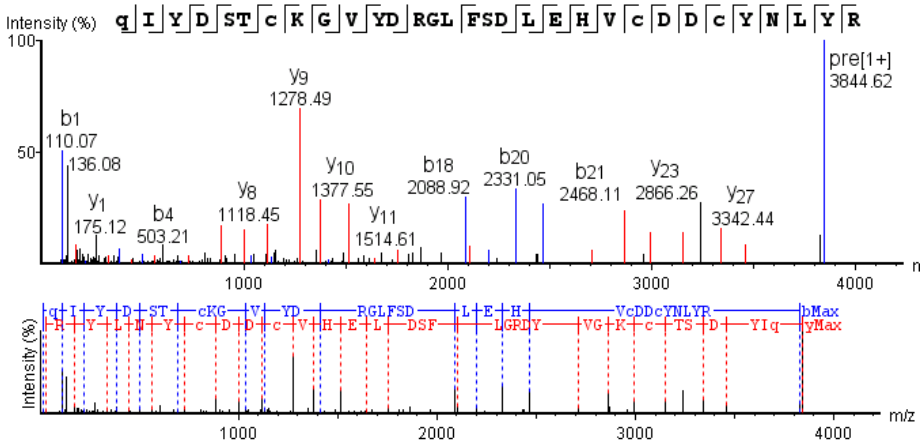


Figure S3. Annotated MS/MS spectrum of tryptic peptide CHH[1-31]. Exported from PEAKS. q, pyro-Gln; c, carbamidomethylated cysteine.

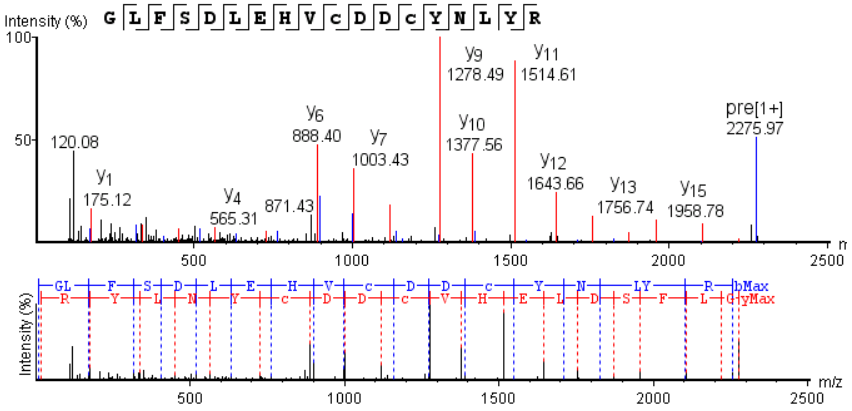


Figure S4. Annotated MS/MS spectrum of tryptic peptide CHH[14-31]. Exported from PEAKS. c, carbamidomethylated cysteine.

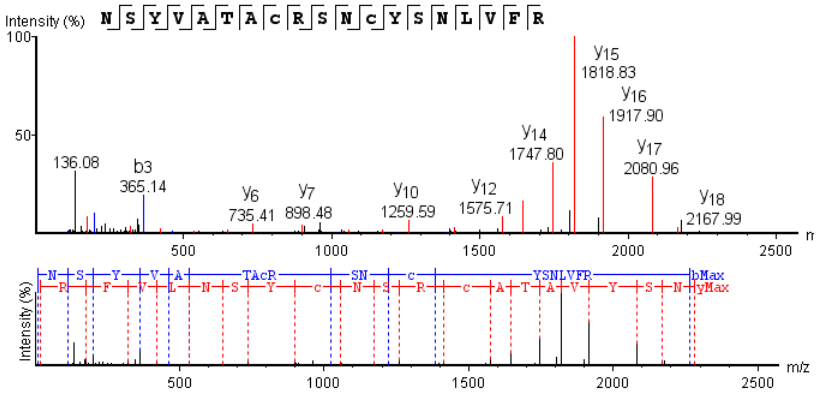


Figure S5. Annotated MS/MS spectrum of tryptic peptide CHH[32-50]. Exported from PEAKS. c, carbamidomethylated cysteine.

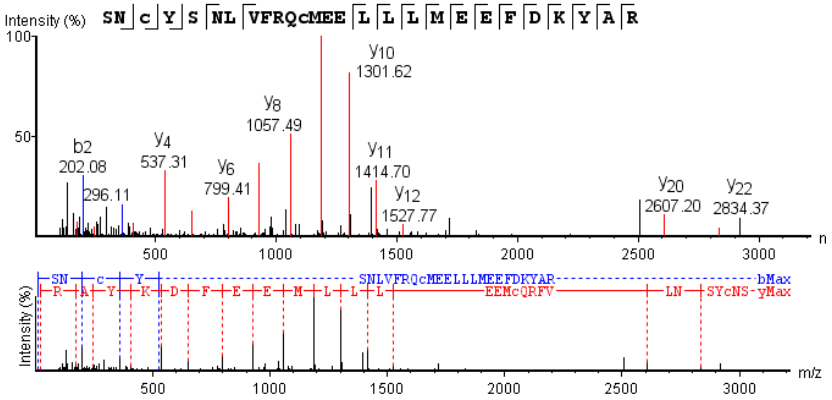


Figure S6. Annotated MS/MS spectrum of tryptic peptide CHH[41-67]. Exported from PEAKS. c, carbamidomethylated cysteine.

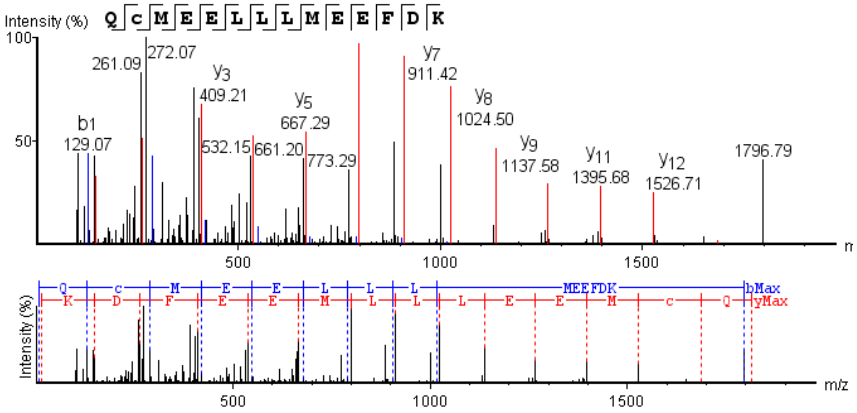


Figure S7. Annotated MS/MS spectrum of tryptic peptide CHH[51-64]. Exported from PEAKS. c, carbamidomethylated cysteine.

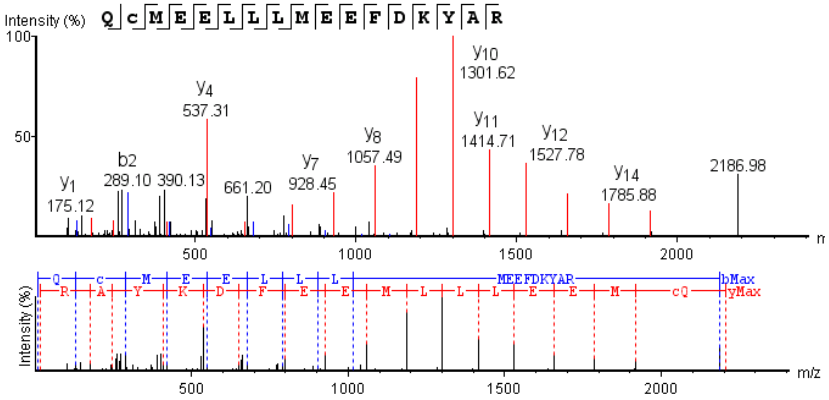


Figure S8. Annotated MS/MS spectrum of tryptic peptide CHH[51-67]. Exported from PEAKS. c, carbamidomethylated cysteine.

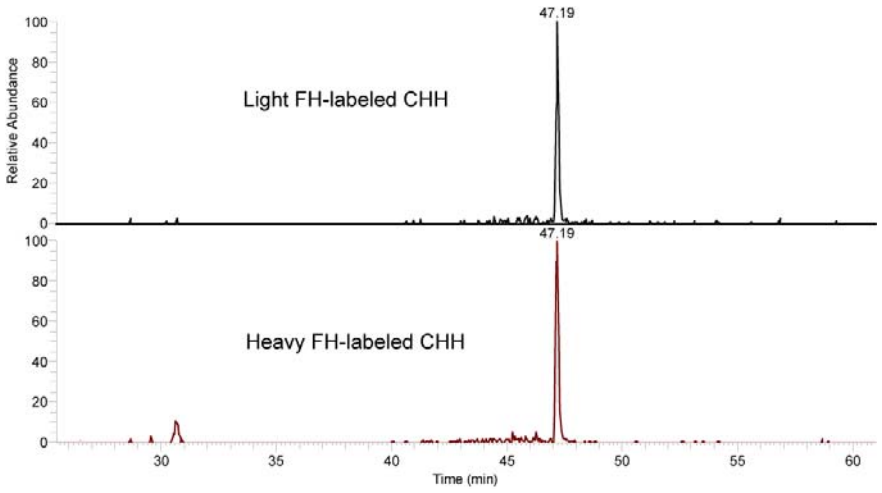


Figure S9. Extract ion chromatograms of light and heavy FH-labeled ^{Cam}-CHH$_{pO}$. The retention time values are identical, indicating that the deuterium effect on LC retention is negligible.

