Mass Spectrometric Application in Neuroscience

By

Jingxin Wang

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Neuroscience)

at the

UNIVERSITY OF WISCONSIN-MADISON

2016

Date of final oral examination: 05/19/16

The dissertation is approved by the following members of the Final Oral Committee:

Lingjun Li, Professor, Pharmaceutical Sciences and Chemistry
Brian Baldo, Associate Professor, Psychiatry
Chris Ikonomidou, Professor, Neurology
Ozioma Okonkwo, Assistant Professor, Medicine
Robert Thorne, Assistant Professor, Pharmacy
Su-Chun Zhang, Professor, Neuroscience and Neurology

Acknowledgements

I would first like to acknowledge my advisor Dr. Lingjun Li, without whom none of the work in my thesis would be accomplished. I would like to thank her for accepting me into her lab firstly. I still remember how thrilled I was the first day I started working in our lab. Dr. Li is one of the most outstanding and diligent scientists I have ever met, who has established a world-class research program in the field of mass spectrometry and neuroscience. She is a great role model for everyone in our lab by devoting herself to pursuing the truth in science. The light in her office has never been off until late night even in the weekend, which has turned into a light tower that guides us to become a future scientist. Dr. Li is also a wonderful mentor who is always encouraging and has guided me to become an independent researcher. I really appreciate the freedom she gave me in delving into the directions and projects that I am interested in. She has always been supportive and patient that helped me overcome the obstacles and challenges in my research. Last but not least, Dr. Li is a very kind person. I'm extremely grateful for all the support and help that she gave me for my residency application and pursuing my future career as a physician scientist.

I also would like to thank my thesis committee, Dr. Hrissanthi Ikonomidou, Dr. Brian Baldo, Dr. Ozioma Okonkwo, Dr. Su-Chun Zhang and Dr. Robert Thorne for providing invaluable support and suggestions throughout my dissertation preparation process. Their guidance in various perspectives and constructive criticism has substantially improved my capacities of critical thinking. Specifically, I would like to thank Dr. Hrissanthi Ikonomidou for teaching me valuable experience in neurology practice and her great help for my residency application. Also I'm very grateful that I spent a year in Dr. Baldo's lab to conduct behavioral studies and I really enjoyed performing neurosurgery on rats. Dr. Okonkwo has led me into the field of Alzheimer's disease

that sparkled my interest in neurodegenerative diseases. In addition, I appreciate Dr. Zhang and Dr. Thorne's time and feedback regarding my thesis proposal and dissertation. All your support and guidance have been instrumental to my pursuit in becoming a better scientist.

Furthermore, I would like extend my gratefulness to Ryan Selleck, Ken Sadeghian, and Dr. Vaishali Bakshi for all their kind help while I was at Dr. Baldo's lab. I also enjoyed the collaboration work with Dr. Yan Zhang and Dr. Shigetoshi Yokoyama in Dr. Xin Sun's lab. In addition, we had a very productive collaboration with Dr. Rikki Hullinger in Dr. Luigi Puglielli's lab. It was my pleasure to have the opportunities to work with these wonderful people. I have learned so much from them and their achievements in the field have inspired me to further expand my research horizon. I also would like to thank our former lab members Dr. Robert Cunningham and Dr. Hui Ye for their guidance when I initially started my research at Li lab. Xiaofang Zhong, Qing Yu, Fengfei Ma, Zichuan Tian have kindly assisted me with projects related to aspects of my dissertation work. In addition, I have benefited from the inspiring discussion with Dr. Chenxi Jia, Dr. Zichuan Zhang, Dr. Zhidan Liang and Shan Jiang. Dr. Cameron Scarlett and Molly Hahn from the Analytical Instrument Center of School of Pharmacy have also helped me with instrument maintenance and troubleshooting. I'm also grateful to the Neuroscience Training Program for all their support and coordination.

Last but not least, I would like to thank my family and friends. The love and support from parents, my grandparents, Gregory, Nancy, Stephen, Bob and Qingyun Zou have kept me moving forward and pursuing my career in the US. I'm so grateful that Dr. Dandan Sun introduced me to Li Lab. I also feel extremely lucky to be surrounded by good friends: Yun Ding, Zhidan Liang, Shan Jiang, Qing Yu, Shiyu Chen, Yan Wang, Ling Hao, Rikki Hullinger, Christian La, Trina Basu, and Erin Gemperline among others in Madison during the past five years.

Table of Contents

Mass Spectrometric Application in Neuroscience

Jingxin Wang

Under the supervision of Professor Lingjun Li

at the University of Wisconsin-Madison

Abstract

Mass spectrometry based comparative proteomics and peptidomics for neuroscience application have been reported in this thesis using different animal and human models. These work are enabled by improved methods that have been developed for neuropeptide characterization and shotgun proteomics. Chapter 1 presents an overview of the background information of each project described in this thesis. In Chapter 2 we employed a combination of cryostat dissection, heat stabilization, neuropeptide extraction and label-free quantitative neuropeptidomics via a liquid chromatography-high resolution mass spectrometry platform to monitor the feeding-induced changes in neuropeptide expression levels within the nucleus accumbens (Acb). We showed feeding caused the expression level changes of ProSAAS. We further performed a microinjection using ProSAAS neuropeptides to the rat Acb. We showed that at high concentration, big LEN significantly increased rats' food and water intake. In addition, both big LEN and PEN at high concentration altered locomotion and rearing.

Chapter 3 details my work in Alzheimer's disease (AD) biomarker discovery. By optimized molecular weight cutoff filtration and in house-constructed database, 645 peptides were identified. Interestingly, 42 proSAAS peptides discovered were exclusively cleaved either from N- or C-termini of the proSAAS protein precursor. Glycoproteins were enriched by lectin affinity

chromatography, enabling identification of 795 glycoproteins. One way ANOVA was applied for statistical analysis, a total 15 proteins were found differentially expressed among the three groups. The dynamic changes of transthyretin that exhibited initial increase in MCI followed by decrease in AD were reported for the first time.

In Chapter 4, a comparative study of membrane protein in mouse hippocampi was conducted by shotgun proteomics. In this study, AT-1 Tg (acetyltransferase) mouse model that selectively overexpresses human AT-1 in neurons. We found a large set of differentially expressed proteins that are involved in neuronal migration and adhesion, neurite outgrowth, transport of synaptic vesicles, assembly of synaptic connections, and regulation of synaptic activity.

Chapters 5 and 6 reported on our method development for neuropeptide characterization and detection. Conclusions and research summaries are presented in Chapter 7. Collectively, this dissertation develops several improved mass spectrometric methods for neuropeptide analysis and showcases several important applications in neuroscience.

Chapter 1

Introduction

1.1 Feeding Behavior

Feeding behavior Feeding behavior is critical for animal survival and is also a fundamental aspect of energy homeostasis. The rising prevalence of obesity and anorexia nervosa further highlights the needs for a thorough understanding of the mechanism involved in the regulation and signaling pathways of food intake and energy homeostasis.[1, 2, 3, 4] Physiological functions such as food intake and energy homeostasis are affected by a bewildering number of endogenous compounds including a multitude of neuropeptides (NPs) and biogenic amines from a highly complex neuroendocrine system. A clear understanding of the neural circuitry that controls feeding behavior requires knowledge of the full cast of molecular components and their mode of action on each of the elements of the circuit. However, the growing number of molecules implicated in feeding behavior, the wide distribution and low concentration of NPs pose significant challenges to elucidate the molecular mechanism at cellular and network level. Despite decades of work on individual neurotransmitter or peptidergic signaling systems, the general principles underlying neuromodulation are still poorly understood. This is partially due to a lack of analytical capabilities to identify and measure these low abundance endogenous signaling molecules in a complex microenvironment. Therefore, I hypothesize that the development of highly sensitive and high resolution analytical tools for NP identification and quantification will enable comprehensive characterization of NPs, including classical NPs and potentially bioactive peptide fragments derived from prohormone precursors, and further elucidate the complex functional roles of low abundance NPs in food intake. The overall goals of my proposed research are to develop new bioanalytical methods to enrich and extract low abundance NPs; to comprehensively identify and quantify the NPs involved in food intake

by an array of novel mass spectrometry (MS) based strategies with improved sensitivity and specificity; and to validate potential targets by selected ion monitoring (SIM) MS strategy combining microinfusion and behavioral tests, and to map the co-localization of targeted NPs by MS imaging (MSI) method. These MS-based approaches coupled with biological and functional analyses will offer an unprecedented opportunity to investigate the intriguing questions.

Neuropeptides NPs are a class of neural signaling molecules that modulate important physiological functions, including feeding, reward, fear, pain, anxiety, sleep, reproduction, and learning and memory. In order to be classified as a NP, the molecule must exhibit bioactivity in the central nervous system with many NPs acting locally in cell-to-cell communication and others acting systemically as hormones. NPs are synthesized in the rough endoplasmic reticulum as immature prepropeptides. The transformation of a prepropeptide into a mature NP requires a variety of processing enzymes including peptidases, sulfatases, and prohormone convertase 1 and 2, which typically cleave the propeptide at dibasic sites to yield a semi-mature neuropeptide with a dibasic extension at the *C*-terminal.[5] This basic residue is then removed by carboxypeptidase E.[6] After the removal of the *C*-terminal basic residue, about half of the known neuropeptides subsequently undergo *C*-terminal amidation, a modification to enhance receptor binding and protect the neuropeptide from extracellular peptidases.[7] A large number of other modifications can also occur in the secretory vesicles, including sulfation, acetylation, phosphorylation and glycosylation, which can affect the biological properties of the NPs. For instance, *N*-terminal acetylation of β - endorphin eliminates its ability to bind the

opioid receptor.[8] The neuropeptide precursors can contain several copies of the same peptide (for example, preprothyrotropin-releasing hormone), but can also hold several different neuropeptides (for example, pro-opiomelanocortin). Once the immature neuropeptides reach the *trans*-Golgi, it is sorted and packaged into secretory vesicles and then transported to the inner leaf of the cell membrane.[9] The regulation of neuropeptide expression is a cell-specific phenomenon. Different processing of these precursors can lead to the generation of distinct neuropeptides with novel biological activities. Upon the proper stimulus, usually a rise in cytoplasmic calcium, the secretory vesicles fuse with the cell membrane, releasing the NPs into extracellular space.[10] Once released into the extracellular space, the NPs bind their G-protein coupled receptors and modulate neuronal activity in conjunction with the co-localized neurotransmitters. Unlike classic neurotransmitters, no known re-uptake or transport mechanisms exist for NPs and thus termination of the NP signal is processed by the cleavage of the NP by extracellular peptidases.[11]

Energy Homeostasis and Hypothalamic NPs Food intake involves several aspects of different behaviors including decision-making and hunting for food. The complexity of feeding behavior is in part reflected in the number of brain areas involved. For example, the orbitofrontal cortex is involved in sensory-specific satiety,[12] while the amygdala is implicated in the evaluation of taste palatability.[13] Hence, feeding behavior can be divided into different phases, such as the appetitive phase, which involves the searching for food, and a consummatory phase, which involves the real eating of the food. Besides the more central effects, feeding behavior is also influenced by a peripheral system, which uses, for instance, sensory and gastrointestinal cues to terminate ingestion. The modulation of food

intake by the central as well as the peripheral system is amongst others achieved by NPs. Regarding the different behaviors involved, it is logical that the NPs controlling feeding behavior are not only implicated in feeding behavior, but also modulate other processes like, memory and analgesia.[14] Despite the large number of brain areas involved in feeding behavior, the hypothalamus is still regarded as the main feeding center of the brain. Food intake is thought to be under the control of two regions in the hypothalamus: a ventromedial region and a lateral region. This 'Dual Center Model' theory was formed due to the discoveries that destruction of the ventromedial nuclei of the hypothalamus produces overeating (hyperphagia) and severe obesity. In contrast, bilateral lesions of the lateral hypothalamus produce severe neglect of eating (aphagia) so that the animal dies unless force-fed. These observations were originally interpreted to suggest that the lateral hypothalamus contains a feeding center and the medial hypothalamus a satiety center. Although this model has been questioned several times, the hypothalamus still appears to be the most important area for regulation of food intake and body weight homeostasis in the brain. The hypothalamus consists of several nuclei involved in food intake, including the arcuate nucleus (ARC), the paraventricular nucleus (PVN), the lateral hypothalamic area (LHA), the ventromedial nucleus (VMH), and the dorsomedial nucleus (DMH). ARC neurons are located at the bottom of the hypothalamus around the third ventricle and are called 'first order neurons' because of their 'direct' contact with peripheral satiety factors like leptin, ghrelin and insulin. This is due to the fact that in the median eminence (ME), which overlies the ARC, the blood brain barrier (BBB) is not present and ARC axon terminals are in direct contact with the bloodstream. However, the neuronal cell bodies in the ARC are protected by the BBB and thus are not in direct contact with the bloodstream.[15]

Thus, the ARC relays energy information from the peripheral organs, such as the gut, liver and adipose tissue, to the higher structures in the brain involved in decision making and motivation. In turn, the ARC is also responsible for sending signals back to the periphery via a specialized, hormone secreting organ – pituitary. This system affords a back and forth communication between the body and higher brain centers about the energy state of the organism. In addition, ghrelin and leptin exhibit opposing bioactivities, with ghrelin stimulating feeding and leptin suppressing feeding. Both of these peripheral signals modulate feeding via NP signaling in the ARC.[16, 17, 18, 19, 20, 21] The ARC contains at least two distinct groups of neurons controlling energy balance, e.g. neurons that contain the orexigenic neuropeptides agouti-gene-related protein (AgRP) and neuropeptide Y (NPY), and neurons that contain the anorexigenic neuropeptides pro-opiomelanocortin (POMC) and cocaine- and amphetamine-regulated transcript (CART). Both NPY/AgRP and POMC/ CART neurons express both leptin and ghrelin receptors. Leptin increase activity in POMC/CART cells while inhibiting activity in the NPY/AgRP cells while ghrelin does the opposite.[22] From the ARC, neurons project to 'second order neurons' in the PVN, VMH, DMH, and LHA.[2] 'Second order neurons' project to the nucleus of the solitary tract (NTS) in the brainstem and the dorsomotor nucleus of the vagus (DMV). This communication between hypothalamic pathways and the caudal brainstem, responding to meal-related satiety signals, is essential for the long-term regulation of energy homeostasis.

Hedonistic Overlap between Feeding and Drug Addiction Motivation and reward have been studied most extensively in regard to drugs of addiction. However, a number of studies have suggested an overlap of neurobiological mechanisms between drug addiction

and the hedonistic aspects of feeding.[23] Electrical stimulation of the LH is inherently rewarding and LH probe-implanted rats work vigorously to obtain stimulation. This stimulation also caused a concurrent increase in food intake.[24] Alternatively, food deprivation increases the self – administration of non-food reward, such as psychostimulants, intracranial stimulation, and heroin, suggesting a significant link between food reward and drug rewards.[25, 26] The fact that endogenous opioids were found to be involved in a wide range of physiological processes, including the modulation of pain, stress, endocrine response, motivation states, feeding and addiction further enforced those linkages.[27] Early studies examining the effects of opiates on rats found morphine injections dramatically increased food intake.[28] These initial experiments led to a number of subsequent studies using opioid agonist and antagonists, with the agonists increasing food consumption and antagonists decreasing food intake, solidifying the specific involvement of endogenous opioids in the regulation of feeding.[29] More recent studies using specific μ − opioid agonists, such as D – Ala2, Nme – Phe4, Gyo15 – enkephalin (DAMGO), have shown the increase in food intake to be modulated by the μ - opioid receptors in the nucleus accumbens (NAc).[30] These experiments were performed with satiated rats which had just eaten, and via the infusion of DAMGO into the NAc the rats resumed eating, consuming up to 400% of their normal intake of food. Additionally, the DAMGO rats displayed a preference for high – fat food.[31] The actions of opioids have upon feeding behavior lies that the neurons in the NAc are primarily GABAergic neurons which express either preproenkephalin or prodynorphin. The mature NP forms of preproenkephalin include both Met – and Leu – enkephalin (μ - opioid receptor agonists) while prodynorphin is processed into dynorphin A and B (κ - opioid receptor agonists). Both types of medium spiny neurons

project to various parts of the brain, including connections to the ARC and the LH. DAMGO infusion into the NAc causes μ- opioid receptor activation and consequent inhibition of GABA release from the affected medium spiny neurons. The projections of NAc to both LH and ARC, as well as to the ventral pallidum (VP) and the ventral tegmental area (VTA) make the opioids agonists increase food intake.

Significance Obesity represents one of the most urgent global health problems as well as one of the leading causes of death throughout industrialized nations. This disease can be caused by disorder of food intake, imbalance of energy storage or energy expenditure, all of which are associated with multiple endocrine alterations as defined by changes in the level of circulating hormones and changes in the pattern of secretion or clearance of these circulating molecules. Of these alterations, some are most certainly secondary to the development of obesity, whereas others may be putative causative factors. However, very little is known about the molecular mechanisms. While much research has focused on mechanisms controlling food intake and appetite, there has been little success in developing effective pharmacological treatments, partly due to the complex interactions between neural networks and the many neurochemical systems that modify the output of these networks. To further elucidate these complex interactions, new tools and tractable model systems are needed. Although many NPs in the mammalian hypothalamus [e.g. NPY, POMC, melanin-concentrating hormone (MCH), neurotensin, cholecystokinin (CCK)] either stimulate or decrease food intake,[4] their interaction with each other, which is pivotal to their functions, is largely unknown due to the limited accessibility of the mammalian central nervous system (CNS).

Challenges for Neuropeptide Analysis Several technical challenges exist for studying NPs, including very low concentrations (pM-nM level *in vivo*) and the presence of highly abundant proteins in brain tissues. Extensive posttranslational modifications and processing often complicates prediction of bioactive product from the DNA sequence. Finally, the extreme chemical diversity of NPs often requires analytical methods with high resolution and specificity. Conventional methods for analysis of NPs and neurotransmitters often employ the Edman degradation or immunohistochemical techniques.[32] Although these methods have been well established and are able to provide accurate sequence and quantification information, they suffer from several limitations. The Edman degradation sequencing technique needs a large amount of sample and extensive purification steps. The immunohistochemistry is based on antibody recognition, where antibody could have cross-affinity to structurally similar molecules and lacks the capability to simultaneously detect multiple analytes.[33] These limitations dramatically lower the analytical specificity and throughput. In the past two decades, mass spectrometry (MS) has emerged as a central tool for the analysis of biological macromolecules and small molecules due to its high performance and speed. A series of MS-based strategies for neuropeptide analysis have been developed in our lab, involving *de novo* peptide sequencing, bottom-up and top-down peptide sequencing, isotopic labeling peptide quantification, dimethylated leucine (DiLeu) isobaric quantification technique, etc. [34, 35] In this project, I will focus on developing MS-based methodologies for NPs sequencing, structure elucidation and quantification in the context of feeding.

1.2 Alzheimer's disease

Alzheimer's disease (AD) is the most common form of dementia in the elderly population and the 6th leading cause of death in the United States. More than 36.5 million people were estimated to suffer from dementia in 2010, and there are 7.7 million new cases each year. In addition, the number of dementia patients will nearly double every 20 years. Due to the aging baby boom generation, it is predicted that the number of people with Alzheimer's disease age 65 and older may triple by 2050, costing an estimated $1.2 trillion.[36] AD is characterized by progressive accumulation of extracellular β-amyloid (Aβ) peptides and intracellular neurofibrillary tangles of protein tau, in addition to synaptic and neuronal loss in the brain.[37] Currently, the diagnosis of AD is mainly based on the history of a patient or an objective cognitive assessment and the exclusion of dementia due to other causes such as Lewy body dementia, vascular dementia, frontotemporal dementia, Creutzfeldt-Jakob disease (CJD) etc.[38] Three AD biomarkers: Aβ 1-42, total tau (t-tau) and phosphorylated tau (p-tau) in the cerebrospinal fluid (CSF) have been used to facilitate the diagnostic criteria.[39, 40, 41, 42] These biomarkers may increase the certainty of AD pathophysiological process only in patients who have already met the critical clinical criteria for probable AD dementia. However, the AD biomarker tests are not advocated routinely for diagnostic purposes these days. The reason is mainly due to the limitation of utilizing these biomarkers to predict and monitor the disease progression. Specifically, the recently reported failure of a phase 3 clinical trial of humanized monoclonal antibody of soluble forms of amyloid[43, 44] necessitates the discovery and establishment of other biomarkers and targets for disease prevention and/or treatment.

CSF Although lumbar puncture is an invasive procedure, CSF bathes the brain and spinal cord which makes it a valuable biological fluid for biomarker studies of neurodegenerative diseases.[45] CSF offers promise to facilitate disease diagnosis, monitoring disease progression, and the efficacy of potential therapies. CSF is generated in the choroid plexus and composed of blood filtrate and brain extracellular matrix. Despite the differences between CSF and plasma the proteome of both share the vast majority of abundant protein identifications, although it is estimated that around 20% of the CSF proteome are produced by brain itself.[46, 47] Compared to a brain biopsy which could cause permanent damage, and approximately 500 ml of CSF is generated each day with a maximum of 20 ml sampled per spinal tap.[48] CSF is reabsorbed into the blood four times per day and due to the constant turnover the dynamic changes of CSF composition is able to reflect the disease status of the central nervous system (CNS). In a clinical setting, the lumbar puncture is a routine procedure for the diagnosis of multiple sclerosis, Guillain-Barré syndrome and CJD etc. Previous large scale proteomic studies have greatly filled in the gap of our knowledge about CSF protein composition by two dimensional gel,[49, 50, 51, 52] chromatography and mass spectrometry.[53, 54, 55, 56, 57]

Glycosylation and AD Glycosylation is one of the most complicated but common forms of post-translational modifications in proteins. It modulates cell-molecule, cell-matrix and cell-cell interactions and facilitates the assembly and development of complex organisms. The major functions of a glycan can be characterized into two categories: (1) structural and modular functions; and (2) recognition of glycans by other molecules.[58] There are two major forms of protein glycosylation: glycans linked to an asparagine (N-glycans) and

glycans linked to a serine/threonine (O-glycans). N-glycosylation occurs in the endoplasmic reticulum (ER) while O-glycosylation is more dynamic occurring in nuclear and cytoplasmic proteins. Lectin affinity chromatography is a useful and practical separation technique in glycoproteomics. It is characterized by selectively binding a specific carbohydrate motif. Among them, concanavalin-A (ConA) and wheat germ agglutinin (WGA) are most commonly used. Specifically, ConA has higher affinity for N-linked glycoproteins with high-mannose and terminal glucose carbohydrates while WGA has a higher affinity for terminal acetylglucosamine and sialic acid with O-linked glycosylation motif. In addition, ConA has a special binding site for excessively hydrophobic proteins while WGA is able to bind both hydrophobic and hydrophilic proteins.[59] Their functional properties are useful to utilize as a combination of both ConA and WGA which can obtain an optimal and impactful glycoprotein enrichment. Previous studies indicate that the global protein glycosylation pattern is altered in AD.[60, 61, 24] Recently, the potential precursor of AD-mild cognitive impairment (MCI) has drawn increased attention since treatments could potentially be initiated before the damage of the neurodegenerative process becomes too extensive.[62, 63] One can investigate patients with cognitive impairment that is worse than age-matched healthy individuals but not severe enough to fulfill the criteria for suspected AD. People with MCI display similar pathophysiologic symptoms as those with AD and have a 10%-15% relative risk of becoming AD annually.[64, 65] Here we reported a comparative glycoproteomic and peptidomic discovery study using CSF from healthy (control), MCI and AD individuals via mass spectrometry (MS). This study demonstrates the usefulness of glycoprotein enrichment in biomarker discovery which facilitates the isolation, identification and

relative quantification of lectin affinity enriched glycoproteins from CSF samples via multidimensional separation and high resolution accurate mass tandem mass spectrometry.

1.3 Acetyl-CoA and neurological disease

Acetyl-CoA is an essential substrate for a wide range of biochemical reactions that occur within the cell. [66] Cytosolic acetyl-CoA is produced predominantly by the conversion of citrate and coenzyme A (CoA) by ATP-citrate lyase (ACLY) and the condensation of free acetate and CoA by acetyl-CoA synthetase (ACESS2, also referred to as AceCS).[67] Acetyl-CoA is then actively imported into the lumen of the endoplasmic reticulum (ER) by the ER membrane transporter AT-1 (also referred to as SLC33A1),[68] where it serves as donor of the acetyl group in the Nε-lysine acetylation of ER-resident and -transiting proteins.[69] Recent reports suggest that lysine acetylation within the ER is required for ER-mediated quality control. Specifically, the ER based acetyltransferases ATase1 and ATase2 associate with the oligosaccharyl transferase complex to acetylate properly folded glycoproteins.[70] In addition, studies performed with two type I membrane proteins indicate that the acetylation status of nascent secretory proteins in the ER regulates the efficiency of transport along the secretory pathway.[71] Decreased influx of acetyl-CoA into the ER lumen leads to aberrant induction of autophagy in both cell-based [68, 72] and animal [73] models. At the mechanistic level, the induction of autophagy is linked to the acetylation status of autophagy-related protein 9A (ATG9A).[72-73] Haploinsufficiency of AT-1 in the animal results in neurodegeneration, inflammation, and propensity to infections and cancer.[73] Heterozygous mutations in AT-1 have been identified in patients affected by an autosomal dominant form of spastic paraplegia[74] while homozygous mutations have been identified

in patients affected by severe developmental delay and childhood death.[75] Chromosomal duplications affecting the 3q25.31 locus harboring *AT-1/SLC33A1* have been associated with autism spectrum disorder (ASD) and intellectual disability (see *SFARI-Autism Database*, http://sfari.org/).[76] Additionally, a gain of approximately 1.1-1.5 Mb in 3q25.2-3q25.31, which contained *AT-1/SLC33A1* and *Guanine Monphosphate Synthase* (*GMPS*, associated with myeloid leukemia) was found in three male children with autism, seizure, abnormal electroencephalogram, and facial dysmorphism (Swisshelm K., et al. ASHG Annual Meeting, October 18-22, 2014, San Diego, CA).

References:

1. Woods, S. C.; Seeley, R. J.; Porte, D., Jr.; Schwartz, M. W., Signals that regulate food intake and energy homeostasis. *Science* **1998,** *280* (5368), 1378-83.
2. Schwartz, M. W.; Woods, S. C.; Porte, D., Jr.; Seeley, R. J.; Baskin, D. G., Central nervous system control of food intake. *Nature* **2000,** *404* (6778), 661-71.
3. Benarroch, E. E., Neural control of feeding behavior: Overview and clinical correlations. *Neurology* **2010,** *74* (20), 1643-50.
4. Hillebrand, J. J.; de Wied, D.; Adan, R. A., Neuropeptides, food intake and body weight regulation: a hypothalamic focus. *Peptides* **2002,** *23* (12), 2283-306.
5. Zhou, A.; Webb, G.; Zhu, X.; Steiner, D. F., Proteolytic processing in the secretory pathway. *J Biol Chem* **1999,** *274* (30), 20745-8.
6. Fricker, L. D.; Snyder, S. H., Enkephalin convertase: purification and characterization of a specific enkephalin-synthesizing carboxypeptidase localized to adrenal chromaffin granules. *Proc Natl Acad Sci U S A* **1982,** *79* (12), 3886-90.
7. Prigge, S. T.; Mains, R. E.; Eipper, B. A.; Amzel, L. M., New insights into copper monooxygenases and peptide amidation: structure, mechanism and function. *Cell Mol Life Sci* **2000,** *57* (8-9), 1236-59.
8. Akil, M.; Lewis, D. A., The dopaminergic innervation of monkey entorhinal cortex. *Cereb Cortex* **1993,** *3* (6), 533-50.
9. Pritchard, L. E.; White, A., Neuropeptide processing and its impact on melanocortin pathways. *Endocrinology* **2007,** *148* (9), 4201-7.
10. Hokfelt, T.; Broberger, C.; Xu, Z. Q.; Sergeyev, V.; Ubink, R.; Diez, M., Neuropeptides--an overview. *Neuropharmacology* **2000,** *39* (8), 1337-56.
11. Shrimpton, C. N.; Smith, A. I.; Lew, R. A., Soluble metalloendopeptidases and neuroendocrine signaling. *Endocr Rev* **2002,** *23* (5), 647-64.
12. O'Doherty, J.; Rolls, E. T.; Francis, S.; Bowtell, R.; McGlone, F.; Kobal, G.; Renner, B.; Ahne, G., Sensory-specific satiety-related olfactory activation of the human orbitofrontal cortex. *Neuroreport* **2000,** *11* (4), 893-7.
13. Nishijo, H.; Uwano, T.; Tamura, R.; Ono, T., Gustatory and multimodal neuronal responses in the amygdala during licking and discrimination of sensory stimuli in awake rats. *J Neurophysiol* **1998,** *79* (1), 21-36.
14. Morley, J. E., Neuropeptide regulation of appetite and weight. *Endocr Rev* **1987,** *8* (3), 256-87.
15. Peruzzo, B.; Pastor, F. E.; Blazquez, J. L.; Schobitz, K.; Pelaez, B.; Amat, P.; Rodriguez, E. M., A second look at the barriers of the medial basal hypothalamus. *Exp Brain Res* **2000,** *132* (1), 10-26.
16. Bagnasco, M.; Kalra, P. S.; Kalra, S. P., Ghrelin and leptin pulse discharge in fed and fasted rats. *Endocrinology* **2002,** *143* (2), 726-9.
17. Horvath, T. L.; Diano, S.; Sotonyi, P.; Heiman, M.; Tschop, M., Minireview: ghrelin and the regulation of energy balance--a hypothalamic perspective. *Endocrinology* **2001,** *142* (10), 4163-9.
18. Kojima, M.; Hosoda, H.; Date, Y.; Nakazato, M.; Matsuo, H.; Kangawa, K., Ghrelin is a growth-hormone-releasing acylated peptide from stomach. *Nature* **1999,** *402* (6762), 656-60.
19. Nakazato, M.; Hanada, R.; Murakami, N.; Date, Y.; Mondal, M. S.; Kojima, M.; Yoshimatsu, H.; Kangawa, K.; Matsukura, S., Central effects of neuromedin U in the regulation of energy homeostasis. *Biochem Biophys Res Commun* **2000,** *277* (1), 191-4.
20. Ahima, R. S.; Saper, C. B.; Flier, J. S.; Elmquist, J. K., Leptin regulation of neuroendocrine systems. *Front Neuroendocrinol* **2000,** *21* (3), 263-307.
21. Frederich, R. C.; Lollmann, B.; Hamann, A.; Napolitano-Rosen, A.; Kahn, B. B.; Lowell, B. B.; Flier, J. S., Expression of ob mRNA and its encoded protein in rodents. Impact of nutrition and obesity. *J Clin Invest* **1995,** *96* (3), 1658-63.
22. Baskin, D. G.; Hahn, T. M.; Schwartz, M. W., Leptin sensitive neurons in the hypothalamus. *Horm Metab Res* **1999,** *31* (5), 345-50.
23. Saper, C. B.; Chou, T. C.; Elmquist, J. K., The need to feed: homeostatic and hedonic control of eating. *Neuron* **2002,** *36* (2), 199-211.
24. Wise, R. A., Lateral hypothalamic electrical stimulation: does it make animals 'hungry'? *Brain Res* **1974,** *67* (2), 187-209.
25. Cabeza de Vaca, S.; Carr, K. D., Food restriction enhances the central rewarding effect of abused

drugs. *J Neurosci* **1998,** *18* (18), 7502-10.
26. Carr, K. D., Feeding, drug abuse, and the sensitization of reward by metabolic need. *Neurochem Res* **1996,** *21* (11), 1455-67.
27. Kelley, A. E.; Bakshi, V. P.; Haber, S. N.; Steininger, T. L.; Will, M. J.; Zhang, M., Opioid modulation of taste hedonics within the ventral striatum. *Physiol Behav* **2002,** *76* (3), 365-77.
28. Martin, W. R.; Wikler, A.; Eades, C. G.; Pescor, F. T., Tolerance To And Physical Dependence on Morphine In Rats. *Psychopharmacologia* **1963,** *4*, 247-60.
29. Majeed, N. H.; Przewlocka, B.; Wedzony, K.; Przewlocki, R., Stimulation of food intake following opioid microinjection into the nucleus accumbens septi in rats. *Peptides* **1986,** *7* (5), 711-6.
30. Zhang, M.; Kelley, A. E., Opiate agonists microinjected into the nucleus accumbens enhance sucrose drinking in rats. *Psychopharmacology (Berl)* **1997,** *132* (4), 350-60.
31. Zhang, M.; Kelley, A. E., Enhanced intake of high-fat food following striatal mu-opioid stimulation: microinjection mapping and fos expression. *Neuroscience* **2000,** *99* (2), 267-77.
32. Chang, E. S.; Keller, R.; Chang, S. A., Quantification of crustacean hyperglycemic hormone by ELISA in hemolymph of the lobster, Homarus americanus, following various stresses. *Gen Comp Endocrinol* **1998,** *111* (3), 359-66.
33. Coon, J. J., Collisions or electrons? Protein sequence analysis in the 21st century. *Anal Chem* **2009,** *81* (9), 3208-15.
34. Chen, R.; Hui, L.; Cape, S. S.; Wang, J.; Li, L., Comparative Neuropeptidomic Analysis of Food Intake via a Multi-faceted Mass Spectrometric Approach. *ACS Chem Neurosci* **2010,** *1* (3), 204-214.
35. Li, L.; Sweedler, J. V., Peptides in the brain: mass spectrometry-based measurement approaches and challenges. *Annu Rev Anal Chem (Palo Alto Calif)* **2008,** *1*, 451-83.
36. Sosa-Ortiz, A. L.; Acosta-Castillo, I.; Prince, M. J., Epidemiology of dementias and Alzheimer's disease. *Arch Med Res* **2012,** *43* (8), 600-8.
37. Blennow, K.; de Leon, M. J.; Zetterberg, H., Alzheimer's disease. *Lancet* **2006,** *368* (9533), 387-403.
38. McKhann, G. M.; Knopman, D. S.; Chertkow, H.; Hyman, B. T.; Jack, C. R., Jr.; Kawas, C. H.; Klunk, W. E.; Koroshetz, W. J.; Manly, J. J.; Mayeux, R.; Mohs, R. C.; Morris, J. C.; Rossor, M. N.; Scheltens, P.; Carrillo, M. C.; Thies, B.; Weintraub, S.; Phelps, C. H., The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* **2011,** *7* (3), 263-9.
39. Olsson, A.; Vanderstichele, H.; Andreasen, N.; De Meyer, G.; Wallin, A.; Holmberg, B.; Rosengren, L.; Vanmechelen, E.; Blennow, K., Simultaneous measurement of beta-amyloid(1-42), total tau, and phosphorylated tau (Thr181) in cerebrospinal fluid by the xMAP technology. *Clin Chem* **2005,** *51* (2), 336-45.
40. Hampel, H.; Burger, K.; Teipel, S. J.; Bokde, A. L.; Zetterberg, H.; Blennow, K., Core candidate neurochemical and imaging biomarkers of Alzheimer's disease. *Alzheimers Dement* **2008,** *4* (1), 38-48.
41. Jack, C. R., Jr.; Albert, M. S.; Knopman, D. S.; McKhann, G. M.; Sperling, R. A.; Carrillo, M. C.; Thies, B.; Phelps, C. H., Introduction to the recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* **2011,** *7* (3), 257-62.
42. Kang, J. H.; Korecka, M.; Toledo, J. B.; Trojanowski, J. Q.; Shaw, L. M., Clinical utility and analytical challenges in measurement of cerebrospinal fluid amyloid-beta(1-42) and tau proteins as Alzheimer disease biomarkers. *Clin Chem* **2013,** *59* (6), 903-16.
43. Doody, R. S.; Thomas, R. G.; Farlow, M.; Iwatsubo, T.; Vellas, B.; Joffe, S.; Kieburtz, K.; Raman, R.; Sun, X.; Aisen, P. S.; Siemers, E.; Liu-Seifert, H.; Mohs, R.; Alzheimer's Disease Cooperative Study Steering, C.; Solanezumab Study, G., Phase 3 trials of solanezumab for mild-to-moderate Alzheimer's disease. *N Engl J Med* **2014,** *370* (4), 311-21.
44. Doody, R. S.; Farlow, M.; Aisen, P. S.; Alzheimer's Disease Cooperative Study Data, A.; Publication, C., Phase 3 trials of solanezumab and bapineuzumab for Alzheimer's disease. *N Engl J Med* **2014,** *370* (15), 1460.
45. Zhang, J.; Goodlett, D. R.; Montine, T. J., Proteomic biomarker discovery in cerebrospinal fluid for neurodegenerative diseases. *J Alzheimers Dis* **2005,** *8* (4), 377-86.
46. Blennow, K.; Hampel, H.; Weiner, M.; Zetterberg, H., Cerebrospinal fluid and plasma biomarkers in Alzheimer disease. *Nat Rev Neurol* **2010,** *6* (3), 131-44.
47. Holtta, M.; Zetterberg, H.; Mirgorodskaya, E.; Mattsson, N.; Blennow, K.; Gobom, J., Peptidome

analysis of cerebrospinal fluid by LC-MALDI MS. *PLoS One* **2012,** *7* (8), e42555.
48. Johanson, C. E.; Duncan, J. A., 3rd; Klinge, P. M.; Brinker, T.; Stopa, E. G.; Silverberg, G. D., Multiplicity of cerebrospinal fluid functions: New challenges in health and disease. *Cerebrospinal Fluid Res* **2008,** *5*, 10.
49. Raymackers, J.; Daniels, A.; De Brabandere, V.; Missiaen, C.; Dauwe, M.; Verhaert, P.; Vanmechelen, E.; Meheus, L., Identification of two-dimensionally separated human cerebrospinal fluid proteins by N-terminal sequencing, matrix-assisted laser desorption/ionization--mass spectrometry, nanoliquid chromatography-electrospray ionization-time of flight-mass spectrometry, and tandem mass spectrometry. *Electrophoresis* **2000,** *21* (11), 2266-83.
50. Sickmann, A.; Dormeyer, W.; Wortelkamp, S.; Woitalla, D.; Kuhn, W.; Meyer, H. E., Identification of proteins from human cerebrospinal fluid, separated by two-dimensional polyacrylamide gel electrophoresis. *Electrophoresis* **2000,** *21* (13), 2721-8.
51. Sickmann, A.; Dormeyer, W.; Wortelkamp, S.; Woitalla, D.; Kuhn, W.; Meyer, H. E., Towards a high resolution separation of human cerebrospinal fluid. *J Chromatogr B Analyt Technol Biomed Life Sci* **2002,** *771* (1-2), 167-96.
52. Yuan, X.; Russell, T.; Wood, G.; Desiderio, D. M., Analysis of the human lumbar cerebrospinal fluid proteome. *Electrophoresis* **2002,** *23* (7-8), 1185-96.
53. Maccarrone, G.; Milfay, D.; Birg, I.; Rosenhagen, M.; Holsboer, F.; Grimm, R.; Bailey, J.; Zolotarjova, N.; Turck, C. W., Mining the human cerebrospinal fluid proteome by immunodepletion and shotgun mass spectrometry. *Electrophoresis* **2004,** *25* (14), 2402-12.
54. Pan, S.; Zhu, D.; Quinn, J. F.; Peskind, E. R.; Montine, T. J.; Lin, B.; Goodlett, D. R.; Taylor, G.; Eng, J.; Zhang, J., A combined dataset of human cerebrospinal fluid proteins identified by multi-dimensional chromatography and tandem mass spectrometry. *Proteomics* **2007,** *7* (3), 469-73.
55. Schutzer, S. E.; Liu, T.; Natelson, B. H.; Angel, T. E.; Schepmoes, A. A.; Purvine, S. O.; Hixson, K. K.; Lipton, M. S.; Camp, D. G.; Coyle, P. K.; Smith, R. D.; Bergquist, J., Establishing the proteome of normal human cerebrospinal fluid. *PLoS One* **2010,** *5* (6), e10980.
56. Wenner, B. R.; Lovell, M. A.; Lynn, B. C., Proteomic analysis of human ventricular cerebrospinal fluid from neurologically normal, elderly subjects using two-dimensional LC-MS/MS. *J Proteome Res* **2004,** *3* (1), 97-103.
57. Xu, J.; Chen, J.; Peskind, E. R.; Jin, J.; Eng, J.; Pan, C.; Montine, T. J.; Goodlett, D. R.; Zhang, J., Characterization of proteome of human cerebrospinal fluid. *Int Rev Neurobiol* **2006,** *73*, 29-98.
58. Varki, A.; Lowe, J. B., Biological Roles of Glycans. In *Essentials of Glycobiology*, 2nd ed.; Varki, A.; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Bertozzi, C. R.; Hart, G. W.; Etzler, M. E., Eds. Cold Spring Harbor (NY), 2009.
59. Di Domenico, F.; Owen, J. B.; Sultana, R.; Sowell, R. A.; Perluigi, M.; Cini, C.; Cai, J.; Pierce, W. M.; Butterfield, D. A., The wheat germ agglutinin-fractionated proteome of subjects with Alzheimer's disease and mild cognitive impairment hippocampus and inferior parietal lobule: Implications for disease pathogenesis and progression. *J Neurosci Res* **2010,** *88* (16), 3566-77.
60. Maguire, T. M.; Gillian, A. M.; O'Mahony, D.; Coughlan, C. M.; Dennihan, A.; Breen, K. C., A decrease in serum sialyltransferase levels in Alzheimer's disease. *Neurobiol Aging* **1994,** *15* (1), 99-102.
61. Fodero, L. R.; Saez-Valero, J.; Barquero, M. S.; Marcos, A.; McLean, C. A.; Small, D. H., Wheat germ agglutinin-binding glycoproteins are decreased in Alzheimer's disease cerebrospinal fluid. *J Neurochem* **2001,** *79* (5), 1022-6.
62. Butterfield, D. A.; Sultana, R., Redox proteomics identification of oxidatively modified brain proteins in Alzheimer's disease and mild cognitive impairment: insights into the progression of this dementing disorder. *J Alzheimers Dis* **2007,** *12* (1), 61-72.
63. Meguro, K., Clinical features of mild cognitive impairment and dementia in a community: an update of the Osaki-Tajiri Project. *Tohoku J Exp Med* **2008,** *215* (2), 125-31.
64. Visser, P. J.; Verhey, F. R., Mild cognitive impairment as predictor for Alzheimer's disease in clinical practice: effect of age and diagnostic criteria. *Psychol Med* **2008,** *38* (1), 113-22.
65. Petersen, R. C.; Doody, R.; Kurz, A.; Mohs, R. C.; Morris, J. C.; Rabins, P. V.; Ritchie, K.; Rossor, M.; Thal, L.; Winblad, B., Current concepts in mild cognitive impairment. *Arch Neurol* **2001,** *58* (12), 1985-92.
66. Pietrocola, F.; Galluzzi, L.; Bravo-San Pedro, J. M.; Madeo, F.; Kroemer, G., Acetyl Coenzyme A: A Central Metabolite and Second Messenger. *Cell metabolism* **2015,** *21* (6), 805-821.
67. (a) Pehar, M.; Puglielli, L., Lysine acetylation in the lumen of the ER: a novel and essential function

under the control of the UPR. *Biochimica et biophysica acta* **2013,** *1833* (3), 686-97; (b) Shi, L.; Tu, B. P., Acetyl-CoA and the regulation of metabolism: mechanisms and consequences. *Curr Opin Cell Biol* **2015,** *33*, 125-31.
68. Jonas, M. C.; Pehar, M.; Puglielli, L., AT-1 is the ER membrane acetyl-CoA transporter and is essential for cell viability. *J Cell Sci* **2010,** *123* (Pt 19), 3378-88.
69. (a) Choudhary, C.; Kumar, C.; Gnad, F.; Nielsen, M. L.; Rehman, M.; Walther, T. C.; Olsen, J. V.; Mann, M., Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* **2009,** *325* (5942), 834-40; (b) Pehar, M.; Lehnus, M.; Karst, A.; Puglielli, L., Proteomic assessment shows that many endoplasmic reticulum (ER)-resident proteins are targeted by N(epsilon)-lysine acetylation in the lumen of the organelle and predicts broad biological impact. *J Biol Chem* **2012,** *287* (27), 22436-40.
70. Ding, Y.; Dellisanti, C. D.; Ko, M. H.; Czajkowski, C.; Puglielli, L., The endoplasmic reticulum-based acetyltransferases, ATase1 and ATase2, associate with the oligosaccharyl-transferase to acetylate correctly folded polypeptides. *J Biol Chem* **2014,** *289*, 32044-32055.
71. (a) Costantini, C.; Ko, M. H.; Jonas, M. C.; Puglielli, L., A reversible form of lysine acetylation in the ER and Golgi lumen controls the molecular stabilization of BACE1. *Biochem J* **2007,** *407* (3), 383-95; (b) Mak, A. B.; Pehar, M.; Nixon, A. M.; Williams, R. A.; Uetrecht, A. C.; Puglielli, L.; Moffat, J., Post-Translational Regulation of CD133 by ATase1/ATase2-Mediated Lysine Acetylation. *Journal of molecular biology* **2014,** *426* (11), 2175-82.
72. Pehar, M.; Jonas, M. C.; Hare, T. M.; Puglielli, L., SLC33A1/AT-1 protein regulates the induction of autophagy downstream of IRE1/XBP1 pathway. *J Biol Chem* **2012,** *287* (35), 29921-30.
73. Peng, Y.; Li, M.; Clarkson, B. D.; Pehar, M.; Lao, P. J.; Hillmer, A. T.; Barnhart, T. E.; Christian, B. T.; Mitchell, H. A.; Bendlin, B. B.; Sandor, M.; Puglielli, L., Deficient Import of Acetyl-CoA into the ER Lumen Causes Neurodegeneration and Propensity to Infections, Inflammation, and Cancer. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **2014,** *34* (20), 6772-89.
74. Lin, P.; Li, J.; Liu, Q.; Mao, F.; Qiu, R.; Hu, H.; Song, Y.; Yang, Y.; Gao, G.; Yan, C.; Yang, W.; Shao, C.; Gong, Y., A missense mutation in SLC33A1, which encodes the acetyl-CoA transporter, causes autosomal-dominant spastic paraplegia (SPG42). *Am J Hum Genet* **2008,** *83* (6), 752-9.
75. Huppke, P.; Brendel, C.; Kalscheuer, V.; Korenke, G. C.; Marquardt, I.; Freisinger, P.; Christodoulou, J.; Hillebrand, M.; Pitelet, G.; Wilson, C.; Gruber-Sedlmayr, U.; Ullmann, R.; Haas, S.; Elpeleg, O.; Nurnberg, G.; Nurnberg, P.; Dad, S.; Moller, L. B.; Kaler, S. G.; Gartner, J., Mutations in SLC33A1 cause a lethal autosomal-recessive disorder with congenital cataracts, hearing loss, and low serum copper and ceruloplasmin. *Am J Hum Genet* **2012,** *90* (1), 61-8.
76. (a) Sanders, S. J.; Ercan-Sencicek, A. G.; Hus, V.; Luo, R.; Murtha, M. T.; Moreno-De-Luca, D.; Chu, S. H.; Moreau, M. P.; Gupta, A. R.; Thomson, S. A.; Mason, C. E.; Bilguvar, K.; Celestino-Soper, P. B.; Choi, M.; Crawford, E. L.; Davis, L.; Wright, N. R.; Dhodapkar, R. M.; DiCola, M.; DiLullo, N. M.; Fernandez, T. V.; Fielding-Singh, V.; Fishman, D. O.; Frahm, S.; Garagaloyan, R.; Goh, G. S.; Kammela, S.; Klei, L.; Lowe, J. K.; Lund, S. C.; McGrew, A. D.; Meyer, K. A.; Moffat, W. J.; Murdoch, J. D.; O'Roak, B. J.; Ober, G. T.; Pottenger, R. S.; Raubeson, M. J.; Song, Y.; Wang, Q.; Yaspan, B. L.; Yu, T. W.; Yurkiewicz, I. R.; Beaudet, A. L.; Cantor, R. M.; Curland, M.; Grice, D. E.; Gunel, M.; Lifton, R. P.; Mane, S. M.; Martin, D. M.; Shaw, C. A.; Sheldon, M.; Tischfield, J. A.; Walsh, C. A.; Morrow, E. M.; Ledbetter, D. H.; Fombonne, E.; Lord, C.; Martin, C. L.; Brooks, A. I.; Sutcliffe, J. S.; Cook, E. H., Jr.; Geschwind, D.; Roeder, K.; Devlin, B.; State, M. W., Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron* **2011,** *70* (5), 863-85; (b) Prasad, A.; Merico, D.; Thiruvahindrapuram, B.; Wei, J.; Lionel, A. C.; Sato, D.; Rickaby, J.; Lu, C.; Szatmari, P.; Roberts, W.; Fernandez, B. A.; Marshall, C. R.; Hatchwell, E.; Eis, P. S.; Scherer, S. W., A discovery resource of rare copy number variations in individuals with autism spectrum disorder. *G3* **2012,** *2* (12), 1665-85; (c) Krumm, N.; O'Roak, B. J.; Karakoc, E.; Mohajeri, K.; Nelson, B.; Vives, L.; Jacquemont, S.; Munson, J.; Bernier, R.; Eichler, E. E., Transmission disequilibrium of small CNVs in simplex autism. *Am J Hum Genet* **2013,** *93* (4), 595-606.

Chapter 2

Mass spectrometric investigation of neuropeptidomic alteration in food intake

Adapted from manuscript: "Label-free Quantitation of Feeding-Induced Neuropeptides Levels Revealed Potential Therapeutics for Food Intake Regulation" **Jingxin Wang**, Hui Ye, Zichuan Tian, James Dowell, Fengfei Ma, Gaoyuan Lu, Brian Baldo, Lingjun Li.

Abstract

Endogenous neuropeptides are important signaling molecules that function as regulators of food intake and body weight. Previously, changes in neuropeptide gene expression within the nucleus accumbens (Acb) were shown to be affected by food intake in rats. In order to directly monitor the feeding-induced changes in neuropeptide expression levels within the Acb, we employed a combination of cryostat dissection, heat stabilization, neuropeptide extraction and label-free quantitative neuropeptidomics using a liquid chromatography-high resolution mass spectrometer. Here we showed that using this methodology over 300 neuropeptides were identified from rat Acb, and feeding caused the expression level changes of multiple neuropeptides, especially those opioid peptides such as enkaphalins and non-opioid peptides including ProSAAS, cholecystokinin and somatostatins. We further investigated the regulatory function of non-opioid neuropeptides from the ProSAAS family by performing a microinjection using the identified ProSAAS neuropeptides to the rat Acb. We showed that at high concentration, big LEN significantly increased rats' food and water intake. In addition, both big LEN and PEN at high concentration altered other behaviors including locomotion and rearing. In addition, we also quantified the feeding-induced changes of neuropeptides from hippocampus, hypothalamus and striatum to reveal the neuropeptide interplay among different anatomical regions. In summary, our study revealed the neuropeptidomics changes in response to food intake in the rat Acb and other key brain regions. More importantly, we demonstrated the orexigenic functions of proSAAS neuropeptides before food intake.

Keywords

label-free quantitation; liquid chromatography- mass spectrometry; neuropeptides; nuclear accumbens; ProSAAS

1. Introduction

Neuropeptides are a class of the most diverse and largest cell-cell signaling molecules that modulate the activities of neural circuits. They function as regulators of a wide range of physiological process like stress, pain, addiction, memory, circadian rhythm, reproduction and food intake. Quantitative description of the expression levels of these peptide neuromodulators aids to understand the underlying mechanisms of such actions, implies its potential as an effective target in developing therapeutic strategies for relative diseases such as eating order, and may even represent a pool of future therapeutics.

Among the various physiological processes neuropeptides regulate, motivation and reward have been studied most extensively in regard to drugs of addiction. Interestingly, a number of studies have suggested an overlap of neurobiological mechanisms between drug addiction and the hedonisitic aspects of feeding.[1-7] Electrical stimulation of the lateral hypothalamus (LH), an area involved in feeding, both increases food intake and produces reward-seeking behavior.[8] Alternatively, food deprivation increases the self-administration of non-food rewards, such as psychostimulants, intracranial stimulation, and heroin, suggesting a significant link between food reward and drug rewards.[9, 10]

Since the discovery of the endogenous opioids and their receptors in the 1970s, there has been a large body of research dedicated to understanding their functional roles in the brain.[11, 12] Subsequent to their discovery, the endogenous opioids were found to be involved in a wide variety of physiological processes, including the modulation of pain, stress, endocrine responses, motivational states, feeding, and addiction.[13] Early studies examining the effects of opiates on rats found morphine injections dramatically increased food intake.[14] These initial experiments led to a number of follow-up studies using both opioid agonists and antagonists, with the

agonists increasing food consumption and antagonists decreasing food intake, solidifying the specific involvement of opioids in the regulation of feeding.[15] More recent studies using specific μ-opioid agonists, such as D-Ala2,Nme-Phe4,Gyol5-enkephalin (DAMGO), have shown the opioid-induced increase in food intake to be modulated by the μ-opioid receptors in the nucleus accumbens (NAc).[7] These experiments were performed with satiated rats which had just eaten and upon infusion of DAMGO into the NAc the rats resumed eating, consuming up to 400% of their normal intake of food.[16] Additionally, the DAMGO-treated rats displayed a preference for high-fat food.[16, 17]

Experiments employing exogenous application of opioids into the nucleus accumbens have established a strong link between μ-opioid receptor activation in the nucleus accumbens (NAc) and hedonistic feeding behavior; however, none of these experiments have examined the reverse effects—the effect of food intake on the release of endogenous opioids in the NAc. In light of this, Will and co-workers examined feeding induced changes in the endogenous opioid system.[18] Interestingly, the expression of preproenkephalin mRNA was significantly down-regulated in the NAc of fed rats when compared to food-denied rats, indicating a feeding-induced modulation of endogenous opioid expression. These data equate enkephalin expression changes with specific motivational states, i.e. hungry versus satiated; however, mRNA expression studies can only reveal transcript-level changes which do not always reflect changes in actual opioid peptide release. Due to the skyrocketing rates of feeding and energy metabolism-related disorders, it is becoming more and more needed for researchers and pharmaceutical companies to elucidate opioid peptides' underlying control mechanisms over humans' appetites.

Besides extensively studies opioid peptides, non-opioid neuropeptides are also involved in the regulation of food intake. For example, neuropeptide Y as the most potent endogenous

orexigenic neurotransmitter has demonstrated its influence on food intake and during the development of eating disorders across classes from fish to mammals such as humans. Similarly, galanin, melanin-concentrating hormone (MCH) and ghrelin are reported to function as orexigenic peptides, while cocaine- and amphetamine-regulated transcript (CART), α-melanocyte stimulating hormone (α-MSH), and neuropeptide W (NPW) are believed to exert anorexigenic effects in animals. Therefore, it is intriguing to speculate that opioid and non-opioid peptide contribute combinatorially to the regulation of animals' food intake behavior.

Although a myriad of studies have shown that food intake behavior is tightly controlled by many opioid and non-opioid peptides, these studies have a limitation that individual investigations have each focused on clarifying the functions of a single peptide or a specific peptide family. Nevertheless, the regulation of food intake is an intricate process that takes place via a complex neural network involving co-regulation of multiple neuropeptides probably via synaptic inputs, whereas currently little information is available regarding how these neuropeptides change comprehensively in response to feeding, even for the well-studies opioid peptides. Moreover, it has been suggested that reward circuitry in NAc and several brain areas is responsible for the regulation of food intake. Although reward-related or feeding-induced signaling of neuropeptides have previously been investigated in these regions, respectively, it has never been reported how neuropeptides respond to natural rewards, food intake, and cross-talk among different neuroanatomical regions.

In order to directly interrogate feeding induced changes in opioid and non-opioid peptides, we have employed a mass spectrometry-based neuropeptidomics approach to compare levels of neuropeptides in NAc, striatum, hippocampus and hypothalamus between fed and unfed rats. Historically, neuropeptides have been predominantly studied employing radioimmunoassay,

immunohistochemistry, and Edman degradation.[19] These techniques continue to be important but recent advances in mass spectrometry (MS) have provided a new and powerful analytical platform for studying neuropeptides. Mass spectrometry-based neuropeptidomics is capable of identifying and quantifying large numbers of neuropeptides simultaneously, yielding higher throughput and specificity, whereas immuno-based approaches suffers from large sample consumption, low throughput and could not accurately identify the neuropeptides with additional post-translational modifications (PTMs) such as acetylation, C-term amidation and pyro-glutamic acid.

