FROM CRABS TO HAMSTERS: BIOANALYTICAL MASS SPECTROMETRY FOR PEPTIDOMIC ANALYSIS AND BIOMARKER DISCOVERY

by

Joshua John Schmidt

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Pharmaceutical Sciences)

at the

UNIVERSITY OF WISCONSIN-MADISON

2007

Dedicated to the two wonderful women in my life,

my mother and my wife.

Thanks to you for all your support and encouragement

throughout my education.

Acknowledgements

First and foremost, I want to acknowledge my graduate mentor, Lingjun Li. Not only has she been a guiding light for my work at the University of Wisconsin, she has also been a great friend. She has also encouraged me in my research by allowing me the liberty to follow my own research interests.

Next, I would like to thank all of the professors I have had during my graduate career at UW-Madison, especially those on my thesis committee. Without their excellent teaching and mentoring, I would not be the scientist that I am today. Through this tutoring, I have become much more knowledgeable about subjects such as mass spectrometry, molecular separation, and neuroscience. I am indebted to them all for the knowledge and wisdom they have imparted to me through their teaching and mentoring.

In addition to these, I must also very deeply thank Dr. Judd Aiken and his whole lab (especially Allen Herbst) for his support, both monetarily and in many other ways. He and his whole lab have been great friends and support and I will greatly miss spending time with them all. Allen in particular has been a great friend who I have shared much with and hope to continue doing so in the future. Also, Dr. David Page and Sean McIlwain must be thanked for their wonderful insight and assistance with my research. Sean has not only made much of my research possible, but has also been a great friend. I hope we can continue to stay in touch beyond my time here.

Additionally, I would like to thank the University of Wisconsin-Madison School of Pharmacy and all the staff that have helped me on my way through

graduate school. Worthy of mention are Linda Frei of the graduate program, Gary Girdaukas in the Analytical Instrumentation Center, Sharon Hildebrandt of the research animal care facility, and Tom O'Connor in the business office. There are many others that have helped and I thank them as well.

I must also acknowledge all my colleagues in the Li lab. All have provided wonderful support both scientifically and as friends. Without them, the lab would not be what it is today. Those that I would like to especially acknowledge are: Kimberly Naber (Kutz), Stephanie Cape (DeKeyser), and James Dowell, who have all been wonderful colleagues and even greater friends.

Furthermore, I have to thank my whole family for their support. My father has always been there for me to talk to about my problems and has greatly helped me through these times (especially when they related to a broken vehicle). My mother has been an inspiration and driving force for me to continue. Since starting school 22 years ago, she has always encouraged me to do my best and continue learning, in school and beyond. Thanks also to my brothers who have always been behind me and supported me greatly.

Finally and mostly, I would like to acknowledge my wife, Lori. When we began our graduate work, we were not married and were attending different schools. Despite that, she said yes and we were married three years ago and have continued to live three hours apart. Through it all, she has been a wonderful encouragement and has kept me on my toes to continue and finish my graduate work. Without her, all of my graduate work would have been so much harder and I am ever thankful for her being in my life.

Table of Contents

Table of Abbreviations

ACE	Angiotensin Converting Enzyme
acn	Anterior Cardial Nerve
ACP	Anterior Cardial Plexus
ADC	Analog to Digital Conversion
aln	Anterior Lateral Nerve
BSE	Bovine Spongiform Encephalopathy
Cab TRP	*CAncer Borealis* Tachykinin Related Peptide
CID	Collisionally Induced Dissociation
CJD	Creutzfeld-Jacob Disease
CNS	Central Nervous System
CoG	Commissural Ganglia
CPRP	Cab TRP Precursor Related Peptide
CSF	Cerebral Spinal Fluid
CWD	Chronic Wasting Disease
DHB	2,5-Dihydroxy Benzoic Acid
dvn	Dorsal Ventricular Nerve
ECD	Electron Capture Dissociation
ELISA	Enzyme-Linked ImmunoSorbent Assay
ESI	ElectroSpray Ionization
EST	Expressed Sequence Tags
FN	False Negative
FP	False Positive
FTICR	Fourier Transform Ion Cyclotron Resonance
FTMS	Fourier Transform Mass Spectrometer
FWHM	Full Width at Half Mass
HPLC	High Pressure Liquid Chromatography
ICA	In Cell Accumulation
InCAS	Internal Calibration on an Adjacent Standard
ion	Inferior Oesophageal Nerve
ivn	Inferior Ventricular Nerve
kHz	KilloHertZ
kNN	k-Nearest Neighbor
LC	Liquid Chromatograph
m/z	Mass to charge (Z) ratio
M/ΔM	Mass / change (Δ) in Mass
MALDI	Matrix Assisted Laser Desorption Ionization
MHz	MegaHertZ
MMA	Mass Measurement Accuracy
MS	Mass Spectrometer

ms	MilliSecond
MSDM	Mass Spectral Data Miner
mvn	Median Ventricular Nerve
OG	Oesophageal Ganglia
OTMS	Over Thirty Months Scheme
PO	Pericardial Organ
Prion	PRoteinaceous Infectious particle (-on by analogy to virion)
PrP	Prion Protein
PrPC	Prion Protein^{Cellular} (normal)
PrP^{Sc}	Prion Protein^{Scrapie} (abnormal)
PTM	PostTranslational Modification
Q-TOF MS	Quadrupole-Time Of Flight Mass Spectrometer
RF	Radio Frequency
RO	Reverse Osmosis
ROC	Receiver Operator Characteristic
S/N	Signal to Noise Ratio
SDS-PAGE	Sodium Dodecyl Sulfate - PolyAcrylamide Gel Electrophoresis
SG	Sinus Gland
son	Superior Oesophageal Nerve
SORI	Sustained Off-Resonance Irradiation
STG	STomatogastric Ganglia
stn	STomatogastric nerve
STNS	STomatogastric Nervous System
SVM	Support Vector Machine
SWIFT	Stored Waveform Inverse Fourier Transform
TAN	Tree Augmented Networks
TG	Thoracic Ganglia
TME	Transmissible Mink Encephalopathy
TN	True Negative
TOF MS	Time Of Flight Mass Spectrometer
TP	True Positive
TSE	Transmissible Spongiform Encephalopathy
UV	Ultra Violet
V	Volts
vCJD	Variant Creutzfeld-Jacob Disease

FROM CRABS TO HAMSTERS: BIOANALYTICAL MASS SPECTROMETRY FOR PEPTIDOMIC ANALYSIS AND BIOMARKER DISCOVERY

By Joshua John Schmidt

Under the supervision of Professor Lingjun Li

At the University of Wisconsin-Madison

Abstract

While sometimes overlooked, peptidomics and peptide diversity is one of the most important areas of research today. This is because of the vital role that naturally created peptides play in cellular and systemic communication. Peptides have been shown to control and regulate many physiological processes, including feeding, sleep, development, and even cell death. Currently there is a lack of knowledge as to the peptide content of different but related species, as well as peptide distribution throughout nervous system within the same species. Additionally, due to integrative activity of bioactive peptides in physiological processes, there is also great potential for their use as biomarkers to indicate disease infection or progression. To begin filling in these gaps in knowledge, mass spectrometric (MS) methods have been developed for both peptidomic analysis and biomarker identification. Specifically, matrix-assisted laser desorption/ ionization (MALDI) Fourier transform mass spectrometry has been used for a majority of this work due to its sensitivity, high mass measurement accuracy, and high resolution. A MALDI time-of-flight/time-of-flight instrument was also used for tissue-specific peptidomic profiling due to its sensitivity and high throughput nature. Furthermore, as the application of these methods

produced a large set of data, bioinformatics methods were developed and used to process and analyze the results of MS analysis. The combination of mass spectrometry with bioinformatics has enabled rapid comparison of peptidomes of five brachyuran crabs and the determination of peptide distribution among the model nervous system and neurohemal organs of the crustacean *C. borealis*. Furthermore, this approach also allowed the discovery of a panel of classifying features (biomarkers) from prion infected hamsters useful in identifying infected animals. Overall, this research has developed improved methodologies and tools for peptidomic analysis and demonstrated their utility in neurochemical signaling and biomarker discovery in biological fluids.

Chapter One

A Mass Spectrometry-based Peptidomics Approach for the Discovery of Bioactive Peptides and Disease Biomarkers

1.1 Introduction

In the last 20 years or so, there have been many -omics fields of science that have developed. From genomics, to proteomics, now to even conformatiomics, there are a wide range of studies, all with the purpose of comprehensive analysis of the complement of a specific molecular aspect (e.g., genes, proteins, etc.) within an organism. In comparison to genomics and proteomics, peptidomics has been largely overlooked, with a great need to develop specific and effective analytical methodologies. Peptidomics aims at a simultaneous visualization and identification of the entire complement of peptides in a cell or tissue, including their post-translational modifications (PTMs). There are currently many methods for the analysis of polypeptides, but these have been primarily directed at the examination of whole proteins and peptides that result from the proteolytic cleavage of proteins. While many of these methods can be attempted in the study of naturally occurring peptides, they often fall short as they are not designed to deal with the many PTMs that occur in most bioactive peptides. As a result, a large portion of the peptidome has been unnoticed and disregarded.

Unfortunately, it is this ignored peptidome that could be considered to hold the most promise as potential drug targets and/or disease indicative biomarkers. As our knowledge about the intracellular function and intercellular communication grew, it became apparent that peptides are much more involved throughout the workings of an organism than were originally thought. Since this realization, peptides have been found in all complexities of organisms, from unicellular

organisms all the way up to mammals. Furthermore, they have also been found to be integrally involved in all bodily functions, from feeding[1-3] to blood pressure regulation[4-6] and development[7-9] to apoptosis.[10-12] It is this importance in systemic functions that require more research to be performed to elucidate and characterize the structures and functions of these bioactive peptides. With the proteomics tools that are already well established, it would seem that this would be a relatively easy proposition, but the differences between a peptide produced strictly through proteolytic cleavage and a naturally derived peptide are great enough that many of the classical proteomic methods simply do not work.

Naturally created peptides differ from cleavage products in that they often contain PTMs such as C-terminal amidation or N-terminal acetylation.[13; 14] These modifications on bioactive peptides give rise to one of the challenges that bar the use of traditional proteomics methods. Because the genomic origin of PTMs is still largely a mystery, we can rarely use genetic data to predict the bioactive peptides present. Furthermore, a great deal of peptidomic research is performed on organisms with a lack of genomic information and making any genetic comparison impossible. Without genetic information assisting researchers, they are left to *de novo* sequence peptides and elucidate PTMs, all while lacking an extensive set of methodologies specific for bioactive peptide research.

Because bioactive peptides are vital to many cellular functions and signals they hold a great deal of promise for potential new drug targets[15] and/or disease indicative biomarkers.[16; 17] For this reason, the aim of peptidomic research should be at the development of sample preparation, analysis, and

characterization methods that succeed in overcoming the challenges that bioactive peptides present.[18] Currently, while many proteomic methods exist, there are limitations when applying them to peptidomics. If peptidomics is to play a more significant role in drug discovery and biomarker identification, an increase in the repertoire of peptidomic methods needs to be observed in the scientific community.

1.2 Proteomics Methods and their Limitations for Peptidomics Research

If one performs a literature search on the term "peptidomics," it is very evident that this term was not even used until after the turn of the century. In fact, even "proteomics" was not in use until after 1996. Although this does not seem a great difference in time since these two closely related areas of research began, proteomics has naturally been many technological steps ahead of peptidomics. While elucidating the genomes of many species was monumental, it was quickly realized that genetic information provided the blueprint but is a far cry from the finished product, proteins. It was at this point that proteomics came on the scene in order to begin describing the cellular machinery in whole, along with all its post-translational modifications. This interest in characterizing protein content led to the development of numerous methods directed towards amino acid sequencing and identifying the PTMs as well as the extended structure of such proteins. To perform this research, investigators began using Edman degradation as the main sequencing method but soon realized the shortcomings of this method (the large amount of pure material necessary and the extended

time required). Fortunately, several other methods existed that were readily adapted to proteomics and continue to be the main methods for this research. These methods include gel electrophoresis, HPLC, tryptic digestion, as well as several mass spectrometry methods. Additionally, two schemes of proteomics research, bottom up[19] and top down proteomics,[20] came about.

One of the first methods for proteomic research was gel electrophoresis which was adapted from genetic research. In order to deal with the extreme chemical complexity encountered in proteomics research, two dimensional gel electrophoresis was developed.[21] This method separates proteins based on two properties (such as isoelectric point and size), providing more information about the protein and higher resolving power than one dimensional electrophoresis. Since its inception, this method has been one of the most widely used proteomics tool as it provides data about two properties of the proteins at the same time. Furthermore, by performing 2D gel electrophoresis on a mixed sample (a cell lysate for example), a map of the proteins contained in that type of cell could be produced and used in the future to identify proteins in similar samples. Unfortunately, while this method has been incredibly useful to proteomic research, it is not suitable for peptidomic research. The reason for this is that peptides are much too small for SDS-PAGE and tend to run off the gel so fast that they are not even observed. It is possible to greatly increase the acrylamide concentration in a gel, but because no protocols have been established for peptide separation on a polyacrylamide gel substrate and peptide

movement is much more difficult to control on a gel, it is not a practical method to employ for peptidomics.

In addition to 2D gel electrophoresis, recent mass spectrometry technologies coevolved with proteomics and thus many of the novel MS methods of that time were aimed at solving the proteomic challenges confounding researchers at the time. Along with mass spectrometry came the schemes of research termed "bottom up" and "top down" proteomics. The difference between these schemes is whether the proteins are proteolytically digested (to produce peptides) prior to MS analysis or whether they are analyzed intact. Interestingly, although these schemes produce different types of samples, both have made extensive use of mass spectrometry and its various methods of ionization, mass analysis, and detection. For bottom up proteomics, matrix assisted laser desorption/ionization (MALDI) – based MS methods are used to perform fingerprinting analysis of relatively simple protein digests while liquid chromatography (LC) coupled to electrospray ionization (ESI) – based MS methods are used to analyze highly complex samples, such as digested cell lysates.[22] Top down proteomics on the other hand, demands slightly different methods but has also made extensive use of both MALDI and LC-ESI technologies for its analysis of intact proteins.[20] A wide range of mass analyzers including time-of-flight, Fourier transform ion cyclotron resonance, ion trap, and quadrupole have been used for both proteomics schemes. Furthermore, when linked to an LC system, both top down and bottom up proteomics benefits greatly from separation and then MS analysis. Whether to separate and analyze a

whole cell lysate or a tryptic digest, LC-MS has become the quintessential tool for proteomics research today, both bottom up and top down.[23-27]

Though some of the methods developed for proteomics can be useful, they do not perform well enough to produce the results necessary for knowledge about bioactive peptides to reach its full potential. For example, one of the most popular 2D separation methods for tryptic peptides is SCX-RP, but will not work for bioactive peptides due to common PTMs such as C-terminal amidation or N-terminal acetylation. Additionally, because of the fact that naturally occurring peptides do not all contain a lysine or arginine C-terminal residue as tryptic peptides do, ionization and fragmentation efficiencies are far less, thus making MS analysis much more difficult. This difference in C-terminal residues also creates problems when searching databases as most of them are currently set to search for tryptic peptides (with their known lysine or arginine C-termini) or a defined C-terminus due to a specific cleavage enzyme instead of naturally produced peptides (with their varying C-termini). To counter this, novel methods will need to be developed with the aim of enabling researchers to detect and characterize bioactive peptides with all of their many unique features.

1.3 Current Bioactive Peptidomic Methodology

Sample Cleanup

Generally, a peptide is considered a polymer from two to twenty or so amino acids (AAs) long but can be as long as 50 or more AAs. As stated above, peptides are commonly produced and analyzed in the course of proteomics

research, these peptides are artificially created by enzymatic digestion and thus differ from the natural peptides that are produced *in vivo* through cellular processes. Because proteolytic cleavage of proteins requires a fairly controlled environment, the samples produced by proteolytic digestion are relatively clean and easy to separate by HPLC and/or analyze by MS. This is not the case with bioactive peptides which are produced or secreted *in vivo* in a dynamic and tissue/cell specific fashion. This in turn, creates samples that are incredibly complex and require extensive preparation prior to being analyzed. Some of the issues that must be dealt with when attempting to analyze bioactive peptides are the pH, the presence of salts (especially when performing research on marine organisms), and the interference from high abundance cellular proteins. Additionally, when preparing a sample, researchers must also account for naturally occurring proteolytic enzymes in tissue which, if left unchecked, will begin cleaving proteins into peptides and making the sample even more complicated. All of these factors require unique methods for the preparation of samples for native peptide analysis. Occasionally, it is possible to avoid these challenges and analyze peptide content directly from tissues of interest, but these analyses often will not reflect the complete peptidome and the sample must be investigated more thoroughly.

One way to circumvent many sample preparation challenges is to extract the peptides directly, thus removing them from other cellular machinery. Because the peptides must be extracted while everything else is left behind, solvents must be used that solubilize peptides but not proteins and lipids. Thus,

solvents such as methanol, ethanol, and acetone are used, often in combination with water and/or some acid. One of the most common extraction solvents is acidified methanol which is comprised of 90:9:1 methanol: glacial acetic acid: water.[28-30] Through a process of tissue grinding, centrifugation, and pellet re-extraction, it is possible to extract a majority of the native peptides while leaving behind the cellular junk. If performed carefully, the resulting extract is quite clear and free of large proteins and most lipids. Occasionally in more lipid rich tissue, these lipids can remain present but separated from the extract via hard centrifugation. Another method requires boiling the tissue in acetone and then centrifuging to remove excess cellular particles.[31] Although several extraction techniques have been developed to isolate native peptides, none of them are perfect and many require modification to suit the needs of the research.

Once an extract has been made, it is often necessary to separate the sample using HPLC in order to enable detection of peptides of lower abundance. If the extract itself is MS analyzed, the resulting spectra frequently contain the more prominent peptides from a sample. By HPLC separating the extract, it is possible to enable the observation of less abundant peptides in the mass spectra of individual fractions. As a result of this benefit, a majority of the peptides in a sample can not only be analyzed, but characterized as well.

In some cases though, peptide extraction from a sample is very difficult, as in the case of biological fluids such as blood or cerebral spinal fluid. Many biological fluids do contain cells that can be lysed and extracted, but generally the peptides of interest are contained in the fluid itself. Unfortunately, the

dynamic range of proteins and peptides in bodily fluids creates many of the difficulties in analysis of these samples.[32] In these cases it may be difficult to extract peptides and it is easier to remove the complicating features of such samples. Obviously, it is relatively easy to adjust pH to the necessary level by adding dilute HCl or NaOH as needed. Instead it is the salts in such samples that can greatly complicate matters. HPLC can remove all the salts, but sample sizes and highly abundant proteins vs. low peptide concentration can often prohibit its use in some situations. In order to address the challenge of analyzing low-abundance species in the presence of highly- abundant proteins in samples, researchers have developed methods of immunodepletion whereby the most abundant proteins are bound by antibody coated beads while the peptides remain in the sample.[33-35] Although this method works well for removing a great deal of the abundant proteins, peptides are also often removed due to the abundant proteins, such as albumin, binding to peptides and acting as carrier proteins.[36; 37] This issue of dynamic range remains to be addressed adequately for bioactive peptide extraction.

Another method that has been developed as a way to remove salts and enrich peptide samples is to pass the sample through a bed of affinity coated beads (i.e. C8, C18, antibodies, etc) similar to those found in an HPLC column.[38-43] By doing so peptides are bound to the beads, everything else (i.e. salts) is washed away, and then the peptides are eluted off the beads. This has the effect of both removing the salts from a sample as well as concentrating the peptides if necessary. Currently, these procedures have been adapted to small

columns and even pipette tips enabling smaller samples to be purified and thus analyzed. Unfortunately, these methods still do not work reliably and often sample loss requires repetition to obtain a large enough sample. A method of complex biological sample cleanup remains to be created for adequate bioactive peptide research.

As is obvious, there are several methods for sample cleanup that have already been developed for peptidomic research, but most of these methods are adapted from proteomics research and thus have many limitations. One example of this is the widely used proteomics method of 2D LC, with one dimension being strong-cation exchange separation and the other dimension a reverse-phase separation.[23; 44] While it can be applied to bioactive peptides, it does not always provide valuable results due to PTMs such as C-terminal amidation or N-terminal acetylation. Overall, methods for peptidomic sample preparation need to be improved and made more amenable to the challenges that bioactive peptides present, especially the small size of peptides and their common PTMs.

Peptidomic Mass Spectrometry

After preparing a sample, whether by extraction, separation, or enrichment, it must be analyzed in some way for the peptides it contains. Unlike proteomics with many detection and characterization methods, peptidomics analysis is currently carried out primarily by mass spectrometry.

Besides providing accurate mass measurement and matching against previously known peptides for peptide identification, another key feature of mass spectrometry is the ability to fragment and detect those resulting product ions. For peptidomics, some of the most useful fragmentation techniques are collisionally induced dissociation (CID) and electron capture dissociation (ECD). CID is useful as it breaks the peptide bond producing a b- and y- ion series that makes peptide sequencing relatively easy. ECD on the other hand breaks the bond between the amide and primary carbon thus producing c- ions and z radicals.[45] While not as straightforward to sequence a peptide, they are equally useful.[46; 47] If necessary, multiple fragmentation methods can be used in a complementary fashion for sequencing.

One of the more prolific instruments currently used in peptidomics research is a MALDI front end with a TOF mass analyzer. This instrument excels in its ability to efficiently ionize molecules in the primary peptide mass range followed by sensitive analysis of complex mixtures of peptides with relatively good resolution and mass accuracy. Additionally, analyses using MALDI-TOF can be performed rapidly thus cutting the data acquisition time down dramatically and increasing throughput. This is important when attempting to analyze a large number of fractions from multiple samples. If time is not an issue and/or high resolution and mass measurement accuracy are desired, a MALDI-FTICR instrument can be used.

Another set of techniques that make use of MALDI technology are those of direct tissue analysis, single cell MS, and MS imaging. Because MALDI only

requires the application of a matrix to a sample, whether it is liquid or solid, cells and tissue can easily be analyzed. Direct tissue methods allow for peptidomic profiling of tissue and whole organs (depending on their size) with no peptide extraction and limited sample preparation.[29] Methods developed for the analysis of individual cells provide extensive peptidomic information from an individual cell without interference from surrounding tissue.[48; 49] As an extension of both of these methods, MS imaging has developed. This method is carried out by applying matrix to a tissue slice and then rastering the MALDI laser across it and collecting spectra from each spot hit by the laser. After all the spectra have been collected, software is used to compile the data and display a heat map of the location of the masses detected. By doing so, researchers can determine the spatial distribution of peptides within a tissue of interest. One example of this is the identification of cancer specific peptides within a tumor in a mouse brain.[50-53] The speed of a MALDI-TOF, along with its ability to analyze tissue directly, makes it very amenable to MS imaging. In the future, peptidomic MS imaging promises to be one of the best methods for identifying peptide biomarkers closely associated with disease morphology.

In addition to MALDI instruments, CapLC-ESI QTOF offers another method of peptidomic analysis. One advantage it offers over MALDI analysis is its ability to analyze HPLC separations in an on-line fashion. While HPLC fractions are easily analyzed off-line using a MALDI instrument, the separations often require a large injection volume, the collection and preparation of samples, followed by the spotting of all fractions onto a MALDI target. By being able to

analyze fractions as they come off a column, the injection volume is greatly reduced and throughput increased as a result. Furthermore, if larger peptides need to be detected, ESI has the ability to multiply charge molecules and thus reduce the observed *m/z* ratio. This multiple charging of molecules also assists in peptide sequencing due to the fact that fragments have a higher chance of retaining a charge and thus being detectable, i.e. more fragment ions are observed.

When performing peptide analysis and characterization, it is difficult to obtain adequate fragmentation spectra allowing for complete determination of the amino acid sequence. This is due to the fact that it requires a varying amount of energy to break the peptide bonds between the different amino acids. This results in some fragments being much more abundant than others which can often make *de novo* sequencing nearly impossible. Thus, it is necessary to perform other tests to learn more about the peptide sequence. One important method for peptidomics research is chemical derivatization whereby the peptide is allowed to react in a known manner with certain reagents. Three of the most popular derivatizations are formaldehyde labeling, methylation, and acetylation.[28; 54; 55] The information that needs to be gathered will determine which of these labeling schemes is used. In addition to improving fragmentation and increasing chemical information, using various derivatizations in conjunction with isotope-encoded reagents also enables relative quantification of peptides in two different samples.[56-59] In general, chemical derivatization offers another tool to peptidomics researchers to assist in gathering the necessary information.

Overall, mass spectrometry remains to be one of the most powerful and versatile technologies currently available for molecular identification and characterization. The flexibility of mass spectrometry stems from the ability of researchers to mix and match ionization techniques with mass analyzers depending on their research and the instruments available. One unifying factor of all mass spectrometers being used for peptidomics is their relative high sensitivity in the peptide mass range (*m/z* 500 to >5000) as well as the capability to produce data that directly correlates to the peptide's sequence. The sensitivity is necessary for the analysis of peptides through the range of abundances found *in vivo*. Additionally, the ability to determine amino acid sequence is vital if investigators are to identify the nature of the peptide and whether it can be utilized as a drug target or disease biomarker. Exactly which instrument is used is often determined by the sample and how it is prepared for analysis. If, for example, peptides are to be analyzed directly from tissue, it is necessary to use MALDI as the ionization technique as it does not require the sample to be in liquid form. Furthermore, a sample that must be separated first may benefit greatly from being analyzed on an LC-ESI MS and thus reducing the required injection volume. In all, there are a great number of instrumental variations to suit many of the various areas of peptidomics research.

Data Analysis and Bioinformatics

In the process of performing peptidomics research, a great deal of data is produced that can make data analysis a daunting task. As proteomics

developed, many software packages were developed that allowed users to submit their data, piece together peptides into complete protein sequences, and then search a database for the encoding gene. This has greatly facilitated proteomics research, but due to the challenges of peptidomics application of these tools has been almost impossible. Fortunately, numerous bioinformatics software packages have developed for peptidomics to combat the large data sets and enabled researchers to more quickly process data and draw the necessary conclusions.[60-62] By using such software, users can input a substantial amount of data, visualize it simultaneously, and carry out the needed functions in a timely and error free manner. Some of these functions include the merging of overlapping data, comparing and contrasting data of varying classes, and the compiling of multiple data sources (i.e. retention time, *m/z* values, relative peak abundance, etc.). Additionally, when dealing with a large amount of data in a scientific setting, it is often necessary to prove statistical relevance, as well as provide some amount of statistical values in evidence of your conclusions. Bioinformatics packages also allow researchers the ability to easily perform the required statistical analyses to make their data scientifically acceptable.[60-62]

Bioinformatics has been one of the main tools in peptidomics research,[16; 17; 63] but unfortunately is still lacking when compared to the tools used in proteomics research. Because a great deal of proteomics research is being performed in common model organisms, researchers have the ability to search their protein sequences against a database of previously identified proteins or theoretical proteins derived from genetic information. Peptidomics on the other

hand has no such database through which to search. Investigators can occasionally search the protein databases for short matches to their peptide sequences, but because these sequences are so short, an inordinate amount of matches are made of which many are incorrect. For many model systems there is no genomic information and this process is completely impossible. As a result, there is a real and apparent need for a bioactive peptide database for the more common model systems. Bioinformatics can and does assist greatly in peptidomics research, but without a better understanding of the peptides and their origins, many vital conclusions cannot be made.

Although the fully sequenced genomic information may not be available, it is possible to obtain the sequences of the expressed portion of genes coding for peptides, or expressed sequence tags (ESTs) by reverse translating the amino acid sequences of peptides to mRNA. Once ESTs have been produced, databases can be searched for similarities in species that have been more thoroughly researched. Furthermore, ESTs can be used to identify and sequence the complete genes coding for these peptides, and their precursors. By creating ESTs and characterizing the associated gene, a great deal of information can be gathered on a species that has otherwise not been genetically described. Overall, ESTs library searching provides an excellent addition to peptidomics research as it allows investigators to gather data on the genomics and proteomics of species that have not been as fully described as well as enabling them to perform phylogenetic research on a species and learn better how they are genetically related to other similar species.[78-81]

1.4 Bioactive Peptidomic Applications

Peptide Family Characterization

One observation that has resulted from peptidomics research is the homology between some groups of peptides, especially across multiple related species. Thus peptides with high homology have been grouped into families. Although peptides within families sometimes have similar function, differences in tissue location and time of presence indicate that many have very different functions.[12; 64-66] As a result, one goal of peptidomics is to identify all the members of these peptide families. Currently, while many peptide families have been identified, the full complement of peptides in each family has not yet been fully identified as is evidenced by the continual additions to peptide families that show up regularly in the literature. Additionally, of high interest are those peptide families that have members in many species. Some species gaps are as far apart as vertebrates and invertebrates as is demonstrated by the RFamide family of peptides. These widespread peptide families point towards the importance of the peptide families in organisms, but what their function of is still quite often a mystery.

Peptides in the same families are thought to have similar functions but are often differentially expressed between species, organs, and occasionally within an organ indicating differences in how and where they work. By gathering data on this distribution of a peptide family, conclusions can be made regarding the function of the peptides as well as their diversification through speciation. Furthermore, by understanding how closely related species can differ in their

peptide family distribution, investigators can gain insight into the differences between species and why drugs affect each species differently. This information would not only be invaluable to pharmaceutical companies, but would also benefit human health by preventing some of the negative drug interactions that can occur when testing new drugs in humans.

Drug Target Elucidation

Due to their importance in systemic signaling, bioactive peptides play an important role as drug targets.[15; 67-70] Through the modification of peptide signaling, it may be possible to adjust the effect peptides have on an organism and thus change the course of diseases. It is thought that peptides have great control over such bodily functions as feeding, sleep, and how the brain communicates to the rest of the body. Because of this control peptides have in the body, efforts should be made to either mimic them or affect how they operate.

One excellent example of peptide modulation is the Angiotensin family of peptides in humans which cause vasoconstriction and thus increase blood pressure. Because this particular peptide and its function was so well understood, investigators developed angiotensin converting enzyme (ACE) inhibitors to reduce the amount of anigiotensin produced in the body. By administering ACE inhibitors, doctors can effectively reduce blood pressure in patients with hypertension.[71-73] This illustration shows the important effect that bioactive peptides play in the body and what is possible through a better understanding of their function and how they are created. Only through

increased research in bioactive peptidomics will these goals be met and new drug targets discovered.

Biomarker Discovery

Finally, because peptides are integrated with a large part of cellular machinery, a change in the functions of a cell will presumably change the peptides produced and/or secreted by the cell. Thus, if that change in peptide composition can be detected, infection can be detected earlier and possibly more reliably. If the peptide can be detected in the blood, or some other bodily fluid, it may even be possible to identify infected organisms without invasive operations. Furthermore, in the case of treating patients, the identification of a biomarker that is associated with disease progression can greatly assist in determining how well a patient is responding to a particular medication. By enabling doctors to identify the efficacy of drugs on patients, they can more quickly administer the drug therapy that will be most effective for that patient. From identifying infected organisms to monitoring how well a drug is treating a disease, peptidomics promises to play an important role in the future of biomarker discovery.[17; 74-77]

1.5 Conclusions

In conclusion, although there has been a great deal of methodologies developed for earlier -omics fields, peptidomics remains lacking many of the methods necessary to reach its full potential. While many proteomics methods have been adapted to peptidomics, the challenges of naturally occurring

bioactive peptides such as their small size and the often present PTMs prevent these methods from working well enough to fully identify and characterize the full complement of an organism's peptides. Obviously though, peptides play an important role in many processes, both cellular and systemic. It is for this reason that they hold great promise as drug targets, disease biomarkers, and a general improved understanding of how the body works. The methods currently being used have set the stage for the future of peptidomics. As mass spectrometry will certainly continue to play a large role in the analysis of peptides, efforts should be focused to further develop methodology of sample preparation, MS analysis, and data processing to gain an adequate knowledge of peptides and their function.

1.6 References

1. Meister, B. (2007) Neurotransmitters in key neurons of the hypothalamus that regulate feeding behavior and body weight. *Physiology & Behavior*

2. Shiraishi, J. I., Yanagita, K., Fujita, M., and Bungo, T. (2007) Central insulin suppresses feeding behavior via melanocortins in chicks. *Domestic Animal Endocrinology*

3. Proekt, A., Vilim, F. S., Alexeeva, V., Brezina, V., Friedman, A., Jing, J., Li, L., Zhurov, Y., Sweedler, J. V., and Weiss, K. R. (2005) Identification of a new neuropeptide precursor reveals a novel source of extrinsic modulation in the feeding system of *Aplysia*. *J Neurosci*25, 9637-9648

4. Freitas, R. R., Lopes, K. L., Carillo, B. A., Bergamaschi, C. T., Carmona, A. K., Casarini, D. E., Furukawa, L., Heimann, J. C., Campos, R. R., and Dolnikoff, M. S. (2007) Sympathetic and renin-angiotensin systems contribute to increased blood pressure in sucrose-fed rats. *American Journal of Hypertension : Journal of the American Society of Hypertension*20, 692-698

5. Wang, Y., and Wang, D. H. (2007) Neural control of blood pressure: Focusing on capsaicin-sensitive sensory nerves. *Cardiovascular & Hematological Disorders Drug Targets*7, 37-46

6. Ben-Abraham, R., Hadad, E., Weinbroum, A. A., Efrati, O., and Paret, G. (2003) Vasopressin in cardiac arrest and vasodilatory shock: A forgotten drug for new indications. *Isr Med Assoc J*5, 272-276

7. Dupin, E., Calloni, G., Real, C., Goncalves-Trentin, A., and Le Douarin, N. M. (2007) Neural crest progenitors and stem cells. *Comptes Rendus Biologies*330, 521-529

8. Pangas, S. A. (2007) Growth factors in ovarian development. *Seminars in Reproductive Medicine*25, 225-234

9. Baggerman, G., Boonen, K., Verleyen, P., De Loof, A., and Schoofs, L. (2005) Peptidomic analysis of the larval *Drosophila melanogaster* central nervous system by two-dimensional capillary liquid chromatography quadrupole time-of-flight mass spectrometry. *J Mass Spectrom*40, 250-260

10. Dinnen, R. D., Drew, L., Petrylak, D. P., Mao, Y., Cassai, N., Szmulewicz, J., Brandt-Rauf, P., and Fine, R. L. (2007) Activation of targeted necrosis by a p53 peptide: A novel death pathway which circumvents apoptotic resistance. *Journal of Biological Chemistry*

11. Ko, J. K., Choi, K. H., Pan, Z., Lin, P., Weisleder, N., Kim, C. W., and Ma, J. (2007) The tail-anchoring domain of Bfl1 and HCCS1 targets mitochondrial membrane permeability to induce apoptosis. *Journal of Cell Science*

12. Pavelic, J., Matijevic, T., and Knezevic, J. (2007) Biological & physiological aspects of action of insulin-like growth factor peptide family. *The Indian Journal of Medical Research*125, 511-522

13. Lloyd D. Fricker, J.L., Hui Pan,Fa-Yun Che. (2006) Peptidomics: Identification and quantification of endogenous peptides in neuroendocrine tissues. *Mass Spectrometry Reviews*25, 327-344

14. Hummon, A. B., Amare, A., and Sweedler, J. V. (2006) Discovering new invertebrate neuropeptides using mass spectrometry. *Mass Spectrom Rev*25, 77-98

15. Hokfelt, T., Bartfai, T., and Bloom, F. (2003) Neuropeptides: Opportunities for drug discovery. *The Lancet Neurology*2, 463-472

16. Geho, D. H., Liotta, L. A., Petricoin, E. F., Zhao, W., and Araujo, R. P. (2006) The amplified peptidome: The new treasure chest of candidate biomarkers. *Current Opinion in Chemical Biology*10, 50-55

17. Schrader, M., and Selle, H. (2006) The process chain for peptidomic biomarker discovery. *Disease Markers*22, 27-37

18. Soloviev, M., and Finch, P. (2006) Peptidomics: Bridging the gap between proteome and metabolome. *Proteomics*6, 744-747

19. Zhu, W., Reich, C. I., Olsen, G. J., Giometti, C. S., and Yates, J. R.,3rd. (2004) Shotgun proteomics of *Methanococcus jannaschii* and insights into methanogenesis. *J Proteome Res*3, 538-548

20. Kelleher, N. L. (2004) Top-down proteomics. *Analytical Chemistry*76, 197A-203A

21. Gygi, S. P., Corthals, G. L., Zhang, Y., Rochon, Y., and Aebersold, R. (2000) Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc Natl Acad Sci U S A*97, 9390-9395

22. Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature*422, 198-207

23. Delahunty, C., and Yates, J. R.,3rd. (2005) Protein identification using 2D-LC-MS/MS. *Methods*35, 248-255

24. Johnson, R. W.,Jr, Ahmed, T. F., Miesbauer, L. J., Edalji, R., Smith, R., Harlan, J., Dorwin, S., Walter, K., and Holzman, T. (2005) Protein fragmentation via liquid chromatography-quadrupole time-of-flight mass spectrometry: The use of limited sequence information in structural characterization. *Anal Biochem*341, 22-32

25. Bischoff, R., and Luider, T. M. (2004) Methodological advances in the discovery of protein and peptide disease markers. *Journal of Chromatography B*803, 27-40

26. McDonald, W. H., and Yates, J. R.,3rd. (2002) Shotgun proteomics and biomarker discovery. *Dis Markers*18, 99-105

27. le Coutre, J., Whitelegge, J. P., Gross, A., Turk, E., Wright, E. M., Kaback, H. R., and Faull, K. F. (2000) Proteomics on full-length membrane proteins using mass spectrometry. *Biochemistry*39, 4237-4242

28. Fu, Q., Kutz, K. K., Schmidt, J. J., Hsu, Y. W., Messinger, D. I., Cain, S. D., de la Iglesia, H. O., Christie, A. E., and Li, L. (2005) Hormone complement of the *Cancer productus* sinus gland and pericardial organ: An anatomical and mass spectrometric investigation. *J Comp Neurol*493, 607-626

29. Kutz, K. K., Schmidt, J. J., and Li, L. (2004) In situ tissue analysis of neuropeptides by MALDI FTMS in-cell accumulation. *Anal Chem*76, 5630-5640

30. Fu, Q., Goy, M. F., and Li, L. (2005) Identification of neuropeptides from the decapod crustacean sinus glands using nanoscale liquid chromatography tandem mass spectrometry. *Biochem Biophys Res Commun*337, 765-778

31. Dowell, J. A., Heyden, W. V., and Li, L. (2006) Rat neuropeptidomics by LC-MS/MS and MALDI-FTMS: Enhanced dissection and extraction techniques coupled with 2D RP-RP HPLC. *Journal of Proteome Research*5, 3368-3375

32. Anderson, N. L., and Anderson, N. G. (2002) The human plasma proteome: History, character, and diagnostic prospects. *Mol Cell Proteomics*1, 845-867

33. Ramstrom, M., Hagman, C., Mitchell, J. K., Derrick, P. J., Hakansson, P., and Bergquist, J. (2005) Depletion of high-abundant proteins in body fluids prior to liquid chromatography fourier transform ion cyclotron resonance mass spectrometry. *J Proteome Res*4, 410-416

34. Greenough, C., Jenkins, R. E., Kitteringham, N. R., Pirmohamed, M., Park, B. K., and Pennington, S. R. (2004) A method for the rapid depletion of albumin and immunoglobulin from human plasma. *Proteomics*4, 3107-3111

35. Wang, Y. Y., Cheng, P., and Chan, D. W. (2003) A simple affinity spin tube filter method for removing high-abundant common proteins or enriching low-abundant biomarkers for serum proteomic analysis. *Proteomics*3, 243-248

36. Mehta, A. I., Ross, S., Lowenthal, M. S., Fusaro, V., Fishman, D. A., Petricoin, E. F.,3rd, and Liotta, L. A. (2003) Biomarker amplification by serum carrier protein binding. *Dis Markers*19, 1-10

37. Duan, X., Yarmush, D., Berthiaume, F., Jayaraman, A., and Yarmush, M. L. (2005) Immunodepletion of albumin for two-dimensional gel detection of new mouse acute-phase protein and other plasma proteins. *Proteomics*5, 3991-4000

38. Lee, W. -., and Lee, K. H. (2004) Applications of affinity chromatography in proteomics. *Analytical Biochemistry*324, 1-10

39. Feng, S., Ye, M., Zhou, H., Jiang, X., Jiang, X., Zou, H., and Gong, B. (2007) Immobilized zirconium ion affinity chromatography for specific enrichment of phosphopeptides in phosphoproteome analysis. *Mol.Cell.Proteomics*

40. Xiao, X., Yang, X., Liu, T., Chen, Z., Chen, L., Li, H., and Deng, L. (2007) Preparing a highly specific inert immunomolecular-magnetic beads for rapid detection and separation of *S. aureus* and group G *Streptococcus*. *Applied Microbiology and Biotechnology*75, 1209-1216

41. Ogata, Y., Charlesworth, M. C., and Muddiman, D. C. (2005) Evaluation of protein depletion methods for the analysis of total-, phospho- and glycoproteins in lumbar cerebrospinal fluid. *J Proteome Res*4, 837-845

42. Terry, D. E., Umstot, E., and Desiderio, D. M. (2004) Optimized sample-processing time and peptide recovery for the mass spectrometric analysis of protein digests. *J Am Soc Mass Spectrom*15, 784-794

43. Baczek, T. (2004) Fractionation of peptides in proteomics with the use of pI-based approach and ZipTip pipette tips. *Journal of Pharmaceutical and Biomedical Analysis*34, 851-860

44. Chen, E. I., Hewel, J., Felding-Habermann, B., and Yates, J. R.,3rd. (2006) Large scale protein profiling by combination of protein fractionation and multidimensional protein identification technology (MudPIT). *Mol Cell Proteomics*5, 53-56

45. Zubarev, R. A. (2004) Electron-capture dissociation tandem mass spectrometry. *Curr Opin Biotechnol*15, 12-16

46. Bakhtiar, R., and Guan, Z. (2005) Electron capture dissociation mass spectrometry in characterization of post-translational modifications. *Biochemical and Biophysical Research Communications*334, 1-8

47. Kjeldsen, F., Haselmann, K. F., Budnik, B. A., Sorensen, E. S., and Zubarev, R. A. (2003) Complete characterization of posttranslational modification sites in the bovine milk protein PP3 by tandem mass spectrometry with electron capture dissociation as the last stage. *Anal Chem*75, 2355-2361

48. Li, L., Garden, R. W., and Sweedler, J. V. (2000) Single-cell MALDI: A new tool for direct peptide profiling. *Trends Biotechnol*18, 151-160

49. Rubakhin, S. S., Greenough, W. T., and Sweedler, J. V. (2003) Spatial profiling with MALDI MS: Distribution of neuropeptides within single neurons. *Anal Chem*75, 5374-5380

50. Chaurand, P., Fouchecourt, S., DaGue, B. B., Xu, B. J., Reyzer, M. L., Orgebin-Crist, M. C., and Caprioli, R. M. (2003) Profiling and imaging proteins in the mouse epididymis by imaging mass spectrometry. *Proteomics*3, 2221-2239

51. Chaurand, P., and Caprioli, R. M. (2002) Direct profiling and imaging of peptides and proteins from mammalian cells and tissue sections by mass spectrometry. *Electrophoresis*23, 3125-3135

52. Caprioli, R. M., Farmer, T. B., and Gile, J. (1997) Molecular imaging of biological samples: Localization of peptides and proteins using MALDI-TOF MS. *Anal Chem*69, 4751-4760

53. Pierson, J., Norris, J. L., Aerni, H. R., Svenningsson, P., Caprioli, R. M., and Andren, P. E. (2004) Molecular profiling of experimental parkinson's disease: Direct analysis of peptides and proteins on brain tissue sections by MALDI mass spectrometry. *J Proteome Res*3, 289-295

54. Dekeyser, S. S., and Li, L. (2006) Matrix-assisted laser desorption/ionization fourier transform mass spectrometry quantitation via in cell combination. *Analyst*131, 281-290

55. Fu, Q., and Li, L. (2005) 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*77, 7783-7795

56. Wei, H., Nolkrantz, K., Parkin, M. C., Chisolm, C. N., O'Callaghan J, P., and Kennedy, R. T. (2006) Identification and quantification of neuropeptides in brain tissue by capillary liquid chromatography coupled off-line to MALDI-TOF and MALDI-TOF/TOF-MS. *Anal Chem*78, 4342-4351

57. Ji, C., and Li, L. (2005) Quantitative proteome analysis using differential stable isotopic labeling and microbore LC-MALDI MS and MS/MS. *J Proteome Res*4, 734-742

58. Hamdan, M., and Righetti, P. G. (2002) Modern strategies for protein quantification in proteome analysis: Advantages and limitations. *Mass Spectrom Rev*21, 287-302

59. DeKeyser, S.S.,and Li, L. (2006) Matrix-assisted laser desorption/ionization fourier transform mass spectrometry (MALDI FTMS) quantitation via in-cell combination (QUICC). *The Analyst*In press.,

60. Hauskrecht, M., Pelikan, R., Malehorn, D. E., Bigbee, W. L., Lotze, M. T., Zeh, H. J., Whitcomb, D. C., and Lyons-Weiler, J. (2005) Feature selection for classification of SELDI-TOF-MS proteomic profiles. *Appl Bioinformatics*4, 227-246

61. Liu, J., Zheng, S., Yu, J. K., Zhang, J. M., and Chen, Z. (2005) Serum protein fingerprinting coupled with artificial neural network distinguishes glioma from healthy population or brain benign tumor. *J Zhejiang Univ Sci B*6, 4-10

62. Mitchell, B. L., Yasui, Y., Lampe, J. W., Gafken, P. R., and Lampe, P. D. (2005) Evaluation of matrix-assisted laser desorption/ionization-time of flight mass spectrometry proteomic profiling: Identification of alpha 2-HS glycoprotein B-chain as a biomarker of diet. *Proteomics*5, 2238-2246

63. Zucht, H. D., Lamerz, J., Khamenia, V., Schiller, C., Appel, A., Tammen, H., Crameri, R., and Selle, H. (2005) Datamining methodology for LC-MALDI-MS based peptide profiling. *Comb Chem High Throughput Screen*8, 717-723

64. Jimenez, C. R., Spijker, S., de Schipper, S., Lodder, J. C., Janse, C. K., Geraerts, W. P., van Minnen, J., Syed, N. I., Burlingame, A. L., Smit, A. B., and Li, K. (2006) Peptidomics of a single identified neuron reveals diversity of multiple neuropeptides with convergent actions on cellular excitability. *J Neurosci*26, 518-529

65. Fu, Q., and Li, L. (2006) Investigation of several unique tandem mass spectrometric fragmentation patterns of NFDEIDR, an orcokinin analog, and its N-terminal dimethylated form. *Rapid Commun Mass Spectrom*20, 553-562

66. Cruz-Bermudez, N. D., Fu, Q., Kutz-Naber, K. K., Christie, A. E., Li, L., and Marder, E. (2006) Mass spectrometric characterization and physiological actions of GAHKNYLRFamide, a novel FMRFamide-like peptide from crabs of the genus cancer. *J Neurochem*97, 784-799

67. Hedner, T., Sun, X., Junggren, I. L., Pettersson, A., and Edvinsson, L. (1992) Peptides as targets for antihypertensive drug development. *J Hypertens Suppl*10, S121-32

68. Pavelic, S. K., and Saban, N. (2007) Evolving '-omics' technologies in the drug development process. *Expert Opinion on Drug Discovery*2, 431-436

69. Adriano M. C. Pimenta,M.E.D.L. (2005) Small peptides, big world: Biotechnological potential in neglected bioactive peptides from arthropod venoms. *Journal of Peptide Science*11, 670-676

70. Adermann, K., John, H., Standker, L., and Forssmann, W. -. (2004) Exploiting natural peptide diversity: Novel research tools and drug leads. *Current Opinion in Biotechnology*15, 599-606

71. Chung, F. M., Shieh, T. Y., Yang, Y. H., Chang, D. M., Shin, S. J., Tsai, J. C., Chen, T. H., Tai, T. Y., and Lee, Y. J. (2007) The role of angiotensin-converting enzyme gene insertion/deletion polymorphism for blood pressure regulation in areca nut chewers. *Translational Research : The Journal of Laboratory and Clinical Medicine*150, 58-65

72. Ohta, Y., Tsuruya, K., Fujii, K., Tokumoto, M., Kanai, H., Matsumura, K., Tsuchihashi, T., Hirakata, H., and Iida, M. (2007) Improvement of blood pressure control in hypertensive patients with renal diseases. *Hypertension Research : Official Journal of the Japanese Society of Hypertension*30, 295-300

73. Update on the efficacy of angiotensin receptor blockers in treatment of hypertension.(2005) *The American Journal of Managed Care*11, S386-91

74. Budde, P., Zucht, H., and Schulz-Knappe, P. (2004) Identification of biomarkers using peptidomics technology. *BIOspektrum*10, 577-579

75. Budde, P., Schulte, I., Appel, A., Neitz, S., Kellmann, M., Tammen, H., Hess, R., and Rose, H. (2005) Peptidomics biomarker discovery in mouse models of obesity and type 2 diabetes. *Combinatorial Chemistry and High Throughput Screening*8, 775-781

76. Juergens, M., Menzel, C., Schulz-Knappe, P., and Zucht, H. (2005) Peptide biomarker in body fluids. *BIOspektrum*11, 778-779

77. Schulte, I., Tammen, H., Selle, H., and Schulz-Knappe, P. (2005) Peptides in body fluids and tissues as markers of disease. *Expert Review of Molecular Diagnostics*5, 145-157

78. Clynen, E., Huybrechts, J., Verleyen, P., De Loof, A., and Schoofs,L. (2006) Annotation of novel neuropeptide precursors in the migratory locust based on transcript screening of a public EST database and mass spectrometry. BMC Genomics, 7, 201.