Chapter 7

Rapid and Sensitive Characterization of Neuropeptidome in Dungeness Crabs on a Q-Exactive High-Resolution Mass Spectrometer and Peptide Alterations in Response to Feeding

Adapted from: Chenxi Jia, Qing Yu, Chuanzi Ouyang, Fengfei Ma, Claire M Schmerberg, and Lingjun Li, Rapid and Sensitive Characterization of Neuropeptidome in Dungeness Crabs on a Q-Exactive High-Resolution Mass Spectrometer and Peptide Alterations in Response to Feeding. *EuPA Open Proteomics*, to be submitted.

Abstract

The Dungeness crab *Cancer magister* is a favored animal model for neurobiological studies due to their relatively simple and well-organized nervous system. However, the neuropeptidome is not fully discovered in this species. In this study, we applied high resolution/accurate mass measurement and high-throughput database searching to elucidate the neuropeptidome in the nervous system of Dungeness crab. In total, 127 peptides were identified with high confidence, including 11 novel ones. In addition, one large mandibular organ-inhibiting hormone was completely *de novo* sequenced by bottom-up strategy. Label-free quantitative analysis reveals that the orcokinin and RFamide peptides were decreased in abundance after food intake.

7.1 Introduction

Neuropeptides are the most diverse and complex class of signaling molecules, which acts as neurotransmitters or hormones regulating many important physiological processes [1, 2]. The peptidomic approach aims at simultaneous visualization and identification of the entire complement of peptides in a cell or tissue, which has been proven successful in neuroendocrine and neurobiology research [2]. Mass spectrometry is a powerful tool and can identify the precise sequences and modifications of neuropeptides without prior knowledge of peptide identity. Application of high resolution/accurate mass (HR/AM) measurement into peptidomics studies greatly improves the identification rate of peptides [3, 4]. The Q Exactive hybrid quadrupole-Orbitrap mass spectrometer exhibits high performance in proteomics studies [5, 6]. The superior quality of tandem mass (MS/MS) data provides ultimate confidence for a wide range of qualitative and quantitative applications. In this study, we tested the performance of this cutting-edge instrument for characterization of neuropeptidome in Dungeness crabs *Cancer magister* and obtained a high identification rate.

Another challenging aspect of peptidomics research in crustaceans is the almost complete lack of genomic information [7]. This will make the database searching strategy difficult for neuropeptide identification. Even for neuropeptide analysis in organisms with genomic sequence information available, database searching is often complicated by the lack of knowledge of the proteases involved. This in turn, results in a much larger library of peptides to be searched and an increased probability of finding a random match. In this study, the HR/AM data was processed by software Peaks against a home-built database [8] including the neuropeptides previously identified and reported in decepod crustaceans, resulting in identificaiton of neuropeptidome in a high-throughput manner.

The accurate molecular information of the neuropeptides faciliates their functional studies. The ultimate goal of peptidomics research is to understand how the neuropeptides function in circuit system. Feeding behavior is a fundamental aspect of energy homeostasis [9]. Neuroendocrine regulation of food intake is the results of an integrated response of neural circuits using a diverse set of neuropeptides. Previous studies have reported that neuropeptide Y, proopiomelanocortin, melanin-concentrating hormone, neurotensin, and cholecystokinin either stimulate or decrease food intake [10, 11]. In our recent studies, the neuropeptide changes in the brains, pericardial organs (POs) and sinus glands (SGs) of decopod crustaceans were monitored after food intake [12]. The RFamides, *Cancer borealis* tachykinin-related peptides, RYamides and pyrokinins exhibited significant alterations. In this study, one of the most important neuroendocrine organs, the sinus gland was investigated and neuropeptide alterations were observed upon feeding.

7.2 Experimentals

The protocols of feeding experiment, tissue dissection and bottom up MS sequencing of large neuropeptides are the same as described in Chapter 6.

For the neuropeptidome discovery, the tissue extract samples were analyzed on a Waters nanoAcquity ultra-performance liquid chromatography (LC) system coupled to a Q Exactive quadrupole orbitrap mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). A 0.5 μL of peptide sample was injected and loaded onto the Waters Symmetry C_{18} trap column (180 μm x 20 mm, 5 μm) using 97% mobile phase A (0.1% of formic acid in water) and 3% mobile phase B (0.1% of formic acid in ACN) at a flow rate of 5 μL/min for 3 min. A Waters BEH 300Å C_{18} reversed phase capillary column (150 mm x 75 μm, 1.7 μm) was used for separation. The gradient started from 3 to 10% B during the first 5 min, increased to 55% B in the next 45 min,

then was kept at 90% B for 20 min. The data was acquired under data dependent mode (DDA, top10). Typical mass spectrometric conditions were as follows: spray voltage, 2.8 kV; no sheath and auxiliary gas flow; heated capillary temperature, 275 °C; normalized high-energy collision dissociation (HCD) collision energy 30%. The settings are: resolution 70,000 for full scan; resolution 17,500 for MS/MS; automatic gain control $2e^5$; maximum ion injection time, 100 ms; isolation window, 2 *m/z*; fixed first mass, 100 *m/z*.

The mass spectral data were processed by Peaks Studio 5.2 (BSI, Canada) [3]. The data was refined and subjected to Auto *De Novo* program for sequencing with the mass tolerance at 0.1 Da. Variable PTMs included C-terminal amidation, pyro-Glu and methionine oxidation. The data from Auto *De Novo* was subsequently subjected to Peaks Database Search against the home-built neuropeptide database [8] with the same settings as above.

The label-free quantification software Sieve was used to process the DDA data, where respectively extracted ion chromatograms of target peptides were directly compared. The parameters are: m/z range 200-2000 Da, time frame 6 min, m/z frame 0.02 Da with a peak intensity threshold 10,000.

7.3 Results and Discussion

In our previous studies, we utilized QTOF mass spectrometer for identification of peptidome in crustaceans [7, 13]. The limited MS/MS sensitivity and mass accuracy results in low identification rate of neuropeptides. In this study, a Q Exactive Quadrupole-Orbitrap mass spectrometer was employed to discover the neuropeptidome in the nervous system of Dungeness crabs. The Q Exactive is equipped with a quadrupole as a front-end mass filter coupled with an Orbitrap mass analyzer. Under MS/MS mode, the quadrupole acts as an ion filter, which only allows a specified mass range of ions to be transferred and accumulated

into the C-trap [5]. Therefore, the MS/MS sensitivity is dramatically improved. In addition, the Orbitrap mass analyzer offers the capability of high resolution/accuracy. Those features greatly improve the MS/MS spectra quality of peptides. In addition to the use of the state-of-the-art instrument, we also constructed a database which contains the crustacean neuropeptides reported previously.

In our experiment, the tissue extract samples of POs, brains, and sinus glands were respectively analyzed by LC-MS/MS under DDA mode (top10 most abundant ions). The data was processed by Peaks and searched against the home-built neuropeptide database. In total, 116 peptides were matched, which were derived from 11 families, including allatostatin (AST)-A, AST-B, AST-C, crustacean hyperglycemic hormone precursor related peptides (CPRP), cryptocyanin, orcokinin, pigment-dispersing hormone (PDH), RFamide, RYamide and tachykinin. Among those, 57 peptides were found in the sinus glands; 66 from the POs; and 59 from brains. Although we set the mass accuracy tolerance as 10 ppm and -10logP score as 30, most of the matched peptide in Table 1 showed mass accuracy better than 2 ppm. Figure 1 shows three representative annotated MS/MS sequencing spectra of peptides, exhibiting high fragmentation efficiency by HCD and resulting peptides were identified with high confidence. More importantly, one LC-MS/MS run on Q Exactive only consumed 2% of the tissue extract from one sinus gland organ, resulting in identification of 116 neuropeptides with high confidence. This result suggests that the use of Q Exactive greatly reduces the detection limit and the sample consumption

Spider algorithm in Peaks is designed to detect peptide mutations and perform cross-species homology search. Zhang and co-workers [3, 4] has successfully applied it to identify the neuropeptides from tree shrew and mouse brains with substituted amino acid residues. In

Chapter 6, we employed this approach to predict the substituted residue of large CHH peptides, which greatly enhanced the throughput of peptide identification. In this work, we also tested this strategy by searching the LC-MS/MS data against our home-built database. However, the Spider algorithm failed to generate any hits for putative peptides. The possible reason is that the similarities of the homologous sequences are low. Herein, we manually inspect the *de novo* sequencing results exported by Peaks according to the specific motifs of crustacean neuropeptide families. For example, the A-type AST family possesses –YXFGLamide C-terminal motif (where X is a variable amino acid). The B-type AST family has –WX_6Wamide C-terminal motif. The RYamide and RFamide families have –RYamide and –RFamide at the C-termini [7]. Table 2 lists the putative novel neuropeptides identified by manually inspection of the results. In total, 11 peptides were found from 4 families. The MS/MS sequencing spectra are shown in Supplementary Materials.

As shown in Table 1, the molecular sizes of matched peptides range from 700 Da to 4,000 Da. The traditional strategy lacks the capability for *de novo* sequencing of peptides with molecular weights larger than 4 kDa. In this study, we applied the high-definition strategy introduced in previous chapters to the sequencing of CHH-family neuropeptides. In the tissue extract, we detected one large neuropeptide with molecular weight of 9242.224 Da. After treatment by dithiothreitol (DTT) and iodoacetamide (IAA), the peptide was digested by trypsin, followed by LC-MS/MS analysis. The data was searched in MASCOT against a NCBI database as described in previous chapters. One homologous mandibular organ-inhibiting hormone (MOIH) peptide from *Cancer pagurus* [14] was matched with a 43% of sequence coverage. Subsequently, the sequence of *Cancer pagurus* MOIH was included in database and the LC-MS/MS data was processed by Spider Search in Peaks against it. Table

3 lists all the tryptic peptides matched. By assembling all the tryptic peptide segments, the complete sequence of this novel MOIH can be determined as shown in Figure 2.