However, one drawback of MS-based neuropeptidomic approaches is the requirement of highly purified and enriched peptide samples. Specifically in mammals, non-specific protein degradation produces a large amount of peptide contamination which can mask the signal of the less abundant endogenous neuropeptides. In order to curtail post-mortem protein degradation, a number of methods have been employed, including the use of transgenic mice lacking carboxypeptidase E activity, processing samples in a boiling buffer, focused microwave irradiation for animal sacrifice and rapid post-sacrificial microwave irradiation.[20-22] Microwave irradiation heat deactivates endogenous proteases and arrests post-mortem protein degradation, resulting in a clean neuropeptide sample for MS analysis. In contrast, our lab has developed a non-microwave protease deactivation method employing snap freezing, crytostat dissection, and snap heat stabilization employing an apparatus Denator.[23] With rapid heating by Denator, we have minimized unspecific protein degradation that might modify actual neuropeptides to other forms and interfere with assignment. We anticipate the neuropeptides we reported in this study represent mostly bioactive mature neuropeptides and intermediate cleavage products that might function as well with minimal artifact.

In this study, we aimed to employ a neuropeptidomic approach to comprehensively characterize opioid and non-opioid peptides in key brain regions involved in reward (NAc, striatum, hypothalamus and hippocampus) in fed and unfed adult rats. We employed a combination of cryostat dissection, neuropeptide extraction and label-free quantitative neuropeptidomics using a liquid chromatography-high resolution mass spectrometer. Such approach enables simultaneous monitoring of a number of neuropeptides from specific brain regions of interest in a high throughput with robustness. Our results successfully identified over a thousand peptides from the sampled regions, and showed a distribution map of nearly 40 neuropeptide families among NAc, striatum, hypothalamus and hippocampus. More importantly, we determined proSAAS peptides displayed significant change following food intake and we picked several upregulated non-opioid peptides from the ProSAAS family and performed micro-injection using these peptides into rats' NAc region. We found big LEN at high concentration induced food intake in rats. This phenomenon strongly implied the regulatory function of food intake and potentially other reward-related behaviors by non-opioid peptides that exhibited significant changes as opioid peptides upon feeding. It also expands present scope in developing effective and safe therapeutic strategies for feeding-related diseases such as eating order and unveils a pool of endogenous neuropeptides as potential therapeutic candidates.

2. Experimental

2.1 Chemicals and reagents

Optima grade formic acid, acetonitrile (ACN), water, and methanol (MeOH) were purchased from Fisher Scientific (Pittsburgh, PA). All other chemicals were obtained from Sigma-Aldrich (St. Louis, MO).

2.2 Animals

Adult male Sprague-Dawley rats (total n = 24, Harlan Sprague Dawley, Indianapolis, IN) weighing 300–400 g were maintained in a temperature- and humidity-controlled animal colony on a 12:12-h light-dark cycle (lights off at 18:00). All subjects were naive and were allowed a minimum of a week adaptation followed by 2 days of daily handling before the beginning of the experiment. Subjects had free access to normal laboratory chow (24% protein, 4% fat) and drinking water was available ad libitum. On the day of the experiment, at 17:00, food (chow) was removed from half of the subjects, while a measured amount of food was given to the other subjects (18 g/cage). The rats were then killed at 19:30, 2.5 h after food was given. The time point was chosen based on the previous mRNA experiment performed in the lab.[18] All experimental procedures were in accordance with protocols approved by the University of Wisconsin Institutional Animal Care and Use Committee.

2.3 Sample preparation

Animal Sacrifice and Cryostat Dissection. Rats were anesthetized with halothane and then sacrificed by decapitation. The brain was then rapidly removed (<90 s) and snap frozen in 2-methylbutane cooled by dry ice. The frozen brain was then sectioned in 300 µm thick slices on a cryostat from Leica (Wetzlar, Germany) with a compartment temperature of -15 °C. The regions corresponding to dorsal striatum, hippocampus, hypothalamus and nucleus accumbens were removed with a 2 mm micro-punch and stored in 1.5 mL tubes at -80 °C until extraction (see Figure 1).

Heat Stabilization and Tissue Extraction. Tissue punches (20-30 mg per rat) were removed from -80 °C and immediately processed by Denator Stabilizor T1 tissue stabilization device (Gothenburg, Sweden) for heat stabilization. Four punches were placed in a Stabilizor

cartridge as a pooled aliquot, inserted into the device, and stabilized using the fresh preserve tissue function. The stabilization process involved uniformly heating the tissue to 95 °C for 30–45 sec (depending on the tissue thickness). After stabilization, tissue was removed from the cartridge, placed in the appropriate extraction solvent, and stored at −80 °C until needed.

The processed tissues were then extracted into ice-cold acidified methanol (90:10:1 / MeOH: water: acetic acid). The samples were then homogenized manually with a glass-glass Dounce homogenizer from Wheaton (Millville, NJ). The homogenized sample was then spun at 20,000 g for 20 minutes at 4 °C to remove the insoluble pellet. Protein concentration of the pellet was determined for each sample using bicinchonic acid (BCA) assay from Pierce (Rockford, IL) and used to adjust the differences in the neuropeptide levels contained in different sample aliquots. The adjusted supernatant was decanted and then dried in a vacuum centrifuge. Extracts were resuspended in 20 μL 0.1% formic acid aqueous solution by vortexing and brief sonication. Subsequently, the reconstituted samples were purified and concentrated by C18 ZipTip (Millipore, Billerica, MA). Briefly, the C18 ZipTip was first wetted using ACN and then pre-equilibrated for sample binding with 0.1% formic acid in water. Subsequently, the tissue extract was loaded on the C18 ZipTip. After being rinsed three times with 0.1% formic acid in water, the sample was eluted with 5 μL of ACN/water/formic acid solution (50:49.9:0.1; vol/vol/vol). Next, the eluent was dried and resuspended in 10 μL of 0.1% formic acid in water and subjected to future LC−MS/MS analysis.

2.4 Liquid chromatography Mass spectrometry

Peptide extracts from each brain sample aliquot were analyzed separately. Peptide samples were dissolved in 0.1% FA before analysis. A Waters nanoAcquity UPLC (Waters Corp., Milford, MA) was coupled to a Thermo Q-Exactive Orbitrap mass spectrometer (San

Jose, CA) for LC-MS/MS analysis. Chromatographic separations were performed on a Waters BEH 130 Å C18 reversed-phase capillary column (150 mm × 75 μm, 1.7 μm). The mobile phases used were: Mobile phase A consisted of water with 0.1% FA, and mobile phase B was composed of ACN with 0.1% FA. Samples were injected and loaded onto the Waters NanoACQ 2G-V/M Sym C18 (20 mm ×180 μm, 5 μm) using 100% A at a flow rate of 5 μL/min for 1 min. Then the peptides were separated using a solvent gradient of 0–10% B over 0.5 min and then 10–30% B over 70 min at a flow rate of 300 nL/min. Data-dependent acquisition (DDA) parameters recorded MS scans in profile mode from m/z 380–1500 at a resolution of 35K. Automatic gain control (AGC) targets of 1 x 106 and maximum injection times (IT) of 100 ms were set. The 15 most intense precursor ions were selected for MS2 higher-energy collisional dissociation (HCD) fragmentation with an isolation window of 2 Da and dynamic exclusion set at 40 s. An AGC target of 1 x 105 and a maximum IT of 150 ms was selected for tandem mass acquisition. The tandem MS spectra were acquired at a resolution of 17.5K in profile mode, with normalized collision energy (NCE) set at 27, and a fixed lower mass at m/z 110.

2.6 Label-free quantitation of neuropeptides

Neuropeptide Identification. The enzyme was specified as none. The peptide mass tolerance was set at xxx ppm and the MS/MS mass tolerance was set at xxx Da. Proteome Discover searches were performed for each sample. The variable modifications were set to include C-terminal amidation, N-terminal acetylation, and methionine oxidation

Neuropeptide Quantitation. Relative expression changes were calculated for each identified peptide pair by comparing extracted ion chromatograms (XIC). The m/z for each identified neuropeptide and its isotope partner was used to create XICs. Each XIC was then integrated to obtain the peak area for each peptide from individual samples by Proteome Discover

automatically. The area values were entered into an Excel spreadsheet and the ratios between XIC areas were calculated. To control for the bias associated with unequal amounts of tissue and differences in extraction efficiencies a normalization scheme was employed. The normalization factor was determined from the average isotope ratio from all identified peptides and then re-applied to each individual ratio to obtain the normalized ratios. or use the protein concentration

2.7 Microinjection to Acb using Significantly-changed Neuropeptides

Subjects

10 Sprague-Dawley rats (Harlan; Madison, WI) weighted 300-400 g when arriving at the laboratory were housed in clear cages in a temperature and light-controlled vivarium (lights on from 0700 to 1900, 12:12 h light-dark cycle). Food and water were available *ad libitum* and all experiments were conducted between 1000-1600 h. The rats were handled daily to reduce stress except the period during recovery from surgery. The facilities and procedures has followed the guidelines of animal use and care from the National Institutes of Health, which were approved by the Institutional Animal Care and Use Committee of the University of Wisconsin.

Surgical Procedure

Rats were anesthetized with isoflurane while undergone the stereotaxic surgery. Stainless-steel cannulae (10-mm long; 25 gauge) were implanted bilaterally into AcbSh (coordinates from bregma with toothbar at +4 mm, + 3.2 anteroposterior (AP); $\pm$ 1.0 lateromedial (LM); –5.2 dorsoventral (DV). The guide cannulae were cemented with dental acrylic (New Truliner, Skokie, IL) and skull screwers (Plastics One, Roanoke, VA). Stylet wires (10.1 mm long, 0.008 in. diameter) were placed in the guide cannulae to prevent blockage. Intramuscular penicillin (0.3 ml of a 300,000 U/ml suspension; Phoenix Pharmaceuticals, St. Joseph, MO) and Meloxicam

were given to the animals after surgery. Animals were placed in a warm recovery cage and returned to their original cages upon awakening. Postoperative recovery and care were 7-10 days with daily health checks before behavioral testing commenced.

Experimental Design

Prior to all the experiments, rats were placed into the testing cages to be habituated for one hour on three sequential days to minimize the stress. 2 hour free-feeding test sessions were conducted in the clear polycarbonate cages (9.5 in. width X 17 in. length X 8 in. height) which were identical to their home cages except for the wire grid floor. After recovery from surgery, animals were accommodated to the microinfusion procedures with preliminary sham and saline injections after exhibiting stable baseline responding (no more than $\pm$ 10% variability over three consecutive testing days). The drug testing commenced after the rats were re-baselined after these injections.

Rats had free access to laboratory chow and water before drug injection. On the testing day, rats (N=10) received intra-AcbSh infusions (3, 10μg/0.5μl PEN, 3, 10μg/0.5μl big LEN) and were placed into a behavior-observation procedure (discussed below). For each drug, doses were administered in accordance to within-subjects Latin square designs, with 2-3 interim days of drug-free baseline testing to separate the drug-infusion days. After finishing the infusion of big LEN and PEN at two dosages in a counterbalanced order, all the rats (N=10) were given another saline injection to re-baseline them in order to exclude the changes of behavioral effect due to other factors. After the re-baseline, all the rats (N=10) were administered an injection of little-SAAS (10μg/0.5μl) to determine its effect since the quantitative results of little-SAAS showed a

smaller fold-changes compared to PEN or big LEN. Finally, DAMGO (0.025μg/0.5μl) was given to all the rats (N=10) to compare the behavioral changes of ProSAAS-derived peptides to opioid receptor agonist.

Behavioral Observation Procedure

After one-hour habituation sessions were carried out for three sequential days (described above), the rats were injected with their specific treatment respectively. After infusion, rats were placed into the wire-bottom polycarbonate cages (described above) with pre-measured laboratory chow pellets displaced on the floor of the testing cages, while pre-measured water was available in overhead bottles. In the meanwhile, they were videotaped with a digital camcorder for 2-hour session. The experimenter was blind to the treatment that were coded with different letters that were viewed by the digital files. The spontaneous motor and ingestive behavior were recorded immediately after replacement of the cage lid. All spontaneous behaviors of 2-hour session were recorded continuously via an event recorder which was interfaced to a PC-based computer (Paul Fray, Ltd, Cambridge, U.K.). The following parameters were scored, including: locomotion (crossing over the cage center), feeding, drinking, rearing, grooming. The total duration (in seconds) and frequency of each behavior were scored except locomotion (only frequency). Total food ingested and food spilled (grams) during a session were measured after the test. In addition to measuring the total effects of treatment, the first hour test period was divided into 12 x 5 min time-bins and the effects of treatment was determined as a function of time during this session.

Microinfusion Procedures and Drugs

Intracerebral microinfusions were performed in accordance to the standard procedures, with details described in Supplementary Materials. The injectors (12.5 mm long) were able to extend 2.5 mm beyond the guide cannulae. Big-Len, PEN and little-SAAS were purchased from Phoenix Pharmaceuticals (Burlingame, CA). DAMGO (μ-opioid agonist) was obtained from Sigma-Aldrich (St. Louis, MO). The dosage-range of big-LEN, PEN and little-SAAS used here were similar to that applied in previous studies. In addition, intra-AcbSh infusion of DAMGO induced a clearly active behavior as expected. All drugs were dissolved and diluted with fresh sterile 0.9% saline immediately prior to the infusions.

Statistical Analyses

Within-subjects factorial ANOVAs were applied for data analyses as required by the experimental designs. Data with significant changes in ANOVAs were subjected to post-hoc comparisons among means via Tukey-Kramer method. To demonstrate the changes of behavioral patterns over time, the first hour continuous recording of bouts and duration of each behavior was divided into 12 x 5 min time-bins. Data was subjected to repeated measurement of two-way ANOVA (treatment x time). Interactions showed significant changes with ANOVA were further analyzed with repeated paired t-tests. $P < 0.05$ was considered as statistically significant for all experiments.

3. Results

In this study, we aimed to extensively characterize the neuropeptide changes due to food intake from four different rat brain regions by comparing 12 fed and unfed rats (Figure 1). Rats fed on a regular diet or starved were all sacrificed 2.5 hours after the feeding time. Rat brains

were immediately dissected following decapitation and the dorsal striatum, hippocampus, hypothalamus and nucleus accumbens regions were collected as tissue punches, followed by rapid heating by DenatorTM to minimize postmortem degradation. The four brain regions were all related to food intake or reward-related neuropeptide signaling, and were therefore selected. The Tissue samples from each three rats were pooled as an aliquot to minimize individual variation and increase the amount of neuropeptides contained in samples. The sample aliquots were then processed individually, and the endogenous neuropeptides were analyzed by an Oribtrap LC-MS/MS. The high accuracy MS and MS/MS data yielded by Orbitrap were search against a home-built database, delivering comprehensive identification and quantification results of the neuropeptides from the rat brain regions of interest. Such results allow us to explore the functions of the food intake-induced peptide changes. Nevertheless, due to previous studies mostly focusing on opioid peptides and food intake, we chose to investigate the putative regulatory roles of several non-opioid ProSAAS peptides that changed significantly upon feeding by performing microinjection experiments on rat NAc.

3.1 Identification of Neuropeptides from Rat Dorsal Striatum, Hippocampus, Hypothalamus and Acb

To understand the underlying mechanism of neuropeptides' impact on food intake, an initial step is to comprehensively identify the neuropeptides present in the targeted rat brain regions. The high resolution, high accuracy MS and MS/MS information acquired from the peptides following nanoLC separation enabled by the Orbitrap detector guarantees high confidence peptide identification. HCD MS/MS fragmentation in conjunction with charge deconvolution and de-isotoping of the MS/MS spectra permit the interpretation of the relatively highly-charged neuropeptides. Figure 2A shows the MS/MS spectra of a 5+ charged peptide ions detected at m/z

679.569. Using a stringent searching criterion, mass tolerance for precursor ions at 10 ppm and product ions at 0.05 Da, the precursor ion in Figure 2A was identified as YIQQVRKAPSGRMSVLKNLQGLDPSHRISD from the neuropeptide precursor cholecystokinin with high confidence. Besides a few highly-charged peptide ions, most of the neuropeptide identified are 2+, 3+ charged. Figure 2B illustrates the identification of a moderately-charged 3+ ion at m/z 1055.198 characterized as GWTLNSAGYLLGPHAIDNHRSFSDKHGL derived from the precursor of galanin. All the LC-MS/MS data was searched against a home-build database that originated from the reviewed Swiss-prot database of rattus and then filtered to retain precursors that contain signal peptide sequences. With this database, we totally reported 1405 endogenous peptide derived from XX precursors from the four brain regions. This is the largest rat brain peptidome reported so far to our knowledge. A complete summary of all the identified peptides is provided in Supplemental Information, Table 1.

With the peptidome of DS, Hippo, HT and Acb being characterized, we then analyzed the overlap of peptides among the four regions in Figure 3A. HT was attributed as the region containing the highest number of neuropeptides. It also contained the largest number of unique peptides (333 peptides, 21% of peptidome) that are not overlapped with other regions. In contrast, DS, Hippo and NAc only had 49 (3.4%), 47 (3.3%) and 35 (2.4%) non-overlapped peptides, respectively. This observation accorded with previous report that compared the neuropeptidome of HT with DS (AC ref). Notably, 439 peptides (30.4% of whole peptidome) were shared among the all four brain areas. Out of the identified 1405 peptides, 280 peptides carry one or multiple PTMs. Approximately 60% of the modified peptides were N-terminal acetylated, whereas about 30% peptides with PTM possess a C-terminal amidation.

Interestingly, both of the two modification states could turn inactive versions of incompletely-processed peptides to active and mature forms of neuropeptides upon modification. For example, the peptides RPKPQQFFGLM and HKTDSFVGLM cannot be translated to active neuropeptides substance P and neurokinin A that modulates immune response, nociception, addiction and so on without C-terminal amidation. In addition to acetylation and amidation, a few methionine oxidation-modified peptides exist. Nevertheless, such modification is often assumed as an artifact introduced to peptides during sample preparation and storage.

Figure 3B-E shows the distribution of variant neuropeptide families identified from each brain regions. As expected, HT consists of the most variant neuropeptide families that 17 families possessing more than 15 unique peptides were identified. These families includes tachykinin, ProSAAS, enkephalin, prothyroliberin, somatostatin, CCK, MCH, CART, NPY, PACAP and neurosecretory protein such as VGF and secretogranin. Of all identified endogenous peptides, tachykinin, enkephalin, CCK, MCH, PACAP are all known to regulate food intake. Besides the well-known mature neuropeptides, many peptides we identified are truncated forms or longer forms of these neuropeptides. These immature forms are normally intermediate neuropeptide products produced during neuropeptide maturation process by processing enzymes such as carboxypeptidases or extracellular peptidases. Although these peptides could also result from post-mortem degradation, the little protein post-mortem markers such as stathmin we observed, thanks to the employment of DenatorTM, reflected a minor impact of such degradation.

DS, Hippo and NAc possessed 10, 12, 12 families of precursors that yielded more than 15 unique peptides, showing less diversity of neuropeptide composition compared to HT. This accorded with the results we showed in Figure 3A, probably due to HT's largest sample area and great heterogeneity than other regions. It is worth noticing that the most peptides from the enkephalin

family are attributed to NAc. It agrees with the fact that NAc controls biosynthesis and reslease of enkephalin, yet is also modulated by the opioid peptides, enkephalins. Besides enkephalin, ProSAAS is another primary component of all the identified neuropeptide families in NAc. Other than its potential role in inhibiting the activities of prohormone convertase, ProSAAS's more prominent physiological roles are demonstrated by evidence such as that ProSAAS deletion leads to the decrease of body mass and over-expression of ProSAAS result in obesity. Our further investigation on whether ProSAAS-derived peptides were changed due to food intake would aid to unravel the postulated functions of ProSAAS peptides in regulating food intake and thereby the body weight and energy homoestasis.

3.2 Label-free Quantitation of Neuropeptides regulated by Feeding

To perform a robust yet accurate label-free quantitative experiment, we randomized the samples and analyzed the pooled tissue samples in four replicates for each region. Previous label-free experiment normally employed a single internal standard or TIC for normalization, we spiked into each sample a mixture of enolase tryptic digest at appropriate concentration. The enolase digest mixture was thus used as internal standards for intensity normalization and retention time alignment.

3.3 Microinjections to rat Nucleus Accumbens by non-opioid ProSAAS peptides

Figure 4 illustrated the histological verification of intra-AcbSh injection placement. Two rats were removed from the experiment due to guide cannulae obstruction or brain abscess. Panel **b** showed reprehensive photomicrographs of injector placements into the AcbSh, where the injector tracks were visible obviously without any unusual damage of the area of interest. The injector tips were found always to be located within the shell of Acb or the adjacent area between the shell and core. Panel **a** depicted the schematic injector tip placement for each rat.

As shown in **Figure 5**, DAMGO significantly elevated chow [F (6, 7) = 18.291, $P < 0.0001$] and water [F (6, 7) = 7.024, $P < 0.0001$] intake (see Figure 2a and 2b). *Post hoc* comparison among means with Tukey-Kramer test confirmed that DAMGO-associated levels of food and water intake were significantly elevated in comparison with saline infusion. Whereas the effect among pre-, mid-, post-saline infusion did not present significant changes in food or water intake ($P > 0.05$ respectively). Although both dosages (3, 10μg/0.5μl) of PEN and big-LEN and high dosage of little-SAAS have increased the mean of food intake, only high dosage (10μg/0.5μl) of big-LEN showed an essentially statistical significance [F (5, 7) = 3.534, $P = 0.0109$]. Similarly, high dosage (10μg/0.5μl) of big-LEN also elevated the water intake in rats significantly compared to saline [F (5, 7) = 3.148, $P = 0.0189$].

For behavioral changes among locomotion, rearing, eating, drinking and grooming, including both duration and bouts, none of them showed a significant alteration among pre-, mid-, post-saline injection ($P > 0.05$ respectively). The mean bouts and duration of each behavior in three saline infusion was calculated and compared to those of two dosages (3, 10μg/0.5μl) of PEN and big-LEN in the first hour of the two hour observational sessions. In Figure 2c and 2d, both high dosage (10μg/0.5μl) of PEN and big-LEN produced a significant increase in the bouts of locomotion [F (4, 7) = 2.782, $P = 0.0460$] and the duration of rearing [F (4, 7) = 3.688, $P = 0.0156$]. In the following *Post hoc* confirmation test, Tukey-Kramer method failed to show a significance of locomotion among the means of each treatment, however, it displayed significant changes between saline and high dosage PEN (10μg/0.5μl), and between saline and high dosage big-LEN (10μg/0.5μl) rearing duration. None significant changes were shown with low dosage

(3μg/0.5μl) PEN or (3μg/0.5μl) big-LEN in any behavioral analysis. Temporal analysis of each behavior generally displayed time-dependent reductions in active behaviors. Significant treatment-time interaction alterations of locomotion, rearing bouts and rearing duration were illustrated in **Figure 6**. Locomotion [F (66, 462) = 2.257, $P < 0.001$] was shown in **Figure 6a** and **6b**. Bar chart in **Figure 6c** and **6d** reflected changes of rearing bouts [F (66, 462) = 2.578, $P < 0.001$] and **Figure 6e** and **6f** displayed alteration of rearing duration [F (66, 462) = 3.018, $P < 0.001$]

4. Discussion

Enkephalin release varies with diurnal rhythms and the rewarding and/or feeding effects of opiates can also be influenced by diurnal rhythms.[24, 25] Rats forage for food directly after waking-up (they are nocturnal and wake-up at the beginning of the dark cycle) and μ-opioid receptor membrane trafficking within the NAc increases during this time.[25] Based on these previous studies, the animals were subjected to food deprivation at the beginning of the dark cycle—their natural time to eat. In this way, we sought to maximize the difference in motivational state, i.e. hungry versus satiated, and presumably the difference in vesicular enkephalin content.

At first glance, an enkephalin increase in the unfed group might seem counterintuitive given that opioid agonists increase feeding behavior. However, in our current experiment, we are assuming the release of enkephalins occurs upon feeding and once released, the enkephalins are degraded. There is a large amount of evidence from in vivo microdialysis studies supporting this assumption.[26-29] In all of these experiments, enkephalin is relatively short-lived in the extracellular space and is degraded within 60 minutes of release. In regards to our experiment,

we sacrificed our rats approximately 150 minutes after feeding and thus the enkephalins we extracted were from the secretory vesicles and not the extracellular space. If this logistical caveat is taken into account, then the relative changes in enkephalin expression fit within our paradigm. Upon eating, enkephalin is released into the extracellular space where it is degraded. In contrast, the food-deprived rats retain their pool of enkephalin within the secretory vesicles. In both groups, the vesicular enkephalin is extracted and analyzed by mass spectrometry. Since the fed rats released a substantial portion of their vesicular enkephalin, the extracts from these animals contain less enkephalin than the unfed animals.

Another consideration in the interpretation of experimental results is the actual location of enkephalin release. As discussed above, μ-opioid receptor agonists infused directly into the NAc induce feeding and feeding itself induces transcriptional changes in preproenkephalin in the NAc.[7, 16-18] However, to date, there has been no direct evidence of feeding-induced enkephalin release in the NAc. In order to definitely identify the site of enkephalin release, we employed a tissue punch to specifically remove the NAc (see Figure 1). After neuropeptide extraction and MS analysis, we found a feeding-induced reduction in vesicular enkephalin within the NAc. Given that the majority of enkephlain is stored in the nerve terminals, we assume the enkephalin release is occurring within the NAc itself. This does not preclude the simulatenous release of enkephalin from the NAc into other areas of the brain, such as the hypothalamus or the ventral tegmental area. To our knowledge this is the first evidence of feeding-induced enkephalin release within the NAc.

Figures

Figure 1.

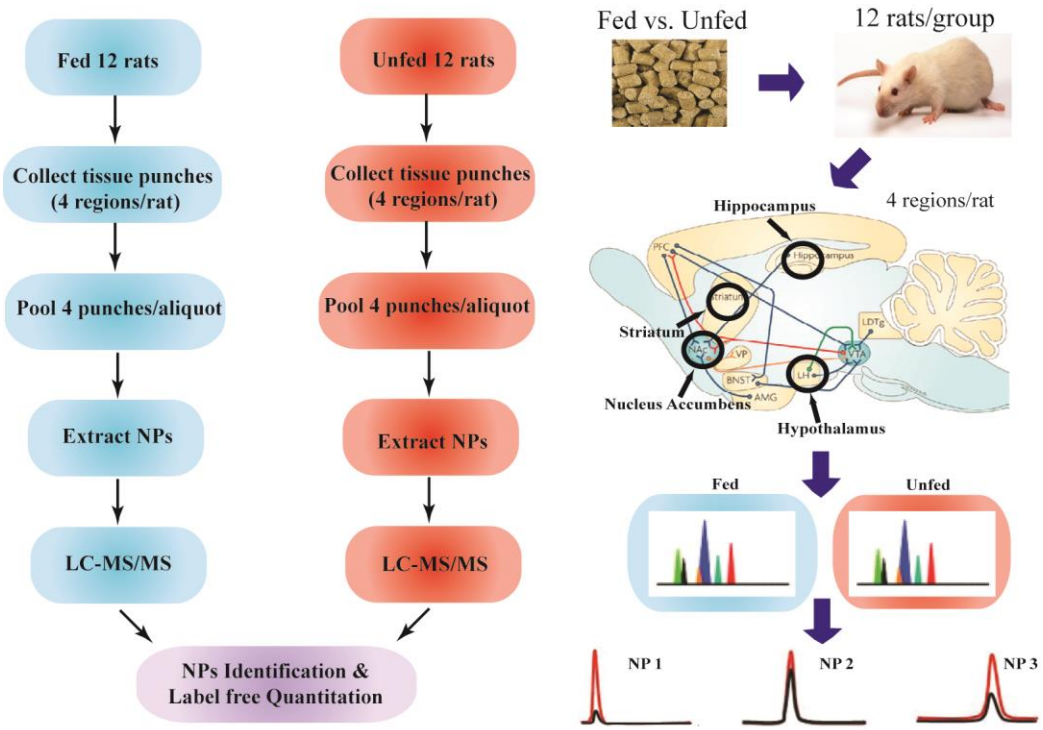


Figure 2. (A) CCK, (B) galanin

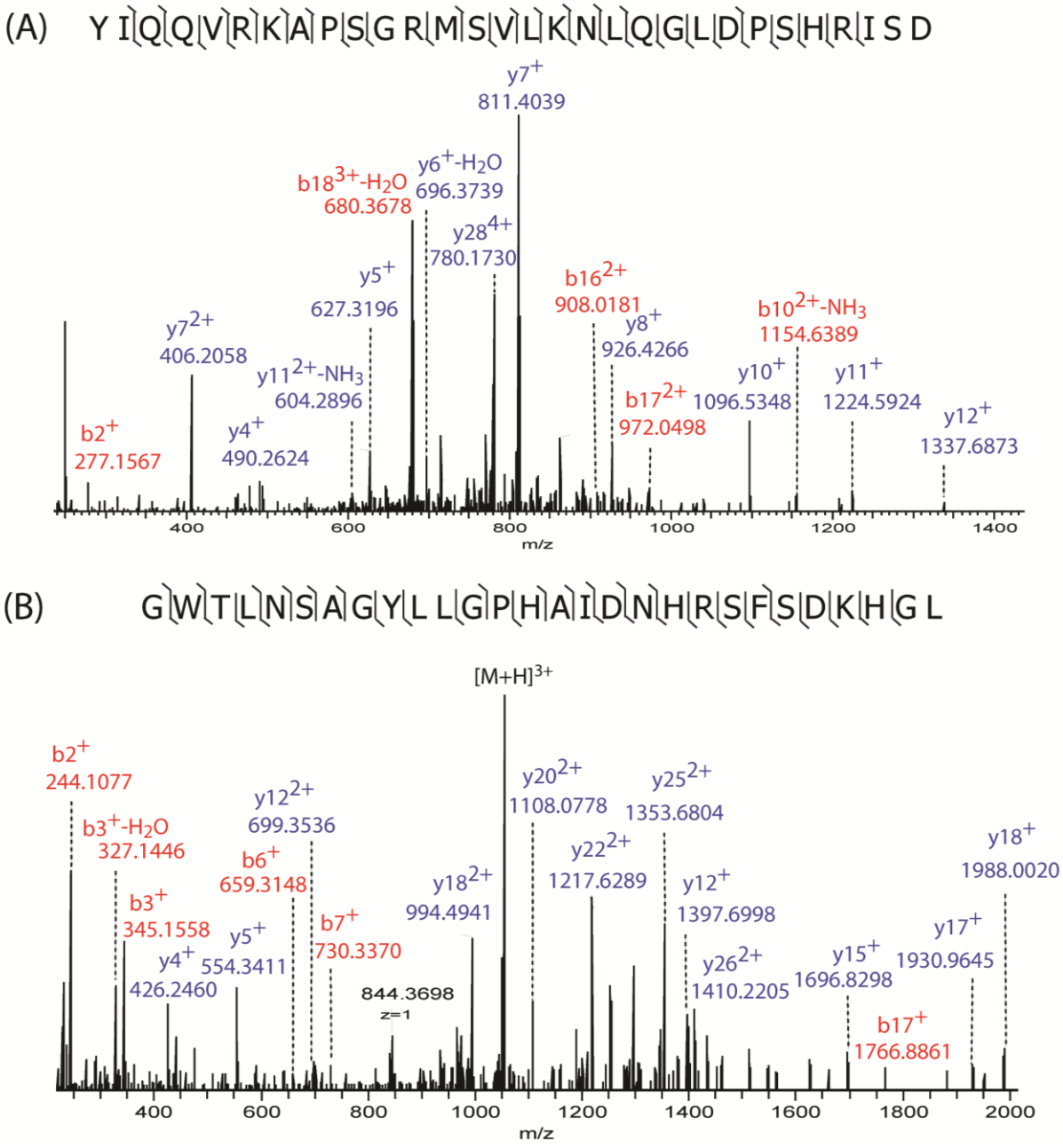
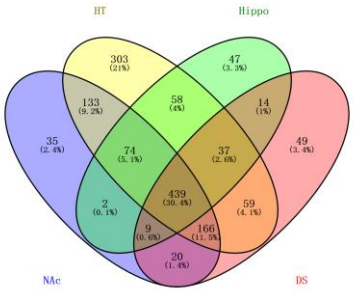
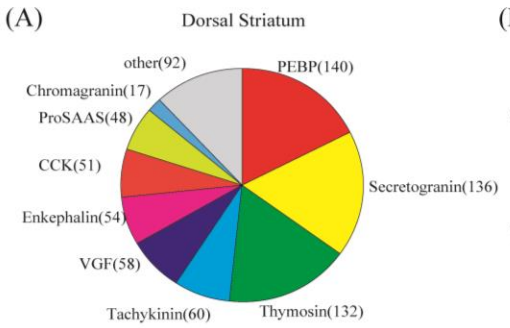
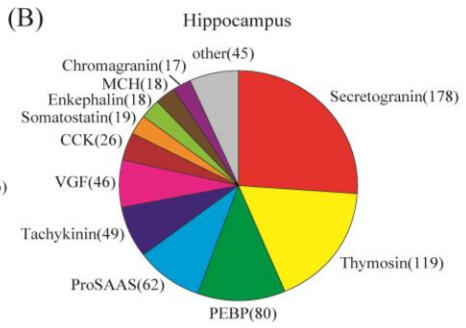


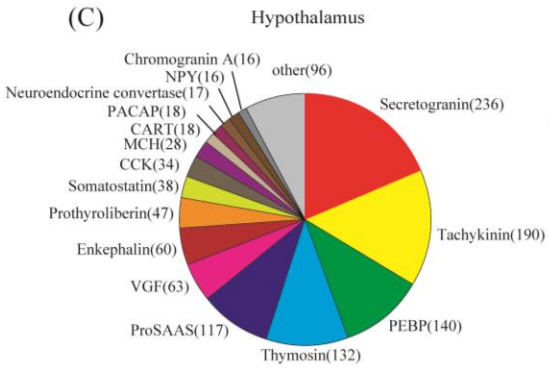
Figure 3.

Figure 3


HT
Hippo
NAc
DS
303
(21%)
47
(3.3%)
133
(9.2%)
58
(4%)
14
(1%)
35
(2.4%)
74
(5.1%)
37
(2.6%)
49
(3.4%)
439
(30.4%)
2
(0.1%)
59
(4.1%)
9
(0.6%)
166
(11.5%)
20
(1.4%)


(A)
Dorsal Striatum
other(92)
PEBP(140)
Chromagranin(17)
ProSAAS(48)
CCK(51)
Secretogranin(136)
Enkephalin(54)
VGF(58)
Tachykinin(60)
Thymosin(132)


(B)
Hippocampus
other(45)
Chromagranin(17)
MCH(18)
Secretogranin(178)
Enkephalin(18)
Somatostatin(19)
CCK(26)
VGF(46)
Tachykinin(49)
Thymosin(119)
ProSAAS(62)
PEBP(80)


(C)
Hypothalamus
Chromogranin A(16)
NPY(16)
other(96)
Neuroendocrine convertase(17)
Secretogranin(236)
PACAP(18)
CART(18)
MCH(28)
CCK(34)
Somatostatin(38)
Prothyroliberin(47)
Tachykinin(190)
Enkephalin(60)
VGF(63)
PEBP(140)
ProSAAS(117)
Thymosin(132)

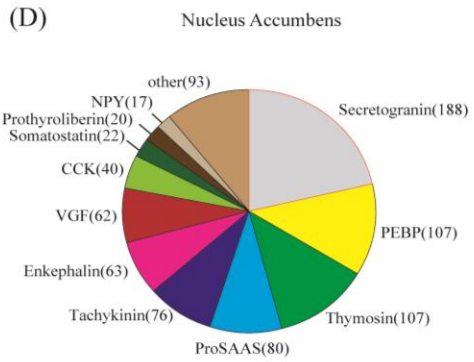
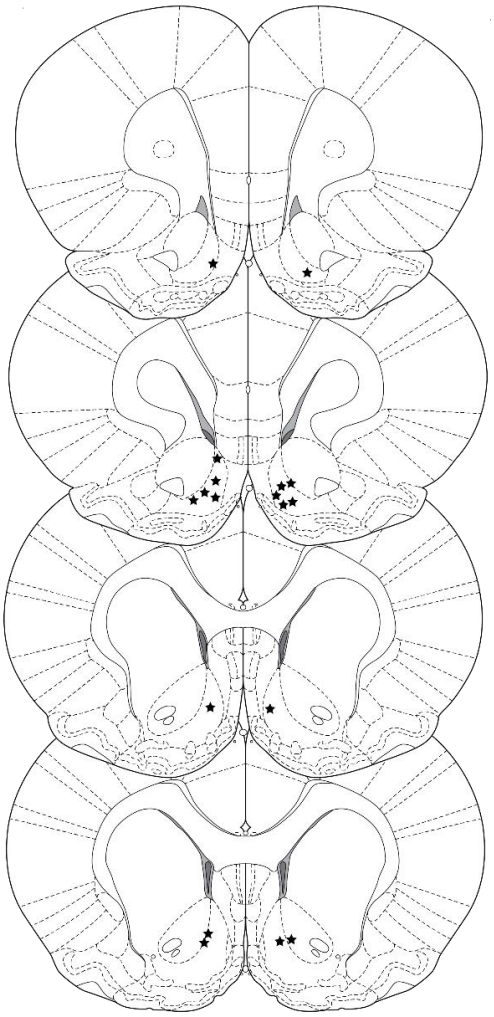

(D)
Nucleus Accumbens
other(93)
NPY(17)
Secretogranin(188)
Prothyroliberin(20)
Somatostatin(22)
CCK(40)
VGF(62)
PEBP(107)
Enkephalin(63)
Tachykinin(76)
Thymosin(107)
ProSAAS(80)

Figure 4

(a)



2.76

2.52

2.28

2.16

(b)

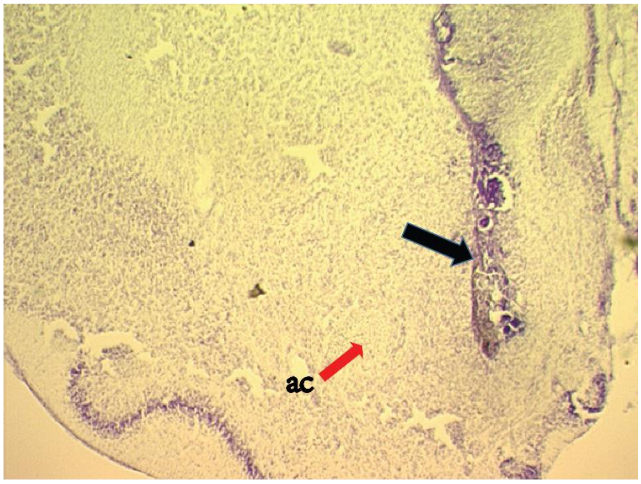


Figure 5

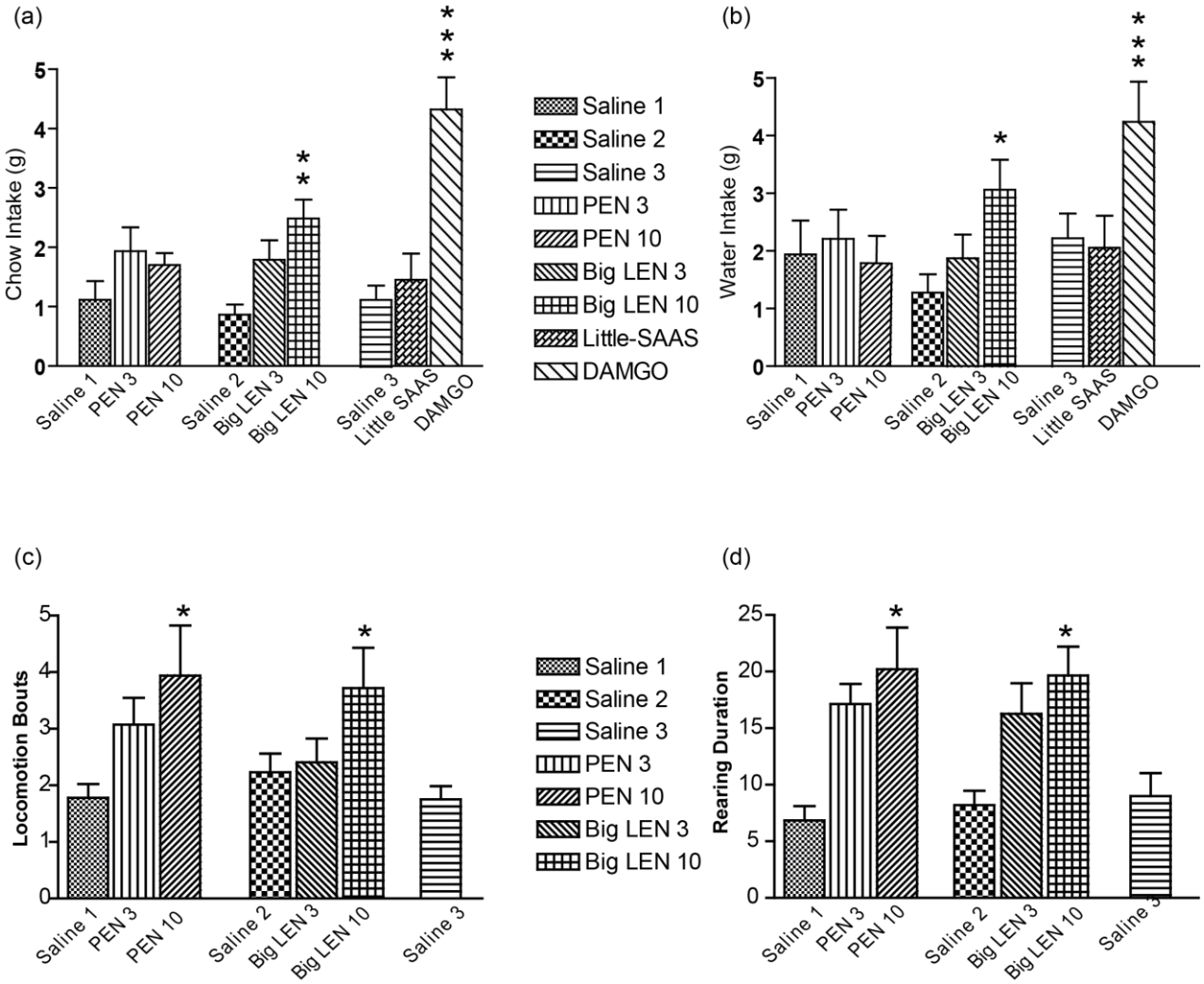
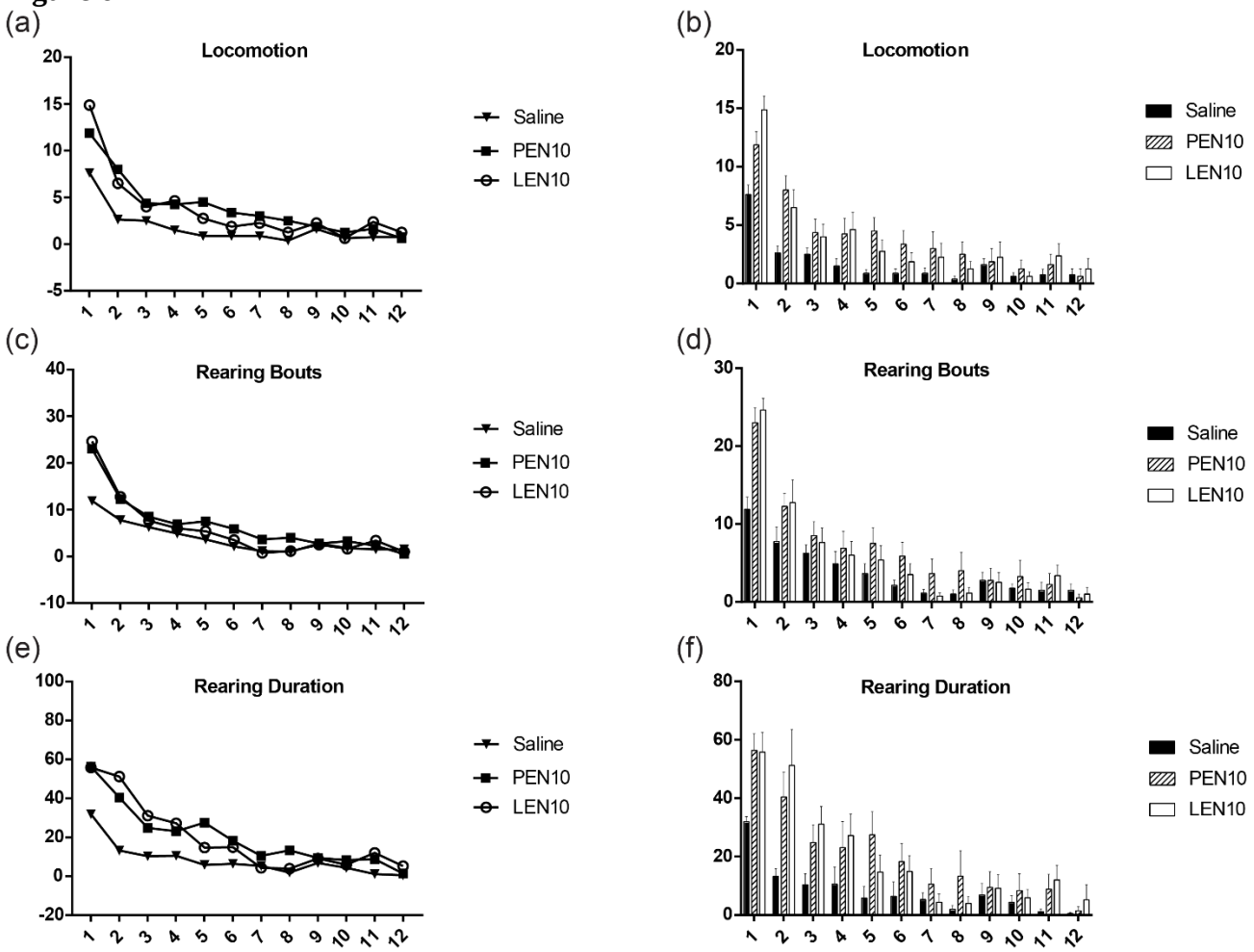

(a)
Chow Intake (g)
0
1
2
3
4
5
***
**
Saline 1
PEN 3
PEN 10
Saline 2
Big LEN 3
Big LEN 10
Saline 3
Little SAAS
DAMGO
(b)
Water Intake (g)
0
1
2
3
4
5
*
***
Saline 1
PEN 3
PEN 10
Saline 2
Big LEN 3
Big LEN 10
Saline 3
Little SAAS
DAMGO
(c)
Locomotion Bouts
0
1
2
3
4
5
*
*
Saline 1
PEN 3
PEN 10
Saline 2
Big LEN 3
Big LEN 10
Saline 3
(d)
Rearing Duration
0
5
10
15
20
25
*
*
Saline 1
PEN 3
PEN 10
Saline 2
Big LEN 3
Big LEN 10
Saline 3
Saline 1
Saline 2
Saline 3
PEN 3
PEN 10
Big LEN 3
Big LEN 10
Little-SAAS
DAMGO

Figure 6


(a)
Locomotion
Saline
PEN10
LEN10
(b)
Locomotion
Saline
PEN10
LEN10
(c)
Rearing Bouts
Saline
PEN10
LEN10
(d)
Rearing Bouts
Saline
PEN10
LEN10
(e)
Rearing Duration
Saline
PEN10
LEN10
(f)
Rearing Duration
Saline
PEN10
LEN10

References:

1. H. Hao, X. Zheng and G. Wang, *Trends Pharmacol Sci*, 2014, **35**, 168-177.
2. X. B. Wang, Y. Feng, N. Wang, F. Cheung and C. W. Wong, *Recent Pat Food Nutr Agric*, 2012, **4**, 91-106.
3. D. D. Soejarto, C. Gyllenhaal, M. R. Kadushin, B. Southavong, K. Sydara, S. Bouamanivong, M. Xaiveu, H. J. Zhang, S. G. Franzblau, G. T. Tan, J. M. Pezzuto, M. C. Riley, B. G. Elkington and D. P. Waller, *Pharm Biol*, 2012, **50**, 42-60.
4. V. Hanot, S. Goscinny and M. Deridder, *J Chromatogr A*, 2015, **1384**, 53-66.
5. L. Wu, P. Gong, Y. Wu, K. Liao, H. Shen, Q. Qi, H. Liu, G. Wang and H. Hao, *J Chromatogr A*, 2013, **1303**, 39-47.
6. G. Dowling, L. Regan, J. Tierney and M. Nangle, *J Chromatogr A*, 2010, **1217**, 6857-6866.
7. H. Cao, A. Zhang, H. Zhang, H. Sun and X. Wang, *Phytother Res*, 2015, **29**, 159-166.
8. S. Naz, M. Vallejo, A. García and C. Barbas, *J Chromatogr A*, 2014, **1353,** 99-105.
9. J. Dong, Y. Zhu, X. Gao, Y. Chang, M. Wang and P. Zhang, *J Pharm Biomed Anal*, 2013, **80**, 50-62.
10. L. Bell, M. J. Oruna-Concha and C. Wagstaff, *Food Chem*, 2015, **172**, 852-861.
11. M. J. Martínez Bueno, S. Uclés, M. D. Hernando and A. R. Fernández-Alba, *Talanta*, 2011, **85,** 157-166.
12. J. F. Xiao, B. Zhou and H. W. Ressom, *Trends Analyt Chem*, 2012, **32**, 1-14.
13. Y. Liang, H. Hao, L. Xie, A. Kang, T. Xie, X. Zheng, C. Dai, K. Hao, L. Sheng and G. Wang, *Drug Metab Dispos*, 2010, **38**, 1747-1759.
14. S. Broecker, S. Herre, B. Wüst, J. Zweigenbaum and F. Pragst, *Anal Bioanal Chem*, 2011, **400**, 101-117.
15. M. Paul, J. Ippisch, C. Herrmann, S. Guber and W. Schultis, *Anal Bioanal Chem*, 2014, **406**, 4425-4441.
16. X. Zhu, Y. Chen and R. Subramanian, *Anal Chem*, 2014, **86**, 1202-1209.
17. A. T. Roemmelt, A. E. Steuer, M. Poetzsch and T. Kraemer, *Anal Chem*, 2014, **86**, 11742-11749.
18. J.-L. Wolfender, G. Marti, A. Thomas and S. Bertrand, *J Chromatogr A*, 2015, **1382**, 136-164.
19. S. Kazuno, M. Yanagida, N. Shindo and K. Murayama, *Anal Biochem*, 2005, **347**, 182-192.
20. G. Hopfgartner, D. Tonoli and E. Varesio, *Anal Bioanal Chem*, 2012, **402**, 2587-2596.
21. D. Moran, T. Cross, L. M. Brown, R. M. Colligan and D. Dunbar, *J Virol Methods*, 2014, **195**, 9-17.
22. S. Liu, X. Chen, Z. Yan, S. Qin, J. Xu, J. Lin, C. Yang and W. Shui, *Proteomics*, 2014, **14**, 169-180.
23. R. P. Dator, K. W. Gaston and P. A. Limbach, *Anal Chem*, 2014, **86**, 4264-4270.
24. J. D. Egertson, A. Kuehn, G. E. Merrihew, N. W. Bateman, B. X. MacLean, Y. S. Ting, J. D. Canterbury, D. M. Marsh, M. Kellmann, V. Zabrouskov, C. C. Wu and M. J. MacCoss, *Nat Methods*, 2013, **10**, 744-746.
25. D. Siegel, A. C. Meinema, H. Permentier, G. Hopfgartner and R. Bischoff, *Anal Chem*, 2014, **86**, 5089-5100.

26. H. Li, Y. Yu, Z. Wang, J. Geng, Y. Dai, W. Xiao and X. Yao, *PLoS One*, 2015, **10**, e0121031.
27. H. Hao, N. Cui, G. Wang, B. Xiang, Y. Liang, X. Xu, H. Zhang, J. Yang, C. Zheng, L. Wu, P. Gong and W. Wang, *Anal Chem*, 2008, **80**, 8187-8194.
28. L.-W. Qi, H.-Y. Wang, H. Zhang, C.-Z. Wang, P. Li and C.-S. Yuan, *J Chromatogr A*, 2012, **1230**, 93-99.
29. W. Liu, J. Chen, T. Xu, W. Tian, Y. Li, Z. Zhang and W. Li, *Am J Hypertens*, 2012, **25**, 250-260.
30. Y. Zou, L. Lin, Y. Ye, J. Wei, N. Zhou, Y. Liang, H. Gong, L. Li, J. Wu, Y. Li, Z. Jia, Y. Wu, J. Zhou and J. Ge, *J Cardiovasc Pharmacol*, 2012, **59**, 268-280.
31. H. Xiao, Y. Song, Y. Li, Y. H. Liao and J. Chen, *Cell Immunol*, 2009, **260**, 51-55.
32. N. Madala, P. Steenkamp, L. Piater and I. Dubery, *Anal Bioanal Chem*, 2012, **404**, 367-372.

Chapter 3

Label Free Quantitative Comparison of Cerebrospinal Fluid Glycoproteins and Endogenous Peptides in Subjects with Alzheimer's disease, Mild Cognitive Impairment and Healthy Individuals

Adapted from manuscript: "Label Free Quantitative Comparison of Cerebrospinal Fluid Glycoproteins and Endogenous Peptides in Subjects with Alzheimer's disease, Mild Cognitive Impairment and Healthy Individuals" **Jingxin Wang**, Robert Cunningham, Henrik Zetterberg, Sanjay Asthana, Cynthia Carlsson, Ozioma Okonkwo, Lingjun Li

Abstract

Alzheimer's disease (AD) is the most common form of dementia in the elderly. It is fundamentally necessary to discovery and establish reliable and accurate biomarkers that are capable of predicting, diagnosing and monitoring disease progression. Here we report a cohort of glycoproteins and endogenous peptides identified and compared in cognitively healthy, mild cognitive impairment and AD via mass spectrometry (MS)-based strategies. By an optimized sub-microgram peptide separation using molecular weight cut-off filtration and in house-constructed database resulted in 645 peptides identified. Glycoproteins were effectively enriched by lectin affinity chromatography, resulting in 795 identified. The discovery and alterations of ProSAAS-derived peptides and transthyretin are described and their roles in AD are discussed.

INTRODUCTION

Alzheimer's disease (AD) is the most common form of dementia in the elderly population and the 6th leading cause of death in the United States. More than 36.5 million people were estimated to suffer from dementia in 2010, and there are 7.7 million new cases each year. In addition, the number of dementia patients will nearly double every 20 years. Due to the aging baby boom generation, it is predicted that the number of people with Alzheimer's disease age 65 and older may triple by 2050, costing an estimated $1.2 trillion.[1] AD is characterized by progressive accumulation of extracellular β-amyloid (Aβ) peptides and intracellular

neurofibrillary tangles of protein tau, in addition to synaptic and neuronal loss in the brain.[2] Currently, the diagnosis of AD is mainly based on the history of a patient or an objective cognitive assessment and the exclusion of dementia due to other causes such as Lewy body dementia, vascular dementia, frontotemporal dementia, Creutzfeldt-Jakob disease (CJD) etc.[3] Three AD biomarkers: Aβ 1-42, total tau (t-tau) and phosphorylated tau (p-tau) in the cerebrospinal fluid (CSF) have been used to facilitate the diagnostic criteria.[4, 5, 6, 7] These biomarkers may increase the certainty of AD pathophysiological process only in patients who have already met the critical clinical criteria for probable AD dementia. However, the AD biomarker tests are not advocated routinely for diagnostic purposes these days. The reason is mainly due to the limitation of utilizing these biomarkers to predict and monitor the disease progression. Specifically, the recently reported failure of a phase 3 clinical trial of humanized monoclonal antibody of soluble forms of amyloid[8, 9] necessitates the discovery and establishment of other biomarkers and targets for disease prevention and/or treatment.