79. Yew, J.Y., Kutz, K.K., Dikler, S., Messinger, D., Li, L., and Stretton, A.O. (2005) Mass spectrometric map of neuropeptide expression in *Ascaris suum*. *J.Comp.Neurol.*, 488, 4, 396-413.

80. Preechaphol, R., Leelatanawit, R., Sittikankeaw, K., Klinbunga, S., Khamnamtong, B., Puanglarp, N., and Menasveta, P. (2007) Expressed sequence tag analysis for identification and characterization of sex-related genes in the giant tiger shrimp *Penaeus monodon*. *J.Biochem.Mol.Biol.*, 40, 4, 501-510.

81. Dreyer, C., Hoffmann, M., Lanz, C., Willing, E.M., Riester, M., Warthmann, N., Sprecher, A., Tripathi, N., Henz, S.R., and Weigel, D. (2007) ESTs and EST-linked polymorphisms for genetic mapping and phylogenetic reconstruction in the guppy, *Poecilia reticulate*. *BMC Genomics*, 8, 1, 269.

Part II

Method Development

Chapter Two

Marine Crustacean Care and Dissection

There are no particular coauthors for this chapter as the work has been performed completely by me, but I should acknowledge some of those that have trained me in crustacean dissection. First and foremost, my professor Lingjun Li initially taught me dissection and has encouraged my increase in experience by sending me twice to the University of Washington-Seattle and Friday Harbor Labs to learn more. Also, Andrew Christie of the University of Washington-Seattle Department of Biology taught me a great deal while I was in Washington. Other than these two, I would also like to recognize Xin Wei and Heidi Behrens, coworkers of mine who took a Saturday morning to assist me in taking the pictures necessary for this manual.

2.1 Introduction

Ever since mankind has been curious about the workings of the human body, they have used other organisms as model systems for their research. The factors that are taken into account when selecting a model system include such things as ease of acquisition, cost, how well it reflects the original system, and many times simplicity (a simple system is easier to understand than an incredibly complex one). It is for this reason that crustaceans have been used for many years as model organisms for nervous system research. Our lab has chosen to work with marine decapod crustacean species to learn more about this model nervous system. While it is possible to obtain a great deal of sample tissue from collaborators, there are many complications that these collaborations can introduce. Furthermore, there are many times when it is necessary to obtain fresh tissue as well as experiments that require the presence of live animals (i.e. microdialysis). It is for these reasons that our lab must not only maintain decapod crustaceans in our lab, but also dissect them and collect the tissue needed.

As can be surmised, obtaining and keeping marine invertebrates in a lab located thousands of miles from any ocean is no easy task. To do this, we were required to order crabs from various vendors and or get them from collaborators on the coasts. The majority of the crabs we worked with were *Cancer borealis* (the Jonah crab) which is found in the North Atlantic Ocean. For our work, we purchased crabs from Marine Biological Laboratories in Woods Hole, MA (www.mbl.edu) and The Fresh Lobster Company

(www.thefreshlobstercompany.com) in Gloucester, MA. Both of these companies worked well for us as they often would catch the crabs and pack and ship them the same day for next day delivery. Even with the relative speed of this process, the animals received a great deal of stress and almost always several would be dead upon arrival or die shortly after being received.

2.2 Tank Maintenance

To improve the survival of the crabs, it is necessary have a chilled seawater tank in which to place the crabs in as soon as they arrive. By doing so, the crabs are placed back into a system as similar to their natural habitat as can be provided for them in Wisconsin. To fit these unique needs, we set up a chilled, seawater tank that is fairly easy to maintain. This tank system was ordered from Aquatic Ecosystems Inc. (www.aquaticecosystems.com) and was specially designed for our needs. While this was a complete filter system, it lacked the necessary plumbing requiring me to complete the system setup. **Figure 2.1** shows the set up and the direction of water flow. In addition the physical setup of the aquariums, it is also necessary to introduce bacteria to the system in order to get the nitrogen cycle started for the processing of harmful ammonia to less harmful nitrogen forms. This is done by initially adding saltwater specific bacteria as well as a small amount of nitrogen and allowing the system to run for two weeks prior to the addition of crabs. After doing this, the tank is ready for the addition of crabs.

Although this aquarium system is fairly maintenance free, we have found it necessary to perform a 10-20% water change prior to the arrival of a shipment of crabs. This is usually done the day immediately before and consists of the following steps:

1. Prepare for the water change by first turning on the RO water by turning the two red valves in the box outside the animal room above and to the left of the door. Also prepare the buckets that are to be used. Generally, it is enough to have two five gallon buckets ready to mix the saltwater in before pouring it into the tanks.
2. Next, refer to **Figure 2.1** and quickly close valves A and B and turn off system by switching the on/off switch on the power strip to off. After this, everything should be stopped and no water should be running anywhere. Make sure that the valves are turned off all the way as they will leak if not shut completely.
3. After everything is shut off, you can now open the C valves to release water to the drain. Make sure to turn both of them on in quick succession so that roughly the same amount of water is flowing out of each of them. Watch the water level and only let out as much as is necessary. One way to determine how much water to let out is to measure, in inches, the water level in one tank when full, then take 10% of that measurement and draw a mark down from the full mark that amount. When letting out water, let it out up to that mark and then close both C valves again in rapid succession.

4. Once water has drained out of the tanks, you can begin making fresh saltwater. To do this, place 2.5 cups of Red Sea Salt into one 5 gallon bucket, and fill the bucket with RO water. The RO water can be obtained from the provided hose and by turning the red valve connected to the water line running along the upper part of the wall. While filling the bucket, make sure to stir the water to make sure the salt is getting fully dissolved. While the first bucket is filling, get the other bucket ready by pouring in the salt; then switch the hose over to the second bucket once the first one is full. For a 10% water change you will only need to fill two buckets, and four buckets for a 20% water change.
5. Once the saltwater is prepared, carefully lift it up and pour an equal number of buckets into each tank, i.e. if performing a 10% water change with two buckets, you should pour one bucket in each tank.
6. Once you have poured all the fresh saltwater into the tanks you are ready to start the system again. To do this, you will reverse the steps you took earlier. Therefore, turn on the power strip, open valves B, then valves A. The system should start flowing normally.
7. Before you finish the mesh filter that the water coming out of the tanks flows into needs to be changed. To do this, move the outlet pipes out of the way and work the dirty mesh filter out and replace with a fresh one. The dirty one can be taken to the dirty cage room or given to Animal Care personnel to be run through the cage washer for cleaning.

After replacing the mesh filter with a fresh filter, make sure to move the outlet pipes back so they are flowing into the filter again.

8. Also, you will need to clean the protein skimmer. This is done by first unhooking the waste tube from the top compartment of the skimmer and turning this compartment to the slots where it can be removed. Take this to the sink, remove the lid by turning it to the appropriate spot and lifting, and empty and clean it thoroughly with the sponge. Before replacing it, make sure to wipe out the inside of the middle compartment of the skimmer with a paper towel, thus removing any protein buildup that can be seen there. Then replace the top compartment and hook up the waste hose again.
9. Finally, check the filter bag that sits in the bottom of the tank to make sure that the beads are not leaking out of the bag. Also note the color; if the beads are somewhat brown, the beads need to be recharged. This can be done by placing the beads (in the bag) into a solution of dilute bleach, then in fresh water with a dechlorinating agent before being placed back in the bottom of the filter tank with the top sticking out the top (see the product bottle for more instructions on this).
10. Finally (do not worry, you are almost done), before you leave, make sure the water level in the filter tank is okay. The water on the left side of the tank (where the dirty water is coming in from the tanks) should be high enough to be going over the first barrier (from the left) and through the biofilter (blue sponge thing), but not higher than the second

barrier. The water should then be flowing over the third barrier into the area on the right where the pumps draw water from. The water on the right should be high enough to cover the tubes that draw water into the pumps, but not so high that it is over the middle barrier. Make sure both pumps are drawing water and that water is flowing out of valve F. Also, although you should not have changed them, valve E1 should be all the way open and valve E2 should be partially open. If you do need to adjust this, it should be open enough to pump water into the protein skimmer just up into the lower cylindrical chamber, but not more than half an inch. Fine tune this as necessary.

11. After all this, you should be done. Make sure you also shut off the RO water in the hall as well as release the pressure in the line by using the valve in the sink in the room next to the crab room. Once you have done this, you are done with the water change.

While this process may seem fairly technical, after performing this process multiple times, it is easier to perform and takes only a half hour. Other than water changes, the tank requires very little maintenance. Once a year is the chiller needs to have water run through it in reverse to wash out any buildup that has gathered throughout the year. This can be done by moving some of the piping around and pumping water from one bucket to another or cycling water through the chiller. Other than these minor things, the aquarium system requires very little maintenance.

2.3 Dissection

2.3.1 Dissection Introduction

Once the tanks are running, you are ready to place crabs in the tank. Although this seems straightforward, there are steps that need to be taken prior to and for the proper placement of crabs in the tank to ensure their survival. First, to make sure the quality of the saltwater is adequate, it is necessary to perform tests to determine the water quality in the tanks, i.e. salinity, ammonia, nitrites, nitrates, etc. Most of these tests can be performed by using the water quality test kits provided or by consulting the animal care facility staff and salinity is tested using a hydrometer. The values for each of the variables that are safe can be found in **Table 2.1**. After confirming adequate water quality, *C. borealis* can be ordered from one of the various east coast seafood vendors, or marine labs as stated above. When the crabs arrive, retrieve the box as quickly as possible and take it to the crab room. As you remove the crabs from the box, separate the dead crabs from the live ones. Place the dead crabs into one of the clear plastic bags provided by the animal care staff for dead animal disposal. Generally (not always), a crab that is moving or has his claws and feet tucked up to its body is alive, while those that are floppy and have no muscle control are dead. Occasionally, a dead or almost dead crab will twitch and move slightly despite being dead. Place live crabs in the tank by holding on to its back completely under water with its bottom (ventral side) against the wall of the tank and its mouthparts up towards you. This should allow air bubbles to escape from the crabs gills through its mouthparts. You need to hold the crab in this position

until all the bubbles have stopped. Then, carefully lower the crab to the bottom of the tank and make sure that it is sitting upright. Continue this process until all live crabs are in the tanks. Afterwards, do not forget to place the dead crabs in the appropriate freezer provided by the animal care facility.

Once you have live crabs in the tanks, you are ready to begin dissecting and obtaining the tissue you need for your experimentation. The process of dissection involves two major procedures, a macrodissection and a microdissection. The macrodissection is the rough dissection without a microscope to remove the areas of the crab that contain the ganglia and neurosecretory organs. Following macrodissection, microdissection under a microscope is performed to remove the nervous system organs from their surrounding tissue. The sequential steps for performing a full brachyuran crab dissection are listed below with pictures outlining the method.

To begin dissection, you must first decide what organs/tissues you want to obtain. This must be established initially as it will determine how much how little of this manual you will need to follow. To make this easier to follow, the instructions have been split into sections in order to make it easier for the reader to find the section needed to perform the dissection they are interested in.

2.3.2 Macrodissection

Drawing Hemolymph

1. To perform the most successful hemolymph collection, crabs must not be previously cooled on ice and should be done first, before dissection.

The hemolymph must also be drawn first so that none is lost when the legs are removed later for dissection. If only hemolymph is needed, not icing the crab greatly improves the odds of survival of the crab after collection. The first step for drawing hemolymph is to prepare an aluminum dissecting pan, a syringe that will hold at least 5 mL (even if you only want 1 mL), a large needle (20 gauge or lower, for the needle), a spatula, and a tube containing your extraction buffer of choice **(Figure 2.2)**. Get the syringe ready by placing the needle on it in preparation of the next step. Make sure to get everything ready so you are completely prepared during the actual collection.

2. Next, lean the crab's dorsal (top) side against the back of the pan (so you are looking at the tail) and then hold the claws of the crab back so they do not pinch you (**Figure 2.3**). Most people do not have large enough hands to hold both claws with one hand and may require someone to assist them in doing this. Initially, crabs will generally cooperate until you carry out the next step when the crab will begin struggling desperately.
3. After you have the crabs claws controlled, use the spatula to pry the tail of the crab away from his body. As soon as you pry the tail up, the crab will become very active and aggressive. Ask the person with you to hold the claws while you hold the tail back. If you happen to release the tail, it will fold against the body, but not so tight as it was originally.

4. Once the tail has been pulled away from the body, you are ready to insert the needle to draw hemolymph. Insert the needle into the crab right at the base of the tail, where it meets the body (**Figure 2.5**). Angle the needle into the crab towards the ventral side, into the cardial chamber of the crab.
5. With the needle in the crab, pull back on the syringe plunger to begin drawing hemolymph. Pull back on the plunger enough that you create a vacuum inside the syringe to assist in the drawing of the hemolymph(**Figure 2.6**). You may need to move the syringe in and out of the crab to find the spot that provides the most hemolymph. This will be evident when hemolymph begins rapidly filling up the syringe. Try not to move the needle around too much inside the crab as you will be damaging the organs inside the crab and will reduce the survivability of the crab. If you are not able to draw enough hemolymph, you can also insert the needle at the base of one of the legs and try to draw hemolymph there (**Figure 2.7**). Generally, drawing hemolymph from the base of the legs is not as successful, but is an alternative.

Once you have drawn as much hemolymph as you need, remove the needle. For crab survival, the smaller the amount of hemolymph that is drawn, the better the chances are. Generally, 2 mls can be drawn without a serious detrimental effect on the animal, and as much

as 10-15 mls can be drawn if the crab is to be sacrificed following hemolymph removal.

6. Once you have removed the needle, let the crab rest and/or place it back into the tank, making sure to allow all the air to burp out of the crab before letting it settle to the bottom.
7. Next, take the syringe and place the appropriate amount of drawn hemolymph into the tube you have prepared. Place this tube on ice and continue with dissection.

Stomatogastric Nervous System (STNS) and Eyestalk Removal

Macrodissection for the removal of the following organs:

Brain

Stomatogastric Ganglia (STG)

Commissural Ganglia (CoG)

Oesophagial Ganglia (OG)

Anterior Cardial Plexus (ACP)

Sinus Gland (SG)

Items needed:

Aluminum dissecting pan

Side cutters or rongeurs

Small scissors

Toothed forceps or locking hemostat

Spatula

Rubber lined dissecting dish

1. Before performing this dissection, you need to place the crab in ice for 15-30 minutes. This anesthetizes the crab as well as makes it easier to deal with initially. Make sure to bury crabs in ice, making sure it is completely covered. If you need to draw hemolyph, follow the directions above, and then place the crab on ice. While waiting for the crab to chill, prepare all the tools you will need for this dissection (**Figure 2.8**). Also, become familiar with the location of each organ in relation to the crab so you better understand what you are dissecting (**Figure 2.9**).
2. Once the crab has been cooled, place it in the dissection pan. Then, use the side cutters to remove the claws and legs at their base where they meet the body of the crab (**Figure 2.10**). Generally, it is best to remove the claws first so the crab can not injure you as you progress.
3. Next, with the ventral side of the crab facing up, use the side cutters to crack away the outer edge of the crab's shell. Crack into the crab roughly 0.5 inch or at least enough to provide space to slip in a spatula (**Figure 2.11 A**). Crack around the crab's shell from the rear edge of the crab, up around the crab almost to the eyestalks (**Figure 2.11 B**).
4. After you have the edge of the shell cracked away, slip the more rounded end of the spatula into the shell and separate the hypodermis of the crab from the crab shell (**Figure 2.12**). The hypodermis is the thin skin that is present just under the shell of the crab. Be careful

when doing this as you do not want to break the skin (although breaking the skin does not mean you can not continue, it does make the dissection more difficult later).

5. Separate an area of the hypodermis from the shell large enough that you can remove a sufficient amount of the shell that you can continue using the spatula to separate the hypodermis all the way across the crab. Be especially careful when separating the hypodermis under the middle section of the crab as this is where the stomatogastric nervous system lies and you can damage it if not careful. You will essentially be alternating between cracking the shell away and separating the hypodermis. Eventually, you want to have most of the front half of the crab's shell removed, except for a small bridge (maybe 1 inch wide) from the eyestalk area to the rear of the crab. The hypodermis should be separated completely from the front portion of the crab (**Figure 2.13**).
6. Once you have the top of the crab prepared, with the appropriate amount of shell removed and the hypodermis separated, turn the crab over and begin removing the mouthparts of the crab. These can be removed by pulling them away from the body of the crab, and twisting until they crack off (**Figure 2.14 A**). There are multiple mouthparts, some of which you will need to pull away from the crab's body with a spatula (**Figure 2.14 B**). Finally, you will get down to the main pair of mandibles over lips and the entrance to the oesophagus. These can

be removed by using the side cutters to cut the base of the mandibles and then levering the mandibles out with the cutters(**Figure 2.14 C**).

7. Next, return the crab back so its ventral side is up again and crack the shell along the anterior (front) edge of the crab, around the eyestalk sockets. Make sure to crack enough of the shell along the very front of the crab (**Figure 2.15 A**). Once you have the front of the shell removed, use the spatula under the shell to make sure the hypodermis is separated as far forward as possible without puncturing it. After this is done, use the cutters to crack the shell all the way across the middle of the crab (where you left shell earlier). Once the shell is cracked all the way across, carefully lift from the rear the remaining middle/front portion of the shell (**Figure 2.15 B**). If you have cracked enough shell away along the front, this piece should lift away fairly easily with no hypodermis or other tissue still attached to it. If hypodermis or other tissue is still stuck to this portion of the shell, either crack more of the shell on the front of the crab, or carefully use the small scissors to cut away the connected tissue as close to the shell as is possible.
8. Once the front portion of the crab has been removed, you should now be able to remove the eyestalks (for SG). They may either be on the piece of shell you just removed, or remain on the crab. Wherever they are, use the small pair of scissors to cut the softer section at the base of the eyestalk to remove it from the rest of the shell (**Figure 2.16**).

Place the eyestalks into saline solution and place on ice until carrying out microdissection (in next section).

9. After removing the eyestalks, use the spatula to separate the hypodermis from the lower shell on either side of the eyestalks (**Figure 2.17 A**). Then, use the small scissors to cut the tissue along the front of the crab. If the shell was cracked correctly, there should be two small muscles that remain attached to the front portion of the crab's lower shell. Cut these, and then continue to cut down and forward along the inside edge of the crab's lower shell. Make sure to keep your scissors as close to the shell as possible, making small cuts as you progress (**Figure 2.17 B**). Use the scissors or the spatula to occasionally scrape the cut tissue down out of the way. Continue doing this to the lower part of the shell where you will eventually reach an arch structure in the shell with a small bump in the shell protruding into the crab's body (**Figure 2.17 C**).
10. Once you have reached the arch in the shell over the oesophagus, use the small scissors to carefully cut around the oesophagus and lips of the crab (**Figure 2.18 A**). Once the lips and oesophagus are free, use the cutters to cut the shell on either side of where the eyestalks were down towards the arch in the shell over the oesophagus (**Figure 2.18 B**). This should release the whole piece of shell over the oesophagus.
11. Now that you have the oesophagus revealed, you will be removing the stomach. This is probably one of the more difficult parts of the

dissection and generally takes practice to become proficient at. First, turn the crab so that you are facing it head on and prop the front of the crab on the front of the dissecting tray so the crab is leaning up toward you. Now, take the toothed forceps or the locking hemostats in your left hand and grab the lips. Gently lift the lips and using the small scissors, cut at the base of the oesophagus (**Figure 2.19 A**). Continue carefully lifting the lips as the oesophagus is cut away, making sure not to tear the lips but to use them to control the oesophagus. As you lift, continue cutting the tissue under the oesophagus, towards the posterior (rear) end of the crab. You should find two small ossicles (bones) just below the oesophagus that you can use to guide your cutting. Keep cutting above these ossicles and below the oesophagus and stomach. Also, avoid cutting into the adjacent sperm ducts as puncturing these will release sperm which will interfere with the dissection. Continue lifting the oesophagus with the forceps and cutting back into the crab until you see the pyloris (two whitish yellow balls) on the underside of the stomach (**Figure 2.19 B**). Once you have reached this, you can cut the hypodermis on either side of the stomach angling back and away from the stomach (**Figure 2.19 C**). Extend this cut towards the back of the stomach, and then cut behind the stomach and lift the whole stomach away from the crab, cutting any other tissue holding the stomach in place (**Figure 2.19 D**). Keep

holding the stomach by the lips with the forceps and go on to the next step.

12. The next step is the removal of the teeth from the stomach and can also be challenging to perform the first several times attempted. To do this, take the stomach that you have just removed and flip it over, turning the forceps in the process. Then, drape the stomach over the index finger of your left hand and then grab and hold the forceps with your left hand as well. Once you have the stomach positioned, use the small scissors to cut straight down the bottom of the oesophagus (that should now be up because the stomach is flipped over) (**Figure 2.20 A**). Cut straight back, through the pyloris (between the two whitish yellow balls) and down on the back side of the stomach so that there is a cut running from front to back on the bottom of the stomach. Then take the scissors and at the widening of the oesophagus into the stomach, make cuts angling back towards the posterior side corners of the stomach (**Figure 2.20 B**). Once these cuts are made, the stomach should almost fall open, revealing the inside of the stomach. Inside the stomach, you may find the remains of the crab's last meal. Wash this out with saline if necessary. You will also find three brown, harder parts in the stomach. These are the teeth the crab uses to grind its food. To remove the teeth, cut them out with the small scissors (**Figure 2.20 C**).

13. Once you are done removing the teeth, flip the stomach back over and place it into a rubberized dissecting dish with enough saline to cover it completely (**Figure 2.21**). Set this aside on ice or in a fridge until you are ready to perform the microdissection to remove the stomatogastric nervous system organs.

After removing the eyestalks and the stomach, you are ready to go on to either perform the remaining macrodissection or the microdissection of the sections you have just dissected.

Heart, Cardial Ridge and Thoracic Ganglia Removal

Macrodissection for removal of the following organs:

Heart

Pericardial Organ (PO)

Thoracic Ganglia (TG)

Items needed:

All items for macrodissection of stomatogastric ganglia

Large scissors (curved or straight)

Forceps, fine and paddle (NOT microdissecting forceps)

Tube prepared for heart and another for TG (if necessary)

1. To perform this dissection, you must have performed dissection of the stomach before beginning. Make sure you have all the supplies prepared before beginning (**Figure 2.22**).

2. First, pick up the crab in your left hand and take the large scissors in your right hand. Using the scissors cut the shell straight back from 1 inch on either side of the midline all the way back to where the top shell meets the bottom shell along the back of the crab (**Figure 2.23 A**). When you cut all the way through on one side, you will be able to break off that side of the shell. Once the shell is removed, use the scissors to cut away the tissue that remains there along with the gills along the cardial ridges. Make sure to remove the tissue on both sides of the crab so that all you have left are the tail section of the crab, the cardial ridges, and the shell over the top of that (**Figure 2.23 B**).
3. Once you have both sides of the crab removed, turn the scissors so the blades open parallel to the shell, open them and slide them under the upper part of the shell with the blades on either side of the top of the cardial ridge (**Figure 2.24 A and B**). Cut the connection between both cardial ridges and the upper portion of the remaining shell. Once these connections are cut, lift the shell away to reveal the heart and the inside of the cardial ridges.
4. To remove the heart for tissue collection, use a clean pair of forceps to carefully tug on the heart until it pulls out of the crab. If necessary, use the small scissors to cut any connections to the surrounding tissue (**Figure 2.25**). Be careful when removing the heart not to damage the inside of the cardial ridges as this is where the PO's reside. Blot the

heart on a Kimwipe to remove any excess hemolymph and place the heart into the tube you have prepared.

5. Once you have removed the heart, lift the tail and pull it away from the remaining portion of the crab (**Figure 2.26 A**). Then use a pair of forceps (I use the paddle ended forceps) to remove all the tissue from between the cardial ridges (**Figure 2.26 B**). This includes removing all the yellow tissue (hepatopancreas) that you can, as well as removing the sperm ducts which run between the cardial ridges under the heart. Again, make sure not to damage the inside of the cardial ridges where the PO's are.
6. After the inside of the cardial ridges is fairly clean, use the large scissors to cut the cardial ridges away from the remaining crab at their bases (**Figure 2.27**). Make sure to get the entire cardial ridge, and place it in saline solution as soon as it is removed from the body of the crab.
7. Once the cardial ridges have been placed in saline, you can go on to work on removing the TG. To begin this process, use the small scissors to cut away the ossicles around the center of the remaining piece of the crab (**Figure 2.28 A**). Be careful as the TG lies directly in the center of the crab. Once you have cut most of the ossicles, use the forceps to remove as much of the ossicles as you can (**Figure 2.28 B**).

8. After removing the ossicles, you should be able to see the TG in the bottom of the crab piece (**Figure 2.29 A**). It looks like a wagon wheel, with ganglia projecting out to the legs and the rest of the crab's body. To remove it, use the small scissors to cut around the TG (**Figure 2.29 B**), and using the fine (and clean) forceps, carefully lift the TG. As you pull it up, use the scissors to cut any remaining connections it has to the surrounding tissue (**Figure 2.29 C**).
9. Once you have the TG removed from the crab, use a Kimwipe to blot the TG and to remove any excess tissue from it. After the TG is clean, place it into the tube you have prepared for it.
10. After removing the TG from the crab, you are finished with the macrodissection of the crab. Now, you should have the eyestalks in saline, the stomach section in saline in a dissecting dish, and the PO in saline, all of these chilled on ice or in a fridge (**Figure 2.30**). Additionally, you should have the hemolymph, heart tissue, and TG, all in separate tubes with the appropriate extraction buffers in them. These tubes should also be placed on ice so as to keep them cool while you perform microdissection to obtain all the other tissues. The final step of the macrodissection process is to clean all the tools you have used in the process. All the crab parts need to be placed in a dead animal bag and placed in the dead animal freezer provided by the animal care facility. All the tools should be rinsed and all of the "gunk" scrubbed off of them. After rinsing in water, rinse them in a

solution of 70% EtOH. Once this is all done, you are all finished with the macrodissection and can continue on to the microdissection.

2.3.3 Microdissection

Sinus gland (SG) dissection from eyestalks

Microdissection for removal of the following organs:

Sinus glands (SGs)

Items needed:

Dissecting microscope

Rubberized dissecting dish

Large spring scissors

Two pairs of the least and most sturdy, fine micro- forceps

Fine micro- forceps

Small spring scissors

Tube ready for SGs

1. Prepare by pouring saline into the dissecting dish and getting all your tools ready and at hand (**Figure 2.31**) near the microscope you will be using.
2. Next, take the eyestalks out of the saline. Using the large spring scissors, slide one blade along the inside edge of the eyestalk and cut first the concave side, and then turn the eyestalk over and cut the convex side (**Figure 2.32**). Do this for both eyestalks and then place them in the dissecting dish.

3. While looking through the microscope, use the two pairs of the most sturdy, fine forceps and carefully pull the two halves of the eyestalk apart (**Figure 2.33 A**). Never use the very fine forceps when there is hard shell present. While prying the two halves apart, watch for a bright white, almost iridescent, spot that seems to be embedded in the tissue on the concave side of the eyestalk; this is the SG (**Figure 2.33 B**). Once you have identified the SG, use one of the forceps to pull the whole piece of tissue out of the eyestalk shell. Do this for both eyestalks. Once you have removed the tissue from the shell, remove the shell from the dissecting dish and place it aside to throw away later.
4. Once you have just the soft tissue, you can grab the fine forceps and the small spring scissors. For microdissection, you should always hold the forceps in your left hand and the scissors in your right hand (vice versa for left handed people) as this allows you the best control with both tools. Once you have the tools, use them to carefully cut away the tissue surrounding the SG (**Figure 2.34 A**). It is almost impossible to get rid of all the surrounding tissue, but you should be able to clean the organ up fairly well. One trick is to grab the tissue with the forceps, then place the open scissors over the very tip of the forceps (above the tissue) and carefully pull the tissue through the blades of the scissors up to the point you want to cut (**Figure 2.34 B**). By doing this multiple

times, it is generally possible to clean most of the extraneous tissue from the SGs.

5. After removing the SGs from the eyestalks, grab them with the forceps and remove them from the saline. Before placing them in the tube, make sure to blot the forceps with a Kimwipe to remove any excess saline solution that is on them. To prevent the loss of SGs by sticking to the Kimwipe, blot just the upper part of the forceps, above where you are holding the SGs. Finally, place the SGs in the tube you have prepared with the appropriate buffer in it.

Pericardial organ (PO) dissection from cardial ridges

Microdissection for removal of the following organs:

Pericardial organs (POs)

Items needed:

Dissecting microscope

Rubberized dissecting dish with two pins

Small scissors

Fine micro- forceps

Small spring scissors

Tube ready for POs

1. To begin this dissection, you again want to fill a dissecting dish with saline solution and get all your tools and materials ready (**Figure 2.35**).

2. Once you are ready to start, you can remove the cardial ridges from the saline you have them stored in from the macrodissection. Take each cardial ridge and identify which side was inside (towards the heart) and which side was out. If you imagine the cardial ridge as having a triangular shape, the inner side makes up one side of the triangle and the outer side makes up one of the angles of the triangle. Because the POs are removed from the inside of the cardial ridge, you normally would have to balance the cardial ridge on the point of the outer side, making the dissection fairly difficult. To simplify this, use the small scissors to cut the point of the outer side off to create a flat side on which to lay the cardial ridge in the dissecting dish (**Figure 2.36**).
3. Once you have cut the outer side of the cardial ridge, place it in the dissecting dish with the inner side up and use the pins to stabilize it from moving. This is best done by placing one pin in the ventral (bottom) side of the cardial ridge, and placing another on the dorsal (top) side, making sure to not to place the top pin through the PO itself by keeping the pin as close to the dorsal ossicle as possible (**Figure 2.37**). These pins should be placed at an angle so that you can work on the cardial ridge between the pins.
4. Next, use the microscope to identify the PO. If there is extraneous tissue remaining from macrodissection covering the PO, carefully use the forceps and the small spring scissors to remove it. To identify the

POs, use the microscope and look for bright white or iridescent strings that run across the main portion of the cardial ridge (**Figure 2.38**). When looking down on the inside of the cardial ridge, you should see a set of ossicles that create a triangular structure, with another ossicle projecting off one of the triangles vertices (this side is the anterior or side). Over the ossicles that project out of the triangle, there are two commissures that run along this ossicle and connect with the main portion of the PO inside the triangular ossicle structure. In the triangular structure, the PO has junctions on the inside of two of the vertices of the triangle. Between these junctions, there are two to three commissures that connect them. These can be difficult to see sometimes so it is often easier to identify the junctions first.

5. Once the POs (junctions or commissures) are identified, use the forceps to gently lift them away from the cardial ridge. It is generally easier to begin by lifting one of the junctions and then using the spring scissor to cut any of the smaller connections to the cardial ridge (**Figure 2.39 A**). If possible, do not cut the large commissures that cross through the triangular ossicle. Once a junction is free, carefully lift it and as you do, the commissures will begin pulling away as well. Because the commissures are generally under a layer of tissue, it is necessary to cut them out. This can be done by either cutting alongside the commissures, or carefully making small cuts at the spot where the commissure is held in the tissue (**Figure 2.39 B**). As there

are two to three commissures connecting the two main junctions, complete removal requires this process to be carried out multiple times. Overall, this is the most tedious part of the PO dissection as well as when there is the most risk for cutting the PO. If cut, continue from the other junction and try to obtain as much of the PO as possible. You have the entire PO when you have two junctions, all the commissures connecting them, and a bit of PO that projects from the anterior junction along the ossicle projecting towards the anterior of the crab.

6. After removing the POs, pull them out of the saline solution and carefully blot the forceps with a Kimwipe to remove any remaining saline. Because the POs will easily stick to the Kimwipe, blot the forceps behind the tips where you are holding the POs. Doing this reduces the amount of saline transferred to the tubes with buffer in them. Once blotted, place the POs in the tube, making sure that the POs actually get into the buffer and do not just stick to the walls of the tube.
7. Now that you have dissected one set of POs, discard the cardial ridge you were just working on and start work on the other one. After completely finishing, make sure to clean up the dish you have been working in as well as any tools that were dirtied in the process.

Stomatogastric nervous system (STNS) dissection

Microdissection for removal of the following organs:

Brain

Commissural Ganglia (CoG)

Oesophageal Ganglia (OG)

Anterior Cardial Plexus (ACP)

Stomatogastric Ganglia (STG)

Items needed:

Dissecting microscope

Rubberized dissecting dish with at least seven pins

Fine micro- forceps

Small spring scissors

Five tubes with buffer for each type of organ

1. To begin the STNS dissection, first prepare all dishes, tools, and tubes (**Figure 2.40**), and then, place the stomach section under the microscope. You should have saline on the stomach from the macrodissection, but make sure that it is fairly clean and is covering the stomach completely. If the saline is dirty, make sure the lips are pinned down first and empty out the dirty saline and replace with clean saline. You want enough saline so that the whole stomach is completely submerged. This makes the dissection much easier as tissues can free float in the saline and will not lie down on top of each

other. Additionally, refer often to the schematic of the STNS (**Figure 2.41**) as it will assist you greatly in identifying the location of many of the organs you will be dissecting.

2. Once ready, you need to pin out the stomach to stretch everything out so that it is easily accessible for dissection. There should already be a pin in the lips holding them in place. Next, place a pin in each of the lower corners of the stomach, slightly stretching the stomach taught (**Figure 2.42 A**). After these pins are placed, put a pin on either side of the stomach about halfway between the lips and the lower pins (**Figure 2.42 B**). When placing these pins, make sure to pin the stomach tissue (possibly hidden under the hypodermis) and not just the hypodermis. Finally, look for where the oesophagus was cut open during macrodissection just below the lips, and place pins there on either side of the oesophagus, pulling up slightly to stretch these up and away from each other (**Figure 2.42 C**). Be careful when placing these last pins as it is easy to stick a pin through the CoG. If you are unsure, wait until you have located the CoGs before placing these pins.
3. After the stomach has been pinned out, you can begin the actual dissection. Initially, you want to identify the brain. There should be tissue protecting the brain that will need to be removed. You ought to see hypodermis from the dorsal (top) side of the crab (a mottled red brown tissue) and from the ventral (bottom) side of the crab (a whitish

yellow tissue). Under where these two tissues meet is roughly where the brain should be located (**Figure 2.43**). If it is not revealed already, carefully use your forceps and spring scissors to cut the hypodermis out of the way until you see the brain. Remove tissue that is interfering. The brain looks like a white iridescent blob with many commissures projecting from it. It should be fairly easy to spot as it is one of the larger structures on top of the stomach. Once you have identified the brain DO NOT remove it, but carefully clear away the tissue surrounding it.

4. **Removing the CoGs:** With the brain identified, you will now dissect out the CoGs. To do this start at the brain and identify one of the two large commissures that project out of the brain at an angle up towards the lips, but away from the midline. Follow these commissures, cutting away tissue as you go, until you reach a junction (**Figure 2.44 A**). The junction will look iridescent white and will have another large commissure projecting from it as well as several smaller commissures (**Figure 2.44 B**). This junction is the commissural ganglia (CoG). Carefully, cut the commissure between the brain and the CoG and use this to pull the CoG up and out of the surrounding tissue. Because there are other commissures connecting it to the rest of the STNS, you will have to make several more cuts to completely remove it. Once you have it removed, move it to a clear area of the dissecting dish and shorten the projecting commissures so that only the CoG remains

(**Figure 2.44 C**). When trimmed down, remove it from the saline and blot the forceps with a Kimwipe to remove the excess saline, again being careful not to lose the CoG in the process. After blotting, place the CoG into the appropriate tube and make sure it gets into the buffer. Finally, repeat this process on the other side to remove the other CoG. One thing to note is that you may have to rotate the dish in order to get a good angle to cut from when dissecting the other CoG, keeping a relative sense of anterior and posterior.

5. **Removing the brain:** To remove the brain, grasping the right commissure that projected towards the CoG and lift slowly. As you lift, you will need to cut the other commissures that are projecting out of the brain (**Figure 2.45 A**). Eventually, you will also have to cut directly underneath the brain to free it from the surrounding tissue (**Figure 2.45 B**). Be careful not to cut too deeply under the brain as other STNS organs lay there. Eventually, you should be able to completely remove the brain. Like the CoG, move the brain to a clear area of the dissecting dish and trim short all the projecting commissures. Once trimmed, remove the brain from the saline, blot it as with the other organs, and place it in the appropriate tube.
6. Once you have dissected out the brain, you are ready to collect the rest of the STNS organs. This next set of steps is probably the most difficult part of the whole dissection. It takes practice and perseverance. Eventually you will get gain the skill required to do it

quickly and efficiently.

Begin cutting straight down (towards the dish) from where the brain was (**Figure 2.46 A**). You will only need to cut down approximately half a centimeter or so. Stop cutting down when you reach an artery that runs along the midline of the crab. This artery will look like an opening in the tissue. It will be deceiving though as there are two muscles that are inside the artery (**Figure 2.46 B**). Watch for these two muscles as well as the artery wall to indicate when you have reached the artery. Once you have reached the artery, do not cut any further or you will make your dissection very difficult.

7. Look down between the two muscles to find a commissure running inside the artery along the midline, this is the stomatogastric nerve (*stn*) (**Figure 2.47 A**). Quite often the stn is difficult to see between the muscles, so carefully cut their connection to the inside wall of the artery on the posterior (rear) end (**Figure 2.47 B**). Again, look for the stn lying along the bottom of the artery. It may be necessary to use the forceps and blindly attempt to grab the *stn* from inside the bottom of the artery. Generally, if you have not seen the *stn* by now, blindly grabbing for it like this will reveal its location. If you do grab it, do so gently and do not pull hard as you can snap the *stn* and make this dissection much more difficult. Once you have identified the *stn*, grab the muscles that you just cut and lift them up. As you lift them, you may have to cut the top of the artery open to continue pulling up on the

muscles. As you reach the end of the muscles, continue pulling on them and cutting towards the lips, all the while paying attention to where the *stn* is and making sure not to cut it (**Figure 2.47 C**). After you have cut almost all the way to the lips, cut the tissue containing the muscles free and discard it. If you have done this correctly, you should have just revealed the majority of the forward section of the STNS.

8. Remember to refer to the STNS schematic (**Figure 2.41**) to assist in identifying organs. After you have removed the main portion of the artery, find the *stn* again. Next, follow the *stn* forward to a thickening of the nerve. Although you may not be able to see them, there should be two projecting nerves (anterior cardial nerves, *acn*) coming off the *stn* at this thickening. Most likely, you will be unable to see the *acn*s at this point, but take note of where this thickening is as we will be returning here shortly to remove the ACPs. As you move anterior, you will find a major intersection where two relatively large commissures project off the stn towards where the CoGs were (the superior oesophageal nerve, *son*). You may find it necessary to clear away any excess tissue that has remained in the way, just make sure when doing so that you do not cut the *stn* in the process. From this intersection, the *stn* should continue forward to a final branch point. At this final branching, there will be many projecting nerves, including two that curve forward and then back towards where the CoGs were (the inferior oesophageal nerves, *ion*), and another that projects straight up towards you (the

inferior ventricular nerve, *ivn*). Just posterior to this junction is where the oesophageal ganglia (OG) is located (**Figure 2.48**).

9. **Removing the OG:** Once you have identified the location of the OG, use the spring scissors to cut the *ion*s that project out of this junction. Then using the forceps, grab either the *ivn* or an *ion* and carefully lift the OG up away from the stomach tissue. As you lift it away, you will need to cut the *stn* below the OG but above the junction of the *stn* and the *son*s (**Figure 2.49**). Once you cut the *stn*, the OG should be free and you can move it to a clear place in the dissecting dish. Do so and as you did with the CoGs and the brain, trim down the projecting nerves before removing the OG from the saline, blotting the forceps to remove excess saline, and placing into the appropriate tube.
10. **Removing the ACPs:** After removing the OG, follow the *stn* posterior from where the OG was to the thickening in the *stn* that you identified earlier, this is where the *acn*s (that contain the ACPs) branch off. If you do not see a thickening, do not worry; you can usually also find the acns by carefully grabbing the *stn* and lifting it gently. By doing so you should see two nerves projecting out the *stn* and winding their way to either side and into the surrounding tissue. Because the ACPs location on the *acn* is vaguely known, it is necessary to take a length of the *acn* to guarantee removal of the ACP. To do this, first grab the *acn* close to the *stn*, but not right at their junction. Then slide the spring scissors alongside the *stn* and cut the *acn* off the *stn* while still holding

onto the *acn* with the forceps (**Figure 2.50**). Finally, lift the *acn* and carefully cut away tissue that is holding it in place until you have a section of *acn* roughly 0.5 centimeters long. Cut this segment out, remove it from the saline, blot it and place it in the appropriate tube. Do the same on the other side of the *acn* and you have obtained the ACPs.

11. **Removing the STG:** To remove the STG, start by following the *stn* posterior from where you found the *acn*s. When you have reached the point that you first cut into the artery, you should see the *stn* starting to thicken into the STG. Occasionally, the STG is further posterior than where you cut into the artery and you will be required to cut open the artery a little more to have uninhibited access to the STG. The crab STG is fairly distinctive as it has two nerves projecting directly out of the sides (the anterior lateral nerves, *aln*) and another that projects posterior (the dorsal ventricular nerve, *dvn*) with a nearby junction where two more nerves angle posterior and to either side (the median ventricular nerves, *mvn*). The STG is also marked by a more opaque spot at its center that is the neuropil, or meeting of all the incoming neurons (**Figure 2.51 A**). Once you have identified the STG, use the forceps to grab the *stn* just anterior of the STG and carefully pull it towards the anterior. Once there is a little tension on the STG, cut the *dvn*, the *mvn*s, the *aln*s, and finally the *stn* anterior to where you are holding it (**Figure 2.51 B and C**). After doing this, the STG should be

free and you can move it to a clear area of the dish. As with other organs, trim the projecting neurons, remove it from the saline, blot the forceps and place it into the appropriate tube.