The well-characterized molecular information of these neuropeptides facilitates physiological studies. In our previous report, a multifaceted approach has been employed to investigate the neuropeptide changes in response to food intake using crustacean models. Two types of neural tissues have been studied: brains and pericardial organs [12]. In this study, we investigated the neuropeptide changes in sinus glands using Dungeness crabs. The abundances of orcokinin and RFamide, including NFDEIDRSSFGV, NFDEIDRSSFGFN, NFDEIDRSGFGF, NFDEIDRSGFGFA and PELDHVFLRFamide, were significantly decreased after feeding. This result indicates that these peptides may be altered upon food intake. In our previous study, the orcokinin peptides did not show changes in brains, while RFamide peptides were significantly increased [12]. To further elucidate the regulation pattern and underlying mechanisms, microdialysis experiments [15] will be carried out to monitor the continuous changes along a certain time course after feeding.

7.4 Conclusions

In this study, we combined the HR/AM measurement with high-throughput algorithm-assisted peptide discovery strategy to characterize the neuropeptidome in the nervous system of Dungeness crabs. In total, 127 peptides were identified with 11 novel ones. Also, a large MOIH peptide was completely sequenced by a bottom-up strategy. The quantitative results indicate that orcokinin and RFamide peptides were decreased in the sinus glands after feeding. This study greatly increases the number of known peptides present in this species and provides a strong foundation for future studies on the physiological roles of these signaling molecules in neural circuits, especially their regulation mechanism in feeding

behavior.

7.5 References

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

[2] 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.

[3] 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.

[4] Ranc V, Petruzziello F, Kretz R, Argandona EG, Zhang X, Rainer G. Broad characterization of endogenous peptides in the tree shrew visual system. J Proteomics. 2012;75:2526-35.

[5] Michalski A, Damoc E, Hauschild JP, Lange O, Wieghaus A, Makarov A, et al. Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. Mol Cell Proteomics. 2011;10:M111 011015.

[6] Peterson AC, Russell JD, Bailey DJ, Westphall MS, Coon JJ. Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. Mol Cell Proteomics. 2012;11:1475-88.

[7] Hui L, D'Andrea BT, Jia C, Liang Z, Christie AE, Li L. Mass spectrometric characterization of the neuropeptidome of the ghost crab Ocypode ceratophthalma (Brachyura, Ocypodidae). General and comparative endocrinology. 2013;184:22-34.

[8] Schmerberg CM. Functional Neuropeptidomics in the Decapod Crustacean: Method Development and Application to Behavioral Neuroscience Research. University of Wisconsin-Madison, Madison, USA. 2012.

[9] Morton GJ, Cummings DE, Baskin DG, Barsh GS, Schwartz MW. Central nervous system control of food intake and body weight. Nature. 2006;443:289-95.

[10] Matsuda K, Sakashita A, Yokobori E, Azuma M. Neuroendocrine control of feeding behavior and psychomotor activity by neuropeptideY in fish. Neuropeptides. 2012;46:275-83.

[11] Frese CK, Boender AJ, Mohammed S, Heck AJ, Adan RA, Altelaar AF. Profiling of diet-induced neuropeptide changes in rat brain by quantitative mass spectrometry. Anal Chem. 2013;85:4594-604.

[12] 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.

[13] Hui L, Zhang Y, Wang J, Cook A, Ye H, Nusbaum MP, et al. Discovery and functional study of a novel crustacean tachykinin neuropeptide. ACS Chem Neurosci. 2011;2:711-22.

[14] Wainwright G, Webster SG, Wilkinson MC, Chung JS, Rees HH. Structure and significance of mandibular organ-inhibiting hormone in the crab, Cancer pagurus. Involvement in multihormonal regulation of growth and reproduction. J Biol Chem. 1996;271:12749-54.

[15] Schmerberg CM, Li L. Mass spectrometric detection of neuropeptides using affinity-enhanced microdialysis with antibody-coated magnetic nanoparticles. Analytical chemistry. 2013;85:915-22.

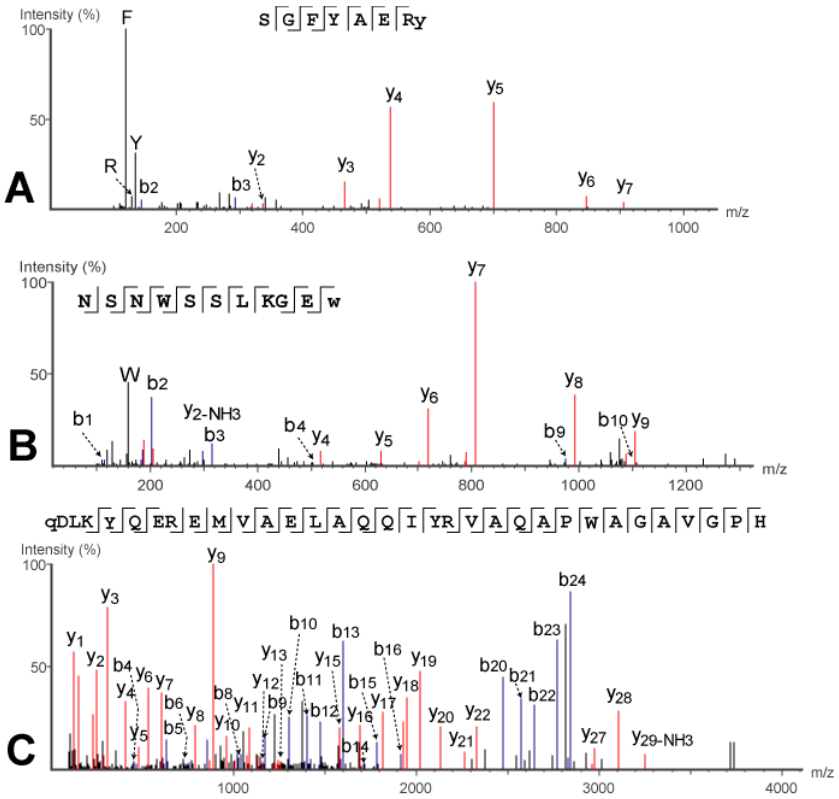


Figure 1. MS/MS sequencing of RYamide (A), AST-B (B) and PDH (C). y, Y-NH_2; w, W-NH_2; and q, pyro-Gln.

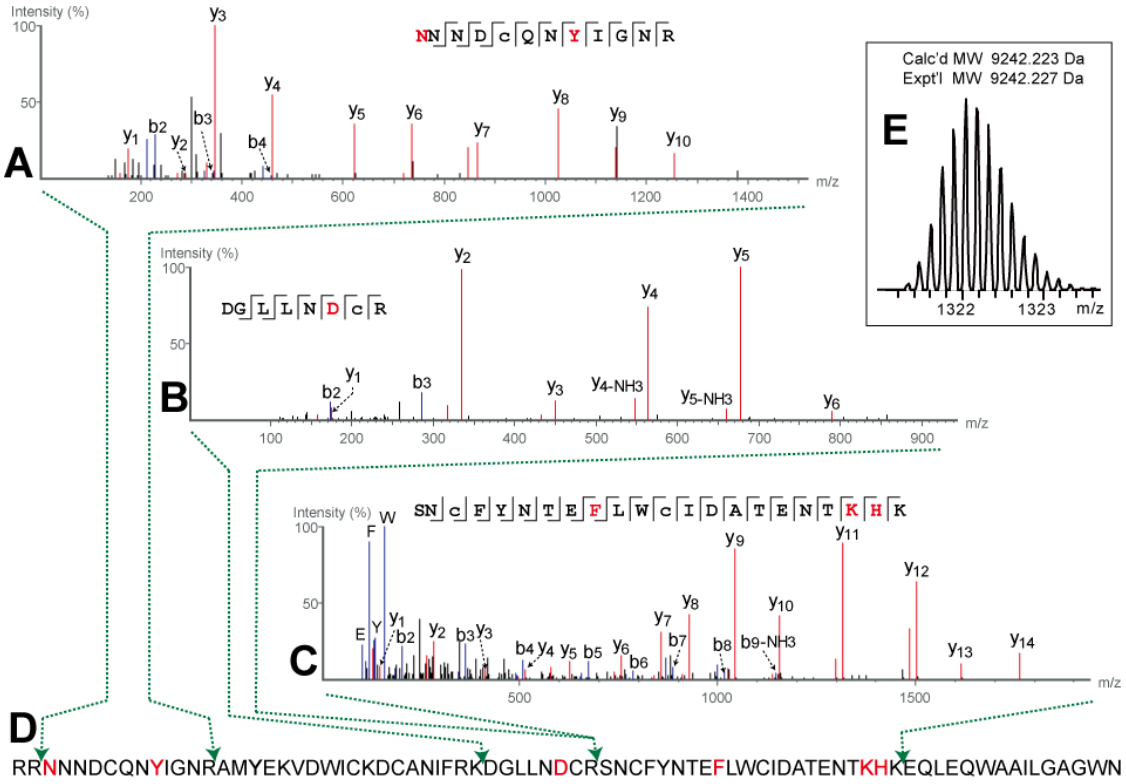


Figure 2. *De novo* sequencing of novel MOIH from *Cancer magister*. (A, B and C) MS/MS sequencing spectra of three tryptic peptides. (D) Sequence assembly. (E) Isotopic distributions of the MOIH. The residues different from the reference sequence of *Cancer pagurus* MOIH are highlighted in red. c, carbaminomethylated cysteine.