Although lumbar puncture is an invasive procedure, CSF bathes the brain and spinal cord which makes it a valuable biological fluid for biomarker studies of neurodegenerative diseases.[10] CSF offers promise to facilitate disease diagnosis, monitoring disease progression, and the efficacy of potential therapies. CSF is generated in the choroid plexus and composed of blood filtrate and brain extracellular matrix. Despite the differences between CSF and plasma the proteome of both share the vast majority of abundant protein identifications,

although it is estimated that around 20% of the CSF proteome are produced by brain itself.[11, 12] Compared to a brain biopsy which could cause permanent damage, and approximately 500 ml of CSF is generated each day with a maximum of 20 ml sampled per spinal tap.[13] CSF is reabsorbed into the blood four times per day and due to the constant turnover the dynamic changes of CSF composition is able to reflect the disease status of the central nervous system (CNS). In a clinical setting, the lumbar puncture is a routine procedure for the diagnosis of multiple sclerosis, Guillain-Barré syndrome and CJD etc. Previous large scale proteomic studies have greatly filled in the gap of our knowledge about CSF protein composition by two dimensional gel,[14, 15, 16, 17] chromatography and mass spectrometry.[18, 19, 20, 21, 22]

Glycosylation is one of the most complicated but common forms of post-translational modifications in proteins. It modulates cell-molecule, cell-matrix and cell-cell interactions and facilitates the assembly and development of complex organisms. The major functions of a glycan can be characterized into two categories: (1) structural and modular functions; and (2) recognition of glycans by other molecules.[23] There are two major forms of protein glycosylation: glycans linked to an asparagine (N-glycans) and glycans linked to a serine/threonine (O-glycans). N-glycosylation occurs in the endoplasmic reticulum (ER) while O-glycosylation is more dynamic occurring in nuclear and cytoplasmic proteins. Lectin affinity chromatography is a useful and practical separation technique in glycoproteomics. It is characterized by selectively binding a specific carbohydrate motif. Among them, concanavalin-A (ConA) and wheat germ agglutinin (WGA) are

most commonly used. Specifically, ConA has higher affinity for N-linked glycoproteins with high-mannose and terminal glucose carbohydrates while WGA has a higher affinity for terminal acetylglucosamine and sialic acid with O-linked glycosylation motif. In addition, ConA has a special binding site for excessively hydrophobic proteins while WGA is able to bind both hydrophobic and hydrophilic proteins.[24] Their functional properties are useful to utilize as a combination of both ConA and WGA which can obtain an optimal and impactful glycoprotein enrichment. Previous studies indicate that the global protein glycosylation pattern is altered in AD.[25, 26, 24] Recently, the potential precursor of AD-mild cognitive impairment (MCI) has drawn increased attention since treatments could potentially be initiated before the damage of the neurodegenerative process becomes too extensive.[27, 28] One can investigate patients with cognitive impairment that is worse than age-matched healthy individuals but not severe enough to fulfill the criteria for suspected AD. People with MCI display similar pathophysiologic symptoms as those with AD and have a 10%-15% relative risk of becoming AD annually.[29, 30] Here we reported a comparative glycoproteomic and peptidomic discovery study using CSF from healthy (control), MCI and AD individuals via mass spectrometry (MS). This study demonstrates the usefulness of glycoprotein enrichment in biomarker discovery which facilitates the isolation, identification and relative quantification of lectin affinity enriched glycoproteins from CSF samples via multidimensional separation and high resolution accurate mass tandem mass spectrometry.

MATERIALS AND METHODS

Participants

Twelve enrollees in the Wisconsin Alzheimer's Disease Research Center (ADRC) participated in this study. This research was conducted in March 5, 2010 to February 13, 2013. The samples comprised of four cognitively normal individuals who enrolled in the Wisconsin ARDC at late middle age without family history of AD and eight patients with cognitive impairment. The cognitively impairment group consisted of four people with amnestic MCI and for people with mild AD. All AD and MCI participants were diagnosed via applicable clinical criteria characterized by standardized, consensus, multidisciplinary consensus conferences,[31, 3] while cognitive normalcy was determined based on intact cognitive performance by a comprehensive battery of neuropsychological tests, lack of functional impairment, and absence of neurological or psychiatric conditions that might impair cognition.[32, 33] Women comprised seven out of a total of twelve samples and the mean (SD) age of the total sample was 75.24 (±3.37) years. Cognitive reserve (CR) was indexed by years of education. Individuals with fewer than 16 years of education were considered as having low CR (n=2), while those with at least 16 years of education were considered as having high CR (n=10).[34] The University of Wisconsin Institutional Review Board approved all study procedures. Each enrollee provided a signed informed consent form before participation.

CSF Collection

A lumbar puncture for CSF sample collection was performed in the morning after a 12-hour fast. A Sprotte 24-gauge or 25-gauge spinal needle was inserted at L3/4 or L4/5 following proper local anesthesia. Each CSF sample was collected via syringes with polypropylene sample tubes. The samples were combined and gently mixed for each individual consisted of approximately 22 mL of CSF. Subsequently, the CSF samples were centrifuged at 2000g for ten minutes and the supernatant were collected in 0.5 mL aliquots in polypropylene tubes, frozen and stored at -80°C. The samples were immunoassayed for Aβ1-42, t-tau and p-tau by enzyme-linked immunosorbent assays (INNOTEST; Fujirebio) by board-certified technicians who were blind to clinical data and used protocols accredited by the Swedish Board for Accreditation and Conformity Assessment as described in a previous report.[35] Table 1 represents the demographic characteristics of the study participants. The mean age of all individuals was 75 with a SD of ±3.4, ranging from 70.3 to 82.5. The mean years of education was 16.4 with a SD of 2.5.

Materials

Iodoacetamide (IAA), N-acetyl-D-glucosamine, methyl-R-D-mannopyranoside, methyl-R-D-glucopyranoside, manganese chloride tetrahydrate and protease inhibitor cocktail were purchased from Sigma-Aldrich (St. Louis, MO). Ammonium bicarbonate, urea, formic acid (FA), Tris hydrochloride, acetonitrile, methanol,

optima LC/MS grade water, sodium chloride (99.5%), calcium chloride, and sodium acetate were obtained from Thermo Fisher Scientific (Pittsburgh, PA). Agarose bound Concanavalin A (Con A, 6 mg lectin/mL gel) and Wheat Germ Agglutinin (WGA, 7 mg lectin/mL gel) were obtained from Vector Laboratories (Burlingame, CA). Sequencing grade modified trypsin and dithiothreitol (DTT) were from Promega (Madison, WI). 660 nm protein assay kit was purchased from Pierce (Rockford, IL) and microplate reader was purchased from BioTek Instrument (Winooski, VT). Amicon Ultra 0.5 mL 10,000 molecular weight cut-off (MWCO) membrane-based centrifugal filters and C18 ZipTips were purchased from Millipore (Billerica, MA). 100 µL Omix Tips packed with C18 reversed-phase resin were obtained from Agilent (Palo Alto, CA).

Peptide separation

Two aliquots of CSF samples (total 1 mL) of each participant were used for this study. Protease inhibitor cocktail was added to the CSF by the recommended ratio immediately before use. Endogenous peptide separations were performed using Amicon Ultra 0.5 mL 10,000 MWCO centrifugal filters. Before the MWCO separation, three washing steps were performed to remove the contaminants from the filter and achieve an optimal peptide coverage. The three washes were 500 µL water: methanol, 500 µL of water, and 400 µL 70:30 aqueous 1 M sodium chloride:methanol. This method yielded a better peptide coverage after separated from a complex protein mixture and was reported by our previous study.[36] The MWCO filter was subjected to centrifugation at 14,000 g for 5 min at 4°C via an

Eppendorf 5415 D microcentrifuge (Brinkmann Instruments Inc., Westbury, NY). The flow-through was then concentrated by a Savant SC 110 SpeedVac concentrator (Thermo Electron Corporation, West Palm Beach, FL) and resuspended in 20 µL 0.1% formic acid. The resulting sample was desalted by C18 ZipTips by the manufacturer's protocol. Specifically, the ZipTips were washed with 100% ACN and then pre-equilibrated with 0.1% formic acid in water for three times respectively. Next, the endogenous peptides were loaded onto the C18 ZipTips repeatedly and gently. The loaded peptides were desalted using 0.1% formic acid in water three times and then eluted in 20 µL of 50% ACN in 0.1% formic acid. Subsequently, the solution was dried down and resuspended in 10 µL 0.1% formic acid in water and subjected to LC MS/MS analysis.

Lectin Affinity Chromatography

Lectin affinity columns were prepared in house by adding 75 µL WGA and 150 µL ConA slurry both bound to agarose beads into an empty Spin Columns-Screw Cap (Pierce, Rockford, IL). The columns were washed two times to remove contaminants with lectin affinity chromatography (LAC) binding buffer (0.15 M NaCl, 0.02 M Tris-HCl, 1 mM CaCl2, pH=7.4). Each time the columns were centrifuged at 400 g for 30 seconds with the bottom plug removed. Subsequently, 500 µL CSF was loaded into the lectin affinity column with the bottom plug placed back. After incubating with continuous mixing at room temperature for two hours, the sample was centrifuged at 400 g for 30 seconds and the unbound flow-through fraction was discarded. The lectin beads were washed with 300 µL LAC binding buffer

twice, centrifuged, and flow-through discarded. The lectin beads with captured glycoproteins were eluted after vortexing for 10 minutes with 300 µL LAC eluting buffer (0.075 M NaCl, 0.01 M Tris-HCl, 0.2 M alphamethyl mannoside, 0.2 M alpha-methyl glucoside, and 0.5 M acetyl-Dglucosamine) and the elution protocol was performed twice and combined for each participant. The protein concentration was determined by 660 nm protein assay and the final protein amount was normalized for all participant samples.

Protein digestion

Each sample consisted of 20 µg of protein and was denatured with 8 M urea in 50 mM ammonium bicarbonate buffer and reduced by DTT (final concentration as 20 mM) via incubation at 37 °C for 1 h. After incubation, the reduced sample was alkylated by IAA and allowed to react in the dark for 15 min, making final concentration of IAA as 60 mM. Afterwards, DTT was added to quench IAA and incubated for 10 min. The sample was diluted to 1 M urea by 50 mM ammonium bicarbonate (pH=8.5) followed by addition of trypsin (trypsin: protein= 1: 40). The digestion was incubated for 18 h at 37°C and subsequently quenched by 2.5 µL 10% formic acid. The solid phase extraction of the tryptic peptides were performed by Varian 100 µL C18 Omix Tips (Palo Alto, CA) by similar steps as described above. The peptides were sequentially eluted with 50% ACN in 0.1% formic acid, dried by SpeedVac and reconstituted with 20 µL 0.1% formic acid.

LC – ESI Orbitrap mass spectrometry data acquisitions

Online reversed-phase liquid chromatography separation of the tryptic peptides was performed on a nanoAcquity UPLC (Waters Corp., Milford, MA) and infused into a Q Exactive quadrupole orbitrap mass spectrometer (Thermo Fisher Scientific, San Jose, CA). The chromatographic separation was carried out via mobile phase A that consisted of 0.1% formic acid in water and mobile phase B consisting of 0.1% formic acid in ACN. One µL of tryptic or endogenous peptides were loaded onto a 2 cm, 150 µm i.d. PLRP-S d_p 5 µm, pore size 1000 Å, trap column at a flow rate of 5 µL/min for 5 minutes at 95% A and 5% B, and subsequently separated by a Waters BEH 130 Å C18 reversed-phase capillary column (150 mm × 75 µm, 1.7 µm). Emitter tips were pulled from capillary tubing 75 µm I.D. (Polymicro Technologies, Phoenix, AZ) using a model P-2000 laser puller (Sutter Instrument Co., Novato, CA). The LC gradient was 5-30% B over 120 minutes at a flow rate of 350 nL/min. Ions were generated under positive electrospray ionization (ESI) at a capillary voltage of 2.8 KV; 275°C capillary temperature; 30% collision energy via high energy collision dissociation (HCD). MS1 scans were acquired over 200–2000 *m/z* at 70 k resolution followed by data dependent selection of the top 10 most abundant precursor ions with an isolation window of 2.0 *m/z.* HCD fragmentation was then performed and analyzed at a resolution of 17,500. Other parameters include: automatic gain control 1e5; maximum ion injection time, 100 ms; dynamic exclusion enabled with unassigned, +1 and greater than +8 charges ignored for MS/MS selection. Each sample was injected three times to acquire technical triplicates.

Data analysis

Endogenous peptide discovery was performed via Peaks Studio 7 software (Bioinformatics Solutions Inc., Waterloo, ON, Canada). All raw LC-MS/MS data were processed by Peaks software for spectral interpretation. The database was in-house constructed based on Uniprot's reference database of *Homo sapiens* (release 06_2014) by predicting neuropeptides sequence based on *in silico* peptide cleavage principles.[37] There were 16690 entries in total. Data Refinement was applied as default to correct the precursor mass and charge states. For De Novo and Peaks Search the parameters were: non-enzyme; MS 1 and MS/MS mass tolerance as 10 ppm and 0.1 Da respectively; amidation (C-terminal), acetylation (N-terminal), oxidation, pyroglutamination of glutamic acid and glutamine as variable PTMs; monoisotopic mass values. Estimation of false positive identification rate was determined by searching all sepctra against a decoy database and with a false discovery rate (FDR) set at ≤1%, and only peptides with -10 log*P* score ≥ 15 were considered as confident identifications.

Tryptic peptide identification was carried out via Proteome Discoverer 1.4 (Thermo Scientific). A FASTA file was downloaded from Uniprot's reference database of *Homo sapiens* (release 06_2014). Other parameters include: allowed missed cleavage, 1; enzyme, trypsin; fixed modification, carbamidomethylation of cysteine (+57.0215 Da); variable modification, oxidation of methionine (+15.9949 Da);

peptide mass tolerance, 10 ppm; fragment mass tolerance, 0.1 Da. q value was set to achieve 1% false discovery rate (FDR) via the Percolator node to verify the identified peptides, and the results were filtered by high confidence peptide identification. Only protein identifications that appeared in at least two technical replicates were considered. Label free quantification (LFQ) was conducted to compare control, MCI and AD glycoproteins. Protein concentration was estimated by the sum of normalized peak areas of its tryptic peptides. Area under curve of acquired data was automatically calculated by Proteome Discoverer 1.4 by an event detection node that was applied at a setting of 4 ppm along with the precursor ion peak detector node to extract ion chromatograms. Only unique peptides to one assigned protein were calculated and summed to achieve the cumulative peak area. One way ANOVA analysis was applied to detect the significant protein differential expression. Online software was applied and data was manually input into Statistica. Protein peak area with significant changes in statistical analysis was subjected to post-hoc comparisons among means via Tukey's HSD test. $P < 0.05$ was considered to be statistically significant. Only proteins that were identified in three out of four individuals' CSF in each group were further analyzed by one way ANOVA.

RESULTS AND DISCUSSION

Endogenous Peptide Discovery

Figure 1 demonstrates the schematic illustration of the experimental design which enabled detection and analysis of large proteins and endogenous peptides simultaneously. It is worth noting that our previous work has provided evidence of significant sample loss when the amount of peptides were below micrograms using MWCO, and it also illustrated a solution to minimize the peptide loss and obtain an optimal peptide recovery.[36] Specifically, MeOH and NaCl was added to wash the MWCO membrane before separation which is usually comprised of regenerated cellulose. Numerous free hydroxyl groups of these cellulose could potentially cause significant non-specific adsorption of endogenous peptides when separated by MWCO. The addition of MeOH and NaCl may disrupt the interaction and non-specific interactions between peptides and hydroxyl groups of the regenerated cellulose. This application has greatly facilitated our peptidomic studies of human body fluids and significantly increased the number of peptide identification.

Peptide identification from tandem MS data is a fundamental task in peptidomic studies. The sensitivity and accuracy greatly impacts the performance of peptide identification as well as downstream analysis. *De novo* sequencing and database searching are two of the most pivotal cornerstones in both proteomic and peptidomic research by MS. In this study, we have improved these two aspects and significantly improve the number of endogenous peptides identified expanding the peptidome of CSF. Our study has greatly surpassed the previously reported 563 peptides in the CSF peptidome by an extensive and combined strategy.[38]

Specifically, our house-constructed database was computed based on *in silico* peptide cleavage principles[37] of protein precursors and compiled of previous reported bioactive endogenous peptides of human. The rationale is functional peptides are produced from their precursor proteins by various processing enzymes through different pathways. Cysteine and subtilisin-like protease are involved in two major peptide processing pathways and they mainly cleave at the dibasic sites.[39, 40] This database construction method was described in detail by a previous study.[37] With the exclusion of non-bioactive peptide precursors, potential novel neuropeptides could be identified more specifically and with improved sensitivity.[41] *De novo* sequencing was mostly used when a protein database was unavailable and require spectra with higher mass accuracy. The Orbitrap mass spectrometer enables both high mass accuracy and increased speed of analysis for proteomics and peptidomics analysis. PEAKS DB was performed for *de novo* sequencing to assist in database searching due to its superior performance at sensitivity and accuracy.[42] In addition, PEAKS provided a more robust result validation and precision in controlling the FDR. A total of 645 endogenous peptides (Supplementary Table 1) were identified with the improved algorithm described above. Proteome Discoverer 1.4 (Thermo Scientific) was also used to search the same raw data file using the identical database and searching parameters, but only 53 peptides were identified among all the samples. The 645 endogenous peptides identified by Peaks Studio 7 were derived from 93 protein precursors. 391 confident identifications were discovered in cognitively healthy individuals, while 261 and 208 were found in MCI and AD patients respectively which was illustrated

by the Venn Diagrams (Figure 2c). The 645 unique peptides covered a wide variety of biological processes (shown in Figure 2d), molecular functions and pathways analyzed by PANTHER gene ontology database search. The top biological functional categories include cellular process (24.8%), biological regulation (13.5%), metabolic process (13.5%) and organismal process (12.8%). Secretory proteins comprised a large portion of the 93 protein precursors and their functions include binding (45.1%), catalytic activity (19.7%) and receptor activity (16.9%). The peptides consisted of several major neuropeptides families: proSAAS, neurosecretory protein VGF, Cholecystokinin (CCK), neuroendocrine protein 7B2, proenkephalin-A, neuroendocrine convertase 1, chromogranin-A, prepronociceptin, neurexophilin-3, neuropeptide Y, secretogranin1, 2, 3, tachykinin-3, proopiomelanocortin, insulin-like growth factor-1, chromogranin-A, orexigenic neuropeptide QRFP, preprotachykinin B etc.

ProSAAS and AD

ProSAAS, a secretory protein, is expressed in neurons throughout the CNS.[43, 44] It is proteolytically processed within the regulated secretory pathway and function as a potent and specific inhibitor of prohormone convertase 1/3 (PC1/3).[45, 46] It is also expressed in many non-PC1/3 cells, and therefore, many potential functions are still being unknown.[43, 47] Recently, proSAAS-derived peptides have drawn more and more attention and various studies have shown that they are involved in a significant number of physiologically behaviors, such as circadian rhythm,[48, 49] food intake,[50] energy balance and fetal neuropeptide processing.[44] Intriguingly,

proSAAS has been mentioned repeatedly in a number of neurodegenerative diseases. ProSAAS immunoreactivity has been observed in neuritic plagues and neurofibrillary tangles in patients with Pick's disease, parkinsonism-dementia complex and AD.[51, 52] Specifically, the N-terminal of proSAAS was found to be co-localized with tau inclusions. Moreover, several proteomic studies have discovered reduced amount of proSAAS-derived peptides in AD and frontotemporal dementia, suggesting it being a potential biomarker.[53, 54] More interestingly, a novel anti-aggregant chaperone function of proSAAS was reported recently and this anti-aggregation function of $A\beta_{42}$ could be achieved sufficiently from residues 97-180.[55]

In our study, forty two proSAAS-derived peptides were discovered in human CSF with peptide confidence level -10 log*P* score ≥ 15 by PEAKS DB, shown in Table 2. Intriguingly, all the observed peptides were either from N-terminus ($SAAS_{42-80}$) and/or C-terminus ($SAAS_{207-242}$, $SAAS_{245-260}$). It is worth pointing out that the intact form of big LEN ($SAAS_{245-260}$) and little LEN ($SAAS_{245-254}$) were discovered, and their sequences are LETPAPQVPARRLLPP and LETPAPQVPA respectively. Various fragments of big SAAS ($SAAS_{34-59}$), little SAAS ($SAAS_{42-59}$), big PEN-LEN ($SAAS_{221-260}$), PEN ($SAAS_{221-242}$) and Big LEN ($SAAS_{245-260}$) were also identified. Figure 3a presents a representative example of MS/MS spectra of an identified fragment of PEN ($SAAS_{222-238}$), DHDVGSELPPEGVLGA (*m/z* 796.3882), with a -10 log*P* score of 51.23. Similar with a tryptic peptide shown in Figure 3b, the endogenous peptide was identified with high confidence. The label free quantification using peak area was also performed on all the discovered

endogenous peptides by Peaks Studio and Proteome Discoverer 1.4, but due to Peaks Studio not display any statistical functions nor the peak area for individual peptides of every single raw data file, the peptide quantification results should be discussed conservatively and cautiously. More than 80% of the discovered proSAAS-derived peptides showed a decrease in MCI and AD group by Peaks Studio. More commonly, a decreasing trend of peak area was observed from healthly controls to MCI to AD. Using Proteome Discoverer two proSAAS-derived peptides were confidently identified and matched our criteria to be quantified, which was when a peptide was present in at least three out of four individuals' CSF within each group. Two peptides, DHDVGSELPPEGVLGA and DHDVGSELPPEGVLG was present in all twelve participants and the peak area was highly abundant and was quantified by Proteome Discoverer. For peptide DHDVGSELPPEGVLGA, the mean peak area among healthy control, MCI and AD were 4.37E+08, 1.73E+08 and 1.41E+08 respectively, and the *P* value was 0.0206. Furthermore, Tukey's test suggested the concentration decrease of this peptide was significant in MCI ($P < 0.05$) and AD ($P < 0.05$), however the decrease in AD compared to MCI was insignificant. For peptide DHDVGSELPPEGVLG, the mean peak area among healthy control, MCI and AD were 1.87E+08, 7.16E+07 and 4.96E+07 respectively, and the *P* value was 0.0297. Tukey's test indicated the concentration decrease of this peptide was significant in AD ($P < 0.05$) compared to control, however the decrease in MCI was not statistically significant.

Glycoprotein identification and quantification

For glycoprotein identification and quantification, Proteome Discoverer 1.4 was used exclusively. To broaden the identification of low abundant proteins in CSF, proteins with one or more unique peptides were considered with a false discovery rate of 0.05 at both the peptide and protein level. Proteins with only one unique peptide were manually examined by checking the peptide XCorr score and the alignment of amino acids. Each sample was subjected to technical triplicates and only proteins identified in at least two technical replicates were considered as a confident match. With the database searching algorithm and criteria discussed above, as shown in the Venn Diagram of Figure 2a, 502, 457 and 386 proteins were identified with high confidence, totaling 795 unique proteins among all 12 samples (Supplementary Table 2). The 795 proteins identified cover a wide variety of biological processes (shown in Figure 2b), molecular functions and pathways analyzed by the PANTHER gene ontology database search. The top biological functional categories include cellular process (23%), metabolic process (20%), development process (12%), and response to stimulus (10%). Secretory proteins comprised a large portion of all identified proteins and their functions include binding (28.3%), catalytic activity (28%) and receptor activity (17.5%) etc. Figure 3b displays a representative example of an MS/MS spectra of a tryptic peptide of transthyretin YTIAALLSPYSYSTTAVVTNPKE (*m/z* 1245.1498).

Label free quantitation using AUC was used to provide relative quantitation among control, MCI and AD glycoproteins. Peak area for each protein was summed and normalized by its assigned tryptic peptides. Only peptide precursors with mass

accuracy ≤4 ppm were subjected to extraction of their ion chromatograms. To avoid repeated use of a peptide for peak area of several proteins only unique peptides were calculated and summed to achieve a more accurate analysis. In order to be quantified and analyzed by ANOVA, proteins had to have been present in at least three out of four individuals' CSF within each group. Two example proteins were neural cell adhesion molecule L1 (L1CAM_HUMAN) and receptor-type tyrosine-protein phosphatase eta (PTPRJ_HUAMN) were only discovered in cognitively healthy individuals. L1CAM was present in all four participants in the control group but was not detected in either the MCI or AD group. PTPRJ was identified in three out of four participants in the control group but none of the MCI or AD groups. Thirteen other proteins displayed significant changes ($P < 0.05$) via one way ANOVA analysis to compare the mean value of the peak area among these three groups. The differentially expressed proteins are shown in Figure 4 and the proteins were transthyretin (TTHY_HUMAN), apolipoprotein E (APOE_HUMAN), contactin-1 (CNTN1_HUMAN), alpha-2-HS-glycoprotein (FETUA_HUMAN), histidine-rich glycoprotein (HRG_HUMAN), leucine-rich alpha-2-glycoprotein (A2GL_HUMAN), cell surface glycoprotein MUC18 (MUC18_HUMAN), cell adhesion molecule 2 (CADM2_HUMAN), neuronal cell adhesion molecule (NRCAM_HUMAN), neural cell adhesion molecule L1-like protein (CHL1_HUMAN), neuronal pentraxin-1 (NPTX1_HUMAN), neuronal growth regulator 1 (NEGR1_HUMAN), and Dickkopf-related protein 3 (DKK3_HUMAN).

Protein Networks

The Mean peak area, one way ANOVA *P* value, post-hoc comparisons via Tukey's HSD test of the remaining differentially expressed proteins are shown in Table 3. Figure 5a illustrates the protein-protein interaction analysis generated by the database and web-tool STRING 10.0 (http://string-db.org/). The total fifteen significantly dysregulated proteins were subjected to this molecular interaction tool with parameters: (a) medium confidence (default); (b) no more than 10 interactors to show. Two functional modules formed tightly connected clusters in the network. Further molecular function analysis identified five categories: (1) Receptor activity (40%); (2) Catalytic activity (30%); (3) Binding capacity (10%); (4) Enzyme activity (10%); (5) Transporter activity (10%).

Transthyretin and AD

Intriguingly, transthyretin (TTR) was found to increase 2.42 times ($P < 0.05$) in MCI and further decrease to 0.53 times in AD ($P < 0.01$) by post-hoc Tukey's HSD test, shown in Table 3. The mean peak area in healthy control, ACI and AD were 5.55E+08, 1.34E+09, 2.96E+08 respectively, with one way ANOVA *P* value 0.0074. TTR is a thyroid hormone carrier and plasma retinol transporter. It is encoded by a single gene copy on chromosome 18 in humans and is expressed in liver, kidney, pancreas, retinal epithelium, leptomeningeal epithelium, choroid plexus and potentially in neurons, which is most importantly related to AD.[56] Anti-TTR antibodies present in brain parenchyma indicate the neuronal endocytosis of TTR

from choroid plexus is active and TTR mRNA is effectively translated.[57, 58] Human studies have shown that TTR co-localized in amyloid plaques and vessels in hippocampi of AD patients.[57, 58] In addition, anti-TTR serum stained the majority of neuronal bodies in AD brains while only 10% of the neurons were stained in age-matched cognitively healthy individuals.[59] A large number of studies have reported a decreased level of TTR in AD patients consistent with our results.[60, 61, 62] However, there is some inconsistency in the literature on TTR levels being increased in AD patients compared to controls in CSF proteomic studies.[10, 63] Nevertheless, other CSF proteomics displayed a decline of TTR in AD patients.[64, 65, 66] The significance of the decreased level of TTR in AD is not clear, but one proposed hypothesis was that the decrease indicated TTR was able to sequester Aβ. Moreover, CSF TTR concentration may be determined partially by neuronal TTR synthesis, and the reduction in TRR could be explained by neuronal loss in AD.[59] An alternative hypothesis is that AD patients may have a genetic disposition or acquired low CSF TTR level independent of Aβ binding which could potentially put them at greater risk.[56] The discovery that TTR bound Aβ and inhibited fibril tangle formation was proved both *in vitro* and *in vivo* as discussed below.

The first *in vitro* observation was made in *C. elegans*, where a phenotype of defective locomotion in Aβ-expressing worms were rescued by wild type human TTR expressed in muscle cells.[67] In addition, the transgenic AD mouse model Tg2576 showed up-regulation of TTR in the hippocampus and cerebral cortex and TTR immunoreactivity was observed as co-localized with Aβ.[68,69] Furthermore,

increased Aβ deposition was seen in one ventricle that was injected with anti-TTR antibodies.[58] In another AD mouse model APP23, TTR and Aβ co-staining in the hippocampi, cortical regions and blood vessels were also observed.[70] The over-expression of wild type human TTR in the mouse strain (APP23/h*TTR*$^{+}$) corrected the cognitive function and special learning as well as the diminished neuropathological changes and amount of Aβ deposition. *In vitro* studies reported that most recombinant TTR variants bound to Aβ and inhibited Aβ aggregation.[71] Wild type human TTR binds to all forms of soluble Aβ, monomer, oligomer and fibrils, but has better affinity with Aβ aggregates than the monomer, and $Aβ_{1-42}$ having higher affinity for binding than $Aβ_{1-40}$.[70, 72, 73] By mass spectrometry interaction analysis, it was reported that the human TTR monomer bound Aβ with higher affinity than its tetramer form; the binding occurd at the inner β-sheet and EF helix of the A strand of TTR.[74] Lastly, various groups have reported that TTR interrupted Aβ aggregation *in vivo*.[75, 70, 72, 73]

More intriguingly, this interaction between TTR and Aβ was found to be mutual, more specifically, Aβ was found to dynamically regulate the expression of TTR as discussed below. The hippocampal slices of Tg2576 AD mice discussed above had more TTR mRNA and protein compared to wild type mice. The same phenomenon was found in the cortex and hippocampus of APP23 mice.[70, 58, 68] It was assumed that this up-regulation of TTR also occurred in human brains since TTR was extensively stained in neurons of AD patients whereas there was minimal detectable neuronal staining in age-matched cognitively normal controls.[59] In

addition, the TRR gene has been indicated to be a specific downstream target of soluble amyloid precursor protein beta(sAPPβ).[76] Thus, the quantitative result of TTR in this study could be better supported with these studies. During the early stages of AD such as preclinical AD or MCI, TTR is most likely sequesters and inhibits Aβ aggregation, and facilitated Aβ clearance. Simultaneously, the *TTR* gene is triggered for upregulation in response to the over-production of Aβ peptides and APP. This hypothesis explains why TTR was upregulated in MCI. Over time, the mount of pathogenic Aβ production exceeds the neuronal capacity to neutralize and remove them by TTR and other major Aβ binding proteins (ApoE and ApoJ) as disease progressed. The neuronal homeostasis network of the Aβ neutralizing capacity is finally disrupted and unable to control the disease. The negative feedback between Aβ and TTR was dysregulated and new TTR was not expressed due to Aβ over-production. Alternatively, neuronal TTR was consumed by Aβ which can cause a decline of TTR in advanced AD and neuronal loss in advanced AD led to reduced production of TTR. It was inferred that TTR was part of the conventional protein homeostasis network in the brain of AD, which includes unfolded protein response, heat shock induced chaperones and their co-chaperones, the proteasome ubiquitin system and autophagic responses.[56]

CONCLUSION

In conclusion, with an optimized sub-microgram peptide separation method using MWCO and in house-constructed bioactive peptide-specific database 645 endogenous peptides in CSF were identified. Interestingly, 42 proSAAS peptides

discovered were exclusively cleaved either from the N- or C-terminal of the proSAAS protein precursor, which represents similar sequences as the bioactive peptides big/little SAAS, PEN and LEN. Among them, the intact form of big LEN and little LEN were identified, which indicates potential bioactivity. Glycoproteins were enriched via lectin affinity chromatography, digested, identified by Proteome Discoverer, and quantified by comparing area under curve between cognitively healthy, MCI and AD individuals. One way ANOVA was applied for statistical analysis, and 15 proteins were found to be differentially expressed among the three groups. The dynamic changes of transthyretin increasing in MCI and then declining in AD was reported for the first time.

ACKNOWLEDGEMENTS

We would like to thank Wisconsin Alzheimer's Disease Research Center for organizing this clinical study and collecting the CSF samples. This work was supported in part by National Institutes of Health through grant 1P50AG033514. We would also like to thank the Analytical Instrumentation Center and Zeeh Pharmaceutical Experiment Station in School of Pharmacy for the shared instrument grant.

Table 1. Characteristics of Study Participants

Characteristic	Control	MCI	AD	Total
Age, mean (SD)	74.0(3.1)	76.0(2.9)	75.7(4.9)	75.0(3.4)
Education, mean (SD)	17.0(2)	16.3(3.1)	16.0(2.8)	16.4(2.5)
MMSE score, mean (SD)	29.5(0.6)	28.8(0.9)	20.3(2.5)	26.2(4.6)
Women, No. (%)	3 (75)	2 (50)	2 (50)	7 (58)
APOE4 positive, No. (%)	2 (50)	3 (75)	3 (75)	8 (75)
Family history of dementia, No. (%)	0 (0)	2 (50)	1 (25)	3 (25)

Abbreviations: MMSE, Mini-Mental State Examination; AD, Alzheimer's disease; MCI, mild cognitive impairment; SD standard deviation

Figure 1

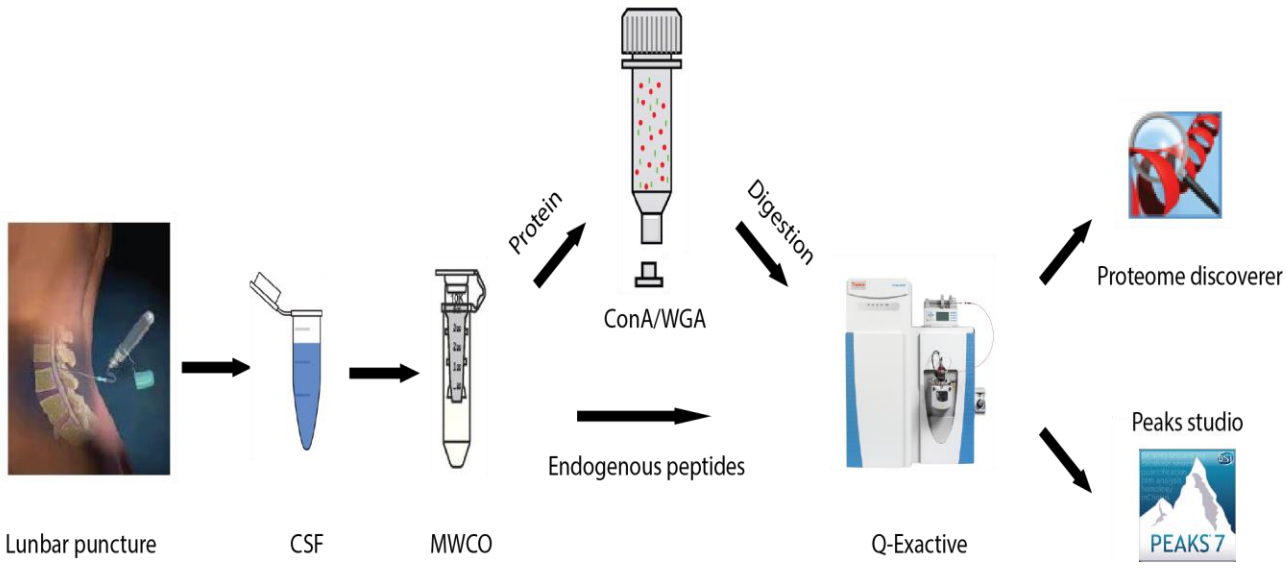

Protein
Digestion
ConA/WGA
Proteome discoverer
Peaks studio
Endogenous peptides
Lunbar puncture
CSF
MWCO
Q-Exactive
PEAKS 7

Figure 2

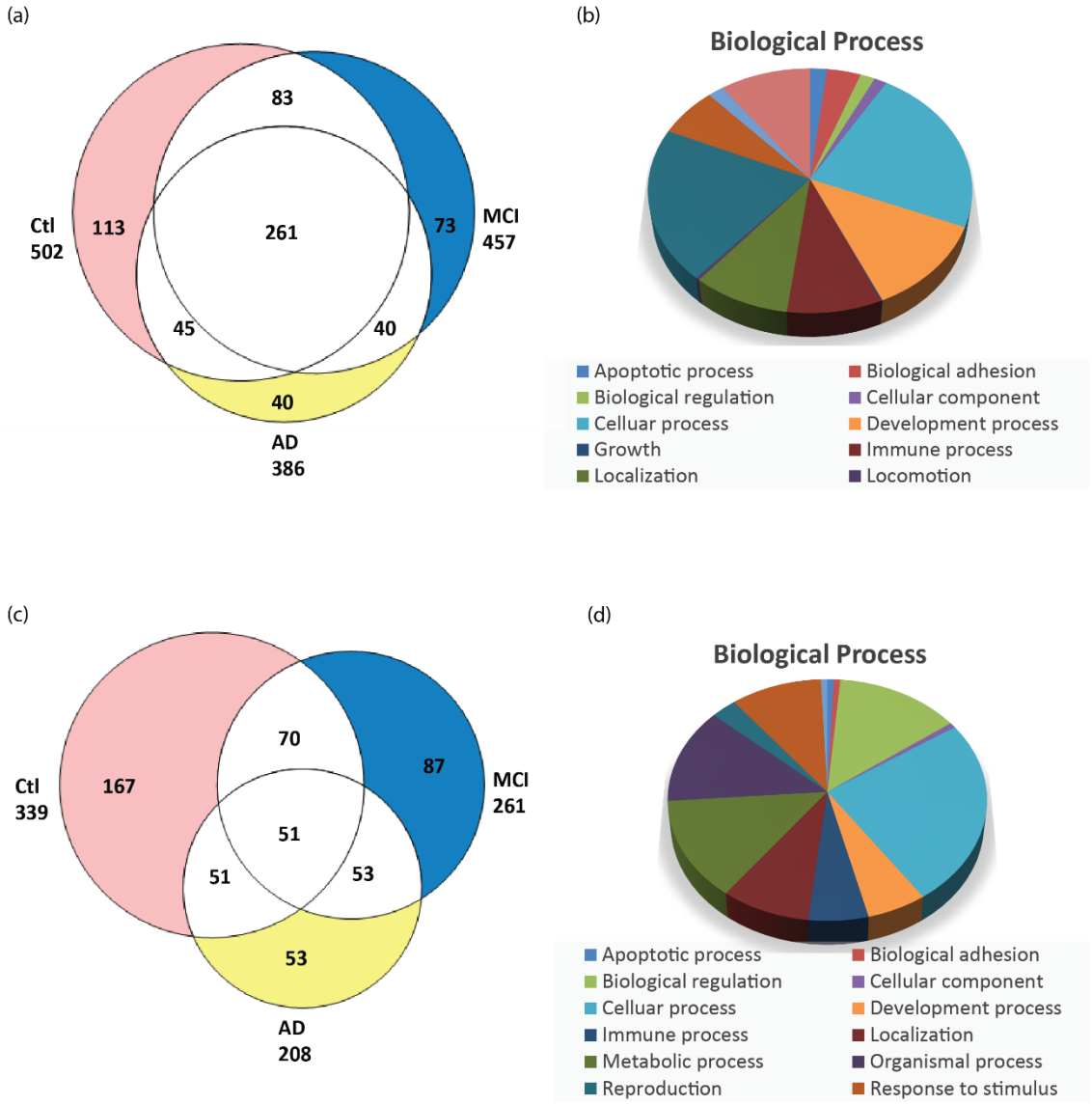

(a)
Ctl
502
113
83
261
73
MCI
457
45
40
40
AD
386
(b)
Biological Process
Apoptotic process
Biological adhesion
Biological regulation
Cellular component
Celluar process
Development process
Growth
Immune process
Localization
Locomotion
(c)
Ctl
339
167
70
87
MCI
261
51
51
53
53
AD
208
(d)
Biological Process
Apoptotic process
Biological adhesion
Biological regulation
Cellular component
Celluar process
Development process
Immune process
Localization
Metabolic process
Organismal process
Reproduction
Response to stimulus

Table 2 ProSAAS peptides identified in CSF

Peptide	-10logP	Mass	PPM	*m/z*	RT	PTM
DHDVGSELPPEGVLGA	51.23	1590.753	5.8	796.3882	35.2	
DHDVGSELPPEGVLG	46.48	1519.716	-1.1	760.8641	33.67	
HDVGSELPPEGVLGA	44.42	1475.7256	-0.9	738.8694	33.56	
DHDVGSELPPEGVL	41.55	1462.694	-2.4	732.3525	36.48	
SPPLAETGAPR	40.13	1094.572	-0.5	548.293	19.56	
SPPLAETG	38.93	770.381	-1.6	771.387	18.1	
AADHDVGSELPPEGVLGA	38.49	1732.8268	-1.4	867.4194	34.41	
DHDVGSELPPEG	37.35	1250.5415	-1.1	626.2773	22.56	
HDVGSELPPEGVLG	37.08	1404.6885	-2	703.3501	32.33	
HDVGSELPPEGVL	35.62	1347.667	-1	674.8401	33.62	
ADHDVGSELPPEGVLGA	35.46	1661.7897	-1	831.9013	34.05	
TPAPQVPARRLLPP	34.97	1511.8936	-1.2	504.9712	30.02	
GLSAASPPLAETGAPR	34.92	1493.7837	0.7	747.8997	27.91	
VGSELPPEGVLGA	34.78	1223.6398	0.3	612.8273	34.71	
HDVGSELPPEG	34.38	1135.5145	1.6	568.7654	21.13	
DHDVGSELPPEGVLGAL	31.99	1703.8365	-0.5	852.9251	43.02	
DVGSELPPEGVLGA	31.33	1338.6666	-0.7	670.3401	37.44	
SAASPPLAETGAPR	30.54	1323.6782	1.5	662.8474	21.46	
SAASPPLAETGAPRRF	28.46	1626.8478	-0.8	543.2894	26.92	
AADHDVGSELPPEGVLG	27.04	1661.7897	1.5	831.9034	33.21	
ETPAPQVPARR	26.73	1220.6625	-1.2	407.8943	16.7	
SPPLAETGAPRRF	26.58	1397.7415	-0.6	466.9208	25.54	
ETPAPQVPARRLLPP	26.44	1640.9362	1.4	547.9868	31.03	
DVGSELPPEG	26.23	998.4556	-1.4	500.2344	25.02	
DVGSELPPEGVL	24.73	1210.608	0.2	606.3114	37.9	
DVGSELPPEGVLG	24.32	1267.6295	1.7	634.8231	36.66	
ADHDVGSELPPEGVLG	24.27	1590.7526	-3.3	796.3809	33.38	
DHDVGSELPPEGVL	23.95	1461.71	-1.3	731.8613	33.14	amidation
LSAASPPLAETGAPRRF	21.23	1739.9318	-0.3	580.9844	29.46	
DHDVGSELPPEGVLGALL	21.18	1816.9207	1.1	909.4686	49.79	
DHDVGSELPPEGVLGALLRV	21	2072.0901	-0.5	691.7036	51.84	
ADSEGVAAPRRLRRAADHDV--GSELPPEG	20.34	2926.4756	-8.6	976.4908	30.44	amidation
LPPEG	19.5	511.2642	-2.1	512.2704	14.01	
PPEGVLGAL	19.14	851.4752	-0.2	426.7448	31.34	
ETPAPQVPA	18	908.4603	-1.4	455.2368	22.48	
F(+42.01)RRSVP	17.92	802.4449	-3.8	402.2282	22.43	acetylation
SVPRGEAAGAVQELARA	17.31	1680.8906	1.6	561.3051	30.01	
LETPAPQVPARRLLPP	16.26	1754.0203	0	585.6807	32.7	
LSAASPPLAETGAPR	16.22	1436.7623	-3.4	719.386	25.52	
LAETGAPR	15.75	813.4344	-4.7	407.7226	14.34	
DHDVGSELPPEGV	15.13	1349.6099	-4.2	675.8094	27.96	
LETPAPQVPA	15.11	1021.5444	0	511.7795	26.94	

Figure 3

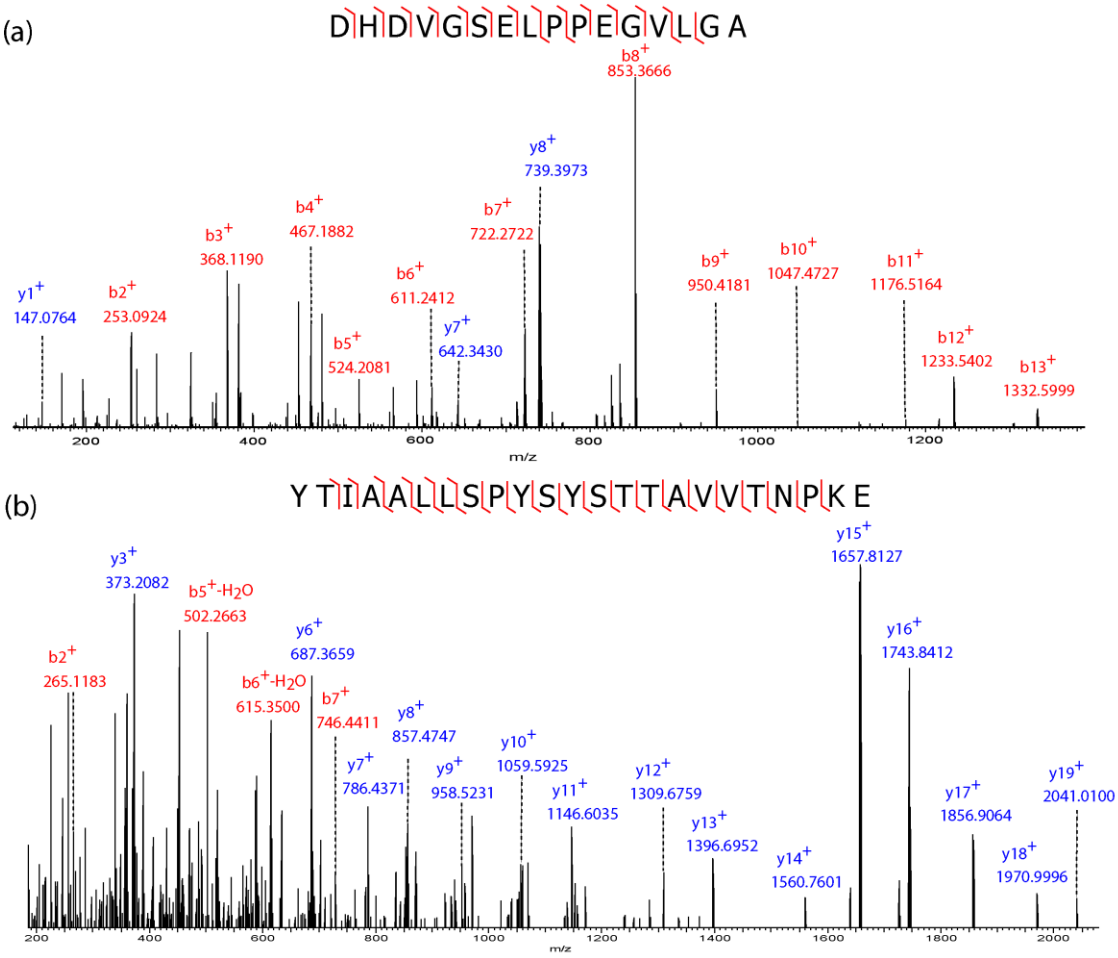
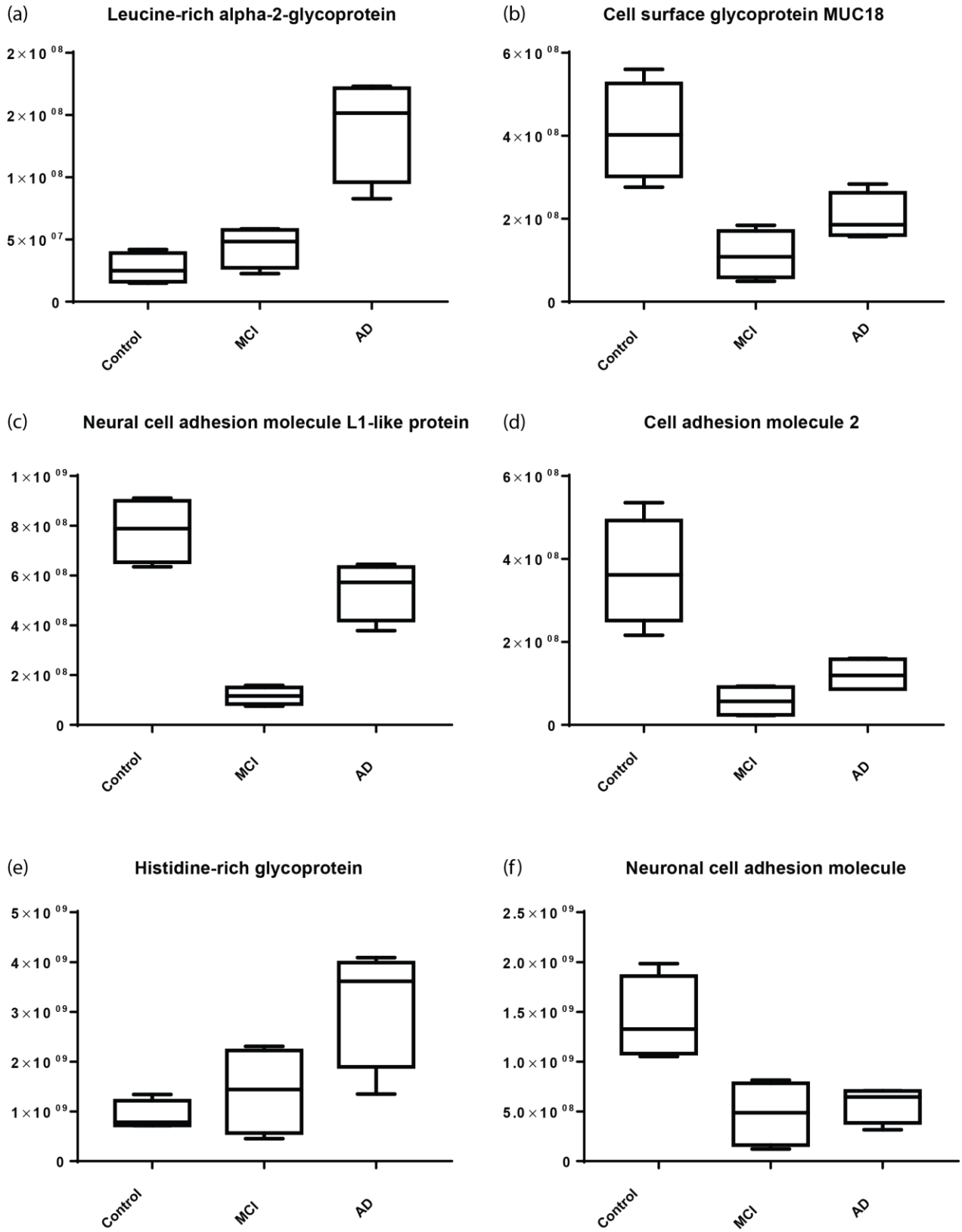
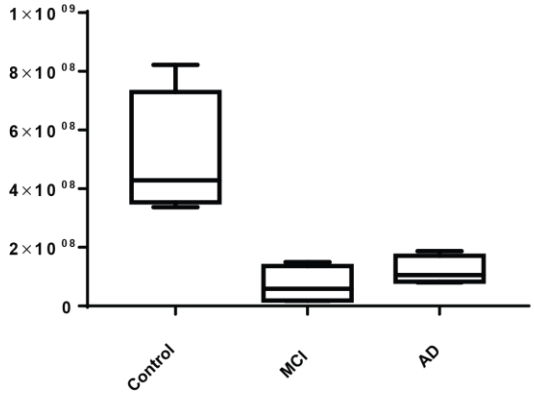

(a)
D H D V G S E L P P E G V L G A
b8+ 853.3666
y8+ 739.3973
b4+ 467.1882
b3+ 368.1190
b7+ 722.2722
b6+ 611.2412
y1+ 147.0764
b2+ 253.0924
b5+ 524.2081
y7+ 642.3430
b9+ 950.4181
b10+ 1047.4727
b11+ 1176.5164
b12+ 1233.5402
b13+ 1332.5999
m/z
(b)
Y T I A A L L S P Y S Y S T T A V V T N P K E
y15+ 1657.8127
y3+ 373.2082
b5+-H2O 502.2663
y6+ 687.3659
y16+ 1743.8412
b2+ 265.1183
b6+-H2O 615.3500
b7+ 746.4411
y8+ 857.4747
y7+ 786.4371
y9+ 958.5231
y10+ 1059.5925
y12+ 1309.6759
y11+ 1146.6035
y17+ 1856.9064
y19+ 2041.0100
y13+ 1396.6952
y14+ 1560.7601
y18+ 1970.9996
m/z

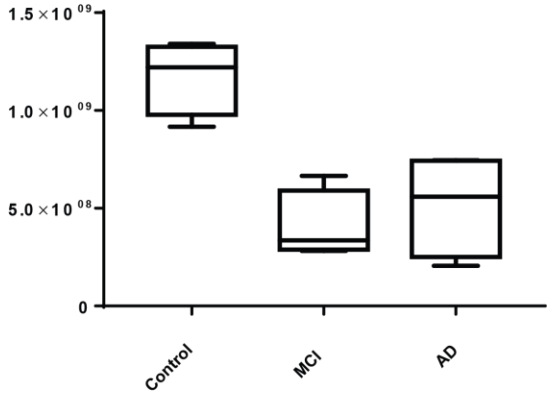
Figure 4


(a)
Leucine-rich alpha-2-glycoprotein
2×10 08
2×10 08
1×10 08
5×10 07
0
Control
MCI
AD
(b)
Cell surface glycoprotein MUC18
6×10 08
4×10 08
2×10 08
0
Control
MCI
AD
(c)
Neural cell adhesion molecule L1-like protein
1×10 09
8×10 08
6×10 08
4×10 08
2×10 08
0
Control
MCI
AD
(d)
Cell adhesion molecule 2
6×10 08
4×10 08
2×10 08
0
Control
MCI
AD
(e)
Histidine-rich glycoprotein
5×10 09
4×10 09
3×10 09
2×10 09
1×10 09
0
Control
MCI
AD
(f)
Neuronal cell adhesion molecule
2.5×10 09
2.0×10 09
1.5×10 09
1.0×10 09
5.0×10 08
0
Control
MCI
AD

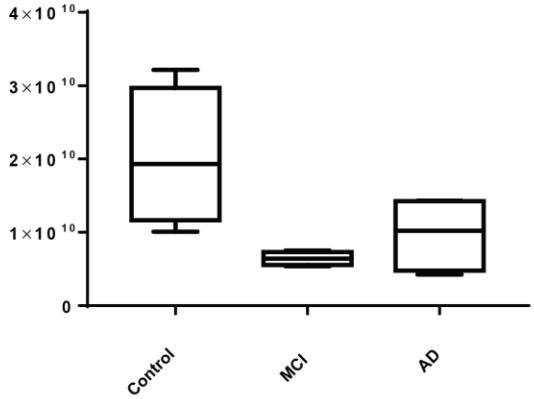
(g) Neuronal pentraxin-1



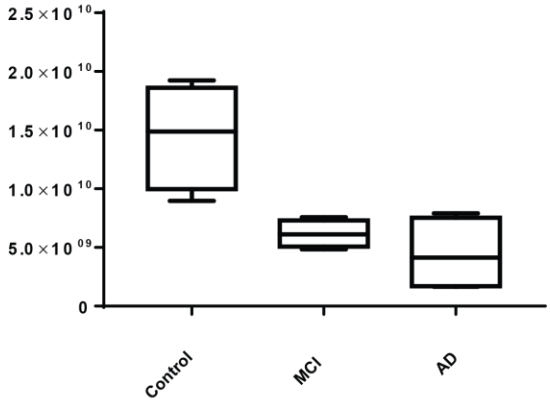
(h) Neuronal growth regulator 1



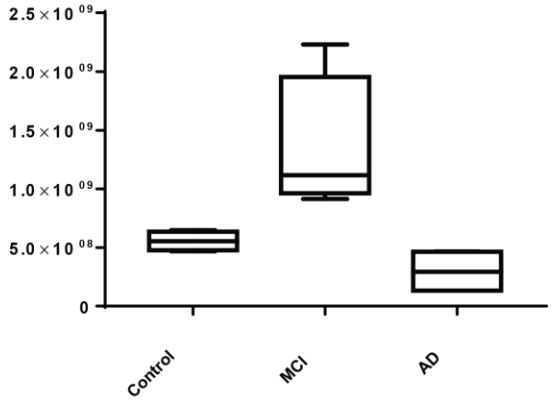
(i) Dickkopf-related protein 3



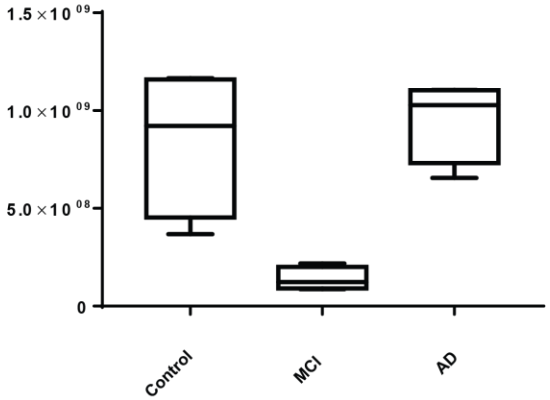
(j) Apolipoprotein E



(k) Transthyretin



(l) Alpha-2-HS-glycoprotein



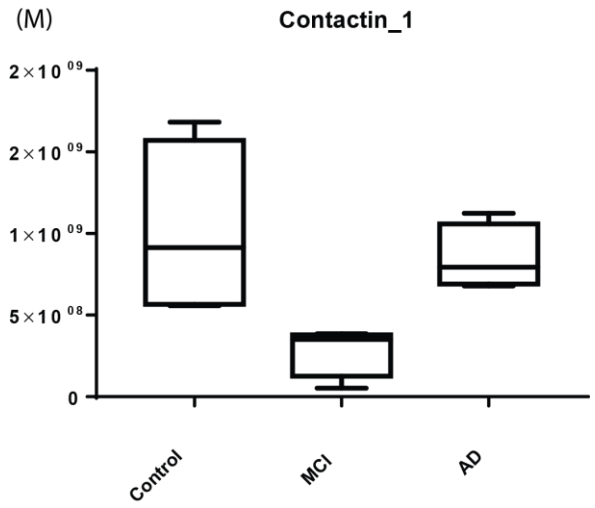

(M)
Contactin_1
2×10 09
2×10 09
1×10 09
5×10 08
0
Control
MCI
AD

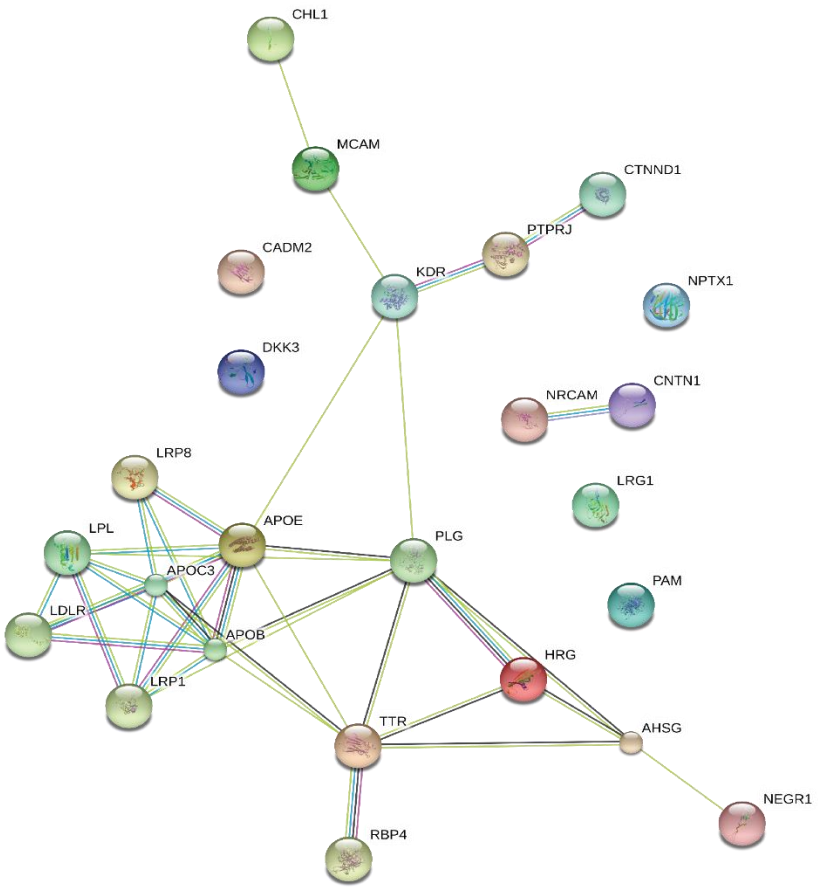
Table 3. Proteins showed differential expression in CSF among healthy control, MCI and AD