12. Now, if you were able to do all that, you have just carried out a complete dissection of the brachyuran crab. Make sure to discard the remaining stomach tissue, clean out the dish, and clean all the tools you have used. As with the macrodissection, rinse and scrub everything with water first, and then rinse with 70% EtOH. Also, make sure that your tubes are chilled and that you place them in the freezer as soon as possible. Once you have cleaned up, you are done with dissection of the crab and can go about enjoying the rest of your day (smelling like the insides of a crab, of course).

2.3.4 Dissection Conclusion

Overall, the dissection of the brachyuran crab is a fairly complex process that takes many times to master. Throughout my graduate career, I have probably dissected well over 1,000 crabs of varying species and only through this have I become as proficient as I have in doing this. Still, though, I even have trouble from time to time and continue to learn and hone my skills. Also, when faced with the dissection of a crustacean that I do not have experience with, i.e. the American lobster (*Homarus americanus*), I can do a great deal, but generally have great difficulty completing the whole dissection as I am inexperienced in this type of dissection. Dissection takes practice, that's all there is to it. So, keep

practicing and trying new things to develop your own ways of dissection and you will soon be able to perform these dissections quickly and efficiently.

2.4 Figures

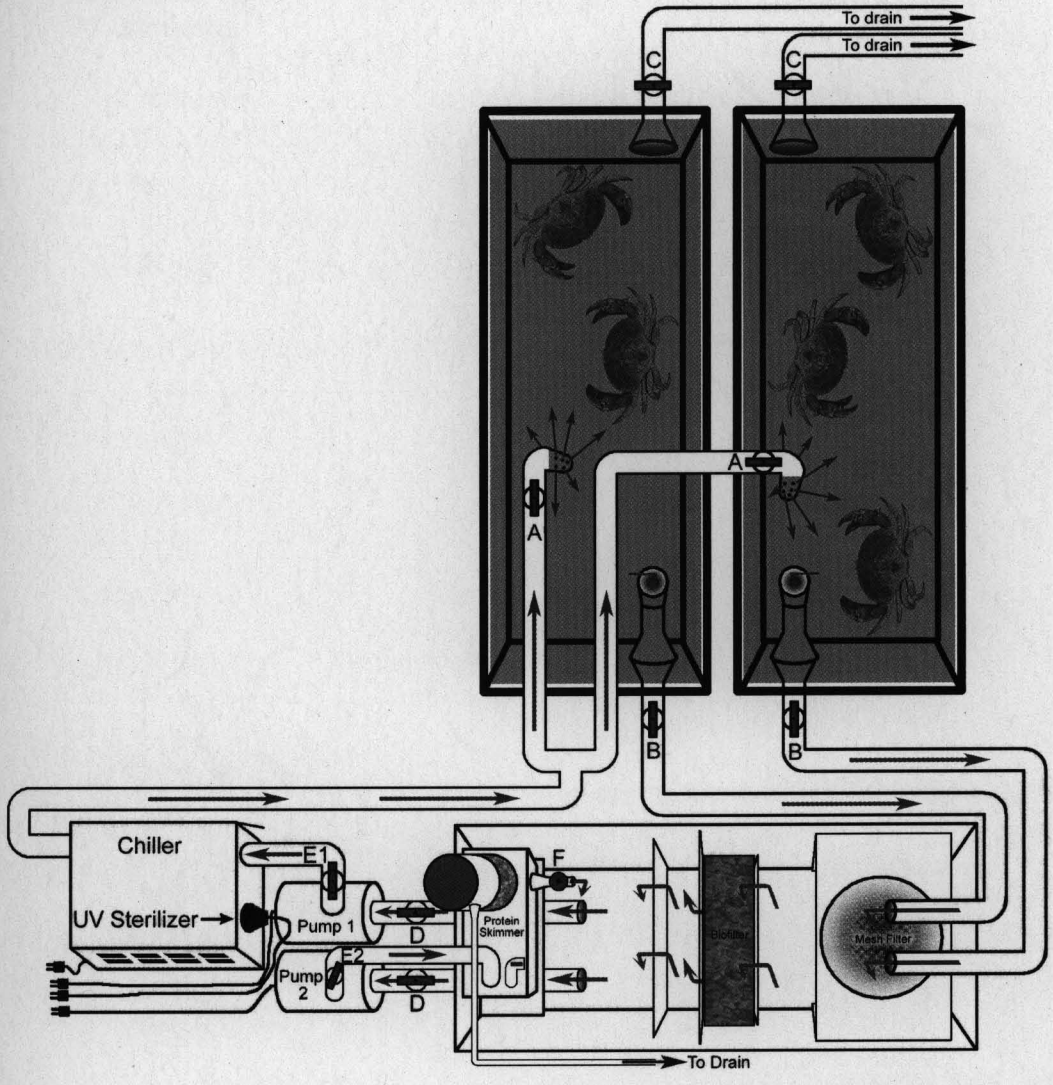

Figure 2.1: Schematic of the artificial seawater tanks in the Li Lab at UW-Madison, School of Pharmacy. The valves shown are labeled, both on the schematic and on the valves. With the valves, when the handle is in line with the piping, the valve is open (valves A, B, D, and E1), whereas when the valve handle is perpendicular to the piping, the valve is closed (valve C). Note: To work properly, the filter tank must be lower than the aquariums in order to allow water to drain from the aquariums into the filter tank.

Water Quality Parameters	Preferred Value	Desired Range
Salinity	1.025 Specific gravity	1.020-1.030 Spg
pH	8.3	8.2-8.4
Alkalinity	2.5 meq/L	1.8-3.0 meq/L
Hardness (Calcium)	400 ppm	350-450 ppm
Phosphate (PO_4)	0 ppm	0-0.1 ppm
Total Ammonia (NH_3/NH_4^+)	0 ppm	< 0.25 ppm
Nitrite (NO_2^-)	0 ppm	< 0.3 ppm
Nitrate (NO_3^-)	0-0.25 ppm	< 5 ppm

Table 2.1: Table of Water quality parameters, the values sought after, and the acceptable range for each. Note: Ppm and mg/L are roughly the same measurement.

Figure 2.2: Materials needed for drawing hemolymph.

Figure 2.3: Method of holding crab for drawing hemolymph.

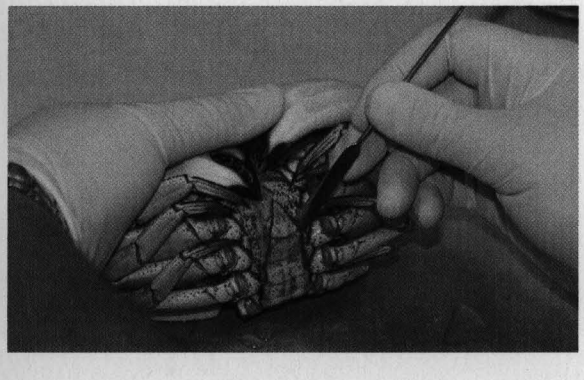

Figure 2.4: Prying the tail away from the crab's body with a spatula.

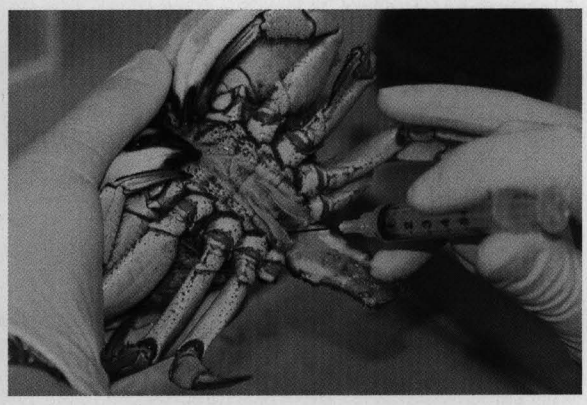

Figure 2.5: Insertion of the needle at the base of the crab's tail for collecting hemolymph.

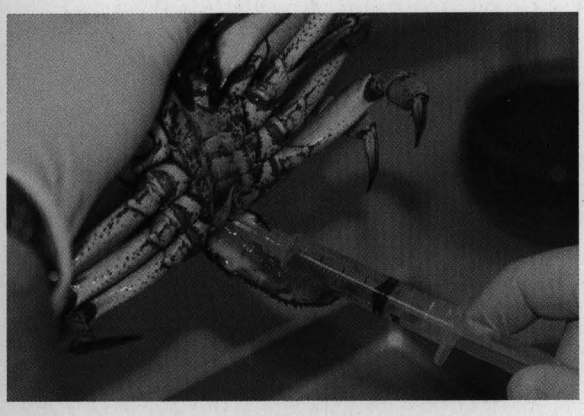

Figure 2.6: Drawing hemolymph.

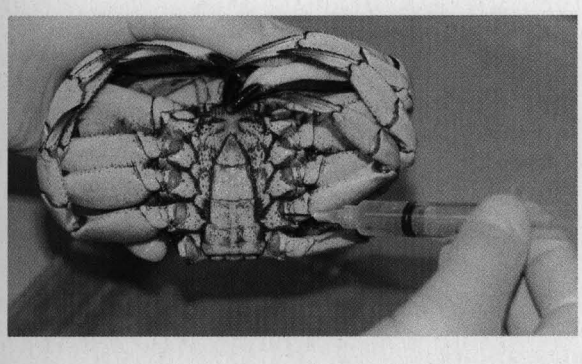

Figure 2.7: Drawing hemolymph from the base of a leg.

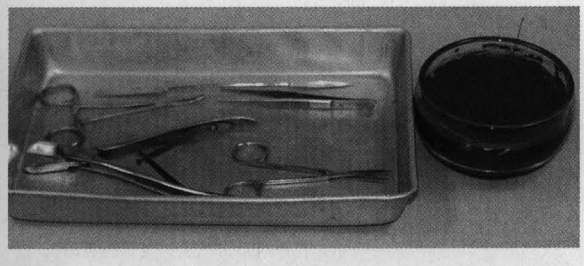

Figure 2.8: Materials needed for first portion of macrodissection.

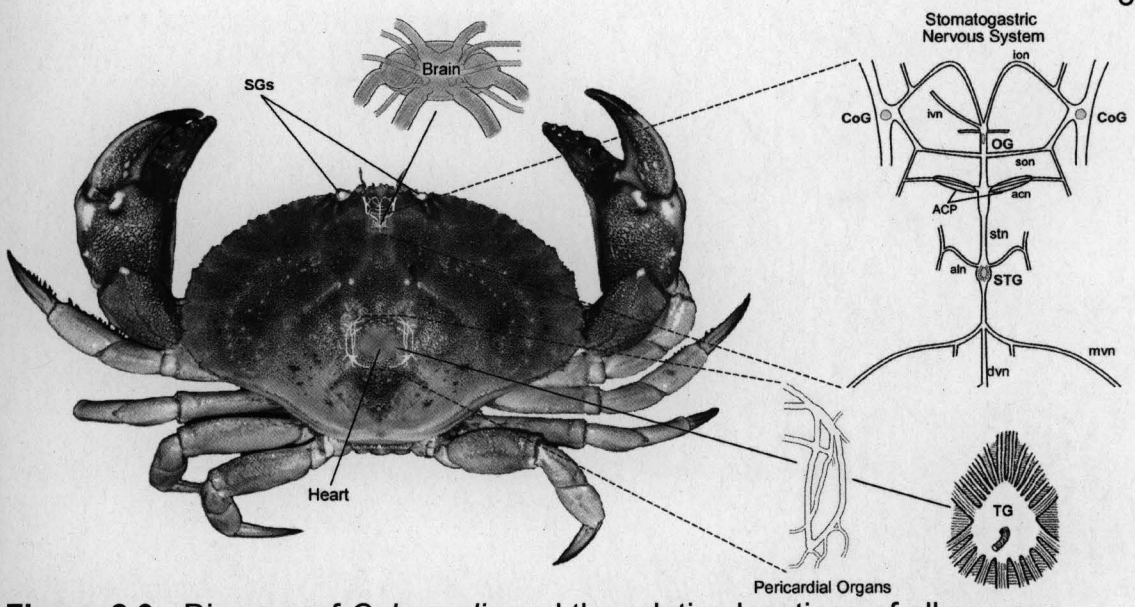

Figure 2.9: Diagram of *C. borealis* and the relative locations of all nervous system and neurohemal organs.

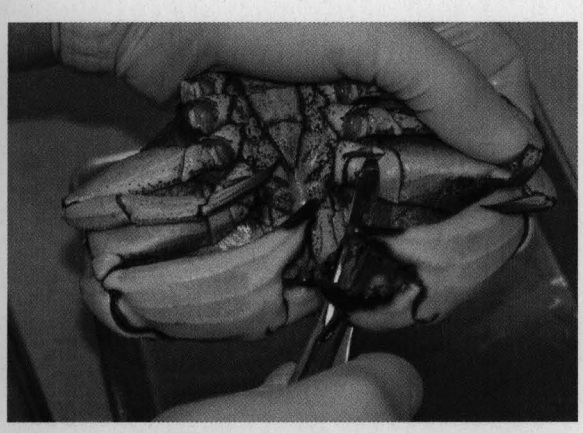

Figure 2.10: Remove claws and legs using side cutters at their base.

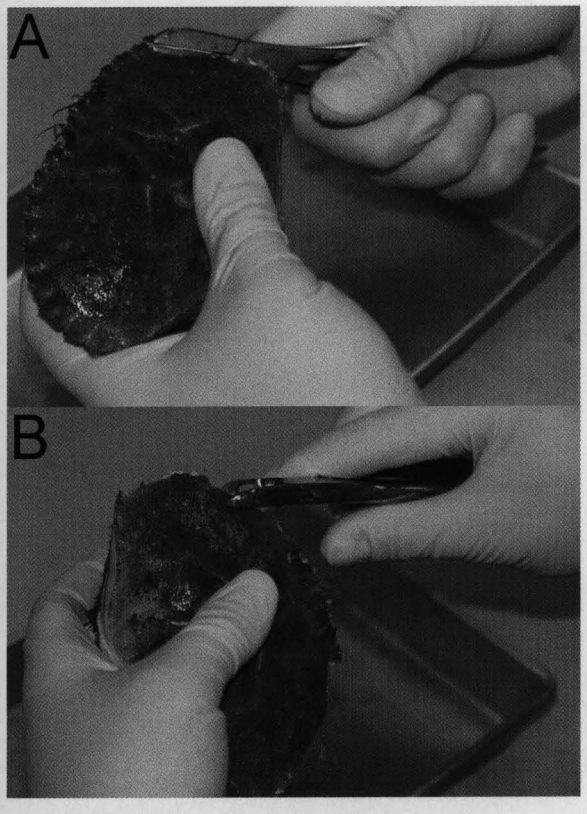

Figure 2.11: A. Use side cutters to crack away outer edge of the shell. **B.** Crack and remove shell around the crab almost to the eyestalks.

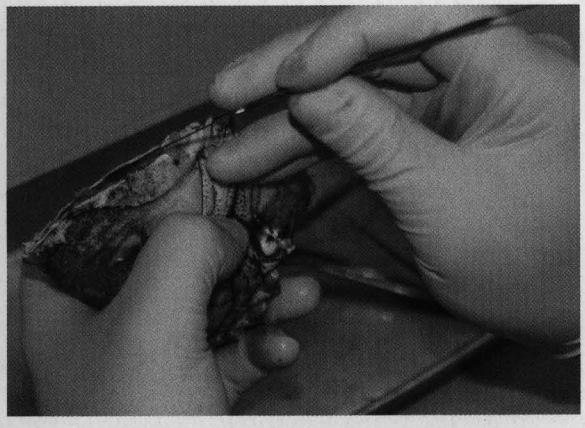

Figure 2.12: Separate hypodermis from the shell using rounded end of spatula.

Figure 2.13: Shell should be cracked away as illustrated and hypodermis separated all the way across crab.

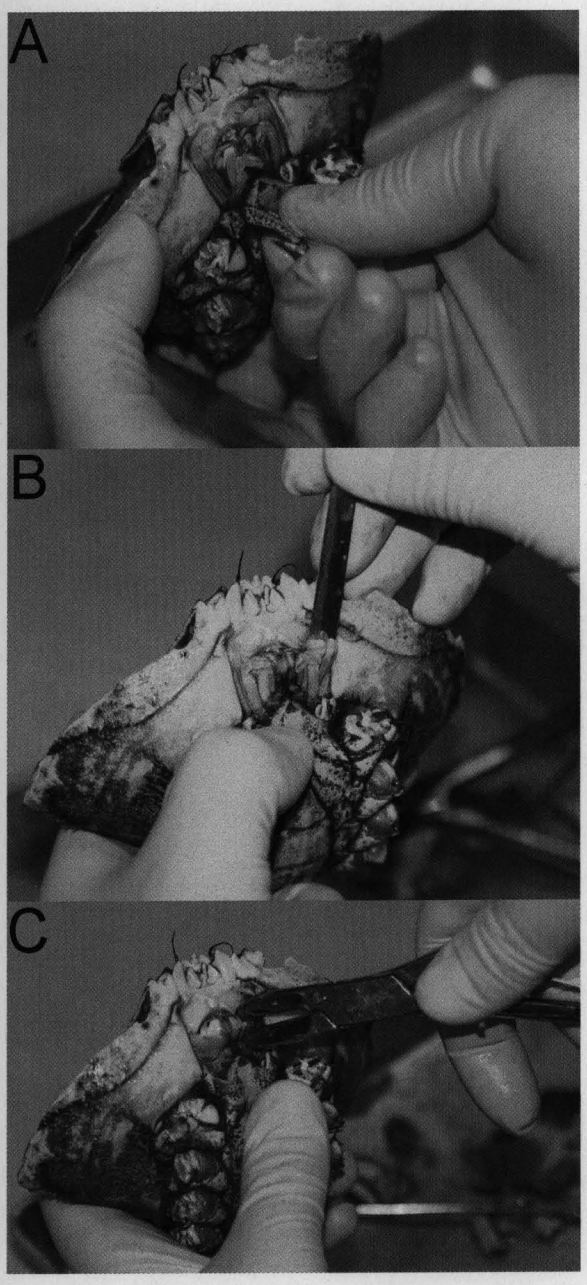

Figure 2.14: **A.** Twist and remove the outer mouthparts. **B.** Use a spatula to pull away and remove more of the mouthparts. **C.** Finally, use the side cutters to remove the last mouthparts over the lips.

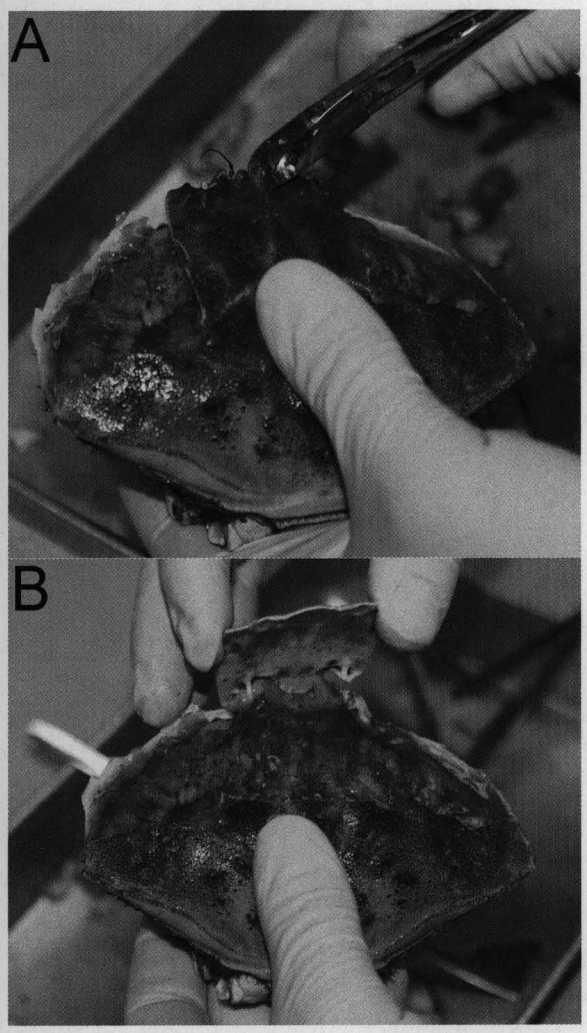

Figure 2.15: A. Crack the shell away from the front of the crab and around the eyestalks. **B.** After cracking it loose, carefully lift the middle of the crab shell away from the crab.

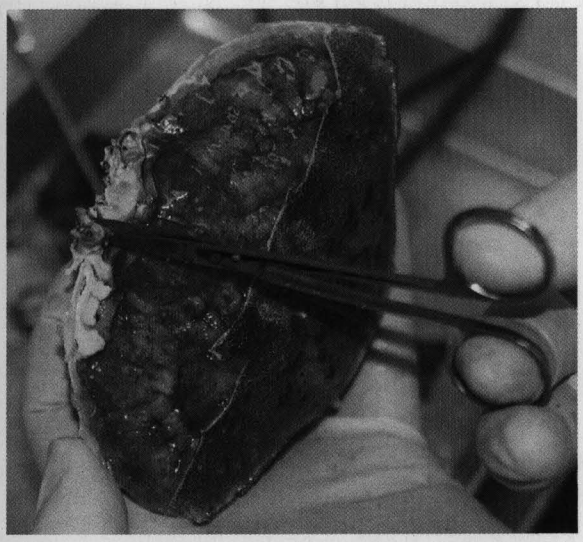

Figure 2.16: Removal of the eyestalks for later SG removal.

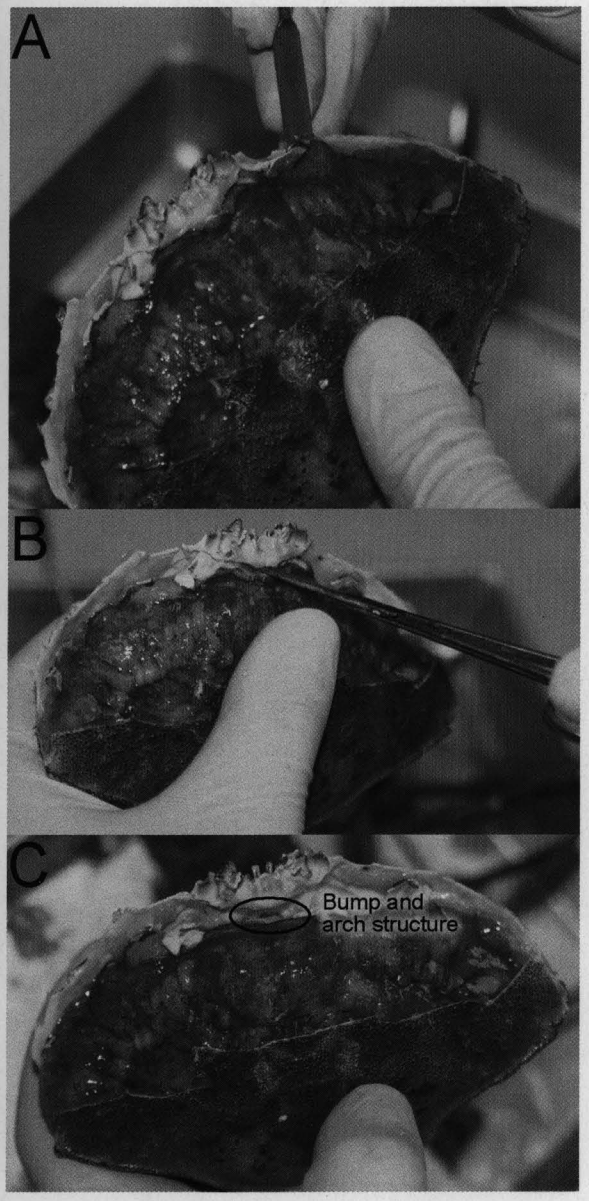

Figure 2.17: **A.** Use the spatula to separate the hypodermis from the lower shell. **B.** Use the scissors against the inside front of the crab to cut tissue away. **C.** Cut and scrape tissue away until a small bump on the inside of the shell and an arch structure created by the shell are revealed.

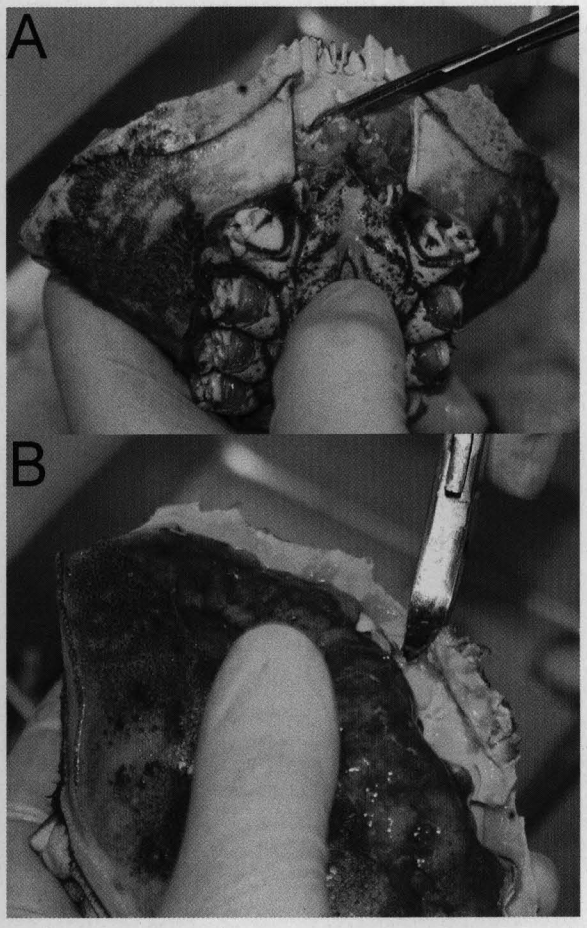

Figure 2.18: **A.** Cut the oesophagus free using the small scissors. **B.** Use the side cutters to cut away the shell from above the oesophagus.

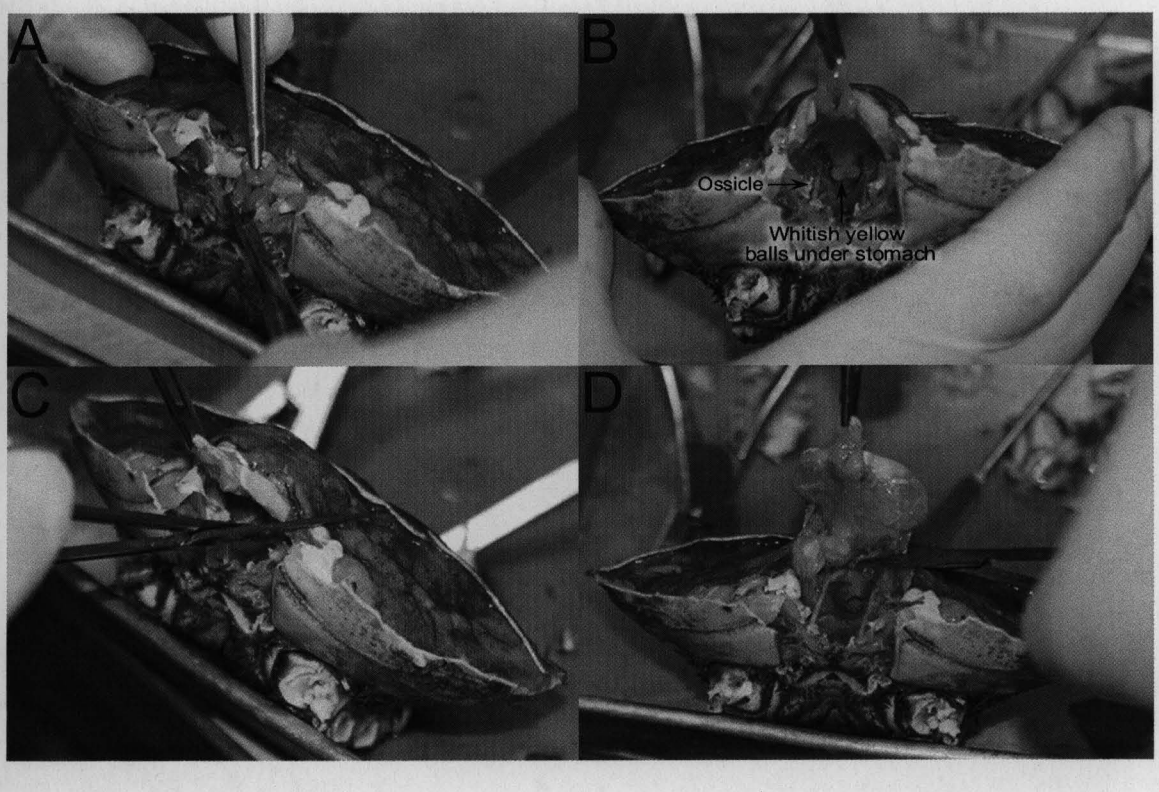

Figure 2.19: **A.** Grasp the lips with the toothed forceps and cut at the base of the oesophagus. **B.** Cut back, into crab, above ossicles, till you see the pylorus (two whitish yellow balls). **C.** Cut the hypodermis along stomach. **D.** Cut behind the stomach to remove it.

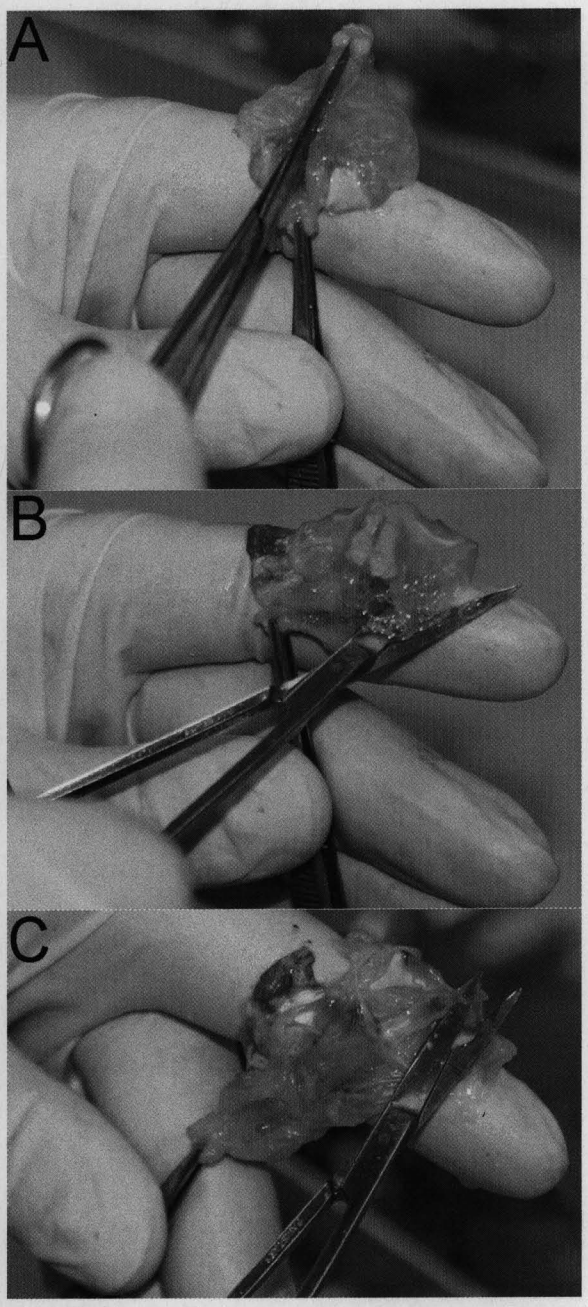

Figure 2.20: **A.** Cut the oesophagus open straight back to the rear of the stomach. **B.** Make two cuts on either side of the stomach angling back to the rear corners of the stomach. **C.** Cut out all three teeth.

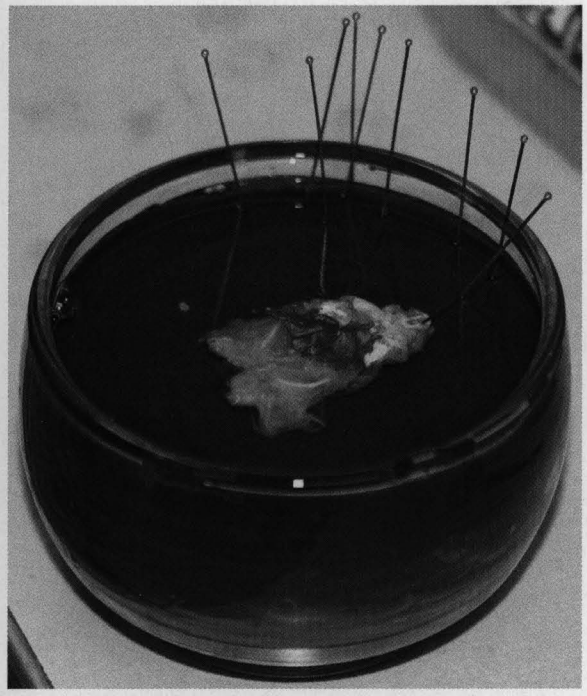

Figure 2.21: Place the stomach section into a rubberized dissecting dish with enough saline solution to cover it.

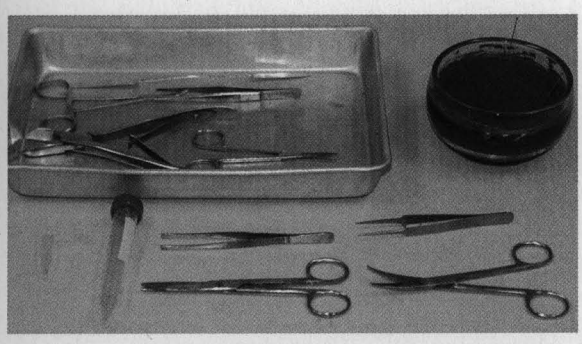

Figure 2.22: Supplies for the second part of the macrodissection.

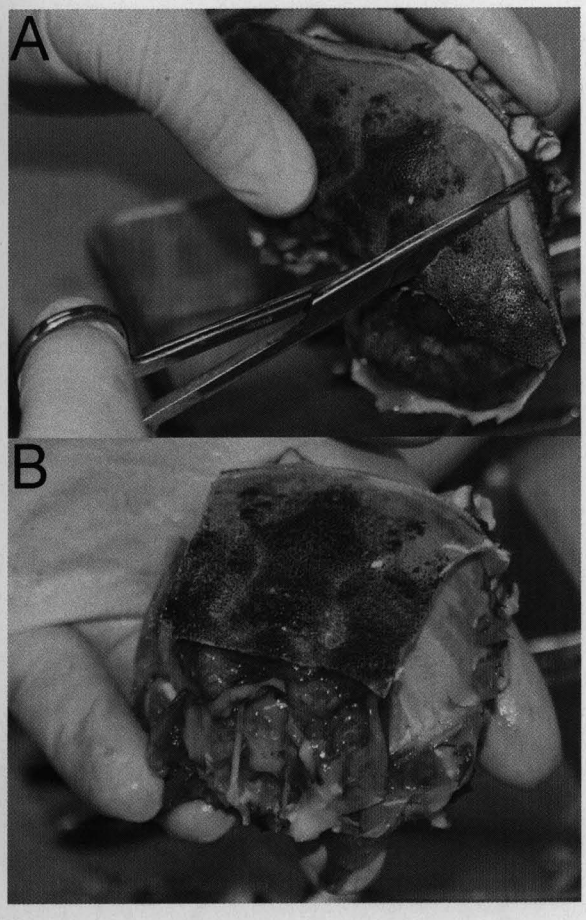

Figure 2.23: **A.** Cut the shell on both sides of the midline. **B.** Remove the tissue from both sides.

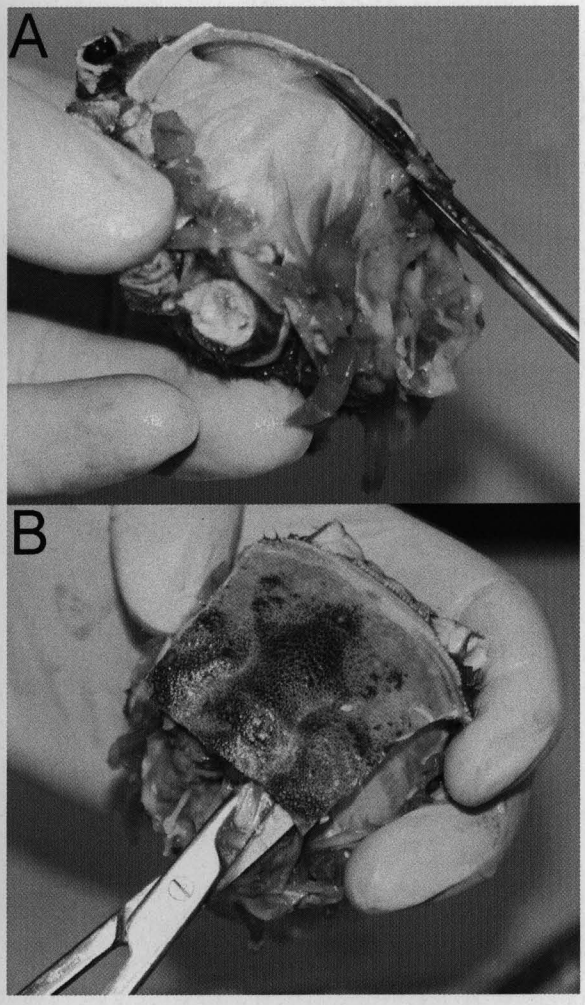

Figure 2.24: A. Cut the connection between the cardial ridge and the upper shell by (**B**) placing the scissors on either side of the cardial ridge and cutting.

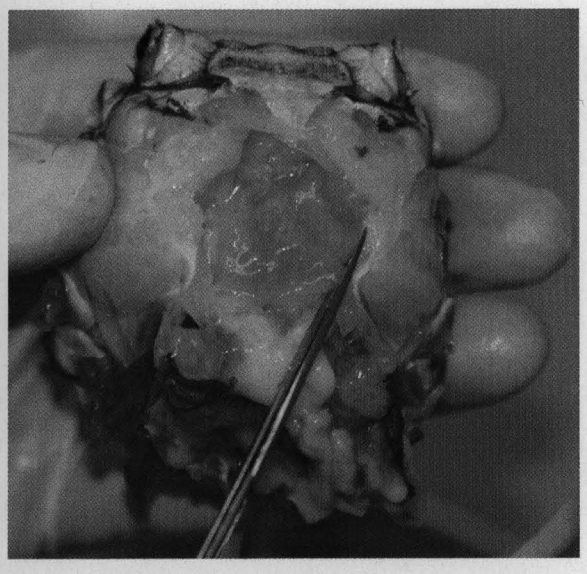

Figure 2.25: Cut away the heart, being careful not to damage the inside of the cardial ridges.

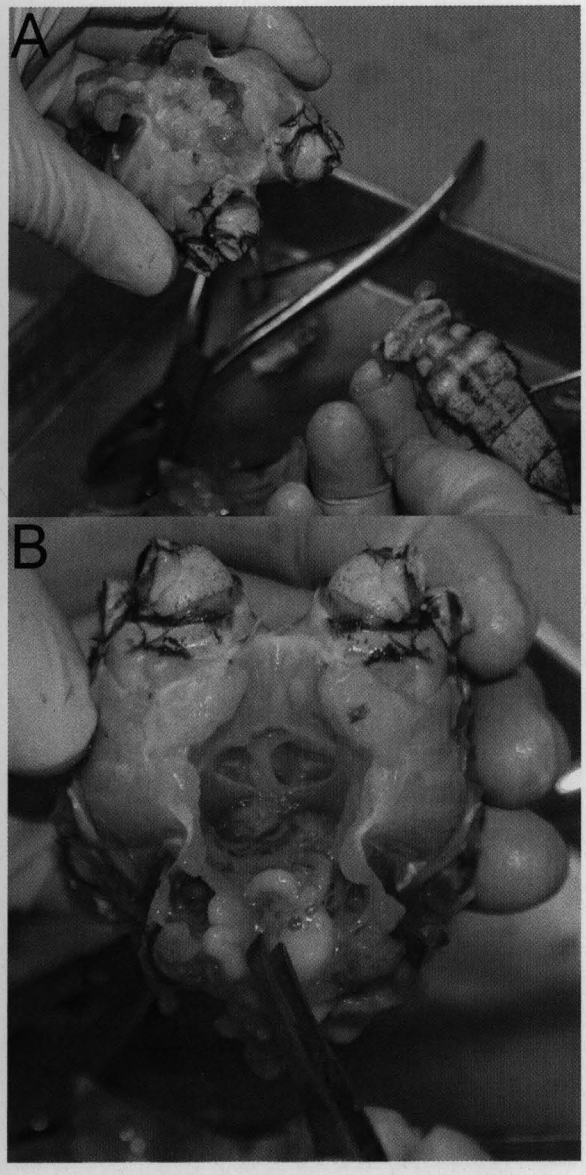

Figure 2.26: **A.** Remove the tail. **B.** Use forceps to clean out the tissue between the cardial ridges.

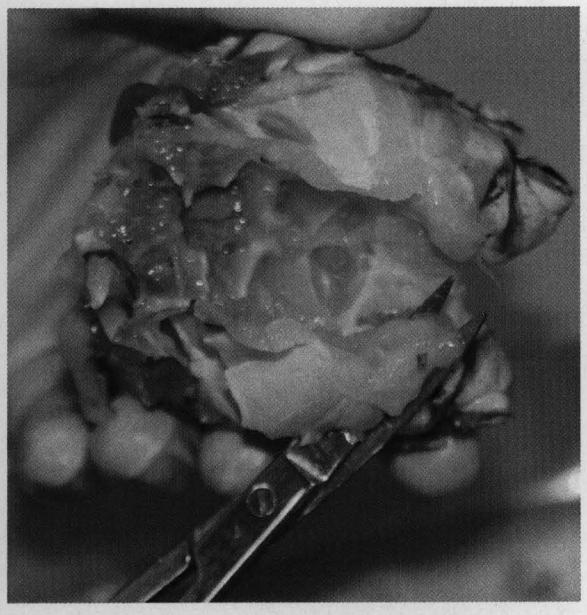

Figure 2.27: Cut the cardial ridges away with the large scissors and place them in saline.

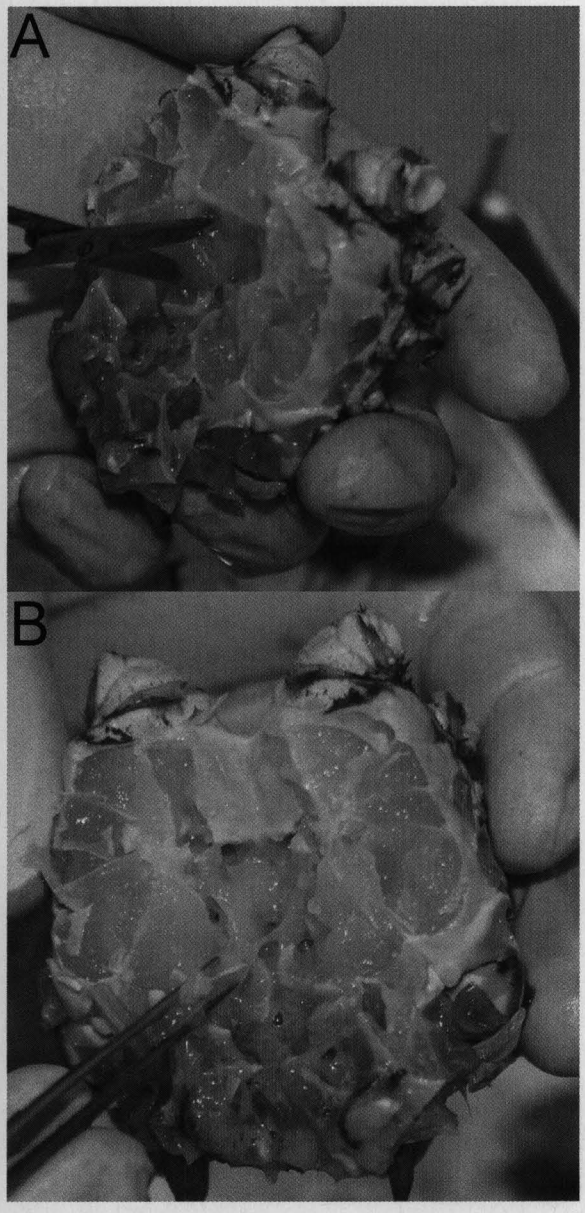

Figure 2.28: **A.** Cut the ossicles around the TG. **B.** Remove the ossicles with forceps.