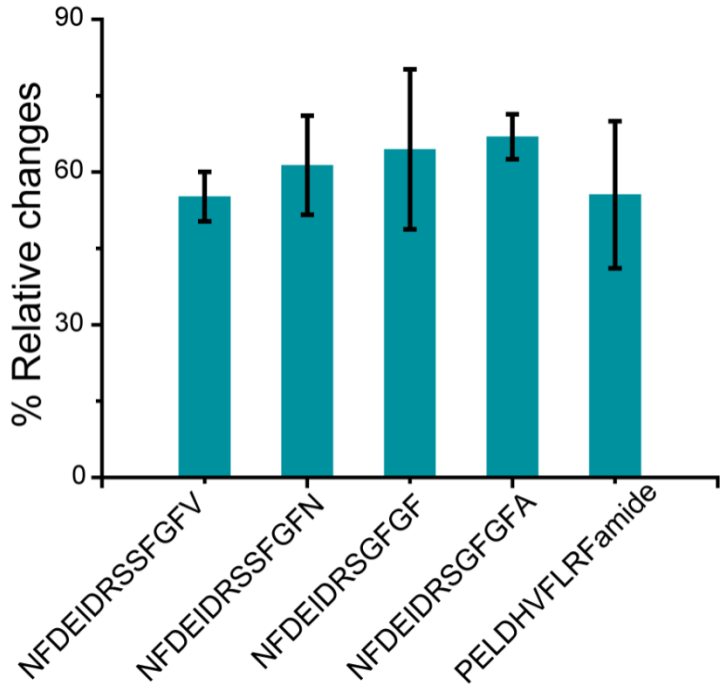


Figure 3. Significantly altered neuropeptides in response to feeding. Error bars indicate the variability of the difference between sample means (standard error, $p<0.05$).

Table 1. Neuropeptides identified by database searching. [a]

Family	Neuropeptides	-10LogP	ppm	m/z	Charge	SG	PO	Br
AST-A	DPYAFGL(-.98)	35.03	1.4	391.2	2		+	+
	EPYAFGL(-.98)	27.67	0.8	398.21	2		+	
	AGPYSFGL(-.98)	27.14	0.8	405.71	2		+	
	AGGAYSFGL(-.98)	30.17	0.9	421.21	2		+	
	GSGQYAFGL(-.98)	32.53	0.7	449.72	2	+	+	
	ARPYSFGL(-.98)	25.63	1.5	455.25	2		+	+
	PDMYAFGL(-.98)	36.48	0.3	456.72	2		+	+
	E(-18.01)RAYSFGL(-.98)	34.52	1	462.24	2		+	
	PDM(+15.99)YAFGL(-.98)	43.07	2.1	464.72	2		+	+
	PRDYAFGL(-.98)	33.4	0.9	469.25	2		+	
	PRTYSFGL(-.98)	27.37	1.7	470.26	2		+	
	QRAYSFGL(-.98)	39.31	0.1	470.75	2	+	+	+
	Q(-17.03)RDYSFGL(-.98)	33.08	2	484.24	2		+	
	QRDYSFGL(-.98)	38.29	0.7	492.7	2	+	+	+
	PADLYEFGL(-.98)	43.93	1.2	512.26	2		+	
	SGGFAFSPRL(-.98)	67.56	0.9	519.28	2			+
	TNFAFSPRL(-.98)	58.6	0	526.29	2			+
	DPYAFGLGKRPDMYAFGL(-.98)	67.42	0.5	1009	2		+	
AST-B	E(-18.01)LNFSPGW(-.98)	32.51	0.3	465.72	2	+		
	GWSSMRGAW(-.98)	31.18	3.3	518.74	2		+	
	AGWSSMRGAW(-.98)	50.72	0.1	554.26	2	+	+	+
	AGWSSM(+15.99)RGAW(-.98)	45.58	0.9	562.26	2		+	
	NWSKFQGSW(-.98)	34.3	1.6	569.78	2		+	
	GKWSNLRGAW(-.98)	47.56	0.9	587.32	2		+	
	TSWGKFQGSW(-.98)	43.86	0.2	591.79	2	+	+	
	GNWNKFQGSW(-.98)	51.17	0.2	611.79	2	+	+	+
	NNWSKFQGSW(-.98)	61.67	0	626.8	2	+	+	+
	NDWSKFGQSW(-.98)	50.18	0	627.29	2		+	
	NNNWSKFQGSW(-.98)	65.33	0.6	683.82	2			+
	VPNDWAHFRGSW(-.98)	73.21	1	735.85	2	+	+	+
AST-C	PDGDMEM(+15.99)TGPWDTI(-.98)	33.54	1.4	790.32	2		+	
CPRP	SYRGAVEPNTPLG	60.36	0.4	680.85	2	+		
	YRGAVEPNTPLGD	64.84	1.2	694.85	2	+		
cryptocyanin	YKIFEPLRE	47.18	0.5	398.89	3		+	
Orcokinin	NFDEIDR	49.76	0.3	454.71	2			+
	NRNFLRF(-.98)	49.17	0.3	483.27	2	+	+	+
	IDRSGFGFA	60.68	1.4	485.24	2	+		+
	NFDEIDRS	40.34	0.7	498.22	2			+
	IDRSSFGFV	60.21	1.1	514.26	2	+		
	EIDRSGFGFA	49.19	1.2	549.77	2	+		

	EIDRSGFGFV	58.19	0.2	563.78	2	+		
	EIDRSSFGFV	57.33	0.1	578.79	2	+		
	EIDRSSFGFN	52.43	1.2	586.27	2	+		
	NSELINSILGL	52.34	0.7	586.83	2	+		
	FDAFTTGFGHS	55.67	2.4	593.76	2		+	+
	NFDEIDRSGF(-.98)	63.32	0.7	599.78	2	+		+
	NFDEIDRSSF(-.98)	65.47	0.5	614.78	2			+
	NFDEIDRSGFG	68.57	1.1	628.78	2	+		+
	NFDEIDRSGFA	56.99	0.4	635.79	2	+		+
	NFDEIDRSSFG	60.18	1.4	643.79	2	+		+
	NFDEIDRSSFA	77.65	2.2	650.79	2			+
	FDEIDRSGFGFA	88.21	0.3	680.81	2			+
	NFDEIDRSGFGF	75.88	0.8	702.31	2	+		+
	NFDEIDRSGFGFA	76.53	0.4	737.83	2	+	+	+
	NFDEIDRSGFGFV	80.22	1.1	751.85	2	+	+	+
	NSELINSLLGISRL	65.96	0	764.94	2	+		+
	NFDEIDRSSFGFV	78.89	0.4	766.85	2	+		+
	NFDEIDRSSFGFN	78.57	0.1	774.34	2	+	+	+
	FDAFTTGFGHS	52.58	0.7	593.76	2	+		
other	HIGSLYR(-.98)	35.43	1.1	422.74	2		+	
PDH	APLEGAGGLPH	75.05	1.8	509.77	2	+		
	Q(-17.03)DLKYQEREM	63.33	0.9	661.8	2	+		
	RVAQAPWAGAVGPH	90.95	0.6	472.92	3	+		+
	Q(-17.03)DLKYQEREMVA	62.72	0.4	746.86	2	+		
	Q(-17.03)DLKYQEREMVAEL	84.65	0.4	867.92	2			+
	NSELINSILGLPKVMNDA(-.98)	69.62	1.1	964.01	2	+		+
	NSELINSLLGIPKVM(+15.99)TDA(-.98)	45.04	2.9	965.52	2	+		
	NSELINSLLGISRLMNEA(-.98)	71.56	0.5	987.03	2			+
	NSELINSLLGLPKFMIDA	41.16	2.1	988.03	2	+		
	AQQIYRVAQAPWAGAVGPH	88.3	1.2	674.02	3	+		
	Q(-17.03)DLKYQEREMVAELAQQIY	86.33	0.3	1169.57	2			+
	Q(-17.03)DLKYQEREMVAELAQQIYRVAQAPWAGAVGPH	107.9	0.3	934.7227	4	+		+
	QDLKYQEREMVAELAQQIYRVAQAPWAGAVGPH	78.94	8.1	938.97	4	+		+
RFamide	RSFLRF(-.98)	33.13	0.5	412.75	2		+	
	RNFLRF(-.98)	39.66	0.5	426.25	2		+	
	GRNFLRF(-.98)	47.53	0.5	454.76	2		+	
	SKNYLRF(-.98)	44.7	3.5	309.51	3		+	
	NRSFLRF(-.98)	38.43	1.8	313.51	3		+	
	Q(-17.03)RNFLRF(-.98)	45.22	1	481.77	2		+	
	GGRNFLRF(-.98)	46.37	0.1	322.52	3	+	+	+
	DRNFLRF(-.98)	51.31	1	483.77	2	+	+	+
	TPALRLRF(-.98)	55.12	1.2	324.87	3			+
	PQGNFLRF(-.98)	35.55	10.2	489.27	2		+	