Protein ID	Mean Peak Area			ANOVA p	Tukey's test		
	Ctl	MCI	AD		Ctl vs MCI	Ctl vs AD	MCI vs AD
Leucine-rich alpha-2-glycoprotein	2.67E+07	4.45E+07	1.40E+08	0.004	NS	<0.01	<0.01
Cell surface glycoprotein MUC18	4.10E+08	1.12E+08	2.03E+08	0.019	<0.01	<0.05	NS
Neural cell adhesion molecule L1-like protein	7.81E+08	1.17E+08	5.42E+08	0.0026	<0.01	<0.05	<0.01
Cell adhesion molecule 2	3.69E+08	5.74E+07	1.21E+08	0.001	<0.01	<0.01	NS
Histidine-rich glycoprotein	9.05E+08	1.41E+09	3.17E+09	0.014	NS	<0.05	<0.05
Neuronal cell adhesion molecule	1.42E+09	4.77E+08	5.78E+08	0.048	<0.01	<0.05	NS
Neuronal pentraxin-1	5.04E+08	7.12E+07	1.20E+08	0.0025	<0.01	<0.01	NS
Neuronal growth regulator 1	1.17E+09	4.05E+08	5.18E+08	0.0013	<0.01	<0.01	NS
Dickkopf-related protein 3	2.02E+10	6.43E+09	9.75E+09	0.0293	<0.05	NS	NS
Apolipoprotein E	1.45E+10	6.17E+09	4.46E+09	0.039	<0.05	<0.01	NS
Transthyretin	5.55E+08	1.34E+09	2.96E+08	0.0074	<0.05	<0.01	<0.01
Alpha-2-HS-glycoprotein	8.45E+08	1.38E+08	9.55E+08	0.0027	<0.01	NS	<0.01
Contactin-1	1.02E+09	2.86E+08	8.48E+08	0.0362	<0.05	NS	NS

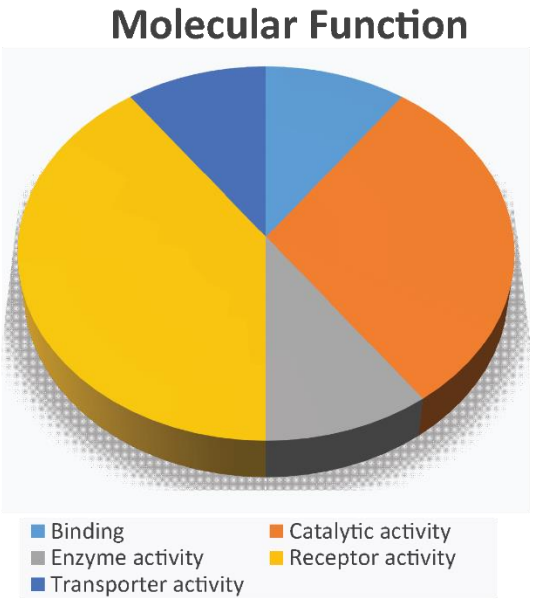
NS: not significant; Ctl: control

Figure 5

(a)


CHL1
MCAM
CTNND1
PTPRJ
CADM2
KDR
NPTX1
DKK3
NRCAM
CNTN1
LRP8
LRG1
LPL
APOE
PLG
APOC3
PAM
LDLR
APOB
HRG
LRP1
TTR
AHSG
NEGR1
RBP4

(b)


Molecular Function
Binding
Catalytic activity
Enzyme activity
Receptor activity
Transporter activity

REFERENCES:

1. Sosa-Ortiz, A. L.; Acosta-Castillo, I.; Prince, M. J., Epidemiology of dementias and Alzheimer's disease. *Arch Med Res* **2012,** *43* (8), 600-8.
2. Blennow, K.; de Leon, M. J.; Zetterberg, H., Alzheimer's disease. *Lancet* **2006,** *368* (9533), 387-403.
3. McKhann, G. M.; Knopman, D. S.; Chertkow, H.; Hyman, B. T.; Jack, C. R., Jr.; Kawas, C. H.; Klunk, W. E.; Koroshetz, W. J.; Manly, J. J.; Mayeux, R.; Mohs, R. C.; Morris, J. C.; Rossor, M. N.; Scheltens, P.; Carrillo, M. C.; Thies, B.; Weintraub, S.; Phelps, C. H., The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* **2011,** *7* (3), 263-9.
4. Olsson, A.; Vanderstichele, H.; Andreasen, N.; De Meyer, G.; Wallin, A.; Holmberg, B.; Rosengren, L.; Vanmechelen, E.; Blennow, K., Simultaneous measurement of beta-amyloid(1-42), total tau, and phosphorylated tau (Thr181) in cerebrospinal fluid by the xMAP technology. *Clin Chem* **2005,** *51* (2), 336-45.
5. Hampel, H.; Burger, K.; Teipel, S. J.; Bokde, A. L.; Zetterberg, H.; Blennow, K., Core candidate neurochemical and imaging biomarkers of Alzheimer's disease. *Alzheimers Dement* **2008,** *4* (1), 38-48.
6. Jack, C. R., Jr.; Albert, M. S.; Knopman, D. S.; McKhann, G. M.; Sperling, R. A.; Carrillo, M. C.; Thies, B.; Phelps, C. H., Introduction to the recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* **2011,** *7* (3), 257-62.
7. Kang, J. H.; Korecka, M.; Toledo, J. B.; Trojanowski, J. Q.; Shaw, L. M., Clinical utility and analytical challenges in measurement of cerebrospinal fluid amyloid-beta(1-42) and tau proteins as Alzheimer disease biomarkers. *Clin Chem* **2013,** *59* (6), 903-16.
8. Doody, R. S.; Thomas, R. G.; Farlow, M.; Iwatsubo, T.; Vellas, B.; Joffe, S.; Kieburtz, K.; Raman, R.; Sun, X.; Aisen, P. S.; Siemers, E.; Liu-Seifert, H.; Mohs, R.; Alzheimer's Disease Cooperative Study Steering, C.; Solanezumab Study, G., Phase 3 trials of solanezumab for mild-to-moderate Alzheimer's disease. *N Engl J Med* **2014,** *370* (4), 311-21.
9. Doody, R. S.; Farlow, M.; Aisen, P. S.; Alzheimer's Disease Cooperative Study Data, A.; Publication, C., Phase 3 trials of solanezumab and bapineuzumab for Alzheimer's disease. *N Engl J Med* **2014,** *370* (15), 1460.
10. Zhang, J.; Goodlett, D. R.; Montine, T. J., Proteomic biomarker discovery in cerebrospinal fluid for neurodegenerative diseases. *J Alzheimers Dis* **2005,** *8* (4), 377-86.
11. Blennow, K.; Hampel, H.; Weiner, M.; Zetterberg, H., Cerebrospinal fluid and plasma biomarkers in Alzheimer disease. *Nat Rev Neurol* **2010,** *6* (3), 131-44.
12. Holtta, M.; Zetterberg, H.; Mirgorodskaya, E.; Mattsson, N.; Blennow, K.; Gobom, J., Peptidome analysis of cerebrospinal fluid by LC-MALDI MS. *PLoS One* **2012,** *7* (8), e42555.
13. Johanson, C. E.; Duncan, J. A., 3rd; Klinge, P. M.; Brinker, T.; Stopa, E. G.; Silverberg, G. D., Multiplicity of cerebrospinal fluid functions: New challenges in health and disease. *Cerebrospinal Fluid Res* **2008,** *5*, 10.
14. Raymackers, J.; Daniels, A.; De Brabandere, V.; Missiaen, C.; Dauwe, M.; Verhaert, P.; Vanmechelen, E.; Meheus, L., Identification of two-dimensionally separated human cerebrospinal fluid proteins by N-terminal sequencing, matrix-assisted laser desorption/ionization--mass spectrometry, nanoliquid chromatography-electrospray ionization-time of flight-mass spectrometry, and tandem mass spectrometry. *Electrophoresis* **2000,** *21* (11), 2266-83.

15. Sickmann, A.; Dormeyer, W.; Wortelkamp, S.; Woitalla, D.; Kuhn, W.; Meyer, H. E., Identification of proteins from human cerebrospinal fluid, separated by two-dimensional polyacrylamide gel electrophoresis. *Electrophoresis* **2000,** *21* (13), 2721-8.
16. Sickmann, A.; Dormeyer, W.; Wortelkamp, S.; Woitalla, D.; Kuhn, W.; Meyer, H. E., Towards a high resolution separation of human cerebrospinal fluid. *J Chromatogr B Analyt Technol Biomed Life Sci* **2002,** *771* (1-2), 167-96.
17. Yuan, X.; Russell, T.; Wood, G.; Desiderio, D. M., Analysis of the human lumbar cerebrospinal fluid proteome. *Electrophoresis* **2002,** *23* (7-8), 1185-96.
18. Maccarrone, G.; Milfay, D.; Birg, I.; Rosenhagen, M.; Holsboer, F.; Grimm, R.; Bailey, J.; Zolotarjova, N.; Turck, C. W., Mining the human cerebrospinal fluid proteome by immunodepletion and shotgun mass spectrometry. *Electrophoresis* **2004,** *25* (14), 2402-12.
19. Pan, S.; Zhu, D.; Quinn, J. F.; Peskind, E. R.; Montine, T. J.; Lin, B.; Goodlett, D. R.; Taylor, G.; Eng, J.; Zhang, J., A combined dataset of human cerebrospinal fluid proteins identified by multi-dimensional chromatography and tandem mass spectrometry. *Proteomics* **2007,** *7* (3), 469-73.
20. Schutzer, S. E.; Liu, T.; Natelson, B. H.; Angel, T. E.; Schepmoes, A. A.; Purvine, S. O.; Hixson, K. K.; Lipton, M. S.; Camp, D. G.; Coyle, P. K.; Smith, R. D.; Bergquist, J., Establishing the proteome of normal human cerebrospinal fluid. *PLoS One* **2010,** *5* (6), e10980.
21. Wenner, B. R.; Lovell, M. A.; Lynn, B. C., Proteomic analysis of human ventricular cerebrospinal fluid from neurologically normal, elderly subjects using two-dimensional LC-MS/MS. *J Proteome Res* **2004,** *3* (1), 97-103.
22. Xu, J.; Chen, J.; Peskind, E. R.; Jin, J.; Eng, J.; Pan, C.; Montine, T. J.; Goodlett, D. R.; Zhang, J., Characterization of proteome of human cerebrospinal fluid. *Int Rev Neurobiol* **2006,** *73*, 29-98.
23. Varki, A.; Lowe, J. B., Biological Roles of Glycans. In *Essentials of Glycobiology*, 2nd ed.; Varki, A.; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Bertozzi, C. R.; Hart, G. W.; Etzler, M. E., Eds. Cold Spring Harbor (NY), 2009.
24. Di Domenico, F.; Owen, J. B.; Sultana, R.; Sowell, R. A.; Perluigi, M.; Cini, C.; Cai, J.; Pierce, W. M.; Butterfield, D. A., The wheat germ agglutinin-fractionated proteome of subjects with Alzheimer's disease and mild cognitive impairment hippocampus and inferior parietal lobule: Implications for disease pathogenesis and progression. *J Neurosci Res* **2010,** *88* (16), 3566-77.
25. Maguire, T. M.; Gillian, A. M.; O'Mahony, D.; Coughlan, C. M.; Dennihan, A.; Breen, K. C., A decrease in serum sialyltransferase levels in Alzheimer's disease. *Neurobiol Aging* **1994,** *15* (1), 99-102.
26. Fodero, L. R.; Saez-Valero, J.; Barquero, M. S.; Marcos, A.; McLean, C. A.; Small, D. H., Wheat germ agglutinin-binding glycoproteins are decreased in Alzheimer's disease cerebrospinal fluid. *J Neurochem* **2001,** *79* (5), 1022-6.
27. Butterfield, D. A.; Sultana, R., Redox proteomics identification of oxidatively modified brain proteins in Alzheimer's disease and mild cognitive impairment: insights into the progression of this dementing disorder. *J Alzheimers Dis* **2007,** *12* (1), 61-72.
28. Meguro, K., Clinical features of mild cognitive impairment and dementia in a community: an update of the Osaki-Tajiri Project. *Tohoku J Exp Med* **2008,** *215* (2), 125-31.
29. Visser, P. J.; Verhey, F. R., Mild cognitive impairment as predictor for Alzheimer's disease in clinical practice: effect of age and diagnostic criteria. *Psychol Med* **2008,** *38* (1), 113-22.
30. Petersen, R. C.; Doody, R.; Kurz, A.; Mohs, R. C.; Morris, J. C.; Rabins, P. V.; Ritchie, K.; Rossor, M.; Thal, L.; Winblad, B., Current concepts in mild cognitive impairment. *Arch Neurol* **2001,** *58* (12), 1985-92.
31. Albert, M. S.; DeKosky, S. T.; Dickson, D.; Dubois, B.; Feldman, H. H.; Fox, N. C.; Gamst, A.; Holtzman, D. M.; Jagust, W. J.; Petersen, R. C.; Snyder, P. J.; Carrillo, M. C.; Thies, B.; Phelps, C. H., The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from

the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* **2011,** *7* (3), 270-9.

32. Almeida, R. P.; Schultz, S. A.; Austin, B. P.; Boots, E. A.; Dowling, N. M.; Gleason, C. E.; Bendlin, B. B.; Sager, M. A.; Hermann, B. P.; Zetterberg, H.; Carlsson, C. M.; Johnson, S. C.; Asthana, S.; Okonkwo, O. C., Effect of Cognitive Reserve on Age-Related Changes in Cerebrospinal Fluid Biomarkers of Alzheimer Disease. *JAMA Neurol* **2015,** *72* (6), 699-706.

33. Okonkwo, O. C.; Xu, G.; Oh, J. M.; Dowling, N. M.; Carlsson, C. M.; Gallagher, C. L.; Birdsill, A. C.; Palotti, M.; Wharton, W.; Hermann, B. P.; LaRue, A.; Bendlin, B. B.; Rowley, H. A.; Asthana, S.; Sager, M. A.; Johnson, S. C., Cerebral blood flow is diminished in asymptomatic middle-aged adults with maternal history of Alzheimer's disease. *Cereb Cortex* **2014,** *24* (4), 978-88.

34. Roe, C. M.; Mintun, M. A.; D'Angelo, G.; Xiong, C.; Grant, E. A.; Morris, J. C., Alzheimer disease and cognitive reserve: variation of education effect with carbon 11-labeled Pittsburgh Compound B uptake. *Arch Neurol* **2008,** *65* (11), 1467-71.

35. Palmqvist, S.; Zetterberg, H.; Blennow, K.; Vestberg, S.; Andreasson, U.; Brooks, D. J.; Owenius, R.; Hagerstrom, D.; Wollmer, P.; Minthon, L.; Hansson, O., Accuracy of brain amyloid detection in clinical practice using cerebrospinal fluid beta-amyloid 42: a cross-validation study against amyloid positron emission tomography. *JAMA Neurol* **2014,** *71* (10), 1282-9.

36. Cunningham, R.; Wang, J.; Wellner, D.; Li, L., Investigation and reduction of sub-microgram peptide loss using molecular weight cut-off fractionation prior to mass spectrometric analysis. *J Mass Spectrom* **2012,** *47* (10), 1327-32.

37. Fouillen, L.; Petruzziello, F.; Veit, J.; Bhattacharyya, A.; Kretz, R.; Rainer, G.; Zhang, X., Neuropeptide alterations in the tree shrew hypothalamus during volatile anesthesia. *J Proteomics* **2013,** *80*, 311-9.

38. Zougman, A.; Pilch, B.; Podtelejnikov, A.; Kiehntopf, M.; Schnabel, C.; Kumar, C.; Mann, M., Integrated analysis of the cerebrospinal fluid peptidome and proteome. *J Proteome Res* **2008,** *7* (1), 386-99.

39. Hook, V.; Funkelstein, L.; Lu, D.; Bark, S.; Wegrzyn, J.; Hwang, S. R., Proteases for processing proneuropeptides into peptide neurotransmitters and hormones. *Annu Rev Pharmacol Toxicol* **2008,** *48*, 393-423.

40. Rholam, M.; Fahy, C., Processing of peptide and hormone precursors at the dibasic cleavage sites. *Cell Mol Life Sci* **2009,** *66* (13), 2075-91.

41. Savitski, M. M.; Falth, M., Peptide fragmentation and phospho-site detection. *Expert Rev Proteomics* **2007,** *4* (4), 445-6.

42. Zhang, J.; Xin, L.; Shan, B.; Chen, W.; Xie, M.; Yuen, D.; Zhang, W.; Zhang, Z.; Lajoie, G. A.; Ma, B., PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. *Mol Cell Proteomics* **2012,** *11* (4), M111 010587.

43. Lanoue, E.; Day, R., Coexpression of proprotein convertase SPC3 and the neuroendocrine precursor proSAAS. *Endocrinology* **2001,** *142* (9), 4141-9.

44. Morgan, D. J.; Mzhavia, N.; Peng, B.; Pan, H.; Devi, L. A.; Pintar, J. E., Embryonic gene expression and pro-protein processing of proSAAS during rodent development. *J Neurochem* **2005,** *93* (6), 1454-62.

45. Fricker, L. D.; McKinzie, A. A.; Sun, J.; Curran, E.; Qian, Y.; Yan, L.; Patterson, S. D.; Courchesne, P. L.; Richards, B.; Levin, N.; Mzhavia, N.; Devi, L. A.; Douglass, J., Identification and characterization of proSAAS, a granin-like neuroendocrine peptide precursor that inhibits prohormone processing. *J Neurosci* **2000,** *20* (2), 639-48.

46. Qian, Y.; Devi, L. A.; Mzhavia, N.; Munzer, S.; Seidah, N. G.; Fricker, L. D., The C-terminal region of proSAAS is a potent inhibitor of prohormone convertase 1. *J Biol Chem* **2000,** *275* (31), 23596-601.

47. Feng, Y.; Reznik, S. E.; Fricker, L. D., Distribution of proSAAS-derived peptides in rat neuroendocrine tissues. *Neuroscience* **2001,** *105* (2), 469-78.
48. Atkins, N., Jr.; Mitchell, J. W.; Romanova, E. V.; Morgan, D. J.; Cominski, T. P.; Ecker, J. L.; Pintar, J. E.; Sweedler, J. V.; Gillette, M. U., Circadian integration of glutamatergic signals by little SAAS in novel suprachiasmatic circuits. *PLoS One* **2010,** *5* (9), e12612.
49. Hatcher, N. G.; Atkins, N., Jr.; Annangudi, S. P.; Forbes, A. J.; Kelleher, N. L.; Gillette, M. U.; Sweedler, J. V., Mass spectrometry-based discovery of circadian peptides. *Proc Natl Acad Sci U S A* **2008,** *105* (34), 12527-32.
50. Wardman, J. H.; Berezniuk, I.; Di, S.; Tasker, J. G.; Fricker, L. D., ProSAAS-derived peptides are colocalized with neuropeptide Y and function as neuropeptides in the regulation of food intake. *PLoS One* **2011,** *6* (12), e28152.
51. Kikuchi, K.; Arawaka, S.; Koyama, S.; Kimura, H.; Ren, C. H.; Wada, M.; Kawanami, T.; Kurita, K.; Daimon, M.; Kawakatsu, S.; Kadoya, T.; Goto, K.; Kato, T., An N-terminal fragment of ProSAAS (a granin-like neuroendocrine peptide precursor) is associated with tau inclusions in Pick's disease. *Biochem Biophys Res Commun* **2003,** *308* (3), 646-54.
52. Wada, M.; Ren, C. H.; Koyama, S.; Arawaka, S.; Kawakatsu, S.; Kimura, H.; Nagasawa, H.; Kawanami, T.; Kurita, K.; Daimon, M.; Hirano, A.; Kato, T., A human granin-like neuroendocrine peptide precursor (proSAAS) immunoreactivity in tau inclusions of Alzheimer's disease and parkinsonism-dementia complex on Guam. *Neurosci Lett* **2004,** *356* (1), 49-52.
53. Davidsson, P.; Sjogren, M.; Andreasen, N.; Lindbjer, M.; Nilsson, C. L.; Westman-Brinkmalm, A.; Blennow, K., Studies of the pathophysiological mechanisms in frontotemporal dementia by proteome analysis of CSF proteins. *Brain Res Mol Brain Res* **2002,** *109* (1-2), 128-33.
54. Finehout, E. J.; Franck, Z.; Choe, L. H.; Relkin, N.; Lee, K. H., Cerebrospinal fluid proteomic biomarkers for Alzheimer's disease. *Ann Neurol* **2007,** *61* (2), 120-9.
55. Hoshino, A.; Helwig, M.; Rezaei, S.; Berridge, C.; Eriksen, J. L.; Lindberg, I., A novel function for proSAAS as an amyloid anti-aggregant in Alzheimer's disease. *J Neurochem* **2014,** *128* (3), 419-30.
56. Li, X.; Buxbaum, J. N., Transthyretin and the brain re-visited: is neuronal synthesis of transthyretin protective in Alzheimer's disease? *Mol Neurodegener* **2011,** *6*, 79.
57. Schwarzman, A. L.; Goldgaber, D., Interaction of transthyretin with amyloid beta-protein: binding and inhibition of amyloid formation. *Ciba Found Symp* **1996,** *199*, 146-60; discussion 160-4.
58. Stein, T. D.; Anders, N. J.; DeCarli, C.; Chan, S. L.; Mattson, M. P.; Johnson, J. A., Neutralization of transthyretin reverses the neuroprotective effects of secreted amyloid precursor protein (APP) in APPSW mice resulting in tau phosphorylation and loss of hippocampal neurons: support for the amyloid hypothesis. *J Neurosci* **2004,** *24* (35), 7707-17.
59. Li, X.; Masliah, E.; Reixach, N.; Buxbaum, J. N., Neuronal production of transthyretin in human and murine Alzheimer's disease: is it protective? *J Neurosci* **2011,** *31* (35), 12483-90.
60. Serot, J. M.; Christmann, D.; Dubost, T.; Couturier, M., Cerebrospinal fluid transthyretin: aging and late onset Alzheimer's disease. *J Neurol Neurosurg Psychiatry* **1997,** *63* (4), 506-8.
61. Hansson, S. F.; Andreasson, U.; Wall, M.; Skoog, I.; Andreasen, N.; Wallin, A.; Zetterberg, H.; Blennow, K., Reduced levels of amyloid-beta-binding proteins in cerebrospinal fluid from Alzheimer's disease patients. *J Alzheimers Dis* **2009,** *16* (2), 389-97.
62. Gloeckner, S. F.; Meyne, F.; Wagner, F.; Heinemann, U.; Krasnianski, A.; Meissner, B.; Zerr, I., Quantitative analysis of transthyretin, tau and amyloid-beta in patients with dementia. *J Alzheimers Dis* **2008,** *14* (1), 17-25.

63. Davidsson, P.; Westman-Brinkmalm, A.; Nilsson, C. L.; Lindbjer, M.; Paulson, L.; Andreasen, N.; Sjogren, M.; Blennow, K., Proteome analysis of cerebrospinal fluid proteins in Alzheimer patients. *Neuroreport* **2002,** *13* (5), 611-5.
64. Korolainen, M. A.; Nyman, T. A.; Nyyssonen, P.; Hartikainen, E. S.; Pirttila, T., Multiplexed proteomic analysis of oxidation and concentrations of cerebrospinal fluid proteins in Alzheimer disease. *Clin Chem* **2007,** *53* (4), 657-65.
65. Puchades, M.; Hansson, S. F.; Nilsson, C. L.; Andreasen, N.; Blennow, K.; Davidsson, P., Proteomic studies of potential cerebrospinal fluid protein markers for Alzheimer's disease. *Brain Res Mol Brain Res* **2003,** *118* (1-2), 140-6.
66. Castano, E. M.; Roher, A. E.; Esh, C. L.; Kokjohn, T. A.; Beach, T., Comparative proteomics of cerebrospinal fluid in neuropathologically-confirmed Alzheimer's disease and non-demented elderly subjects. *Neurol Res* **2006,** *28* (2), 155-63.
67. Link, C. D., Expression of human beta-amyloid peptide in transgenic Caenorhabditis elegans. *Proc Natl Acad Sci U S A* **1995,** *92* (20), 9368-72.
68. Stein, T. D.; Johnson, J. A., Lack of neurodegeneration in transgenic mice overexpressing mutant amyloid precursor protein is associated with increased levels of transthyretin and the activation of cell survival pathways. *J Neurosci* **2002,** *22* (17), 7380-8.
69. Wu, Z. L.; Ciallella, J. R.; Flood, D. G.; O'Kane, T. M.; Bozyczko-Coyne, D.; Savage, M. J., Comparative analysis of cortical gene expression in mouse models of Alzheimer's disease. *Neurobiol Aging* **2006,** *27* (3), 377-86.
70. Buxbaum, J. N.; Ye, Z.; Reixach, N.; Friske, L.; Levy, C.; Das, P.; Golde, T.; Masliah, E.; Roberts, A. R.; Bartfai, T., Transthyretin protects Alzheimer's mice from the behavioral and biochemical effects of Abeta toxicity. *Proc Natl Acad Sci U S A* **2008,** *105* (7), 2681-6.
71. Schwarzman, A. L.; Tsiper, M.; Wente, H.; Wang, A.; Vitek, M. P.; Vasiliev, V.; Goldgaber, D., Amyloidogenic and anti-amyloidogenic properties of recombinant transthyretin variants. *Amyloid* **2004,** *11* (1), 1-9.
72. Giunta, S.; Valli, M. B.; Galeazzi, R.; Fattoretti, P.; Corder, E. H.; Galeazzi, L., Transthyretin inhibition of amyloid beta aggregation and toxicity. *Clin Biochem* **2005,** *38* (12), 1112-9.
73. Liu, L.; Murphy, R. M., Kinetics of inhibition of beta-amyloid aggregation by transthyretin. *Biochemistry* **2006,** *45* (51), 15702-9.
74. Du, J.; Murphy, R. M., Characterization of the interaction of beta-amyloid with transthyretin monomers and tetramers. *Biochemistry* **2010,** *49* (38), 8276-89.
75. Costa, R.; Goncalves, A.; Saraiva, M. J.; Cardoso, I., Transthyretin binding to A-Beta peptide--impact on A-Beta fibrillogenesis and toxicity. *FEBS Lett* **2008,** *582* (6), 936-42.
76. Li, H.; Wang, B.; Wang, Z.; Guo, Q.; Tabuchi, K.; Hammer, R. E.; Sudhof, T. C.; Zheng, H., Soluble amyloid precursor protein (APP) regulates transthyretin and Klotho gene expression without rescuing the essential function of APP. *Proc Natl Acad Sci U S A* **2010,** *107* (40), 17362-7.

Figure Caption

Figure 1 Schematic illustration of the experimental design.

Figure 2 Number of proteins and endogenous peptides identified and their biological functions. (a) Venn Diagram displaying 502, 457 and 386 proteins identified with high confidence in control, MCI and AD participants biological process. (b) Biological process of total identified 795 proteins by PANTHER gene ontology search. (c) Venn Diagram showing 391, 261 and 208 endogenous peptides found in control, MCI and AD individuals. (d) Biological process of total identified 93 protein precursors by PANTHER gene ontology search.

Figure 3 Tandem mass spectra of representative endogenous peptides and tryptic peptides (a) Endogenous peptide DHDVGSELPPEGVLGA derived from proSAAS protein precursor (b) Tryptic peptide of transthyretin YTIAALLSPYSYSTTAVVTNPKE

Figure 4 One way ANOVA analysis results of differentially altered glycoproteins via bar chart. (a) leucine-rich alpha-2-glycoprotein (b) cell surface glycoprotein MUC18 (c) neural cell adhesion molecule L1-like protein (d) cell adhesion molecule 2 (e) histidine-rich glycoprotein (f) neuronal cell adhesion molecule (g) neuronal pentraxin-1 (h) neuronal growth regulator 1 (i) Dickkopf-related protein 3 (j) apolipoprotein E (k) transthyretin (l) alpha-2-HS-glycoprotein (M) contactin-1

Figure 5 (a) Protein-protein interaction analysis of fifteen significantly altered glycoproteins by STRING 10.0. (b) Molecular functions of significantly changed glycoproteins by PANTHER gene ontology search.

Chapter 4

Increased expression of AT-1/SLC33A1 causes an autistic-like phenotype in mice by affecting dendritic branching and spine formation

Adapted from manuscript: "Increased expression of AT-1/SLC33A1 causes an autistic-like phenotype in mice by affecting dendritic branching and spine formation" Rikki Hullinger, Mi Li, **Jingxin Wang**, Yajing Peng, James Dowell, Ewa Bomba-Warczak, Heather A. Mitchell, Corinna Burger, Edwin R. Chapman, John M. Denu, Lingjun Li, Luigi Puglielli, Accepted by *J Exp Med*

Abstract

The import of acetyl-CoA into the lumen of the endoplasmic reticulum (ER) by AT-1/SLC33A1 regulates Nε-lysine acetylation of ER-resident and -transiting proteins. Specifically, lysine acetylation within the ER appears to influence the efficiency of the secretory pathway by affecting ER-mediated quality control. Mutations or duplications in *AT-1/SLC33A1* have been linked to diseases such as familial spastic paraplegia, developmental delay with premature death, and autism spectrum disorder with intellectual disability. In the present study, we generated an AT-1 Tg mouse model that selectively overexpresses human AT-1 in neurons. These animals demonstrate cognitive deficits, autistic-like social behavior, aberrations in synaptic plasticity, an increased number of dendritic spines and branches, and widespread proteomic changes. We also found that AT-1 activity regulates acetyl-CoA flux causing epigenetic modulation of the histone epitope H3K27 and mitochondrial adaptation. In conclusion, our results indicate that increased expression of AT-1 can cause an autistic-like phenotype by affecting key neuronal metabolic pathways.

INTRODUCTION

Acetyl-CoA is an essential substrate for a wide range of biochemical reactions that occur within the cell (Pietrocola et al., 2015). Cytosolic acetyl-CoA is produced predominantly by the conversion of citrate and coenzyme A (CoA) by ATP-citrate lyase (ACLY) and the condensation of free acetate and CoA by acetyl-CoA synthetase (ACESS2, also referred to as AceCS) (reviewed in (Pehar and Puglielli, 2013; Shi and Tu, 2015)). Acetyl-CoA is then actively imported into the lumen of the endoplasmic reticulum (ER) by the ER membrane transporter AT-1 (also referred to as SLC33A1) (Jonas et al., 2010), where it serves as donor of the acetyl group in the Nε-lysine acetylation of ER-resident and -transiting proteins (Choudhary et al., 2009; Pehar et al., 2012b).

Recent reports suggest that lysine acetylation within the ER is required for ER-mediated quality control. Specifically, the ER based acetyltransferases ATase1 and ATase2 associate with the oligosaccharyl transferase complex to acetylate properly folded glycoproteins (Ding et al., 2014). In addition, studies performed with two type I membrane proteins indicate that the acetylation status of nascent secretory proteins in the ER regulates the efficiency of transport along the secretory pathway (Costantini et al., 2007; Mak et al., 2014).

Decreased influx of acetyl-CoA into the ER lumen leads to aberrant induction of autophagy in both cell-based (Jonas et al., 2010; Pehar et al., 2012a) and animal (Peng et al., 2014) models. At the mechanistic level, the induction of autophagy is linked to the acetylation status of autophagy-related protein 9A (ATG9A) (Pehar et al., 2012a; Peng et al., 2014). Haploinsufficiency of AT-1 in the animal results in neurodegeneration, inflammation, and propensity to infections and cancer (Peng et al., 2014). Heterozygous mutations in AT-1 have been identified in patients affected by an autosomal dominant form of spastic paraplegia (Lin et

al., 2008) while homozygous mutations have been identified in patients affected by severe developmental delay and childhood death (Huppke et al., 2012). Chromosomal duplications affecting the 3q25.31 locus harboring *AT-1/SLC33A1* have been associated with autism spectrum disorder (ASD) and intellectual disability (see *SFARI-Autism Database*, http://sfari.org/; (Krumm et al., 2013; Prasad et al., 2012; Sanders et al., 2011)). Additionally, a gain of approximately 1.1-1.5 Mb in 3q25.2-3q25.31, which contained *AT-1/SLC33A1* and *Guanine Monphosphate Synthase* (*GMPS*, associated with myeloid leukemia) was found in three male children with autism, seizure, abnormal electroencephalogram, and facial dysmorphism (Swisshelm K., et al. ASHG Annual Meeting, October 18-22, 2014, San Diego, CA).

In the present study, we sought to investigate the consequences of increased activity of AT-1. Specifically, we generated an AT-1 transgenic (Tg) mouse model that selectively overexpresses human AT-1 in neurons. These animals demonstrate cognitive deficits, autistic-like social behavior, aberrations in synaptic plasticity, increased number of dendritic spines and branches, and widespread proteomic changes. The synaptic phenotype appears to be caused by increased trafficking of nascent proteins along the secretory pathway. In addition, we found that AT-1 Tg animals display increased expression of mitochondrial enzymes related to acetyl-CoA production, suggesting that increased movement of acetyl-CoA into the ER causes downstream compensatory mechanisms in mitochondrial activity. Furthermore, this apparent mitochondrial adaptation appears to be driven by changes in the acetylation/methylation status of Lys27 on the histone protein H3 suggesting that this site acts as a sensor to adapt supply of citrate from the mitochondria to rapidly compensate for changes in cytosolic acetyl-CoA, as induced by increased AT-1 activity.

In conclusion, our results indicate that increased expression of AT-1 can cause an autistic-like phenotype by affecting key neuronal metabolic pathways.

RESULTS

AT-1 Tg animals display cognitive deficits and autistic-like behaviors

To explore the role of AT-1 in the brain, we generated transgenic mice with an inducible neuron-specific overexpression Tet-Off system driven by the CamK2 promoter (Fig. 1A). For the purpose of this study the animals (referred to as AT-1 Tg) were maintained in the absence of doxycycline (Dox); as such, they overexpressed human AT-1 during development as well as after birth. Direct assessment of both mRNA and protein levels in neurons isolated from the adult brain confirmed successful upregulation of AT-1 in the transgenic mice (Fig. 1, B-D).

To determine whether selective overexpression of AT-1 in neurons would impact learning and memory, mice were assessed with the fear conditioning (FC), novel object recognition (NOR), and Morris water maze (MWM) tasks. We found that AT-1 Tg animals displayed significant deficits in each of these hippocampus-dependent tasks (Fig. 2, A-F). Specifically, we observed significant impairments in contextual FC (Fig. 2A) and NOR (Fig. 2B). The animals did not display differences in freezing behavior during the initial exposure to the behavior arena (Fig. 2A; see *Training*) and did not display increased anxiety in the open field test (data not shown). In the NOR test, wild-type (WT) mice spent over 40% of their total investigation time exploring the novel object, while AT-1 Tg mice displayed no preference for the novel object relative to the trained objects. During the training phase of the MWM task, transgenic mice had a significantly higher escape latency (Fig. 2C) and swam greater distances before locating the hidden platform than WT controls (Fig. 2D). Furthermore, they spent less

time in the target quadrant (Fig. 2E) and displayed fewer platform crossings during the probe trial than WT animals (Fig. 2F).

Given the possible association of *AT-1/SLC33A1* with ASD and intellectual disability we decided to test AT-1 Tg mice with the marble burying and social preference paradigms, two commonly used models of autistic behavior. We found that transgenic mice displayed decreased repetitive behavior in the marble burying task (Fig. 2G) and failed to show a preference for investigating the novel mouse in the social preference paradigm (Fig. 2H). These tests indicate aberrations in repetitive behavior and social tendencies, suggesting that our transgenic mouse line shares similarities with behavioral models of ASD (Dere et al., 2014; Gkogkas et al., 2013; Kouser et al., 2013; Lugo et al., 2014; Pucilowska et al., 2015; Speed et al., 2015). To assess whether the above behavioral changes were solely due to developmental events, we also generated AT-1 Tg mice in the presence of Dox to repress expression of the transgene. The Dox was suspended at weaning and the animals were analyzed 6 months later (Fig. S1A). Under these conditions, we did not observe any differences with WT mice (Fig. S1, B-D) indicating that the aberrations in repetitive behavior and social tendencies described in Figure 2 are solely -or mostly- developmental.

Although there is no specific behavioral profile that defines ASD in the mouse, the combined deficiency in FC, NOR, MWM, marble burying and social preference paradigms observed in AT-1 Tg mice has been described in other mouse models of ASD (Dere et al., 2014; Gkogkas et al., 2013; Kouser et al., 2013; Lugo et al., 2014; Pucilowska et al., 2015; Speed et al., 2015) . Therefore, given the fact that chromosomal duplications of the 3q25.31 locus, which harbors *AT-1/SLC33A1*, have been associated with ASD and intellectual disability (see also later in the Discussion section), we conclude that AT-1 Tg mice display an autistic-like phenotype.

Transgenic mice demonstrate changes in neuronal morphology and aberrant synaptic plasticity

We next sought to investigate whether increased expression of AT-1 in neurons would lead to morphological and/or synaptic changes that could explain the behavioral phenotype. We found that cultured neurons from AT-1 Tg mice displayed a drastic increase in the number of dendritic branches and spines compared to WT controls (Fig. 3, A-D). Similar results were obtained *in vivo* by analyzing hippocampal tissue from 6-month-old mice (Fig. 3E), thus confirming that overexpression of AT-1 causes drastic changes in neuronal morphology.

Alterations in neuronal function were assessed by hippocampal slice electrophysiology. Long term potentiation (LTP) and its counterpart, long term depression (LTD), are believed to be correlated with cognitive function, and it is has been demonstrated that a balance between LTP and LTD must be present in order for normal learning and memory formation to occur (Feldman, 2009; Huganir and Nicoll, 2013). We found that AT-1 Tg mice display increased levels of hippocampal LTP elicited by Theta burst stimulation (3xTBS) (Fig. 3F) and significantly reduced LTD induced by paired pulse low-frequency facilitation (PP-LFS) (Fig. 3G) when compared to non-transgenic WT controls. Changes in spine morphology and concomitant altered synaptic plasticity similar to AT-1 Tg mice have been observed in a number of animal models of neurological disorders (Amini et al., 2013; Auerbach et al., 2011; Auffret et al., 2009; Neuhofer et al., 2015). Importantly, when exposed to Dox during development to repress expression of the transgene, AT-1 Tg mice displayed neither electrophysiological nor morphological changes (Fig. S2).

To better understand the mechanisms driving the observed changes in neuron structure and activity in AT-1 Tg mice, we selected target proteins known to be involved in neuronal outgrowth and synaptic plasticity, and compared expression levels in the hippocampi of transgenic and WT mice (Fig. 3H). Neurexin1, Neuroligin3, and Rap2a were selected as indicators of dendritic branching and synapse formation (Craig and Kang, 2007; Kawabe et al., 2010). AMPA2/3/4, mGluR5, synaptogyrin1, synapsin3, and Rab12 were selected as markers of pre- and post-synaptic plasticity (Alvarez-Castelao and Schuman, 2015; Chia et al., 2013; Martenson and Tomita, 2015). We found that a majority of these proteins were significantly upregulated in post-synaptic densities of transgenic animals compared to controls (Fig. 3, I and J). Similar results were also obtained by mass spectrometry (see later in Table S2).

AT-1 Tg animals display widespread proteomic changes

Recognized functions of Nε-lysine acetylation include regulation of expression, activity, molecular stabilization, and conformational assembly of a protein (Kouzarides, 2000). Proteomic studies have shown that a wide range of proteins undergoes Nε-lysine acetylation within the lumen of the ER; they include both ER-resident and -transiting proteins with a variety of biological functions (Choudhary et al., 2009; Pehar et al., 2012b). We have previously shown that changes in expression and/or activity of AT-1 have profound effects on the ER acetylation profile; specifically, increased activity leads to increased acetylation of ER cargo proteins (Jonas et al., 2010), while reduced activity has the opposite effect (Peng et al., 2014).

To dissect the molecular mechanism(s) of the AT-1 Tg phenotype, we first resolved the neuron-specific ER acetylome of the adult AT-1 Tg brain. The analysis identified 395 acetylation sites on 152 proteins (Table S1) involved in different biological pathways, as assessed by the

PANTHER classification system (Fig. 4A). STRING analysis identified six major clusters (Fig. 4B), all highly relevant to our observed phenotype, suggesting that the increased acetylation of proteins may have consequences for neuronal outgrowth and adhesion, synaptic activity and structure, cell surface transporters, and protein synthesis.

To further elucidate the impact of increased AT-1 activity on protein expression, we conducted large-scale liquid chromatography coupled to tandem mass spectrometry (MS) analysis to identify global changes in the proteome of AT-1 Tg animals. We found that in hippocampal tissue, 476 proteins were upregulated in transgenic mice compared to CamK2 controls (Table S2). CamK2 mice were selected as the control group to correct for possible non-specific variations in protein expression. Identified proteins were responsible for a variety of biological functions (Fig. 5A). STRING analysis at 90% confidence revealed eight clusters that were highly correlated to the observed clusters in the acetylome data (Fig. 5B).

When viewed globally, the acetylome (Fig. 4) and proteome (Fig. 5) data identify four major biological functions: (1) protein translation and quality control; (2) transport of synaptic vesicles, assembly of synaptic connections, and regulation of synaptic activity; (3) neuronal migration, outgrowth and adhesion; (4) mitochondria activity, specifically electron transport chain, tricarboxylate acid cycle, and acetyl-CoA metabolism. Therefore, these data suggest that the increased influx of acetyl-CoA into the ER lumen in AT-1 Tg mice affects proteins related to synaptic plasticity, neuron structure, transporter activity, and protein synthesis.

AT-1 Tg mice display increased efficiency of the secretory pathway

The proteome data (see Fig. 5) indicate that 163 out of 476 proteins that were found upregulated in AT-1 Tg mice are either membrane or secreted proteins that normally are

translated on the ER surface and inserted into the secretory pathway (Table S2.B). Although involved in different biochemical and cellular functions (Fig. 6A), STRING analysis revealed that the great majority of them are tightly related to synaptic plasticity (data not shown), therefore highlighting the significance of these proteins for our observed phenotype.

We recently reported that the ER-based acetyltransferases, ATase1 and ATase2, associate with the oligosaccharyl transferase complex to acetylate correctly folded glycoproteins (Ding et al., 2014). Studies conducted with two type I membrane glycoproteins suggest that Nε-lysine acetylation might regulate efficiency of transport along the secretory pathway (Costantini et al., 2007; Mak et al., 2014). To investigate whether increased influx of acetyl-CoA into the ER lumen can indeed cause global changes in the efficiency of transport in the secretory pathway, we used human neuroglioma (H4) cells overexpressing AT-1. Cells were labeled with azide-modified mannosamine (ManNAz), which is metabolically incorporated into sialic acid-containing glycoproteins while they traffic through the secretory pathway (Laughlin and Bertozzi, 2007). Sialic acid is the last sugar to be attached to the oligosaccharide chain; both the Golgi-membrane cytidine monophosphate (CMP)-sialic acid transporter and the Golgi-resident sialyltransferase are located in the *trans*-Golgi (reviewed in (Hirschberg et al., 1998)). Only nascent glycoproteins that have successfully reached the *trans*-Golgi can be sialylated (reviewed in (Hirschberg et al., 1998)). Therefore, ManNAz can specifically determine efficiency of trafficking of newly-synthesized glycoproteins. The results show a dramatic increase in the levels of nascent glycoproteins as a result of AT-1 overexpression (Fig. 6B). To account for possible changes in protein translation, we also labeled cells with O-propargyl-puromycin (OPP), an alkyne analog of puromycin that incorporates into newly translated proteins. The results revealed a significant increase in rate of protein-biosynthesis (Fig. 6C). However, normalization

of ManNAz (rate of transport) and OPP (rate of biosynthesis) labeling still showed increased delivery of secretory glycoproteins to the cell surface (Fig. 6D). Importantly, OPP incorporation does not discriminate between secretory and non-secretory proteins while ManNAz does. Therefore, we can conclude that although increased translation can partially account for the increased levels of secretory proteins in AT-1 Tg mice, increased efficiency of transport along the secretory pathway is likely to play a major role. To support the studies conducted with ManNAz, we also determined levels of cell-surface biotinylated proteins. Again, we found a significant increase in AT-1 overexpressing cells (Fig. 6E). Using a cell-impermeable reagent to biotinylate primary amines can be used as a reliable readout of steady-state levels of cell-surface proteins. Biotin labeling does not visualize secreted proteins or proteins that localize to intracellular organelles (i.e.; Golgi apparatus, lysosomes, etc.). Therefore, in contrast to ManAz labeling, it underestimates the efficiency of the secretory pathway. However, the biotin read-out (Fig. 6E) is consistent with the ManAz results (Fig. 6B). Finally, to confirm the above conclusions we also determined ManNAz incorporation in mouse embryo fibroblasts (MEFs) generated from AT-$1^{S113R/+}$ mice. AT-1_{S113R} is a mutant version of AT-1 associated with a familial form of spastic paraplegia (Lin et al., 2008) and is deficient in acetyl-CoA transport activity (Peng et al., 2014). As a result, AT-$1^{S113R/+}$ mice represent a model of AT-1 haploinsufficiency (Peng et al., 2014). As expected, we observed a significant reduction in levels of cell-surface glycoproteins in AT-$1^{S113R/+}$ MEFs (Fig. 6F), thus supporting the conclusion that influx of acetyl-CoA into the ER lumen regulates efficiency of trafficking of secretory glycoproteins.

When taken together, the above results suggest that AT-1 Tg mice have increased delivery of ER cargo proteins to the cell surface. The same animals also have increased levels of

163 proteins that are translated on the ER and insert into the secretory pathway. It is likely that these two findings are functionally related and that they play an important role in the observed mouse phenotype.

Increased activity of AT-1 leads to mitochondria adaption

In addition to changes in the efficiency of the secretory pathway, the proteomics results (see Fig. 5 and Table S2) revealed a marked upregulation of proteins related to three important mitochondrial functions: electron-chain transport, tricarboxylate acid cycle, and acetyl-CoA metabolism (Figure S3). It is worth remembering that the ER imports acetyl-CoA from the cytosol and that the cytosolic pool of acetyl-CoA is largely supplied by the conversion of mitochondria-derived citrate into acetyl-CoA (Pehar and Puglielli, 2013; Pietrocola et al., 2015). Given the increased influx of cytosolic acetyl-CoA into the ER lumen, caused by the overexpression of AT-1, we hypothesized that levels of cytosolic acetyl-CoA would be decreased in AT-1 Tg mice, forcing the mitochondria to compensate by generating more citrate for cytosolic conversion. To test this hypothesis, we initially determined acetyl-CoA levels in the cytosol of AT-1 overexpressing H4 cells. The results showed a significant reduction when compared to control cells (Fig. 7A), supporting the conclusion that the increased flux into the ER lumen is able to affect the cytosolic pool of acetyl-CoA. To assess whether this conclusion could be extended to the *in vivo* settings, we also analyzed the cytosol recovered from total hippocampal tissue (Fig. 7B) as well as isolated adult cortical neurons (Fig. 7C) of AT-1 Tg mice. Again, we observed a significant reduction of acetyl-CoA levels when compared to WT mice (Fig. 7, B and C). Importantly, two crucial proteins that allow “cross-talk” between mitochondria and cytosolic acetyl-CoA were also found to be upregulated in our proteomics

analysis: the mitochondrial carrier citrate transporter (Slc25a1) and Acly (Table S2). Slc25a1 is the mitochondria membrane transporter that translocates citrate to the cytosol while Acly is a cytosolic-based enzyme that converts mitochondria-derived citrate into acetyl-CoA by using cytosolic CoA and ATP (Pehar and Puglielli, 2013; Pietrocola et al., 2015). Interestingly, Acss2 (also known as AceCS), which contributes to the synthesis of cytosolic acetyl-CoA through the condensation of acetate and CoA (Pehar and Puglielli, 2013; Pietrocola et al., 2015), was not found to be upregulated (Table S2). It is worth stressing that although both Acly and Acss2 can generate acetyl-CoA, they carry out two different biochemical reactions; they also respond to different inputs and elicit different biological functions (Wellen et al., 2009; Xu et al., 2014). Therefore, our results would suggest that Acly, but not Acss2, is activated to ensure mitochondria adaptation in the AT-1 Tg mice. To investigate this further, we also determined mRNA levels of *Slc25a1*, *Acly*, and *Acss2* in the hippocampus of AT-1 Tg mice. Again, *Slc25a1* and *Acly* showed significant upregulation while *Acss2* did not (Fig. 7D; also discussed later).

Taken together, the above data (Fig. 7, A-D) suggest that the increased transport of acetyl-CoA into the lumen of the ER reduces levels of cytosolic acetyl-CoA, leading to increased generation and delivery of citrate from the mitochondria, followed by increased conversion of citrate into acetyl-CoA (see Fig. 7E). Therefore, we interpret the upregulation of mitochondria-related pathways as a compensatory mechanism (here defined as *mitochondrial adaptation*) to the increased expression of AT-1.

The above findings raise the question of how mitochondria can “sense” cytosolic levels of acetyl-CoA. A possible explanation is that changes in the cytosolic pool of acetyl-CoA cause epigenetic changes leading to transcriptional activation of targeted mitochondria enzymes. To test this possibility, we utilized a MS-based strategy to assess the post-translational modification

profile of histone proteins within the hippocampus of WT and AT-1 Tg animals. Previous studies have correlated acetyl-CoA availability to global epigenetic changes (Wellen et al., 2009). The MS results showed a significant increase in the acetylation of H3K27 accompanied by a significant decrease in H3K27 methylation, indicative of increased transcriptional activation in our transgenic animals (Fig. 7F; Table S3). Importantly, the MS data were confirmed by direct assessment of H3K27acetyl and H3K27trimethyl levels by immunoblotting (Fig. 7G), suggesting that a significant portion of chromatin has "sensed" the altered levels of acetyl-CoA.

When taken together, the above results suggest that the H3K27 modification might serve as a "sensor" to rapidly respond to changes in cytosolic levels of acetyl-CoA by activating key mitochondria genes. To test this hypothesis we conducted chromatin immunoprecipitation (ChIP) analysis and probed for *SLC25A1* and *ACLY*, which are key to the mitochondrial adaptation observed in AT-1 Tg mice (see Fig. 7E). Again, we used *ACSS2* as negative control. Immunoprecipitation with an anti-H3K27ac antibody resolved *SLC25A1* and *ACLY* but not *ACSS2* (Fig. 7H). When coupled to quantitative real-time PCR (qPCR), the ChIP analysis revealed increased H3K27 acetylation of *SLC25A1* and *ACLY* in AT-1 over-expressing H4 cells (Fig. 7I). These results correlate to the increased mRNA levels of both genes shown in Figure 7D. To assess whether these findings could be extended to the *in vivo* settings, we performed the same analysis with total hippocampal tissue (Fig. 7J) and isolated adult cortical neurons (Fig. 7K) of AT-1 Tg mice. Again, we found increased H3K27 acetylation of *Slc25a1* and *Acly*, when compared to WT mice.

In conclusion, these findings suggest that the enhancement in acetylation at the H3K27 site directly modifies expression of genes related to mitochondrial activity, providing a selective, targeted mitochondrial adaptation to changes in cytosolic levels of acetyl-CoA.

DISCUSSION

Here, we report that mice that overexpress human AT-1 display cognitive deficits, autistic-like social behavior, aberrations in synaptic plasticity, increased number of dendritic spines and branches, and widespread proteomic changes. We also report that AT-1 activity regulates the efficiency of the secretory pathway as well as cytosolic levels of acetyl-CoA, which -in turn- leads to epigenetic modulation of the histone epitope H3K27 and mitochondrial adaptation. In conclusion, our results indicate that increased expression of AT-1 can cause an autistic-like phenotype by affecting key neuronal metabolic pathways.

N^{ε}-lysine acetylation occurs on both ER-transiting and ER-resident proteins with a wide array of biological functions ((Choudhary et al., 2009; Pehar et al., 2012b); see also Fig. 4). Lysine acetylation is emerging as a novel mechanism that contributes to ER-based quality control: proteins that are properly folded are acetylated and proceed through the secretory pathway, whereas proteins that are not acetylated are retained and degraded (Costantini et al., 2007; Ding et al., 2014; Mak et al., 2014). The ER-based acetylation machinery requires the membrane transporter AT-1(Jonas et al., 2010) and at least two acetyltransferases (Ko and Puglielli, 2009).

Heterozygous and homozygous mutations in *AT-1/SLC33A1* have been identified in patients with familial spastic paraplegia (Lin et al., 2008) and developmental delay/premature death (Huppke et al., 2012), respectively. Chromosomal duplications affecting the 3q25.31 locus harboring *AT-1/SLC33A1* have been associated with ASD and intellectual disability (see *SFARI-Autism Database*; (Krumm et al., 2013; Prasad et al., 2012; Sanders et al., 2011)). Additionally, a short chromosomal gain that covered only two genes, *AT-1/SLC33A1* and *GMPS*, was found in

children with autism, seizure, abnormal electroencephalogram, and facial dysmorphism (Swisshelm K., et al. ASHG Annual Meeting, October 18-22, 2014, San Diego, CA). Although not definitive, the above genetic association supports a causative role of AT-1 in ASD. Importantly, chromosomal duplications of 17p13.1 (harboring *SLC13A5*), 22q11.21 (harboring *SLC25A1*), and 17q21.2 (harboring *ACLY*) have also been associated with ASD (see *SFARI-Autism Database*). SLC13A5 is the cell surface citrate transporter that translocates citrate from the extracellular *milieu* to the cytosol; SLC25A1 is the mitochondria membrane citrate transporter that translocates citrate from the mitochondria to the cytosol; and ACLY is the enzyme responsible for converting cytosolic citrate into acetyl-CoA. In conclusion, these individual genetic associations affecting *SLC14A5*, *SLC25A1*, *ACLY*, and *AT-1/SLC33A1* seem to suggest a model where increased supply and movement of acetyl-CoA into the ER lumen is tightly linked to ASD. This model is supported by the phenotype of AT-1 Tg mice reported here. It is also worth mentioning that, like *SLC33A1* (Huppke et al., 2012), mutations in *SLC13A5* and *SLC25A1* have also been associated with developmental delay of the brain (Hardies et al., 2015; Prasun et al., 2015; Thevenon et al., 2014). In addition, single variants on several other genes that are directly or indirectly related to the citrate-acetyl-CoA metabolic pathway have been identified in ASD cohorts. Some of them have also been associated with developmental delay and epileptic encephalopathy. These include *SLC16A3*, *SLC16A7*, *SLCA1*, *SLCA2*, *SLC25A12*, *SLC25A14*, *SLC25A24*, and *SLC25A27* (see *SFARI-Autism Database*).