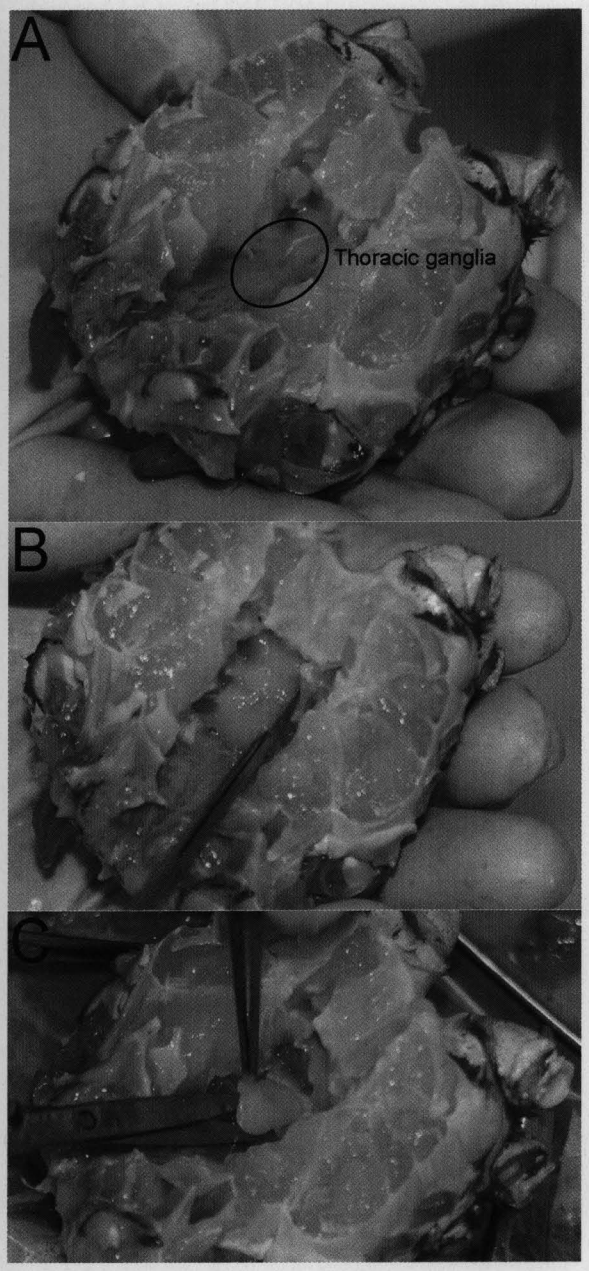

Figure 2.29: **A.** The thoracic ganglia looks like a wagon wheel with ganglia projecting out of it. **B.** Cut around the TG to free it. **C.** Grab it with clean forceps and lift it out, cutting any remaining tissue as you do so.

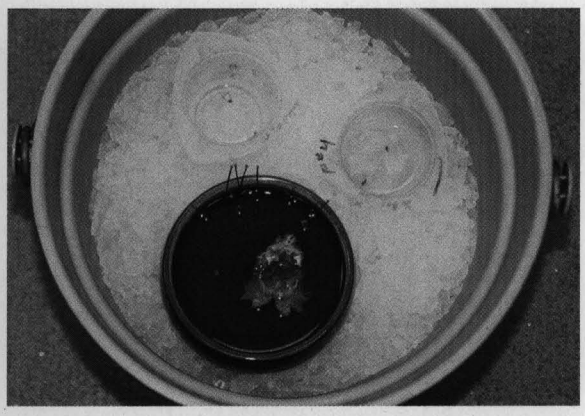

Figure 2.30: Make sure to place the stomach section, the eyestalk, and the cardial ridges on ice or in the fridge.

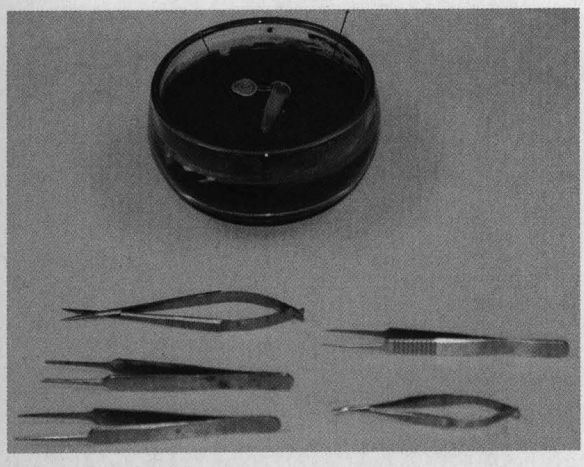

Figure 2.31: Materials needed for microdissection of SGs.

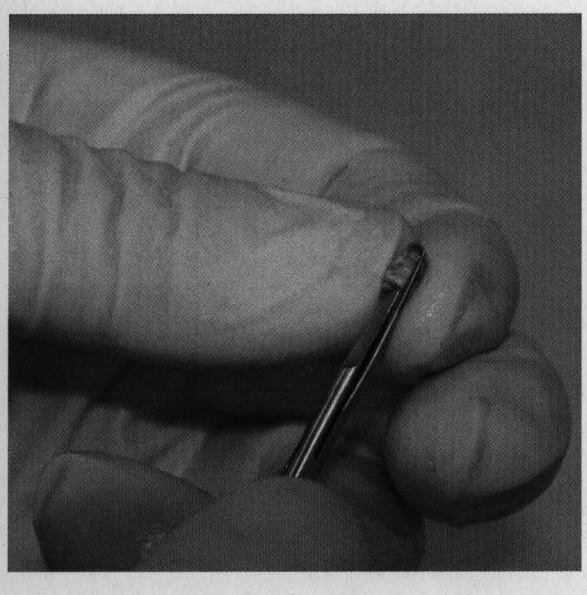

Figure 2.32: Using the large spring scissors, cut open the eyestalks on both the concave and convex sides.

Figure 2.33: **A.** Using the sturdy, fine forceps, separate both sides of the eyestalk. **B.** Watch for a bright, iridescent spot embedded in the tissue.

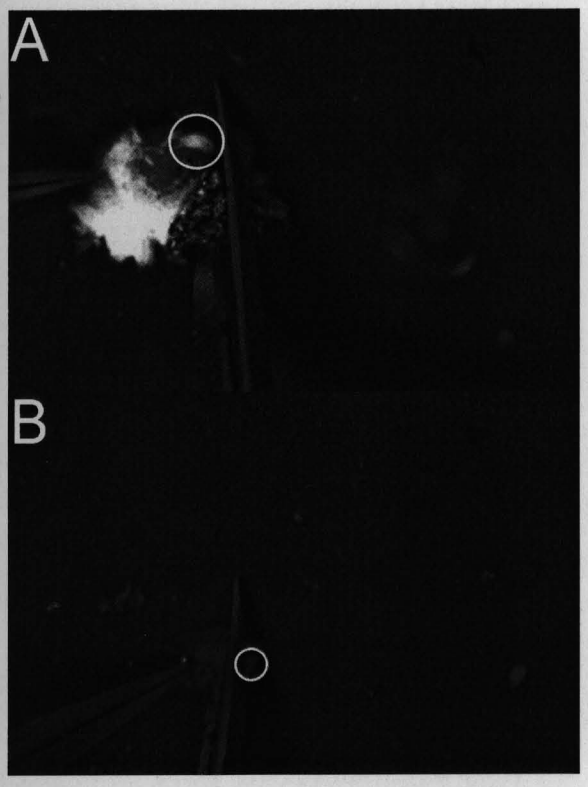

Figure 2.34: **A.** Cut the tissue away from around the SG. **B.** Pull the tissue through the scissors to better trim the SG. Circles indicate the SG.

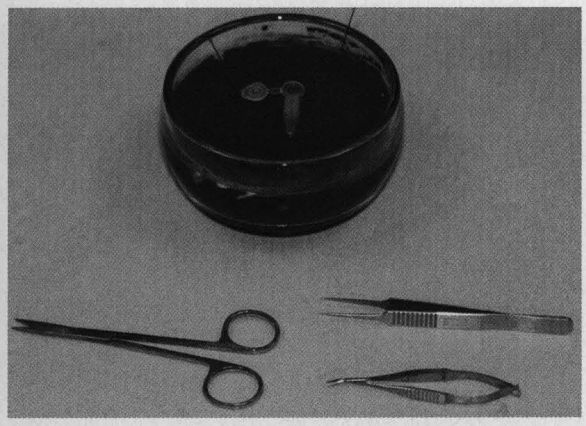

Figure 2.35: Materials needed for PO dissection.

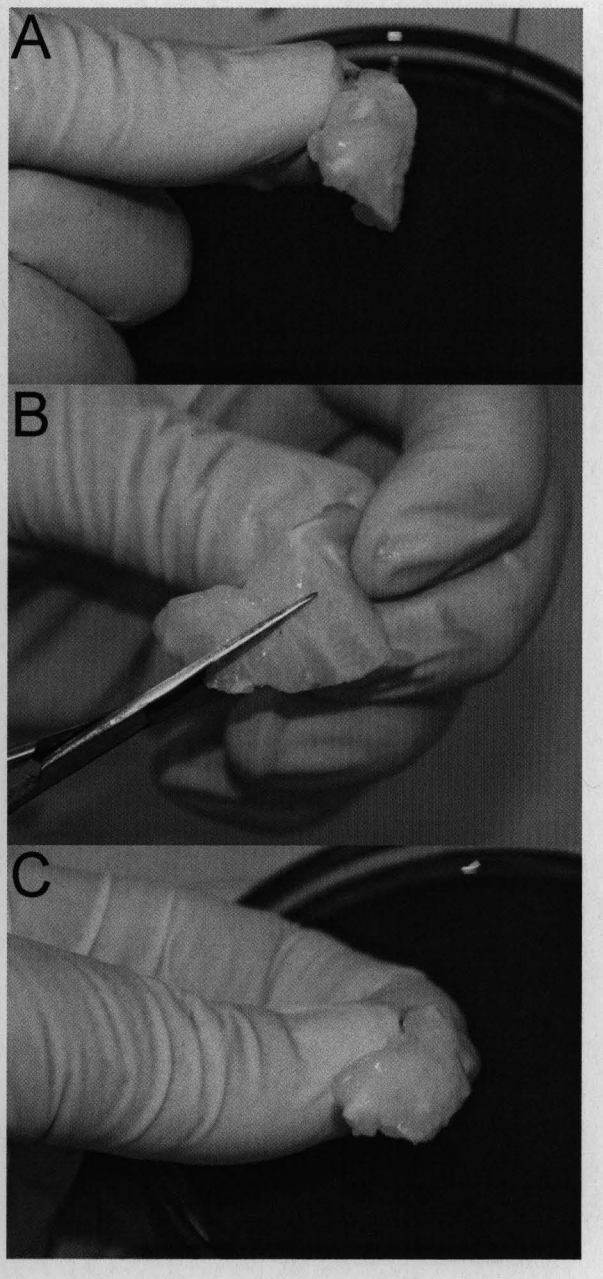

Figure 2.36: **A.** The outer side of the cardial ridge creates an edge that makes dissection difficult. **B.** Using the small scissors cut away the outside edge of the cardial ridge. **C.** Doing this creates a flat side on which to lay the cardial ridge in the dissecting dish.

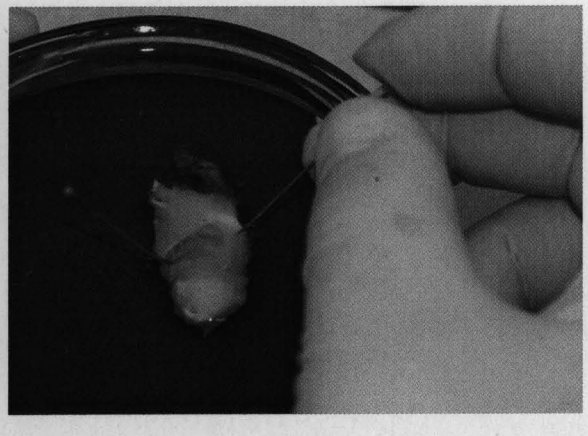

Figure 2.37: Place pins in the ventral and dorsal sides of the cardial ridge, making sure not to damage the PO.

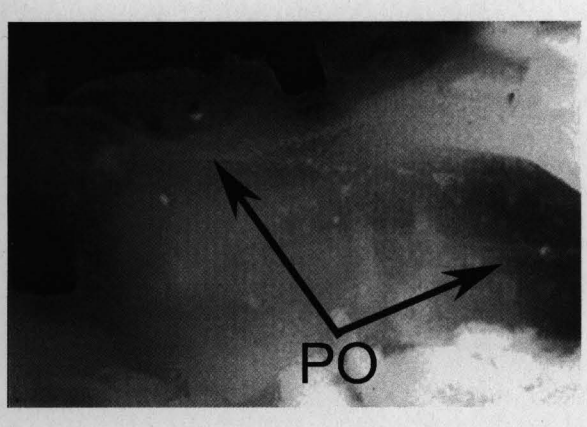

Figure 2.38: The PO can be identified as highly iridescent ganglia spanning the inside of the cardial ridge.

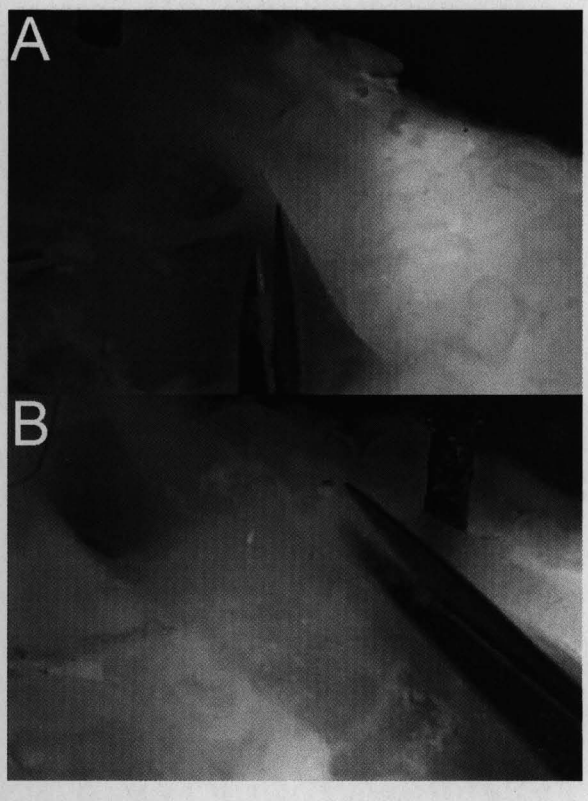

Figure 2.39: A. Start removing the POs by cutting out the junctions first. **B.** Continue cutting out the PO by cutting alongside the commissures.

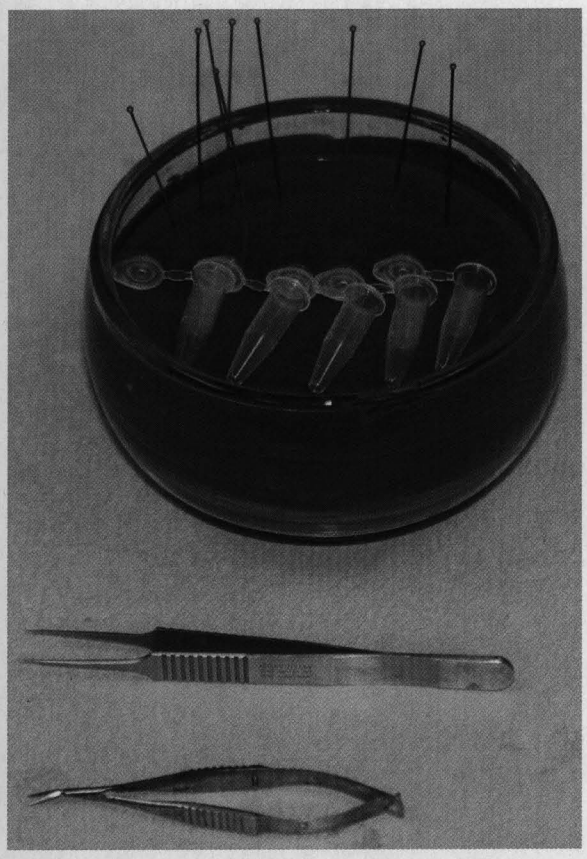

Figure 2.40: Tools and materials needed for STNS microdissection.

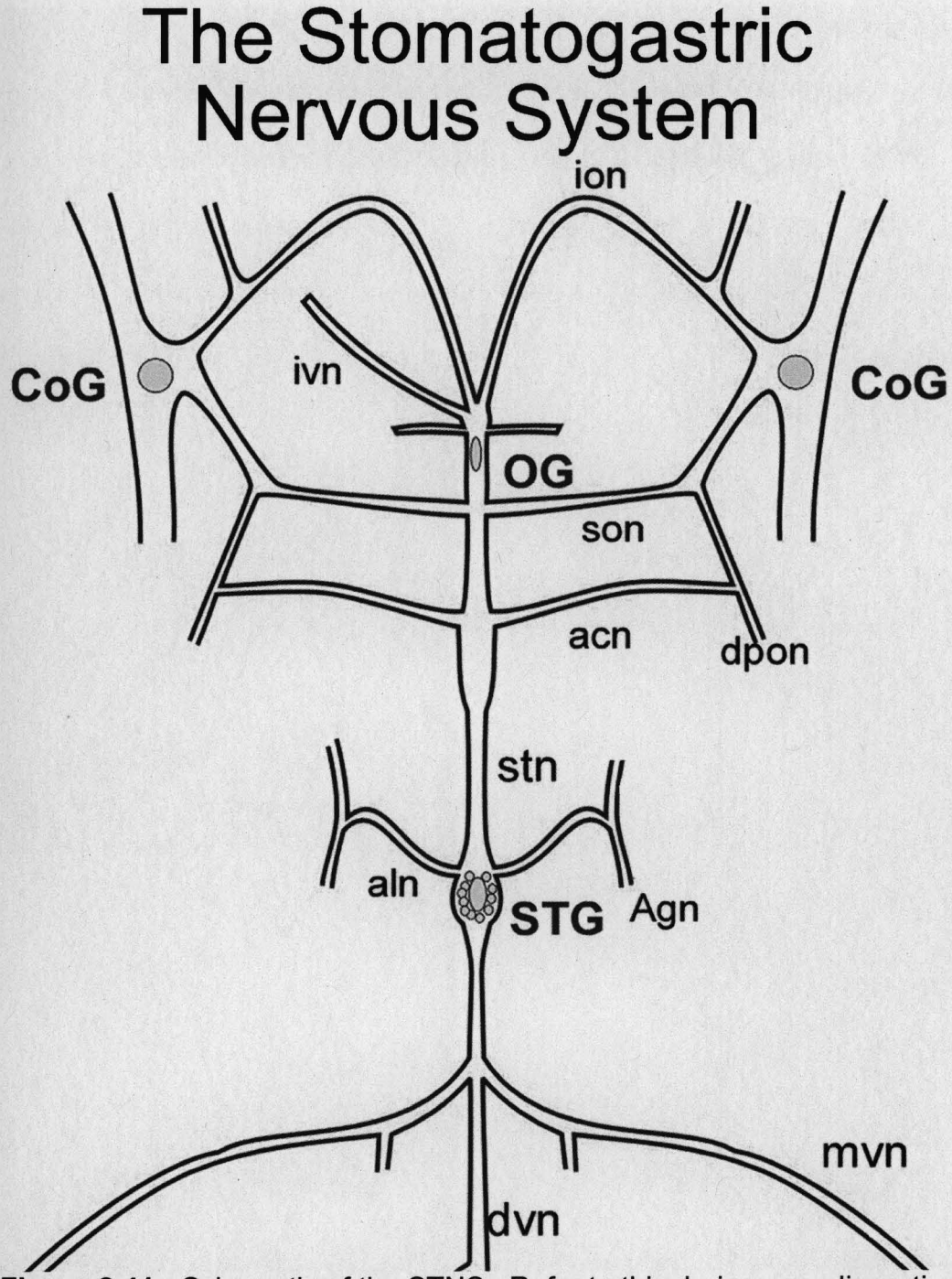

Figure 2.41: Schematic of the STNS. Refer to this during your dissection.

Image taken from Farzan Nadim's Lab website at Rutgers University.

(http://cancer.rutgers.edu/stg_lab/protocols.htm)

CoG -- Commissural Ganglia OG -- Oesophageal Ganglia

STG	-- Stomatogastric Ganglia	ACP	-- Anterior Cardial Plexus
ion	-- inferior oesophageal nerve	*ivn*	-- inferior ventricular nerve
son	-- superior oesophageal nerve	*acn*	-- anterior cardial nerve
stn	-- stomatogastric nerve	*aln*	-- anterior lateral nerve
dvn	-- dorsal ventricular nerve	*mvn*	-- median ventricular nerve

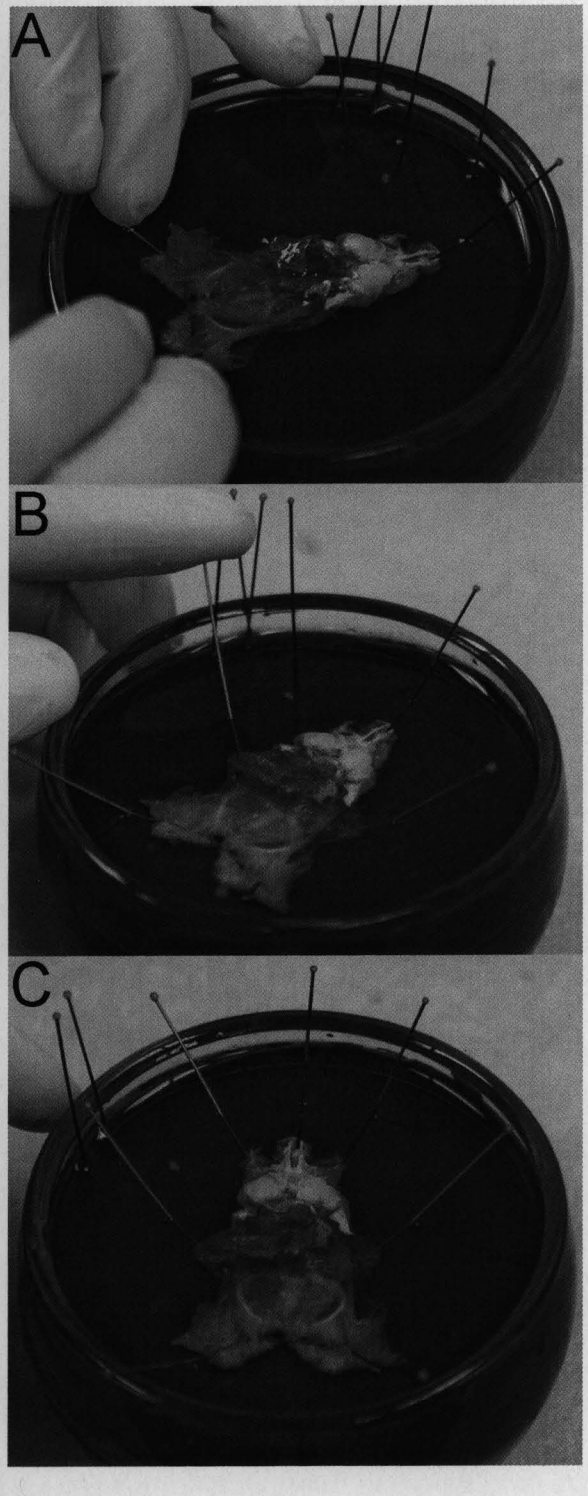

Figure 2.42: **A.** Pin the bottom of the stomach, pulling it slightly tight. **B.** Next, pin the sides of the stomach, again slightly tight. **C.** Finally, pin either side of the oesophagus, just below the lips.

Figure 2.43: The brain (circled) embedded in the tissue above the stomach.

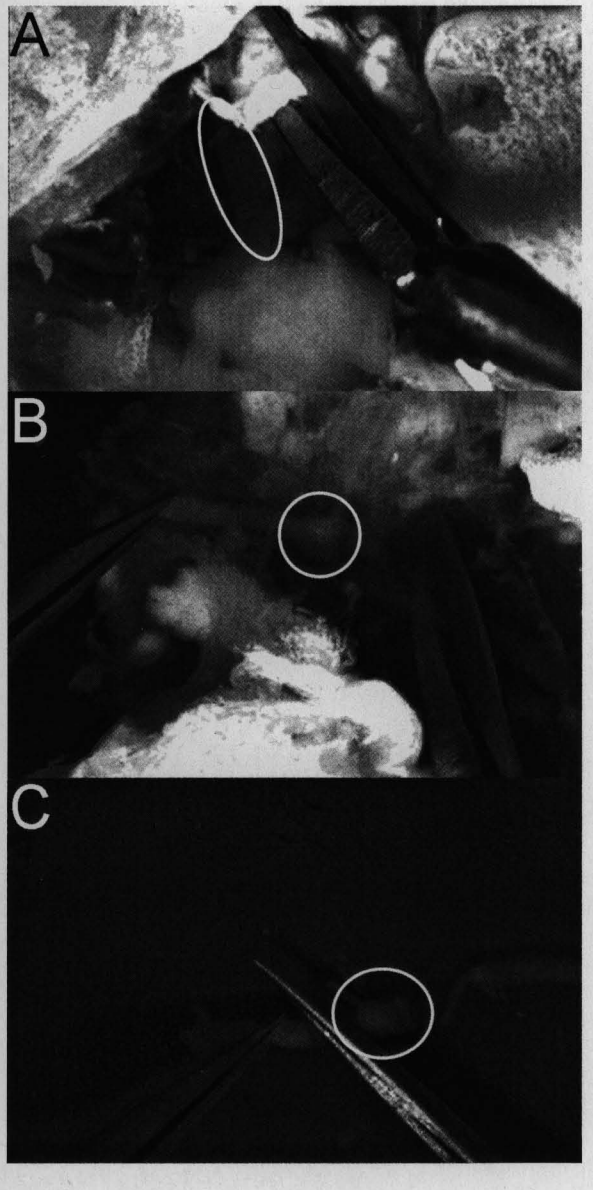

Figure 2.44: **A.** Follow the large commissure (circled) away from the brain towards the lips. **B.** Identify the CoG and cut it free. **C.** Trim the commissures from the CoG before placing it in the appropriate tube.

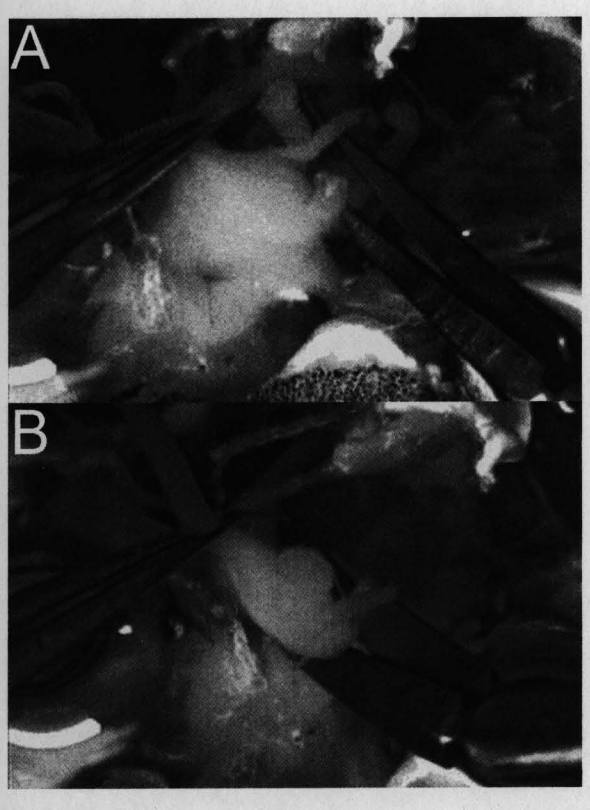

Figure 2.45: **A.** Cut all the commissures leading to the brain. **B.** Free the brain by cutting any connections between the brain and the tissue below.

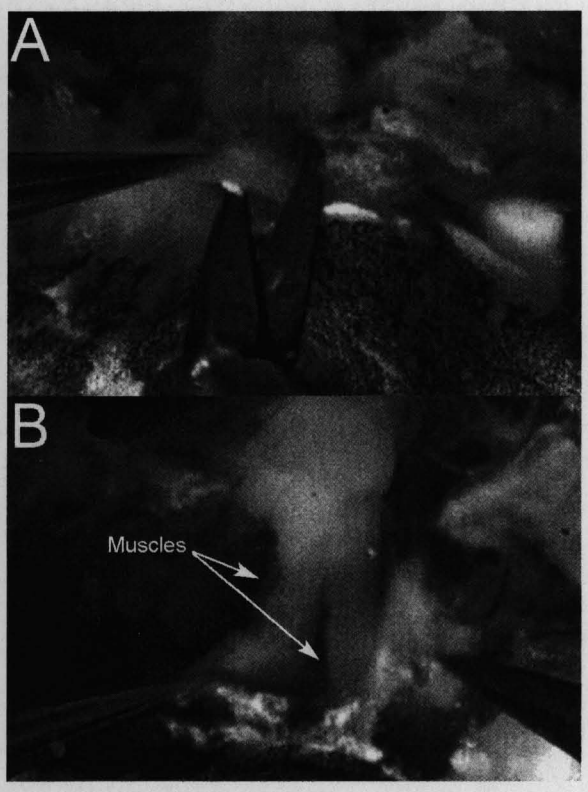

Figure 2.46: **A.** Cut straight down towards the dish from where the brain was. **B.** Watch for two muscles inside an artery running along the midline.

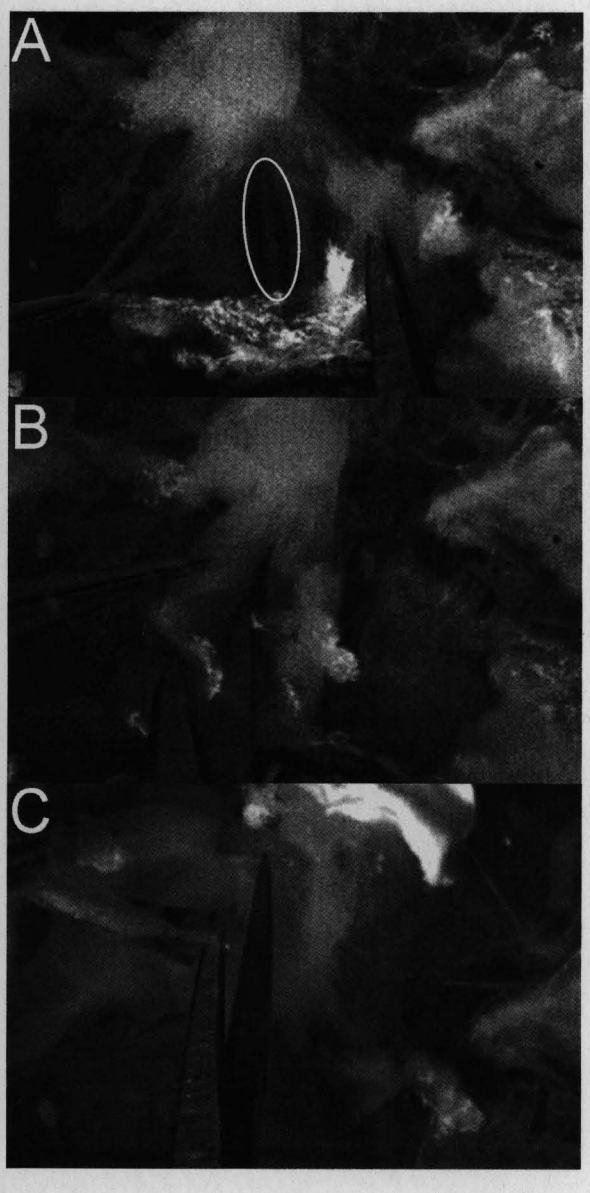

Figure 2.47: A. Look down between the muscles for the *stn* (circled). **B.** Carefully cut out the muscles. **C.** Continue cutting forward under the muscles, making sure not to cut the *stn*.

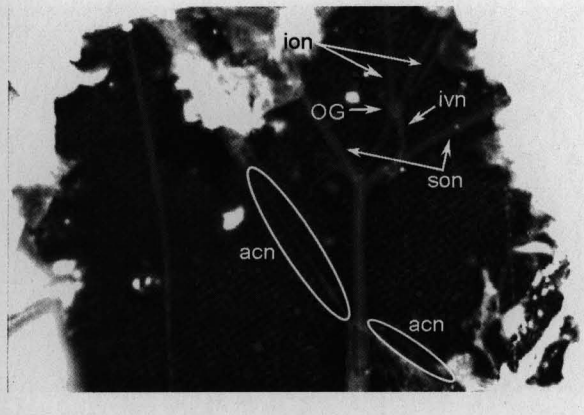

Figure 2.48: The forward (rostral) portion of the STNS showing the *acn*, *son*, *ion*, *ivn*, and the OG. Note: A black piece of cardboard was placed behind the STNS for better contrast.

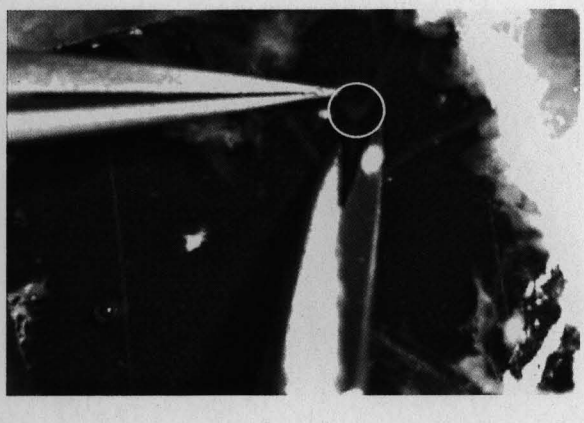

Figure 2.49: Cut out the OG (circled) by grasping the *ivn* and cutting the projecting commissures.

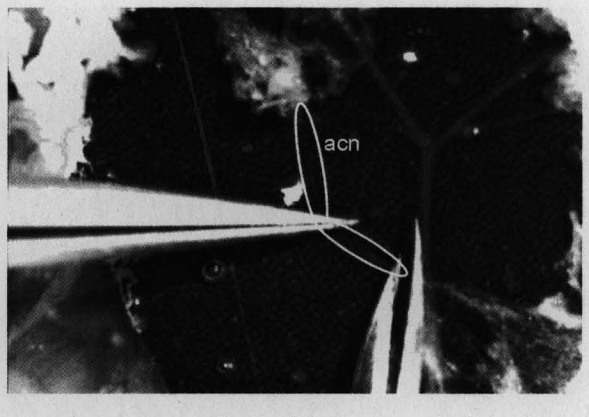

Figure 2.50: Removal of the ACPs which are on the *acn*s.

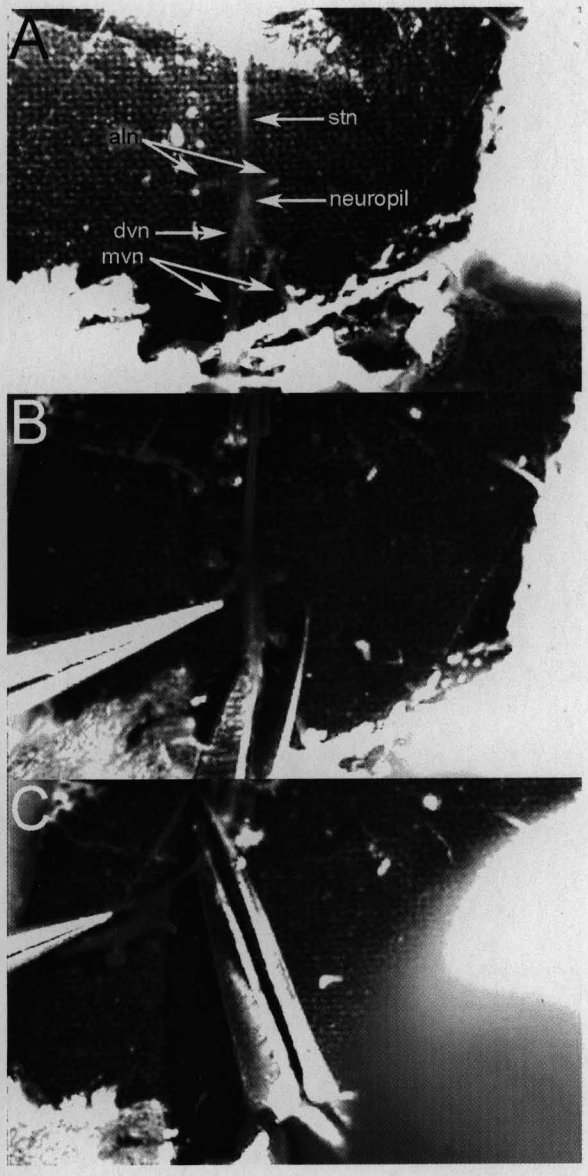

Figure 2.51: **A.** STG with the neuropil and the projecting commissures. **B.** Cut the *mvn*s to begin removing the STG. **C.** Finally, cut the *stn* to remove the STG.

Chapter Three

In situ Tissue Analysis of Neuropeptides by MALDI FTMS In-Cell Accumulation

The coauthors and work for this chapter was completed in cooperation with Kimberly K. Naber (Kutz) and Lingjun Li.

3.1 Abstract

Herein we report the first application of Fourier transform mass spectrometry (FTMS) for the analysis of neuropeptides directly from neuronal tissues. Sample preparation protocols and instrumentation conditions are developed to allow *in situ* neuropeptide analysis of the neuroendocrine organs freshly isolated from a marine organism *Cancer borealis*. The utility of a previously developed in-cell accumulation (ICA) technique is extended for peptide analysis in complex tissue samples. With the ICA procedure, ion signals from multiple laser shots are accumulated in the analyzer cell prior to detection. This procedure allows the accumulation of ion signals without accumulating noise, thus improving the signal-to-noise ratio and enhancing the sensitivity for the detection of trace-level endogenous neuropeptides. *De novo* sequencing of peptides directly from tissue samples becomes more feasible through this improvement. Additionally, an integrated pulse sequence is constructed to cover a wide mass range from *m/z* 215 – 9,000 by centering quadrupole collection of ions at different masses for successive laser shots. Finally, improved mass measurement accuracy (2 ppm) for tissue peptide analysis is achieved using ICA by incorporating calibrants on a separate spot from the sample of interest without premixing calibration standards with the analytes.

3.2 Introduction

Neuropeptides, a class of important chemical messengers, are ubiquitous in the nervous systems at all levels of organizations, ranging from hydrozoans to humans. As the most diverse and complex group of signaling molecules, neuropeptides induce and regulate many important physiological processes throughout the animal kingdom[1-3]. Knowledge of complete neuropeptide profiles at the cellular and network levels is critical for a better understanding of peptidergic signaling in a neuronal circuitry. Classical methods for peptide identification often involve the use of immunocytochemical techniques[4; 5]. This approach, while a powerful first step in determining the presence of neuropeptides, suffers from limitations including cross-reactivity with structurally similar peptides and the inability to analyze large numbers of peptides simultaneously and to discern between structurally similar peptides. Alternatively, one can pool tissues of interest from a large number of animals to generate a tissue extract, followed by multiple steps of fractionation and purification for conventional peptide sequencing, such as Edman degradation[6; 7]. This method, while well established and highly accurate, requires a large amount of starting material and is very time-consuming. It is also difficult for the discovery of low-abundance endogenous neuropeptides from complex tissue samples.

Because of these analytical challenges and difficulties, biological mass spectrometry has become the method of choice for neuropeptide analysis due to its speed, high sensitivity, chemical specificity and capability for complex mixture

analysis[8-12]. In particular, matrix-assisted laser desorption/ionization (MALDI) time-of-flight mass spectrometry (TOF MS) has evolved as a viable tool for peptide profiling directly from biological tissues[13-20], single cells[21-24], and even individual organelles[25]. Although a highly sensitive and rapid molecular weight profiling technique, MALDI-TOF has limited mass measurement accuracy, resolving power and lacks the capability to perform effective tandem mass spectrometry experiments for peptide sequencing. Several other mass analyzers such as quadrupole ion trap and TOF-TOF coupled with MALDI have been reported for peptide structural analysis in biological samples[26-28].

Among various types of mass analyzers, Fourier transform mass spectrometer (FTMS) offers several unique advantages including high mass resolving power, high mass measurement accuracy (MMA), and multi-stage MS/MS capabilities[29-33]. Furthermore, the ability to trap and store ions while allowing additional ions to be introduced in the ion cyclotron resonance (ICR) cell makes it possible to accumulate ions from multiple MALDI ionization events in the FT ICR cell prior to detection. Such in-cell ion accumulation (ICA) methods produce significant signal-to-noise improvement over the more commonly used signal averaging methods. This improvement is due to the fact that ion signals are accumulated over several laser shots without the accumulation of noise. Variations of such ICA schemes have been reported for FTMS instruments equipped with both internal and external MALDI sources[34-36]. Additionally, the ICA pulse sequences provide flexibility to incorporate calibrant ions, desorbed

from a separate sample spot, in the ICR cell to allow effective internal mass calibration without premixing the calibrant and analyte[35; 37].

Although FTMS has become increasingly popular in the field of proteomics[38-47], and the detection of seven ribosomal proteins from whole bacteria cells by MALDI FTMS was recently reported[48], the use of this high performance MS technique for *in situ* analysis of trace-level endogenous neuropeptides directly from some of the most complex samples – neuronal tissues, has not been investigated. As neuropeptides of interest generally fall in the *m/z* 500 – 10,000 range, MALDI FTMS appears to be well-suited for the detection of these peptides.

In the current study, we demonstrate the utility of MALDI FTMS in conjunction with the ICA technique for neuropeptide analysis directly from neuroendocrine tissue isolated from the model system *Cancer borealis*, or the Jonah crab. Improved sample preparation methods have also been developed to allow effective detection of neuropeptides by MALDI FTMS through microextraction using acidified methanol. The mass accuracy and resolution remain excellent for these unconventional samples as well as an enhancement of sensitivity for low-abundance peptide detection being observed. We also report the first example of *de novo* peptide sequencing directly from neuroendocrine tissue using the ICA method with MALDI FTMS. Additionally, variation of ICA pulse sequences to include internal calibrants (on a separate spot) significantly improves MMA. This in turn, permits high confidence peptide identification in complex neuronal samples. Finally, a broad-band ion accumulation and MS

detection covering a wide mass range is accomplished by further modification of the ICA pulse sequences.

3.3 Experimental Section

3.3.1 Animals and Dissection

Jonah crabs, *Cancer borealis*, were shipped from Marine Biological Laboratories (Woods Hole, Massachusetts) and maintained without food in an artificial seawater tank at 10-12 °C. Animals were cold-anesthetized by packing in ice for 15-30 min prior to dissection. They were dissected by removing the stomach section, eye stalks, and pericardial ridges located on either side of the heart. The pericardial organs (PO) were removed from the pericardial ridges and the sinus glands (SG) were removed from the eye stalks. The commissural ganglion (CoG) and the stomatogastric ganglion (STG), containing the neuropil, were removed from the stomach portion of the crab. The SG, STG, and CoG are on the order of a few square millimeters. The PO is approximately 0.5 mm by two to three cm in length. All dissection was carried out in chilled physiological saline (composition in mM: NaCl, 440; KCl, 11; $MgCl_2$, 26; $CaCl_2$, 13; Trizma base, 11; maleic acid, 5; pH 7.45).

3.3.2 Sample Preparation

Unless otherwise stated, tissue was briefly rinsed in acidified methanol (90% methanol (Fisher Scientific), 9% glacial acetic acid (Fisher Scientific), and 1% deionized water), immediately following dissection. Tissue was then desalted

in 10 mg/mL 2,5-dihydroxybenzoic acid (DHB) (ICN Biomedical Inc.) prepared in deionized water. A spot containing 0.6 μL of saturated DHB (150 mg/mL in 50:50 water: purge trap grade methanol (Fisher Scientific), v/v), was placed on one facet of the IonSpec MALDI sample probe. Before the DHB crystallized fully, the tissue was placed carefully on the facet and an additional 0.6 μL drop of saturated DHB was placed on top of the tissue to help affix the tissue to the target. The saturated DHB was then allowed to crystallize at room temperature.

3.3.3 Instrumentation

MALDI experiments were performed on an IonSpec Fourier transform mass spectrometer (Lake Forest, CA) equipped with a 7.0 Tesla actively-shielded super conducting magnet. The FTMS instrument contains an external ion source utilizing a quadrupole ion guide to transfer the ions to the ICR cell, which is differentially pumped. The sample probe is a ten-faceted stainless steel target. A 337 nm nitrogen laser (Laser Science, Inc., Franklin, MA) was used for ionization/desorption. All mass spectra were collected in positive ion mode.

3.3.4 In Cell Accumulation (ICA)

The ICA method was written using IonSpec version 7.0 software. Several types of ICA methods were used, the simplest being a pulse sequence where each of the seven laser ionization/desorption events was optimized for a specific mass for transport down the quadrupole. This method was used to increase the

ion concentration and observe the ions present. The optimized mass used most frequently for neuropeptide analysis was at *m/z* 1000. As shown in **Figure 3.1**, for every desorption/ionization event there was a corresponding quadrupole radio frequency (RF) burst. The first event was at 70 ms, with subsequent events exactly 1000 ms apart, each with a width of 50 ms. The frequency corresponding to the transport of *m/z* 1000 down the quadrupole was 2165 kHz and the base to peak amplitude was 378.675 V. There was an initial surface preparation laser pulse at 0 ms with the duration of 1 ms and 50% laser intensity. Following this there were seven laser pulses with the first at 100 ms and the rest exactly 1000 ms apart, with 1 ms duration and 100% laser intensity. The MALDI valve was pulsed seven times; the first pulse occurred at 50 ms with a width of 250 ms, and the following six pulses were 1000 ms apart with the same duration. The inner trapping rings were held at 0.5 V throughout the experiment but were pulsed once to -9.0 V at -5 ms with a width of 5 ms. The filament trapping plate was initialized to 20 V, pulsed once at -5 ms for a duration of 1 ms to a voltage of 30 V, and ramped to 0 V at 12000 ms and a width of 1000 ms. The quadrupole trapping plate was initialized to 20 V. A pulse of the quadrupole trapping plate to -30 V at -5 ms with a width of 1 ms was used to quench the cell prior to ionization/desorption. The second pulse was set to 0 V at 98 ms with a width of 2.599 ms, the width being mass specific. The following six pulses were set to 4 V and exactly 1000 ms apart with the same width. Following all the pulses, the quadrupole trapping plate, along with the filament trapping plate, was ramped to 0 V at 12000 ms (just before detection) with a width of 1000 ms to reduce base

line distortion of peaks. Electron energy was initialized to 2 V, and the left and right source plates were initialized to 100 V. The extractor plate and the quadrupole offset were initialized to -30 V and held constant throughout analysis. Ions were excited with a RF Sweep beginning at 13000 ms with a width of 2.900 ms, amplitude of 150 V base to peak, and mass range of *m/z* 100- 9000. Detection was performed in broadband mode from *m/z* 108.000 to 2000.000, starting at 13006.996 ms and acquiring for 1048.576 ms. There were 2097152 transient points collected with an analog to digital conversion (ADC) rate of 2 MHz and an ADC gain of one.