	QRNFLRF(-.98)	54.15	0.1	327.19	3		+	
	GNRNFLRF(-.98)	64.34	0.1	511.78	2	+	+	+
	AHKNYLRF(-.98)	47.07	2.3	349.87	3		+	
	PQRNFLRF(-.98)	38.67	2.9	359.54	3		+	
	YSKNYLRF(-.98)	54.67	0.7	545.3	2		+	
	YNRSFLRF(-.98)	47.03	0.4	551.3	2		+	
	GAHKNYLRF(-.98)	61.14	0.9	368.87	3	+	+	+
	SMPSLRLRF(-.98)	53.85	1.2	553.32	2			+
	SM(+15.99)PSLRLRF(-.98)	49.64	0.7	374.54	3			+
	YSQVSRPRF(-.98)	48.08	1.5	380.21	3			+
	GYSKNYLRF(-.98)	77.36	0.5	573.81	2	+	+	+
	APQRNFLRF(-.98)	62	0.1	574.33	2	+	+	+
	SENRNFLRF(-.98)	61.58	1.2	394.54	3	+	+	+
	YSNLNYLRF(-.98)	35.89	9.4	594.82	2		+	+
	DENRNFLRF(-.98)	59.66	0.5	605.31	2		+	
	YSDRNFLRF(-.98)	40.13	0.9	608.82	2		+	
	Q(-17.03)DLDHVFLRF(-.98)	66.95	0.1	636.33	2	+	+	+
	QDLDHVFLRF(-.98)	78.18	0.7	644.84	2	+	+	+
	GYRKPPFNGSIF(-.98)	67.22	0.2	691.37	2	+		+
	DKYYSQVSRPRF(-.98)	52.58	0	515.6	3	+		+
RYamide	FYANRY(-.98)	33.74	3.2	416.71	2		+	
	HLGSLRY(-.98)	39.11	4.1	282.17	3	+		+
	FYSQRY(-.98)	41.69	1.6	431.71	2		+	
	GFYANRY(-.98)	35.02	0.1	445.22	2		+	
	GFYSQRY(-.98)	45.37	0.6	460.22	2		+	
	SGFYANRY(-.98)	62.12	0.1	488.74	2	+	+	+
	Q(-17.03)GFYSQRY(-.98)	42.58	0.9	515.74	2	+	+	
	QGFYSQRY(-.98)	43.57	0.5	524.25	2		+	
	SSRFVGGSRY(-.98)	64.52	1.1	557.79	2		+	
Tachykinin	APSGFLGMR(-.98)	58.55	0.7	467.75	2	+		+
	APSGFLGM(+15.99)R(-.98)	56.47	1.9	475.75	2			+
	TPSGFLGMR(-.98)	61.86	0.3	482.76	2			+
	APSGFLGMR(-.98)	66.11	0.7	933.48	2	+		

[a] The modifications are annotated as a number in bracket. (-.98) C-terminal amidation; (-17.03) pyro-Gln; (+15.99) oxidation; (-18.01) pyro-Glu.

Table 2 Putative novel neuropeptides identified in this study. [a]

Family	Neuropepides [b]	[M+H]+ [c]	m/z [d]	z	ppm	TLC	ALC	SG	PO	Br
AST-A	SAGFGFYGFGL(-.98)	1121.542	561.274	2	1.1	7.5	68		+	
AST-B	NSPSHASYGFGL(-.98)	1235.580	618.294	2	0.7	9.2	77	+	+	
	NSNWSSLKGEW(-.98)	1306.618	653.818	2	8.2	6.9	62	+	+	+
RFamide	RACPLRF(-.98)	861.488	431.246	2	4.5	2.5	36		+	
	NRPM(+15.99)LRF(-.98)	948.520	474.762	2	2.8	4.3	62		+	
	DLRTPALRLRF(-.98)	1356.822	452.95	3	1.3	7.8	71			+
	NAYFALAGRPRF(-.98)	1381.749	691.378	2	0	8.4	70			+
RYamide	SGFYAERY(-.98)	991.463	496.236	2	1.2	7.5	93		+	
	SGAPALASQRY(-.98)	1119.591	560.299	2	1.0	6.7	61		+	
	SGSGFYANRY(-.98)	1120.517	560.762	2	0.7	8.6	86		+	
	TGDTRQLTERSGFYANRY(-.98)	2134.042	534.266	4	0.7	15.3	85		+	+

[a] Modifications: (-.98) C-terminal amidation; (+15.99) oxidation. SG, sinus gland; PO, pericardial organ; Br, brain. TLC, total local confidence; ALC, average local confidence.
[b] The order of the underlined residues is not determined.
[c] Theoretical values.
[d] Experimental values.

Table 3. Tryptic peptides of the novel MOIH identified by LC-MS/MS and database searching.[a]

Tryptic peptides	Score	ppm	m/z	z	Start	End
NNNDC(+57.02)QNYIGNR	33.50 [b]	1	741.3115	2	3	14
AMYEKVDWIC(+57.02)K	81.23 [c]	0.6	721.8435	2	15	25
AMYEKVDWIC(+57.02)KDC(+57.02)ANIFR	89.74 [c]	0.9	773.6948	3	15	32
AMYEKVDWIC(+57.02)KDC(+57.02)ANIFRK	79.07 [c]	0.8	816.3931	3	15	33
VDWIC(+57.02)KDC(+57.02)ANIFR	81.00 [c]	0.9	848.8996	2	20	32
VDWIC(+57.02)KDC(+57.02)ANIFRK	84.42 [c]	2.8	912.9453	2	20	33
DGLLNDC(+57.02)R	23.00 [b]	2.1	481.7219	2	34	41
DGLLNDC(+57.02)RSNC(+57.02)FYNTEFLWC(+57.02)IDATENTKHK	63.56 [b]	3	745.1365	5	34	62
SNC(+57.02)FYNTEFLWC(+57.02)LDATENTK	55.50 [b]	1	838.3629	3	42	60
SNC(+57.02)FYNTEFLWC(+57.02)LDATENTKHK	56.50 [b]	0.6	695.3123	4	42	62
LWC(+57.02)IDATENTK	21.00 [b]	1.4	675.8198	2	51	60
WC(+57.02)IDATENTK	22.00 [b]	1	619.2794	2	52	60
EQLEQWAAILGAGWN	62.67 [c]	1.1	562.6136	3	64	78

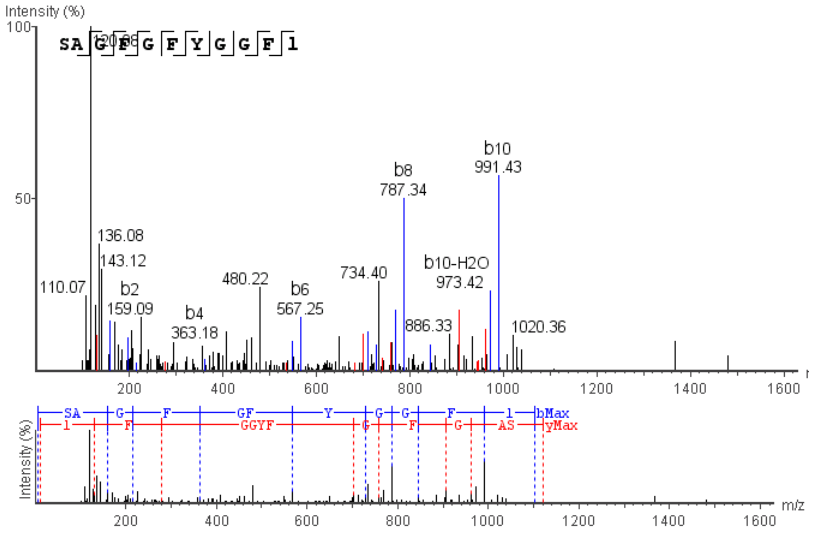
[a] Modifications: (+57.02) carbamidomethylation. The residues different from the reference sequence of *Cancer pagurus* MOIH are underlined.

[b] Spider search score exported by PEAKS.

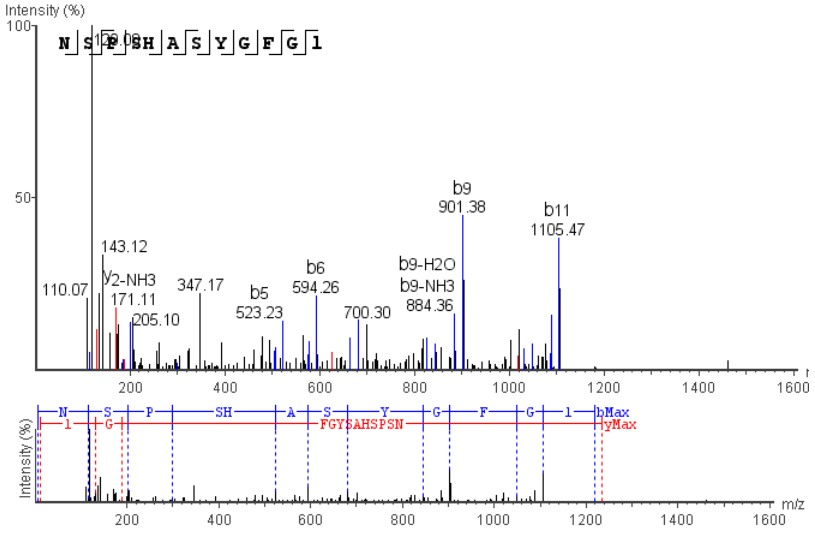
[c] -10logP score exported by PEAKS.