The molecular basis behind the observed AT-1 Tg phenotype involves increased transport of acetyl-CoA into the lumen of the ER, likely causing changes in the ER acetylome as well as in the efficiency of the secretory pathway. The ability of Nε-lysine acetylation to regulate activity and/or levels of modified proteins is well documented (Kouzarides, 2000; Pehar and

Puglielli, 2013; Pietrocola et al., 2015). Both the acetylome (see Fig. 4) and the proteome (see Fig. 5) identified a large set of proteins that are implicated in cellular functions that could immediately explain the Tg mouse phenotype, specifically: neuronal migration and adhesion, neurite outgrowth, transport of synaptic vesicles, assembly of synaptic connections, and regulation of synaptic activity. Therefore, it is conceivable to assume that the increased influx of acetyl-CoA, and consequent changes in the ER acetylome and in the efficiency of the secretory pathway, is the primary cause of ASD-like phenotype.

The characterization of the AT-1 Tg phenotype also revealed that the animals display decreased levels of cytosolic acetyl-CoA and, as a result, mitochondrial adaptation. This adaptive mechanism appears to be driven by epigenetic changes on H3K27. Importantly, we did not detect widespread epigenetic changes (see Table S3). This is likely due to the compensatory activity of Acly; indeed, silencing of *ACLY* in mammalian cells causes global changes in histone acetylation (Wellen et al., 2009). Whether the mitochondria adaptation participates in the observed phenotype remains to be fully determined. The fact that genes such as *SLC13A5* and *ACLY*, which can alter cytosolic levels of acetyl-CoA without affecting mitochondria activity directly, might be linked to ASD (discussed above) would suggest that mitochondria have limited impact. However, genes such as *SLC25A12* and *SLC25A1*, which immediately influence acetyl-CoA production or delivery from the mitochondria, have also been linked to ASD (discussed above). Therefore, the role of mitochondria might be more complex than expected. It is also possible that cross-talk exists between mitochondria, cytosol and ER so that either increased supply (from mitochondria and/or cytosol) or movement of acetyl-CoA into the ER lumen causes ASD. In conclusion, the specific role of the mitochondria citrate pathway and of the cytosolic citrate/acetyl-CoA pools in ASD remains to be fully dissected.

Although increased efficiency of the secretory pathway seems to be logically connected to increased expression of proteins related to synaptic plasticity and learning and memory formation, a majority of proteins identified by mass spectrometry do not insert into the secretory pathway. However, the dramatic increase in dendritic spine and branch formation observed in our animals would require targeted (and secondary) upregulation of cytosolic and nuclear proteins related to neurite outgrowth. Establishing synaptic densities would also require the activation of cytosolic scaffolding and adaptor proteins that support the structural integrity of the synapse as well as regulate signal transduction pathways. Indeed, our proteomic assessment found significant increase of cytosolic proteins that are involved in different aspects of the above mentioned biological pathways.

In conclusion, our studies have revealed that both decreased (Peng et al., 2014) and increased (present study) AT-1 activity affect neuronal biology. In particular, the data shown here have linked increased AT-1 activity to ASD, and elucidated a novel aspect of acetyl-CoA metabolism that affects learning and memory, synaptic plasticity, protein expression, mitochondrial activity, and epigenetics. Further studies will now be necessary to determine whether changes in AT-1 activity at various points in the lifespan of the mouse will induce significant changes in neuronal morphology and/or behavior.

MATERIALS AND METHODS

Generation of AT-1 Tg mice

cDNA encoding human AT-1 was isolated by BamHI and EcoRV digestion from Topo-AT-1 constructs and subcloned into pTRE-Tight vector (Clontech). TRE-AT-1 transgenic lines were generated by injection of linearized pTRE-Tight vector (Clontech) containing AT-1 cDNA

into pronucleus of fertilized eggs from FVB mice. Monogenic pTRE-AT-1 mice were back-crossed to wild-type C57BL/6 mice for 5 generations and then bred to CamK2a-tTA mice (B6.Cg-Tg(Camk2a-tTA)1Mmay/DboJ; The Jackson Laboratory, stock number 007004) generating non-transgenic wild-type (WT), Camk2a-tTA monogenic, TRE-AT-1 monogenic, and CamK2a-tTA;TRE-AT-1 (referred to as AT-1 Tg) mice. Genotyping from tail DNA was performed using the following primers: AT-1-forward (AAT CTG GGA AAC TGG CCT TCT), AT-1-reverse (TAT TACCGC CTT TGA GTG AGC TGA), Camk2a-tTA-forward (CGCTGTGGGGCATTTTACTTTAG), and Camk2a-tTA-reverse (CATGTCCAGATCGAAATCGTC).

Animal experiments were carried out in accordance with the NIH Guide for the Care and Use of Laboratory Animals and were approved by the Institutional Animal Care and Use Committee of the University of Wisconsin-Madison and the Madison Veterans Administration Hospital. Unless specified, WT littermates were used as control throughout our study.

Cell cultures

Mouse Embryonic Firoblasts (MEFs) from WT and AT-$1^{S113R/+}$ mice were described in (Peng et al., 2014); AT-1 overexpressing H4 (human neuroglioma) cells were described in (Jonas et al., 2010; Pehar et al., 2012b). MEFs and H4 cells were maintained in DMEM supplemented with 10% Fetal Bovine Serum (FBS) and 1% Penicillin/Streptomycin/Glutamine solution (Mediatech, Inc). G418 was used as selection marker in $H4_{AT-1}$ cells (Jonas et al., 2010; Pehar et al., 2012b).

Adult neurons were isolated as described (Cahoy et al., 2008). Briefly, mouse brains were quickly removed and digested with Neural Tissue Dissociation Kit (Miltenyi Biotec) and

gentleMACS Dissociators (Miltenyi Biotec). After enzymatic and mechanical dissociation, the cell suspension was passed through 40µm cell strainer and washed with HBSS. The pellets were spun down at 300g for 10 min and re-suspended with buffer from Neuron Isolation Kit (Miltenyi Biotec). Non-Neuron Cells were removed by using Non-Neuron Cells Biotin-Antibody Cocktail (Miltenyi Biotec) followed by Anti-Biotin MicroBeads (Miltenyi Biotec). After magnetic separation with QuadroMACS Separator (Miltenyi Biotec), the neuronal cells were collected as the unlabeled fraction from the flow-through part.

Primary neuronal cultures were generated from E17-E18 mice. The isolated hippocampus was digested in 0.25% trypsin-EDTA (Gibco BRL) for 15 min at 37 °C, followed by trituration into a single-cell suspension with fire-polished glass pipettes. The dissociated cells were resuspended in the plating medium (Dulbecco's modified Eagle's medium with 10% fetal bovine serum). The cells were plated onto coverslips pre-coated with poly-D-lysine (100 µg/ml, Sigma) at a density of 150 cells per mm^2. Four hours after plating, media was changed to a serum-free neurobasal medium supplemented with 2% B27 (Gibco BRL) and 0.5 mM GlutaMAX (Invitrogen). Media was changed every 3 days. Cultures were maintained in 37ºC and 5% CO2 incubator for up to 21 days. For immunocytofluorescence analysis, cultures were fixed with 4% paraformaldehyde and immunostained with NeuN antibody (clone A60, 1:100, Chemicon) and tetramethyl rhodamine isothiocyanate (TRITC)-labeled phalloidin (0.5 µg/ml, Sigma). Nuclei were counterstained with DAPI.

Behavioral analysis

All behavioral tests were conducted and analyzed by an experimenter blind to the genotype of the mice. Unless indicated, mice were approx. 6 months-old when tested.

Morris Water Maze (MWM). The test consisted of 8 blocks of testing over 4 days, followed by a probe trial. The first day, mice were briefly exposed to the pool to test swimming capabilities. Each mouse received 2 blocks of testing per day for 4 days, each block was comprised of 4 sequential trials, for a total of 32 training trials. Each trial, the mouse was placed into the pool at a different start location and allowed to swim until it either located the platform, or reached the end of the 60 second trial. Behavior was monitored by an overhead camera and Noldus Ethovision software. If the mouse did not locate the platform within the 60 second time limit, it was gently placed on the platform and allowed to rest and look around for a few seconds before beginning the next trial. On the final day of testing, after completing the 8^{th} block of testing, mice were exposed to the pool for a probe trial. For this trial, the platform was removed, and swimming behavior was monitored for 60 seconds.

Novel Object Recognition (NOR). Each mouse received 5 sequential 6 minute trials, with a 3 minute inter-trial interval. The first trial allowed the mouse to acclimate to the empty arena, with no objects presented. Trials 2-4 were training trials with 4 objects presented within the arena. For the testing trial, one randomly selected object was removed and replaced with a novel object. This trial was videotaped and subsequently analyzed by a trained observer blind to both mouse condition and the identity of the novel object. Time spent investigating all objects was used as total investigative time; the time spent exploring each object was expressed as percent total investigative time. All objects used were pre-tested to guarantee no innate preference, and the novel object was rotated through a pool of these objetcs.

Fear Conditioning (FC). Mice were tested in the delay FC paradigm to assess contextual memory formation. On day 1, mice were exposed to a 6 minute training protocol, which consisted of 2 pairings of 30 sec of 87 dB white noise and a 1.5 sec 0.7 mA shock, with an

intertrial interval of 2 minutes. On day 2, 22 hours after exposure to the training protocol, mice were tested for conditioning to context. Mice were placed back in the chambers in which they originally received the shock for 5 minutes, but no stimuli (noise or shock) were presented. Percent freezing during each portion of the fear conditioning paradigm was calculated by the Coulbourn FreezeFrame 3 software using video feed and a motion index.

Marble Burying. The marble burying paradigm was used to assess repetitive behaviors. Each mouse was placed into a clean cage with extra bedding and 20 black marbles arranged in a grid pattern. After 30 minutes of exploration in the cage, the number of marbles completely buried in the bedding and the number of marbles buried greater than 50% were counted.

Social Interaction. The social interaction test was used to assess autistic-like social behaviors in the AT-1 Tg mouse model. The test was conducted in a three-chambered arena with a central compartment and 2 side compartments. The test consisted of 3 trials, each 10 minutes in length, run consecutively. In trial 1, the experimental mouse is exposed to an empty arena and allowed to explore freely. In trial 2, a male juvenile mouse was placed in a mesh cup in one side chamber of the box and an identical empty mesh cup was placed in the opposite side chamber of the box to provide a neutral object control. The mesh cup allowed for visual and scent cues, but prevented aggressive behavior between the mice. During trial 3, the juvenile stimulus mouse from trial 2 was kept in the test arena, and a novel juvenile male mouse was placed in a cup in the opposite chamber. All trials were videotaped and used for subsequent analysis. Investigation time was scored by a trained observer blind to genotype.

Assessment of neuronal morphology

Confocal images of cultured neurons were uploaded and analyzed in Imaris 8 (Oxford Instruments, Concord MA) using FilamentTracer module. Filaments and spines were traced, segmented, and assigned using a Semi Automatic and Manual detection methods. 2-D rendering of cells was created based on FilamentTracer analysis in Imaris, using cylinder display. Data was extracted into Prism 6 (Graphpad Software, Inc, LaJolla, CA) and statistical significance was determined by performing student t-test and ANOVA.

Golgi staining

Brains were harvested from 6 month old mice and Golgi stained using the FD Neurotech Golgi Kit. Briefly, brains were immersed in equal parts solution A and solution B and allowed to sit for 2 weeks. Brains were moved into solution C for 72 hours, and then sliced to 100μM thickness using a sliding microtome.

Electrophysiology

Extracellular recordings of field excitatory postsynaptic potentials (fEPSP), were measured from acute hippocampal slices (400μm) prepared as previously described (Pehar et al., 2010). Briefly, enameled bipolar platinum–tungsten stimulating electrodes were placed along Schaffer collaterals, and fEPSPs were recorded with ACSF filled recording electrodes (2-4 MOhm) from area CA1 stratum radiatum. Baseline responses were set to an intensity that evoked an fEPSP with a slope of 50% that of the maximum evoked response. For LTP, 3x theta burst (3xTBS) consisted of a total of three trains of 10 bursts (each burst consisting of four stimulations at a frequency of 100 Hz) with an interburst interval of 200 ms. For LTD, paired pulse low frequency stimulation consisted of 900 pairs of stimuli with a 50ms paired pulse

interval. Input/output curves were performed before and at the end of LTD recordings to ensure slice health. Data were analyzed by two-way ANOVA (treatment and time) with repeated measures (mixed model) and Bonferroni post hoc tests.

Post-synaptic densities

Hippocampal tissue was homogenized in Tris-acetate buffer (50 mM, pH 7.4) containing 100 μM EGTA, 0.32 M sucrose and protease and phosphatase inhibitors. Homogenate was centrifuged at 1,000×g for 10 min. Supernatants were then centrifuged at 14,000×g for 20 min and the resulting pellet was defined as the crude synaptosomal fraction. Synaptosomes were resuspended in Tris-acetate buffer. To purify the post-synaptic densities (PSD), the synaptosomal fraction was diluted with 20 mM Tris-HCl, pH 6.0, 0.1 mM $CaCl_2$ containing 1% Triton X-100, mixed for 20 min at 4°C on a plate rocker and then centrifuged at 25 p.s.i. for 20 min using an air-driven ultracentrifuge (Airfuge; Beckman). The pellet was resuspended in 20 mM Tris-HCl, pH 8.0, 0.1 mM $CaCl_2$ containing 1% Triton X-100. Samples were mixed again for 20 min at 4°C on the plate rocker and then centrifuged on the Airfuge (25 p.s.i.) for 20 minutes. The insoluble pellet containing the PSD fraction was suspended in Tris-acetate buffer (50 mM, pH 7.4, 100 μM EGTA, 0.32 M sucrose and proteases inhibitors) and stored at −80°C until use.

Immunoblotting

Western blotting was performed on a 4–12% Bis-Tris SDS-PAGE system (NuPAGE, Invitrogen) as described (Costantini et al., 2006; Pehar et al., 2010; Peng et al., 2014). The following primary antibodies were used in this study: Neurexin 1 (1:1000, Pierce, #18730) Neuroligin 3 (1:1000, Pierce, #18849); Rap2a (1:1000, Pierce, #23298); Synapsin 3

(1:1000, Pierce, #25658); AMPA2/3/4 (1:1000, Cell Signaling Technologies, #2460), mGluR5 (1:3000, Millipore, #76316); Gapdh (1:5000, Millipore, #2302); Synaptogyrin 1 (1:1000, Abcam, #113886); Rab12 (1:1000, Abcam, #170046); Psd-95 (1:1000, Cell Signaling Technologies, #2507); S100 (1:1000, Abcam, #4066); Iba1 (1:1000, Wako, #016-20001); βIII Tubulin (1:1000, Abcam, #18207); L1CAM (1:1000, Abcam, #24345); NFL160 (1:5000, Abcam, #9034); H3K27ac (1:1000, Active Motif, #39134); H3K27met (1:1000, Active Motif, #39157). HRP-conjugated anti-mouse or anti-rabbit secondary antibodies were used for chemiluminescent detection (ImageQuant LAS4000; GE Healthcare); goat anti-rabbit Alexa Fluor 680-conjugated or anti-mouse Alexa Fluor 800-conjugated secondary antibodies were used for infrared imaging (LICOR Odyssey Infrared Imaging System; LI-COR Biosciences).

Proteomics studies and mass spectrometry

ER Acetylome. The neuron specific acetylome was determined as before (Pehar et al., 2012b). Briefly, adult neurons were isolated with the gentle MACS Dissociator. Proteins were digested with trypsin prior to high-resolution high-accuracy LC-MS/MS analysis at the Mass Spectrometry facility at the University of Wisconsin-Madison Peptides and proteins were identified with the Mascot search engine (Matrix Science, London, UK) via automated database searching of all tandem mass spectra. The output of the MS analysis was further processed to select proteins that insert into the secretory pathway.

Proteomics. Hippocampi of 3 pairs of transgenic and control mice were dissected and two hippocampi from the same mouse were placed into the same 1.5 mL tube and stored in -80 °C until homogenization. A 200 μl aliquot of lysis buffer (8 M urea, 100 mM Tris – HCl, protease inhibitor cocktail) was added to each tube whose bottom was immersed in the ice bath. Then

tissues were immediately sonicated by Fisher Scientific Model 120 Sonic Dismembrator at 120 W output with 3 bursts that each lasted 45 s respectively. Tissues were allowed to cool on the ice bath for 30 s after each burst. The protein concentration was determined by 660 nm protein assay and 10 μg protein of each sample was used for tryptic digestion.

The protein of each sample was denatured with 8 μL of 8 M urea that were diluted by 50 mM ammonium bicarbonate buffer and reduced by 1 μL of 0.50 M DTT by incubation at 37 °C for 1 h. After incubation, the reduced sample was alkylated by 2.7 μL of 0.55 M IAA and kept in the dark for 15 min. 1 μL of 0.50 M DTT was applied to quench the IAA alkylating reaction for 10 min. 70 μL of 50 mM ammonium bicarbonate was used to dilute the urea to a final concentration of 1 M and 0.5 μg trypsin was added into the solution subsequently. The protein digestion reaction was incubated for 18 h at 37 °C and subsequently quenched by 2.5 μL of 10% formic acid. The solid phase extraction of the tryptic peptides were performed by Varian 100 μL C 18 Omix Tips (Palo Alto, CA). The peptides were sequentially eluted with 60 μL 50% ACN in 0.1% formic acid, dried by SpeedVac and reconstituted with 9 μL 0.1% formic acid. 1μL 1 μM enolase digestion standards were spiked into each sample as internal standards.

Online reversed phase liquid chromatography (RPLC) separation of the tryptic peptides was performed on a nanoAcquity UPLC (Waters Corp., Milford, MA) and subsequently analyzed with Q Exactive quadrupole orbitrap mass spectrometer (Thermo Fisher Scientific, San Jose, CA). The chromatographic separation was carried out by mobile phase A that consisted of 0.1% formic acid in water and mobile phase B with 0.1% formic acid in ACN. 1 μL of each sample was injected onto a custom – packed column that was made by silica tubing (360 μm o.d., 75 μm i.d.) with emitter tips pulled by a laser puller. The column was packed with Waters 150Å, 1.7 μm, BEH C18 material. The LC gradient started with 5% B and increased to 30% B

over 120 min at a flow rate of 350 nL/min. Ions were generated under positive electrospray ionization (ESI) at 2.8 KV capillary voltage; 275 °C capillary temperature; 30% collision energy via high – energy collision dissociation (HCD). MS1 scans were acquired over *m/z* 200 – 2000 at 70 k resolution and data – dependent tandem MS scan were acquired via selection of the top 10 most abundant precursor ions by HCD fragmentation with an isolation window of *m/z* 2.0 at a resolution of 17,500. Other parameters include: automatic gain control 1e5; maximum ion injection time, 100 ms; dynamic exclusion enabled with unassigned, + 1 and greater than + 8 charges ignored for MS/MS selection. Each sample was analyzed as technical triplicates.

Tryptic peptide identification was carried out via Proteome Discoverer 1.4 (Thermo Scientific). FASTA file was downloaded from Uniprot's reference database of *Mus musculus* (release 2014_08) with manually added yeast enolase fasta file (SwissProt P00924). Other parameters include: allowed missed cleavage, 1; enzyme, trypsin; fixed modification, carbamidomethylation of cysteine (+57.0215 Da); variable modification, oxidation of methionine (+15.9949 Da); peptide mass tolerance, 10 ppm; fragment mass tolerance, 0.1 Da. q value was set to achieve 1% false discovery rate (FDR) via Percolator to verify the identified peptides, and the results were filtered by high confidence peptide identification. For label free quantification of acquired data, SIEVE (Version 2.1, Thermo Scientific) was applied. The following parameters were used in the peak alignment and frame generation: m/z min=300, m/z max=1500, frame time width= 6.0 min, frame m/z width=0.02 Da, retention time start=15 min, retention time stop=90 min, peak intensity threshold=10,000. Alignment was validated using the spiked enolase tryptic peptides with accurate m/z and retention time. Proteins with p value < 0.05 were selected.

Histone extraction and LC-MS/MS analysis. Quantification of histone post-translational modifications was performed as recently described (Krautkramer et al., 2015). Hippocampi were

dounced in ice-cold nuclear isolation buffer (10 mM Tris-HCl, 10 mM NaCl, 3 mM $MgCl_2$, 0.15% NP-40) and nuclei pelleted by centrifugation. Histones were acid extracted from nuclei, derivatized with propionic anhydride and trypsinized as described (Lin and Garcia, 2012). Histone peptides were injected onto a Dionex Ultimate3000 nanoflow HPLC with a Waters NanoEase C18 column (100 µm x 15 cm, 3 µm) coupled to a Thermo Scientific Q-Exactive mass spectrometer at 700 nL/min using a 2-35% gradient of acetonitrile. Histone modifications were identified and quantified using the MaxQuant (v 1.4.1.2). Spectra were searched against the human SwissProt database using a 20 ppm mass tolerance for the first search and a 4.5 ppm mass tolerance for the main search. The enzyme was specified as ArgC with zero missed cleavages. Variable modifications were set as follows: acetyl(K), butyryl(K), dimethyl(K), propionlyl(K), trimethyl(K) and propionyl(peptide n-terminus). A reverse decoy database was generating within MaxQuant and the False Discovery Rate (FDR) was set to <0.01 for peptide spectrum matches (PSMs), proteins, and modification sites.

Changes in protein expression as identified by mass spectrometry were further analyzed using the Protein ANalysis THrough Evolutionary Relationship (PANTHER) gene list analysis (http://www.pantherdb.org/) (Mi et al., 2013) and STRING 10 software (http://string-db.org/) (Szklarczyk et al., 2015) to highlight protein-protein interactions.

Acetyl-CoA determinations

Cytosol was isolated from total hippocampal tissue, isolated adult cortical neurons, and H4 cells using differential centrifugation as previously described (Ko and Puglielli, 2007; Pehar et al., 2014). Acetyl-CoA was measured using the PicoProbe Kit (Abcam).

Trafficking of secretory proteins

Nascent glycoproteins were labeled using Click-It ManNAz (Life Technologies) and visualized using the Click-It Cell Reaction Kit (Life Technologies). Biotinylated cell surface proteins were isolated using the Cell Surface Protein Isolation Kit (Pierce) and visualized by Western blot with Streptavidin-HRP (1:10,000). Rate of protein biosynthesis was assessed using the Click-It Plus OPP Alexa Fluor 488 Protein Synthesis Assay Kit (Life Technologies).

Chromatin Immunoprecipitation (ChIP) analysis

Control and AT-1 overexpressing H4 cells were cultured in complete medium in 150-mm Petri dishes until ~70% confluent. The cells were then fixed by the addition of 280 μl of 37% formaldehyde (Sigma) to 10 ml of culture medium for 10 min at 37 °C, harvested, and processed for ChIP using a commercially available kit (Active Motif). H3K27-DNA immune complexes were precipitated with a polyclonal antibody against H3K27ac (Active Motif). PCR was carried out using primer sets for *Acly*, *Slc25a1*, and *Acss2* centered on the first exon region. Primer sets used were: *Acly*: GCT TAG CCT GTG AGC TGA T (Sense), AGG TGG TGC AGA TGT ACT TG (AntiSense); *Slc25a1*: GAA TGG GTC GTG GTC TCA GTA G (Sense), CTG CTA GGA TTG CCT TCC C (AntiSense); *Acss2*: AAGAGCGGCAGTGGAAG (Sense), TTCGATCCAGCACGTTGTAG (AntiSense). No antibody and normal IgG served as internal negative controls.

Real-time PCR

Real-time PCR was carried out using the Roche 480 lightcycler and Sybr Green Real Time PCR Master Mix (Life Technologies). The cycling parameters were: 95 °C, 10 s; 52 °C,

20s; 72 °C, 30s, for a maximum of 45 cycles. Controls without reverse transcription were included in each assay. PCR primers specific to each gene were: *Slc25a1* and *Acly* (BioRad, Prime PCR assay); *Acss2*: 5'-AAGAGCGGCAGTGGAAG-3' (Sense), 5'-TTCGATCCAGCACGTTGTAG-3' (AntiSense); *Gapdh*: 5′-ACCACAGTCCATGCCATCAC-3′ (Sense), 5′-TCCACCACCCTGTTGCTGTA-3′ (AntiSense). Expression levels were normalized against *Gapdh* levels and are expressed as percent of control.

Statistical analysis

Data analysis was performed using GraphPad InStat 3.06 statistical software (GraphPad Software Inc.). Data is expressed as mean ± SD. Unless specified, comparison of the means was performed using Student's t-test or one-way ANOVA followed by Tukey-Kramer multiple comparisons test. Differences were declared statistically significant if $P<0.05$. Unless specified, throughout the paper the following statistical significance is used: *, $P<0.05$; **, $P<0.005$; #, $P<0.0005$.

Online supplemental material

Fig. S1 shows behavioral analysis of Dox-treated AT-1 Tg mice. Fig. S2 shows LTP and Golgi staining of hippocampal sections in Dox-treated AT-1 Tg mice. Fig. S3 shows STRING analysis and clustering of mitochondria-related proteins that were found upregulated in AT-1 Tg mice. Table S1 lists all acetylated peptides identified in the adult brain of AT-1 Tg mice. Table S2 provides the entire list of proteins identified by MS in WT and AT-1 Tg mice. Table S3 lists all post-translational modifications of histone proteins identified in in WT and AT-1 Tg mice.

ACKNOWLEDGEMENTS

We thank Dr. Qiang Chang and Dr. Albee Messing for critical reading of an early version of this manuscript. This work was supported by: VA Merit Award (BX001638), NIH (NS094154; GM065386), and a core grant to the Waisman Center from NICHD (P30 HD03352). E.R.C. is an Investigator of the Howard Hughes Medical Institute. R.H. was supported by a National Science Foundation Graduate Research Fellowship. The authors declare no competing financial interests.

Author Contribution

RH, ML, JW, YP, EBW, JD, HAM, JMD, LL, and LP conceived and designed experiments. RH, ML, JW, YP, EBW, JD, and HAM performed experiments and analyzed data. CB, ERC, JMD, LL, and LP analyzed data. RH and LP wrote the paper. All authors contributed to writing and revision of the paper.

REFERENCES

Alvarez-Castelao, B., and E.M. Schuman. 2015. The Regulation of Synaptic Protein Turnover. *J Biol Chem* 290:28623-28630.

Amini, M., C.L. Ma, R. Farazifard, G. Zhu, Y. Zhang, J. Vanderluit, J.S. Zoltewicz, F. Hage, J.M. Savitt, D.C. Lagace, R.S. Slack, J.C. Beique, M. Baudry, P.A. Greer, R. Bergeron, and D.S. Park. 2013. Conditional disruption of calpain in the CNS alters dendrite morphology, impairs LTP, and promotes neuronal survival following injury. *J Neurosci* 33:5773-5784.

Auerbach, B.D., E.K. Osterweil, and M.F. Bear. 2011. Mutations causing syndromic autism define an axis of synaptic pathophysiology. *Nature* 480:63-68.

Auffret, A., V. Gautheron, M. Repici, R. Kraftsik, H.T. Mount, J. Mariani, and C. Rovira. 2009. Age-dependent impairment of spine morphology and synaptic plasticity in hippocampal CA1 neurons of a presenilin 1 transgenic mouse model of Alzheimer's disease. *J Neurosci* 29:10144-10152.

Cahoy, J.D., B. Emery, A. Kaushal, L.C. Foo, J.L. Zamanian, K.S. Christopherson, Y. Xing, J.L. Lubischer, P.A. Krieg, S.A. Krupenko, W.J. Thompson, and B.A. Barres. 2008. A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J Neurosci* 28:264-278.

Chia, P.H., P. Li, and K. Shen. 2013. Cell biology in neuroscience: cellular and molecular mechanisms underlying presynapse formation. *J Cell Biol* 203:11-22.

Choudhary, C., C. Kumar, F. Gnad, M.L. Nielsen, M. Rehman, T.C. Walther, J.V. Olsen, and M. Mann. 2009. Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* 325:834-840.

Costantini, C., M.H. Ko, M.C. Jonas, and L. Puglielli. 2007. A reversible form of lysine acetylation in the ER and Golgi lumen controls the molecular stabilization of BACE1. *Biochem J* 407:383-395.

Costantini, C., H. Scrable, and L. Puglielli. 2006. An aging pathway controls the TrkA to p75(NTR) receptor switch and amyloid beta-peptide generation. *EMBO J* 25:1997-2006.

Craig, A.M., and Y. Kang. 2007. Neurexin-neuroligin signaling in synapse development. *Curr Opin Neurobiol* 17:43-52.

Dere, E., L. Dahm, D. Lu, K. Hammerschmidt, A. Ju, M. Tantra, A. Kastner, K. Chowdhury, and H. Ehrenreich. 2014. Heterozygous ambra1 deficiency in mice: a genetic trait with autism-like behavior restricted to the female gender. *Front Behav Neurosci* 8:181.

Ding, Y., C.D. Dellisanti, M.H. Ko, C. Czajkowski, and L. Puglielli. 2014. The endoplasmic reticulum-based acetyltransferases, ATase1 and ATase2, associate with the oligosaccharyl-transferase to acetylate correctly folded polypeptides. *J Biol Chem* 289:32044-32055.

Feldman, D.E. 2009. Synaptic mechanisms for plasticity in neocortex. *Ann Rev Neurosci* 32:33-55.

Gkogkas, C.G., A. Khoutorsky, I. Ran, E. Rampakakis, T. Nevarko, D.B. Weatherill, C. Vasuta, S. Yee, M. Truitt, P. Dallaire, F. Major, P. Lasko, D. Ruggero, K. Nader, J.C. Lacaille, and N. Sonenberg. 2013. Autism-related deficits via dysregulated eIF4E-dependent translational control. *Nature* 493:371-377.

Hardies, K., C.G. de Kovel, S. Weckhuysen, B. Asselbergh, T. Geuens, T. Deconinck, A. Azmi, P. May, E. Brilstra, F. Becker, N. Barisic, D. Craiu, K.P. Braun, D. Lal, H. Thiele, J. Schubert, Y. Weber, R. van 't Slot, P. Nurnberg, R. Balling, V. Timmerman, H. Lerche,

S. Maudsley, I. Helbig, A. Suls, B.P. Koeleman, P. De Jonghe, and E.R.E.S.C. autosomal recessive working group of the Euro. 2015. Recessive mutations in SLC13A5 result in a loss of citrate transport and cause neonatal epilepsy, developmental delay and teeth hypoplasia. *Brain* 138:3238-3250.

Hirschberg, C.B., P.W. Robbins, and C. Abeijon. 1998. Transporters of nucleotide sugars, ATP, and nucleotide sulfate in the endoplasmic reticulum and Golgi apparatus. *Annu Rev Biochem* 67:49-69.

Huganir, R.L., and R.A. Nicoll. 2013. AMPARs and synaptic plasticity: the last 25 years. *Neuron* 80:704-717.

Huppke, P., C. Brendel, V. Kalscheuer, G.C. Korenke, I. Marquardt, P. Freisinger, J. Christodoulou, M. Hillebrand, G. Pitelet, C. Wilson, U. Gruber-Sedlmayr, R. Ullmann, S. Haas, O. Elpeleg, G. Nurnberg, P. Nurnberg, S. Dad, L.B. Moller, S.G. Kaler, and J. Gartner. 2012. Mutations in SLC33A1 cause a lethal autosomal-recessive disorder with congenital cataracts, hearing loss, and low serum copper and ceruloplasmin. *Am J Hum Genet* 90:61-68.

Jonas, M.C., M. Pehar, and L. Puglielli. 2010. AT-1 is the ER membrane acetyl-CoA transporter and is essential for cell viability. *J Cell Sci* 123:3378-3388.

Kawabe, H., A. Neeb, K. Dimova, S.M. Young, Jr., M. Takeda, S. Katsurabayashi, M. Mitkovski, O.A. Malakhova, D.E. Zhang, M. Umikawa, K. Kariya, S. Goebbels, K.A. Nave, C. Rosenmund, O. Jahn, J. Rhee, and N. Brose. 2010. Regulation of Rap2A by the ubiquitin ligase Nedd4-1 controls neurite development. *Neuron* 65:358-372.

Ko, M.H., and L. Puglielli. 2007. The sterol carrier protein SCP-x/pro-SCP-2 gene has transcriptional activity and regulates the Alzheimer disease gamma-secretase. *J Biol Chem* 282:19742-19752.

Ko, M.H., and L. Puglielli. 2009. Two Endoplasmic Reticulum (ER)/ER Golgi Intermediate Compartment-based Lysine Acetyltransferases Post-translationally Regulate BACE1 Levels. *J Biol Chem* 284:2482-2492.

Kouser, M., H.E. Speed, C.M. Dewey, J.M. Reimers, A.J. Widman, N. Gupta, S. Liu, T.C. Jaramillo, M. Bangash, B. Xiao, P.F. Worley, and C.M. Powell. 2013. Loss of predominant Shank3 isoforms results in hippocampus-dependent impairments in behavior and synaptic transmission. *J Neurosci* 33:18448-18468.

Kouzarides, T. 2000. Acetylation: a regulatory modification to rival phosphorylation? *EMBO J* 19:1176-1179.

Krautkramer, K.A., L. Reiter, J.M. Denu, and J.A. Dowell. 2015. Quantification of SAHA-Dependent Changes in Histone Modifications Using Data-Independent Acquisition Mass Spectrometry. *J Proteome Res* 14:3252-3262.

Krumm, N., B.J. O'Roak, E. Karakoc, K. Mohajeri, B. Nelson, L. Vives, S. Jacquemont, J. Munson, R. Bernier, and E.E. Eichler. 2013. Transmission disequilibrium of small CNVs in simplex autism. *Am J Hum Genet* 93:595-606.

Laughlin, S.T., and C.R. Bertozzi. 2007. Metabolic labeling of glycans with azido sugars and subsequent glycan-profiling and visualization via Staudinger ligation. *Nat Protoc* 2:2930-2944.

Lin, P., J. Li, Q. Liu, F. Mao, R. Qiu, H. Hu, Y. Song, Y. Yang, G. Gao, C. Yan, W. Yang, C. Shao, and Y. Gong. 2008. A missense mutation in SLC33A1, which encodes the acetyl-

CoA transporter, causes autosomal-dominant spastic paraplegia (SPG42). *Am J Hum Genet* 83:752-759.

Lin, S., and B.A. Garcia. 2012. Examining histone posttranslational modification patterns by high-resolution mass spectrometry. *Methods Enzymol* 512:3-28.

Lugo, J.N., G.D. Smith, E.P. Arbuckle, J. White, A.J. Holley, C.M. Floruta, N. Ahmed, M.C. Gomez, and O. Okonkwo. 2014. Deletion of PTEN produces autism-like behavioral deficits and alterations in synaptic proteins. *Front Mol Neurosci* 7:27.

Mak, A.B., M. Pehar, A.M. Nixon, R.A. Williams, A.C. Uetrecht, L. Puglielli, and J. Moffat. 2014. Post-Translational Regulation of CD133 by ATase1/ATase2-Mediated Lysine Acetylation. *J Mol Biol* 426:2175-2182.

Martenson, J.S., and S. Tomita. 2015. Synaptic localization of neurotransmitter receptors: comparing mechanisms for AMPA and GABAA receptors. *Curr Opin Pharmacol* 20:102-108.

Mi, H., A. Muruganujan, and P.D. Thomas. 2013. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res* 41:D377-386.

Neuhofer, D., C.M. Henstridge, B. Dudok, M. Sepers, O. Lassalle, I. Katona, and O.J. Manzoni. 2015. Functional and structural deficits at accumbens synapses in a mouse model of Fragile X. *Front Cell Neurosci* 9:100.

Pehar, M., M.C. Jonas, T.M. Hare, and L. Puglielli. 2012a. SLC33A1/AT-1 protein regulates the induction of autophagy downstream of IRE1/XBP1 pathway. *J Biol Chem* 287:29921-29930.

Pehar, M., M.H. Ko, M. Li, H. Scrable, and L. Puglielli. 2014. P44, the 'longevity-assurance' isoform of P53, regulates tau phosphorylation and is activated in an age-dependent fashion. *Aging Cell* 13:449-456.

Pehar, M., M. Lehnus, A. Karst, and L. Puglielli. 2012b. Proteomic assessment shows that many endoplasmic reticulum (ER)-resident proteins are targeted by N(epsilon)-lysine acetylation in the lumen of the organelle and predicts broad biological impact. *J Biol Chem* 287:22436-22440.

Pehar, M., K.J. O'Riordan, M. Burns-Cusato, M.E. Andrzejewski, C.G. del Alcazar, C. Burger, H. Scrable, and L. Puglielli. 2010. Altered longevity-assurance activity of p53:p44 in the mouse causes memory loss, neurodegeneration and premature death. *Aging Cell* 9:174-190.

Pehar, M., and L. Puglielli. 2013. Lysine acetylation in the lumen of the ER: a novel and essential function under the control of the UPR. *Biochim Biophys Acta* 1833:686-697.

Peng, Y., M. Li, B.D. Clarkson, M. Pehar, P.J. Lao, A.T. Hillmer, T.E. Barnhart, B.T. Christian, H.A. Mitchell, B.B. Bendlin, M. Sandor, and L. Puglielli. 2014. Deficient Import of Acetyl-CoA into the ER Lumen Causes Neurodegeneration and Propensity to Infections, Inflammation, and Cancer. *J Neurosci* 34:6772-6789.

Pietrocola, F., L. Galluzzi, J.M. Bravo-San Pedro, F. Madeo, and G. Kroemer. 2015. Acetyl Coenzyme A: A Central Metabolite and Second Messenger. *Cell Metab* 21:805-821.

Prasad, A., D. Merico, B. Thiruvahindrapuram, J. Wei, A.C. Lionel, D. Sato, J. Rickaby, C. Lu, P. Szatmari, W. Roberts, B.A. Fernandez, C.R. Marshall, E. Hatchwell, P.S. Eis, and S.W. Scherer. 2012. A discovery resource of rare copy number variations in individuals with autism spectrum disorder. *G3* 2:1665-1685.

Prasun, P., S. Young, G. Salomons, A. Werneke, Y.H. Jiang, E. Struys, M. Paige, M.L. Avantaggiati, and M. McDonald. 2015. Expanding the Clinical Spectrum of Mitochondrial Citrate Carrier (SLC25A1) Deficiency: Facial Dysmorphism in Siblings with Epileptic Encephalopathy and Combined D,L-2-Hydroxyglutaric Aciduria. *JIMD Reports* 19:111-115.

Pucilowska, J., J. Vithayathil, E.J. Tavares, C. Kelly, J.C. Karlo, and G.E. Landreth. 2015. The 16p11.2 deletion mouse model of autism exhibits altered cortical progenitor proliferation and brain cytoarchitecture linked to the ERK MAPK pathway. *J Neurosci* 35:3190-3200.

Sanders, S.J., A.G. Ercan-Sencicek, V. Hus, R. Luo, M.T. Murtha, D. Moreno-De-Luca, S.H. Chu, M.P. Moreau, A.R. Gupta, S.A. Thomson, C.E. Mason, K. Bilguvar, P.B. Celestino-Soper, M. Choi, E.L. Crawford, L. Davis, N.R. Wright, R.M. Dhodapkar, M. DiCola, N.M. DiLullo, T.V. Fernandez, V. Fielding-Singh, D.O. Fishman, S. Frahm, R. Garagaloyan, G.S. Goh, S. Kammela, L. Klei, J.K. Lowe, S.C. Lund, A.D. McGrew, K.A. Meyer, W.J. Moffat, J.D. Murdoch, B.J. O'Roak, G.T. Ober, R.S. Pottenger, M.J. Raubeson, Y. Song, Q. Wang, B.L. Yaspan, T.W. Yu, I.R. Yurkiewicz, A.L. Beaudet, R.M. Cantor, M. Curland, D.E. Grice, M. Gunel, R.P. Lifton, S.M. Mane, D.M. Martin, C.A. Shaw, M. Sheldon, J.A. Tischfield, C.A. Walsh, E.M. Morrow, D.H. Ledbetter, E. Fombonne, C. Lord, C.L. Martin, A.I. Brooks, J.S. Sutcliffe, E.H. Cook, Jr., D. Geschwind, K. Roeder, B. Devlin, and M.W. State. 2011. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron* 70:863-885.

Shi, L., and B.P. Tu. 2015. Acetyl-CoA and the regulation of metabolism: mechanisms and consequences. *Curr Opin Cell Biol* 33:125-131.

Speed, H.E., M. Kouser, Z. Xuan, J.M. Reimers, C.F. Ochoa, N. Gupta, S. Liu, and C.M. Powell. 2015. Autism-Associated Insertion Mutation (InsG) of Shank3 Exon 21 Causes Impaired Synaptic Transmission and Behavioral Deficits. *J Neurosci* 35:9648-9665.

Szklarczyk, D., A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, M. Simonovic, A. Roth, A. Santos, K.P. Tsafou, M. Kuhn, P. Bork, L.J. Jensen, and C. von Mering. 2015. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* 43:D447-452.

Thevenon, J., M. Milh, F. Feillet, J. St-Onge, Y. Duffourd, C. Juge, A. Roubertie, D. Heron, C. Mignot, E. Raffo, B. Isidor, S. Wahlen, D. Sanlaville, N. Villeneuve, V. Darmency-Stamboul, A. Toutain, M. Lefebvre, M. Chouchane, F. Huet, A. Lafon, A. de Saint Martin, G. Lesca, S. El Chehadeh, C. Thauvin-Robinet, A. Masurel-Paulet, S. Odent, L. Villard, C. Philippe, L. Faivre, and J.B. Riviere. 2014. Mutations in SLC13A5 cause autosomal-recessive epileptic encephalopathy with seizure onset in the first days of life. *Am J Hum Genet* 95:113-120.

Wellen, K.E., G. Hatzivassiliou, U.M. Sachdeva, T.V. Bui, J.R. Cross, and C.B. Thompson. 2009. ATP-citrate lyase links cellular metabolism to histone acetylation. *Science* 324:1076-1080.

Xu, M., J.S. Nagati, J. Xie, J. Li, H. Walters, Y.A. Moon, R.D. Gerard, C.L. Huang, S.A. Comerford, R.E. Hammer, J.D. Horton, R. Chen, and J.A. Garcia. 2014. An acetate switch regulates stress erythropoiesis. *Nature Med* 20:1018-1026.

FIGURE LEGENDS

Figure 1. AT-1 Tg mice selectively overexpress AT-1 in forebrain neurons.

(A) AT-1 Tg mice were generated with an inducible neuron-specific overexpression Tet-Off system driven by the CamK2 promoter.

(B) Western blot showing isolation of neurons from the brain of adult AT-1 Tg mice. Total cells dissociated from the brain (Input), neuronal, and non-neuronal fractions are shown. The following markers were used: Iba1 for microglia; S100 for astrocytes; βIII Tubulin, L1CAM, and NFL 160 for neurons.

(C-D) mRNA (C) and protein (D) levels of AT-1 in AT-1 Tg mice. Data in (C) represent 3 independent determinations of 7 mice/group.

**, $P<0.005$; #, $P<0.0005$. Student's t-test and one-way ANOVA followed by Tukey-Kramer multiple comparisons test. Bars represent mean $\pm$ SD.

Figure 2. AT-1 Tg animals display cognitive deficits and autistic-like social behaviors.

(A) Contextual fear conditioning (FC) of WT and AT-1 Tg mice (WT n=6; AT-1 Tg n=6).

(B) Novel object recognition (NOR) of WT and AT-1 Tg mice (WT n=11; AT-1 Tg n=12).

(C-F) Morris Water Maze (MWM) of WT and AT-1 Tg mice. (WT n=10; AT-1 Tg n=10).

(G) Marble burying task of WT and AT-1 Tg mice (WT n=13; AT-1 Tg n=13). Number of marbles that were at least 50% covered (total) and number of marbles that were completely buried (100%) is shown.

(H) Social interaction test of WT and AT-1 Tg mice. Preference for investigating novel *vs* familiar mouse is shown. (WT n=13; AT-1 Tg n=13).

*, P<0.05; **, P<0.005. Student's t-test and one-way ANOVA followed by Tukey-Kramer multiple comparisons test. Bars represent mean ± SD.

Figure 3. AT-1 Tg animals display changes in neuronal morphology and imbalanced synaptic plasticity.

(A-D) Morphological assessment of hippocampal neurons in culture. Phalloidin staining (upper panels) and unbiased computer-driven reconstruction (lower panels) of both WT and AT-1 Tg neurons (A and B) are shown. Sholl analysis (C) and spines quantification (D) are shown. Scale bar, 25 µm. Number of determinations for Sholl analysis: WT n=4; AT-1 Tg n= 5. Number of independent segments analyzed for spine density: WT n=16; AT-1 Tg n=36.

(E) Golgi staining of the hippocampus from 6 month old WT and AT-1 Tg mice. Scale bar, 50 µm.

(F) 3xTBS induced LTP in WT and AT-1 Tg mice.

(G) PP-LFS LTD in WT and AT-1 Tg mice.

(H-J) Western blot of selected relevant proteins. List and function of targeted proteins (H), quantification (I; n=5), and representative Western blots (J) are shown. GAPDH served as loading control.

*, P<0.05; **, P<0.005; #, P<0.0005. Student's t-test and one-way ANOVA followed by Tukey-Kramer multiple comparisons test; for LTP and LTD, repeated measures ANOVA followed by Bonferroni post hoc tests. Bars represent mean ± SD.

Figure 4. Analysis of the brain-specific ER acetylome in AT-1 Tg mice.

(A) PANTHER analysis of identified acetylated ER-cargo proteins. Proteins were grouped based on biological functions.

(B) Clusters of highly related proteins identified by STRING analysis. Proteins were grouped based on biological function.

Analysis was done in triplicate**.** See also Table S1.

Figure 5. AT1 -Tg animals display widespread proteomic changes.

(A) PANTHER analysis of upregulated proteins.

(B) Clusters of related proteins identified by STRING analysis at high confidence (90%). Proteins were grouped based on biological functions.

Analysis was done in triplicate**.** See also Table S2.

Figure 6. AT-1 Tg mice display increased efficiency of the secretory pathway.

(A) PANTHER analysis of upregulated proteins that insert into the secretory pathway were grouped based on cellular functions.

(B) Incorporation of ManNAz into sialic acid-containing glycoproteins to assess efficiency of trafficking of newly-synthesized glycoproteins. H4(-), transfected with empty vector; H4(AT-1_{wt}), transfected with AT-1_{wt}. (H4(-) n=8; H4(AT-1_{wt}) n=9).

(C) Incorporation of OPP to determine rate of protein biosynthesis. (H4(-) n=21; H4(AT-1_{wt}) n=25).

(D) Normalization of ManNAz labeling per rate of OPP incorporation.

(E) Western blot assessment of cell surface biotinylated proteins. *Left panel*, representative images; *right panel*, quantitation. (H4(-) n=2; H4(AT-1_{wt}) n=2).

(F) Incorporation of ManNAz into sialic acid-containing glycoproteins to assess efficiency of trafficking of newly-synthesized glycoproteins into MEFs generated from WT and AT-$1^{S113R/+}$ mice. (MEF(WT) n=35; MEF(AT-$1^{S113R/+}$) n=43).

*, $P<0.05$; #, $P<0.0005$. Student's t-test. Bars represent mean $\pm$ SD.

Figure 7. Mitochondrial adaptation in AT-1 Tg mice.

(A) Cytosolic levels of acetyl-CoA in H4 cells. H4(-), control/empty vector; H4(AT-1_{WT}), overexpressing wild-type AT-1. Results are expressed as percent of H4(-). (H4(-), n=10; H4 (AT-1_{WT}) n=11).

(B) Cytosolic levels of acetyl-CoA in total hippocampal tissue. Results are expressed as percent of WT. (WT n=6; AT-1 Tg n=6).

(C) Cytosolic levels of acetyl-CoA in isolated adult neurons. Results are expressed as percent of WT. (WT n=3; AT-1 Tg n=3).

(D) mRNA levels of *Slc25a1*, *Acly*, and *Acss2* in the hippocampus of AT-1 Tg mice. Results are expressed as percent of WT. (WT n=7; AT-1 Tg n=6).

(E) Schematic summary of results showing the mitochondrial adaptation that results from the increased influx of acetyl-CoA into the ER in AT-1 Tg mice.

(F) Mass spectrometry quantification of post-translational modifications of histone proteins in the hippocampus of AT-1 Tg mice. Global changes on H3K27/K36 are shown in the left panel while changes in the acetylation and methylation profile of H3K27 are shown in the right panel. (WT n=3; AT-1 Tg n=3).

(G) Western blot assessment showing acetylation and methylation of H3K27 in AT-1 overexpressing H4 cells. Selected images are shown in the left panel while quantification (n=4) is shown in the right panel.

(H) ChIP analysis of *SLC33A1*, *ACLY* and *ACSS2* following immunoprecipitation (IP) with an anti-H3K27ac antibody. IP was done in H4 cells. *Left panel* shows sheared DNA; *right panel* shows ChIP-amplification. Analysis was performed in quadruplicate.

(I) ChIP-qPCR showing H3K27 acetylation of *SLC25A1* and *ACLY* in H4 cells. Results are expressed as fold of H4 (-). (H4(-), n=5; H4(AT-1_{WT}) n=5).

(J) ChIP-qPCR showing H3K27 acetylation of *Slc25a1* and *Acly* in total hippocampal tissue. Results are expressed as fold of WT. (WT n=3; AT-1 Tg n=3).

(K) ChIP-qPCR showing H3K27 acetylation of *Slc25a1* and *Acly* in isolated adult neurons. Results are expressed as fold of WT. (WT n=3; AT-1 Tg n=3).

IP in panels (I-K) was performed with an anti-H3K27ac antibody.

*, $P<0.05$; **, $P<0.005$; #, $P<0.0005$. Student's t-test and one-way ANOVA followed by Tukey-Kramer multiple comparisons test. Bars represent mean $\pm$ SD.