3.3.5 Full Mass Scan

The full mass scan was an ICA pulse sequence optimized for several different masses combined into one pulse sequence containing seven total mass dependent pulses. Two quadrupole RF pulses were used at *m/z* 1000, one pulse at *m/z* 2500, two pulses at *m/z* 5000, and two pulses at *m/z* 8000. The experimental setup was much the same as the *m/z* 1000 ICA program, with minor changes to the quadrupole RF bursts, the quadrupole trapping plate, and the detection event. The frequency and amplitude used for each mass of interest for the quadrupole bursts were: *m/z* 1000 corresponding to 2165.000 kHz and 378.675 V; *m/z* 2500 at 1031.000 kHz and 256.732 V; *m/z* 5000 corresponding to 1031.000 kHz and 376.732 V; *m/z* 8000 at 1031.000 kHz and 476.732 V. The

quadrupole trapping plate pulse widths also depend on the mass of interest. The widths used were: 2.599 ms for *m/z* 1000; 2.899 ms for *m/z* 2500; 3.215 ms for *m/z* 5000; 3.511ms for *m/z* 8000. The detection event was modified by changing the mass window covering *m/z* 215-12,000.

3.3.6 Internal Calibration on Adjacent Samples (InCAS)

InCAS experiments also employed gated trapping with pulsed gas cooling. However, two separate pulse sequences were loaded sequentially as shown in **Figure 3.2**. The first pulse sequence employed a single shot ionization/desorption event centered at *m/z* 1000 to trap and store ions from calibrant standards in the ICR cell. One notable difference is the filament trapping plate and the quadrupole trapping plate were not ramped to 0 V, but instead are maintained at 20 V to retain the ions in the ICR cell. Another difference was the inclusion of "dummy" excitation and detection events wherein the amplitude of ion excitation was set at 0 V, the detection transient points were limited to 16384, and the mass window was 215.000 to 215.010. The second pulse sequence was the same as the ICA sequence at *m/z* 1000 except that the cell quenching events that normally take place at -5 ms were omitted. Since the ions already in the cell were standards to be added to the analyte and were desired in the following detection, these events were deleted. The procedure of the experiment was to first load the ICR cell with calibrant ions using the single shot pulse sequence and then rotate the target and perform a second ICA pulse sequence which loaded the cell with analyte ions from the tissue samples prior to

detection of the resulting mixture. The acquired spectra contained both the calibrant ions and the analyte ions from the sample.

3.3.7 SORI-CID

Fragmentation of the peptides was accomplished by sustained off resonance irradiation and collision induced dissociation (SORI-CID)[49]. As shown in **Figure 3.1 B**, this was done by modifying the ICA pulse sequence to include an arbitrary waveform to isolate the ion of interest, a SORI RF pulse for excitation, and an additional MALDI valve pulse. Also, due to software limitations, the seventh ionization/desorption event could not be included, leaving a total of only six desorption/ionization events. Nitrogen gas was used to cool the ions and dissociate the peptides. The SORI event (1000 ms in duration) employed a frequency offset ranging from 1000 – 600 Hz and amplitudes of 4.0-5.0 V (base to peak) depending on the desired level of fragmentation and the energy needed to dissociate the peptide.

3.4 Results and Discussion

While peptide profiling directly from tissue samples and even single cells has been demonstrated with MALDI-TOF MS previously[21; 50-53], the extremely complex peptide profiles observed in bundles of neuronal processes and neurosecretory structures in these crustacean model systems[15; 54; 55] suggest a possible advantage of using FTMS-based instrumentation. The key features of

using a high performance MALDI FTMS instrument compared to previous work done on MALDI TOF instruments are several fold: (1) The high mass measurement accuracy achievable with FTMS significantly increases the confidence of peptide identification by molecular weight profiling; (2) The high mass resolving power inherent to FTMS allows differentiation of peptides with similar masses; and (3) The high quality MS/MS data allow confirmation of putative neuropeptide identity and discovery of novel neuropeptides or sequence tags by *de novo* sequencing.

Sample preparation often plays a crucial role in obtaining high quality MALDI MS data from tissue samples. Improvement upon a previously established sample preparation protocol results in enhanced signal quality, as illustrated in **Figure 3.3**. Briefly, a piece of neurohemal tissue, pericardial organ (PO), was dissected and cut in half. One half was rinsed briefly (2-3 seconds) in acidified methanol while the other was not rinsed. Both PO sections were then desalted in DHB matrix in an identical manner and mounted onto a MALDI sample probe for analysis. As shown in **Figure 3.3**, the PO rinsed with acidified methanol showed significant improvement of spectral quality and detection enhancement of many low abundance ions, as compared to the PO prepared with the conventional method. The expanded *m/z* 900-1300 region in **Figure 3.3 B** shows the detection of many of the ions not visible in **Figure 3.3 A** and an enhanced signal-to-noise ratio compared to the ions seen in **Figure 3.3 A**. Most notably is the detection of many high mass ions between *m/z* 4000 and 6000 that are not detectable in the PO sample rinsed only in DHB. Rinsing solutions were

tested and no significant peptide signal was observed (data not shown), however; rinsing solutions are present in large excess and dilution effects may mask any small amount of peptides lost from the tissue into the rinsing solutions. The negative effect of the loss of a small amount of peptide in the rinsing solution is out weighed by the improved spectral quality.

The improvement could be due to an enhanced protonation of peptides resulting from the use of acetic acid in the methanol extraction solution, therefore improving the ionization efficiency of the peptides. Alternatively, this improvement could also be due to a more efficient micro-scale extraction of peptides from the tissue samples, as the acidified methanol solution is commonly used for peptide extraction and isolation from biological tissue samples for HPLC separation[56-58]. An important question to address is whether the extra peaks observed after acidified methanol rinsing are due to a sample preparation artifact or the improved detection of lower abundance ions. Specifically, one might wonder if the extraneous peaks are the result of proteolytic cleavage or breakdown products of peptides in the presence of acidified methanol. However, the detection of higher mass ions that were not observed in regular DHB preparation suggests the likelihood of a sample preparation artifact is low. To further test this, a mixture of peptide standards was incubated in acidified methanol for a prolonged period (two weeks) and then analyzed. No cleavage products from these peptides were observed (data not shown). This suggests that the improved detection of endogenous neuropeptides from tissue samples is

not due to additional cleavage, but is a result of a more efficient extraction/ionization.

It is important to note that reproducible tissue-specific peptide profiles were obtained for several types of tissues examined. As shown in **Figure 3.4**, distinct peptide patterns are displayed for three different types of tissues isolated from the same species; (**A**) stomatogastric ganglion (STG) neuropil, (**B**) commissural ganglion (CoG), and (**C**) pericardial organ (PO). The three different neural tissue types were all analyzed under the same experimental conditions consisting of a brief rinse in acidified methanol followed by desalting in DHB. The instrumental parameters were set identically, with five acquisitions of an in-cell accumulation centered at *m/z* 1000. As can be observed, the resulting profiles are different in composition and relative abundance of the peptides present, with each type of tissue giving a unique fingerprint, characteristic of that specific organ. Several peptides were found in common in all three types of tissues, while some were uniquely present in only one particular organ. The resolving power (M/ΔM at FWHM) typical of these spectra ranged from 70,400 for the lower mass ions at *m/z* 599.510 to 30,300 for the higher mass ions at *m/z* 1474.702 due to a broadband acquisition covering a wide mass range from *m/z* 108-2000.

To further improve the sensitivity for detecting peptides present at much lower concentrations and in smaller tissue pieces, we adapted the ICA technique wherein ions were accumulated over multiple laser shots in the FT ICR cell prior to detection. As depicted in **Figure 3.1 A**, for ICA experiments, instead of the

front trapping plate being dropped to ground potential as is typically done, it was only dropped from the gated trapping potential of 20 V to 4 V so that ions trapped in previous MALDI events could not escape. With the MALDI pulse valve open (pressure of ~$1x10^{-6}$ Torr), the ions were collisionally cooled to the center of the cell prior to the next gating event. This method allows accumulation of ion signals without accumulating noise so that the signal-to-noise (S/N) ratio is much improved provided that the space-charge limit of the total number of ions in the ICR cell is not exceeded. Due to the limitation of IonSpec software no more than 10 events are allowed in each pulse sequence. We devised a "cell-loading" sequence modified from a previously published stitched sequence[34] (identical to the one shown in **Figure 3.1 A** with the addition of a dummy excitation/detection event). This was run multiple times depending on the number of laser shots desired followed by loading the excitation/detection pulse sequence to acquire the data.

Figure 3.5 compares spectra obtained from a small piece of PO tissue using a single laser shot versus 18 laser shots. Significant improvement of S/N is achieved with the 18-shot acquisition for ICA from the same sample. The multi-shot experiment allows identification of more neuropeptides and detection of additional isotopic clusters. While 194 peaks were observed in the single-shot experiment, with five of these peaks being known neuropeptides, the 18-shot ICA acquisition revealed 356 isotopic peaks with eight previously known peptides (labeled with red dots in the figure).

ICA is a very powerful tool in enhancing analyte signal. It works very well in direct tissue analysis where the endogenous peptides are in limited concentration. This technique has many advantages over signal averaging methods. These benefits include the fact that the amount of ions (and analyte signal) is increased while the electronic noise is not. This gain is furthered by a reduction in acquisition time for the ICA method (since it is only one "scan") over the signal averaged spectra (comprised of multiple "scans"). In the signal averaged spectra, electronic noise is present at the same frequency over the time period of the multiple scans and is therefore additive. Additionally, ICA increases the S/N linearly while signal averaging increase the S/N as a function of the square root of the number of scans. Finally, signal averaging can lead to peak broadening and/or peak splitting due to space-charge induced frequency shifts that occur from acquisition to acquisition.

Using the ICA technique, improvement in signal is great enough that MS/MS experiments can be performed on several relatively abundant peptides directly from the tissue sample. **Figure 3.6** shows an example of *de novo* sequencing of *m/z* 1030 and 1104 from a PO tissue sample. In **Figure 3.6 A**, a mass spectrum of multiple peptides present in the PO is shown. Using a stored waveform inverse Fourier transform (SWIFT) waveform isolation[59], the ion of interest at *m/z* 1030 was isolated (shown in the left inset of **Figure 3.6 A**). Note the detection of a doubly and triply charged 1030 peptide at *m/z* 515.233 and *m/z* 343.893. This phenomenon can also be seen in the isolation of 1104 (right inset of **Figure 3.6 A**). The SORI-CID fragmentation spectrum of the parent ion at *m/z*

1030 is shown in the middle panel, **Figure 3.6 B**. Extensive fragmentation is produced to allow the derivation of a sequence tag of FYSQRYamide, which matched one of the recently sequenced novel RYamides in the PO from *C. borealis*, pEGFYSQRYamide[15]. A short sequence tag was also acquired for a recently characterized neuropeptide covering six of the nine amino acids in GAHKNYLRFamide[67], **Figure 3.6 C**. This was accomplished by signal averaging ten ICA pulse sequences. This peptide is present at approximately ten percent of the abundance of 1030.

In most cases, neuropeptide identification by MS/MS requires pooling many tissues followed by extraction and fractionation to gain enough material to reveal even a partial sequence. The fact that the signal can be enhanced enough through ICA to allow MS/MS directly from neuronal tissue samples is significant. Direct tissue MS/MS can allow a sequence tag to be determined for database searching, if a database is available. Also, the sequence tag, in combination with high accuracy mass measurement, can improve confidence in the assignment of a known peptide.

Through the use of ICA the sensitivity and detection limit of FTMS can be comparable to that of TOF. Developments in high pressure MALDI have addressed the problems such as metastable fragmentation that plague vacuum MALDI and limit sensitivity. Using external accumulation in a hexapole and high pressure MALDI FTMS a detection limit of 300 zeptomoles has been reported for phosphopeptide RRREEE(pS)EEEAA[60]. External trapping allows shorter analysis time as ions can be collisionally cooled and fragmented in the hexapole

and the ICR cell can be maintained at low vacuum though out the experiment eliminating the time required for the ICR to be pumped down. However, matrix and salt clusters are also accumulated in the hexapole and can lead to space-charge interactions. In ICA these matrix and salt clusters can be selectively ejected from the ICR cell before detection. ICA also allows MS^n to be easily performed.

While ICA can enhance ion signals dramatically, there is a possibility of MMA degradation due to reaching the space-charge limit of the ICR cell. To avoid this, space-charge induced frequency shifts can be corrected by means of external calibration[61; 62]. Alternatively, internal calibration can be used to eliminate space-charge induced frequency shifts between calibrant and analyte spectra. In most types of mass spectrometry, internal calibration is commonly performed by premixing known calibration standards with the analyte prior to mass spectral analysis. With ICA and the decoupled nature of ionization and detection processes in FTMS, it is possible to introduce calibrants from a separate sample spot, followed by the accumulation of analyte ions in the ICR cell prior to MS detection. Because the trapped ions are simultaneously analyzed, any space-charge induced shifts or other drifts in the mass calibration are effectively counteracted.

As outlined in **Figure 3.2**, two separate pulse sequences are designed. The first, usually a single laser pulse, does not include the voltage ramp to zero for the cell trapping plates, allowing the ions to remain in the cell. The second pulse sequence does not include the initial ICR cell quench to remove the ions,

but is an ICA pulse sequence in every other respect. **Figure 3.7** demonstrates the utility of such an internal calibration on adjacent samples (InCAS) method for neuropeptide mapping in the STG neuropil. A mixture of three peptide standards each with concentrations of 10^{-6} M, including FMRFamide (*m/z* 599.31280), substance P (*m/z* 1347.73597) and somatostatin (*m/z* 1637.72448), was deposited on a separate spot from that containing the STG neuropil. Substance P and somatostatin were used for calibration of the direct tissue spectrum (both peaks were denoted with stars in **Figure 3.7 A**). The standards were accumulated in the cell first with one laser shot followed by five acquisitions of seven shots each from the tissue mounted on a separate sample spot. **Figure 3.7 A** reveals a highly complex spectrum obtained from the STG neuropil, with two calibrant ions labeled with stars. The two *m/z* expansion regions (**Figures 3.7 B** and **3.7 C**) highlight the detection of several known peptides (labeled with red dots) used to evaluate the MMA. As summarized in **Table 1**, an average mass measurement error was calculated for six known neuropeptides found in the direct tissue spectrum. This method improved the MMA from an average of 27.9 ppm to 5.7 ppm for calibration using only substance P. Using both substance P and somatostatin improved the average error to 2.04 ppm. A notable MMA of 0.1498 ppm was obtained for *Cancer borealis* tachykinin-related peptide 1a (Cab TRP1a). These results were comparable to that observed with one acquisition of the ICA, with a two standard calibration using substance P and FMRFamide as calibrants. The average errors

for the uncalibrated spectra were 24 ppm and 1.4 ppm for the spectra calibrated with two calibrants, respectively (**Table 1**).

This method has the unique advantage of adding calibrants to a sample without premixing, thus also allowing the flexibility to adjust the concentration of calibrants in the cell by varying the number of laser shots prior to analyte acquisition and detection. This is particularly useful for tissue sample analysis as the tissue is heterogeneous and the peptide content can change from one sample to another making the premixing of calibrants (in the right concentration) with the tissue very difficult. This indirect addition allows the calibrants to sample signal ratio to be adjusted through the number of desorption/ionization events acquired to eliminate analyte suppression by the calibrants. As a result, very little has to be known about the peptide profiles before an experiment begins. For the best MMA the standards should not overlap any of the analyte peaks, but be in the same mass range. Even if no information is available on the sample, several calibrants or calibrant mixtures can be spotted prior to the experiment and the best can be used for the resulting peptide profile. With MMA in the range of 2 ppm, previously characterized neuropeptides can be unambiguously identified even in a highly complex mixture of peptides as is typical for neuronal tissue samples.

For profiling a complex mixture of peptides spanning a wide mass range in tissue samples, MALDI FTMS is limited in its *m/z* window by the time-of-flight effect[63-65]. This effect is particularly prominent in external source instruments, as a time-of-flight mass separation of ions occurs during the transfer of ions into the

magnetic field region which can result in mass discrimination in the collection of ions in the analyzer. Consequently, only ions over a narrow mass range can be captured. Several mass spectra must be acquired at different trapping times optimized for a specific *m/z* range to compensate for this distortion to obtain a complete profile of peptides present in a wide mass region. A number of solutions to this problem have been proposed including in-cell MALDI[35; 65; 66] and shifted accumulation[34; 63; 64].

Using the flexibility of the ICA pulse sequence, an integrated pulse sequence can be constructed to accumulate a varying number of MALDI events optimized for a specific *m/z* range. By centering the quadrupole RF frequency and amplitude and adjusting the time-of-flight duration set by the width of the quadrupole trapping plate for different masses during each desorption/ionization event, a full mass scan was developed to quickly survey the peptides present in a sample. For the study of neuropeptides in crustaceans, two events were optimized for *m/z* 1000, one for *m/z* 2500, two for *m/z* 5000, and two for *m/z* 8000. Only one RF pulse was optimized for *m/z* 2500 because in this range ions are also able to be transmitted in to the ICR cell from the events optimized for *m/z* 1000 and 5000. Also, ions in this range are more efficiently transported into the ICR cell. This pulse sequence allows a mass range from *m/z* 215 to 9,000 to be observed in a single acquisition. This can be seen in a full mass scan of a sinus gland micro-dissected from *C. borealis*, **Figure 3.8**. High mass peaks are readily observable at *m/z* 8513, middle mass peaks are detected with high intensity at *m/z* around 3979 (including a number of peaks accounted for loss of

neutrals), and examples of lower mass peaks include *m/z* 1474.74 (Ala^{13}-orcokinin) and 844.51. Several electrical noise peaks (single spikes with no isotopic distribution) are present in the spectra (labeled with stars). The resolving power of mass spectral peaks in the range of *m/z* 3800-4015 is between 10,200 and 42,400, with the most intense peak (*m/z* 3979.632) displaying a resolving power of 22,900. Remarkably, a resolving power of 10,200 was found for the high mass peak at *m/z* 8513.

The full mass scan is very useful for an initial survey of the peptides found in a sample. Although the ultra-high resolution and therefore high mass accuracy are sacrificed to some extent when observing such a wide mass range, it provides an effective means for high throughput screening. This is very useful when analyzing a large number of HPLC fractions from a tissue extract, or analyzing tissue directly for neuropeptides prior to more in-depth and directed experiments such as MS/MS experiments.

3.5 Conclusions

We have demonstrated in this study the first application of MALDI FTMS for neuropeptide identification and characterization directly from complex neuronal tissue samples. Evidently, analytical performance is not compromised due to the unconventional samples. The ability to store ions and increase their concentration in the cell prior to detection through ICA offers a great deal of

flexibility and versatility for *in situ* neuropeptide analysis. Several analytical figures-of-merit provided by ICA for direct tissue analysis were evaluated. The enhanced sensitivity for trace-level analytes due to the improved signal-to-noise ratio obtained by multiple desorption/ionization events results in greatly improved spectral quality and allows the detection of novel peptides. MS/MS experiments directly from tissue samples have become feasible for lower abundance ions, allowing significant sequence information to be obtained. Additionally, inclusion of external standards as internal calibrants without premixing significantly improves MMA to allow high confidence peptide identification from extremely complex mixtures. Finally, ICA offers the flexibility to construct integrated pulse sequences for MS detection covering a wide mass range. This permits high throughput initial peptide screens for more in-depth and targeted experiments. Collectively, ICA with MALDI FTMS provides a powerful alternative for neuropeptide analysis in complex tissue samples where high mass resolving power, high mass accuracy and high sensitivity are simultaneously required.

3.6 References

1. Taghert, P. H., and Veenstra, J. A. (2003) Drosophila neuropeptide signaling. *Adv Genet*49, 1-65

2. Skiebe, P. (2001) Neuropeptides are ubiquitous chemical mediators: Using the stomatogastric nervous system as a model system. *J Exp Biol*204, 2035-2048

3. Hokfelt, T., Broberger, C., Xu, Z. Q., Sergeyev, V., Ubink, R., and Diez, M. (2000) Neuropeptides--an overview. *Neuropharmacology*39, 1337-1356

4. Keller, R. (1992) Crustacean neuropeptides: Structures, functions and comparative aspects. *Experientia*48, 439-448

5. Kobierski, L. A., Beltz, B. S., Trimmer, B. A., and Kravitz, E. A. (1987) FMRFamidelike peptides of homarus americanus: Distribution, immunocytochemical mapping, and ultrastructural localization in terminal varicosities. *J Comp Neurol*266, 1-15

6. Spittaels, K., Schoofs, L., Grauwels, L., Smet, H., Van Damme, J., Proost, P., and De Loof, A. (1991) Isolation, identification and synthesis of novel oviductal motility stimulating head peptide in the colorado potato beetle, leptinotarsa decemlineata. *Peptides*12, 31-36

7. Christie, A. E., Lundquist, C. T., Nassel, D. R., and Nusbaum, M. P. (1997) Two novel tachykinin-related peptides from the nervous system of the crab cancer borealis. *J Exp Biol*200 (Pt 17), 2279-2294

8. Nilsson, C. L., Karlsson, G., Bergquist, J., Westman, A., and Ekman, R. (1998) Mass spectrometry of peptides in neuroscience. *Peptides*19, 781-789

9. Desiderio, D. M., and Zhu, X. (1998) Quantitative analysis of methionine enkephalin and beta-endorphin in the pituitary by liquid secondary ion mass spectrometry and tandem mass spectrometry. *J Chromatogr A*794, 85-96

10. Baggerman, G., Verleyen, P., Clynen, E., Huybrechts, J., De Loof, A., and Schoofs, L. (2004) Peptidomics. *Journal of Chromatography B*803, 3-16

11. Andren, P. E., and Caprioli, R. M. (1999) Determination of extracellular release of neurotensin in discrete rat brain regions utilizing in vivo microdialysis/electrospray mass spectrometry. *Brain Research*845, 123-129

12. Kennedy, R. T., Watson, C. J., Haskins, W. E., Powell, D. H., and Strecker, R. E. (2002) In vivo neurochemical monitoring by microdialysis and capillary separations. *Curr Opin Chem Biol*6, 659-665

13. Li, L., Moroz, T. P., Garden, R. W., Floyd, P. D., Weiss, K. R., and Sweedler, J. V. (1998) Mass spectrometric survey of interganglionically transported peptides in aplysia. *Peptides*19, 1425-1433

14. Li, L., Pulver, S. R., Kelley, W. P., Thirumalai, V., Sweedler, J. V., and Marder, E. (2002) Orcokinin peptides in developing and adult crustacean stomatogastric nervous systems and pericardial organs. *J Comp Neurol*444, 227-244

15. Li, L., Kelley, W. P., Billimoria, C. P., Christie, A. E., Pulver, S. R., Sweedler, J. V., and Marder, E. (2003) Mass spectrometric investigation of the neuropeptide complement and release in the pericardial organs of the crab, cancer borealis. *J Neurochem*87, 642-656

16. Wang, F., Li, W., Emmett, M. R., Marshall, A. G., Corson, D., and Sykes, B. D. (1999) Fourier transform ion cyclotron resonance mass spectrometric detection of small ca(2+)-induced conformational changes in the regulatory domain of human cardiac troponin C. *J Am Soc Mass Spectrom*10, 703-710

17. Jimenez, C. R., van Veelen, P. A., Li, K. W., Wildering, W. C., Geraerts, W. P., Tjaden, U. R., and van der Greef, J. (1994) Neuropeptide expression and processing as revealed by direct matrix-assisted laser desorption ionization mass spectrometry of single neurons. *J Neurochem*62, 404-407

18. Predel, R., Linde, D., Rapus, J., Vettermann, S., and Penzlin, H. (1995) Periviscerokinin (pea-PVK): A novel myotropic neuropeptide from the perisympathetic organs of the american cockroach. *Peptides*16, 61-66

19. Caprioli, R. M., Farmer, T. B., and Gile, J. (1997) Molecular imaging of biological samples: Localization of peptides and proteins using MALDI-TOF MS. *Anal Chem*69, 4751-4760

20. Chaurand, P., Stoeckli, M., and Caprioli, R. M. (1999) Direct profiling of proteins in biological tissue sections by MALDI mass spectrometry. *Anal Chem*71, 5263-5270

21. Li, L., Garden, R. W., and Sweedler, J. V. (2000) Single-cell MALDI: A new tool for direct peptide profiling. *Trends Biotechnol*18, 151-160

22. Garden, R. W., Shippy, S. A., Li, L., Moroz, T. P., and Sweedler, J. V. (1998) Proteolytic processing of the aplysia egg-laying hormone prohormone. *Proc Natl Acad Sci U S A*95, 3972-3977

23. Li, L., Romanova, E. V., Rubakhin, S. S., Alexeeva, V., Weiss, K. R., Vilim, F. S., and Sweedler, J. V. (2000) Peptide profiling of cells with multiple gene products: Combining immunochemistry and MALDI mass spectrometry with on-plate microextraction. *Anal Chem*72, 3867-3874

24. Jimenez, C. R., Li, K. W., Dreisewerd, K., Spijker, S., Kingston, R., Bateman, R. H., Burlingame, A. L., Smit, A. B., van Minnen, J., and Geraerts, W. P. (1998) Direct mass spectrometric peptide profiling and sequencing of single neurons reveals differential peptide patterns in a small neuronal network. *Biochemistry*37, 2070-2076

25. Rubakhin, S. S., Garden, R. W., Fuller, R. R., and Sweedler, J. V. (2000) Measuring the peptides in individual organelles with mass spectrometry. *Nat Biotechnol*18, 172-175

26. Qin, J., and Chait, B. T. (1996) Matrix-assisted laser desorption ion trap mass spectrometry: Efficient isolation and effective fragmentation of peptide ions. *Anal Chem*68, 2108-2112

27. Qin, J., Fenyo, D., Zhao, Y., Hall, W. W., Chao, D. M., Wilson, C. J., Young, R. A., and Chait, B. T. (1997) A strategy for rapid, high-confidence protein identification. *Anal Chem*69, 3995-4001

28. Yew, J. Y., Dikler, S., and Stretton, A. O. (2003) De novo sequencing of novel neuropeptides directly from ascaris suum tissue using matrix-assisted laser desorption/ionization time-of-flight/time-of-flight. *Rapid Commun Mass Spectrom*17, 2693-2698

29. Gross, M. L., and Rempel, D. L. (1984) Fourier transform mass spectrometry. *Science*226, 261-268

30. Li, Y., McIver, R. T.,Jr, and Hunter, R. L. (1994) High-accuracy molecular mass determination for peptides and proteins by fourier transform mass spectrometry. *Anal Chem*66, 2077-2083

31. Marshall, A. G., Hendrickson, C. L., and Jackson, G. S. (1998) Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrom Rev*17, 1-35

32. McLafferty, F. W., Kelleher, N. L., Begley, T. P., Fridriksson, E. K., Zubarev, R. A., and Horn, D. M. (1998) Two-dimensional mass spectrometry of biomolecules at the subfemtomole level. *Curr Opin Chem Biol*2, 571-578

33. Castoro, J. A. W.,C.L. (1994) High resolution laser desorption/ionization fourier transform mass spectrometry. *Trends Anal Chem*13, 229-233

34. O'Connor, P. B., and Costello, C. E. (2000) Application of multishot acquisition in fourier transform mass spectrometry. *Anal Chem*72, 5125-5130

35. Mize, T. H., and Amster, I. J. (2000) Broad-band ion accumulation with an internal source MALDI-FTICR-MS. *Anal Chem*72, 5886-5891

36. Baykut, G., Jertz, R., and Witt, M. (2000) Matrix-assisted laser desorption/ionization fourier transform ion cyclotron resonance mass spectrometry with pulsed in-source collision gas and in-source ion accumulation. *Rapid Commun Mass Spectrom*14, 1238-1247

37. O'Connor, P. B., and Costello, C. E. (2000) Internal calibration on adjacent samples (InCAS) with fourier transform mass spectrometry. *Anal Chem*72, 5881-5885

38. Ge, Y., Lawhorn, B. G., ElNaggar, M., Strauss, E., Park, J. H., Begley, T. P., and McLafferty, F. W. (2002) Top down characterization of larger proteins (45 kDa) by electron capture dissociation mass spectrometry. *J Am Chem Soc*124, 672-678

39. Johnson, J. R., Meng, F., Forbes, A. J., Cargile, B. J., and Kelleher, N. L. (2002) Fourier-transform mass spectrometry for automated fragmentation and identification of 5-20 kDa proteins in mixtures. *Electrophoresis*23, 3217-3223

40. Smith, R. D., Pasa-Tolic, L., Lipton, M. S., Jensen, P. K., Anderson, G. A., Shen, Y., Conrads, T. P., Udseth, H. R., Harkewicz, R., Belov, M. E., Masselon,

C., and Veenstra, T. D. (2001) Rapid quantitative measurements of proteomes by fourier transform ion cyclotron resonance mass spectrometry. *Electrophoresis*22, 1652-1668

41. Shen, Y., Tolic, N., Zhao, R., Pasa-Tolic, L., Li, L., Berger, S. J., Harkewicz, R., Anderson, G. A., Belov, M. E., and Smith, R. D. (2001) High-throughput proteomics using high-efficiency multiple-capillary liquid chromatography with on-line high-performance ESI FTICR mass spectrometry. *Anal Chem*73, 3011-3021

42. Smith, R. D., Anderson, G. A., Lipton, M. S., Masselon, C., Pasa-Tolic, L., Shen, Y., and Udseth, H. R. (2002) The use of accurate mass tags for high-throughput microbial proteomics. *Omics*6, 61-90

43. Bergquist, J., Palmblad, M., Wetterhall, M., Hakansson, P., and Markides, K. E. (2002) Peptide mapping of proteins in human body fluids using electrospray ionization fourier transform ion cyclotron resonance mass spectrometry. *Mass Spectrom Rev*21, 2-15

44. Meng, F., Cargile, B. J., Patrie, S. M., Johnson, J. R., McLoughlin, S. M., and Kelleher, N. L. (2002) Processing complex mixtures of intact proteins for direct analysis by mass spectrometry. *Anal Chem*74, 2923-2929

45. Pasa-Tolic, L., Lipton, M. S., Masselon, C. D., Anderson, G. A., Shen, Y., Tolic, N., and Smith, R. D. (2002) Gene expression profiling using advanced mass spectrometric approaches. *J Mass Spectrom*37, 1185-1198

46. Chalmers, M. J., Hakansson, K., Johnson, R., Smith, R., Shen, J., Emmett, M. R., and Marshall, A. G. (2004) Protein kinase A phosphorylation characterized by tandem fourier transform ion cyclotron resonance mass spectrometry. *Proteomics*4, 970-981

47. Li, L., Masselon, C. D., Anderson, G. A., Pasa-Tolic, L., Lee, S. W., Shen, Y., Zhao, R., Lipton, M. S., Conrads, T. P., Tolic, N., and Smith, R. D. (2001) High-throughput peptide identification from protein digests using data-dependent multiplexed tandem FTICR mass spectrometry coupled with capillary liquid chromatography. *Anal Chem*73, 3312-3322

48. Jones, J. J., Stump, M. J., Fleming, R. C., Lay, J. O.,Jr, and Wilkins, C. L. (2003) Investigation of MALDI-TOF and FT-MS techniques for analysis of escherichia coli whole cells. *Anal Chem*75, 1340-1347

49. Gauthier, J. W. T.,T.R., and Jacobson, D. B. (1991) Sustained off-resonance irradiation for collision-activated dissociation involving fourier transform mass spectrometry. collision-activated dissociation technique that emulates infrared multiphoton dissociation. *Anal Chimica Acta*246, 211-225

50. Li, L., Romanova, E. V., Rubakhin, S. S., Alexeeva, V., Weiss, K. R., Vilim, F. S., and Sweedler, J. V. (2000) Peptide profiling of cells with multiple gene products: Combining immunochemistry and MALDI mass spectrometry with on-plate microextraction. *Anal Chem*72, 3867-3874

51. Li, L., Garden, R. W., Romanova, E. V., and Sweedler, J. V. (1999) In situ sequencing of peptides from biological tissues and single cells using MALDI-PSD/CID analysis. *Anal Chem*71, 5451-5458

52. Li, L., Garden, R. W., Floyd, P. D., Moroz, T. P., Gleeson, J. M., Sweedler, J. V., Pasa-Tolic, L., and Smith, R. D. (1999) Egg-laying hormone peptides in the aplysiidae family. *J Exp Biol*202 (Pt 21), 2961-2973

53. Li, L., Floyd, P. D., Rubakhin, S. S., Romanova, E. V., Jing, J., Alexeeva, V. Y., Dembrow, N. C., Weiss, K. R., Vilim, F. S., and Sweedler, J. V. (2001) Cerebrin prohormone processing, distribution and action in aplysia californica. *J Neurochem*77, 1569-1580

54. Li, L., Pulver, S. R., Kelley, W. P., Thirumalai, V., Sweedler, J. V., and Marder, E. (2002) Orcokinin peptides in developing and adult crustacean stomatogastric nervous systems and pericardial organs. *J Comp Neurol*444, 227-244

55. Skiebe, P., Dreger, M., Meseke, M., Evers, J. F., and Hucho, F. (2002) Identification of orcokinins in single neurons in the stomatogastric nervous system of the crayfish, cherax destructor. *J Comp Neurol*444, 245-259

56. Baggerman, G., Cerstiaens, A., De Loof, A., and Schoofs, L. (2002) Peptidomics of the larval drosophila melanogaster central nervous system. *J Biol Chem*277, 40368-40374

57. Verleyen, P., Huybrechts, J., Sas, F., Clynen, E., Baggerman, G., De Loof, A., and Schoofs, L. (2004) Neuropeptidomics of the grey flesh fly, neobellieria bullata. *Biochemical and Biophysical Research Communications*316, 763-770

58. Weimann, J. M., Marder, E., Evans, B., and Calabrese, R. L. (1993) The effects of SDRNFLRFamide and TNRNFLRFamide on the motor patterns of the stomatogastric ganglion of the crab cancer borealis. *J Exp Biol*181, 1-26

59. Wang, T. C., Ricca, T. L., and Marshall, A. G. (1986) Extension of dynamic range in fourier transform ion cyclotron resonance mass spectrometry via stored waveform inverse fourier transform excitation. *Anal Chem*58, 2935-2938

60. Moyer, S. C., Budnik, B. A., Pittman, J. L., Costello, C. E., and O'Connor, P. B. (2003) Attomole peptide analysis by high-pressure matrix-assisted laser desorption/ionization fourier transform mass spectrometry. *Anal Chem*75, 6449-6454

61. Bruce, J. E., Anderson, G. A., Brands, M. D., Pasa-Tolic, L., and Smith, R. D. (2000) Obtaining more accurate fourier transform ion cyclotron resonance mass measurements without internal standards using multiply charged ions. *J Am Soc Mass Spectrom*11, 416-421

62. Easterling, M. L. M.,T.H., and Amster, I. J. (1999) Routine part-per-million mass accuracy for high-mass ions: Space-charge effects in FT-ICR. *Anal Chem*71, 624-632

63. O'Connor, P. B., Duursma, M. C., van Rooij, G. J., Heeren, R. M. A., and Boon, J. J. (1997) Correction of time-of-flight shifted polymeric molecular weight distributions in matrix-assisted laser desorption/ionization fourier transform mass spectrometry. *Anal Chem*69, 2751-2755

64. Sze, E. T. C., T.W. (1999) Time-of-flight effects in matrix-assisted laser desorption/ionization fourier transform mass spectrometry. *Rapid Commun Mass Spectrom*13, 398-406

65. Dey, M. C.,J.A., and Wilkins, C. L. (1995) Determination of molecular weight distributions of polymers by MALDI-FTMS. *Anal Chem*67, 1575-1579

66. Castoro, J. A., Koster, C., and Wilkins, C. L. (1993) Investigation of a "screened" electrostatic ion trap for analysis of high mass molecules by Fourier transform mass spectrometry. *Anal Chem*65, 784-788

67. Cruz-Bermudez, N.D., Fu, Q., Kutz, K. K., Christie, A. E., Li, L.,and Marder, E. (2006) Mass spectrometric characterization and physiological actions of GAHKNYLRFamide, a novel FMRFamide-like peptide from crabs of the genus Cancer. J Neurochem.97(3):784-99.

3.7 Figures

Five Acquisitions							
Peptide	**[M+H]+ Calc'd**	**[M+H]+ Uncal.**	**Ppm**	**[M+H]+ Cal.(1 Std)**	**ppm**	**[M+H]+ Cal. (2 Std (P+S))**	**ppm**
Cab TRP1a APSGFLGMRa	934.4933	934.5128	20.9204	934.4815	12.6593	934.4931	0.1498
NRNFLRFa	965.5433	965.5642	21.6251	965.5318	11.9104	965.5430	0.3314
orcokinin[1-12] NFDEIDRSGFGF	1403.6232	1403.6664	30.7989	1403.6203	2.0162	1403.6180	3.7118
[Ala13]-orcokinin NFDEIDRSGFGFA	1474.6603	1474.7096	33.4314	1474.6614	0.7324	1474.6557	3.1126
[Val13]-orcokinin NFDEIDRSGFGFV	1502.6916	1502.7421	33.6263	1502.6930	0.9649	1502.6860	3.7266
FDAFTTGFGHS	1186.5169	1186.5490	27.0793	1186.5097	6.0935	1186.5155	1.2136
Average Error			27.9136		5.7294		2.0410
One Acquisition							
Peptide	**[M+H]+ Calc'd**	**[M+H]+ Uncal.**	**Ppm**	**[M+H]+ Cal. (1 Std)**	**ppm**	**[M+H]+ Cal (2 Std(F+P))**	**ppm**
Cab TRP1a APSGFLGMRa	934.4933	934.5094	17.2179	934.4852	8.6036	934.4930	0.3531
NRNFLRFa	965.5433	965.5611	18.3834	965.5362	7.3844	965.5436	0.2589
orcokinin[1-12] NFDEIDRSGFGF	1403.6232	1403.6585	25.1991	1403.6234	0.1425	1403.6218	0.9760
[Ala13]-orcokinin NFDEIDRSGFGFA	1474.6603	1474.7060	30.9766	1474.6692	6.0353	1474.6654	3.4991
[Val13]-orcokinin NFDEIDRSGFGFV	1502.6916	1502.7368	30.0927	1502.6994	5.1973	1502.6947	2.1029
FDAFTTGFGHS	1186.5169	1186.5449	23.5648	1186.5147	1.8457	1186.5185	1.3738
Average Error			24.2391		4.8681		1.4273

Table 3.1: Mass Measurement Accuracy for Peptide Identification in *C. borealis* STG Neuropil using InCAS Method

1 Std Calibrant is Substance P

P = Substance P

S = Somatostatin

F = FMRFamide

35 shots for five acquisitions and 7 shots for one acquisition

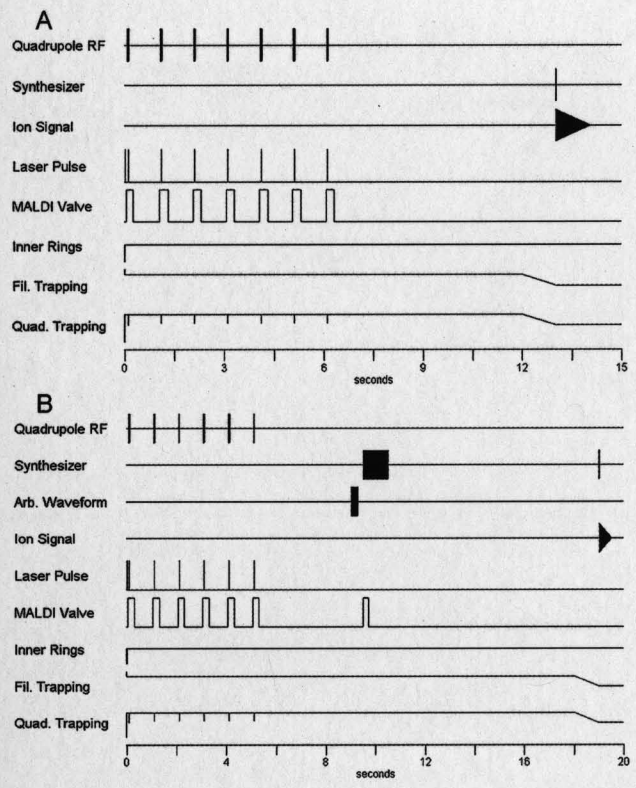

Figure 3.1: Examples of pulse sequence used for (**A**) multi-shot in-cell accumulation (ICA) experiment and (**B**) MS/MS (SORI-CID) experiment in conjunction with ICA.

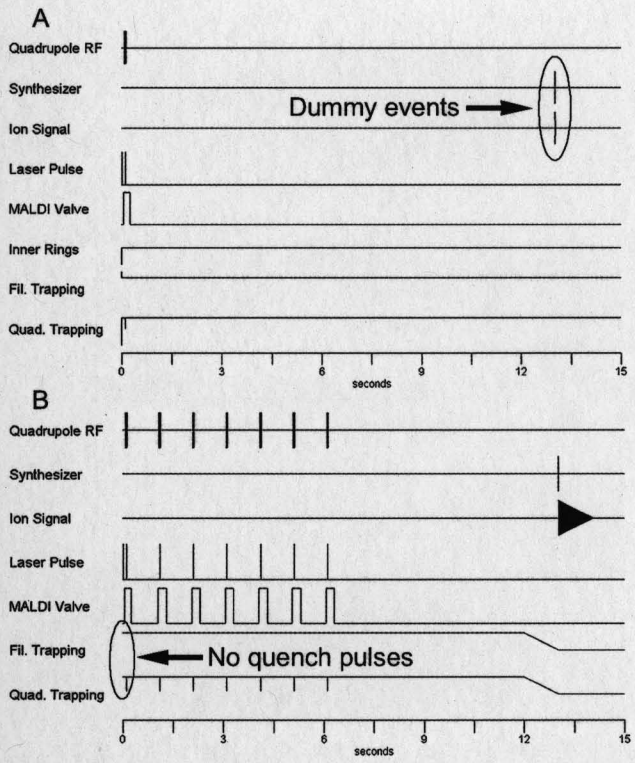

Figure 3.2: Schematic of pulse sequences for InCAS experiment. (**A**) Single-shot calibrant ion loading pulse sequence. Note the inclusion of "dummy" excitation and detection events. (**B**) Multi-shot analyte ion loading sequence, note the absence of initial quench pulses on trapping plates.

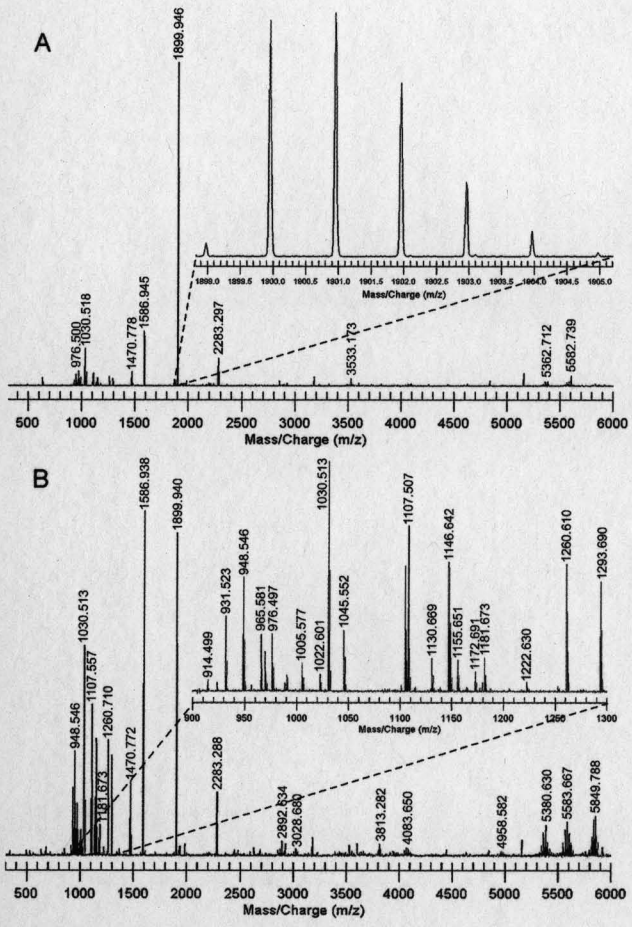

Figure 3.3: Effects of sample preparation on MALDI FTMS of pericardial organs (PO) from *Cancer borealis*. (**A**) PO section rinsed with DHB only. (**B**) PO section rinsed with acidified methanol followed by DHB rinsing. Insets show corresponding *m/z* expansion region.