Supplementary Materials for Chapter 7

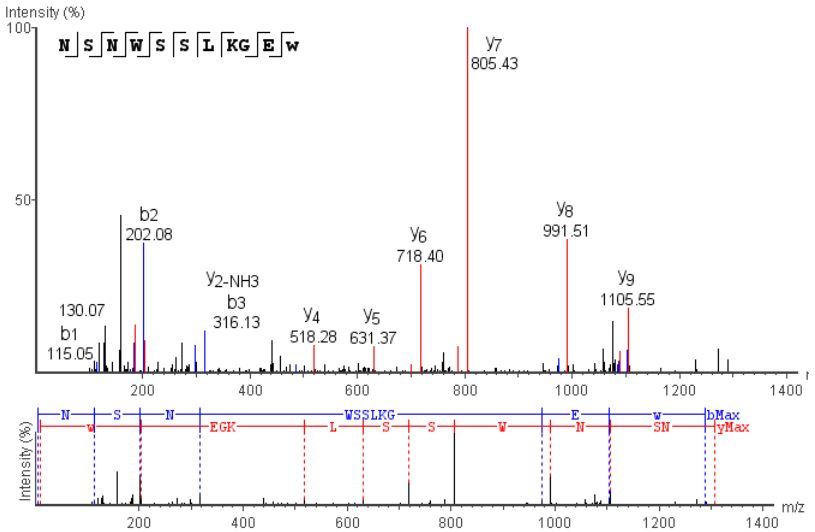
The MS/MS sequencing spectra of the novel neuropeptides (Table 2) found in this study. The spectra are directly exported by Peaks.