Figure 1

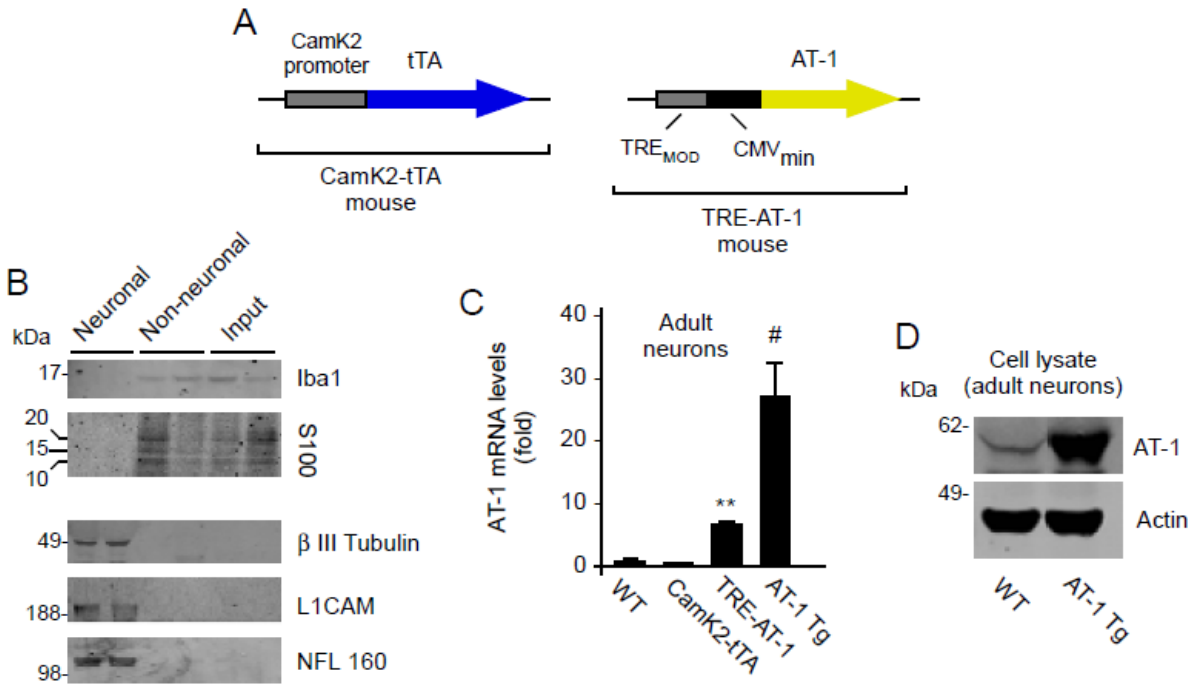

A
CamK2 promoter
tTA
CamK2-tTA mouse
AT-1
TRE$_{MOD}$
CMV$_{min}$
TRE-AT-1 mouse
B
kDa
Neuronal
Non-neuronal
Input
17-
Iba1
20
15
10
S100
49-
β III Tubulin
188-
L1CAM
98-
NFL 160
C
AT-1 mRNA levels (fold)
Adult neurons
#
**
WT
CamK2-tTA
TRE-AT-1
AT-1 Tg
D
Cell lysate (adult neurons)
kDa
62-
AT-1
49-
Actin
WT
AT-1 Tg

Figure 2

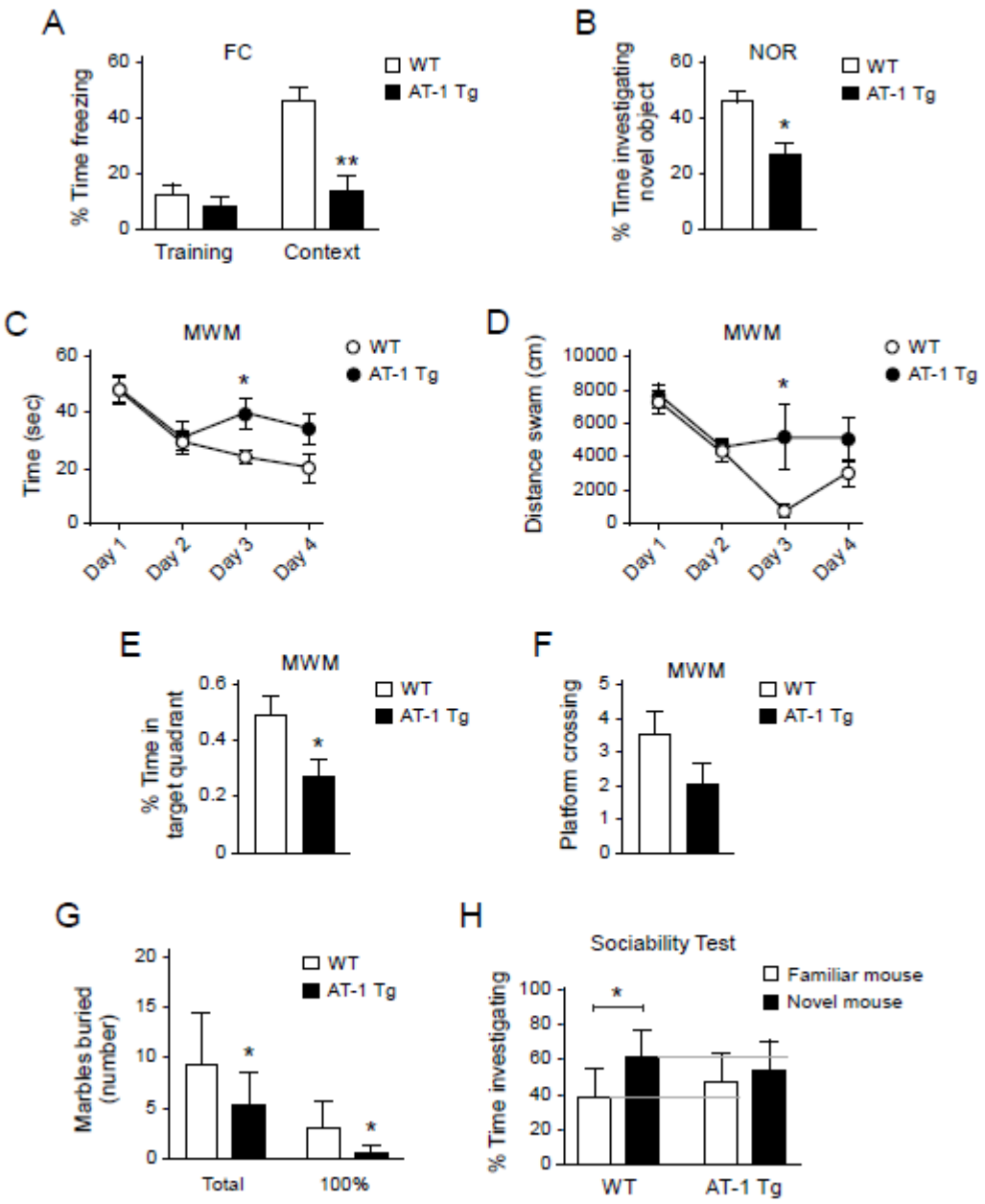
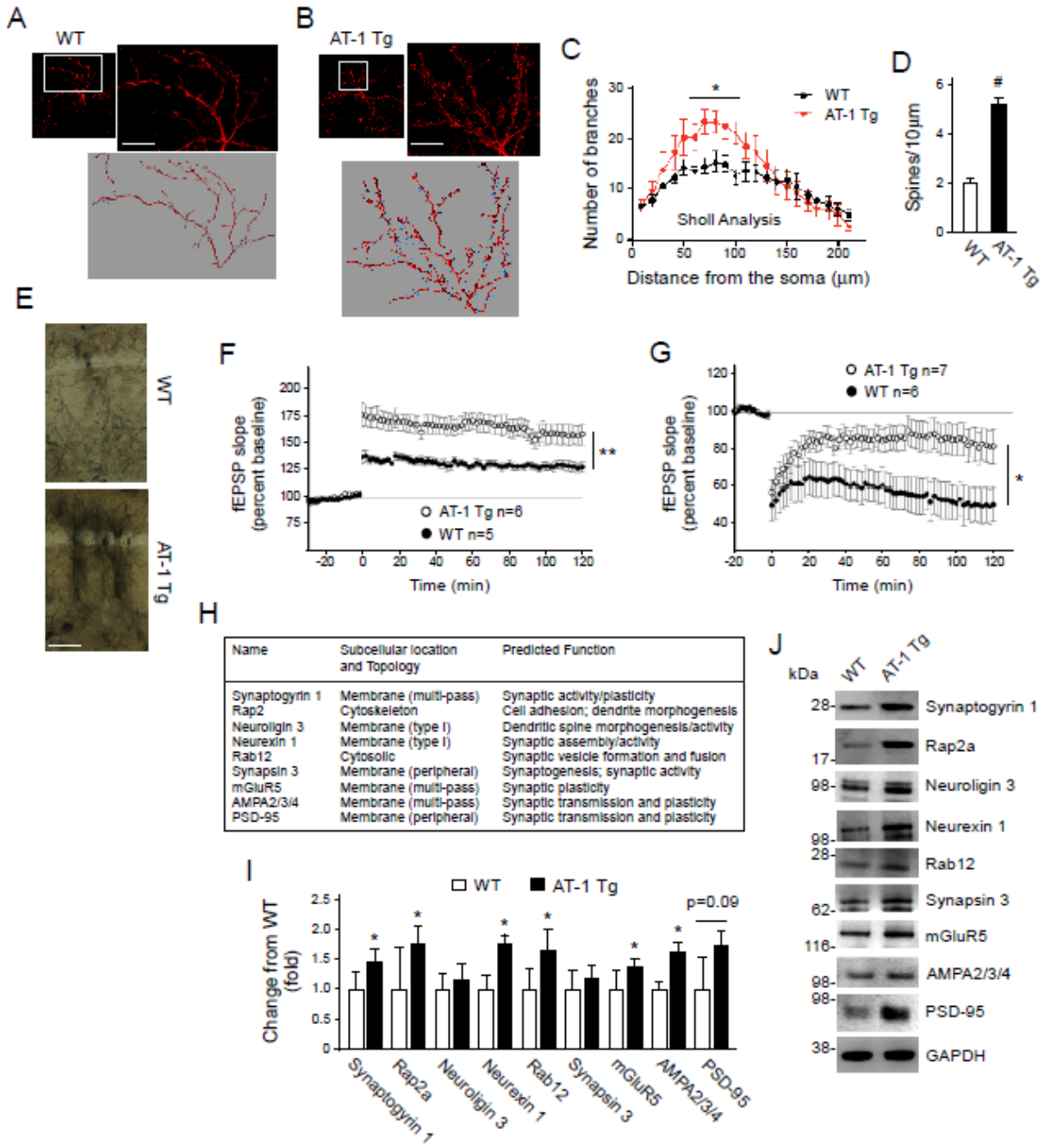

A
FC
% Time freezing
WT
AT-1 Tg
Training
Context
B
NOR
% Time investigating novel object
C
MWM
Time (sec)
Day 1
Day 2
Day 3
Day 4
D
MWM
Distance swam (cm)
E
MWM
% Time in target quadrant
F
MWM
Platform crossing
G
Marbles buried (number)
Total
100%
H
Sociability Test
Familiar mouse
Novel mouse
% Time investigating

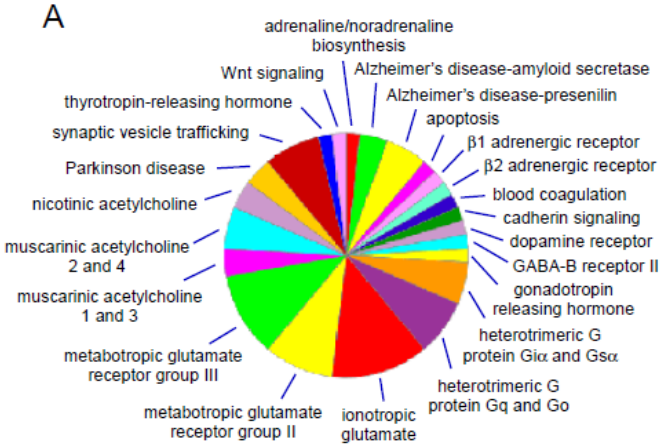
Figure 3



Name	Subcellular location and Topology	Predicted Function
Synaptogyrin 1	Membrane (multi-pass)	Synaptic activity/plasticity
Rap2	Cytoskeleton	Cell adhesion; dendrite morphogenesis
Neuroligin 3	Membrane (type I)	Dendritic spine morphogenesis/activity
Neurexin 1	Membrane (type I)	Synaptic assembly/activity
Rab12	Cytosolic	Synaptic vesicle formation and fusion
Synapsin 3	Membrane (peripheral)	Synaptogenesis; synaptic activity
mGluR5	Membrane (multi-pass)	Synaptic plasticity
AMPA2/3/4	Membrane (multi-pass)	Synaptic transmission and plasticity
PSD-95	Membrane (peripheral)	Synaptic transmission and plasticity

Figure 4

A

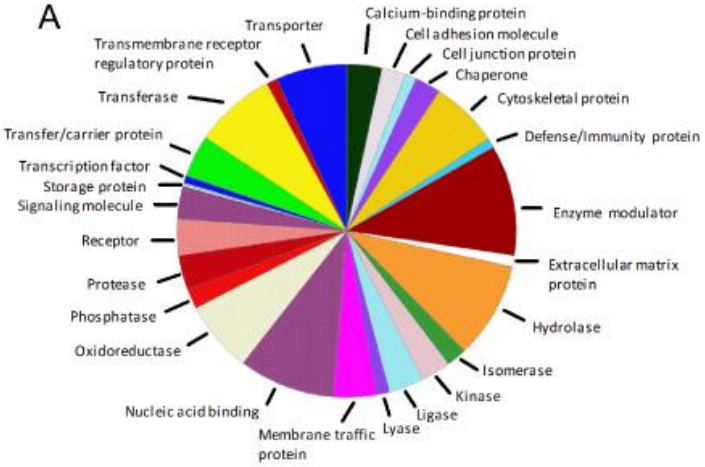


B

Cluster	Overall function
a.	Protein synthesis and folding
b.	Lysosomal acidification membrane transporters
c.	Cell surface Na^+/K^+ transporters
d.	Synaptic activity and structure
e.	Neuronal adhesion
f.	Neurite outgrowth

Figure 5

A



B

Cluster	Overall Function
A	Regulation of synaptic vesicle activity and neurotransmitter release
B	Nucleotide biosynthesis
C	Transmembrane receptor signaling
D	Protein translation
E	Cell migration and growth, endocytosis
F	Proteosome and ERAD
G	Acetyl-CoA metabolism and related pathways
H	Electron transport on the mitochondrial respiratory chain

Figure 6

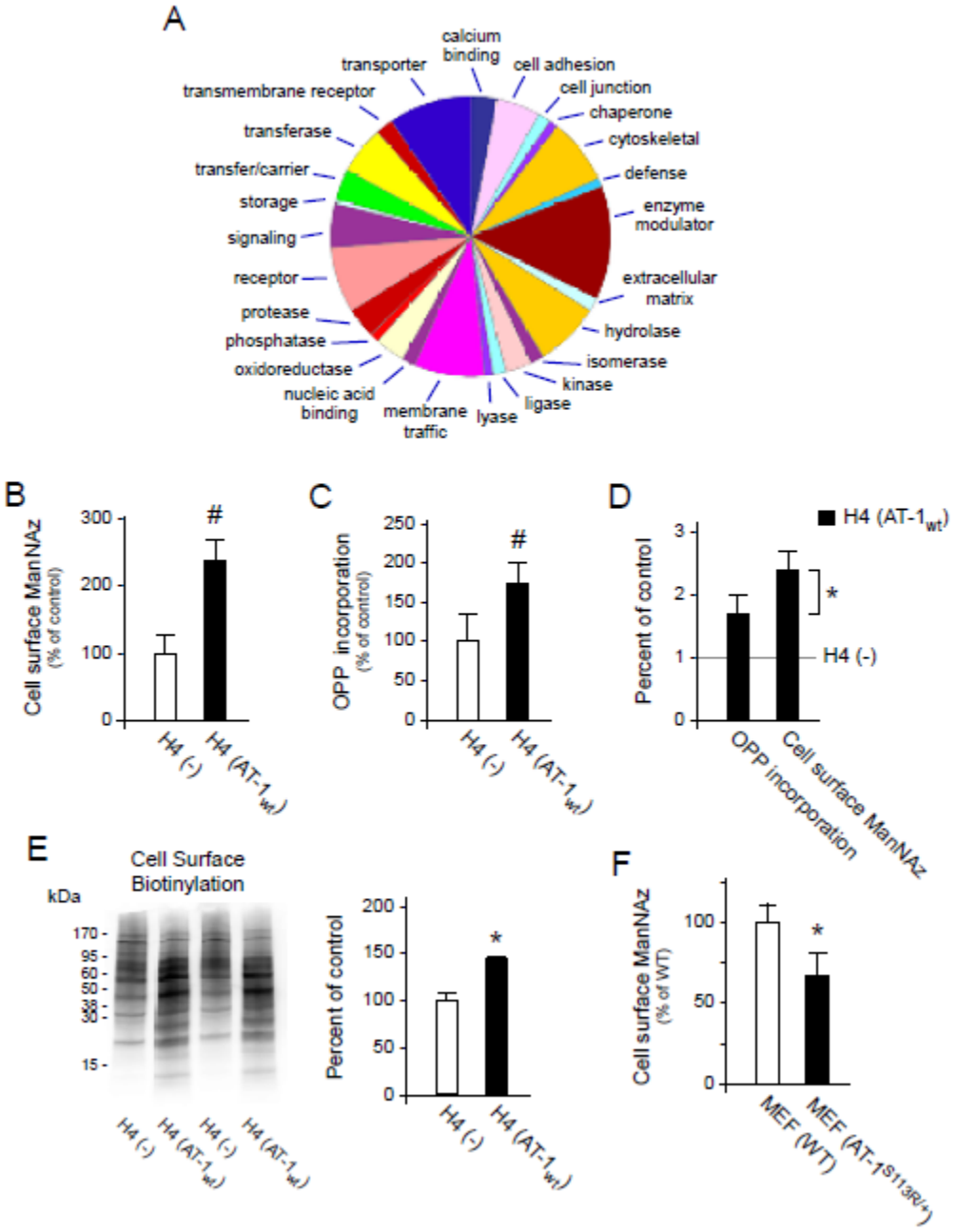
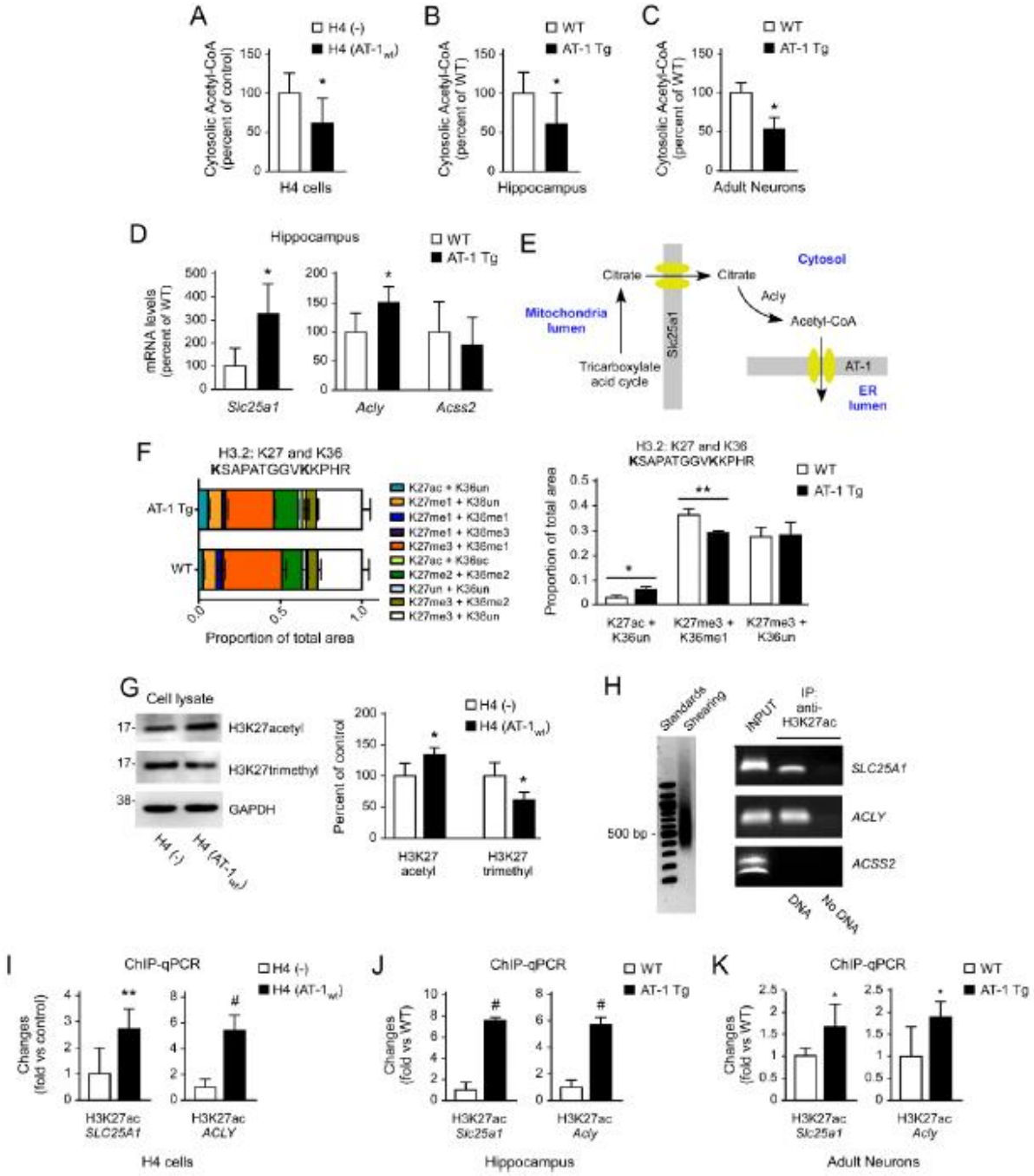

A
calcium binding
cell adhesion
cell junction
chaperone
cytoskeletal
defense
enzyme modulator
extracellular matrix
hydrolase
isomerase
kinase
ligase
lyase
membrane traffic
nucleic acid binding
oxidoreductase
phosphatase
protease
receptor
signaling
storage
transfer/carrier
transferase
transmembrane receptor
transporter
B
Cell surface ManNAz (% of control)
#
H4 (-)
H4 (AT-1wt)
C
OPP incorporation (% of control)
#
H4 (-)
H4 (AT-1wt)
D
H4 (AT-1wt)
Percent of control
*
H4 (-)
OPP incorporation
Cell surface ManNAz
E
Cell Surface Biotinylation
kDa
170 -
95 -
60 -
50 -
38 -
30 -
15 -
H4 (-)
H4 (AT-1wt)
H4 (-)
H4 (AT-1wt)
Percent of control
*
H4 (-)
H4 (AT-1wt)
F
Cell surface ManNAz (% of WT)
*
MEF (WT)
MEF (AT-1S113R/+)

Figure 7


A
H4 (-)
H4 (AT-1wt)
Cytosolic Acetyl-CoA (percent of control)
H4 cells
B
WT
AT-1 Tg
Cytosolic Acetyl-CoA (percent of WT)
Hippocampus
C
WT
AT-1 Tg
Cytosolic Acetyl-CoA (percent of WT)
Adult Neurons
D
Hippocampus
WT
AT-1 Tg
mRNA levels (percent of WT)
Slc25a1
Acly
Acss2
E
Cytosol
Citrate
Citrate
Acly
Acetyl-CoA
Mitochondria lumen
Tricarboxylate acid cycle
Slc25a1
AT-1
ER lumen
F
H3.2: K27 and K36
KSAPATGGVKKPHR
AT-1 Tg
WT
K27ac + K36un
K27me1 + K36un
K27me1 + K36me1
K27me1 + K36me3
K27me3 + K36me1
K27ac + K36ac
K27me2 + K36me2
K27un + K36un
K27me3 + K36me2
K27me3 + K36un
Proportion of total area
H3.2: K27 and K36
KSAPATGGVKKPHR
WT
AT-1 Tg
Proportion of total area
K27ac + K36un
K27me3 + K36me1
K27me3 + K36un
G
Cell lysate
H3K27acetyl
H3K27trimethyl
GAPDH
H4 (-)
H4 (AT-1wt)
Percent of control
H4 (-)
H4 (AT-1wt)
H3K27 acetyl
H3K27 trimethyl
H
Standard Shearing
INPUT
IP: anti-H3K27ac
500 bp
SLC25A1
ACLY
ACSS2
DNA
No DNA
I
ChIP-qPCR
H4 (-)
H4 (AT-1wt)
Changes (fold vs control)
H3K27ac SLC25A1
H3K27ac ACLY
H4 cells
J
ChIP-qPCR
WT
AT-1 Tg
Changes (fold vs WT)
H3K27ac Slc25a1
H3K27ac Acly
Hippocampus
K
ChIP-qPCR
WT
AT-1 Tg
Changes (fold vs WT)
H3K27ac Slc25a1
H3K27ac Acly
Adult Neurons

Chapter 5

Defining the Neuropeptidome of the Spiny Lobster *Panulirus interruptus* Brain Using a Multi-Dimensional Mass Spectrometry-Based Platform

Adapted from: " Defining the Neuropeptidome of the Spiny Lobster *Panulirus interruptus* Brain Using a Multi-Dimensional Mass Spectrometry-Based Platform" H. Ye, **J. Wang**, Z. Zhang, C. Jia, C. Schmerberg, A. Catherman, P. Thomas, N. Kelleher, D. Baro, and L. Li. *J Proteome Res*. **2015** Nov 6; 14(**11**):4776-91.

ABSTRACT

Decapod crustaceans are favored animal models for neurobiologists, whereas characterization of their endogenous neuropeptides is challenging due to limited known preprohormones. Nevertheless, great conservation is observed among decapod species shown by our home-built decapod neuropeptide database. We thereby characterized the brain extract from *P. interruptus*, an important aquaculture decapod species with few known preprohormones, with high mass accuracy and resolution data acquired on a LTQ-Orbitrap and conducted a ProSight search using a flexible algorithm that allows for sequence discrepancy from known sequence against our database. In addition to this streamlined semi-automated sequencing strategy, we characterized *P. interruptu*s by dimethylation-assisted fragmentation and manual *de novo* sequencing using nanoLC-ESI-Q-TOF, and direct tissue analysis on MALDI-TOF-TOF, further improving the coverage and resulting in an overall detection of 55 neuropeptides with 34 novel ones. The high discovery rate from this species demonstrated the usefulness of the neuropeptide discovery pipeline we developed and highlighted the advantage of utilizing multiple mass spectrometers. The localization of brain neuropeptides were further elucidated with MALDI-MSI. Collectively, our study expands the catalog of crustacean neuropeptides present in *P. interruptus*, and more importantly, presents an approach that can be adapted to explore neuropeptidomes from other species that possess limited sequence information.

KEYWORDS

Neuropeptide, *Panulirus interruptus*, *de novo* sequencing, ProSight, high-resolution and high-accuracy mass spectrometry, dimethylation-assisted fragmentation, mass spectrometric imaging

INTRODUCTION

The actions of neural circuits regulate and modify an organism's functions and behaviors, whereas these circuits are extensively modulated by cell-cell signaling molecules. Neuropeptides represent a largest and most diverse group of signaling molecules in nervous system and have shown to affect a wide range of physiological processes like hunger, pain, stress and reproduction.[1-3] To understand the underlying mechanisms of such actions, an initial step is to comprehensively characterize the peptide neuromodulators present in the system. Biologically active neuropeptides start from being cleaved from large protein precursors and then undergo multiple enzymatic processing steps in various orders, resulting in a diverse array of physical and chemical properties[4, 5]. Therefore, significant challenges are present in the detection and identification of endogenous neuropeptides, although the capability to unveil novel neuropeptides could enable a more advanced understanding of peptidergic signaling in the nervous system.

Historically, Edman degradation has served as the predominant technique to sequence novel neuropeptides by sequential removal of amino acids from the N-terminus[6]. However, this method suffers from low throughput and large sample amount required for such analysis. In recent years, mass spectrometry (MS) has revolutionized the analysis of neuropeptides . MS allows for accurate measurements of the peptide precursor ions in MS mode and the resulting fragment ions in tandem MS (MS/MS) mode. The coupling of liquid chromatography (LC) to MS/MS has permitted analysis of complex samples, including neuronal homogenate. Subsequently, the pattern of peptide fragment ions acquired in MS/MS spectra are matched to theoretical peptide spectra, which are predicted from preprohormone sequences arising from cDNA or a sequenced genome. Therefore, this workflow relies heavily on knowledge of the cleavage characteristics of the enzymes that process preprohormones to

neuropeptides and, more crucially, the preprohormone sequences. To date, the workflow of LC-MS/MS analysis in conjunction with subsequent database search has become prevalant in neuropeptidomics, a thriving field that is devoted to the characterization of neuropeptides expressed in an organism, by precisely describing hundreds of diverse forms of neuropeptides from various species including rat[7-10], mouse[11-13], invertebrates including the house cricket *Acheta domesticus*[14], the American cockroach *Periplaneta americana*[15], the fruit fly *Drosophila melanogaster*[19] , the crabs *Cancer borealis*[16], *Callinectes sapidus*[17] and *Ocypode ceratophthalma*[18], and the nematodes *Caenorhabditis elegans*[20], *Ascaris suum*[21], among others.[22]

However, applying this database searching pipeline to identify neuropeptides from MS/MS spectra is not always straightforward when the species of interest have few or virtually no preprohormone sequences available. For example, decapod crustaceans are favored animal models for neurobiologists due to their relatively simple, well-organized nervous system and intriguing behavior patterns mediated by particular neuropeptides[4, 23]. Nevertheless, complete genome sequencing has never been attempted for these decapod crustacean species so far. Therefore, previously reported crustacean neuropeptide sequences were mostly obtained by interpreting MS/MS spectra acquired from crustacean neuronal extract without known preprohormone sequences, which is termed *de novo* sequencing[24, 25]. This process is more challenging than assigning an experimental spectrum to a predicted peptide, since a spectrum of rich fragment ions at high mass accuracy is a prerequisite to read off the sequence. Despite the challenges presented, researchers in this field have successfully exploited LC-MS/MS method to identify hundreds of neuropeptides via *de novo* sequencing from Decapoda[16-18, 24-30], and previously reported neuropeptides and cDNA sequences that encoded predicted or known crustacean neuropeptides (UniProt) have been compiled to assemble a decapod crustacean neuropeptide (CNP) database. Interestingly, a significant degree of NP sequence conservation among species is observed

within the CNP database. A number of identical or similar motifs in neuropeptide families, such as tachykinin-related peptide (TRP), orcokinin and SIFamide, are shared among different crustacean species. We thereby developed a semi-automatic *de novo* sequencing strategy by searching the LC-MS/MS data obtained from the spiny lobster *Panulirus interruptus*, a decapod crustacean species that has been seldomly characterized, against the compiled CNP database as a representative model here. In this study, we coupled nanoLC to a hybrid LTQ-Orbitrap via electrospray ionization (ESI) for the acquisition of LC-MS/MS data, and employed a database searching software designed for neuropeptide discovery, ProSight PC 2.0, to accommodate potential sequence variability and potential post-translational modifications (PTMs) present in *P. interruptus*. The combination of HRAM MS data and rich peptide MS/MS fragments, in conjunction with the flexible scoring/searching software system that translates MS/MS spectra into peptide identifications results in remarkable certainty of identification from the complex *P. interruptus* brain homogenate with minimal manual intervention.

We further improved the comprehensive neuropeptide discovery workflow by employing other sample preparation techniques like N-terminal dimethylation-assisted fragmentation (DAF) and direct tissue analysis, and multiple instruments such as ESI-quadrupole-time-of-flight (Q-TOF) and matrix-assisted laser desorption/ionization (MALDI) time of flight (TOF)-TOF for the most complete coverage of *P. interruptus* neuropeptidome. Using this workflow, we discovered 55 endogenous neuropeptides from the *P. interruptus* brain. Although 21 neuropeptides have been reported previously in the crustacean species *C. borealis* and *C. sapidus*, 34 novel neuropeptides were sequenced for the first time from *P. interruptus* with high confidence. The newly gained information displays the diversity and conservation of crustacean neuropeptide sequences, and provides an important link to their potential biological functions. The inclusion of mass spectrometry imaging (MSI) to the workflow adds another dimension of knowledge to our investigation of the *P. interruptus* brain neuropeptidome. Overall, the

multi-faceted workflow demonstrated in this study serves as a basis for successful discovery of endogenous neuropeptides from *P. interruptus*, as well as a uniform approach that can be adapted to explore neuropeptidomes from crustacean and other species that have only limited sequence information.

EXPERIMENTAL SECTION

Chemical and Materials

Glacial acetic acid, borane pyridine, formaldehyde, 2,5-dihydroxybenzoic acid (DHB) and ammonium bicarbonate were obtained from Sigma-Aldrich (St. Louis, MO). Optima grade formic acid, acetonitrile (ACN), water, and methanol (MeOH) were purchased from Fisher Scientific (Pittsburgh, PA).

Animals and Dissection

California spiny lobsters were purchased from Catalina Offshore Products (San Diego, CA) and maintained in a circulating artificial seawater tank at 14–16 °C before use. Lobsters were anesthetized by packing them in ice for 30minutes, after which the dorsal carapace was removed from each individual and its supraoesophageal ganglia (brain) were dissected free from surrounding muscle and connective tissues. The details of dissection were described elsewhere.[25] Following dissection, brain samples were immediately placed in acidified MeOH (90% MeOH: 9% glacial acetic acid: 1% deionized water) and stored at -80 °C until utilized for peptide extraction.

Tissue Extraction

Brain was homogenized and extracted with acidified MeOH buffer that was also used for tissue storage. The resulting extract was then concentrated in a Savant SC 110 Speedvac concentrator (Thermo Electron Corporation, West Palm Beach, FL, USA) and re-suspended in 0.1 % formic acid.

The re-suspended extracts were then vortexed and briefly centrifuged. The resulting solution was purified and concentrated with C_{18} ZipTip (Millipore, Billerica, MA, USA). Briefly, the C_{18} ZipTip was first wetted using ACN and then pre-equilibrated for sample binding with 0.1 % formic acid in water. Subsequently, the tissue extract was loaded on the C18 ZipTip. After rinsing with 0.1 % formic acid in water for three times, the sample was eluted with 5 μL ACN:water:formic acid solution (50:49.9:0.1; vol/vol/vol). Next, the eluent was dried and re-suspended in 10 μL of 0.1 % formic acid in water, and subjected to future LC-MS/MS and MALDI-MS analysis.

N-terminal Dimethylation via Formaldehyde labeling

For formaldehyde labeling, the neuropeptide extract was re-suspended in 5 μL of 0.1 % formic acid in water followed by the addition of 1.5 μL of borane pyridine (120 mM in MeOH) and 1.5 μL of formaldehyde (4% in H_2O). The labeling mixture was then incubated in 37°C water bath for 15 min. Excess formaldehyde was quenched via the addition of 3 μL of ammonium bicarbonate buffer (0.2 M). The resulting solution was dried and re-suspended in 10 μL of 0.1 % formic acid in water, and subjected to future LC-MS/MS analysis.

Tissue Preparation for MALDI-MS

Sample preparation for direct tissue analysis: Direct tissue analysis was performed as reported previously. Briefly, the accessory lobe (AcN) and olfactory lobe (ON)[31] were designated as region 1 (R1) and region 2 (R2), and dissected out followed by a brief rinse in acidified MeOH buffer and subsequent desalting in 10 mg/mL aqueous DHB solution. The processed tissues were then placed on the MALDI target with 0.3 μL of 100 mg/ml DHB matrix solution (50:49.9:0.1 MeOH:water:formic acid, v/v) deposited on top of the tissue, which allows extraction of neuropeptides from tissue prior to forming analyte-doped crystals.

Sample preparation for MSI

The freshly dissected brain was rinsed briefly in deionized water to eliminate excess salt, embedded in an aqueous solution of gelatin (100 mg/ml) and then snap-frozen in -80° C freezer for further processing. Brain sections were acquired at a thickness of 12 μm on a cryostat (HM525, Thermo Fisher Scientific, Waltham, MA) at -20 °C and thaw-mounted onto indium tin oxide (ITO) coated conductive glass slides (Delta Technologies, Loveland, CO). The sections were then allowed to dry for 30 min under vacuum. For imaging purposes, 150 mg/ml DHB dissolved in 50% MeOH, 0.1% formic acid (v/v) was sprayed onto brain sections homogenously using an airbrush (Paasche Airbrush Company, Chicago, IL, USA). Five coats were applied and the spray duration for each coat was 30 s with 1 min dry time between each cycle.

On-line Top-down MS/MS on Nano-LC-ESI-LTQ-Orbitrap Elite and Data Analysis

The brain neuropeptide extract was further analyzed by on-line top-down MS on an Ultimate 3000 RSLCnano system coupled to an Orbitrap Elite mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). An aliquot of 1 μL was injected onto a 2 cm, 150 μm i.d. PLRP-S (d_p 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 a data-dependent mode, performing higher-energy collision dissociation (HCD)-MS^2 (scan 1), collision-induced dissociation (CID)-MS^2 (scan 2) and electron transfer dissociation (ETD)-MS^2 (scan 3) on each of the Top 3 precursors (selected by intact mass) in a FT-MS precursor scan. The specific conditions for the three fragmentation methods are as follows: HCD: Isolation width: 15 *m/z*, normalized collision energy: 30, activation time: 0.1 ms; CID: Isolation width: 15 *m/z*, normalized collision energy: 41, activation q: 0.4, activation time: 100 ms; ETD: Isolation width: 15 *m/z*, normalized collision energy: 35, activation q: 0.25, activation time: 100 ms.

Data were deisotoped with Xtract using the cRAWler algorithm (ThermoFisher, Bremen, Germany) and searched with a custom 168-core ProSightPC 3.0 cluster using an iterative search tree. ProSightPC Warehouses were created against our home-built CNP database (https://uwmadison.box.com/lilabNP) with all the crustacean neuropeptides previously reported in literature and cDNA sequences filtered in UniProt for *crustaceans* (TaxId: 6657). The data was first searched in an "absolute mass" method, in which the mass tolerance for precursor ions was set at 200 Da and that for fragment ions was 10 ppm with Δ m mode on. A minimum match of 5 fragment ions was required. Then a "biomarker" search was performed, in which the mass tolerances for precursor ions and fragment ions are both set at 10 ppm with Δ m mode off. A minimum match of 4 fragment ions was required. The semi-automatic *de novo* sequencing was performed by manually reviewing all the peptide hits and modifying the sequences with possible PTMs or amino acids when necessary.

NanoLC-ESI-Q-TOF

A Waters nanoAcquity UPLC system was coupled to a Synapt G2 HDMS mass spectrometer (Waters Corp., Milford, MA) for LC-MS/MS analysis of the brain neuropeptide extract. 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 3 µL of brain neuropeptide extract was injected and loaded onto the Waters NanoACQ 2G-V/M SymC18 (180 um X 20mm, 5 µm) using 99% mobile phase A and 1% mobile phase B at a flow rate of 5 µL/min for 1 min. Following this, a linear gradient from 5 to 45% mobile phase B at a flow rate of 300 nL/min was performed over 90 minutes at 35°C. Data dependent acquisition (DDA) was employed with 3 precursors selected for MS/MS at once . The MS scan range was from *m/z* 300–2000 and the MS/MS scan was from *m/z* 50–2000. The MS/MS sequencing was performed with a combination of manual sequencing and PepSeq software (Waters

Corp., Milford, MA) to assist in *de novo* sequencing.

CE fractionation for MALDI-MS analysis

The offline CE-MALDI-MS was performed with a home-built apparatus featuring a sheathless membrane-covered joint. Briefly, a 65 cm long fused-silica capillary (75 μm i.d./190 μm o.d.) was employed with a cellulose acetate membrane-coated porous joint made 3 cm from the outlet ends of the capillary.[32] This joint was inserted into a buffer cell filled with 0.5% ammonium acetate in pH 4.9, with the negative electrode connected. About 2 cm of capillary was stretched out of the buffer cell from a small hole with screw on the bottom. For the inlet end, a 0.6 mL plastic vial was filled with the same CE buffer with positive electrode and capillary inserted. A 10 kV high voltage was supplied by TriSep-2100 HV power supplier from Unimicro Technologies (Pleasanton, CA) for CE separation, while a high voltage of 9 kV was applied for sample loading. The CE flow was then collected on the MALDI plate every 1 min for 25 min.

MALDI-MS analysis

Direct tissue analysis and MSI of the *P. interruptus* brain together with the analysis of CE fractions were all performed on an autoflex III MALDI-TOF/TOF mass spectrometer (Bruker Daltonics, Billerica, MA, USA) equipped with a 200 Hz smartbeam. The following parameters were adopted in the positive reflectron mode for acquisition: ion source 1 voltage 19.00 kV, ion source 2 voltage 16.62 kV, reflector 1 voltage 20.90 kV, reflector 2 voltage 9.64 kV, and lens voltage 8.70 kV. The mass spectra data were acquired over a mass range of *m/z* 600–2000 where the majority of the neuropeptides were detected. For direct issue analysis, each spectrum was accumulated over 2000 laser shots for the profiled region. Additionally, the following parameters were adopted in the positive linear mode at a mass range of 3-10 kDa to profile the larger neuropeptides for direct tissue analysis: ion source 1 voltage 20.00 kV, ion source 3 voltage 18.55 kV, lens voltage 6.80 kV and pulsed ion extraction 130 ns. For

MSI acquisition, each spectrum consisted of 500 laser shots acquired in positive reflectron mode and the array of spectra was collected at 50 μm intervals in both x and y dimensions across the surface of the brain section.

For analysis of CE fractions, each spectrum was summed from 1000 laser shots acquired in positive reflectron mode. Spectrum smoothing and baseline subtraction were performed to process the spectra using flexAnalysis software (Bruker Daltonics, Billerica, MA, USA), and the resulting peaks with S/N > 3 were matched with LC-MS/MS assignments within a mass error of 0.03 Da for neuropeptide identification from resulting spectra of direct tissue and CE fraction analysis. The imaging files were processed and MS images of neuropeptides were generated using the flexImaging software package (Bruker Daltonics, Billerica, MA, USA).

RESULTS AND DISCUSSION

Comprehensive Neuropeptidomics Study from the Brain of *Panulirus interruptus* by a Multi-Dimensional MS-Based Platform

In this study, we employed three commonly used mass spectrometers, LC-ESI LTQ-Orbitrap, LC-ESI Q/TOF and MALDI TOF/TOF, for endogenous neuropeptide discovery. The results obtained from different mass spectrometers were compared and combined (**Table 1**) since each instrument offers overlapping yet distinct peptide coverage due to their specific ionization preferences and sequencing capabilities. For instance, although LC-ESI LTQ Orbitrap and LC-ESI Q/TOF share the same ionization platform, the former instrument offers higher ion transmission efficiency and more versatile MS/MS fragmentation techniques that enable a more comprehensive coverage of high-mass peptides and proteins.[25] Furthermore, spectra of high spectral resolution and mass accuracy contributed by a high-end instrument like LTQ-Orbitrap is a prerequisite for our semi-automated *de novo* sequencing

strategy as a mass tolerance of 10 ppm was set when querying the observed *m/z* values of *P. interruptus* neuropeptide precursors and fragments against the CNP database in ProSight. This permitted precise determination of possible PTMs and amino acid substitutions, alone and in combination with others. This semi-automated *de novo* sequencing approach revolutionizes the *de novo* sequencing process for decapod crustacean species by dramatically decreasing the amount of time required for analysis without sacrifice of identification confidence. It resulted in the discovery of 34 neuropeptides. On the other hand, ESI-Q/TOF provides spectra of decent quality and excels at detection of low-mass immonium ions, delivering a complementary set of neuropeptides compared to LTQ-Orbitrap. In addition to the LC-ESI-MS-based platform, the characterization of *P. interruptus* neuropeptides was further facilitated by the MALDI-MS-based approach due to rapid analysis, good sensitivity and a high tolerance to salt, providing the affirmed presence of 24 neuropeptides directly from the brain tissue. In total, we characterized and identified 55 neuropeptides, the majority of which were not reported in previous *P. interruptus* studies[33-36] and 34 of which are first sequenced for decapod crustacean species in this study. In **Table 1**, neuropeptides detected using different instruments are detailed, highlighting the combinatorial utilization of multiple powerful mass spectral techniques for the discovery of novel neuropeptides from the spiny lobster *P. interruptus* brain.

High Spectral Resolution and Mass Accuracy MS and MS/MS Sequencing

Neuropeptide extracts of *P. interruptus* brain were analyzed on a high spectral resolution and mass accuracy HRAM mass spectrometer, ESI-LTQ Orbitrap, with the top 3 precursors subjected to top-down HRAM MS/MS using a complementary suite of fragmentation techniques: HCD, CID and ETD. The ultra-high resolution (120,000 in MS^1 and 60,000 in MS^2 at *m/z* 400) and accuracy (mass error lower than 5 ppm) precursor mass together with the rich sequence-specific fragment ions produced by the complementary fragmentation techniques were used to search the CNP database within ProSightPC 2.0.

Two types of searches were carried out in ProSight. The first type of search, "biomarker", is employed to find any peptides that were identical as they appeared in the CNP database. **Figure 1** shows the fragmentation spectra and assignment of neuropeptides derived from a known *P. interruptus* tachykinin related peptide (TRP) preprohormone included in the CNP database. **Figure 1a** displays a CID MS/MS spectrum arising from the *m/z* 1179.866 (4+) ion that can be matched with partial sequence of preproTRP, DAAAPLNEVDEASANDYPILPDPIA-ARLYAFRNGNAPVGLAVPLRa, by ProSight (**Figure 1b)**. The accurately measured and completely resolved precursor *m/z* resulted in a mass error of merely 0.85 ppm from the expected *m/z* at 1179.867 as shown in **Figure 1c**. Over 48 fragment ions including 15 b- and y- ion pairs were observed in the representative MS/MS spectrum that could be matched to theoretical fragments of the TRP precursor-related peptide (TPRP), confirming the identity of ions assigned by ProSight search. The combinatorial use of different fragmentation techniques is also highlighted in the study. A fragmentation map showing fragment ions acquired in HCD, CID and ETD MS/MS spectra is shown in **Figure 1d**, with the illustration of how these fragments were produced in **Figure 1e**. As expected, the most comprehensive coverage was obtained with the use of all three fragmentation techniques at one time. The LTQ-Orbitrap employed in this study also features a great dynamic range and higher ion transmission efficiency of analytes, allowing for a comprehensive coverage of neuropeptides spanning the mass range of detection. In comparison, none of the TPRPs or their truncated forms were observed on LC-ESI Q-TOF, highlighting the importance of utilizing different instrumentation platforms for improved coverage.

Other than the capability to detect and fragment the relatively large neuropeptides as shown in **Figure 1a**, its capability to identify peptides of small size was demonstrated in **Figure 1f**. In **Figure 1f**, extensive fragmentation along the peptide backbone was observed with the use of HCD for the TRP APSGFLGMRa. Therefore, the presence of the TRP, previously proposed as a single, conserved

neuropeptide across decapod crustacean species[34], was demonstrated in *P. interruptus* brain. Indeed, no other TRP isoforms like TPSGFLGMRa—previously detected in the crab *Cancer borealis*—were seen in this study, consistent with the fact that seven identical copies of APSGFLGRMa are encoded in the *P. interruptus* preproTRP.[35] The *P. interruptus* TRP precursor undergoes several proteolytic processing steps, including cleavage at pro-hormone convertase processing sites, subsequent removal of the dibasic residues and the final amidation of the exposed Gly residual at the C-terminal, to the final peptide product.[35] Interference with this enzymatic processing cascade could possibly explain the detection of the truncated form of TRP, APSGFLGM, and the incompletely processed APSGFLGMRG, although the latter could alternatively have originated from the C-terminal mRNA coding sequence APSGFLGMR that is directly followed by a stop codon.[34] In addition, a methionine-oxidized form of TRP, APSGFLGM(O)Ra, was also identified, which could be an artifact of the sample preparation steps since no reducing agents have been added to the extract since the isolation of *P. interruptus* brain. In addition to variants of TRP, five truncated forms of the relatively large TPRP identified in **Figure 1a** were detected. The presence of multiple forms of the conserved decapod TRP or TPRP are postulated to have important impacts on the physiological functions of the neuropeptides due to different affinity to their receptors[34]. All the *P. interruptus* preproTRP-derived peptides identified via ProSight "biomarker" search are underlined with different colors in the deduced TRP precursor sequence (accession number AB 113378 in GeneBank/EMBL/DDBJ) as shown in **Figure 1g**. More TPRP are predicted to exist than were identified here, which could be explained by their relatively low abundance, comparably poor ionization efficiency or insufficient chromatographic separation.[37]

Besides TPRP and TRP, two neuropeptides belonging to the AST-A family, ADPYAFGLa and PRNYAGFLa, and five RFamides (PSLRLRFa, PSMRLRFa, SMPSLRLRFa, APQRNFLRFa and QDLDHVFLRFa) were identified based on the HRAM MS together with the rich fragment ions

delivered by the complementary fragmentation techniques using “biomarker” search.

Semi-Automated *de novo* Sequencing of Novel *P. interruptus* Neuropeptides

Although the ProSight “biomarker” search of high quality HRAM MS data against the known neuropeptides or sequenced preprohormones generates a number of neuropeptide hits, the second type of search, the “absolute mass” search, enhances novel neuropeptide discovery rate for *P. interruptus*. This search consists of two searches that tolerate a 20 Da and 200 Da mass difference, respectively, between the detected mass of unknown peptides with the theoretical mass of peptides in CNP database, provided at least 5 fragment ions are matched within a mass error of 10 ppm. This flexible scoring algorithm thus retrieves known neuropeptides from the CNP database that displayed similar fragmentation patterns as the unknown ions, accommodating the sequence variations that arise from amino acid variation, PTMs or truncation of the previously known peptide entries. After the “absolute mass” search displayed the matches, the results were taken to Sequence Gazer, a manual peptide/protein characterization tool, to investigate the site of variation including substitution, addition or deletion of amino acid sequence of the *P. interruptus* neuropeptide compared to the known one present in CNP database. This semi-automated *de novo* sequencing strategy took the advantage of great sequence conservation in decapod neuropeptide isoforms, significantly reducing the labor-intensive workloads of manual *de novo* sequencing and improving the certainty of identification since a cross-species comparison is done. In addition, peptides of large size, such as the TPRP in **Figure 1a,** that pose challenges for manual *de novo* sequencing could also be matched with peptides in CNP database as long as consensus sequences are identified within the query peptide.

A representative example is shown in **Figure 2**, where the full sequence of decapod calcitonin-like diuretic hormone (CLDH) was first characterized by MS without the knowledge of its cDNA sequence. **Figure 2a** displays a representative HCD MS/MS spectrum of the *P. interruptus* CLDH detected at the

m/z of 986.1756 (z=3). The rich fragment ions produced from this ion and the flexible scoring algorithm employed in ProSight "absolute mass" search led to the resulting match with the American lobster *Homarus americanus* CLDH cDNA sequence of GLDLGLGRGFSGSQAAKHLMGLAAANFAGGPa. When compared with the observed fragmentation pattern, a series of b ions, from b_2 to b_{30} with the exception of b_{17}, predicted from the *H. americanus* CLDH sequence were all matched to the dominant peaks in the MS/MS spectrum. Nevertheless, a peak with an *m/z* difference of 81.533 was observed following the b_{26} ion as shown in **Figure 2b**. The zoomed-in view in **Figure 2c** exhibits the ion's completely resolved isotopic pattern and its monoisotopic mass calculated as 2656.349 Da, delivering an incremental mass of 163.066 Da (corresponding to tyrosine) rather than 147.068 Da (corresponding to phenylalanine) following b_{26}. This Δm of 7.999 between the observed and predicted doubly charged b_{27} ion could thereby be explained by the replacement of amino acid residual phenylalanine (F) as encoded in *H. americanus* CLDH sequence by tyrosine (Y). Moreover, the consecutively assigned b_{28}, b_{29} and b_{30} ions as in **Figure 2a** all show an *m/z* difference of 7.993, 7.992 and 7.993 from the predicted values, validating the site of amino acid variation as the 27^{th} residue. Therefore, the *P. interruptus* CLDH sequence was deduced as GLDLGLGRGFSGSQAAKHLMGLAAAN**Y**AGGPa, whereas the monoisotopic mass of its b_{27} ion is expected as 2656.349 Da, agreeing well with the observed mass (**Figure 2c**). This change could correspond to a single nucleotide change in the gene encoding CLDH, as tyrosine is encoded by UAU or UAC and phenylalanine is encoded by UUU or UUC, since such a vital hormone is unlikely to be changed drastically between these two related species. Our result presented here is the first observation of mature CLDH without prior knowledge of the corresponding cDNA sequence, demonstrating the conservation of neuropeptide sequences in decapods and the robustness of the HRAM MS/MS technique and the partnered semi-automatic *de novo* strategy.

CLDH has been previously identified in insects including *Rhodnius prolixus*[38], *Daploptera punctata*[39], *Drosophila melanogaster*[40] etc., and crustaceans such as *Balanus amphitrite, Daphnia pulex*[41] and *H. americanus*[36]. Functional studies of CLDH have shown that this family is implicated in the control of diuresis in insects. CLDH also intrinsically modulates the cardiac neuromuscular system in *H. americanus*, and thus has been proven to be bioactive in Decapoda[36]. Interestingly, a high degree of sequence similarity and a C-terminal amidation are observed in different isoforms of this 31 amino acid in insects and crustaceans, as illustrated in **Figure 2d**. The extensive sequence homology among different species supports the postulation that not only CLDHs but also their receptors are conserved, and might play critical roles across evolution.

Similarly, two ions were detected and matched with the preproneuropeptide F (NPF)-derived peptides (NPFP) predicted from *Penaeus vannamei* and *Melicertus marginatus* (UniProt: F6KM62 and F6KM63) EST transcripts, SDYPMPSGDALMEASERLLET and SDYPMPSGDALMEASERLLETA[26], both with a Δm of 7.94 Da. Manual characterizations were performed in Sequence Gazer on these two ions, deducing the actually detected peptides' sequences as SDYP**L**P**P**GDALMEASERLLET and SDYP**L**P**P**GDALMEASERLLETA, both with a substitution of M to L at the 5th position and S to P at the 7th position. Prior to our study, no authentic NPF isoforms or NPFP have been identified or validated from crustacean species using a MS-based platform. The NPF are viewed as the invertebrate homologs of the vertebrate hormone neuropeptide Y (NPY), which has been associated with the regulation of appetite and feeding behavior[42]. It is noteworthy that the administration of vertebrate NPY dramatically increased food intake in invertebrate *Marsupenaeus japonicus* as well[43], suggesting a conserved role for NPY and possibly a degree of sequence homology in crustacean species, most likely in the part of the hormone that interacts with its receptor. Intriguingly, conservation of the NPF sequence in crustaceans is further evidenced by the sequence homology of the truncated NPF and NPFP

between *P. interruptus* and two other species, *Penaeus vannamei* and *Melicertus marginatus*, that represent both a derived and a basal taxon. The physiological significance of this conservation awaits discovery. It will be interesting to investigate the impact of endogenous NPF molecules on food intake and growth in crustacean species, especially those of economic value to the aquaculture industry. In addition to this potentially important discovery, our work provides a framework for future investigations of NPF and its physiological functions in crustacean species in that it characterizes this important neuropeptide without prior knowledge of its sequence.

Other than the flexible searching algorithm to account for the unknown sequence variation not housed in the CNP database, our semi-automatic *de novo* sequencing strategy also enables "multiplexed search," allowing for identification of multiple peptide precursors using multiplexed MS/MS spectra,[44] also known as chimeric spectra. A mass window of ± 7.5 Da was chosen to isolate peptide precursors subjected to MS/MS . This significantly wide isolation window improves sensitivity in hybrid LTQ Orbitrap without sacrificing identification confidence, since the ProSight multiplexing search option is designed to handle chimeric MS/MS spectra with the aid of HRAM MS/MS data. **Figure 3** highlights the identification of two novel neuropeptide isoforms of SIFamide simultaneously from one multiplexed search in ProSight. In **Figure 3a**, two peptides of *m/z* 465.9091 and 466.5807 with overlapping isotopic distributions and another peptide at *m/z* 471.9124 were observed in a MS precursor scan. Appropriate assignments of their charge state and the resulting monoisotopic masses were automatically done by cRAWler. The resulting deconvoluted masses of 1396.722, 1394.709 and 1412.716 Da were then searched independently in ProSight using all the HRAM MS/MS fragments acquired in the single ETD MS/MS spectrum as shown in **Figure 3b**, all leading to a homologous match with a highly conserved decapod crustacean neuropeptide GYRKPPFNGSIFa[30]. With manual examination, the fragment ions of the peptide at 1396.722 Da were matched to c_6 to c_{11} of GYRKPPFNGSIFa . For the z

ions, a consistent increase of 16.00 Da was observed for all of the assigned z ions, including z_6, z_9, z_{10} and z_{11}. This suggests an amino acid substitution of F to Y at the 12th position, and thus the deduced sequence is GYRKPPFNGSI**Y**a as in **Figure 3c**. Similarly, the other SIFamide isoform of 1412.716 Da is suggested to have another Y at the 7th position and sequenced as GYRKPP**Y**NGSI**Y**a due to the observation of a constant Δm 15.994 Da between several pairs of fragment ions from c_7 to c_{11} but an overlapping c_6 when compared to the fragment ions of GYRKPPFNGSIYa. In addition, the mass difference of 15.994 Da between all the observed z ions including z_6, z_9, z_{10} and z_{11} when compared to GYRKPPFNGSIYa further validates our deduced sequence based on c ions. No other sequences or fragment ions could possibly explain the peptide with a mass of 1394.709 Da, suggesting it as an intermediate with loss of water from GYRKPPYNGSIYa. This work thererefore, demonstrates that multiplexing fragmentation and tailored software for chimeric spectra interpretation allow for conclusive identification of two peptides from a single MS/MS isolation. The HRAM MS and MS/MS scans thereby allow one to confidently identify multiple peptides per DDA MS/MS spectra, increasing the absolute number of peptides identified without a compromise in duty cycle.

Intriguingly, the previously characterized neuropeptide GYRKPPFNGSIFa (Gly-SIFamide) was also identified in *P. interruptus*, agreeing well with its conserved presence in other decapod crustaceans including crabs and crayfish[30]. It is striking that the substitution of F to Y was observed in the two SIFamides as well as CLDH in *P. interruptus*. Future studies that isolate and sequence the cDNA encoding the SIFamides from this species could validate the amino acid substitution observed from *P. interruptus* neuropeptidome using a MS-based approach in this work. In addition to Gly-SIFamide, another SIFamide isoform, VYRKPPFNGSIFa (Val-SIFamide), a truncated form RKPPFNGSIFa, and two incompletely processed forms GYRKPPFNGSIF and GRYKPPFNGSIFG were detected from *P. interruptus*. Collectively, our strategy not only significantly expanded the catalog of SIFamide family

peptides in decapod crustaceans to contain NPs with the C-terminal sequence motif SIYamide, but also laid a foundation for future functional studies of various *P. interruptus* neuropeptide isoforms in this important species.

Discovery of a Novel Motif of Orcokinin Family

Although the hybrid LTQ Orbitrap and the tailored database search algorithm enables semi-automatic *de novo* sequencing of 37 neuropeptides, including 21 novel ones, manual *de novo* sequencing of data acquired on SYNAPT G2 HDMS with the aid of proprietary software PepSeq also generates a list of 38 neuropeptides with 19 ones that have not been reported before. Taken together with the novel identifications achieved on LTQ Orbitrap, this work has revealed a novel sequence motif for the decapod crustacean orcokinin neuropeptide family. A myriad of orcokinins have previously been discovered from crustaceans, insects and other invertebrate species with a conserved N-terminal motif of NFDEIDR. More specifically, decapod orcokinins share a consensus sequence of NFDEISRSX(G/S)FGFX(H/N/V/A) (S-orcokinin) as in **Figure 4a,** with the exception of the Asp-Orcokinins observed in the blue crab *Callinectes sapidus*[28]. Interpretation of MS/MS spectra produced by orcokinin peptides relied heavily on the y ion series, due to this N-terminal consensus sequence. After sequential loss of the conserved residues from the N-terminus that can be viewed as fingerprints of the orcokinin family, information regarding the variable sequence appearing at the C-terminus is retained in the remaining y ions. Based upon accurate mass measurement and the characteristic fingerprints provided by y ions, we identified nine orcokinin peptides sharing a motif of NFDEIDRAGX(F/L/I/V)X(G/A/V)FX(H/N) (A-orcokinin) as summarized in **Figure 4b** (L is used in the figure for the purpose of alignment). In **Figure 4c**, the clean y_2-y_7 ion series made direct reading of the N-terminus sequence NFDEID from the MS/MS spectrum possible. The accurate mass measurement of the y_2 ion determines the C-terminus combination as RA, indicating a novel truncated

form of orcokinin, NFDEIDRA. The fingerprints of unique fragmentation patterns belonging to orcokinin family members further facilitated our identification in the representative MS/MS spectra shown in **Figure 4d-f**. In **Figure 4d**, the high-abundance y_5 to y_{10} ions immediately reveal its identity as orcokinin with an N-terminus sequence NFDEID, whereas the remaining y_1 to y_4 ions determined that the rest of C-terminus variable sequence is RAGL/IA, despite these ions being detected at relatively low abundances. Similarly, in **Figure 4e**, the y_4 to y_{10} ions again exhibit the fingerprint fragmentation pattern of orcokinins, delivering the conserved N-terminus sequence NFDEIDRA, whereas y_1 to y_3 ions aid in deducing the complete sequence as NFDEIDRAGL/IGF. A complete series of y ions with relatively high abundance starting from y_1 to y_{12} is also presented in **Figure 4f**, supporting the identity of this peptide as another orcokinin variant with a sequence of NFDEIDRAGL/IGFH. Nevertheless, it is striking that the MS/MS fragmentation pattern of the orcokinin NFDEIDRAGL/IGFH appears distinct from the orcokinin with a single amino acid variation at the C-terminus, NFDEIDRAGL/IGF. As shown in **Figure 4f**, the b ion series from b_4 to b_{11} were not detected in NFDEIDRAGL/IGFH, although a number of internal fragments were observed and assigned based on the deduced sequence. These internal fragments complicated the MS/MS spectrum, raising the challenge to interpret the fragment ions even with the assistance of the *de novo* sequencing software PepSeq.