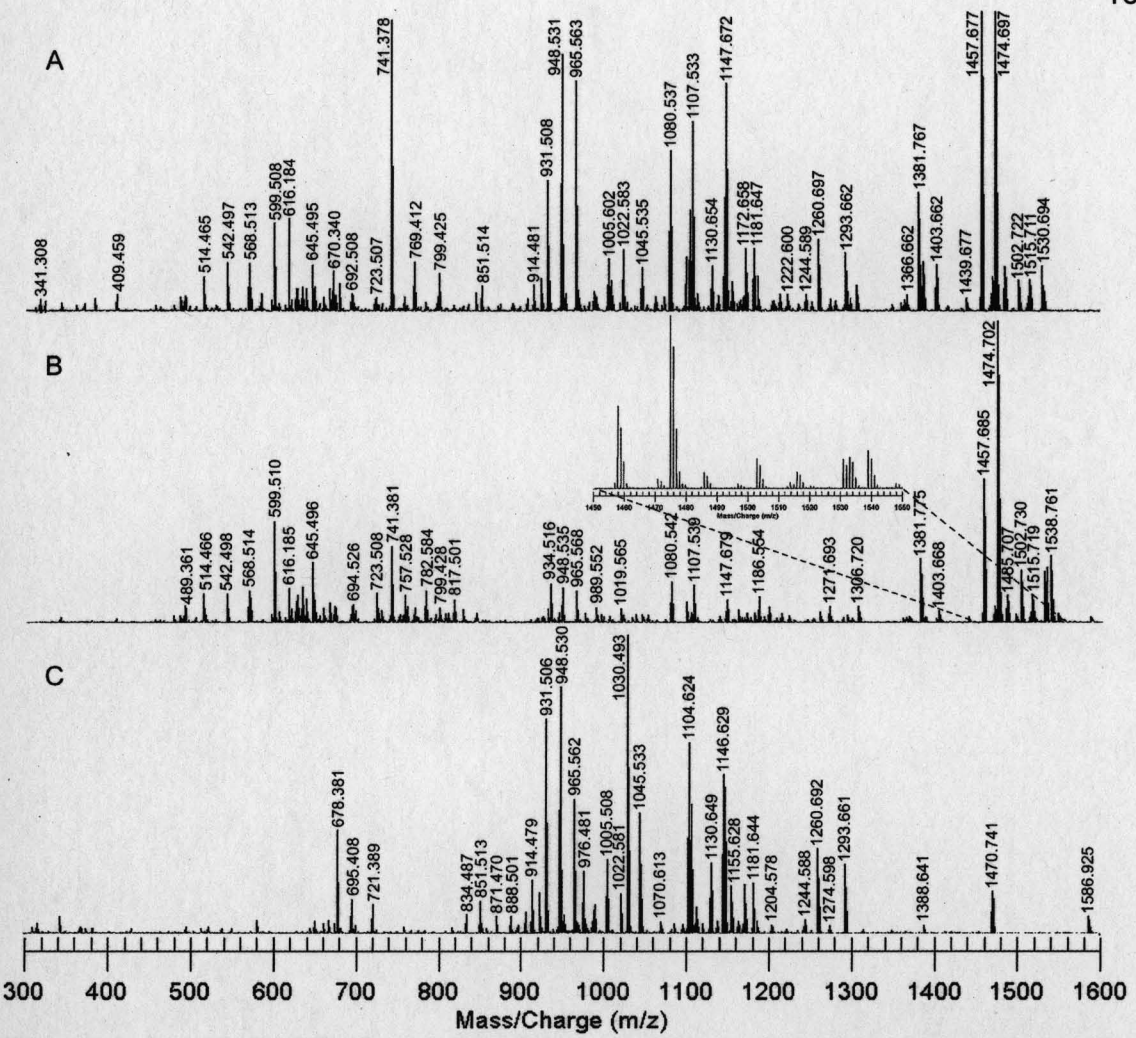

Figure 3.4: MALDI FTMS spectra of neuronal tissue samples isolated from (**A**) the stomatogastric ganglion (STG) neuropil, (**B**) the commissural ganglion (CoG) and (**C**) the pericardial organ (PO) in *C. borealis*. All spectra were acquired using ICA optimized for *m/z* 1000.

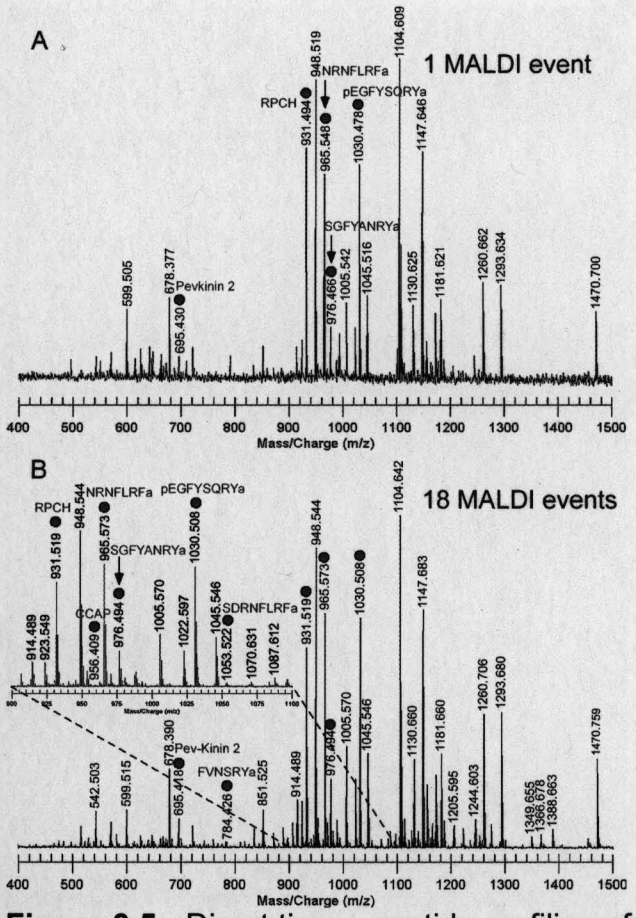

Figure 3.5: Direct tissue peptide profiling of *C. borealis* pericardial organs by MALDI FTMS showing the S/N improvement by multi-shot acquisition for ICA. The peaks labeled with red dots are identified peptides. (**A**) Single-shot acquisition, with the detection of 194 isotopic peaks and 5 previously known peptides. (**B**) 18-shot acquisition, with the detection of 356 peaks and 8 previously known peptides. CCAP is crustacean cardioactive peptide. RPCH is red pigment concentrating hormone.

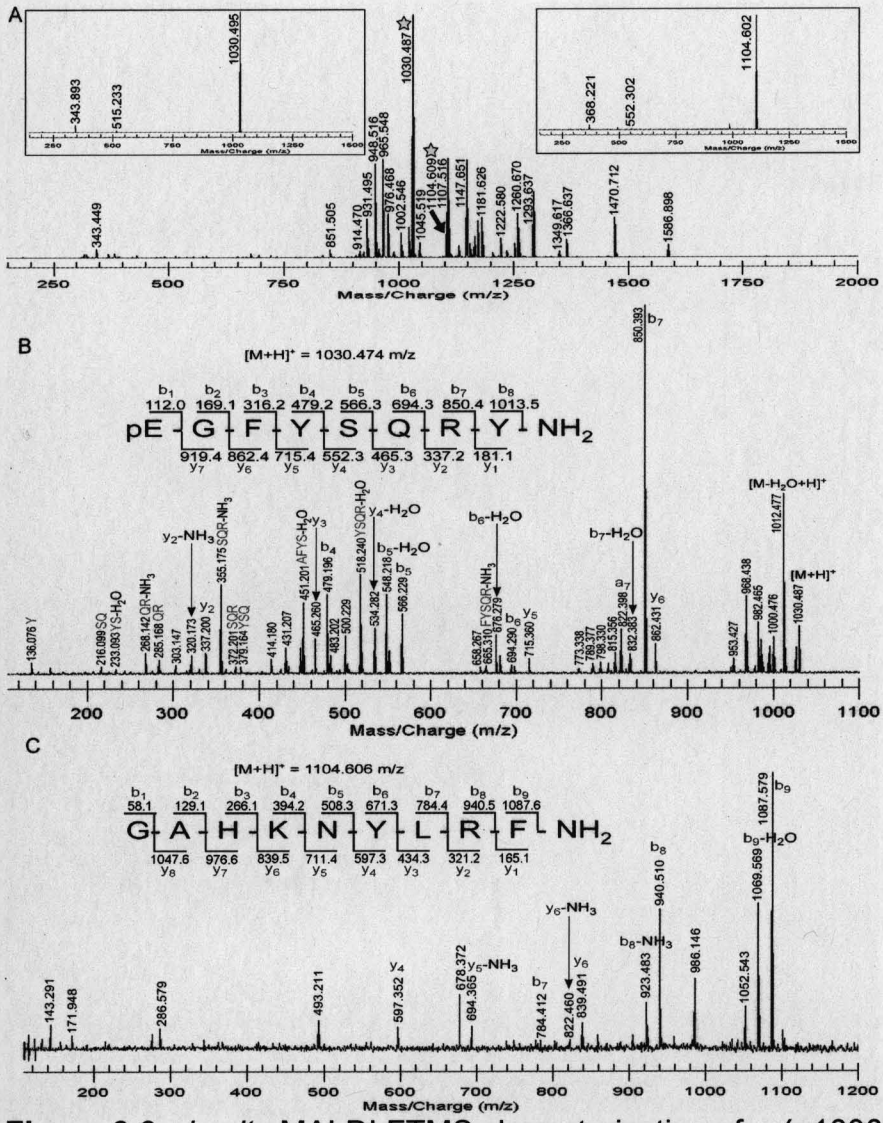

Figure 3.6: *In situ* MALDI FTMS characterization of *m/z* 1030.5 and 1104.6 from the PO in *C. borealis*. (**A**) MALDI FTMS peptide profile of PO, with 1030 and 1104 peaks labeled with a star. (Insets) Isolation of 1030 and 1104 peptides using SWIFT. (**B**) SORI-CID fragmentation spectrum of $[M+H]^+$ at *m/z* 1030.5, with b/y ions and internal fragment ions labeled. The derived amino acid sequence is shown above the spectrum. (**C**) SORI-CID fragmentation spectrum of $[M+H]^+$ at *m/z* 1104.6, with b/y ions labeled. This peak is present at ten percent the abundance of *m/z* 1030.

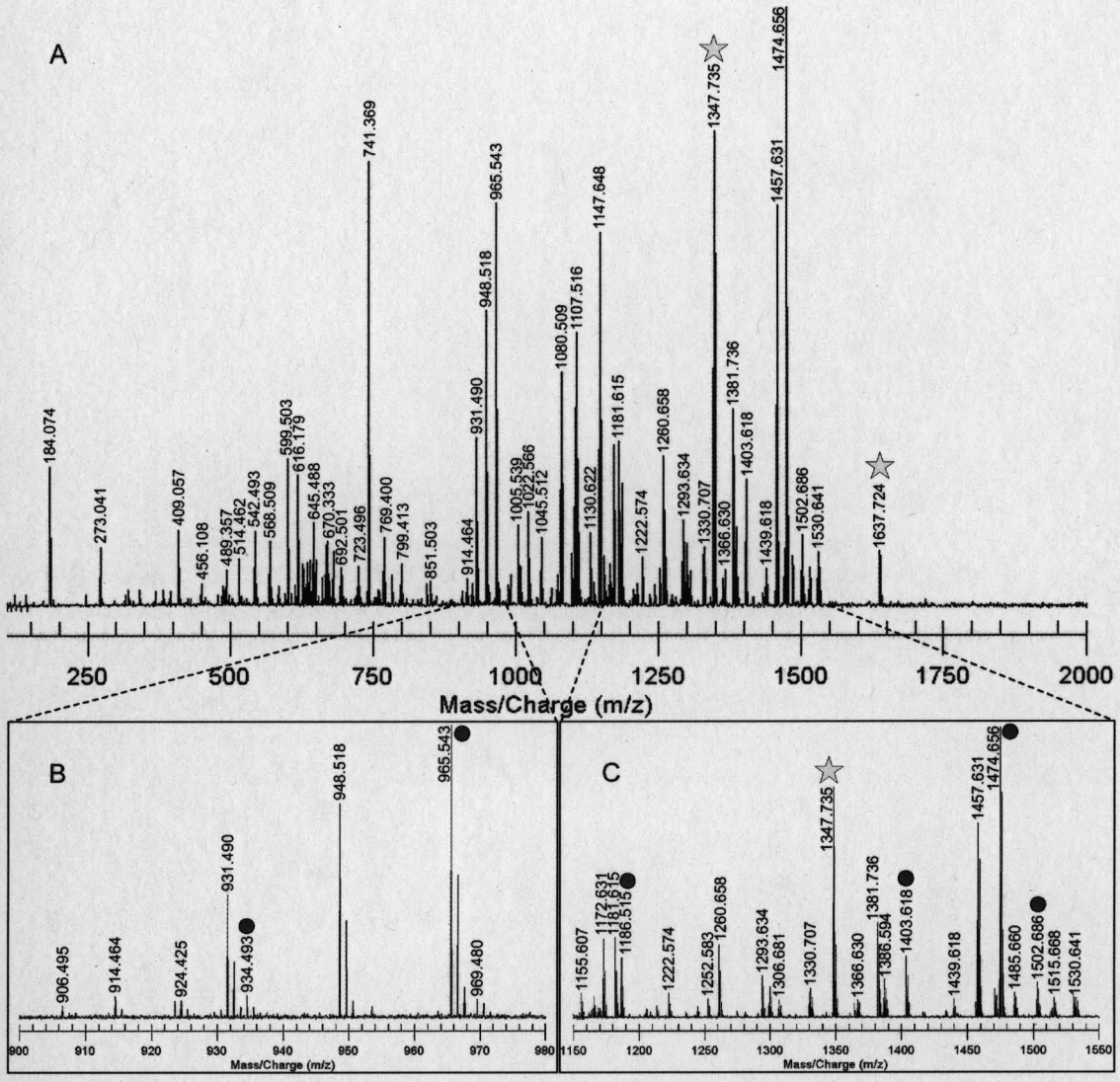

Figure 3.7: InCAS of the STG neuropil with substance P and somatostatin added as calibrants. (**A**) *In situ* MALDI FTMS spectrum of the STG neuropil, with calibrant peaks labeled with stars. Expansions of *m/z* 900-980 (**B**) and *m/z* 1150-1550 (**C**) show the known peptides used for mass measurement accuracy calculations. Known peptides are highlighted with red dots.

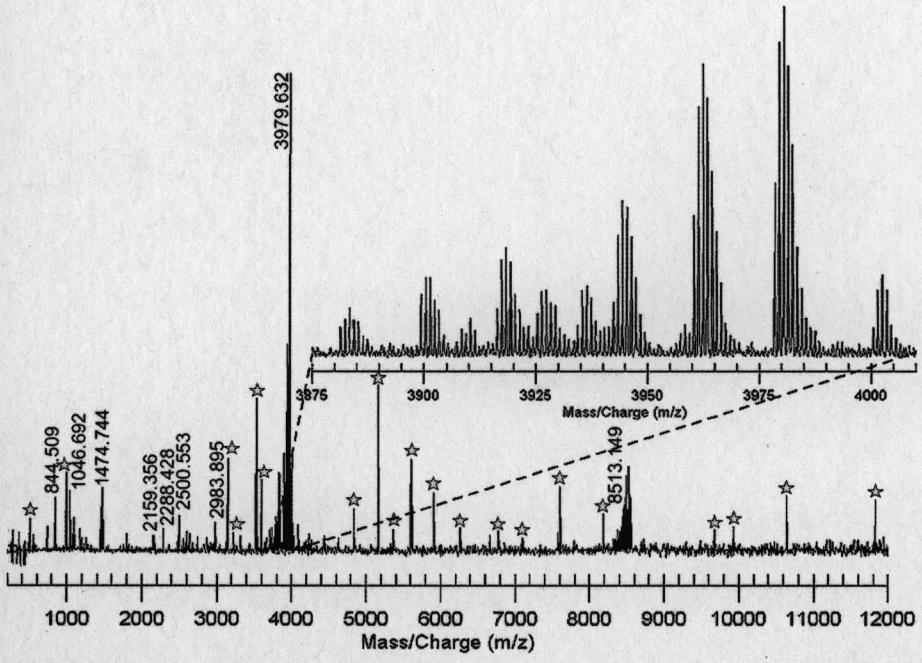

Figure 3.8: *In situ* MALDI FTMS spectrum of sinus gland from *C. borealis* using a full mass scan covering *m/z* 215-12000. Inset shows expansion of *m/z* 3875-4000 with a resolving power of 22,900 for $[M+H]^+$ ion at *m/z* 3979.632. Noise peaks are noted with stars.

Chapter Four

Bioinformatics Software Development

This work was performed in collaboration with Sean McIlwain and David Page of the UW-Madison Department of Biostatistics. Sean in particular spent many hours with me explaining how various bioinformatics processes worked as well as rewriting the bioinformatics program we developed every time I had an idea for some change that should be made. Also, Lingjun Li has been key in performing this work as she has helped guide me and thus the program and all that we eventually did with it.

4.1 Introduction

As is outlined many times in this thesis, there is a current need for more peptidomic information both from multiple tissue types within individual species (see Chapter 6), as well as across a variety of species (see Chapter 5). Furthermore, biomarker identification is another area of research that has become very popular within the last several years (see chapter 7). A common theme of these two areas of research is the extensive use of mass spectrometry to perform the analysis and the need to manage and process large datasets acquired during this research.

For peptidomic work, many varying organs, or a variety of species must be analyzed in order to make meaningful conclusions. Furthermore, for the best possible analysis this means extracting each sample and then separating using HPLC. By doing this, the extract is simplified and more peptides can be resolved and observed as those with lower abundances are no longer suppressed by higher abundance peptides. Through this process, though, a huge amount of data is produced as it is necessary to analyze every fraction. In some cases each fraction will need to be analyzed multiple times depending on the instrument and the information to be gained. In all, this process of exhaustive peptidomic analysis can quickly create thousands of samples, making data analysis a daunting task.

On the other hand, for biomarker investigations, it is vital to provide statistical relevance for any conclusions to be recognized in the scientific community. To produce the statistical numbers necessary for this, the larger the

data set, the better. Thus, even if small numbers of animals are used for an infection trial, the samples obtained need to be analyzed multiple times in order to increase the statistical relevance of such a study. Through this method, a sizeable amount of data is created that must also be processed in order to make any meaningful conclusions.

Thus, we must turn to computers and their software to manage this problem. Through the use of appropriate software, we can knowledgeably combine the data into an amount that is manageable and compare and contrast data sets to highlight similarities and differences for comparative peptidomics and biomarker research. Additionally, for biomarker research we can take these varying features and use them to make predictions about the infection status of individual samples.

4.2 Original Aims and Current Uses of Software Package

Through the progression of my graduate career, I have been integral in the development of a bioinformatics software package named SpecPlot. This work was done together with Sean McIlwain of the UW-Madison Department of Computer Sciences. This custom-designed software package has many unique features unique and make it one thing that I am rather proud of resulting from my graduate career.

Originally, this program stemmed from some of the needs we were realizing as we worked on the prion biomarker identification research (see Chapter 7). For this research there was a need to create statistically relevant

conclusions. Because of the small number of samples, to increase the number of samples in a more virtual manner, we spotted each sample three times and then analyzed each spot three times, thus creating nine spectra for each sample. Afterward collecting the data, each spectra was used as an individual sample for contrast to samples of the opposite infection status in order to identify any classifying features. To simplify this method, there were several operations we needed to perform on the spectra in order to completely process them.

Firstly, because isotopic peaks are not essential to biomarker discovery, and often complicate matters, we wanted to remove the isotopic peaks and use only the monoisotopic masses for feature identification. Thus, deisotoping was the very first procedure written. One complicating factor is that the most intense isotopic peak is not always the monoisotopic mass. Another factor is when isotopic distributions overlap. When this happens, identifying both monoisotopic masses can be difficult. As a result of all these issues, the deisotoping module has been evolving throughout the software development. This module, even though changing often, was used for both the biomarker research and the comparative peptidomics research. Additionally, by utilizing the fact that electronic noise does not have isotopic distribution, we added a feature to the deisotoping module which also removes electronic noise (see **Figure 5.6**).

After developing a method for deisotoping the peak lists, we found it necessary to visualize the mass spectral data in some way. Furthermore, while there are several commercially available bioinformatics software packages, all of

them were rather expensive and none of them would accept the data format we were producing from the IonSpec© MALDI-FTMS software that we were using. Thus, SpecPlot was born as a way to not only visualize the data (see **Figure 7.4**), but also to provide us with the platform to carry out the statistically relevant operations necessary for our research. The SpecPlot software was programmed such that for each necessary operation a "module" was created. This allowed us the flexibility of modifying certain portions of modules while retaining others for use in the future. Through this process, SpecPlot has developed into a novel and informative method of representing data and provides great flexibility and usefulness for bioinformatics research when using instruments that are not supported by other commercial software.

Another module that was designed allowed us to contrast two classes of different data, which was useful to identify differences between infected and uninfected spectra (thus identifying classifying features, see Chapter 7). These differences are the biomarkers and can be used as classifying features that we eventually use to train learning algorithms and make infectivity predictions (see Chapter 7 and **Figure 7.5**). Due to the features produced, this module became one of the most important modules in SpecPlot.

Furthermore, by identifying similarities and differences in the peptidomes of related but different samples, conclusions can be made regarding peptide family evolution (when comparing related species, see Chapter 5), as well as possibly providing some information regarding peptide function (when comparing organs from the same species, see Chapter 6). Although the module for

identifying differences between samples was useful for the identification of classifying features (biomarkers) we needed a way to identify peptides that were present in the multispecies peptide data. To do this, we were required to alter the feature identification module to compare to a theoretical list of known crustacean peptides instead of contrast, and thus considered this a PeakMatching module instead. By using this module, we were able to quickly make peptide identifications in all species' thoracic ganglia extract samples simultaneously and therefore draw conclusions from a large dataset in a much shorter time-scale (see **Figure 5.9** and **Figure 6.6**). Overall, these Feature Identification and PeakMatching modules have been some of the most useful modules of the SpecPlot software.

Another challenge that we encountered when performing our comparative peptidomic research was the fact that we not only had 70 fractions for each species we studied, but to detect the full range of possible peptide masses in each fraction, we analyzed each fraction three times in three varying mass ranges (see Chapter 5). Even though we had three spectra for each fraction, we needed a way to merge the peak lists from each of these fractions into one peak list for each fraction. Furthermore, once we had the peak lists for each fraction, we used SpecPlot to visualize them, identified the fractions with meaningful data in them, and then used the same module to merge these fractions into one all inclusive peak list indicative of all of the detected peptides for a particular sample (see **Figure 5.7**). The same module was also used to combine all the peak lists from each fraction from the organ samples that were analyzed (see Chapter 6).

Because of its ability to simplify and reduce the size of data sets while still retaining all the important information, this module has been used extensively for a great deal of the comparative peptidomics research.

An additional simpler module that was created was one that allows the user to reduce the peak list size by cutting it at certain high and low values. While this may seem trivial, it was necessary in many cases as it allowed us to reduce the size of the peak list as well as remove data that was not necessary to our research. This is especially useful when using an instrument such as a time-of-flight instrument that produces a great deal of low mass noise that can be discarded as it is not only useless, but can at times make data analysis more complicated than is necessary.

Although SpecPlot has many other modules that do an assortment of operations, another helpful one that I helped design is a module that aligns the spectra either to each other or to a set of known calibration peaks. While software for instruments will often externally calibrate spectra, these calibrations often are not accurate and can also often cause variations in how the masses are adjusted. This means that some peaks (while technically the same peaks) will appear in two spectra at slightly different masses. To adjust for this, it is possible to align the peak lists so that they are much closer to their actual mass than the peak lists directly from the instrumentation software. As mentioned, when creating peak lists, you may or may not have known masses to calibrate on. Additionally, there are times when you have not included calibrants but are able to use known peaks that are already present in the peak lists. By aligning to

these peaks, one is able to increase the mass measurement accuracy greatly. Unfortunately, not all peak lists contain known peaks with which calibration can be made. Although not as good as having known peaks, it is possible to align the peak lists with each other so at least all peak lists have consistent masses between them. There are different methods by which this method of alignment or calibration is carried out, but the main goal of all the methods is to increase the mass measurement accuracy and/or to line up all peak lists with each other for future comparison to the theoretical list of known peptides.

4.3 Operation of Modules

As stated above, SpecPlot has many modules that are necessary for the processing of large data sets. There are modules that are based in the same basic principle, but the majority of the modules each have a different method of operation. In this section, I have highlighted how each of the modules mentioned above works in a simpler, more understandable manner.

Peak List Visualization

Although it may seem rather simple, visualization of the peak lists is one factor that we had to decide on when designing this software. For display purposes and because we were only working with the peak lists and not the complete spectra, we decided to display peaks as lines. For this display, each peak list was stacked and its identity was indicated on the left and the *m/z* scale

indicated along the bottom. While intensity of the peaks has never been used as factor in identifying classifying features or making peptide identifications, it was displayed using different colors to indicate how intense a peak was, blue being least intense, green a medium intensity, and red a high intensity. Furthermore, knowing what SpecPlot would be used for, each peak was converted to a feature vector that could later be manipulated by the software. By doing this, we were able to display multiple spectra individually and still retain the information for each peak necessary to carrying out further manipulations.

Deisotoping

To deisotope each peak list, we perform simple ^{13}C deisotoping. For this process, the spectra are analyzed for peaks that are distributed *m/z* 1.0078 apart, with a user entered value of error allowed. We also require that the distribution of peaks contain only one local maximum per group of isotopic peaks. The isotopes are then collapsed into one peak, taking the *m/z* value from the peak that has the minimum *m/z* value, i.e. the monoisotopic peak. Additionally, to remove peaks due to electronic noise, we remove all peaks that do not have a corresponding isotopic distribution.

Merging Peak Lists

The merging of peak lists may seem a simple operation, but due to run to run variations in the data, merging peak lists can be a complicated procedure (obviously simplified for us by the use of the software). The purpose of this

module is to collapse redundant peak lists (such as overlapping *m/z* ranges) from the same sample into one cohesive peak list. Obviously, there is some error that must be allowed, and this value is entered by the user. Depending on the instrument and calibration methods used, it is often necessary for the user to try a range of error tolerances. If the value is too small, SpecPlot will not combine data that should be combined, and if the value is too large erroneous peak combinations may occur. Thus, it is a process of trial and error while observing the effect of different error values on a known peak that is present in a majority of the peak lists. We accomplish this peak merging using hierarchical clustering upon the *m/z* values across different peak lists using an algorithm similar to Tibshirani et al, 2004 (see Chapter 5 and **Figure 5.3**). The algorithm assures that each cluster midpoint is no more than +/- the error ppm from the midpoint of another cluster. For this application, the groups identified by the hierarchical clustering algorithm for the spectra to be compressed are then merged into one peak list with peak values assigned by taking the *m/z* value for the peak with the greatest intensity out of the spectra. We then generate a peak list using the midpoints of the clusters as the new *m/z* value and the most intense peak within the cluster as the new intensity value. Because this module allows us to reduce the amount of data we have to deal with, it is one of the most useful modules and has applications to all large data sets.

Peak List Alignment and Calibration

When using any mass spectrometry instrumentation, there will be some run to run variations, no matter how well the instrument is tuned. Unless every spectrum is internally mass calibrated, there will be slight variations in the masses from one run to the next. For this reason, it is necessary to have an alignment module that can line up peaks from spectrum to spectrum. To perform this alignment, a theoretical peak list of alignment peaks is created of *m/z* values for peptides that were found in a majority of the species peak lists. Given a list of alignment peaks and an error window (either daltons or ppm), the algorithm we use searches for these peaks within each of the species' peak lists. The search then finds the most intense peak within each alignment peak window. Using the list of found peaks, a least-squares linear fit is performed to derive the constants for the correction equation (Corrected = A + B×(Original) where A is the absolute *m/z* shift and B is the change in scale as the *m/z* changes). The corrected value for one peak's *m/z* value is calculated using the linear equation from the peak's original *m/z* value. The algorithm calculates the values of A and B for each spectrum, then corrects the *m/z* values of the peaks within that spectrum using the linear equation. If no alignment peaks are found, the spectrum is left untouched. If only one alignment peak is found, B is set to one and A is the *m/z* delta between the original and alignment peak. By utilizing this module, we can somewhat improve the mass measurement accuracy of a large number of peak lists, even those that are made up of data that has been merged together from more than one peak list. If used in the right situation, this increase in mass

measurement accuracy can enhance the ability of SpecPlot to make accurate peptide identifications.

Peak Matching

Once composite peak lists for each sample have been created and they have all been aligned we can begin comparing the data and identifying known peptides. This is done by first entering a list of *m/z* values for any known peptides we are interested in finding. By entering a +/- ppm value, a window is created, with the known peptide *m/z* values at the midpoints. The samples' peak lists are then compared to these windows and if a peak exists within the windows, identification is indicated by a blue square in the resulting heat map, whereas a black square results from no identification being made. By doing this, we can indicate which peptide *m/z* peaks are present within each species' composite mass spectrum. While this is a simple process, this module makes it possible for us to identify the presence or absence of a list of known peptides in a large data set. This could be performed by hand, but through the use of SpecPlot, we are able to perform it in a matter of seconds as well as removing any human error that might be otherwise introduced.

4.4 Conclusions

Bioinformatics, in combination with mass spectrometry, has played an important role in my graduate research. Since mass spectrometry can quickly

produce large data sets, there is a need for bioinformatics to process that data and make necessary conclusions. This bioinformatics approach is also necessary when performing such research as biomarker identification, comparative peptidomics, and much more. While there are a number of bioinformatics software suites commercially available, none of them were able to fit our needs of accepting spectral information from our unique instrument, being flexible in all ways, and being cost effective. Overall, the process of developing SpecPlot has taught me a great deal about higher level statistics, bioinformatics, and how extensively computers can be utilized in the processing of large data sets.

Part III

Biological Applications

Chapter Five

Combining MALDI-FTMS and Bioinformatics for Rapid Peptidomic Comparisons

The coauthors for this chapter are Sean McIlwain and David Page of the University of Wisconsin-Madison Department of Biostatistics and Medical Informatics, Andrew Christie of both the University of Washington Department of Biology and Mount Desert Island Biological Laboratory in Salisbury Cove, Maine; and Mingming Ma and Lingjun Li of the University of Wisconsin-Madison School of Pharmacy.

5.1 Abstract

Increasing research efforts in large-scale analyses of peptides and proteins have led to many advances in technology and method development for collecting data and improving the quality of data. However, the resultant large datasets often pose significant challenges that require effective handling and processing these large datasets and extracting useful information in a high throughput manner. We describe one such method whereby we performed peptidomic analyses of nervous tissue extracts from several decapod crustacean species using high performance liquid chromatography (HPLC) fractionation off-line coupled to high resolution matrix-assisted laser desorption/ionization Fourier transform mass spectrometry (MALDI-FTMS). Following data acquisition, the data collected from discrete LC fractions was compiled and analyzed using an in-house developed software package that deisotoped, compressed, calibrated, and matched peaks to a list of known crustacean neuropeptides. By processing these data via bioinformatics tools such as hierarchical clustering we report the first set of peptidomic studies on these crustacean species to date. More than 110 neuropeptides that belong to 14 peptide families were mapped in five crustacean species. Overall, we demonstrate the utility of MALDI-FTMS in combination with a bioinformatics software package for the elucidation and comparison of peptidomes of varying crustacean species. This study established an effective methodology and will provide the basis for future investigations into more comprehensive comparative peptidomics with larger collection of species

and phyla in order to gain a deeper understanding of the evolution and diversification of peptide families through speciation.

5.2 Introduction

Peptidomics and the identification of novel peptides from a variety of species has become an intensive research focus due to the importance of neuropeptides as signaling molecules and the rapid development of mass spectrometry (MS)-based tools and methodologies to identify these intriguing molecules.[1-8] Mass spectrometry excels as one of the most informative methods for neuropeptide analysis as it enables sensitive identification of multiple peptides/proteins simultaneously, can provide sequence information via gas-phase fragmentation, and can do this on complex biological samples in a high throughput manner.[9, 10] In addition, the coupling of MS with various separation techniques offers expanded dynamic range and enhanced capability for complex mixture analysis.[9, 11-20] The attribute of MS in generating a large amount of data in a short period of time also poses a great demand for an effective means of data analysis to extract pertinent information out of the large datasets that are acquired. To counter this problem, there are bioinformatics software programs that aim at simplifying this data into manageable information.[21-24] Although these software programs are valuable tools, several limitations include high cost, their inability to accept many forms of MS data, and the inflexibility of many of the data processing variables. To address some of these limitations, we developed an in-house written software program, SpecPlot, to facilitate the processing and

analysis of huge datasets that MS can now produce. The advantages of our program are such that almost all MS data formats can be input and analyzed, as well as a number of variables that can be adjusted to better reflect the similarities and differences between samples and the data obtained from them.

In addition to peptide identification, it is often of significant interest in comparing and understanding the peptide/protein content similarities and differences between species, individuals, or tissue types. By determining the peptide/proteins content in each sample, and how they are similar and different from each other, insight into the function of such peptides/proteins can be derived. From a broader standpoint, it may be more useful to highlight the groups of peptides, or peptide families when making such comparisons. By examining peptide isoforms within peptide families in a specific species or across multiple related species, significant insight can be gained about function and evolution of a peptide family.[25-31] Obviously, this type of research produces large amounts of data that must be processed and analyzed before conclusions can be made. The bioinformatics processing program we describe in this report provides an attractive alternative as it has been designed to handle large datasets, as well as compare and contrast them to each other without introducing a great deal of human error.

For this report, we have used matrix assisted laser desorption/ionization - Fourier transform mass spectrometry (MALDI-FTMS) to obtain data both directly from tissues and HPLC fractions resulting from crude extracts of thoracic ganglia of several crustacean species originated from the Puget Sound and San Juan

Archipelago region of Washington State. Crustaceans provide excellent model systems for the study of neuroendocrine regulation and neuromodulation of behavior, with many recent studies focusing on the determination of their neuropeptide contents.[32-39] The following crustacean species were used in this study: *Cancer productus* (the red rock crab), *Hyas lyratus* (the Pacific lyre crab), *Oregonia gracilis* (the graceful decorator crab), *Hemigrapsus nudus* (the purple shore crab), and *Telmessus cheiragonus* (the helmet crab) (**Figure 5.1**). These species are all decapod crustaceans from the brachyuran family of crabs (**Figure 5.2**). From each species, the thoracic ganglia were collected and analyzed as these organs are integral to the central nervous system with rich peptide content. Furthermore, these organs are well conserved morphologically from species to species.[40]

Once the organs of interest were isolated, pooled, and extracted, HPLC fractionation of crude extracts followed by high resolution MALDI FTMS screening of individual LC fractions from each species were performed. We then processed and analyzed the data using bioinformatics in the form of an in-house written program that addresses some of the current limitations of commercial software. By using this method, we demonstrate the processing of hundreds of mass spectra from several different extracts, how they can be quickly analyzed through comparison and, how new peptides can be identified while performing this method. Overall, the flexibility and speed of this method make it an appealing approach to compare and contrast peptidomes from multiple species.

5.3 Methods

Species and tissue collection

All animals were collected by hand and were then maintained in flow through seawater tables at 12 °C until being dissected. Thoracic ganglia were removed from each species and immediately placed into acidified methanol (90:9:1, MeOH : glacial acetic acid : deionized H_2O). Samples from each species were combined and stored in microcentrifuge tubes. These tubes were then stored in a -80 °C freezer until peptide extraction.[41]

Peptide Extraction and HPLC Separation

Microcentrifuge tubes containing thoracic ganglia (TG) were removed from a -80 °C freezer and immediately placed on ice for peptide extraction. Fifteen thoracic ganglia were then removed from the tube and placed into a 1 ml tissue grinder (Wheaton, Inc.) along with ~500 μl acidified methanol. The tissue was then ground until no apparent large particles remained and all tissue was suspended. The acidified methanol with the suspended tissue was then removed from the tissue grinder and placed into a new microcentrifuge tube. The extract was then centrifuged at 16,100 rcf for 10 minutes on an Eppendorf 5415 D microcentrifuge (Brinkmann Instruments Inc., Westbury, NY). After centrifugation, the supernatant was removed and placed into another new microcentrifuge tube. Following this, another 500 μl of acidified methanol was used to rinse the tissue grinder after which the acidified methanol was transferred to the tube with the original pellet. The pellet was then resuspended,

reextracted, centrifuged again and the supernatant was removed and added to the original supernatant. This process was repeated again producing a final volume of ~1.5 ml of extract supernatant. The pellet was retained and stored in a -80 °C freezer. The 1.5 ml of supernatant was then concentrated using a Savant SC 110 Speed-Vac concentrator (Thermo Electron Corporation, West Palm Beach, FL). The extract was then resuspended with ~50 μl of ddH_2O with 0.1% formic acid (Sigma Aldrich) and vigorous vortexing.

After resuspension, extracts were ready for HPLC separation. For HPLC separation, a 20 μl injection loop was used. The column used was a 2.1 mm I.D., 250 mm long, C18 column (Alltech, Inc.) with a guard column on a Rainin Dynamax HPLC system (Rainin Instrument Inc., Woburn, Massachusetts, USA) equipped with a Dynamax UV-D II absorbance detector. Mobile phases used were A: dH_2O w/ 0.1% formic acid and B: acetonitrile (HPLC grade) w/ 0.1% formic acid. The gradient used was 5-60% B from 0-100 minutes, 60-95% B from 100-110 minutes, and 95% B from 110-120 minutes with a flow rate of 0.2 ml/min. Fractions were collected every two minutes with a Rainin Dynamax FC-4 fraction collector producing 60, 400 μl fractions. Fractions were then concentrated to dryness with the speed-vac. Each fraction was then resuspended with 20 μl ddH_2O w/ 0.1% formic acid.

Chemical Derivatization of HPLC Fractions

To confirm the sequence of the new peptide at *m/z* 1256 in *H. lyratus*, several chemical labeling methods were performed. The formaldehyde labeling was

carried out by spotting 0.3 μL of the appropriate *H. lyratus* HPLC fraction on the MALDI plate, followed by the addition and mixing of 0.3 μL of 26 mM $NaBH_3CN$ (Sigma-Aldrich, St. Louis, MO), and subsequent addition of 0.3 μL of formaldehyde (20% in H_2O vol/vol, Sigma-Aldrich). The droplet was then allowed to react at room temperature (25 °C) for five minutes and then quenched by the addition of 0.3 μL of 50 mM ammonium bicarbonate solution.[46] To methylate the appropriate *H. lyratus* TG fraction, 5 μL was dried down in a speed-vac (Savant, Inc.) followed by the addition of 50 μL of methanolic HCl. The methylation reaction was allowed to take place at 37 °C for two hours. The methanolic HCl was removed by evaporation in a speed-vac, and then the sample was reconstituted in 30% methanol with 0.1% formic acid.[32] Finally, acetylation was carried out by spotting 0.3 μL of the appropriate *H. lyratus* TG HPLC fraction, followed by the addition of 0.3 μL of 3:1 methanol:acetic anhydride. The solution was mixed with 0.3 μL of 50 mM ammonium bicarbonate solution and left at room temperature for three minutes.[47, 48, 3] After each of these labeling methods was performed on the fraction (and peptides) of interest, MALDI-FTMS was used to analyze the results.

MALDI FTMS

After being resuspended, fractions were analyzed by matrix assisted laser desorption/ionization – Fourier transform mass spectrometry (MALDI-FTMS). Prior to analysis, 0.5 μl of saturated DHB (150 mg 2,5-dihydroxy benzoic acid (ICN Chemicals) in 1 ml 50:50 MeOH (purge and trap grade, Sigma Aldrich) :

ddH_2O) was spotted followed by the addition of 0.5 μl of the fraction on the sample probe and allowed to crystallize at room temperature. Fractions were analyzed using an IonSpec Fourier transform mass spectrometer (Varian/IonSpec Corporation, Lake Forest, California, USA) equipped with a 7.0 Tesla actively shielded superconducting magnet.. Ions were formed through the use of a nitrogen laser producing ultraviolet light at 337 nm. Ions were then carried down a quadrupole ion guide into the ion cyclotron resonance cell where they can be detected. Each fraction was analyzed three times in order to cover wide range of masses. The first analysis was focused on *m/z* 1000 and ranged from *m/z* 108-2500, the second analysis was focused on *m/z* 2500 and ranged from *m/z* 215-4500, and the final analysis had ionization events focused on *m/z* 1000 (2 events), *m/z* 2500 (1 event), *m/z* 5000 (2 events), *m/z* 8000 (2 events) and ranged from *m/z* 215-12000. To obtain a peak list, we used the IonSpec99© software (used to acquire the data; IonSpec, Corp.) to differentiate between noise and signal and then produce a peak list of all peaks in the spectra.

Sustained off-resonance irradiation - collisionally induced dissociation was performed by first using an arbitrary waveform with a notch 10 *m/z* wide (+/- 5 *m/z*) to eject all other ions. While this window may seem wide, the peaks were observed and it was confirmed that there were no other peaks within this window in the sample analyzed. If other peaks were found within the window then the width was adjusted to exclude these other peaks. The ions of interest were then excited with 125% amplitude of their adjusted resonant frequency into a cloud of

nitrogen injected into the cell for 200 milliseconds at the time of excitement. The resulting fragments were then detected and the data analyzed.

Nanoflow HPLC ESI-QTOF MS/MS

Nanoscale LC QTOF MS/MS was performed using a Waters capillary LC system coupled to a QTOF Micro mass spectrometer (Waters Corp., Milford, MA, USA). Chromatographic separations were performed on a C18 reverse phase capillary column (75 µm internal diameter × 150 mm length, 3 µm particle size; Micro-Tech Scientific Inc.). The mobile phases used were: A, deionized H_2O with 5% acetonitrile and 0.1% formic acid; B, acetonitrile with 5% deionized H_2O and 0.1% formic acid; C, deionized H_2O with 0.1% formic acid. An aliquot of 6.4 µL of *H. lyratus* TG HPLC fraction was injected and loaded onto the trap column (PepMap^{TM} C18; 300 µm column internal diameter × 1 mm, 5 µm particle size; LC Packings, Sunnyvale, CA, USA) using mobile phase C at a flow rate of 30 µL/min for 3 min. Following this, the stream select module was switched to a position at which the trap column became in line with the analytical capillary column, and a linear gradient of mobile phases A and B was initiated. A splitter was added between the mobile phase mixer and the stream select module to reduce the flow rate from 12 µL/min to 200 nL/min.

The nanoflow ESI source conditions were set as follows: capillary voltage 3800 V, sample cone voltage 35 V, extraction cone voltage 1 V, source temperature 120°C, cone gas (N_2) 13 L/hr. A data dependent acquisition was employed for the MS survey scan and the selection of precursor ions and

subsequent MS/MS of the selected parent ions. The MS scan range was from *m/z* 300-2000 and the MS/MS scan was from *m/z* 50-1800. The MS/MS *de novo* sequencing was performed with a combination of manual sequencing and automatic sequencing by PepSeq software (Waters Corp.).

With the combination of these methods, we were able to not only confirm the presence of many peptides, but also to sequence new peptides that we identified in the course of our study.

Data Analysis and Software Development

For each peak list, we perform simple ^{13}C deisotoping. For this process, the spectra are analyzed for peaks that are distributed *m/z* 1.0078 apart, +/- 25 ppm, which is the user entered value of error allowed. We further require that the distribution of peaks contain only one local maximum per group of isotopic peaks. The isotopes are then collapsed into one peak, taking the *m/z* value from the peak that has the minimum *m/z* value, i.e. the monoisotopic peak. To remove peaks due to electronic noise, we remove all peaks that do not have a corresponding isotopic distribution.

Because of the wide range of masses that could be present in each fraction, it was necessary to analyze each fraction at multiple ranges so as to insure that all ions were detected due to a time-of-flight effect encountered in MALDI FTMS.[42] As a result, each fraction had three spectra at different mass ranges which required the combination of each spectrum into one peak list for each fraction. After doing this, it was necessary to produce a peak list of all

peaks in all the relevant fractions from the TG of each species. Therefore, the peak lists from multiple spectra had to be combined into one peak list (once for each fraction and then another time for the complete species extract). We accomplished this using hierarchical clustering upon the *m/z* values across different peak lists using an algorithm similar to Tibshirani et al, 2004 (**Figure 5.3**).[43] The algorithm assures that each cluster midpoint is no more than +/- 25 ppm from the midpoint of another cluster. For this application, the groups identified by the hierarchical clustering algorithm for all three spectra are then compressed into one peak list with peak values assigned by taking the *m/z* value for the peak with the greatest intensity out of the three spectra. We then generate a peak list using the midpoints of the clusters as the new *m/z* value and the most intense peak within the cluster as the new intensity value. This same method is used to compress all of a species' fraction peak lists into one large peak list indicative of all peaks found in a species TG extract.

After the peak lists were created for each species, it was necessary to align the peak lists because the spectra were externally calibrated. To perform this alignment, a theoretical peak list of alignment peaks was created of *m/z* values for peptides that were found in a majority of the species peak lists. Given a list of alignment peaks and an error window (either daltons or ppm), the algorithm we use searches for these peaks within each of the species' peak lists. The search then finds the most intense peak within each alignment peak window. Using the list of found peaks, a least-squares linear fit is performed to derive the constants for the correction equation (Corrected = A + B×(Original)

where A is the absolute *m/z* shift and B is the change in scale as the *m/z* changes). The corrected value for one peak's *m/z* value is calculated using the linear equation from the peak's original *m/z* value. The algorithm calculates the values of A and B for each spectrum, then corrects the *m/z* values of the peaks within that spectrum using the linear equation. If no alignment peaks are found, the spectrum is left untouched. If only one alignment peak is found, B is set to one and A is the *m/z* delta between the original and alignment peak.

Once we have built the composite peak list for each species under study, we can begin identifying known peptides. This is done by first entering a list of *m/z* values for any known peptides we are interested in finding. By entering a +/- ppm value, a window is created, with the known peptide *m/z* values at the midpoints. The species' peak lists are then compared to these windows and if a peak exists within the windows, identification is indicated by a blue square in the resulting heat map, whereas a black square results from no identification being made. By doing this, we can indicate which peptide *m/z* peaks are present within each species' composite mass spectrum.