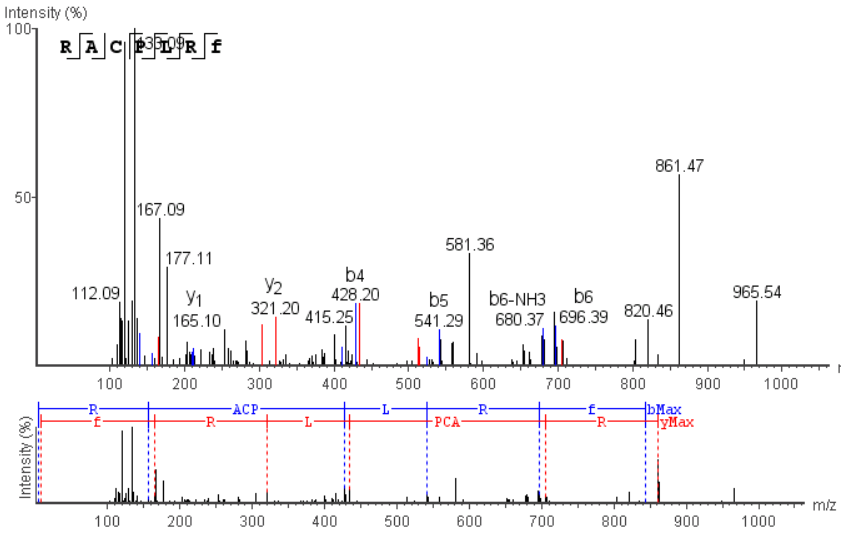
SAGFGFYGGFl



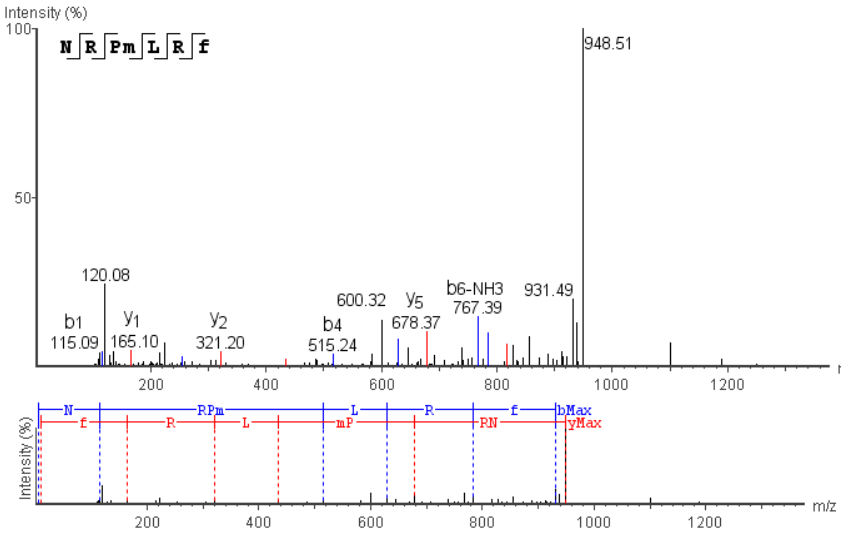
NSPSHASYGFGl



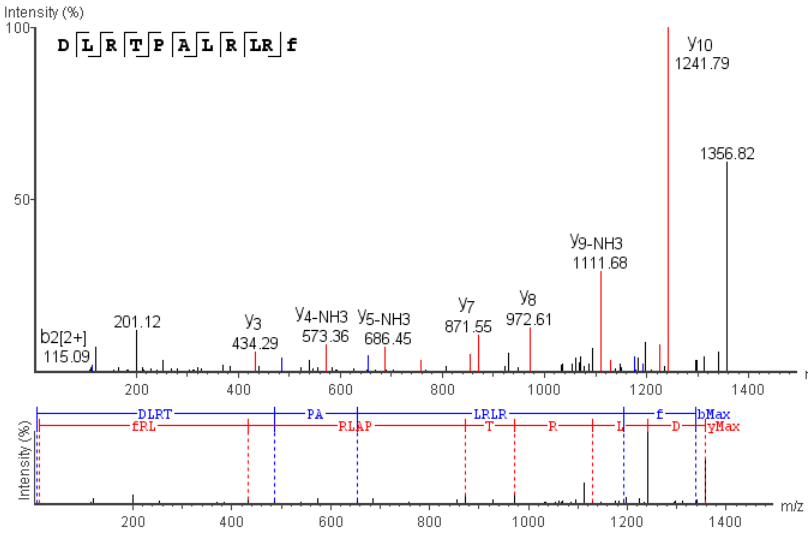
NSNWSSLKGEw



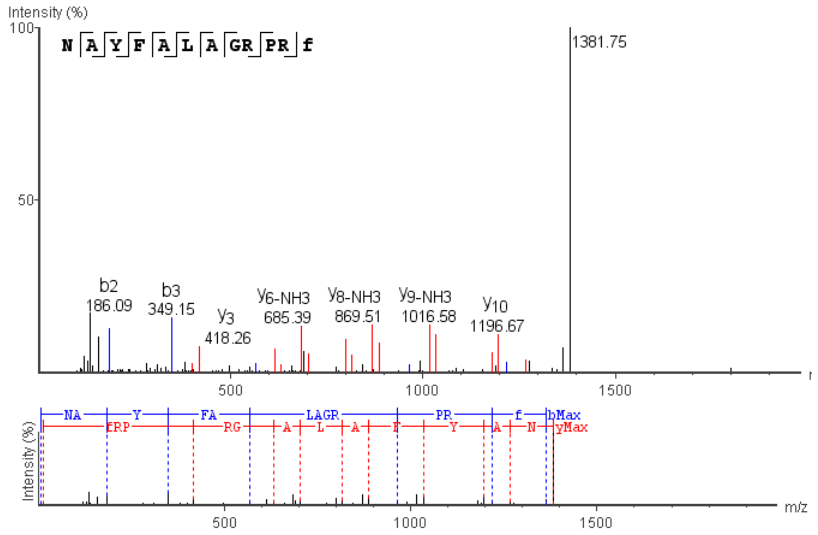
RACPLRf



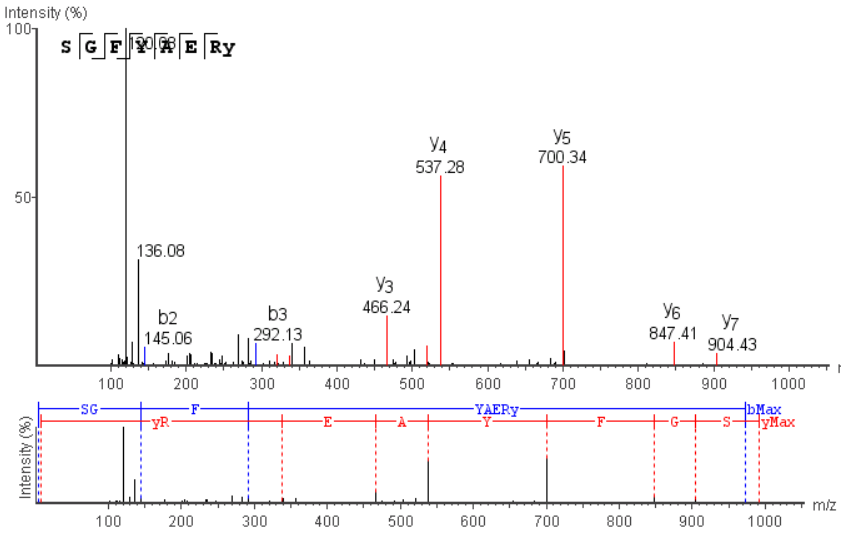
NRPmLRf



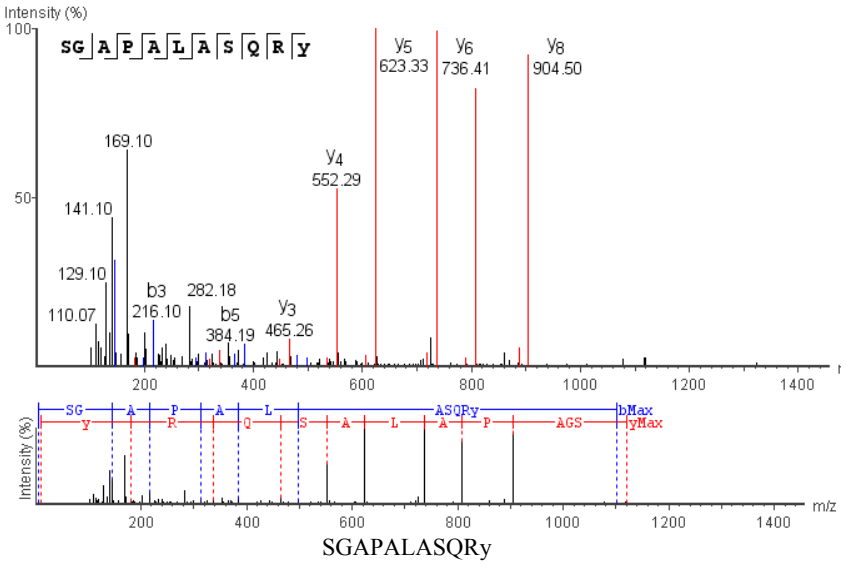
DLRTPALRLRf



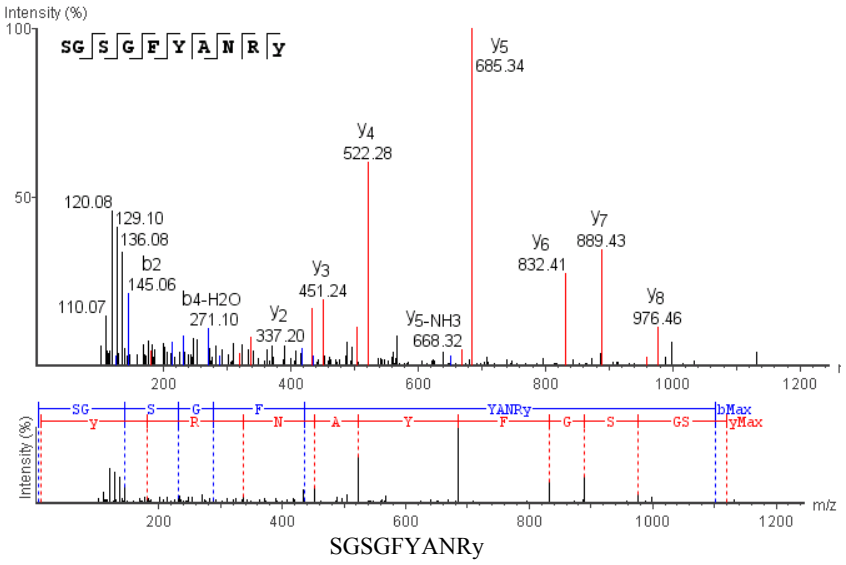
NAYFALAGRPRy



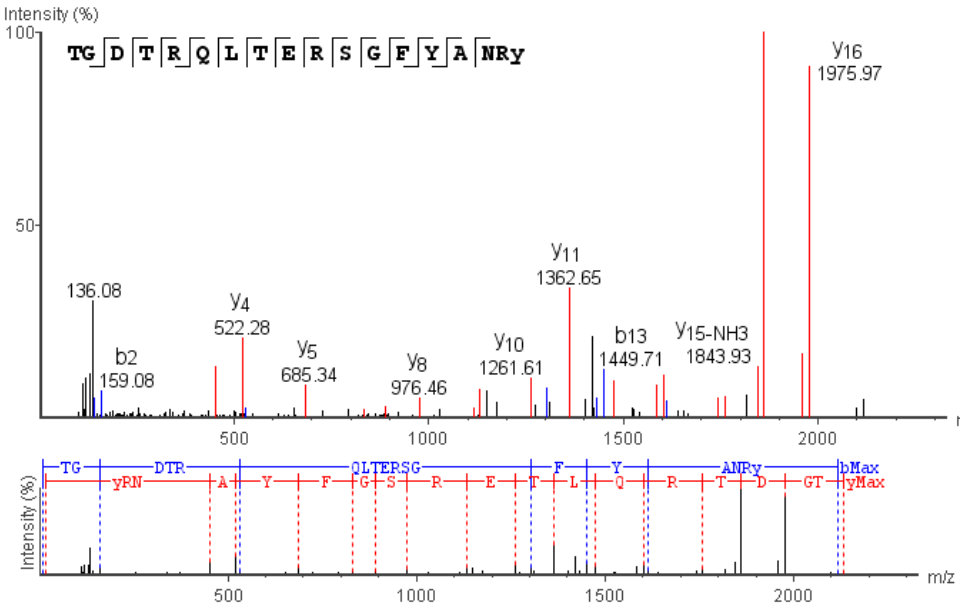
SGFYAERy



SGAPALASQRy



SGSGFYANRy



TGDTRQLTERSGFYANRy

Chapter 8

Conclusions and Future Directions

8.1. Conclusions

In this dissertation, a neuropeptide discovery pipeline has been developed for efficient peptidomic research, which targets at elucidating the sequences and structures of neuropeptides and improving our understanding of their functional roles in neural circuits. The implementation of a suite of mass spectrometry (MS)-based tools results in confident identification and characterization of a set of novel neuropeptides. The elucidated structural and quantitative information of these signaling molecules improves our understanding of their bioactivities, gas-phase fragmentation chemistry, and the functional roles they play in regulating behavior.

The work described in Chapter 2 represents a new route to discovery and characterization of large neuropeptides. The multifaceted MS-based strategy involves a comprehensive and systematic implementation of peptide candidate scanning, *in silico* homology searching, *de novo* sequencing, distribution mapping, and conformation analysis. The accurate sequence, post-translational modifications (PTMs), spatial distribution pattern, and conformational analysis of the crustacean hyperglycemic hormone (CHH)-family neuropeptides were elucidated with a combination of MS tools. The high-definition *de novo* sequencing strategy combines well-established bottom-up, on-line top-down and off-line top-down methods, offering complementary sequence information and in-depth characterization of PTMs at residue level. Because CHH-family peptides represent the typical molecular features of large neuropeptides, this multifaceted strategy is applicable to the characterization of large endogenous signaling peptides in other biological systems.

The study described in Chapter 3 systematically evaluates 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 the spiny lobster *P. interruptus* sinus gland. The results provide a foundation for future functional and mechanistic studies of these novel neuropeptides.

In Chapter 4, a new strategy was developed for site-specific characterization of peptide epimers, which allows for rapid and precise localization of D-amino acids in D-amino acid-containing peptide (DAACP) candidates. The analysis can be finished in a single LC-MS/MS-IMS run, followed by data processing in a simple and straightforward manner. The efficiency and utility of the strategy were demonstrated by analysis of a set of peptides with various molecular sizes, the D-amino acid containing peptides Melanocyte-stimulating hormones, Deltorphin and Achatin-I and their all-L form counterparts. In the CHHs isolated from American lobster, our data indicates that the peptides contain an isomerization site of L- to D-Phe at the third residue from the N-terminus. This study demonstrates a novel approach to obtaining positional information of amino acid isomerizations in peptides by elucidating the IMS data of peptide fragment ions. By coupling with efficient screening approaches, the developed strategy is potentially applicable to large-scale discovery of DAACPs; and the proposed concept is transferable to characterization of other post-translational isomerizations in large biological molecules.

In Chapter 5, ion isomer analysis of the three representative b ions has demonstrated the utility of a novel strategy combining experimental analysis via ETD, IMS and chemical labeling and theoretical calculation using molecular dynamic simulation. In comparison with previous studies, the method described in Chapter 5 has several advantages: (1) all the isomers (^{ol}b, ^{c}b, and ^{rl}b) can be simultaneously probed; (2) the experiments can be conducted using commercially

available mass spectrometers; and (3) the method is applicable for large-scale screening. The detailed analyses of gas-phase ion isomer components and their dynamic conversions reveal the proposed fragmentation mechanism of b ion cyclization and their potential impact on large-scale proteomic studies.

The study in Chapter 6 demonstrates the utility of an improved peptide discovery pipeline for qualitative and quantitative analysis of large neuropeptide hormones and related physiological study. First, by incorporating the high resolution/accurate mass (HR/AM) capability of Orbitrap instruments with PEAKS spider homology search, the substituted residues of target large peptides can be accurately and rapidly predicted according to homologous sequences. Second, we developed a novel quantitative top-down tandem MS strategy with assistance of isotopic labeling, greatly improving the quantitation accuracy. Third, by applying the qualitative and quantitative approaches to peptide discovery in Dungeness crab, novel CHHs were confidently *de novo* sequenced without assistance of genome information, and quantitative monitoring of the abundance changes at the tissue level was achieved. Forth, the quantitative analysis of CHH peptides in response to food intake indicates alteration of CHH abundances in the sinus glands upon feeding suggesting potential link of CHH expression and secretion with feeding behavior. Collectively, this peptide discovery pipeline highlights the ability for qualitative and quantitative analysis of large neuropeptides and is expected to aid in future research efforts to unravel large complex signaling peptides employed by other biological systems.