Interestingly, a previous study by Stemmler et al. reported the approach of use metastable decay and sustained off-resonance irradiation to identify orcokinin neuropeptides in crustaceans[33]. They identified the orcokinin isoform NFDEIDRAGLGF in *P. interruptus*, which is in agreement with our newly reported motif, and predicted the presence of more orcokinin variants that might show distinct C-terminus variation from previously reported orcokinin peptides[33]. To the best of our knowledge, our study with multiple mass spectral techniques presented here demonstrates the most comprehensive repertoire of crustacean orcokinin variants that contain this novel motif, NFDEIDRAG, including

NFDEIDRAGLG, NFDEIDRAGLA, NFDEIDRAGVV, NFDEIDRAGFG, NFDEIDRAGFA, NFDEIDRAGLAFNa, a truncated form NFDEIDRAGLa and the three orcokinins shown in **Figure 4**. The specificity of this orcokinin motif to *P. interruptus* is noteworthy and a plausible explanation is a single codon changed on orcokinins for this species from UCU/C/A/G (corresponding to S) to GCU/C/A/G (corresponding to A). Simply changing the first nucleotide of the codon encoding the amino acid at that position from a U to a G could lead to subsitution of A for S in the final sequence. This may suggest that A-orcokinins have another physiological function different from the S-orcokinins, ie that *P. interruptus* has developed another pathway in which they are involved in signaling. Alternatively, this could also mean that the A-orcokinins are simply redundant extra copies of the orcokinin gene with a small error that occured in DNA replication in an ancestor that was propagated down to current day. In this case, the extra set of A-orcokinins might simply amplify the signal of the other orcokinins. This would need further exploration on whether the A-orcokinins have different receptor affinities and/or different receptors and/or downstream targets of the receptors. This is expected to generate interesting information regarding the impact of structural variance on neuropeptide biological functions in decapod crustaceans for people who study molecular evolution and speciation.

Dimethylation-Assisted Fragmentation (DAF) and *de novo* Sequencing

A reductive N-terminal dimethylation labeling method has previously been shown to facilitate the fragmentation and ultimately *de novo* sequencing of unknown neuropeptides[27, 45]. A schematic illustration of the reaction with formaldehyde and reducing reagent borane-pyridine complex is shown in **Figure 5a**. Briefly, the efficient and quick derivatization method labels a peptide's N-terminus and ϵ-amino groups of lysine, if any are present, through reductive dimethylation, resulting in an incremental mass of 28.03 Da for each derivatized site[45]. Moreover, substantial signal enhancement of the a_1 ion after labeling can be utilized to resolve sequence ambiguity in the N-terminus, since the b_1 ion is usually

missing in a MS/MS spectrum[27]. In addition, the labeling method has also been employed to facilitate sequencing of singly charged peptides by increasing the signal intensities of the a- and b-ion series and reducing the complexity of fragmentation patterns.[27] We observed the same simplification effect in the interpretation of native orcokinin NFDEIDRAGL/IGFH with the approach of DAF. As previously shown in **Figure 4f**, a highly complex MS/MS spectrum was obtained for this peptide. Upon labeling, the mass of the native peptide at *m/z* 745.85 (z = +2) was shifted to *m/z* 759.86 (**Figure 5b**), corresponding to the mass difference induced by N-terminus dimethylation. Intriguingly, the formaldehyde labeled orcokinin displayed dramatically cleaner MS/MS spectrum, containing primarily b- and y-ion series, as shown in **Figure 5b**. Other than the complete y-ion series from y_1 to y_{12}, the b-ion series were significantly enhanced, resulting in detection of b_4, b_5, b_7, b_8, b_{10}, b_{11}, b_{12} ions, all of which were missing in the unlabeled sample. In addition to enhancing b ions, internal fragments were suppressed, and thus a simplified fragmentation profile was obtained, enabling the identification of this novel orcokinin with remarkable confidence and ease.

Unambiguous identification of amino acid residues at the N-terminus of the sequence could be achieved via the DAF method as well. As an example, **Figure 5c** shows the MS/MS sequencing of an RFamide QDLDHVFLRFa, a peptide that has a close mass to KDLDHVFLRFa and isobaric mass with AG(or GA)DLDHVFLRFa. A 28.05 Da difference was detected upon labeling, as indicated by the mass shift from *m/z* 430.22 to 439.58 (z = +3) shown as the insert. This excludes the possible existence of lysine in the peptide, as it too would have been dimethylated for a total mass shift of 56.06 Da. Although the combined mass of two amino acid residues A+G is equal to that of Q (129.10 Da), no a_1 ion corresponding to G (58.07 Da) or A (72.08 Da) was seen even though the dimethylated a_1 ion should be significantly enhanced upon labeling. In contrast, an a_1 ion with *m/z* corresponding to dimethylated Q was detected with high abundance at *m/z* 129.10, supporting our assignment of this

peptide as N-terminally labeled QDLDHVFLRFa.

N-terminal pyroglutamate and pyroaspartate modifications are among the most common PTMs in neuropeptides. The reductive dimethylation method aids in the characterization of neuropeptides with N-terminal blockage. Peptides with pyroglutamate modification, as shown in **Figure 5d**, can be therefore easily assigned based on their inability to react at the N-terminus due to the lack of free N-terminal amine group. After labeling, identical mass and similar retention times were observed compared to the native sample for the peptide at *m/z* 636.32 in the labeled sample, suggesting the presence of an N-terminal blockage in this peptide. Together with the accurate mass measurement of b- and y-ion series, particularly a_1, the peptide was confidently assigned as pQDLDHVFLRFa. Other than pyro- modification, the application of reductive dimethylation also applies to N-terminal acetylation. A representative example is the novel peptide sequenced as (Ac)GYRKPPFNGSIFa with a mass of 1422.74 Da. However, it possesses a mass close to the previously reported VYRKPPFNGSIFa, which is found at 1422.78 Da. This ambiguity could be clarified by the observation that it is theonly peptide with a single dimethylation at K rather than the expected dimethylation at both N-terminus and K, due to the N-terminal blockage. Nevertheless, it is worth noticing that this peptide was identified unambiguously without labeling relying on the HRAM data provided by LTQ-Orbitrap.

The observed enhancement in MS/MS interpretation facility after reductive dimethylation could possibly be explained by the adjusted proton affinities (PA) of the fragments[46]. Upon labeling, the primary amine group at the N-terminus becomes a tertiary amine, possessing higher PA compared to the native form with primary amine. Therefore, the abundances of b-ions are generally improved in the resulting MS/MS spectra of labeled peptides compared to the unlabeled one. Moreover, internal fragments with a minimum of one b-type and one y-type bond cleavage are produced with greater

frequency when samples are not labeled[27]. With the increased PA in the N-terminus, the fragmentation of y-type bonds is suppressed, resulting in a cleaner MS/MS fragmentation pattern. Overall, the utilization of DAF for *de novo* sequencing not only provided validation for our assignments of *P. interruptus* neuropeptides based on their unmodified forms, but also resolved some sequence ambiguity in the N-terminus or in the middle of the peptide, yielding identification with improved confidence on a mass spectrometer with medium spectral resolution and mass accuracy.

To further improve our detection coverage and strengthen the confidence of our assignments, an offline CE-MALDI-MS platform was utilized to fractionate the brain extract to reduce signal suppression induced by the large number of peptides present with sample consumption of merely nanoliters. As demonstrated in Supplemental Figure **Figure S1**, the sample background was much cleaner upon separation as demonstrated by the representative fractions #8 and # 14 in **Figure S1a** and **1b** compared to the direct profiling of the extract in **Figure S1c**. All the peptides profiled via CE-MALDI-MS were indicted in Table 1, whereas those also detected by direct tissue analysis were noted as “d”.

Region-Specific Localization of Neuropeptides in the Brain of *P. interruptus*

Using our multi-faceted MS-based strategy, we have discovered a number of novel neuropeptides from *P. interruptus* that have shared homology yet variations in amino acid residues or PTMs from previously reported neuropeptides or preprohormones. Examples of these changes include the amino acid substitution of S to A in *P. interruptus* orcokinins, and F to Y substitutions in *P. interruptus* SIFamide and CLDH. To provide further evidence that the origins of these sequence variation result from genetic differences among species rather than artifacts from sample preparation procedures[25, 28], we performed direct tissue analysis on *P. interruptus* brain to examine the resulting peptide profiles with minimal sample preparation. As highlighted in **Figure 6a**, 8 orcokinins were observed, including

peptides that comply to the common motif NFDEIDRSG such as NFDEIDRSGFG (*m/z* 1256.55), NFDEIDRSGFA (*m/z* 1270.57), NFDEIDRSGFGFNa (*m/z* 1516.68) and NFDEIDESGFGFH (*m/z* 1540.68), and the orcokinins of the novel motif NFDEIDRAG such as NFDEIDRAGLG (*m/z* 1206.57), NFDEIDRAGFG (*m/z* 1240.56), NFDEIDRAGLGF (*m/z* 1353.64), and NFDEIDRAGLGFH (*m/z* 1490.70). Unfortunately, not all of the sequenced orcokinins were seen via direct tissue analysis on MALDI-TOF/TOF, probably due to their relatively low abundances and the complex sample matrix that is the unprocessed tissue. The complexity of this matrix is proven by the myriad of lipid peaks shown in **Figure 6a**. Nevertheless, the observation of these novel orcokinins directly from tissue validated their existence in the *P. interruptus* brain and thereby the novel motif we discovered. In addition, the substitution of F to Y exemplified by CLDH at *m/z* 2956.50 was also detected via direct tissue analysis of the *P. interruptus* brain in linear mode, excluding the possibility of the sequence variation due to sample processing procedures.

Using knowledge gained from pioneering studies of the brains of spiny lobsters dating back to the 1960's[47], substructures of the *P. interruptus* brain were examined in detail using direct tissue profiling Crustacean brains consist of more complex structures compared to other neuronal organs in crustacean species, including numerous areas of neuropil and neuronal clusters[31]. Previously, neuropeptides that displayed distinguishable distributions have been linked to different biological functions in crustacean species like *C. sapidus*[48] and *C. borealis*[24, 49]. Therefore, it is crucial to study the neuropeptides' localization in *P. interruptus* brain due to the correlations that can often be made between brain location and function. One interesting finding is that neuropeptides desorbed/ionized from region 1 (R1) in **Figure 6a** exhibited overlapping yet distinct peptide profiles compared to those from region 2 (R2) in **Figure 6b**. In **Figure 6a**, orcokinin peptides were detected at high abundances, whereas the SIFamide peptide at *m/z* 1381.74 was observed at a modest level when compared to the base peak APSGFLGMRa

at *m/z* 934.49 in R1. In contrast, SIFamide rose to a dominant peak in the profiling spectrum of R2, whereas orcokinins were detected at a low level in this region compared to R1. The differential neuropeptide patterns from these two regions' profiling spectra demonstrate the complexity of the *P. interruptus* brain structure and the potentially distinct regulatory roles neuropeptides play in different regions of the brain.

Mapping Distribution of Various Brain Neuropeptides via MALDI-MSI

Direct tissue analysis yields "snapshots" of neuropeptide profiles corresponding to targeted regions in *P. interruptus* brain, providing information about spatial distribution of neuropeptides in a high throughput manner. Nonetheless, the need to document detailed and accurate localizations of neuropeptides in a complex and intricate structure like brain cannot be satisfied with direct tissue analysis. Alternatively, MSI, a newly emerged technique, has demonstrated its capability to characterize and localize neuropeptides from neural tissue of heterogeneous structure at high spatial resolution[17, 22, 50-52]. The high-resolution mapping of *P. interruptus* brain neuropeptides is thus accomplished here via MSI with representative neuropeptides from major families, tachykinin, SIFamide, orcokinin, AST-A and RFamide, shown in **Figure 7**. In **Figure 7a**, the TRP neuropeptide of *m/z* 934.49, detected at high abundance in both regions in **Figure 6,** is shown to be distributed throughout the major regions in brain, including anterior medial protocerebral neuropils (AMPN), accessory lobes (AcN), and olfactory lobes (ON) as illustrated by the rostral view of *Panulirus* brain (**Figure 7**)[31]. Its distribution in AcN and ON agrees well with our previous knowledge that decapod TRP has been identified as a neuromodulator in the olfactory neural pathway[53]. Interestingly, two SIFamide peptides, GYRKPPFNGSIFa at *m/z* 1381.74 (**Figure 7b**) and (Ac)GYRKPPFNGSIFa at *m/z* 1423.75 (**Figure 7c**), show concentrated localization in the AcN, which agrees well with its significantly higher abundance in R2 compared to R1 as displayed in **Figure 6**. In contrary, orcokinin peptides at

m/z 1254.47 and 1490.70, also revealed by regional profiling results in **Figure 6**, exhibited higher level of abundance in the ON compared to AcN and AMPN in **Figure 7d-e**. The co-localization of various neuropeptide isoforms from the same family, particularly in the AcN and ON, and potential links to function warrant further investigation. Moreover, the AST-A peptide ADPYAFGLa at *m/z* 852.43 also displays a higher concentration in ON, illustrated in **Figure 7f**. In addition, the RFamide SMPSLRLRFa at *m/z* 1105.63 localized mostly in AcN and ON, yet present in lower quantities in the AMPN in **Figure 7g.** This observation is also in agreement with its presence in both R1 and R2 profiling spectra.

The MALDI-MSI results successfully mapped neuropeptides of different families in the brain, presenting useful and interesting spatial information about endogenous neuropeptides and providing insights into their regulatory functions in the brain of *P. interruptus*. OL and AcN are dominant deutocerebral neuropils (DCN), and are best visualized in the rostral view. A unique feature of *P. interruptus* brain is that the large AcNs are located medial to the slightly smaller ONs, an organization that is distinct from that previously characterized in the lobster *Homarus americanus* [25] and other decapod species like crayfish or crabs[31]. OL receives primary olfactory input from olfactory receptor neurons on antennae, whereas AcN not only receives secondary olfactory signals from the OL but also higher multimodal signals through synaptic contact with neurons derived from tactile and visual sensory systems[53]. Therefore, the distributions of neuropeptides within DCN suggest their possible role as neuromodulators involved in the function of the olfactory system, the tactile system, and/or the visual sensory system. An alternative function could be in the integration of multiple sensory modalities. Moreover, the information about colocalization of neuropeptides gained in this study implies a relationship between their functions—those that are found in close proximity to each other may have similar or related roles in cell-cell signaling. Further in-depth study might provide a clear answer

regarding the possibly synergistic, complementary, or antagonistic relationships that could be present among these co-localized peptide modulators.

To the best of our knowledge, our study presents the first investigation of the spatial distribution of neuropeptides in the brain of *P. interruptus*. Although MSI generally has a limitation in resolution compared to conventional immunohistochemical methods, the capability to distinguish neuropeptide isoforms, exemplified by SIFamide peptides in **Figure 7b-c** and orcokinin peptides in **Figure 7d-e**, in a high throughput manner enables unambiguous mapping of neuropeptides in large-scale. In addition, the overlay feature provides great ease for highly multiplexed simultaneous study of neuropeptide localization (**Figure 7h**) and thereby may provide information on the relationship of their functions.

CONCLUSION

In this study, we employed a suite of mass spectrometric approaches for the discovery of neuropeptides in brain of *P. interruptus*, an important aquaculture species that has not been extensively characterized. Collectively, 55 neuropeptides were sequenced in this work, including 34 novel ones that are *de novo* sequenced. The use of HRAM MS/MS data with various fragmentation methods (CID, HCD and ETD) and tailored ProSight search against our home-built CNP database highlighted the possibility of streamlining the peptide discovery process with highly confidnet identifications. Although the hybrid LTQ-Orbitrap excels in the detection of larger peptides, Q-TOF provides complementary coverage in interpretation of medium-sized neuropeptides. Furthermore, the application of DAF aids in the *de novo* sequencing process and improves identification certainty of neuropeptides on a medium-resolution instrument. In addition, direct tissue analysis was applied to analyze the brain tissue with minimal sample preparation, better preserving native neuropeptide expression profiles in *P. interruptus* brain. Finally, MALDI-MSI was employed to map the distribution

of multiple neuropeptides simultaneously from a brain slice to yield information on localization that may be important in determining the function of these NPs. In summary, our study not only presents a comprehensive characterization of neuropeptide expression and distribution in the brain of *P. interruptus* and provides insights for physiological functions of these endogenous neuropeptides, but also demonstrates the application of a multi-dimensional MS-based platform that will be useful for future neuropeptidome discovery studies on crustacean species without the prerequisite of known preprohormones.

ACKNOWLEDGEMENT

The authors wish to thank Professor Deborah Baro at the Georgia State University for the critical reading and insightful suggestions of the manuscript. This work is supported in part by the National Institutes of Health grants (R01DK071801 and 1S10RR029531 to LL) and the National Science Foundation grant (CHE-1413596 to LL). N.L.K. thanks support from NIH Grant No. R01 GM067193. L.L. acknowledges an H. I. Romnes Faculty Research Fellowship and a Vilas Distinguished Achievement Professorship with funding provided by the Wisconsin Alumni Research Foundation and University of Wisconsin- Madison School of Pharmacy. The authors also wish to thank Bruker Daltonics for graciously loaning the Autoflex III MALDITOF/TOF mass spectrometer.

FIGURE LEGNEDS

Figure 1. Sequencing of *P. interruptus* TRP and TPRP with the aid of ProSight. (a) A HRAM MS/MS spectrum of *P. interruptus* TPRP acquired in HCD. (b) Fragment ion assignment of the TPRP MS/MS spectrum shown in (a). (c) Zoomed-in raw MS spectrum of the precursor TPRP that shows the HRAM *m/z* information. (d) Sequence coverage of the TPRP, combining fragments observed in HCD, CID and ETD spectra. (e) An illustration of how the fragment ions of HCD and CID (b, y ions) and ETD (c,z ions) are produced and annotated in the fragmentation map in (b) and (f). (f) A HCD spectrum of *P. interruptus* TRP and its assignment of fragmentation. (g) Previously reported *P. interruptus* preprotachykinin cDNA sequence with detected peptides highlighted in different colors.

Figure 2. *De novo* sequencing of *P. interruptus* CLDH. (a) A HRAM CID MS/MS spectrum of *P. interruptus* CLDH observed at *m/z* 986.1756. Its fragmentation map and sequence coverage are shown in the insert. (b) The accurate assignment of amino acid substitution at position 27 in CLDH by registering the mass discrepancy observed in ProSight to b_{27} ion in the raw MS/MS spectrum. (c) A zoomed-in view of the well-resolved b_{27} ion detected in the original CID spectrum. (d) Sequence alignment of the newly sequenced *P. interruptus* CLDH with previously reported CLDH from other species, highlighting the sequence conservativeness of this peptide family. Species names in purple indicate insects, blue indicates crustacean. Sequence letters in green are shared between species.

Figure 3. *De novo* sequencing of two novel P. interruptus neuropeptides as homologues of SIFamide. (a) A zoomed-in MS spectra showing the calulcated deconvoluted masses of the three precursors identified in single MS/MS event with ProSight search with the ion at *m/z* 465.9091 assigned as the product of neutral loss from the ion at *m/z* 471.9124. Isotopic peaks from different precursors are indicated with colored dots. (b) The multiplexed MS/MS spectra that exhibits the fragmentation pattern of the two neuropeptides, with the fragments coming from each color identified by

corresponding colors. (c) The fragmentation map and sequence coverage of the two SIFamide homologues.

Figure 4. Discovery of a novel crustacean orcokinin motif from *P. interruptus* orcokinins. (a) A consensus sequence of NFDEISRSX(G/S)FGFX(H/N/V/A) shared by crustacean orcokinin neuropeptides. (b) The novel motif NFDEIDRAG summarized based on *P. interruptus* orcokinin neuropeptides. (c)-(f) *De novo* sequencing of representative A-orcokinins NFDEIDRA, NFDEIDRAGLA and NFDEIDRAGLAF, NFDEIDRAGLGFH with MS/MS spectra acquired on Q-TOF.

Figure 5. Reductive dimethylation-assisted *de novo* sequencing of *P. interruptus* neuropeptides. (a) A schematic illustration of the formaldehyde labeling reaction. (b) *De novo* sequencing of dimethylated orcokinin NFDEIDRAGLGFH. (c) *De novo* sequencing of dimethylated RFamide QDLDHVFLRFa. (d) *De novo* sequencing of RFamide pQDLDHVFLRFa that cannot be labeled due to N-terminus blockage.

Figure 6. Direct tissue analysis of *P. interruptus* brain region 1 and 2. (a) Profiling spectrum of neuropeptides desorbed/ionized from region 1 (R1) that exhibited overlapped yet distinct peptide profiles to that from region 2 (R2) in (b).

Figure 7. MSI of *P. interruptus* brain shows the localization of several neuropeptides from tachykinin, SIFamide, orcokinin, AST-A and RFamide families. The distribution maps of neuropeptides visualized by MSI technique shown in (a)-(g) could be correlated with the schematic illustration of *Panulirus* brain in rostral view. (h) An overlay of the distribution maps of a SIFamide at m/z 1381.74 in (b) and an orcokinin at m/z 1490.70 in (e).

Figure 1. Sequencing of *P. interruptus* TRP and TPRP with the aid of ProSight.

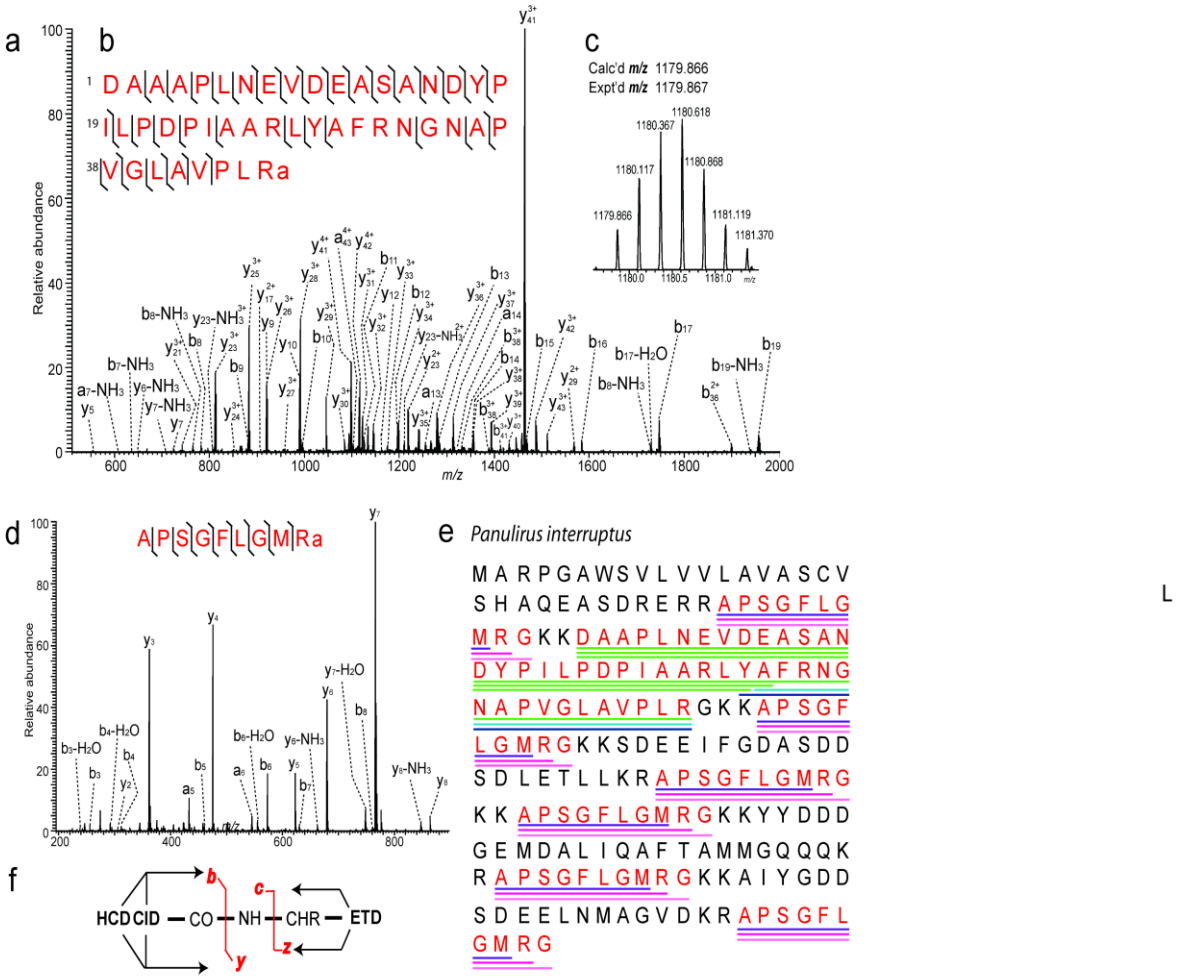
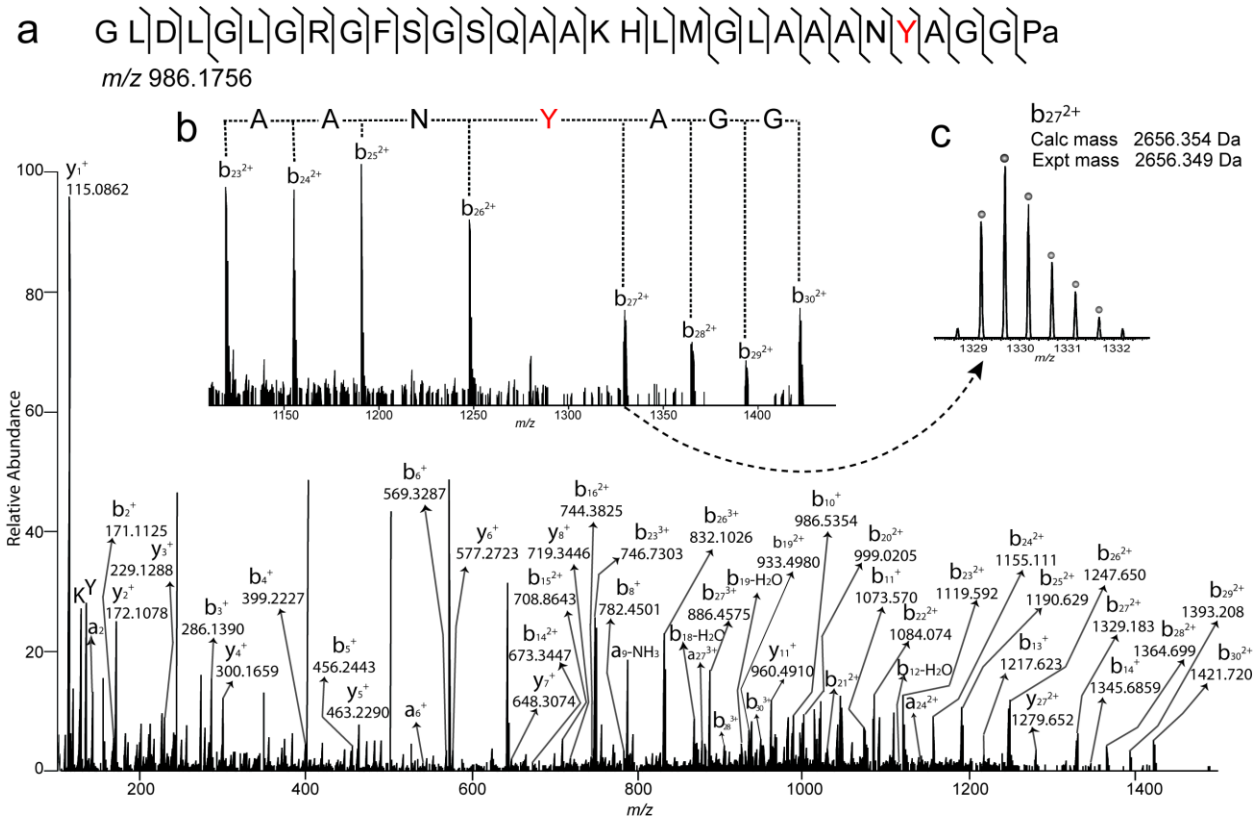


Figure 2. De novo sequencing of *P. interruptus* CLDH.



d

Rhodnius prolixus	F2VU11	GLDLGLSRGFSGSQAAKHLMGLAAANYAGGPa
Balanus amphitrite	I3VN98	GLDFGLGRGFSGSQAAKHMMGLAAANFAGGPa
Drosophila melanogaster	Q9VLK4	TVDFGLARGYSGTQEAKHRMGLAAANFAGGPa
Daploptera punctata	P82373	GLDLGLSRGFSGSQAAKHLMGLAAANYAGGPa
Daphnia pulex	———	GVDFGLGRGYSGSQAAKHLMGLAAANYAIGPa
Homarus americanus	D2IJD5	GLDLGLGRGFSGSQAAKHLMGLAAANFAGGPa
Panulirus interruptus	———	GLDLGLGRGFSGSQAAKHLMGLAAANYAGGPa

Figure 3. *De novo* sequencing of two novel P. interruptus neuropeptides as homologues of SIFamide.

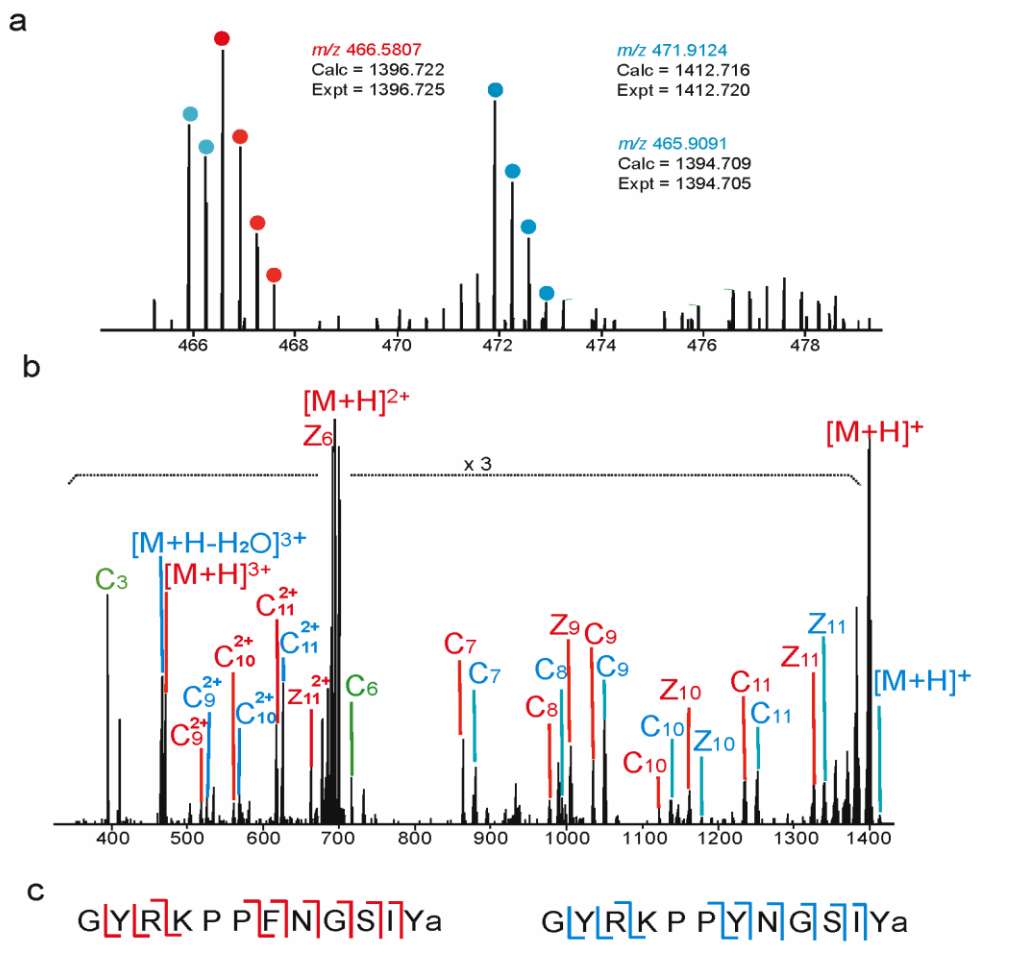


Figure 4. Discovery of a novel crustacean orcokinin motif from *P. interruptus* orcokinins.

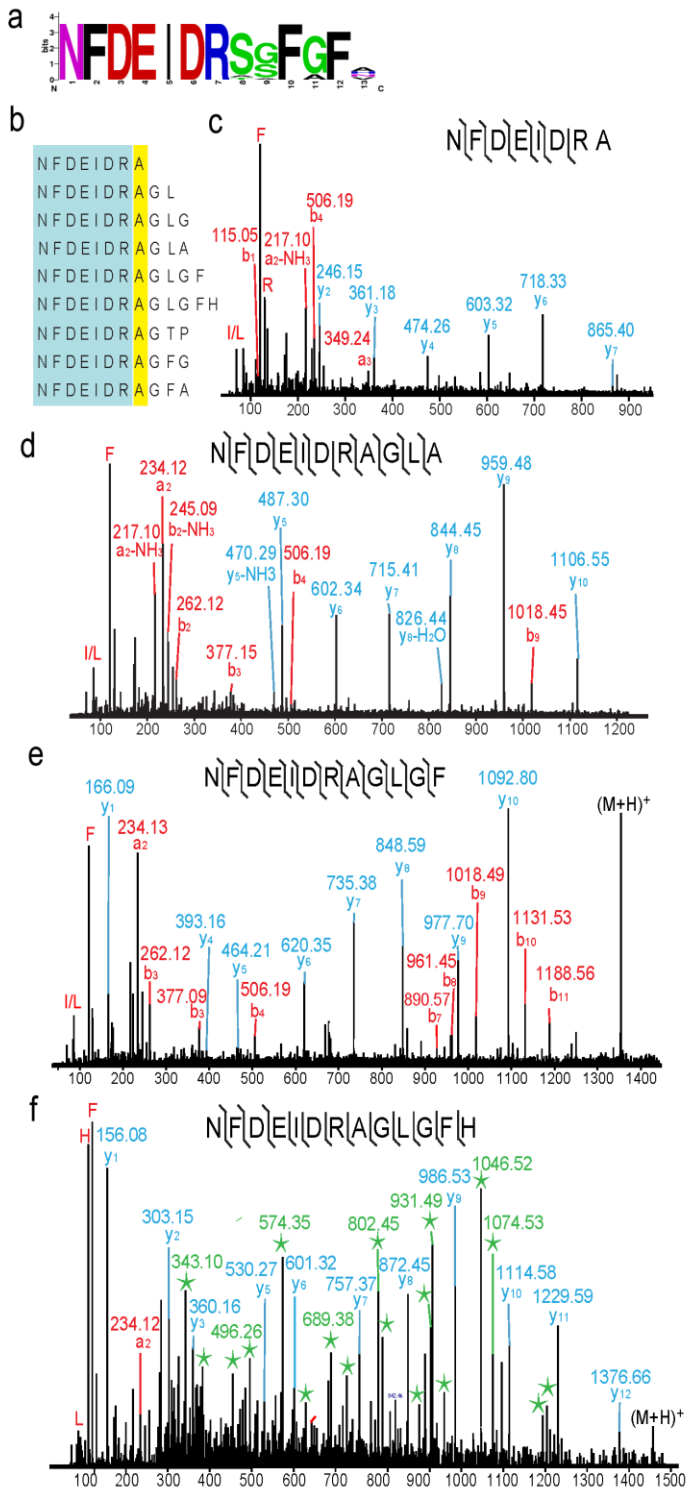


Figure 5. Reductive dimethylation-assisted *de novo* sequencing of *P. interruptus* neuropeptides.

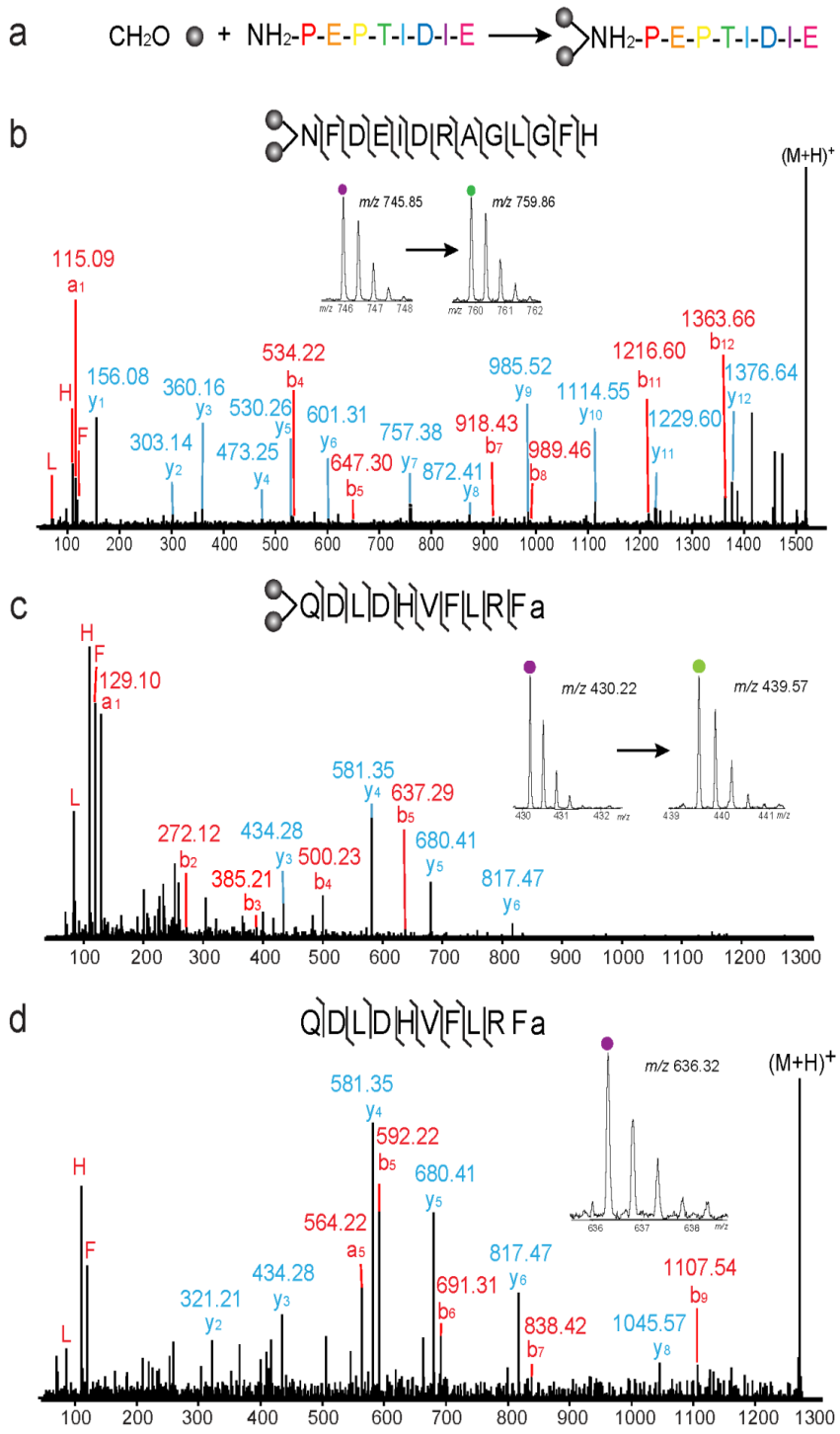


Figure 6. Direct tissue analysis of *P. interruptus* brain region 1 and 2.

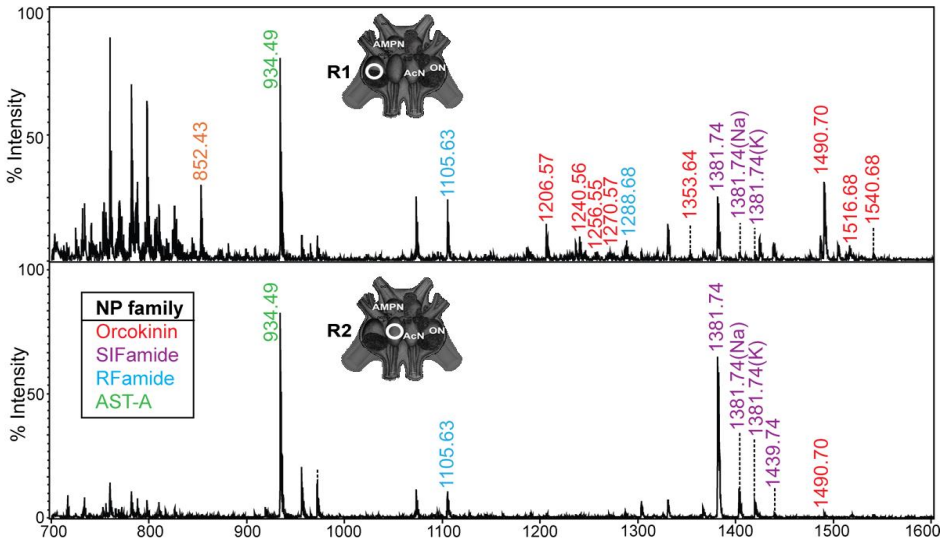
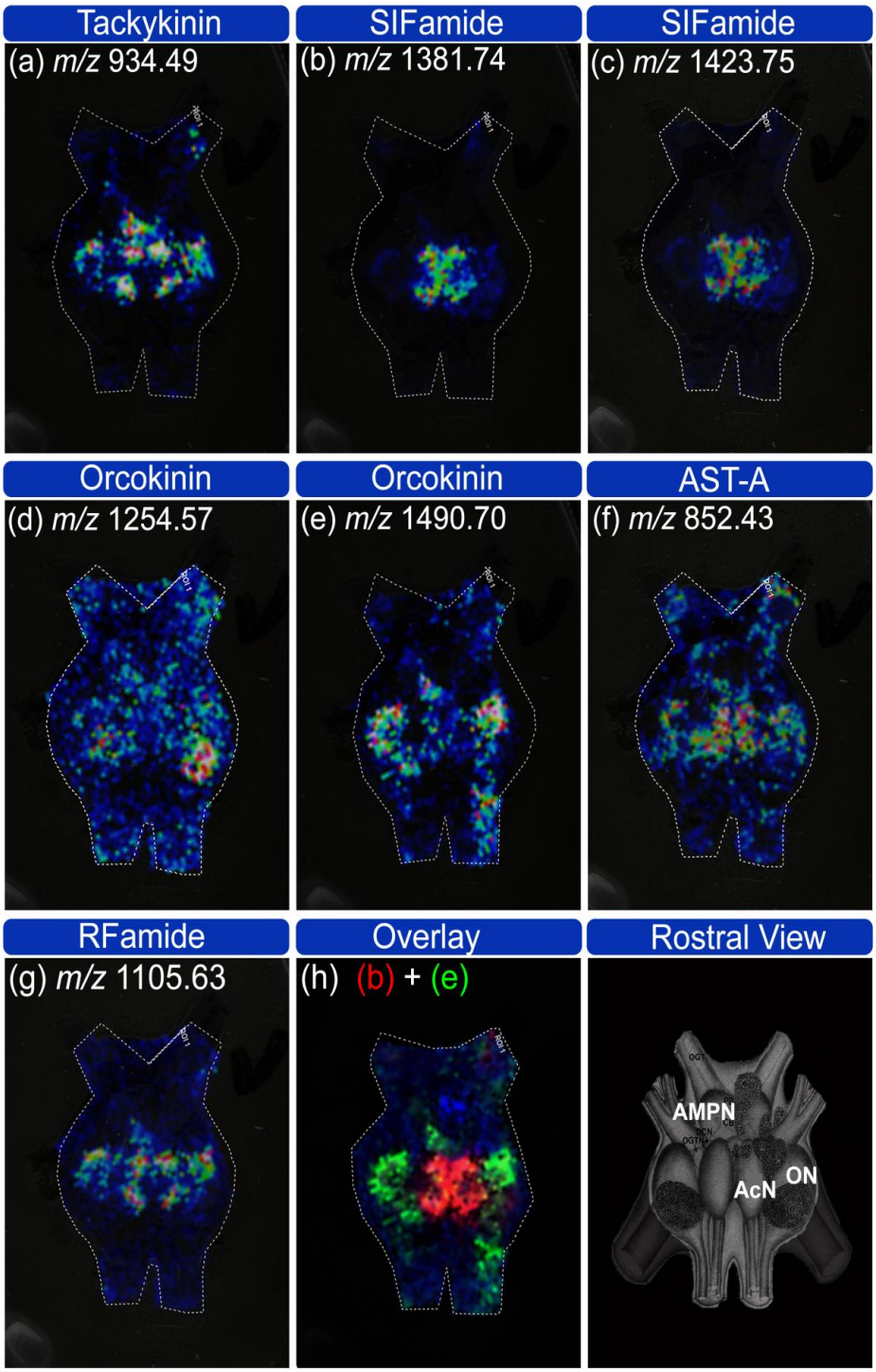


Figure 7. MSI of *P. interruptus* brain shows the localization of several representative neuropeptides from tachykinin, SIFamide, orcokinin, AST-A and RFamide families.



SUPPLEMENTAL FIGURES

SI Figure 1. Profiling of *P. interruptus* brain extract on MALDI-TOF/TOF. MALDI-MS profiling spectra of (a) CE Fraction #8 and (b) CE Fraction #14 displayed significantly reduced background compared to the regular profiling of (c) brain extract without CE fractionation.

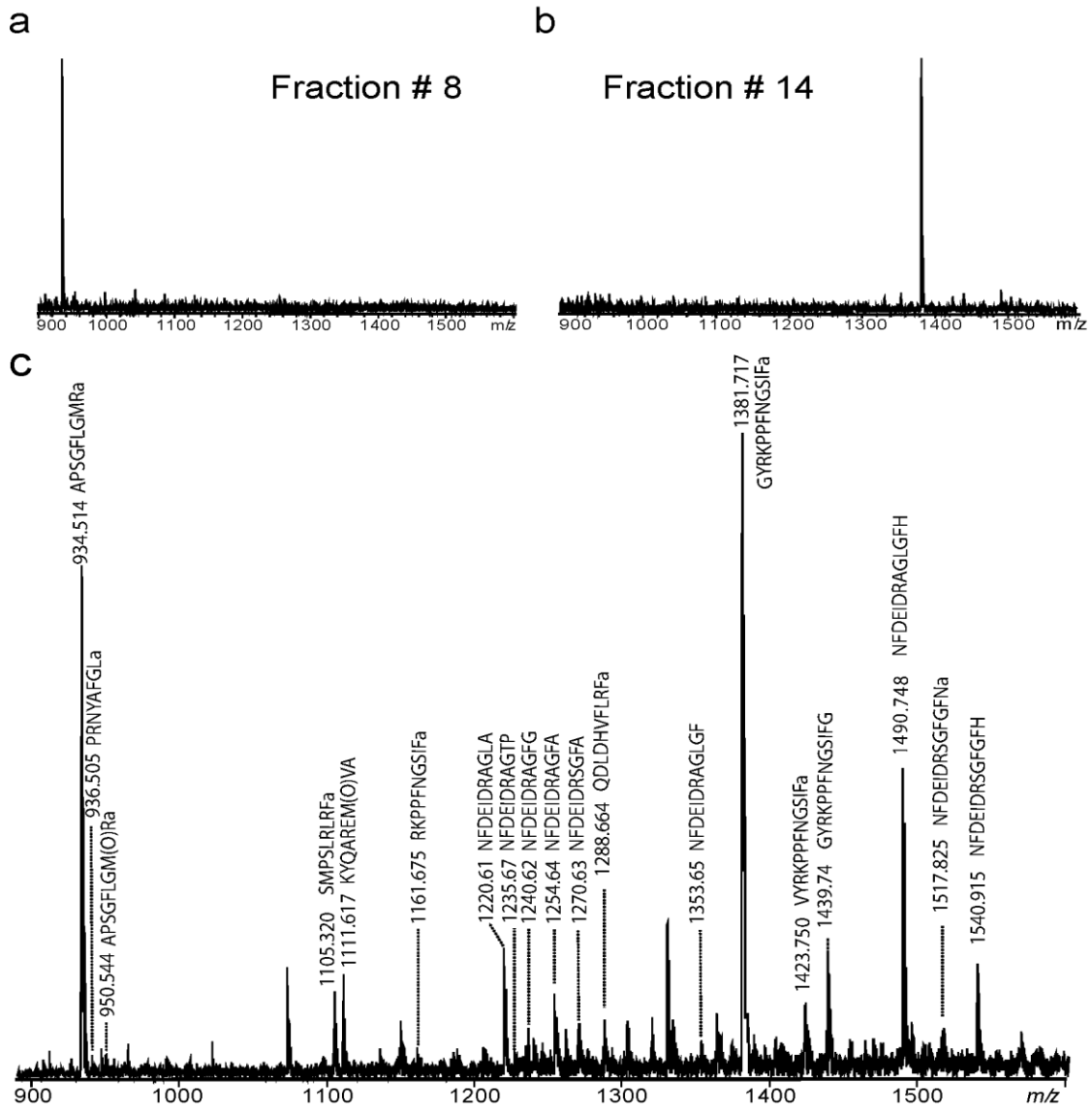


Table 1.