5.4 Results and Discussion

To begin the development of this method, we initially analyzed thoracic ganglia from each species using the MALDI-FTMS direct tissue analysis method developed in our lab.[41] By doing this we were able to quickly generate a snapshot of the peptidome for each species, and thus make an initial comparison of peptides present in each species (**Figure 5.4**). However, a direct analysis of

tissue does not provide a complete profile of an organ's peptidome due to the small amount and the heterogeneity of tissue that is analyzed. Because of the complex structure of the thoracic ganglia, it is necessary to create an extract, then reduce chemical complexity and expand dynamic range by separation using HPLC prior to MS analysis.

Fractions from each species thoracic ganglia extract were analyzed three times by MALDI-FTMS. The resulting spectra included the mass ranges *m/z* 108-2000, *m/z* 400-4500, and *m/z* 215-12000 (**Figure 5.5**). By analyzing these ranges, most peptides present in the samples will be detected. With three spectra for each fraction, and 60-70 fractions for each species extract, there are 180-210 spectra for each species. Currently, with only five species fully explored, more than 1000 spectra need to be processed. With the further addition of species to this comparative analysis of neuropeptide complements in a large collection of species, an evident challenge is how to process this sizeable data set. To solve this problem, we developed an in-house written bioinformatics software program called SpecPlot. This software not only enables the management of large dataset, but also helps to reduce human error from manual analysis, although during software development and validation manual checking has been performed and confirmed the computer generated results. As is evidenced by the identification of unknown peptides with masses very close to those of known peptides, it is necessary to exercise caution for peptide assignment by performing further mass spectral analyses (ion fragmentation,

internal calibration) or other confirmation experiments with complementary information (dot-blotting with antibodies).

After collecting the spectra and using the IonSpec99© software to produce a peak list, we can begin using the SpecPlot software to analyze the data. Initially, peak lists are deisotoped to leave only monoisotopic masses in the list as well as remove most noise peaks (because electronic noise peaks do not have an isotopic distribution these peaks are discarded). After creating a list of only monoisotopic peak masses all three peak lists are stacked and then collapsed together to eliminate the overlap of the three spectra (**Figure 5.6**). Included in the peak lists produced by the IonSpec99© software is information about the relative abundance of each peak in the list. This information is currently used by SpecPlot only to display the peaks with varying colors (blue for least intense, green for medium intensity, and red for most intense). By collapsing three peak-lists of each individual fraction into a single peak list we reduce the amount of data that need to be processed by three fold.

The next step in data analysis requires stacking of each fraction's peak list to display them graphically. Because fraction number is a function of elution time during the HPLC run, stacking their peak lists in order elucidates during which fractions peptides were eluted off of the column (**Figure 5.7**). Additionally, this stacking also highlights any noise peaks and chemical background that made it through the deisotoping step. These show up as peaks appearing in all fractions, thus creating a continuous line from the top of the display to the bottom (**Figure 5.7**). Although a noise peak can not be selectively removed, knowing its mass

prevents the assignment of a peptide identity to this mass in further processing of the peak lists.

As with the multiple peak lists for each fraction, the fractions' peak lists now are combined to remove redundant peaks and create a master peak list. Although it is possible to collapse all fractions into this final peak list it is neither necessary nor desirable as many fractions that contain only electronic and chemical noise would be included. By observing the stacked peak lists from all fractions, it is obvious which fractions contain meaningful information (**Figure 5.7**) and should be included in the final peak list. Once the appropriate fractions have been identified, they are collapsed into a peak list that then represents the full complement of peptides in each species' thoracic ganglia (**Figure 5.7**).

Once there is a peak list for each species, a comparison between them can be performed in order to elucidate similarities and differences between peptide complements. Although the peak lists can now be compared, such a contrast has little meaning without identities for the peaks. In order to identify peptides in these peak lists, a peak list (**Table 1**) of several known crustacean peptide families was created from sequences found in current literature[29-32] and loaded into SpecPlot. Once the theoretical peak list of known crustacean peptides was formed, all species peak lists were stacked in order to observe similarities and major pattern differences between species (**Figure 5.8**). From this comparison many similarities can be observed, with multiple lines aligned with each other, but an overall comparison is difficult due to the number of peaks

present. To make these comparisons easier and less subject to human error, we have again employed the SpecPlot software.

To identify the presence of peptides in the species extracts, the created peak list of known peptides was entered into SpecPlot and compared to each of the species' peak lists individually. Identification of a peptide within a species was made by first creating a window of a designated width (+/- ppm) on either side of the peaks of the theoretical peak list. Each of the species peak list was then examined by the software to determine whether a peak exists within the windows created from the theoretical peak list. If a peak was observed within a window, identification was made and represented by a blue square in the heat map, as opposed to the black square indicating no match was present. Because we used external calibration for this work, we tried varying error values (5, 10, 25, 50, 100, and 200 ppm) in order to demonstrate how the number of identifications changes as the window was widened. External calibration was used due to the large number of spectra obtained, the wide range of these spectra, and difficulty in using internal calibrants with the instrumentation at the time. In the future we will incorporate methods such as internal calibration on adjacent standards (InCAS) to improve this mass measurement accuracy so that a smaller window can be used without loss of identified peptides.

From this comparison, the map of detected peptides is created for each of the crustacean species without regard to peptide intensity (**Figure 5.9**). Once peptide identities have been assigned to the masses, the information regarding their presence in the species can be grouped together into their respective

peptide families. By grouping the masses in this way, patterns regarding peptide presence in species becomes evident and can be used to make further conclusions about the peptide family and its distribution among crustacean species (**Figure 5.9**).

To confirm the peptide identification using our SpecPlot program, we manually checked for the presence of the peptide HL/IGSL/IYRamide (*m/z* 844.48) and [Val 13]-orcokinin (*m/z* 1502.69) in all species analyzed. To find the spectra containing the appropriate peaks, the peaks were found and their respective fractions identified in the map of the HPLC run (**Figure 5.7**). Then, the corresponding spectra were checked by hand for the presence of the peaks as well as their isotopic distribution (**Figure 5.10**). It can be observed that the *m/z* 844.48 and *m/z* 1502.69 peaks are present in all species, and furthermore, that they are present in varying levels in different species. Finally, it is also noted that the HL/IGSL/IYRamide and [Val 13]-orcokinin eluted within one fraction from sample to sample, indicating reproducible HPLC elution conditions. To further confirm the identities of these peptides, sequence-specific fragmentation analyses were performed with two representative resulting fragmentation spectra of HL/IGSL/IYRamide and [Val 13]-orcokinin shown at the bottom (**Figure 5.10**). Several other peptides were confirmed in multiple species in a similar manner.

In the process of confirming the presence of orcokinins (NFDEIDRSGFG (*m/z* 1256.56) and NFDELDRSGFGFH (*m/z* 1540.68)) several peaks were noticed very near (within 0.1 Da) these two peaks and also co-eluted in the same fraction. To identify what these other peaks were, we fragmented them and

found fragmentation patterns that did not match the corresponding putative orcokinin fragmentation patterns. Thus, it was obvious that these ions were not orcokinins, although the two putative peptides are related as is evidenced by several peaks they both share in their fragmentation spectra (b-ions up to b11) (**Figure 5.11**). To further sequence these peptides, these fractions were run on the CapLC Q-TOF and the resulting fragmentation spectrum (**Figure 5.11 B**) was compared to the FTMS data to confirm the proposed sequence. The proposed sequence for the *m/z* 1256 peptide is AQAPFI/LSQPNI/LA. Both MALDI-FTMS and CapLC Q-TOF data corroborate this sequence (**Figure 5.11**).Obviously, using these mass spectrometric methods, it is impossible to differentiate the leucine residues from the isoleucine residues. Additionally, to further confirm this sequence formaldehyde labeling, methylation, and acetylation were used to react with the varying functional groups and confirm the presence or absence of certain amino acids in the sequence. After these methods were applied, MALDI-FTMS was used to assess the results of these reactions. Through this process we were able to confirm this sequence. To further show that this peak was not [1-11]orcokinin and was actually a new peptide, we performed an internal calibration on an adjacent standard (InCAS) experiment where the de novo sequenced new peptide was combined, in the detection cell, with authentic orcokinin [1-11] at *m/z* 1256 as well as other peptide standards (Angiotensin II, *m/z* 1046.5418; Arg-Vasopressin, *m/z* 1084.4078; Angiotensin I, *m/z* 1296.7046; and Substance P, *m/z* 1347.7370). The spectrum was then calibrated on these standards with a mass measurement accuracy of 0.4 ppm for orcokinin [1-11]

(with a calculated *m/z* 1256.5542). The resulting mass measurement accuracy of the other peak to our new peptide sequence (of AQAPFI/LSQPNI/LA with an exact mass of 1256.6634 m/z) is 0.6 ppm. These experiments highlight the utility of high resolution MALDI FTMS for distinguishing two peptides with similar masses using high MMA offered by InCAS and sequence-specific fragmentation analysis (**Figure 5.11**). As stated above, the 1256 m/z and 1540 m/z were evidently related as is evidenced in their similar fragmentation patterns (**Figure 5.11**). Once the sequence of the 1256 m/z was identified, we compared the peaks from the fragmented AQAPFI/LSQPNI/LA to that of the fragmented 1540 m/z peptide. It was found that both spectra contained many of the same b-ion peaks up to the b11 peak of the AQAPFI/LSQPNI/LA (*m/z* 1167.62). From there, the fragmentation spectra was manually sequenced and the complete sequence was determined to be AQAPFI/LSQPNI/LNGI/LA (*m/z* 1540.8118) (**Figure 5.11**).

Finally, information regarding the conservation of peptide families among different crustacean species can be extracted from this comparative peptidomic study. By observing the heat map of peptide matches (**Figure 5.9**), it is obvious that several peptide families, i.e. FLRFamides, RYamides, and orcokinins, are very well conserved among different species. By contrast, peptide families such as the allatostatins (A-, B- types), the CPRP's (crustacean hyperglycemic hormone precursor-related peptides) exhibit greater diversity through speciation. Although a limited set of species investigated here, the methodology established in this study will enable examination of a larger collection of species in a high

throughput manner. The preliminary data presented in this report provides a foundation for future evaluation of the hypothesis regarding variation of neuropeptidome as a function of speciation.[44, 45]

5.5 Conclusions

We describe a new method using MALDI-FTMS in conjunction with bioinformatics for the exploration of peptidomics. As this field grows, such tools will become ever more important to handle and manage the large data sets that are produced. The advantages of sensitivity and high mass measurement accuracy of the MALDI-FTMS show the value of using such an instrument for peptidomics research. Bioinformatics has allowed us to widen our range of analysis, reduce the amount of human error introduced, and efficiently understand the results more fully. The software tool SpecPlot developed in this study can be easily tailored for other applications and to process datasets produced by other mass spectrometers (i.e. mass list including ion abundance).

The combined MALDI FTMS and bioinformatics tools allowed us to collect data from multiple species and then analyze that data in a manageable way. A rapid comparison of peptide content in the nervous system of these species can be made with known peptides highlighted. This method, while not a complete analysis of all peptides present in all species, can work as an excellent first-pass analysis of data to identify the known peptides present as well as provide an initial glimpse at the peptide complement among different species. Additionally, while there are a great number of peptides in unexplored species that can be

matched to known peptides, many more unknown and novel peptides await to be further investigated to gain better understanding of peptide diversity and conversation in the context of neuropeptide families. This research has provided an effective method for future comparative peptidomic investigations using multiple species to understand how peptide diversity has been affected through speciation.

5.6 Acknowledgements

We are grateful to the University of Wisconsin School of Pharmacy Analytical Instrumentation Center for access to the MALDI-FTMS instrument. We would like to thank Dr. Michael Nusbaum (University of Pennsylvania School of Medicine) for his generous donation of the Rainin Dynamax HPLC system to the Li laboratory. In addition, we wish to thank Daniel I. Messinger and Christina T. Ngo from the Christie laboratory at University of Washington and Friday Harbor Laboratories for their assistance in collecting and dissecting some of the crustacean specimens used in this study. Also, thanks to Dr. Peter O'Connor (Boston University, School of Medicine) for the use of his FTMS spectral analysis software BUDA (Boston University Data Analysis) in making figures with FTMS data. This research was funded by NSF CAREER award CHE-0449991, Alfred P. Sloan Research Foundation (LL), and a MDIBL New Investigator Award (from the Salisbury Cove Research Fund provided through the Thomas H. Maren Foundation [to A.E.C.]). J.J.S. acknowledges a predoctoral fellowship from the American Foundation for Pharmaceutical Education.

5.7 References:

1. Hummon, A. B., Amare, A., and Sweedler, J. V. (2006) Discovering new invertebrate neuropeptides using mass spectrometry. *Mass Spectrom Rev* 25, 77-98
2. Li, L., Garden, R. W., and Sweedler, J. V. (2000) Single-cell MALDI: a new tool for direct peptide profiling. *Trends Biotechnol* 18, 151-160
3. Fricker, L. D., Lim, J., Pan, H., and Che, F. Y. (2006) Peptidomics: identification and quantification of endogenous peptides in neuroendocrine tissues. *Mass Spectrom Rev* 25, 327-344
4. Baggerman, G., Verleyen, P., Clynen, E., Huybrechts, J., De Loof, A., and Schoofs, L. (2004) Peptidomics. *Journal of Chromatography B* 803, 3-16
5. Neupert, S., Johard, H. A., Nassel, D. R., and Predel, R. (2007) Single-cell peptidomics of drosophila melanogaster neurons identified by Gal4-driven fluorescence. *Anal Chem* 79, 3690-3694
6. Predel, R., and Gade, G. (2005) Peptidomics of neurohemal organs from species of the cockroach family Blattidae: how do neuropeptides of closely related species differ? *Peptides* 26, 3-9
7. Parkin, M. C., Wei, H., O'Callaghan, J. P., and Kennedy, R. T. (2005) Sample-dependent effects on the neuropeptidome detected in rat brain tissue preparations by capillary liquid chromatography with tandem mass spectrometry. *Anal Chem* 77, 6331-6338

8. Svensson, M., Skold, K., Svenningsson, P., and Andren, P. E. (2003) Peptidomics-based discovery of novel neuropeptides. *J Proteome Res* 2, 213-219

9. Wei, H., Nolkrantz, K., Parkin, M. C., Chisolm, C. N., O'Callaghan J, P., and Kennedy, R. T. (2006) Identification and Quantification of Neuropeptides in Brain Tissue by Capillary Liquid Chromatography Coupled Off-Line to MALDI-TOF and MALDI-TOF/TOF-MS. *Anal Chem* 78, 4342-4351

10. Schulz-Knappe, P., Schrader, M., and Zucht, H. D. (2005) The peptidomics concept. *Comb Chem High Throughput Screen* 8, 697-704

11. Haskins, W. E., Watson, C. J., Cellar, N. A., Powell, D. H., and Kennedy, R. T. (2004) Discovery and neurochemical screening of peptides in brain extracellular fluid by chemical analysis of in vivo microdialysis samples. *Anal Chem* 76, 5523-5533

12. Wei, H., Dean, S. L., Parkin, M. C., Nolkrantz, K., O'Callaghan, J. P., and Kennedy, R. T. (2005) Microscale sample deposition onto hydrophobic target plates for trace level detection of neuropeptides in brain tissue by MALDI-MS. *J Mass Spectrom* 40, 1338-1346

13. Floyd, P. D., Li, L., Moroz, T. P., and Sweedler, J. V. (1999) Characterization of peptides from Aplysia using microbore liquid chromatography with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry guided purification. *J Chromatogr A* 830, 105-113

14. Ji, C., and Li, L. (2005) Quantitative proteome analysis using differential stable isotopic labeling and microbore LC-MALDI MS and MS/MS. *J Proteome Res* 4, 734-742

15. Zhong, H., Marcus, S. L., and Li, L. (2005) Microwave-assisted acid hydrolysis of proteins combined with liquid chromatography MALDI MS/MS for protein identification. *Journal of the American Society for Mass Spectrometry* 16, 471-481

16. Young, J. B., and Li, L. (2006) An Impulse-Driven Liquid-Droplet Deposition Interface for Combining LC with MALDI MS and MS/MS. *Journal of the American Society for Mass Spectrometry* 17, 325-334

17. Chen, H. S., Rejtar, T., Andreev, V., Moskovets, E., and Karger, B. L. (2005) Enhanced characterization of complex proteomic samples using LC-MALDI MS/MS: exclusion of redundant peptides from MS/MS analysis in replicate runs. *Anal Chem* 77, 7816-7825

18. Chen, H. S., Rejtar, T., Andreev, V., Moskovets, E., and Karger, B. L. (2005) High-speed, high-resolution monolithic capillary LC-MALDI MS using an off-line continuous deposition interface for proteomic analysis. *Anal Chem* 77, 2323-2331

19. Orsnes, H., Graf, T., Degn, H., and Murray, K. K. (2000) A rotating ball inlet for on-line MALDI mass spectrometry. *Anal Chem* 72, 251-254

20. Musyimi, H. K., Narcisse, D. A., Zhang, X., Stryjewski, W., Soper, S. A., and Murray, K. K. (2004) Online CE-MALDI-TOF MS using a rotating ball interface. *Anal Chem* 76, 5968-5973

21. Andreev, V. P., Rejtar, T., Chen, H. S., Moskovets, E. V., Ivanov, A. R., and Karger, B. L. (2003) A universal denoising and peak picking algorithm for LC-MS based on matched filtration in the chromatographic time domain. *Anal Chem* 75, 6314-6326

22. Falkner, J., and Andrews, P. (2005) Fast tandem mass spectra-based protein identification regardless of the number of spectra or potential modifications examined. *Bioinformatics* 21, 2177-2184

23. Bensmail, H., Golek, J., Moody, M. M., Semmes, J. O., and Haoudi, A. (2005) A novel approach for clustering proteomics data using Bayesian fast Fourier transform. *Bioinformatics* 21, 2210-2224

24. Titulaer, M. K., Siccama, I., Dekker, L. J., van Rijswijk, A. L., Heeren, R. M., Sillevis Smitt, P. A., and Luider, T. M. (2006) A database application for pre-processing, storage and comparison of mass spectra derived from patients and controls. *BMC Bioinformatics* 7, 403

25. Veelaert, D., Schoofs, L., Verhaert, P., and De Loof, A. (1997) Identification of two novel peptides from the central nervous system of the desert locust, Schistocerca gregaria. *Biochem Biophys Res Commun* 241, 530-534

26. Dircksen, H., Bocking, D., Heyn, U., Mandel, C., Chung, J. S., Baggerman, G., Verhaert, P., Daufeldt, S., Plosch, T., Jaros, P. P., Waelkens, E., Keller, R., and Webster, S. G. (2001) Crustacean hyperglycaemic hormone (CHH)-like peptides and CHH-precursor-related peptides from pericardial organ neurosecretory cells in the shore crab,

Carcinus maenas, are putatively spliced and modified products of multiple genes. *Biochem J* 356, 159-170

27. Jimenez, C. R., Spijker, S., de Schipper, S., Lodder, J. C., Janse, C. K., Geraerts, W. P., van Minnen, J., Syed, N. I., Burlingame, A. L., Smit, A. B., and Li, K. (2006) Peptidomics of a single identified neuron reveals diversity of multiple neuropeptides with convergent actions on cellular excitability. *J Neurosci* 26, 518-529

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

29. Fu, Q., Kutz, K. K., Schmidt, J. J., Hsu, Y. W., Messinger, D. I., Cain, S. D., de la Iglesia, H. O., Christie, A. E., and Li, L. (2005) Hormone complement of the Cancer productus sinus gland and pericardial organ: an anatomical and mass spectrometric investigation. *J Comp Neurol* 493, 607-626

30. Fu, Q., Christie, A. E., and Li, L. (2005) Mass spectrometric characterization of crustacean hyperglycemic hormone precursor-related peptides (CPRPs) from the sinus gland of the crab, Cancer productus. *Peptides* 26, 2137-2150

31. Fu, Q., Goy, M. F., and Li, L. (2005) Identification of neuropeptides from the decapod crustacean sinus glands using nanoscale liquid

chromatography tandem mass spectrometry. *Biochem Biophys Res Commun* 337, 765-778

32. Fu, Q., and Li, L. (2005) 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* 77, 7783-7795

33. Cruz-Bermudez, N. D., Fu, Q., Kutz-Naber, K. K., Christie, A. E., Li, L., and Marder, E. (2006) Mass spectrometric characterization and physiological actions of GAHKNYLRFamide, a novel FMRFamide-like peptide from crabs of the genus Cancer. *J Neurochem* 97, 784-799

34. Stemmler, E. A., Provencher, H. L., Guiney, M. E., Gardner, N. P., and Dickinson, P. S. (2005) 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* 77, 3594-3606

35. Skiebe, P. (2003) Neuropeptides in the crayfish stomatogastric nervous system. *Microsc Res Tech* 60, 302-312

36. Li, L., Pulver, S. R., Kelley, W. P., Thirumalai, V., Sweedler, J. V., and Marder, E. (2002) Orcokinin peptides in developing and adult crustacean stomatogastric nervous systems and pericardial organs. *J Comp Neurol* 444, 227-244

37. Billimoria, C. P., Li, L., and Marder, E. (2005) Profiling of neuropeptides released at the stomatogastric ganglion of the crab, Cancer borealis with mass spectrometry. *J Neurochem* 95, 191-199

38. Saideman, S. R., Ma, M., Kutz-Naber, K. K., Cook, A., Torfs, P., Schoofs, L., Li, L., and Nusbaum, M. P. (2007) Modulation of Rhythmic Motor Activity by Pyrokinin Peptides. *J Neurophysiol* 97, 579-595

39. Stemmler, E. A., Peguero, B., Bruns, E. A., Dickinson, P. S., and Christie, A. E. (2007) Identification, physiological actions, and distribution of TPSGFLGMRamide: a novel tachykinin-related peptide from the midgut and stomatogastric nervous system of *Cancer* crabs. *J Neurochem* 101, 1351-1366

40. Huybrechts, J., Nusbaum, M. P., Bosch, L. V., Baggerman, G., De Loof, A., and Schoofs, L. (2003) Neuropeptidomic analysis of the brain and thoracic ganglion from the Jonah crab, Cancer borealis. *Biochem Biophys Res Commun* 308, 535-544

41. Kutz, K. K., Schmidt, J. J., and Li, L. (2004) In situ tissue analysis of neuropeptides by MALDI FTMS in-cell accumulation. *Anal Chem* 76, 5630-5640

42. O'Connor, P. B., Duursma, M. C., van Rooij, G. J., Heeren, R. M. A., and Boon, J. J. (1997) Correction of time-of-flight shifted polymeric molecular weight distributions in matrix-assisted laser desorption/ionization Fourier transform mass spectrometry. *Anal Chem* 69, 2751-2755

43. Tibshirani, R., Hastie, T., Narasimhan, B., Soltys, S. G., Shi, G., Koong, A. C., and Le, Q. T. (2004) Sample classification from protein mass spectrometry, by 'peak probability contrasts'. *Bioinformatics* 20, 3034-3044

44. Stretton, A. O., Fishpool, R. M., Southgate, E., Donmoyer, J. E., Walrond, J. P., Moses, J. E., and Kass, I. S. (1978) Structure and physiological activity of the motoneurons of the nematode Ascaris. *Proc Natl Acad Sci U S A* 75, 3493-3497

45. Yew, J. Y., Kutz, K. K., Dikler, S., Messinger, L., Li, L., and Stretton, A. O. (2005) Mass spectrometric map of neuropeptide expression in Ascaris suum. *J Comp Neurol* 488, 396-413

46. Dekeyser, S. S., and Li, L. (2006) Matrix-assisted laser desorption/ ionization fourier transform mass spectrometry quantitation via in cell combination. *Analyst*131, 281-290

47. Kindy, J. M., Taraszka, J. A., Regnier, F. E., and Clemmer, D. E. (2002) Quantifying peptides in isotopically labeled protease digests by ion mobility/time-of-flight mass spectrometry. *Anal Chem*74, 950-958

48. Barr, J. M., and Van Stipdonk, M. J. (2002) Multi-stage tandem mass spectrometry of metal cationized leucine enkephalin and leucine enkephalin amide. *Rapid Commun Mass Spectrom*16, 566-578

5.8 Figures

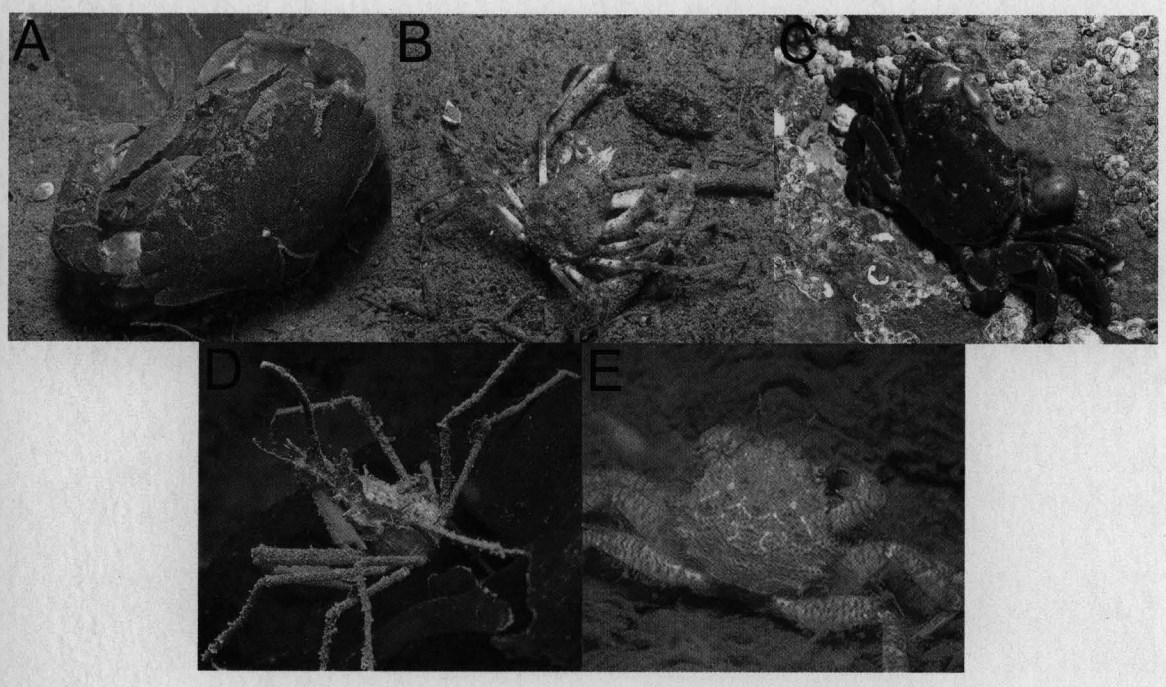

Figure 5.1: The crustaceans used for this research include: (**A**) *Cancer productus* or the Red Rock Crab, (**B**) *Hyas lyratus* or the Pacific Lyre Crab, (**C**) *Hemigrapsus nudus* or the Purple Shore Crab, (**D**) *Oregonia gracilis* or the Graceful Decorator Crab, and (**E**) *Telmessus cheiragonus* or the Helmet Crab. Images taken from *Pacific Coast Crabs and Shrimps*, by Gregory C. Jensen, (c) 1995 Sea Challengers (Monterey, CA).

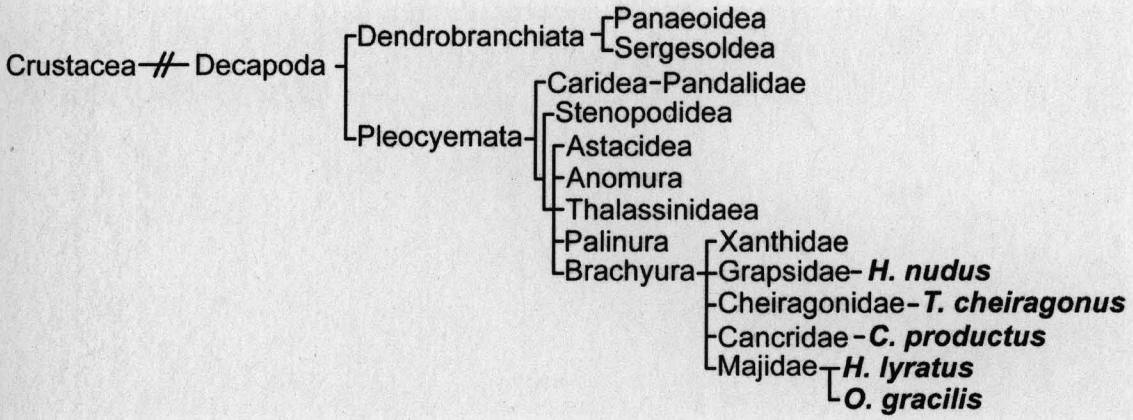

Figure 5.2: Illustration of the location of the crustacean species used in this research on the taxonomic table. Note that all species examined are from the brachyuran family of decapod crustaceans.

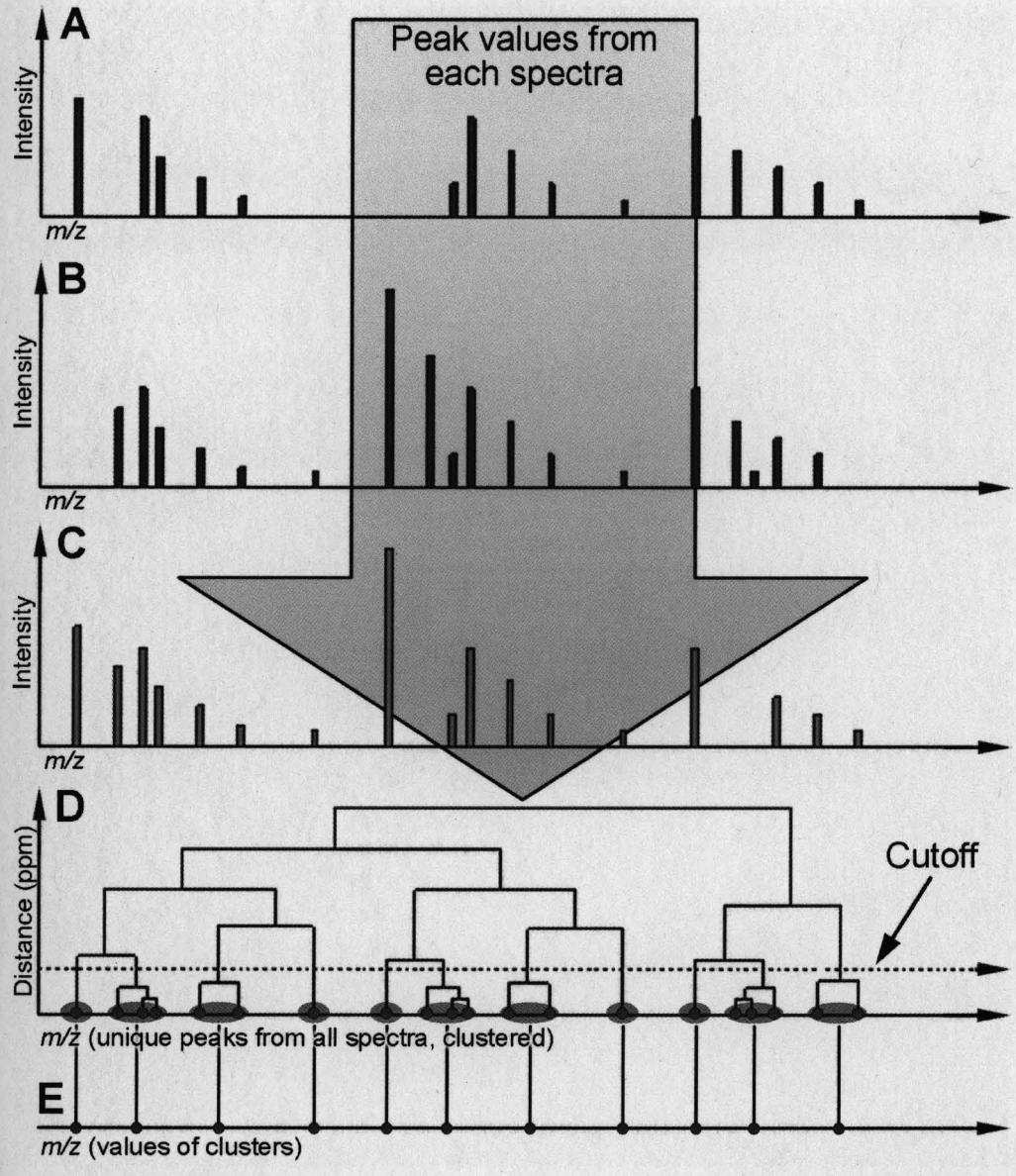

Figure 5.3: An example of how hierarchical clustering is performed. **A**, **B**, and **C** are three separate theoretical peak lists for this example. The points on the x-axis in **D** are the *m/z* values for spectra **A**, **B**, and **C** and each is initially placed into its own cluster (blue, under red). The clustering algorithm then iteratively finds the two closest clusters and combines them to make a new cluster. In **D**, these new cluster are indicated by lines connecting the points where the ppm required to make the cluster is indicated by the height of the line joining the

cluster points. Using a cutoff (at a user entered ppm value) determines the number of clusters, i.e. those clusters that are connected under the cutoff line are grouped together (larger red circles in **D**) and vice versa for those that are connected above the cutoff line. In **E**, the value of each of these clusters is then determined by a user entered method (i.e. average value of peaks in cluster, the value of the most intense peak) and a peak list is created. Intensity can be included if necessary.

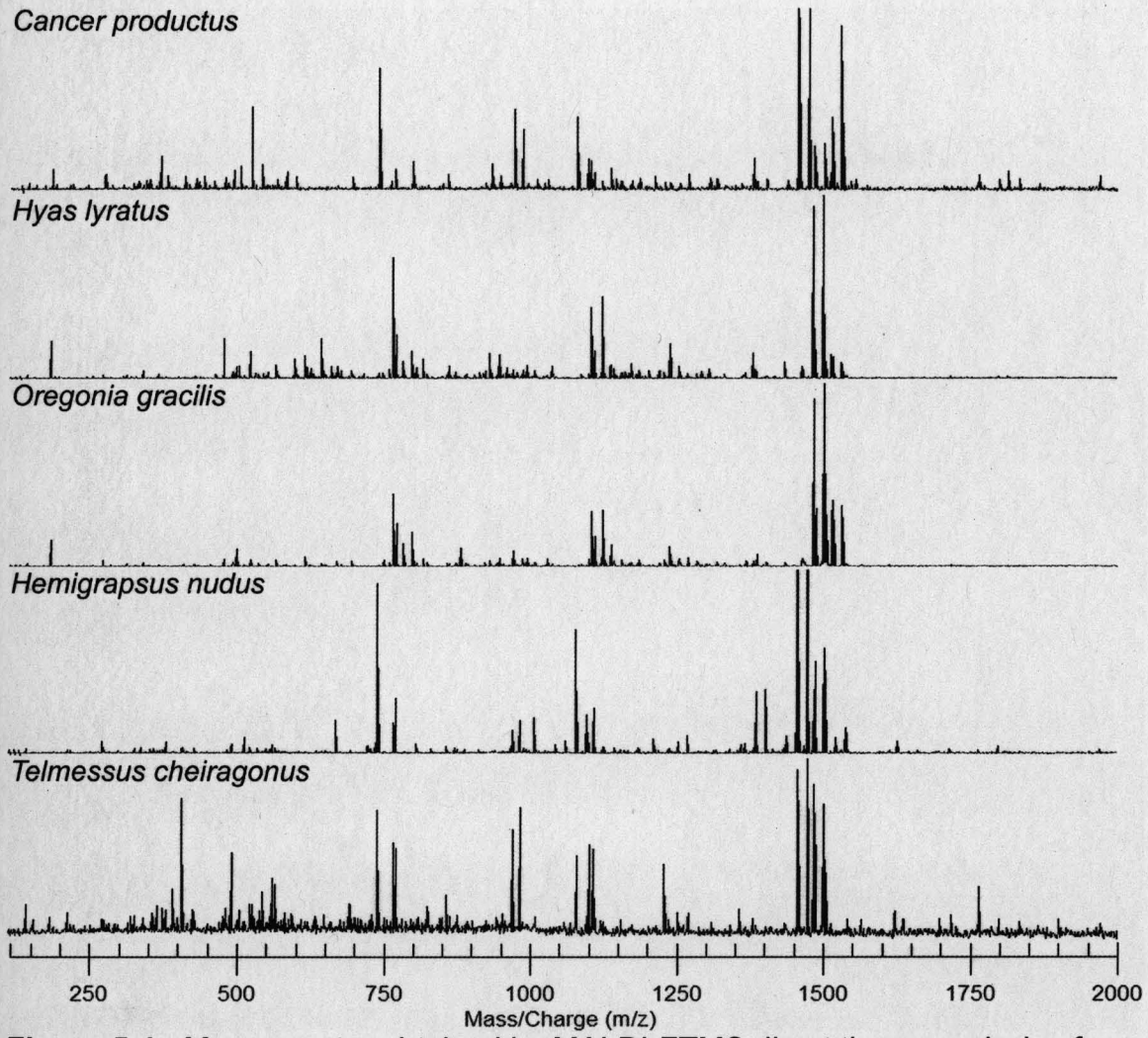

Figure 5.4: Mass spectra obtained by MALDI-FTMS direct tissue analysis of small sections of thoracic ganglia dissected from each species, stored in acidified methanol and then spotted with concentrated DHB. As can be seen, there are both similarities and differences between the species' peptide content.

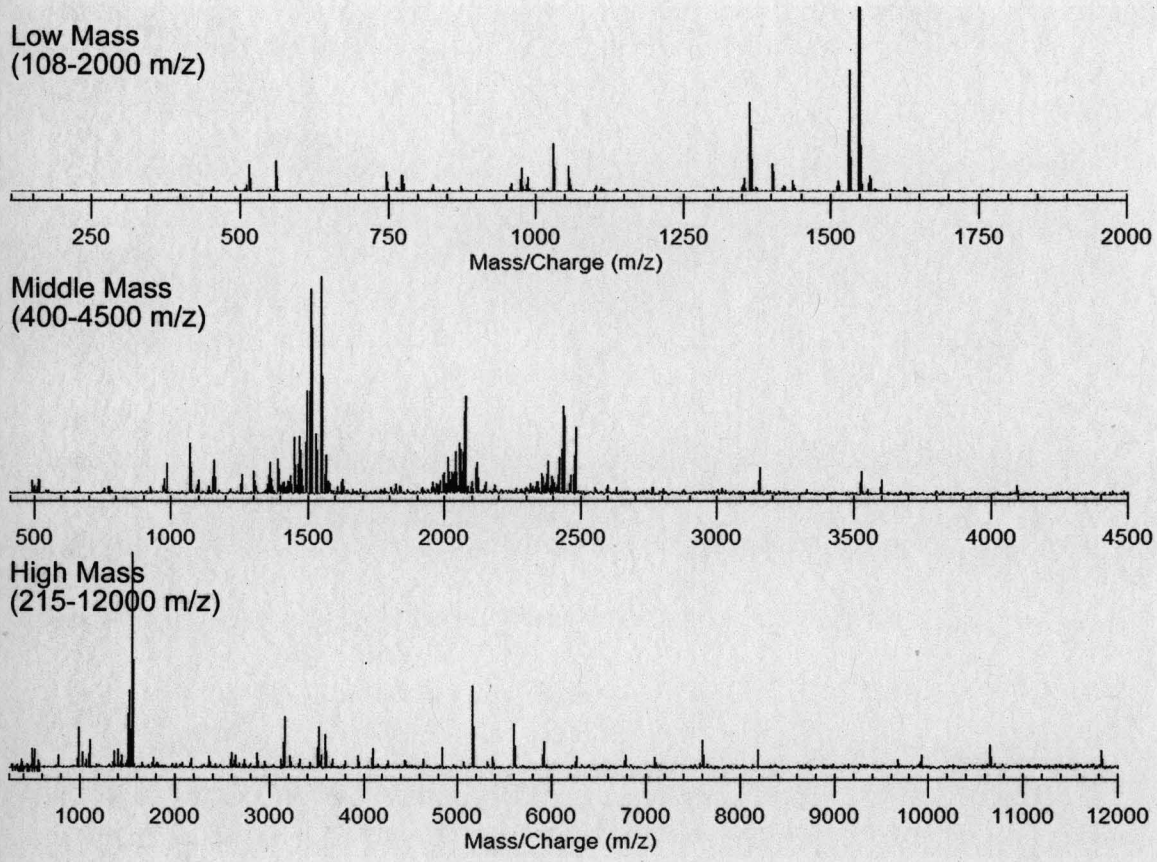

Figure 5.5: An example of the three mass ranges acquired in order to more fully detect all peptides in each fraction. These spectra were all obtained from *H. lyratus* thoracic ganglia extract LC fraction number fifteen.

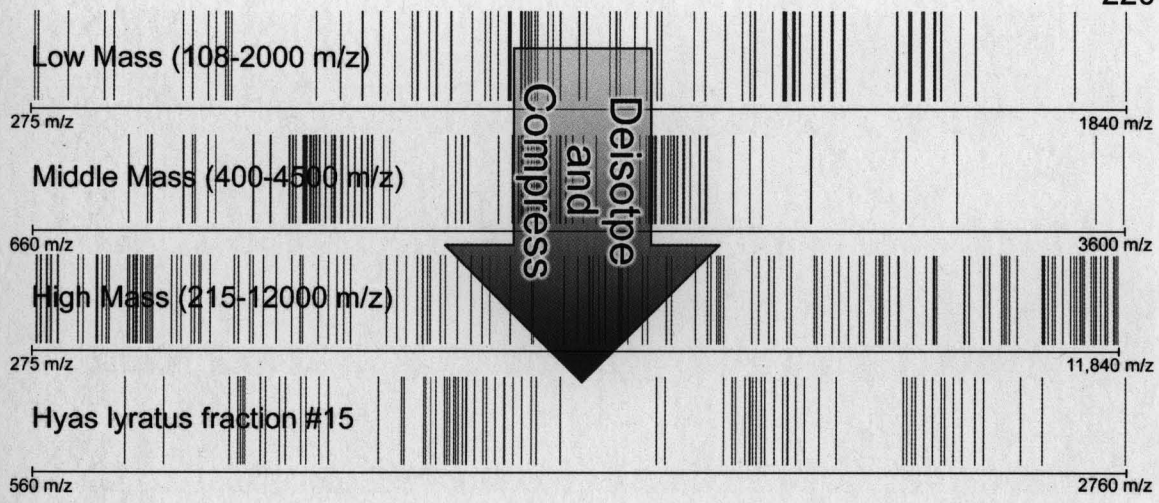

Figure 5.6: Illustration of how peak lists are displayed in SpecPlot; blue lines represent the least intense peaks, green represents medium intensity peaks, and red represents the most intense peaks in a spectrum. This figure illustrates how each of the different mass range spectra are deisotoped and compressed into one peak list representing all peaks for that fraction. Please note the *m/z* scale does not match the different mass ranges because no peaks were identified below the lowest and highest mass indicated by the scale bar.

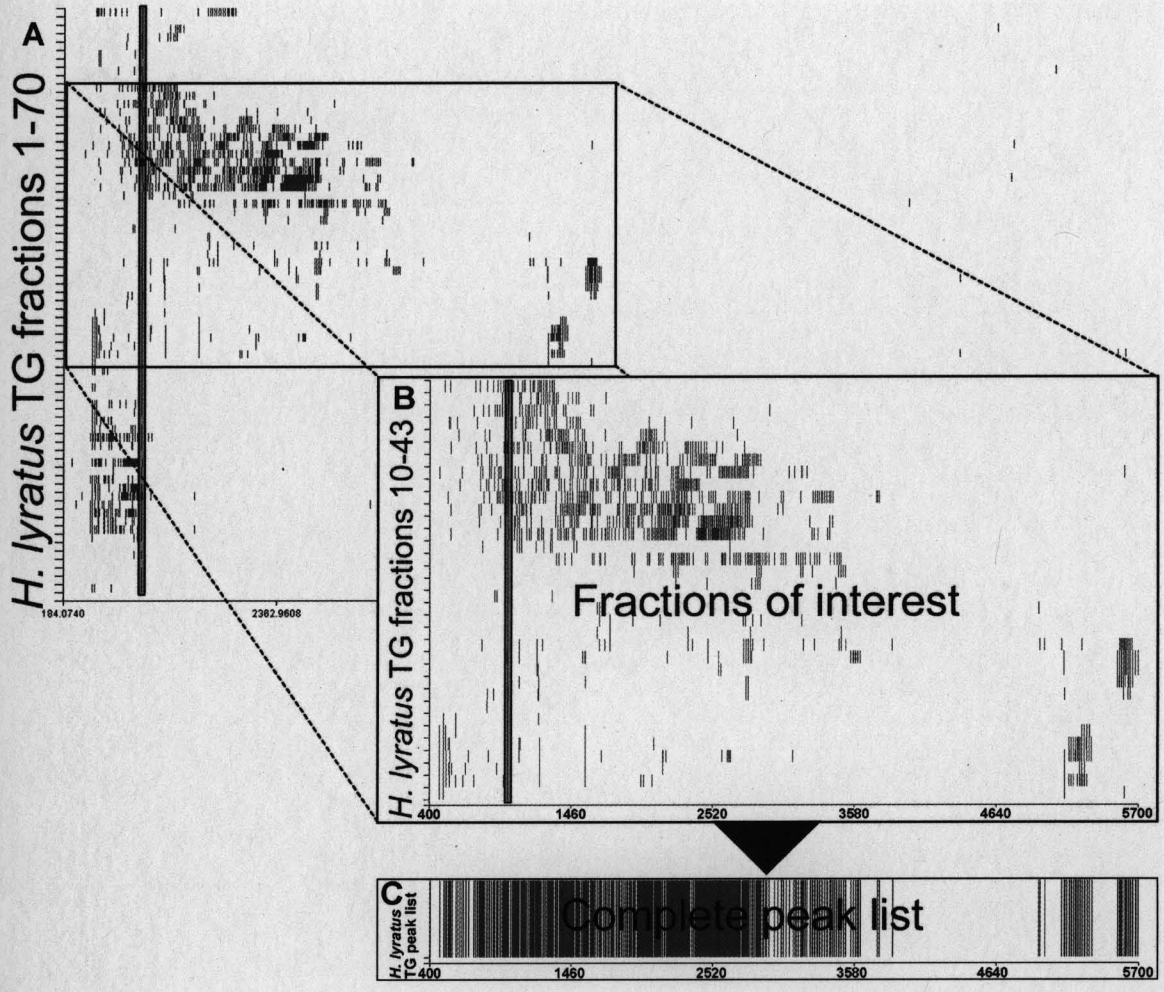

Figure 5.7: Each fraction's peak list is stacked in order (**A**). Also note the peaks at *m/z* 986 (highlighted in red box in **A** and **B**) are found in all fractions, indicating an electronic noise peak that was not eliminated by deisotoping due to its complexity. Because fraction number is related to elution time, the period where peptides eluted off of the column can be observed (**B**). The compressed peak lists from fractions 8-30 (**C**) of the species extract have been combined to create this complete peak list of all peaks detected in the TG extract.