In Chapter 7, I combined the HR/AM measurement with high-throughput algorithm-assisted peptide discovery strategy to characterize the neuropeptidome in the nervous system of Dungeness crabs. In total, 127 peptides were identified with 11 novel ones. Also, the complete sequence of a large MOIH peptide was obtained by a bottom-up strategy. The quantitative results

indicate that orcokinin and RFamide peptides were decreased in the sinus glands after feeding. This study greatly increases the number of known peptides present in this species and provides a strong foundation for future studies on the physiological roles of these signaling molecules in neural circuits, especially their regulation mechanism in feeding behavior.

8.2. Future Directions

This work can be further extended and also provides the foundation for many new directions. Although the methods established in my dissertation are based on crustacean models, they are transferable to the nervous system of other species. One of our on-going projects is to characterize the neuropeptides in mouse brain. Our preliminary data was obtained by LC-MS/MS analysis and database searching of mouse tissue samples, which showed that more than 100 bioactive peptides have been identified in the small pituitary gland. Fricker [1] reported that a large number of bioactive peptides from mouse brains have molecular weights larger than 4 kDa. Therefore, the high-definition approach established in Chapter 2 will be applied to elucidate the large neuropepitdes and their isoforms in mouse brains. In addition to the sequence analysis, the method described in Chapter 4 will be applied for screen for DAACPs in mouse model. The DAACPs can be determined by showing different retention times in LC-MS analysis. The putative D-amino acid position will be identified by the site-specific strategy.

Our collaborator reported that synaptotagmin (Syt) protein family bind Ca^{2+} and trigger exocytosis, but the Syt IV protein appears to have no Ca^{2+}-dependent actions and its biological functions remain obscure.[2] It is hypothesized that Syt IV levels could change during these endocrine transitions to alter the release of oxytocin and vasopression. Therefore, it will be interesting to monitor the global neuropeptide level changes induced by Syt IV variation. Based on the preliminary data obtained above, we will quantify the neuropeptidomes in the posterior

and anterior lobes of pituitary gland between the wild type and Syt IV knockout mice. The results will provide an overview of the neuropeptide alteration and offer new insights that may help better understanding the functions of Syt IV.

Microdialysis is a minimally-invasive sampling technique that is used for continuous measurement of analyte concentrations in the extracellular fluid of tissues and organs. The neuropeptides can be sampled while the animal is behaving to correlate dynamic changes with behavior. Our laboratory has developed a new technique that utilizes antibody-coated magnetic nanoparticles for affinity-enhanced microdialysis, which significantly improve the peptide recovery and reduce the background noise [3]. In Chapter 6 and 7, our results indicate that a number of neuropeptides are up-regulated because of feeding. To further investigate their expression and release from neuroendocrine organs, we will apply the affinity-enhanced microdialysis to monitor these neuropeptides upon feeding at continuous time course. The combination of peptide secretion and expression changes would provide a more complete picture of signaling peptides involved in feeding behavior.

In Chapter 4, we developed a site-specific strategy for localization of D-amino acids in peptide epimers. The protonated peptide fragment ions were examined to compare the drift time difference between the two peptide epimers. In our recent study, we found that in some cases the sodium adducts of peptide epimers display larger drift time shift than the protonated forms. Therefore, it will be interesting to investigate the fragment ions of metal ion adducts of peptide epimers by IMS, which may offer a more efficient approach for determination of D-amino acid positions and obtain more useful structural information.

Previous reports have established a collision cross-section (CCS) database of protonated peptide ions. These high confidence level peptide CCS values define the regions of conformation

space occupied by peptide $[M+H]^+$ ions, which offer a screening method for identifying isomeric or post-translationally modified peptides. In Chapter 6 and 7, we have established a method for accurate measurement of the CCSs of peptide fragment ions, raising a question if these CCSs of peptide fragment ions can be included in a database and improve the identification accuracy of neuropeptides. Herein, we propose to establish a specialized database containing the CCS values of the neuropeptide fragment ions. The tissue extracts of crustacean neural organs will be analyzed by LC-MS/MS in a data-dependent acquisition (DDA) mode followed by on-line IMS analysis. The data will be searched using software PEAKS against a home-built crustacean neuropeptide database (described in Chapter 7) to identify the neuropeptides and generate a peak list file containing the masses of peptides and fragment ions. The drift times of the fragment ions can be extracted from the raw data by our home-built algorithm pepCCScal and then the drift times can be calibrated to generate CCSs by the same algorithm. The resulting CCS dataset of neuropeptide fragment ions will be included into the database to facilitate identification.

8.3 References

[1] Fricker LD. Analysis of mouse brain peptides using mass spectrometry-based peptidomics: implications for novel functions ranging from non-classical neuropeptides to microproteins. Mol Biosyst. 2010;6:1355-65.

[2] Zhang Z, Bhalla A, Dean C, Chapman ER, Jackson MB. Synaptotagmin IV: a multifunctional regulator of peptidergic nerve terminals. Nature neuroscience. 2009;12:163-71.

[3] Schmerberg CM, Li L. Mass spectrometric detection of neuropeptides using affinity-enhanced microdialysis with antibody-coated magnetic nanoparticles. Analytical chemistry. 2013;85:915-22.

Appendix

List of Publications and Presentations

Publications:

1. **Jia C.**, Yu Q., Ouyang C., Ma F., Schmerberg C.M., and Li L., Rapid and Sensitive Characterization of Neuropeptidome in Dungeness Crabs on a Q-Exactive High-Resolution Mass Spectrometer and Peptide Alterations in Response to Feeding. ***EuPA Open Proteomics***, to be submitted.

2. **Jia C.**, Yu Q., Wang J., and Li L., Qualitative and Quantitative Top-down Mass Spectral Analysis of Crustacean Hyperglycemic Hormones in Response to Feeding. ***Proteomics***, under revision. (A special issue for top-down proteomics)

3. **Jia C.**, Wu Z., Lietz C.B., Liang Z., Cui Q., and Li L., Gas-phase Ion Isomer Analysis Reveals the Mechanism of Peptide Sequence Scrambling. ***Anal Chem***, under revision.

4. **Jia C.**, Lietz C.B., Yu Q., and Li L., Site-Specific Characterization of D-Amino Acid-Containing Peptide Epimers by Ion Mobility Spectrometry. ***Anal Chem***, submitted.

5. **Jia C.**, Lietz C.B., Ye H., Hui L., Yu Q., Yoo S., and Li L., A Multi-scale Strategy for Discovery of Novel Endogenous Neuropeptides in the Crustacean Nervous System. ***Journal of Proteomics***, 2013, 91:1-12.

6. **Jia C.**, Hui L., Cao W., Lietz C.B., Jiang X., Chen R., Catherman A.D., Thomas P.D., Ge Y., Kelleher N.L. and Li L., High-definition De Novo Sequencing of Crustacean Hyperglycemic Hormone (CHH)-family Neuropeptides. ***Molecular and Cellular Proteomics***, 2012, 11:1951-64.

7. Ye H., Wang J., Zhang Z., **Jia C.**, Schmerberg C., Catherman A. D., Thomas P. M., Kelleher N. L., Baro D. J., Li L. Defining the Neuropeptidome of the Spiny Lobster *Panulirus interruptus* Brain Using a Multi-Dimensional Mass Spectrometry-Based Platform. ***Journal of Proteome Research***, to be submitted.

8. Zhang Z., **Jia C.**, Li L.. Neuropeptide Analysis with Liquid Chromatography-Capillary Electrophoresis-Mass Spectrometric Imaging. ***Journal of Separation Science***. 2012, 35: 1779-1784. (This is an invited contribution).

9. Hui, L., D'Andrea, B. T., **Jia, C.**, Liang, Z., Christie, A. E., and Li, L. Mass Spectrometric Characterization of the Neuropeptidome of the Ghost Crab *Ocypode Ceratophthalma* (Brachyura, Ocypodidae). ***Gen Comp Endocrinol***, 2013,184: 22-34.

10. Hui L., Cunningham R., Zhang Z., Cao W., **Jia C.**, Li L. Exploring the CPRP and Orcokinin Peptides from the *Callinectes sapidus* Sinus Glands Using a Multifaceted Mass Spectrometry Approach. ***Journal of Proteome Research***. 2011, 10: 4219-4229.

Presentations:

1. **Jia C.**, Wang J., Hui L., Schmerberg C., Ye H. and Li L., "Mass Spectral Analysis of Crustacean Hyperglycemic Hormone Precursor Related Peptides," 60th ASMS Annual Conference, May 24-29, **2012**, Vancouver, Canada.
2. **Jia C.**, Jiang X., Xiang F., Liang Z., Hui L., and Li L., "Quantitation study of biogenic amines and crustacean hyperglycemic hormone (CHH)-family peptides in crustacean nervous system by novel DiLeu labeling technique," 59th ASMS Annual Conference, June 5-9, 2011, Denver, CO.

3. **Jia C.**, Ye H., Ane J. M., and Li L., "A Multi-faceted MS Strategy for *De Novo* Sequencing of the Nodule-Specific Cysteine-Rich Peptides in *Medicago truncatula*," PittCon, March 13-17, Atlanta, GA, **2011**.

4. **Jia C.**, Hui L., Chen R., Zhang Y., and Li L., "Mapping of Crustacean Hyperglycemic Hormone (CHH) Family Neuropeptides and Their Roles in Response to Environmental Stimuli," Poster Presentation, P*roceedings of the 58th ASMS Conference on Mass Spectrometry and Allied Topics,* Salt Lake City, UT, **2010**.

5. Wang J., **Jia C.**, Chen R., and Li L., "Disulfide-bridged neuropeptide screening by off-line CE/MALDI-MS," Poster presentation, PittCon, Chicago, IL, **2009**.