Family	$[M+H]^+$	Sequence	Q/TOF	Orbitrap	TOF/TOF
AST-A	852.43	ADPYAFGLa	+		
	936.51	PRNYAFGLa	+	+	+
RFa	887.56	PSLRLRFa	+	+	
	905.51	PSMRLRFa		+	
	1104.61	GAHKNYLRF	+		
	1105.63	SMPSLRLRFa	+	+	+
	1147.65	APQRNFLRFa	+	+	
	1271.65	pQDLDHVFLRFa	+		
	1288.68	QDLDHVFLRFa	+		
Orcomytropin	976.44	FDAFTTGFA		+	
	1186.52	FDAFTTGFGHS	+	+	
Orcokinin	979.45	NFDEIDRA	+		
	1148.57	NFDEIDRAGLa		+	+
	1206.57	NFDEIDRAGLG		+	+
	1220.59	NFDEIDRAGLA	+		+
	1234.57	NFDEIDRAGVV		+	+
	1240.56	NFDEIDRAGFG	+		+
	1254.57	NFDEIDRAGFA	+	+	+
	1256.55	NFDEIDRSGFG		+	+
	1270.57	NFDEIDRSGFA	+		+
	1353.64	NFDEIDRAGLGF	+		+
	1490.70	NFDEIDRAGLGFH	+	+	+
	1516.68	NFDEIDRSGFGFNa	+		+
	1540.68	NFDEIDRSGFGFH	+		+
SIFa	1161.65	RKPPFNGSIFa	+		+
	1381.74	GYRKPPFNGSIFa	+	+	+
	1382.72	GYRKPPFNGSIF	+		
	1395.75	GYRKPPYNGSIYa(-H_2O)		+	
	1397.78	GYRKPPFNGSIYa	+	+	

	1413.74	GYRKPPYNGSIYa		+	
	1423.75	(Ac)GYRKPPFNGSIFa	+	+	+
	1431.72	GYRKPPFNGSYFa	+		
	1439.74	GYRKPPFNGSIFG	+	+	+
PDH	887.52	NSLLGISAL	+		
	1012.62	ELAAQILRV		+	
	1111.56	KYQAREM(O)VA	+		+
	1133.56	AREM(O)VAELAQ	+		
	1927.03	NSELINSILGLPKVMNDAa	+	+	
	1928.02	NSELINSILGLPKVMNDA	+		
	1890.01	QQIYRVAQAPQAGAVGPHa	+		
Preprotachykinin	605.30	APSGFQa	+		
	779.38	APSGFLGM	+		
	934.49	APSGFLGMRa	+	+	+
	950.49	APSGFLGM(O)Ra	+		+
	992.50	APSGFLGMRG	+	+	
	1650.96	AFRNGNAPVGLAVPLRa		+	+
	1651.94	AFRNGNAPVGLAVPLR			+
	1814.02	YAFRNGNAPVGLAVPLRa		+	
	3084.52	DAAAPLNEVDEASANDYPILPDPIAARLY		+	
	3155.53	DAAAPLNEVDEASANDYPILPDPIAARLYA		+	
	4716.45	DAAAPLNEVDEASANDYPILPDPIAARLYAFRNGNAPVGLAVPLRa		+	
Preproneuropeptide F	2304.11	SDYPLPPGDALMEASERLLET		+	
	2488.23	SDYPLPPGDALMEASERLLETLA		+	
PreproCLDH	2956.52	GLDLGLGRGFSGSQAAKHLMGLAAANYAGGPa	+	+	+

REFERNCES

1. Morton, G. J.; Cummings, D. E.; Baskin, D. G.; Barsh, G. S.; Schwartz, M. W., Central nervous system control of food intake and body weight. *Nature* **2006,** 443, (7109), 289-95.
2. Hokfelt, T.; Bartfai, T.; Bloom, F., Neuropeptides: opportunities for drug discovery. *Lancet Neurol.* **2003,** 2, (8), 463-72.
3. Solway, B.; Bose, S. C.; Corder, G.; Donahue, R. R.; Taylor, B. K., Tonic inhibition of chronic pain by neuropeptide Y. *Proc. Natl. Acad. Sci. U. S. A.* **2011,** 108, (17), 7224-9.
4. Li, L.; Sweedler, J. V., Peptides in the brain: mass spectrometry-based measurement approaches and challenges. *Annu. Rev. Anal. Chem.* **2008,** 1, 451-83.
5. Devi, L., Consensus Sequence for Processing of Peptide Precursors at Monobasic Sites. *FEBS Lett.* **1991,** 280, (2), 189-194.
6. Edman, P., Sequence determination. *Mol. Biol. Biochem. Biophys.* **1970,** 8, 211-55.
7. Romanova, E. V.; Lee, J. E.; Kelleher, N. L.; Sweedler, J. V.; Gulley, J. M., Comparative peptidomics analysis of neural adaptations in rats repeatedly exposed to amphetamine. *J. Neurochem.* **2012,** 123, (2), 276-287.
8. Karlsson, O.; Kultima, K.; Wadensten, H.; Nilsson, A.; Roman, E.; Andrén, P. E.; Brittebo, E. B., Neurotoxin-Induced Neuropeptide Perturbations in Striatum of Neonatal Rats. *J Proteome Res* **2013**, Epub ahead of print.
9. Mabrouk, O. S.; Kennedy, R. T., Simultaneous oxytocin and arg-vasopressin measurements in microdialysates using capillary liquid chromatography-mass spectrometry. *J. Neurosci. Methods* **2012,** 209, (1), 127-133.
10. Ljungdahl, A.; Hanrieder, J.; Falth, M.; Bergquist, J.; Andersson, M., Imaging mass spectrometry reveals elevated nigral levels of dynorphin neuropeptides in L-DOPA-induced dyskinesia in rat model of Parkinson's disease. *PLoS One* **2011,** 6, (9), e25653.
11. Zhang, X. Z.; Petruzziello, F.; Zani, F.; Fouillen, L.; Andrén, P. E.; Solinas, G.; Rainer, G., High Identification Rates of Endogenous Neuropeptides from Mouse Brain. *J. Proteome Res.* **2012,** 11, (5), 2819-2827.
12. Wardman, J. H.; Berezniuk, I.; Di, S.; Tasker, J. G.; Fricker, L. D., ProSAAS-derived peptides are colocalized with neuropeptide Y and function as neuropeptides in the regulation of food Intake. *PLoS One* **2011,** 6, (12), e28152.
13. Guenther, S.; Römpp, A.; Kummer, W.; Spengler, B., AP-MALDI imaging of neuropeptides in mouse pituitary gland with 5 µm spatial resolution and high mass accuracy. *Int. J. Mass Spectrom.* **2011,** 305, (2-3), 228-237.
14. Verhaert, P. D.; Conaway, M. C. P.; Pekar, T. M.; Miller, K., Neuropeptide imaging on an LTQ with vMALDI source: The complete 'all-in-one' peptidome analysis. *Int. J. Mass Spectrom.* **2007,** 260, (2-3), 177-184.
15. Verhaert, P. D.; Pinkse, M. W.; Strupat, K.; Conaway, M. C., Imaging of similar mass neuropeptides in neuronal tissue by enhanced resolution MALDI MS with an ion trap - Orbitrap hybrid instrument. *Methods. Mol. Biol.* **2010,** 656, 433-49.
16. DeKeyser, S. S.; Kutz-Naber, K. K.; Schmidt, J. J.; Barrett-Wilt, G. A.; Li, L. J., Imaging mass spectrometry of neuropeptides in decapod crustacean neuronal tissues. *J. Proteome Res.* **2007,** 6, (5), 1782-1791.
17. Ye, H.; Hui, L. M.; Kellersberger, K.; Li, L. J., Mapping of neuropeptides in the crustacean

stomatogastric nervous system by imaging mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2013,** 24, (1), 134-147.
18. Hui, L.; D'Andrea, B. T.; Jia, C.; Liang, Z.; Christie, A. E.; Li, L., Mass spectrometric characterization of the neuropeptidome of the ghost crab *Ocypode ceratophthalma* (Brachyura, Ocypodidae). *Gen. Comp. Endocrinol.* **2013,** 184, 22-34.
19. Wegener, C.; Neupert, S.; Predel, R., Direct MALDI-TOF mass spectrometric peptide profiling of neuroendocrine tissue of Drosophila. *Methods Mol. Biol.* **2006,** 615, 117-27.
20. Husson, S. J.; Clynen, E.; Boonen, K.; Janssen, T.; Lindemans, M.; Baggerman, G.; Schoofs, L., Approaches to identify endogenous peptides in the soil nematode *Caenorhabditis elegans*. *Methods Mol. Biol.* **2010,** 615, 29-47.
21. Jarecki, J. L.; Frey, B. L.; Smith, L. M.; Stretton, A. O., Discovery of neuropeptides in the nematode *Ascaris suum* by database mining and tandem mass spectrometry. *J. Proteome Res.* **2011,** 10, (7), 3098-3106.
22. Ye, H.; Greer, T.; Li, L. J., Probing neuropeptide signaling at the organ and cellular domains via imaging mass spectrometry. *J. Proteomics* **2012,** 75, (16), 5014-5026.
23. Christie, A. E.; Stemmler, E. A.; Dickinson, P. S., Crustacean neuropeptides. *Cell. Mol. Life Sci.* **2010,** 67, (24), 4135-4169.
24. Cape, S. S.; Rehm, K. J.; Ma, M.; Marder, E.; Li, L. J., Mass spectral comparison of the neuropeptide complement of the stomatogastric ganglion and brain in the adult and embryonic lobster, *Homarus americanus*. *J. Neurochem.* **2008,** 105, (3), 690-702.
25. Chen, R. B.; Jiang, X. Y.; Conaway, M. C. P.; Mohtashemi, I.; Hui, L. M.; Viner, R.; Li, L. J., Mass spectral analysis of neuropeptide expression and distribution in the nervous system of the lobster Homarus americanus. *J. Proteome Res.* **2010,** 9, (2), 818-832.
26. Christie, A. E.; Cashman, C. R.; Brennan, H. R.; Ma, M. M.; Sousa, G. L.; Li, L. J.; Stemmler, E. A.; Dickinson, P. S., Identification of putative crustacean neuropeptides using in silico analyses of publicly accessible expressed sequence tags. *Gen. Comp. Endocrinol.* **2008,** 156, (2), 246-264.
27. Fu, Q.; Li, L. J., *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, (23), 7783-7795.
28. 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, (9), 4219-29.
29. Li, L. J.; Pulver, S. R.; Kelley, W. P.; Thirumalai, V.; Sweedler, J. V.; Marder, E., Orcokinin peptides in developing and adult crustacean stomatogastric nervous systems and pericardial organs. *J. Comp. Neurol.* **2002,** 444, (3), 227-244.
30. Vázquez-Acevedo, N.; Rivera, N. M.; Torres-Gonzalez, A. M.; Rullan-Matheu, Y.; Ruíz-Rodríguez, E. A.; Sosa, M. A., GYRKPPFNGSIFamide (Gly-SIFamide) modulates aggression in the freshwater prawn Macrobrachium rosenbergii. *Biol. Bull.* **2009,** 217, (3), 313-26.
31. Sandeman, D.; Sandeman, R.; Derby, C.; Schmidt, M., Morphology of the brain of crayfish, crabs, and spiny lobsters: a common nomenclature for homologous structures. *Biol. Bull.* **1992,** 183, (2), 304-326.
32. Zhang, Z. C.; Ye, H.; Wang, J. H.; Hui, L. M.; Li, L. J., Pressure-assisted capillary electrophoresis coupling with matrix-assisted laser desorption/ionization-mass spectrometric

imaging for quantitative analysis of complex peptide mixtures. *Anal. Chem.* **2012,** 84, (18), 7684-7691.
33. Stemmler, E. A.; Provencher, H. L.; Guiney, M. E.; Gardner, N. P.; Dickinson, P. S., Matrix-assisted laser desorption/ionization Fourier transform mass spectrometry for the identification of orcokinin neuropeptides in crustaceans using metastable decay and sustained off-resonance irradiation. *Anal. Chem.* **2005,** 77, (11), 3594-3606.
34. Yasuda-Kamatani, Y.; Yasuda, A., APSGFLGMRamide is a unique tachykinin-related peptide in crustaceans. *Eur. J. Biochem.* **2004,** 271, (8), 1546-56.
35. Christie, A. E.; Cashman, C. R.; Stevens, J. S.; Smith, C. M.; Beale, K. M.; Stemmler, E. A.; Greenwood, S. J.; Towle, D. W.; Dickinson, P. S., Identification and cardiotropic actions of brain/gut-derived tachykinin-related peptides (TRPs) from the American lobster *Homarus americanus*. *Peptides* **2008,** 29, (11), 1909-18.
36. Christie, A. E.; Stevens, J. S.; Bowers, M. R.; Chapline, M. C.; Jensen, D. A.; Schegg, K. M.; Goldwaser, J.; Kwiatkowski, M. A.; Pleasant, T. K.; Shoenfeld, L.; Tempest, L. K.; Williams, C. R.; Wiwatpanit, T.; Smith, C. M.; Beale, K. M.; Towle, D. W.; Schooley, D. A.; Dickinson, P. S., Identification of a calcitonin-like diuretic hormone that functions as an intrinsic modulator of the American lobster, *Homarus americanus*, cardiac neuromuscular system. *J. Exp. Biol.* **2010,** 213, (1), 118-127.
37. Gelman, J. S.; Dasgupta, S.; Berezniuk, I.; Fricker, L. D., Analysis of peptides secreted from cultured mouse brain tissue. *Biochim. Biophys. Acta* **2013**, doi: 10.1016/j.bbapap.2013.01.043.
38. Brugge, V. A.; Schooley, D. A.; Orchard, I., Amino acid sequence and biological activity of a calcitonin-like diuretic hormone (DH31) from *Rhodnius prolixus*. *J. Exp. Biol.* **2008,** 211, (Pt 3), 382-90.
39. Furuya, K.; Milchak, R. J.; Schegg, K. M.; Zhang, J.; Tobe, S. S.; Coast, G. M.; Schooley, D. A., Cockroach diuretic hormones: characterization of a calcitonin-like peptide in insects. *Proc. Natl. Acad. Sci. U. S. A.* **2000,** 97, (12), 6469-74.
40. Coast, G. M.; Webster, S. G.; Schegg, K. M.; Tobe, S. S.; Schooley, D. A., The *Drosophila melanogaster* homologue of an insect calcitonin-like diuretic peptide stimulates V-ATPase activity in fruit fly Malpighian tubules. *J. Exp. Biol.* **2001,** 204, (Pt 10), 1795-804.
41. Gard, A. L.; Lenz, P. H.; Shaw, J. R.; Christie, A. E., Identification of putative peptide paracrines/hormones in the water flea *Daphnia pulex* (Crustacea; Branchiopoda; Cladocera) using transcriptomics and immunohistochemistry. *Gen. Comp. Endocrinol.* **2009,** 160, (3), 271-87.
42. Xi, L.; Jin, Y.; Parker, E. A.; Josh, P.; Jones, A.; Wijffels, G.; Colgrave, M. L., Challenges in mass spectrometry-based quantification of bioactive peptides: a case study exploring the neuropeptide Y family. *Biopolymers* **2012,** 98, (4), 357-66.
43. Kiris, I. G.; Eroldogan, O. T.; Kir, M.; Kumlu, M., Influence of neuropeptide Y (NPY) on food intake and growth of penaeid shrimps *Marsupenaeus japonicus* and *Penaeus semisulcatus* (Decapoda: Penaeidae). *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* **2004,** 139, (2), 239-44.
44. Boyne, M. T.; Garcia, B. A.; Li, M.; Zamdborg, L.; Wenger, C. D.; Babai, S.; Kelleher, N. L., Tandem mass spectrometry with ultrahigh mass accuracy clarifies peptide identification by database retrieval. *J. Proteome Res.* **2009,** 8, (1), 374-9.
45. Hsu, J. L.; Huang, S. Y.; Shiea, J. T.; Huang, W. Y.; Chen, S. H., Beyond quantitative proteomics: signal enhancement of the a1 ion as a mass tag for peptide sequencing using

dimethyl labeling. *J. Proteome Res.* **2005,** 4, (1), 101-8.
46. Paizs, B.; Suhai, S., Fragmentation pathways of protonated peptides. *Mass Spectrom. Rev.* **2005,** 24, (4), 508-48.
47. Maynard, D. M., Integration in crustacean ganglia. *Symp. Soc. Exp. Biol.* **1966,** 20, 111-49.
48. Hui, L. M.; Zhang, Y. Z.; Wang, J. H.; Cook, A.; Ye, H.; Nusbaum, M. P.; Li, L. J., Discovery and functional study of a novel crustacean tachykinin neuropeptide. *ACS Chem. Neurosci.* **2011,** 2, (12), 711-722.
49. Chen, R. B.; Hui, L. M.; Sturm, R. M.; Li, L. J., Three dimensional mapping of neuropeptides and lipids in crustacean brain by mass spectral imaging. *J. Am. Soc. Mass Spectrom.* **2009,** 20, (6), 1068-1077.
50. Hanrieder, J.; Ljungdahl, A.; Falth, M.; Mammo, S. E.; Bergquist, J.; Andersson, M., L-DOPA-induced dyskinesia is associated with regional increase of striatal dynorphin peptides as elucidated by imaging mass spectrometry. *Mol. Cell. Proteomics* **2011,** 10, (10).
51. Zimmerman, T. A.; Rubakhin, S. S.; Sweedler, J. V., MALDI mass spectrometry imaging of neuronal cell cultures. *J. Am. Soc.Mass Spectrom.* **2011,** 22, (5), 828-836.
52. Römpp, A.; Guenther, S.; Schober, Y.; Schulz, O.; Takats, Z.; Kummer, W.; Spengler, B., Histology by mass spectrometry: label-free tissue characterization obtained from high-accuracy bioanalytical imaging. *Angew. Chem. Int. Edi.* **2010,** 49, (22), 3834-3838.
53. Yasuda-Kamatani, Y.; Yasuda, A., Characteristic expression patterns of allatostatin-like peptide, FMRFamide-related peptide, orcokinin, tachykinin-related peptide, and SIFamide in the olfactory system of crayfish *Procambarus clarkii*. *J. Comp. Neurol.* **2006,** 496, (1), 135-47.

Chapter 6

Investigation and reduction of sub-microgram peptide loss using molecular weight cut-off fractionation prior to mass spectrometric analysis

Adapted from: "Investigation and reduction of sub-microgram peptide loss using molecular weight cut-off fractionation prior to mass spectrometric analysis" Cunningham R, **Wang J**, Wellner D, Li L. *Journal of Mass Spectrometry*. **2012**, Oct; 47(**10**): 1327-32.

ABSTRACT

This work investigates the introduction of methanol and a salt modifier to molecular weight cut-off membrane-based centrifugal filters (MWCO) to enrich sub-microgram peptide quantities. Using a neuropeptide standard, bradykinin, sample loss was reduced over two orders of magnitude with and without undigested protein present. Additionally, a bovine serum albumin (BSA) tryptic digestion was investigated. Twenty-seven tryptic peptides were identified from MALDI mass spectra after enriching with methanol while only two tryptic peptides were identified after the standard MWCO protocol. The strategy presented here enhances recovery from MWCO separation for sub-µg peptide samples**.**

INTRODUCTION

Molecular weight cut-off membrane-based centrifugal filter devices (MWCO) are commonly used to desalt and concentrate large molecular weight proteins [1]. Greeing and Simpson recently investigated various MWCO membranes for large amounts of starting material (~6 mg), focusing on optimal conditions for the sub 25 kDa protein fraction [2]. The authors recovered 200 µg to 2.9 mg of protein from multiple MWCO experiments and demonstrated that a 10:90 acetonitrile (ACN):H_2O elution solvent produced optimal results [3]. In addition, Manza *et. al.* provided an alternative approach to isolate proteins with a 5 kDa MWCO by using NH_4HCO_3 and recovering the retained proteins [4]. Alternatively, the elution from MWCOs can be collected to recover only low molecular weight peptides. Multiple peptidomic studies have utilized MWCOs for peptide isolation during the first few steps of sample preparation [5, 6]. When sample

amount is limited or peptide content is below 1 µg, sample loss is a significant concern when using MWCOs to isolate endogenous peptides. Optimized protocols have been investigated using ACN [3, 7], salt [4, 8], SDS [5], or native sample [6, 9, 10], but these experiments primarily focused on large sample amounts rather than sub-microgram peptide quantities.

MWCOs separate large molecules from small molecules. The small molecule fraction may be rich with signaling peptides (SP), cytokines, and other small molecules involved in cell-cell signaling. Signaling peptides perform various functions in the body, including cell growth, cell survival, and hormonal signaling between organs [11]. Individual SP contribute to different aspects of behavior, such as pain (enkephalins) [12], feeding (neuropeptide Y) [13], and blood pressure (bradykinin) [14]. MWCO separations can be used to enrich biologically important SP and explore the peptide content from biological fluids with relatively low peptide content like blood or cerebrospinal fluid (CSF). In a recent investigation, the detection of neuropeptides and standards in crustacean hemolymph was improved when methanol and protease inhibitors were present before performing MWCO neuropeptide isolation. The impact of methanol on MWCO sample loss was not investigated in the study [15]. In another study, a large-scale mass fingerprinting protocol of endogenous peptides from CSF used a combination of salts before MWCO fractionation, but the impact of adding salts was not discussed [16]. The most commonly used brand of MWCO in the publications and in peptidomic studies is Millipore. Therefore, Millipore MWCOs (using regenerated cellulose as the membrane) were used in the present study. The purpose of this work is to provide an

optimized sample preparation technique for MWCO filtering to reduce sample loss and allow sub-µg detection of peptides using MALDI mass spectrometry.

MATERIALS AND METHODS

Materials and Chemicals

Water, acetonitrile, methanol (optima LC/MS grade) and sodium chloride (99.5%) were purchased from Fisher Scientific (Fair Lawn, NJ). The α-cyano-4-hydroxy-cinnamic acid (99%), formic acid (FA) (≥98%), and bovine serum albumin (≥96%) were purchased from Sigma-Aldrich (St. Louis, MO). Amicon Ultra 0.5 mL 10,000 MWCO centrifugal filters and ZipTips packed with C18 reversed-phase resin were purchased from Millipore (Billerica, MA). Trypsin-digested bovine serum albumin (BSA) was purchased from Waters (Milford, MA). Bradykinin was purchased from American Peptide Company (Sunnyvale, CA).

MALDI MS Instrumentation

An AutoFLEX III MALDI TOF/TOF mass spectrometer (Bruker Daltonics, Billerica, MA) was operated in positive ion reflectron mode. The MALDI MS instrument is equipped with a proprietary smart beam (Bruker Daltonics, Billerica, MA) laser operating at 200 Hz. The instrument was internally calibrated over the mass range of *m*/*z* 500−2500 using a standard peptide mix. Two thousand laser shots were collected per sample spot at an accelerating voltage of 19 kV and a constant laser power using random shot selection. The acquired data were analyzed using FlexAnalysis software (Bruker

Daltonics, Billerica, MA). Mass spectrometry data acquisition was obtained by averaging 2000 laser shots.

Molecular weight cut off separation procedure

The MWCO separations were performed using Amicon Ultra 0.5 mL 10,000 MWCO centrifugal filters (Billerica, MA). Before MWCO separation, three washing steps were performed to remove contaminants from the filter. The three washes were 500 μL of 50:50 H_2O:MeOH, 500 μL of H_2O, and 400 μL of the solution used for MWCO separation. For the 100% H_2O solution, 1 μg of BSA or bradykinin was used for separation. All the other MWCO separation experiments used 500 ng of BSA or 100 ng or less of bradykinin. The MWCO filter was then centrifuged at 14,000 rcf for 5 min at room temperature in an Eppendorf 5415 D microcentrifuge (Brinkmann Instruments Inc., Westbury, NY). The filtrate was concentrated in a Savant SC 110 SpeedVac concentrator (Thermo Electron Corporation, West Palm Beach, FL) and acidified. The resulting sample was desalted according to the manufacturer using C18 ZipTips from Millipore (Billerica, MA) by washing the ZipTip with three times 100% ACN, three aqueous washes of 0.1% FA, binding the peptides from the solution and one aqueous wash of 0.1% FA. Peptides were then eluted from the ZipTips using 15 μL of 50% ACN in 0.1% FA.

Matrix deposition

Equal volumes of 0.5 μL from the 15 μL sample solution (including standards not subject to MWCO filtering) and α-cyano-4-hydroxy-cinnamic acid (CHCA) matrix solution in 50% ACN:50% H_2O were mixed using a dried-droplet method and spotted on

a MALDI target. The resulting droplets were allowed to air dry prior to mass spectrometry acquisition.

RESULTS AND DISCUSSION

Analysis of two orders of magnitude increase for bradykinin standard

Bradykinin was selected to assess the potential peptide loss in the flow-through after performing MWCO separation. As shown in Figure 1, 1 μg of bradykinin standard does not produce a detectable signal by MALDI mass spectrometry analysis after a 10 kDa MWCO separation in water (performed in triplicate). For comparison, 1 ng of bradykinin standard diluted to 15 μL produced an intense signal on the MALDI mass spectrometer, suggesting significant sample loss occurs when the target analyte is low in quantity (data not shown, performed in triplicate). Figure 1 shows that the addition of a salt, in this case NaCl, improves the limits of detection and decreases sample loss when 70:30 water:methanol was compared to 70:30 aqueous 1 M NaCl:methanol. The reproducible results gave a relative standard deviation (RSD) of 6% for peak intensity. Figure 1 shows that even when starting with 1 μg of bradykinin, too much sample is lost during the MWCO separation in water to detect the remainder. However, an intense signal is observed for 10 ng of bradykinin after the MWCO separation when 70:30 aqueous 1 M NaCl:methanol is used as an elution solvent. Sample loss from zip-tipping was estimated in triplicate by calculating the decreases in peak intensity. When 10 ng of bradykinin was used, sample loss was ~41%. In comparison, the calculated yield for 10 ng of bradykinin after the 70:30 aqueous 1 M NaCl:methanol MWCO separation and zip-

tip desalting showed an estimated sample loss of 63%, meaning more loss can be attributed to sample clean-up than MWCO filtration.

A series of experiments were performed to determine if 70:30 aqueous 1 M NaCl:methanol is an optimal solution to recover peptides during a MWCO separation (data not shown). A 50:50 aqueous 1 M NaCl:methanol and a 50:50 water:methanol elution were performed in duplicate, but signal intensity of the resulting bradykinin was poor and numerous polymer peaks were detected in the flow through. A lower salt concentration, 0.1 M NaCl, was used but also produced greatly reduced signal when compared to aqueous 1 M NaCl. To assess the optimized method's compatibility with lower sample amounts, the 70:30 aqueous 1 M NaCl:methanol solution was added to 1 ng of bradykinin for MWCO separation, but no signal was obtained (data not shown). Using a neuropeptide standard, the addition of methanol and NaCl salt significantly improved the sample recovery in sub-µg amounts.

BSA tryptic peptide mixture analysis

After demonstrating the importance of using an optimized solution for MWCO separations with an individual peptide, the new method was applied to 500 ng of BSA tryptic digest to investigate its utility with more complex peptide mixtures . Table 1 lists the BSA tryptic peptides identified in the MALDI MS analysis from different solution conditions processed by MWCO separation. As shown in Table 1, a directly spotted BSA tryptic peptide standard in the absence of any MWCO filtration enabled identification of 39 tryptic peptides by accurate peptide mass measurements. Once again, when using 100% H_2O for MWCO separations, the starting amount was doubled to 1 µg

(also done with 500 ng, data not shown). However, many tryptic peptides were not detected due to low signal intensities and non-optimal elution conditions. Instead of H_2O, a 1 M NaCl solution was used for the MWCO elution, but only two tryptic peptides were identified (Table 1). The addition of 30% methanol into the sample before MWCO filtration produced the first increase in identified BSA tryptic peptides. The remaining data from Table 1 shows improved BSA tryptic peptide identifications as the sample (elution) conditions were further optimized. Figure 2 shows the actual mass spectra associated with the three most promising elution solutions along with 100% H_2O.

The BSA tryptic peptide intensities are shown in Figure 2A, and the most intense tryptic peptide, YLYEIAR, *m/z* 927.49 was observed in the four different solutions shown in Figure 2B but not in the 100% H_2O or 100% 1 M NaCl solutions (data not shown). In Figure 2A, all mass spectra are normalized to an intensity of 7×10^4 to illustrate two points. First, the MWCO filtering step still produced sample loss regardless of the solvent conditions chosen. Second, there is a noticeable increase in peptide peak intensity using the optimized solvent, 60:40 aqueous 1 M NaCl:methanol (Figure 2A). Figure 2C displays a zoomed-in view of a BSA tryptic peptide signal, LKEC$^{\#}$C$^{\#}$DKPLLEK, *m/z* 1532.66 ($^{\#}$: carbamidomethyl) observed only in the optimized solvent. The detection of the *m/z* 1532.66 peptide in Figure 2C highlights the potential gain in sample and detectable peptides by using an optimized salt/MeOH combination. A non-optimized salt/MeOH combination will still reduce sample loss, but further minimizing sample loss during sample preparation will always be desirable in any analytical protocol.

MWCO composition

The purpose of this application note is to provide evidence of sub-µg sample loss during MWCO separations of peptide samples and a solution to overcome this limitation. The explanation of why adding MeOH and NaCl to the sample solution provides a significant reduction in sample loss is beyond the scope of this application note. Regardless, Supplemental Table S1 is an expanded version of Table 1, showing the amino acid sequence, hydrophobicity calculated using GRAVY scores, and pI of the identified peptides in this study. No discernible trend was obtained from the data. The membrane of commonly used MWCO in peptidomics and for this study is comprised of chemically treated (regenerated) cellulose, which is a polysaccharide containing β (1→4) linked D-glucose. Glucose has numerous free hydroxyl groups which could non-specifically adsorb peptides flowing through the MWCO. The addition of MeOH has the most significant effect on signal, which could be due to disrupting the interaction between peptides and hydroxyl groups from glucose. NaCl has a less significant effect on sample recovery compared to MeOH, but a detectable reduction in sample loss is noted. This improvement in sample recovery could be analogous to the use of NaCl in immunodepletion protocols to reduce non-specific binding, which is accomplished by adding 150 mM NaCl [17].

Analysis of bradykinin in the presence of undigested BSA

When using MWCO for peptide isolation, proteins are typically present in the samples, usually in larger amounts. Figure 3 shows the effect that adding BSA protein to a 10 ng bradykinin solution before MWCO fractionation has on the resulting recovery of

bradykinin. Adding 10 µg of BSA to the optimized 70:30 aqueous 1 M NaCl:methanol solution slightly decreased bradykinin's signal with a RSD of 10%. More severe signal reduction occurred after adding 100 µg BSA with a RSD of 2% (N=2). It is not unexpected that more signal reduction due to sample loss would occur, especially in the 100 µg BSA sample because the BSA protein has a molar ratio of 160 BSA protein molecules to one bradykinin peptide. Figure 3 shows the usefulness of the MWCO method with samples containing large amounts of proteins.

Recommendation/Conclusion

The present work provides solutions to reduce sample loss from the use of MWCO for sub-µg peptide isolation with or without non-digested proteins in the sample. Despite its widespread utility, significant sample loss often occurs during the MWCO fractionation step, which is particularly problematic when analyzing low-abundance peptides from limited starting material. This application note aims to reduce sample loss during MWCO separations, specifically for sub-µg peptide isolation. If complex samples are processed with MWCO separation, the authors recommend eluting the sample with 60:40 aqueous 1 M NaCl:methanol solution as a starting point to minimize sample loss. This application note provides a viable alternative for sub-µg peptide MWCO separation, circumventing the need to increase the starting material by minimizing sample loss from using a MWCO membrane-based centrifugal filter device.

References

[1] H.M. Georgiou, G.E. Rice, M.S. Baker. Proteomic analysis of human plasma: failure of centrifugal ultrafiltration to remove albumin and other high molecular weight proteins. *Proteomics*. **2001**, 1, 1503.

[2] D.W. Greening, R.J. Simpson. Low-molecular weight plasma proteome analysis using centrifugal ultrafiltration. *Methods Mol Biol*. **2011**, 278, 109.

[3] D.W. Greening, R.J. Simpson. A centrifugal ultrafiltration strategy for isolating the low-molecular weight (<or=25K) component of human plasma proteome. *J Proteomics*. **2010**, 73, 637.

[4] L.L. Manza, S.L. Stamer, A.J. Ham, S.G. Codreanu, D.C. Liebler. Sample preparation and digestion for proteomic analyses using spin filters. Proteomics. **2005**, 5, 1742.

[5] D. Theodorescu, D. Fliser, S. Wittke, H. Mischak, R. Krebs, M. Walden, M. Ross, E. Eltze, O. Bettendorf, C. Wulfing, A. Semjonow. Pilot study of capillary electrophoresis coupled to mass spectrometry as a tool to define potential prostate cancer biomarkers in urine. *Electrophoresis.* **2005**, 26, 2797.

[6] K. Antwi, G. Hostetter, M.J. Demeure, B.A. Katchman, G.A. Decker, Y. Ruiz, T.D. Sielaff, L.J. Koep, D.F. Lake. Analysis of the plasma peptidome from pancreas cancer patients connects a peptide in plasma to overexpression of the parent protein in tumors. *J Proteome Res*. **2009**, 8, 4722.

[7] L.P. Aristoteli, M.P. Molloy, M.S. Baker. Evaluation of endogenous plasma peptide extraction methods for mass spectrometric biomarker discovery. *J Proteome Res*. **2007**, 6, 571.

[8] A. Zougman, B. Pilch, A. Podtelejnikov, M. Kiehntopf, C. Schnabel, C. Kumar, M. Mann. Integrated analysis of the cerebrospinal fluid peptidome and proteome. *J Proteome Res*. **2008**, 7, 386.

[9] X. Yuan, D.M. Desiderio. Human cerebrospinal fluid peptidomics. *J Mass Spectrom.* **2005**, 40, 176.

[10] X. Zheng, H. Baker, W.S. Hancock. Analysis of the low molecular weight serum peptidome using ultrafiltration and a hybrid ion trap-Fourier transform mass spectrometer. *J Chromatogr A*. **2006**, 1120, 173.

[11] L. Li, J.V. Sweedler. Peptides in the brain: mass spectrometry-based measurement approaches and challenges. *Annu Rev Anal Chem.* **2008**, 1, 451.

[12] G.B. Stefano, G. Fricchione, Y. Goumon, T. Esch. Pain, immunity, opiate and opioid compounds and health. *Med Sci Monit*. **2005**, 11, MS47.

[13] J. Jensen. Regulatory peptides and control of food intake in non-mammalian vertebrates. *Comp Biochem Physiol A Mol Integr Physiol*. **2001**, 128, 471.

[14] A. Kuoppala, K.A. Lindstedt, J. Saarinen, P.T. Kovanen, J.O. Kokkonen. Inactivation of bradykinin by angiotensin-converting enzyme and by carboxypeptidase N in human plasma. *Am J Physiol Heart Circ Physiol.* **2000**, 278, H1069.

[15] R. Chen, M. Ma, L. Hui, J. Zhang, L. Li. Measurement of neuropeptides in crustacean hemolymph via MALDI mass spectrometry. *J Am Soc Mass Spectrom*. **2009**, 20, 708.

[16] H. Jahn, S. Wittke, P. Zurbig, T.J. Raedler, S. Arlt, M. Kellmann, W. Mullen, M. Eichenlaub, H. Mischak, K. Wiedemann. Peptide fingerprinting of Alzheimer's disease in cerebrospinal fluid: identification and prospective evaluation of new synaptic biomarkers. *PLoS One*. **2011**, 6, e26540.

[17] N.A. Cellar, A.S. Karnoup, D.R. Albers, M.L. Langhorst, S.A. Young. Immunodepletion of high abundance proteins coupled on-line with reversed-phase liquid chromatography: a two-dimensional LC sample enrichment and fractionation technique for mammalian proteomics. *J Chromatogr B Analyt Technol Biomed Life Sci*. **2009**, 877, 79.

Table 1. Identified BSA tryptic peptides from various MWCO separation conditions

BSA tryptic peptide (MH^+)	100% H_2O 1µg	100% 1 M NaCl	70% H_2O	80% 1 M NaCl	70% 1 M NaCl	60% H_2O	60% 1 M NaCl
508.3	✓	✓	✓		✓	✓	✓
545.3			✓	✓	✓	✓	✓
689.4				✓	✓	✓	✓
712.4			✓	✓		✓	✓
898.5	✓	✓	✓	✓	✓	✓	✓
927.5			✓	✓	✓	✓	✓
1034.5							✓
1072.5				✓			✓
1138.5				✓			✓
1163.6					✓	✓	✓
1249.6							
1283.7					✓		✓
1305.7			✓		✓	✓	✓
1399.7							
1415.7							
1419.7							✓
1439.8							
1463.6				✓		✓	✓
1479.8					✓	✓	✓
1502.6						✓	✓
1511.8							✓
1532.8							✓
1554.7					✓		✓
1567.7							
1576.8						✓	✓
1639.9					✓	✓	✓
1667.8							✓
1673.8							
1724.8							
1740.8							
1747.7							✓
1749.7					✓	✓	✓
1880.9					✓	✓	✓
1889.0							
1901.9					✓		✓
1907.9					✓		
2045.0							
2113.9							
2247.9							
Total: 39	2	2	6	8	15	15	27

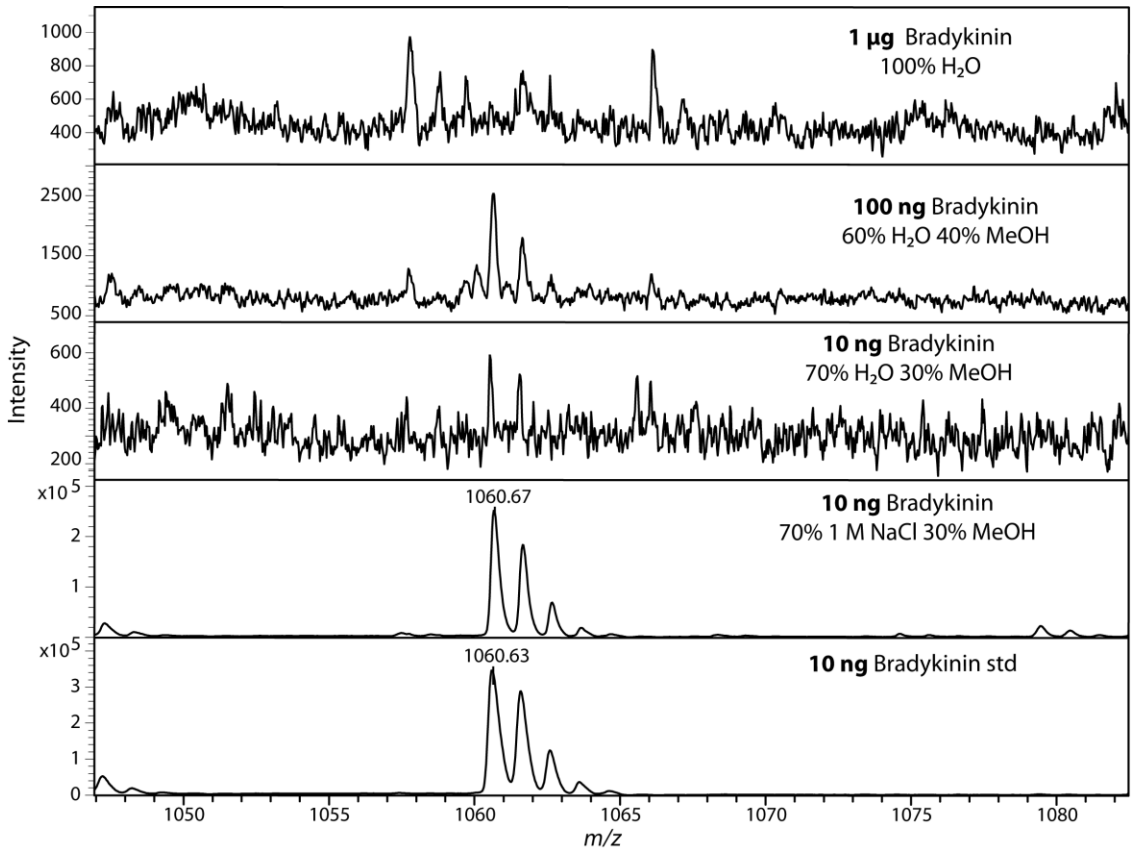


Figure 1: Representative MALDI mass spectra after MWCO separation of a bradykinin standard showing improvement over two orders of magnitude in detection limits. Each MWCO separation was performed at minimum in triplicate with representative spectrum selected for each with a calculated RSD from the peak heights. Three different amounts of bradykinin were tested to assess the magnitude of sample loss under different MWCO solvent conditions. The top panel shows 1 µg of bradykinin standard after MWCO separation with 100% H_2O elution produced no signal. The addition of 40% or 30% MeOH produced very low bradykinin signals for both 100 ng (RSD of 28%) and 10 ng (S/N<3, no RSD calculated) respectively. In the bottom two spectra, each showed very large intensity, but the 70:30 aqueous 1 M NaCl:methanol 10 ng bradykinin was processed with a 10kDa MWCO and zip-tipped and was reproducible with a RSD of 6%.

The last spectrum was from 10 ng bradykinin (RSD of 3%) which was diluted to an equivalent volume as all the other experiments and directly spotted onto the MALDI plate.

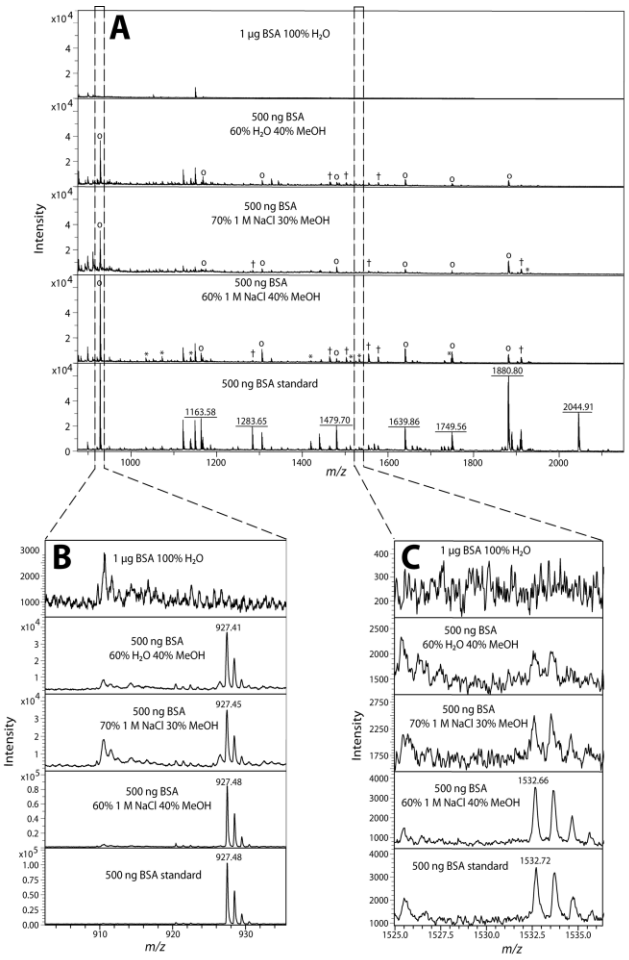


Figure 2: Representative MALDI mass spectra from MWCO separation of a BSA tryptic peptide standard showing sample loss. Stacked mass spectra from *m/z* range 875-2150, normalized to 7 x 10^4 intensity, representing the detection difference from a BSA tryptic peptide standard from different MWCO separation conditions (A). It should be noted that when the solvent for MWCO elution was 100% H_2O, 1 µg of BSA tryptic peptides was processed instead of 500 ng. A zoomed in view of the most abundant BSA tryptic peptide detected, YLYEIAR, *m/z* 927.49 (B). Various percentages of MeOH produced significant signal, but addition of a salt (1 M NaCl) increases the signal, which is closest to a directly spotted BSA tryptic peptide standard. A zoomed in view of a representative low intensity BSA tryptic peptide detected, LKEC$^{\#}$C$^{\#}$DKPLLEK, *m/z*

1532.66 (C). The optimized solution to be used for MWCO filtration, 60:40 aqueous 1 M NaCl:methanol, was the only procedure that enabled the detection of the tryptic peptide in Figure 2C which was also detected in the directly spotted BSA tryptic peptide standard. All experiments were performed a minimum of two times with nearly identical results.

#) Carbamidomethyl amino acid modification

º) Tryptic peptide identified in three of the spectra in Figure 2A

†) Tryptic peptide identified in two of the spectra in Figure 2A

*) Tryptic peptide identified in a single spectrum in Figure 2A

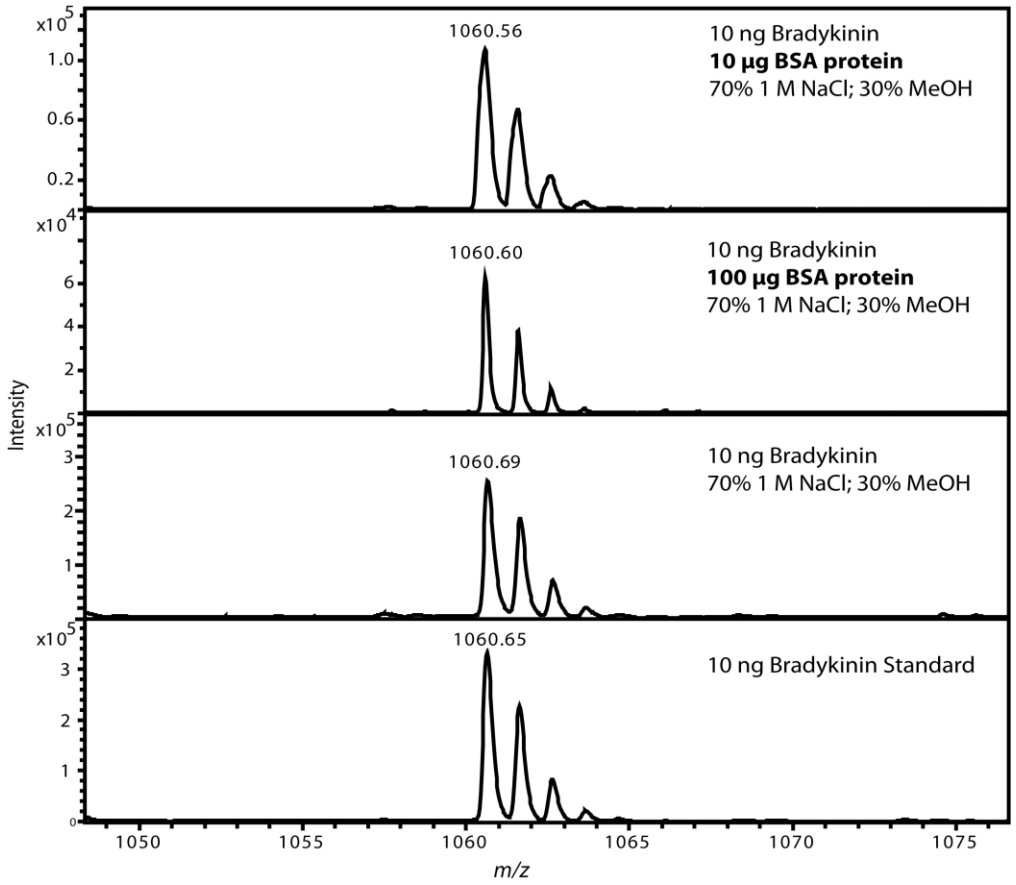


Figure 3: Representative MALDI mass spectra after MWCO separation of a bradykinin standard with a BSA protein present showing optimized solvent conditions minimized samples losses. Each experiment was performed in duplicate. Two different amounts of BSA protein were tested to assess the magnitude of sample loss caused by the presence of a protein. The top panel shows 10 µg of BSA protein and the second panel shows 100 µg of BSA protein added while only 10 ng of bradykinin was added. Detectable sample loss was observed when the BSA protein was added, but panel two shows that the amount of BSA protein was 1 x 10^4 greater (equivalent to 160 fold molar excess) than bradykinin. The last two spectra were controls using a MWCO with the optimized solution in panel 3 and panel 4 using 10 ng bradykinin which was diluted to an equivalent volume as all the other experiments and directly spotted onto the MALDI plate.

BSA tryptic peptide (MH^+)	GRAVY score	Theoretical pI	Sequence	100% H_2O 1µg	100% 1 M NaCl	70% H_2O	80% 1 M NaCl	70% 1 M NaCl	60% H_2O	60% 1 M NaCl
508.3	N/A	N/A	FGER	✓	✓	✓		✓	✓	✓
545.3	0.900	9.72	VASLR			✓	✓	✓	✓	✓
689.4	0.267	9.79	AWSVAR				✓	✓	✓	✓
712.4	-0.950	6.47	SEIAHR			✓	✓		✓	✓
898.5	0.529	6.74	LcVLHEK	✓	✓	✓	✓	✓	✓	✓
927.5	-0.071	6.00	YLYEIAR			✓	✓	✓	✓	✓
1034.5	-0.725	6.74	NEcFLSHK							✓
1072.5	-0.211	5.38	SHcIAEVEK				✓			✓
1138.5	0	5.99	ccTESLVNR				✓			✓
1163.6	0.130	4.53	LVNELTEFAK					✓	✓	✓
1249.6	-1.250	5.45	FKDLGEEHFK							
1283.7	0.264	6.75	HPEYAVSVLLR					✓		✓
1305.7	-0.582	5.32	HLVDEPQNLIK			✓		✓	✓	✓
1399.7	0.567	4.37	TVMENFVAFVDK							
1415.7	0.567	4.37	TVmENFVAFVDK							
1419.7	0.058	5.30	SLHTLFGDELcK							✓
1439.8	-0.133	8.75	RHPEYAVSVLLR							
1463.6	-0.515	4.65	TcVADESHAGcEK				✓		✓	✓
1479.8	0.292	6.00	LGEYGFQNALIVR					✓	✓	✓
1502.6	-0.625	4.09	EYEATLEEccAK						✓	✓
1511.8	0.207	5.97	VPQVSTPTLVEVSR							✓
1532.8	-0.617	6.17	LKEccDKPLLEK							✓
1554.7	-0.823	4.41	DDPHAcYSTVFDK					✓		✓
1567.7	-0.085	4.37	DAFLGSFLYEYSR							
1576.8	-0.985	4.56	LKPDPNTLcDEFK						✓	✓
1639.9	-0.067	8.75	KVPQVSTPTLVEVSR					✓	✓	✓
1667.8	0.064	4.37	MPCTEDYLSLILNR							✓
1673.8	-1.723	5.50	QEPERNEcFLSHK							
1724.8	0.064	4.37	MPcTEDYLSLILNR							
1740.8	0.064	4.37	mPcTEDYLSLILNR							
1747.7	-0.914	4.14	YNGVFQEccQAEDK							✓
1749.7	-0.621	4.10	EccHGDLLEcADDR					✓	✓	✓
1880.9	-0.537	6.06	RPcFSALTPDETYVPK					✓	✓	✓
1889.0	-0.567	6.74	HPYFYAPELLYYANK							
1901.9	-1.275	4.66	NEcFLSHKDDSPDLPK					✓		✓
1907.9	0.044	4.54	LFTFHADIcTLPDTEK					✓		
2045.0	-0.812	8.39	RHPYFYAPELLYYANK							
2113.9	-0.682	4.80	VHKEccHGDLLEcADDR							
2247.9	-0.458	4.23	EccHGDLLEcADDRADLAK							
Total: 39				2	2	6	8	15	15	27

Supplemental Table 1: Expanded Table 1 including grand average of hydropathicity (GRAVY) score, theoretical pI, and the sequence from the underlying amino acid sequence for the peptides identified in the BSA digest. The GRAVY and pI scores were obtained from ExPASy Bioinformatics and modifications were not taken into consideration. (http://web.expasy.org/protparam). Peptide assignments to the recorded peaks were done by mass matching with no tandem mass spectrometry performed.

Lower case amino acids indicate a modification present in the peptide of carbamidomethyl (c) or oxidation (m).

Chapter 7

Conclusion

The work demonstrated in this thesis utilized an array of novel mass spectrometry-based strategies in order to answer a large variety of important questions in the field of neuroscience. The discussed work includes novel peptide discovery in food intake in rats, novel biomarker discovery in Alzheimer's disease and mild cognitive impaired individuals, detection of global proteomic alterations in a transgenic mouse model of autism, and novel peptide discovery in spiny lobster *Panulirus interruptus*. The developed comparative shotgun proteomics approach based on label-free, area-under-the-curve, using a NanoLC–ESI –Orbitrap, has been applied to study brain tissue and CSF in mouse, rats and humans. To enable the comprehensive neuropeptide characterization and quantification a multidimensional platform using multiple analytical techniques, including MALDI–TOF/TOF, NanoLC–ESI–Orbitrap, and NanoLC–ESI –QTOF, was employed. In addition, to further assist in the detection of low abundance endogenous peptides, improved sample preparation methods using molecular weight cut-offs were developed and have been reported. More specifically, I conclude for each project:

In Chapter 2 I employed a combination of cryostat dissection, heat stabilization, neuropeptide extraction and label-free quantitative neuropeptidomics using liquid chromatography coupled to a high-resolution mass spectrometer in order to directly monitor the feeding-induced changes in neuropeptide expression levels within the nucleus accumbens (Acb). In this study we showed that by using this methodology, over 300 neuropeptides were identified from rat Acb, and feeding caused the expression levelof multiple neuropeptides to change, especially opioid peptides, such as enkaphalins, and non-opioid peptides, including ProSAAS, cholecystokinin and somatostatins. We further investigated the regulatory function of non-opioid neuropeptides from the ProSAAS family

by microinjecting the identified ProSAAS neuropeptides into the rat Acb. We showed that at high concentration, big LEN significantly increased food and water intake in the rats. Additionally, both big LEN and PEN at high concentration altered other behaviors in the rats including locomotion and rearing. Furthermore, we also quantified the feeding-induced changes of neuropeptides from the hippocampus, hypothalamus, and striatum to reveal the neuropeptide interplay among different anatomical regions of the brain. In summary, our study revealed the neuropeptidomics changes in response to food intake in the rat Acb and other key brain regions. More importantly, we demonstrated the orexigenic functions of proSAAS neuropeptides before food intake.

The work presented in chapter 3 aimed to discover novel biomarkers for Alzheimer's disease. We used mass spectrometry to compare glycoproteins and endogenous peptides in cognitively healthy individuals, individuals with mild cognitive impairment (MCI), and individuals with AD. By optimizing sub-microgram peptide separation via molecular weight cutoff filtration and using the custom-constructed database, 645 peptides were identified. Interestingly, 42 proSAAS peptides were exclusively cleaved either from the N- or C-terminal of the proSAAS protein precursor, representing similar sequences to the bioactive peptides big/little SAAS, PEN and LEN. Among them, the intact form of big LEN and little LEN were identified, indicating potential bioactivity. Glycoproteins were effectively enriched by lectin affinity chromatography, resulting in the detection of 795 glycoproteins. The glycoproteins were further quantified by comparing the area-under-the-curve among cognitively heathy, MCI and AD individuals. One way ANOVA was applied for statistical analysis, a total of 15 proteins were found to be differentially expressed

among the three groups. The dynamic changes of transthyretin, which first increased in MCI and then declined in AD, were reported and discussed.

In Chapter 4, a comparative study of membrane proteins in mouse hippocampi was conducted. Mutations or duplications in *AT-1/SLC33A1* have been linked to diseases such as familial spastic paraplegia, developmental delay with premature death, and autism spectrum disorder with intellectual disability. In the presented study, Dr. Puglielli's lab generated an AT-1 Tg mouse model that selectively overexpresses human AT-1 in neurons. These animals demonstrate cognitive deficits, autistic-like social behavior, aberrations in synaptic plasticity, an increased number of dendritic spines and branches, and widespread proteomic changes. We found a large set of differentially expressed proteins that are involved in neuronal migration and adhesion, neurite outgrowth, transport of synaptic vesicles, assembly of synaptic connections, and regulation of synaptic activity. Therefore, it is conceivable to assume that the increased influx of acetyl-CoA, and consequent changes in the ER acetylome and in the efficiency of the secretory pathway, is the primary cause of the observed ASD-like phenotype. In conclusion, our results indicate that increased expression of AT-1 can cause an autistic-like phenotype by affecting key neuronal metabolic pathways.

Chapters 5 and 6 detail our method development for neuropeptide characterization and detection. In Chapter 5, we characterized the brain extract from *P. interruptus*, an important aquaculture decapod species with few known preprohormones, with high mass accuracy and resolution data acquired on an LTQ-Orbitrap. A ProSight search was conducted using a flexible algorithm that allows for sequence discrepancy from known sequences against our custom database. In addition to this streamlined, semi-automated sequencing strategy,

we characterized *P. interruptus* by dimethylation-assisted fragmentation and manual *de novo* sequencing using a nanoLC-ESI-Q-TOF, and direct tissue analysis on a MALDI-TOF-TOF, further improving the coverage and resulting in an overall detection of 55 neuropeptides, including 34 novel neuropeptides. The high discovery rate from this species demonstrated the usefulness of the neuropeptide discovery pipeline we developed and highlighted the advantage of utilizing multiple mass spectrometers. The localization of brain neuropeptides were further elucidated with MALDI-MSI. Collectively, our study expands the catalog of crustacean neuropeptides present in *P. interruptus*, and more importantly, presents an approach that can be adapted to explore the neuropeptidomes of other species that possess limited sequence information. Chapter 5 investigates the introduction of methanol and a salt modifier to molecular weight cut-off membrane-based centrifugal filters (MWCO) to enrich sub-microgram peptide quantities. Using a neuropeptide standard, bradykinin, sample loss was reduced over two orders of magnitude with and without undigested protein present. Additionally, a bovine serum albumin (BSA) tryptic digestion was investigated. Twenty-seven tryptic peptides were identified from MALDI mass spectra after enriching with methanol while only two tryptic peptides were identified after the standard MWCO protocol. The strategy presented here enhances recovery from MWCO separation for sub-µg concentration peptide samples.

In summary, the majority of the work described in this dissertation demonstrated the merits of applying mass spectrometry-based proteomic and peptidomic strategies for identifying the important roles of specific proteins and neuropeptides in physiological and pathological processes. The progress in proteomics and peptidomics in the past decades have offered new perspectives and challenges that enabled us to delve further into some

classical questions in neuroscience, such as feeding behavior. Furthermore, these techniques are able to contribute to the diagnosis and treatment of neurodegenerative diseases, such as Alzheimer's disease. These strategies also provided novel tools to elucidate the post-translational modifications in specific disease status, such as autism. I expect that mass spectrometry-based strategies will be widely applied in clinical neurology practice and clinical neuroscience research.

Appendix

List of Publications and Presentations

Publications:

J. Wang, R. Cunningham, H. Zetterberg, S. Asthana, C. Carlsson, O. Okonkwo, L. Li. Label Free Quantitative Comparison of Cerebrospinal Fluid Glycoproteins and Endogenous Peptides in Subjects with Alzheimer's disease, Mild Cognitive Impairment and Healthy Individuals. Ready for submission

J. Wang, H. Ye, Z. Tian, F. Ma, R, Selleck, B. Baldo, L. Li. Mass spectrometric investigation of neuropeptidomic alteration in food intake. Ready for submission

J. Wang, X. Zhong, S. Asthana, C. Carlsson, O. Okonkwo, L. Li. Probing the global proteomic changes and targeted absolute quantification by iDileu in cerebrospinal fluid of preclinical Alzheimer's disease. In preparation

J. Wang, Q. Yu, S. Asthana, C. Carlsson, O. Okonkwo, L. Li. Dynamic proteomic changes of various stages of Alzheimer's disease by 12-plex Dileu. In preparation

J. Wang, S. Yokoyama, R. Cunningham, X. Sun, and L. Li. Comparative analysis of the proteomic changes of amniotic fluid in different gestational age for lung development. Ready for submission

Hullinger, R; Li, M; **Wang, J**; Peng, Y; Dowell, J; Bomba, E; Michelle, H; Burger, C; Chapman, E; Denu, J; Li, L; Puglielli, L.. Neuronal overexpression of the ER membrane acetylCoA transporter, AT-1/SLC33A1, causes an autistic-like phenotype in the mouse. Accepted by *J Exp Med*

H. Ye, **J. Wang**, Z. Zhang, C. Jia, C. Schmerberg, A. Catherman, P. Thomas, N. Kelleher, D. Baro, and L. Li. Defining the Neuropeptidome of the spiny lobster panulirus interruptus brain using a multi-dimensional mass spectrometry-based platform. *J Proteome Res*. **2015** Nov 6; 14(**11**):4776-91.

J. Chen, Q. Yu, **J. Wang**, L. Li. Qualitative and quantitative top-down mass spectral analysis of crustacean hyperglycemic hormones in response to feeding. *Proteomics*. **2014**, May; 14(10): 1185-94.

H. Ye, **J. Wang**, T. Greer, K. Strupat, and L. Li. Visualizing neurotransmitters and metabolites in the central nervous system by high resolution and high accuracy mass spectrometric imaging. *ACS Chem Neurosci*. **2013**, Jul; 4(7): 1049-56.

R. Cunningham, **J.Wang**, D. Wellner, and L. Li. Investigation and reduction of sub-microgram peptide loss using molecular weight cut-off fractionation prior to mass spectrometric analysis.. *Journal of Mass Spectrometry*. **2012**, Oct; 47(10): 1327-32.

Posters:

J. Wang, H. Ye, F. Ma, R, Selleck, B. Baldo, L. Li. (2015). Mass spectrometric

investigation of neuropeptidomicalteration in food intake Poster presented at: American Society for Mass Spectrometry; St. Louis, MO.

J. Wang, O. Okonkwo, L. Li. (2014). Probing the global proteomic changes of cerebrospinal fluid in preclinical Alzheimer's disease. Poster presented at: American Society for Mass Spectrometry; Baltimore, MD.

J. Wang, D. Frost, R. Cunningham, O. Okonkwo, L. Li. (2013). Label Free Quantitative Comparison of Cerebrospinal Fluid Glycoproteins and Endogenous Peptides in Subjects with Alzheimer's disease, Mild Cognitive Impairment and Healthy Individuals. Poster presented at: Society for Neuroscience; San Diego, CA.

J. Wang, S. Yokoyama, R. Cunningham, X. Sun, and L. Li. (2013). Comparative analysis of the proteomic changes of amniotic fluid in different gestational age for lung development. Poster presented at: American Society for Mass Spectrometry; Minneapolis, MN.

B. Chen, H. Ye, M. Gautam, **J. Wang**, R. Thorne, L. Li. (2013). Mapping the Distribution of Intranasally Administered Oxytocin (OXT) in Rat Brain Using MALDI Imaging Mass Spectrometry Poster presented at: American Society for Mass Spectrometry; Minneapolis, MN.

D. Frost, **J. Wang**, L. Li. (2013). Application of DiLeu Isobaric Tandem-Mass Tags to Quantitative Proteomic Analyses of Cerebrospinal Fluid from Alzheimer's Disease Patients Poster presented at: American Society for Mass Spectrometry; Minneapolis, MN.

F. Ma, **J. Wang**, L. Li. (2015). An integrated and high throughput approach for in situ protein digestion, peptide imaging and sequence verification Poster presented at: American Society for Mass Spectrometry; St. Louis, MO.

Q. Yu, K. Branchfield, C. Lietz, **J. Wang**, X. Sun, L. Li. (2014). Mass Spectrometric Characterization of the Neuropeptidome in the Mouse Lung (Mus musculus) Poster presented at: American Society for Mass Spectrometry; Baltimore, MD.

X. Zhong, Q. Yu, **J. Wang**, T. Greer, O. Okonkwo, L. Li. (2015). Novel iDiLeu labeling coupled with high-resolution mass spectrometry for absolute quantification of candidate biomarkers in preclinical Alzheimer's disease Poster presented at: American Society for Mass Spectrometry; St. Louis, MO.

Oral presentation:

J. Wang, H. Ye, L. Li. Defining the Neuropeptidome of the spiny lobster panulirus interruptus brain using a multi-dimensional mass spectrometry-based platform. Oral presentation at: Society for Neuroscience; San Diego, CA.