Peptide	Theoretical Mass	Peptide	Theoretical Mass	Peptide	Theoretical Mass
FLRFamides		**Allatostatins A-type**		**Orcokinins**	
RSFLRFa	824.49	YAFGLa	569.31	EIDRSGFGFA	1098.52
RNFLRFa	851.50	YSFGLa	585.30	EIDRSSFGFN	1171.54
GRNFLRFa	908.52	GQYAFGLa	754.39	NFDEIDRSGF	1199.53
SKNYLRFa	926.52	GGAYSFGLa	770.38	DEIDRSGFGFA	1213.55
NRSFLRFa	938.53	DPYAFGLa	781.39	NFDEIDRSGFG	1256.55
pERNFLRFa	962.53	EPYAFGLa	795.40	NFDEIDRSGFA	1270.57
NRNFLRFa	965.54	AGPYSFGLa	810.41	NFDEIDRSSFG	1286.56
GGRNFLRFa	965.54	GDPYAFGLa	838.41	NFDEIDRSSFA	1300.58
DRNFLRFa	966.53	AGGAYSFGLa	841.42	NFDEIDRSGFGF	1403.62
APRNFLRFa	1019.59	PSMYGFGLa	884.43	NFDEIDRSGFGFA	1474.66
GNRNFLRFa	1022.56	PDMYGFGLa	898.41	NFDEIDRSGFGFV	1502.69
RDRNFLRFa	1122.63	GSGQYAFGLa	898.44	DFDEIDRSGFGFV	1503.68
GYSKNYLRFa	1146.61	PDMYAFGLa	912.43	NFDEIDRSGFGFN	1517.67
APQRNFLRFa	1147.65	pERAYSFGLa	923.47	NFDEIDRSSFGFV	1532.70
AYNRSFLRFa	1172.63	SSGQYAFGLa	928.45	NFDELDRSGFGFH	1540.68
SENRNFLRFa	1181.62	PRDYAFGLa	937.49	NFDEIDRSSFGFN	1547.68
PELDHVFLRFa	1271.69	QRAYSFGLa	940.50		
		pERPYSFGLa	949.49	**CPRP**	
RYamides		DRPYSFGLa	953.48	Cabo CPRP III	3958.02
FVGGSRYa	784.41	pERTYSFGLa	953.48	Cabo CPRP I	3963.05
FYANRYa	832.41	APQPYAFGLa	962.51	Capr CPRP I	3975.98
FYSQRYa	862.42	ERPYSFGLa	967.50	Cabo CPRP II	3977.07
LFVGGSRYa	897.49	PADLYEFGLa	1023.51	Capr CPRP III	3988.02
RFVGGSRYa	940.51	PATDLYAFGLa	1066.56	Cabo CPRP IV	3991.08
SGFYANRYa	976.46			Capr CPRP II	4003.06
pEGFYSQRYa	1030.47	**Allatostatins B-type**		Capr CPRP IV	4017.08
SSRFVGGSRYa	1114.58	NWNKFQGSWa	1165.55		
		TSWGKFQGSWa	1182.57	**Truncated CPRPs**	
Orcomyotropin		GNWNKFQGSWa	1222.58	PGGLVHPVE	904.49
FDAFTTGFGHS	1186.52	NNWSKFQGSWa	1252.59	RSAQGMGKMEHL	1344.65
FDAFTTGFGHN	1213.53	STNWSSLRSAWa	1293.63	RSAQGMGKMERL	1363.69
		NNNWSKFQGSWa	1366.63	RSAQGMGKMEHLL	1457.74
PDHs				RSAQGMGKMEHLLA	1528.77
NSELINSILGLPKVMNDAa	1927.03	**Allatostatin Combos**		[13-38] Capr CPRP II	2658.39
NSELINSILGLSRLMNEAa	1973.05	DPYAFGLGKRPADL	1519.79		
		EPYAFGLGKRPATDL	1634.85	**Others**	
		DPYAFGLGKRPDMYGFGLa	2002.98	RYLPT	649.37
		DPYAFGLGKRPDMYAFGLa	2017.00	DFSAWAa	695.31
		GSGQYAFGLGKKAGGAYSFGLa	2035.04	HIGSLYRa	844.48
		DPYAFGLGKRPADLYEFGLa	2128.09	pELNFSPGWa	930.45
		EPYAFGLGKRPATDLYAFGLa	2185.14	APSGFLGMRa	934.49
				PFCNAFTGCa	956.38
				GFKNVEMMTARGFa	1486.73

Table 5.1: List of known crustacean neuropeptides

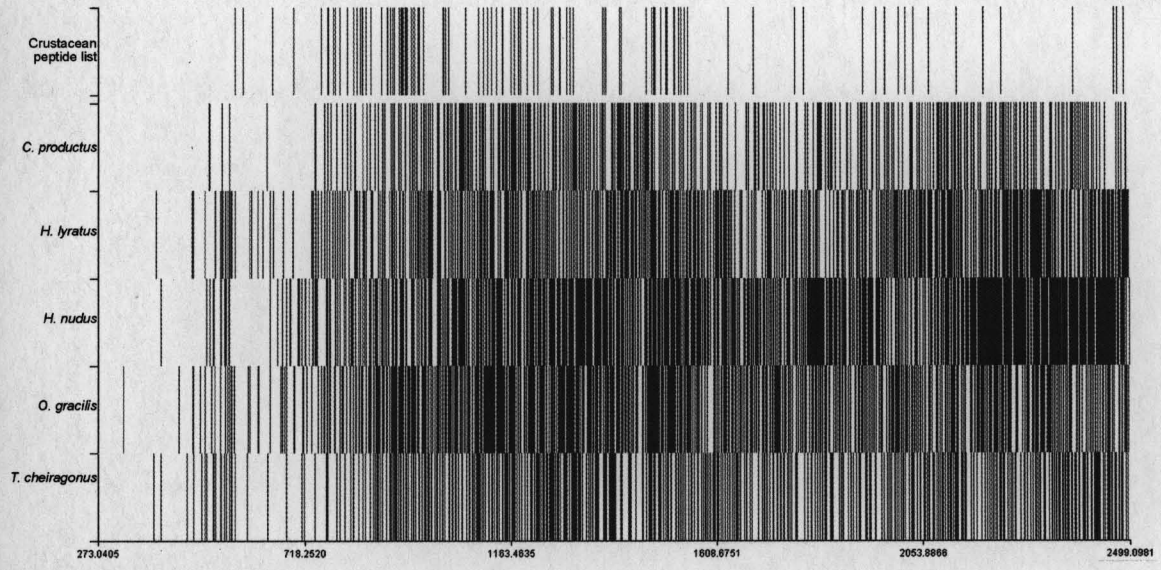

Figure 5.8: Peak lists of each species stacked to show similarities and differences between them. This display has been cut-off at *m/z* 2500 for display purposes. For data analysis and peak matching, all peaks, including those over *m/z* 2500 were used.

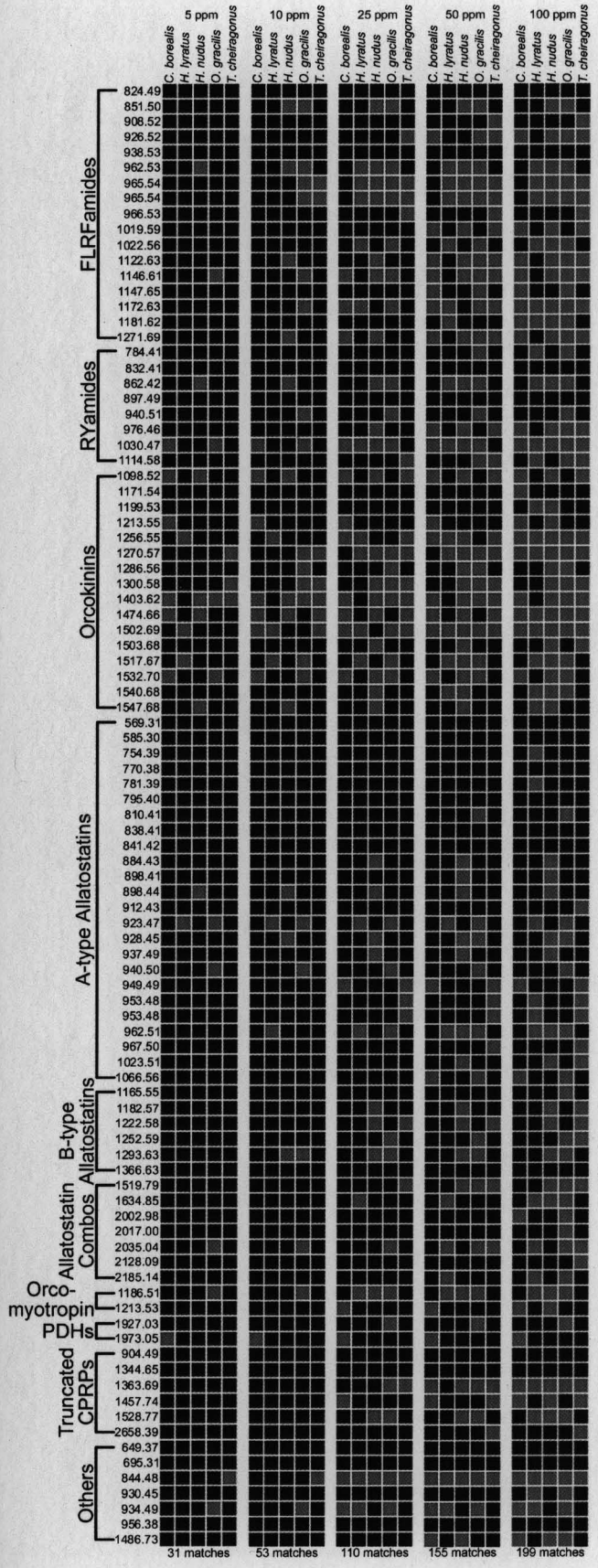

Figure 5.9: A comparison of each species to the list of known crustacean peptides (**Table 1**) is displayed as a heat map where the blue boxes indicate a match and black boxes indicate no match. Using SpecPlot we are able to designate the window of error allowed for these matches, as is indicated above each column.

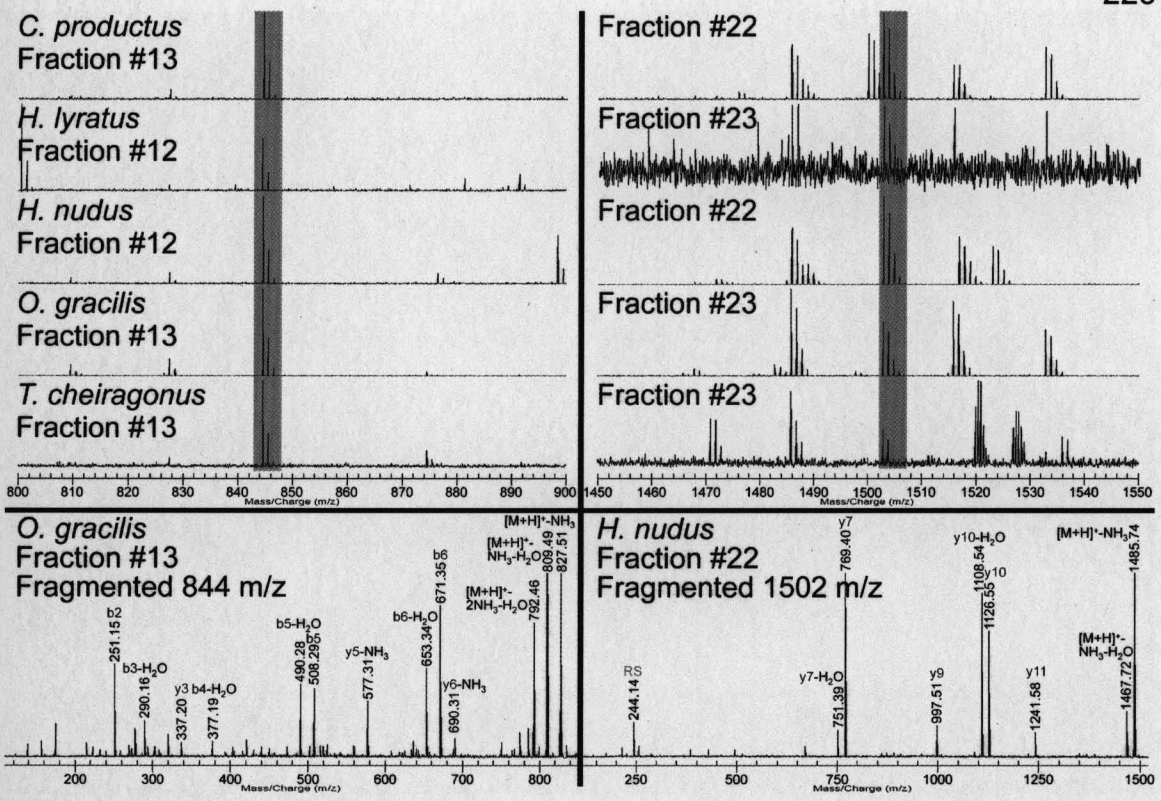

Figure 5.10: Verification and confirmation of putative peptides via fraction alignment and gas-phase fragmentation. As can be seen on the left, a peak with *m/z* 844.5 was identified in all five species in almost the same fraction for each. On the right, the peak *m/z* 1502.7 was identified in all five species, also in almost the same fraction. On the bottom, to confirm the identity of each of these peptides, MS/MS via SORI-CID was performed on the fraction with the most abundance of each peptide and sequence-specific fragment ions and peptide tags labeled.

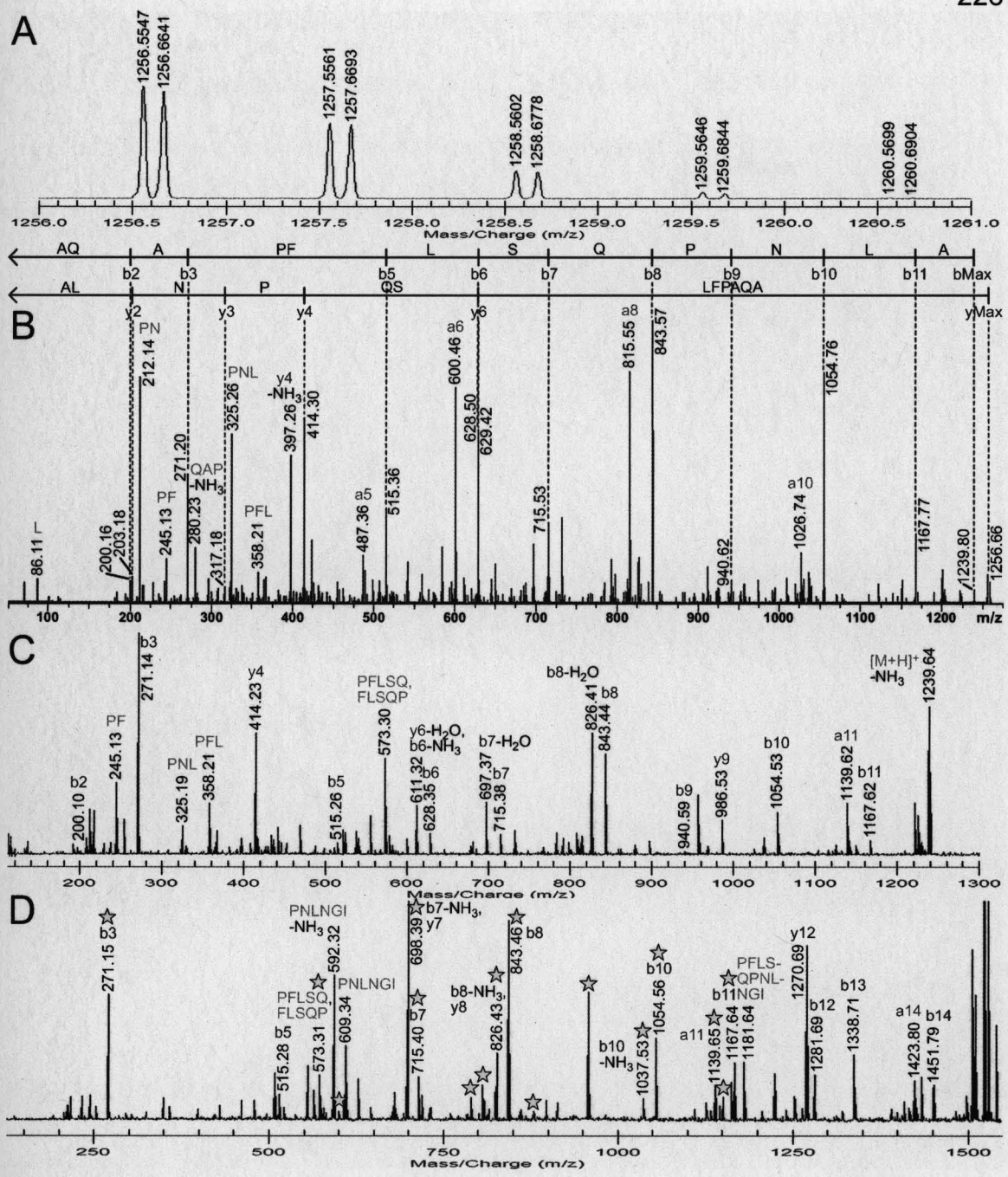

Figure 5.11: **A.** Spectrum of an internal calibration on an adjacent standard (InCAS) experiment with native (confirmed) [1-11] orcokinin with the new *m/z* 1256 peptide showing a definite difference in mass. **B.** The CapLC Q-TOF fragmentation spectrum of the new *m/z* 1256 showing the corresponding

sequence. **C.** The MALDI-FTMS fragmentation spectrum of the new *m/z* 1256 showing the peptide sequence. **D.** The MALDI-FTMS fragmentation of the new *m/z* 1540 showing the similarities to the new *m/z* 1256 (indicated by the yellow stars) and the continuation of the sequence.

Chapter Six

Comparative Analysis of Tissue-Specific Expression of Neuropeptides Using LC-MALDI TOF/TOF Mass Spectrometry

This chapter is coauthored by Sean McIlwain and David Page of the University of Wisconsin Department of Biostatistics and Lingjun Li.

6.1 Introduction

Peptidomics has evolved into a highly active research area due to substantial advancements in mass spectrometry and biological significance of neuropeptides. Interest is especially on bioactive peptides that are involved in cell/cell signaling as well as systemic signaling.[1; 2] One of the organisms that is of special interest that has a model nervous system is the crustacean *Cancer borealis* (the Jonah crab). There are many reasons that *C. borealis* is used as a model system and one of the most important is the peptide diversity contained within.[3; 4] Although many peptides in multiple peptide families have been identified in *C. borealis*, functions of many of these peptides are still have not been fully described and remain a mystery.[5-7]

In addition to a lack of knowledge about function, there is also a lack of data regarding peptide distribution among the nervous system and neurosecretory organs. While some research has described the peptide content of individual or a few organs[8-10], there has never been a thorough investigation into all of the organs simultaneously. Additionally, the exploration of peptidomes of all the organs can provide a great deal of important information, but even more significant is a comparison of the peptide content between all of the organs. By performing such a comparative peptidomic analysis, and tying this information to the known function of each organ, some conclusions about peptide function can be derived.

In this report, we describe the first full investigation of peptide content from all of the important nervous system and neurosecretory organs in the crab *C.*

borealis. Mass spectrometry analysis was carried out using a MALDI TOF/TOF instrument.[11] Previously, we described the comparison of multiple crab species' thoracic ganglia peptidome through the combination of a MALD FTMS instrument with bioinformatics.[16] For this research we chose to use a MALDI TOF/TOF due to its sensitivity and its high throughput capability. Furthermore, because we utilized off-line LC MALDI analysis, we are able to reanalyze the fraction spots containing peptides of interest. This allows for the opportunity to perform MS/MS experiments to obtain fragmentation data and enable either known peptide confirmation or novel peptide sequencing.[12; 13] Moreover, by using an in-house written bioinformatics software package to process the MS data we have compared each organs' peptidome to a list of known peptides. This combination of MALDI TOF/TOF with bioinformatics provides a sensitive and rapid method of identifying known neuropeptides from a large dataset of varied samples.[14]

6.2 Methods

Animal care and tissue collection

Cancer borealis (Jonah crabs) were obtained from the Fresh Lobster Company (www.thefreshlobstercompany.com) in Gloucester, MA and were maintained without food in a recirculated artificial seawater aquarium at 10-14 °C until being dissected. Details of the animal treatment and dissection are described in Kutz et al.[17] For the purposes of this study, the crabs were cold-anesthetized, and the organs were dissected in chilled physiological saline

(composition in mM: NaCl, 440; KCl, 11; MgCl2, 26; CaCl2, 13; Trizma base, 11; maleic acid, 5; pH 7.45). Organs were removed from each animal and immediately placed into acidified methanol (90:9:1, MeOH : glacial acetic acid : deionized H_2O) in microcentrifuge tubes. The number of organs and what part of the crab they were removed from are as follows: two sinus glands (SG) from the eyestalks; two anterior cardial plexi (ACP), the brain, two commissural ganglia (CoG), the oesophageal ganglia (OG), and the stomatogastric ganglia (STG) from the stomatogastric nervous system (STNS); two pericardial organs (PO) from the cardial ridges; and the thoracic ganglia (TG) from ventral to the cardial chamber. Organs were combined into their own microcentrifuge tubes (one for each type of organ) and then stored in a -20 °C freezer until peptide extraction.

Peptide Extraction and HPLC Separation

Microcentrifuge tubes containing organs were removed from the -20 °C freezer and immediately placed on ice for peptide extraction. Organs from 25 animals were then removed from the tube and placed into an appropriately sized tissue grinder (Wheaton, Inc.) along with an adequate amount of acidified methanol to perform the extraction. For the ACP, OG, STG, and SG a 100 μl tissue grinder was used and 100 μl of acidified methanol was used for each extraction and reextraction. The brain, CoG, and PO required a 1 ml tissue grinder and 1 ml of acidified methanol was used for each extraction and reextraction. Finally, a 15 ml tissue grinder was used to extract the TGs along with 7 ml of acidified methanol for each extraction and reextraction. Each organ

was then ground until no apparent large particles remained and all tissue was suspended. The acidified methanol with the suspended tissue was then removed from the tissue grinder and placed into a new microcentrifuge tube(s). The extract was then centrifuged at 16,100 rcf for 10 minutes on an Eppendorf 5415 D microcentrifuge (Brinkmann Instruments Inc., Westbury, NY). After centrifugation, the supernatant was removed and placed into a new micro-centrifuge tube(s). Following this, the appropriate amount of acidified methanol was used to rinse the tissue grinder after which the acidified methanol was transferred to the tube(s) with the original pellet(s). The pellet(s) were then resuspended, reextracted, centrifuged again and the supernatant removed and added to the original supernatant. This process was repeated again producing the final extract. The pellets were retained and stored in a -20 °C freezer. The extracts were then concentrated using a Savant SC 110 Speed-Vac concentrator (Thermo Electron Corporation, West Palm Beach, FL). For extracts with volumes higher than 2 ml and thus in multiple tubes, the volume was observed during concentration and combined when able to all fit in one tube. The extracts for the ACP, CoG, OG, STG, and SG were then resuspended with ~50 μl of ddH_2O with 0.1% formic acid (Sigma Aldrich) and the brain and PO and were resuspended with 80 μl of ddH_2O with 0.1% formic acid all followed by vigorous vortexing. The TG extact was resuspended by taking 20 μl of the extract and adding 80 μl of ddH_2O with 0.1% formic acid and vigorous vortexing.

After resuspension, extracts were ready for HPLC separation. For HPLC separation, a 100 μl injection loop was used, but not always filled with extract.

Injection volumes were as follows: 30 μl was injected for ACP, CoG, OG, STG, and SG, 50 μl was injected for brain and PO, and 100 μl for the TG extract. The column used was a 2.1 mm I.D., 250 mm long, C18 column (Alltech, Inc.) with a guard column on a Rainin Dynamax HPLC system (Rainin Instrument Inc., Woburn, Massachusetts, USA) equipped with a Dynamax UV-D II absorbance detector. Mobile phases used were A: dH_2O w/ 0.1% formic acid and B: acetonitrile (HPLC grade) w/ 0.1% formic acid. The gradient used was 5-70% B from 0-60 minutes, 70-95% B from 60-70 minutes, and 95% B from 70-80 minutes with a flow rate of 0.2 ml/min. Fractions were collected every two minutes with a Rainin Dynamax FC-4 fraction collector producing 40, 400 μl fractions. Fractions were then concentrated to dryness with the speed-vac. Each fraction was then resuspended with 20 μl ddH_2O w/ 0.1% formic acid.

Data Acquisition and Processing

After fractions were resuspended, they were spotted on an ABI Opti-TOF 384-well insert MALDI target plate (Applied Biosystems, Framingham, MA). A 0.5 μl spot of α-cyano-4-hydroxycinnamic acid (CHCA (Sigma Aldrich Inc., St. Louis, MO), saturated solution in 50:50 acetonitrile/0.1% aqueous formic acid, v/v) was placed on the plate and then 0.5 μl of each fraction was added to the spot and then allowed to cocrystallize at room temperature. In addition to the fractions, a 10^{-4} M mixture of six standards was spotted after every 10 fraction spots. The standards used can be found in **Table 6.1**. The standard spots were placed to provide external calibration data if necessary.

A model 4800 MALDI TOF/TOF analyzer (Applied Biosystems, Framingham, MA) equipped with a 200 Hz, 355 nm Nd:YAG laser (spot diameter of ~75 μm) was used for all mass spectral analyses. Acquisitions were performed in positive ion reflectron mode. Instrument parameters were set using the 4000 Series Explorer Software (Applied Biosystems). Each mass spectrum was generated by averaging 1200 laser shots over the mass range m/z 500-4000. Mass spectra were externally calibrated using peptide standards applied directly to the stainless steel MALDI target.

Data Analysis and Software Development

Spectra were analyzed using Data Explorer© Software v. 4.9 (Applied Biosystems). Each fraction's spectrum was reviewed and the peak list generated by the software was converted to a text file that could be imported into the in-house developed bioinformatics software, SpecPlot.

After doing this, it was necessary to produce a peak list of all peaks in all the relevant fractions from each of the organ extracts. Therefore, the peak lists from multiple spectra had to be combined into one peak list representing the peptides in each extract. We accomplished this using hierarchical clustering upon the *m/z* values across different peak lists using an algorithm similar to Tibshirani et al, 2004.[15] For further explanation of hierarchical clustering and this algorithm, please see Chapter 5. This method was used to compress all of the organs' fraction peak lists into one large peak list indicative of all peaks found in an organs' extract.

After the peak lists were created for each organ, it was necessary to align the peak lists because the spectra were externally calibrated. To perform this alignment, a list of alignment peaks was created of *m/z* values for the standards in the calibration spots and included in with the organs' peak lists (**Table 6.1**). Because four spectra of the standards were included in each organs' complete peak list, alignment on these peaks was possible. The method used for this alignment is also described in our previous work.[16] Instead of using peaks naturally found in the peak lists as in our species comparison work, we were able to use the standard peaks from spectra collected during the analysis of the fractions. Using the list of standard peaks, a least-squares linear fit is performed to derive the constants for the correction equation (Corrected = A + B×(Original) where A is the absolute *m/z* shift and B is the adjustment of the scale as the *m/z* changes). The corrected value for one peak's *m/z* value is calculated using the linear equation from the peak's original *m/z* value. The algorithm calculates the values of A and B for each spectrum, then corrects the *m/z* values of the peaks within that spectrum using the linear equation. Also included in this algorithm is an error value to create a tolerance window on either side of the entered standard masses. For this work we used an error value of 100 ppm to create the error window for this alignment algorithm. Because we were using a MALDI TOF/TOF instrument and externally calibrating, we chose 100 ppm as our tolerance error.

Once we have built the composite peak list for each organ being studied, we can begin identifying known peptides. This is done by first entering a list of

m/z values for any known peptides we are interested in identifying in the organs. By entering a +/- ppm value, an error tolerance window is created, with the known peptide *m/z* values at the midpoints. The organs' peak lists are then compared to these windows and if a peak exists within the windows, identification is indicated by a blue square in the resulting heat map, whereas a black square results from no identification being made. By doing this, we can indicate which peptide *m/z* peaks are present within each organs' composite mass spectrum.

6.3 Results and Discussion

We began this research by collecting and pooling organs from 25 *C. borealis* crabs. As the focus of this research was to identify all of the peptides contained in each of these organs, a direct tissue analysis of these organs would not suffice to fully elucidate the complete peptide complement of each organ. Therefore these organs were pooled and their peptides were extracted. The extracts were then concentrated and resuspended. Afterwards, the extracts were HPLC separated using a gradient method that produced 40 fractions for each organ extract. Many of the HPLC traces showed a number of peaks (**Figure 6.1**) and therefore it was obvious that most fractions contained many peptides. To determine the peptides in those fractions, mass spectral analysis was carried out using a MALDI-TOF/TOF instrument.

By using this instrument, we were able to analyze all fractions rapidly and with high sensitivity. Furthermore, because the TOF/TOF can sensitively analyze a wide mass range, each fraction was analyzed from 400-4000 m/z. By

analyzing each fraction with this range, the majority of peptides will be detected. In addition to analyzing each fraction, we also analyzed a spot containing six standards that had been spotted after every ten fraction spots (**Figure 6.2**). These standard spots all contained non-crustacean peptides enabling us to include their spectra in with the fractional spectra when they were combined into a complete peptide list for each organ. By including these standard peaks with the native organ's peptide peaks, we were able to align (calibrate) the spectra on the six standard peaks of known masses.

For many of the organs a great number of peptides were detected. This is obvious if one observes several of the middle fractions from most of the organs' extracts. Examples of several of the most peptide rich fractions from the PO extract are shown in **Figure 6.3**. Although most organs did show a large number of peptides, due to small organ size and lower peptide abundances, the OG and ACP did not exhibit the high quantity of peptides that the other organs did (**Figure 6.4**). Even so, several peptide identifications were made for each of these organs.

One factor to consider when performing an analysis of such varying tissue types (from huge clusters of neurons such as in the TG to axonal lengths in the ACP) the consistency of the HPLC separation may come under scrutiny. As has been stated in the methods, all organ extracts were separated using the same gradient method and all produced impressive HPLC traces as detected by the UV-Vis detector at 286 nm. A demonstration of this is the fact that the peptide HIGSLYRa with a mass of *m/z* 844.48 was identified in similar fractions in all

organs except the OG (**Figure 6.5**). This resemblance of fractional content indicates the relative consistency of the HPLC separations despite the differences in tissue types. It also highlights the wide distribution of this peptide throughout *C. borealis* nervous system.

After producing peak lists for each organ and aligning them with the included standard peaks, we compared the peak lists to a list of known crustacean peptides (**Table 6.2**). For this comparison, we used a 100 ppm window as the tolerance error to create the window on either side of the known peptide mass in which peaks needed to be to count as identification. This error was chosen by taking into account the general mass measurement accuracy of MALDI TOF/TOF instruments as well as the error that was observed between the measured and theoretical values of the standard peaks. Because we are not interested in peptide abundance for this investigation, identifications were made regardless of peak intensity. The identifications made are displayed in a heat map where a blue square indicates the presence of that peptide in that organs peak list and a black square indicates the opposite (**Figure 6.6**). Using the 100 ppm error window size we made 274 peptide identifications in these eight nervous system and neurosecretory organs of the crab *C. borealis*.

By observing where certain peptides were identified and knowing something about the function of the organs they were found in, we can gain insight into the function of these peptides. For this to be true, we are working under the safe assumption that a peptide is only present in organ's peptidome if the organ requires that peptide for a part of its function. Therefore, we can make

the conclusion that because so many peptides were identified in the PO each of these is important to the systemic communication of the crustacean nervous system. On the other hand, a peptide that is identified in only one organ, such as RDRNFLRFa (*m/z* 1122.63) found in the brain, could indicate that this peptide is used for a purpose that is specific to that organ. Overall, the understanding of each organ's peptidome and the peptide distribution among organs can lead to an enhanced comprehension of the function of both the organs and the peptides they employ for this function. Additionally, by analyzing the peptidome of these organs using an off-line LC-MALDI approach, it is likely that novel peptides will be discovered that can then be characterized thus leading to an even more complete knowledge base about these organs and peptides.

6.4 Conclusions

As we have shown, through the use of off-line LC-MALDI TOF/TOF analysis combined with bioinformatics, it is possible to rapidly identify the peptidomes of a variety of tissue types. In this case, we have explored the *C. borealis* nervous system and neurohemal organ peptidome and made many peptide identifications in each organ. Altogether, we made 274 known crustacean peptide identifications in the eight organs we studied. These identifications were made using an in-house written bioinformatics software package that allows us the ability to flexibly manipulate necessary values to optimize for the highest number of accurate peptide assignments as possible.

In the future, one area to be developed is the method by which peaks are picked from the TOF/TOF spectra as it has an inherent large amount of noise that is difficult to work with at times. Future work should also be focused on the sequencing and characterization of several unknown peaks that were found in the course of this investigation. Finally, although the HPLC method for these separations worked well, it may be possible to optimize methods for each organ type individually and thus allow for better separation and detection ability. Overall, while this survey has provided a great deal of informative data, there is more that could be done to further extend this research in the future.

6.5 Acknowledgements

We are grateful to the University of Wisconsin Biotechnology Center for access to the MALDI-TOF/TOF instrument. We would like to thank Dr. Michael Nusbaum (University of Pennsylvania School of Medicine) for his generous donation of the Rainin Dynamax HPLC system to the Li laboratory. This research was funded by NSF CAREER award CHE-0449991, Alfred P. Sloan Research Foundation (LL). J.J.S. acknowledges a predoctoral fellowship from the American Foundation for Pharmaceutical Education.

6.6 References

1. Fricker, L. D., Lim, J., Pan, H., and Che, F. Y. (2006) Peptidomics: Identification and quantification of endogenous peptides in neuroendocrine tissues. *Mass Spectrom Rev*25, 327-344

2. Li, L., Pulver, S. R., Kelley, W. P., Thirumalai, V., Sweedler, J. V., and Marder, E. (2002) Orcokinin peptides in developing and adult crustacean stomatogastric nervous systems and pericardial organs. *J Comp Neurol*444, 227-244

3. Skiebe, P. (2001) Neuropeptides are ubiquitous chemical mediators: Using the stomatogastric nervous system as a model system. *J Exp Biol*204, 2035-2048

4. Billimoria, C. P., Li, L., and Marder, E. (2005) Profiling of neuropeptides released at the stomatogastric ganglion of the crab, cancer borealis with mass spectrometry. *J Neurochem*95, 191-199

5. Fu, Q., Christie, A. E., and Li, L. (2005) Mass spectrometric characterization of crustacean hyperglycemic hormone precursor-related peptides (CPRPs) from the sinus gland of the crab, cancer productus. *Peptides*26, 2137-2150

6. Mercier, A. J., Friedrich, R., and Boldt, M. (2003) Physiological functions of FMRFamide-like peptides (FLPs) in crustaceans. *Microsc Res Tech*60, 313-324

7. Chung, J. S., and Webster, S. G. (2006) Binding sites of crustacean hyperglycemic hormone and its second messengers on gills and hindgut of the

green shore crab, carcinus maenas: A possible osmoregulatory role. *Gen Comp Endocrinol*147, 206-213

8. Rubakhin, S. S., Greenough, W. T., and Sweedler, J. V. (2003) Spatial profiling with MALDI MS: Distribution of neuropeptides within single neurons. *Anal Chem*75, 5374-5380

9. Fu, Q., Christie, AE,Li, L. (2005) Mass spectrometric characterization of the crustacean hyperglycemic hormone precursor related peptides (CPRPs) from the sinus gland of the crab, cancer productus. *Peptides*26,

10. Fu, Q., Kutz, K. K., Schmidt, J. J., Hsu, Y. W., Messinger, D. I., Cain, S. D., de la Iglesia, H. O., Christie, A. E., and Li, L. (2005) Hormone complement of the cancer productus sinus gland and pericardial organ: An anatomical and mass spectrometric investigation. *J Comp Neurol*493, 607-626

11. Wei, H., Nolkrantz, K., Parkin, M. C., Chisolm, C. N., O'Callaghan J, P., and Kennedy, R. T. (2006) Identification and quantification of neuropeptides in brain tissue by capillary liquid chromatography coupled off-line to MALDI-TOF and MALDI-TOF/TOF-MS. *Anal Chem*78, 4342-4351

12. Yergey, A. L., Coorssen, J. R., Backlund, P. S.,Jr, Blank, P. S., Humphrey, G. A., Zimmerberg, J., Campbell, J. M., and Vestal, M. L. (2002) De novo sequencing of peptides using MALDI/TOF-TOF. *J Am Soc Mass Spectrom*13, 784-791

13. Zhang, N., Li, N., and Li, L. (2004) Liquid chromatography MALDI MS/MS for membrane proteome analysis. *J Proteome Res*3, 719-727

14. Zucht, H. D., Lamerz, J., Khamenia, V., Schiller, C., Appel, A., Tammen, H., Crameri, R., and Selle, H. (2005) Datamining methodology for LC-MALDI-MS based peptide profiling. *Comb Chem High Throughput Screen*8, 717-723

15. Tibshirani, R., Hastie, T., Narasimhan, B., Soltys, S. G., Shi, G., Koong, A. C., and Le, Q. T. (2004) Sample classification from protein mass spectrometry, by 'peak probability contrasts'. *Bioinformatics* 20, 3034-3044.

16. Schmidt, J.J., Christie, A.E., McIlwain, S.E., Page, D., Li, L. Combining MALDI-FTMS and Bioinformatics for Rapid Peptidomic Comparisons. Manuscript submitted to *J of Protemic Res*, July 2007.

17. Kutz, K. K., Schmidt, J. J., and Li, L. (2004) In situ tissue analysis of neuropeptides by MALDI FTMS in-cell accumulation. *Anal Chem* 76, 5630-5640

6.7 Figures

Peptide Name	Sequence	Theoretical Mass (*m/z*)
FMRFa	FMRFa	599.31225
Bradykinin	RPPGFSPFR	1060.56867
Arginine Vasopressin	CYFQNCPRGa (Disulfide bridge Cys1-Cys6)	1084.44513
Angiotensin I	DRVYIHPFHL	1296.68477
Substance P	RPKPQQFFGLMa	1347.73542
Somatostatin	AGCKNFFWKTFTSC (Disulfide bridge Cys3-Cys14)	1637.72390

Table 6.1: A list of the standards that were spotted after every ten fraction spots. The peaks resulting from these spectra were included in the peak lists for each organ and acted as alignment peaks for external calibration.

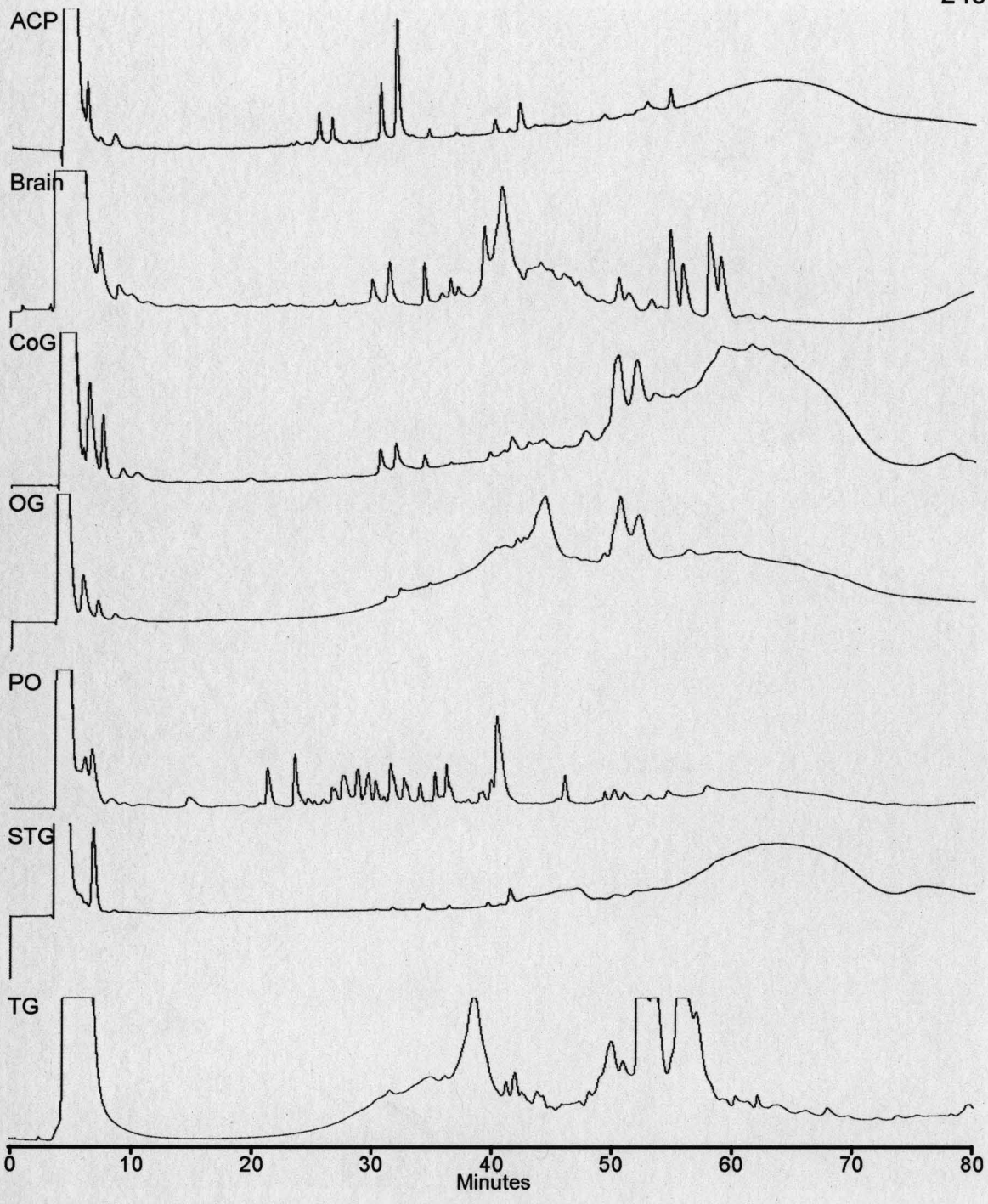

Figure 6.1: HPLC traces for each of the organs indicating the relative peptide content in each organ.

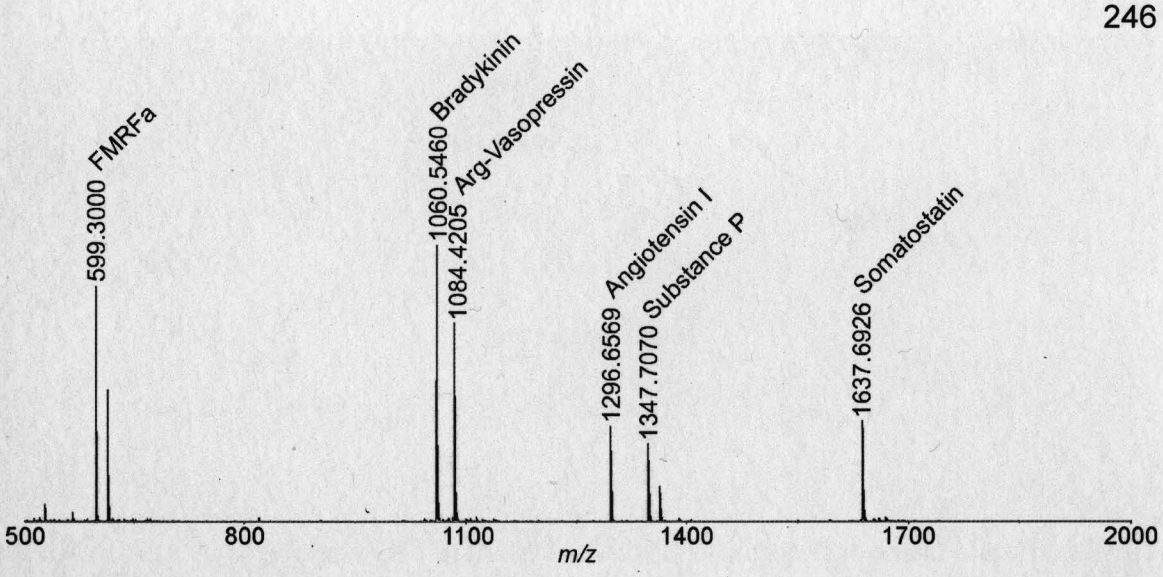

Figure 6.2: An example of one of the spectra of the standard spots spotted after every 10 fraction spots.

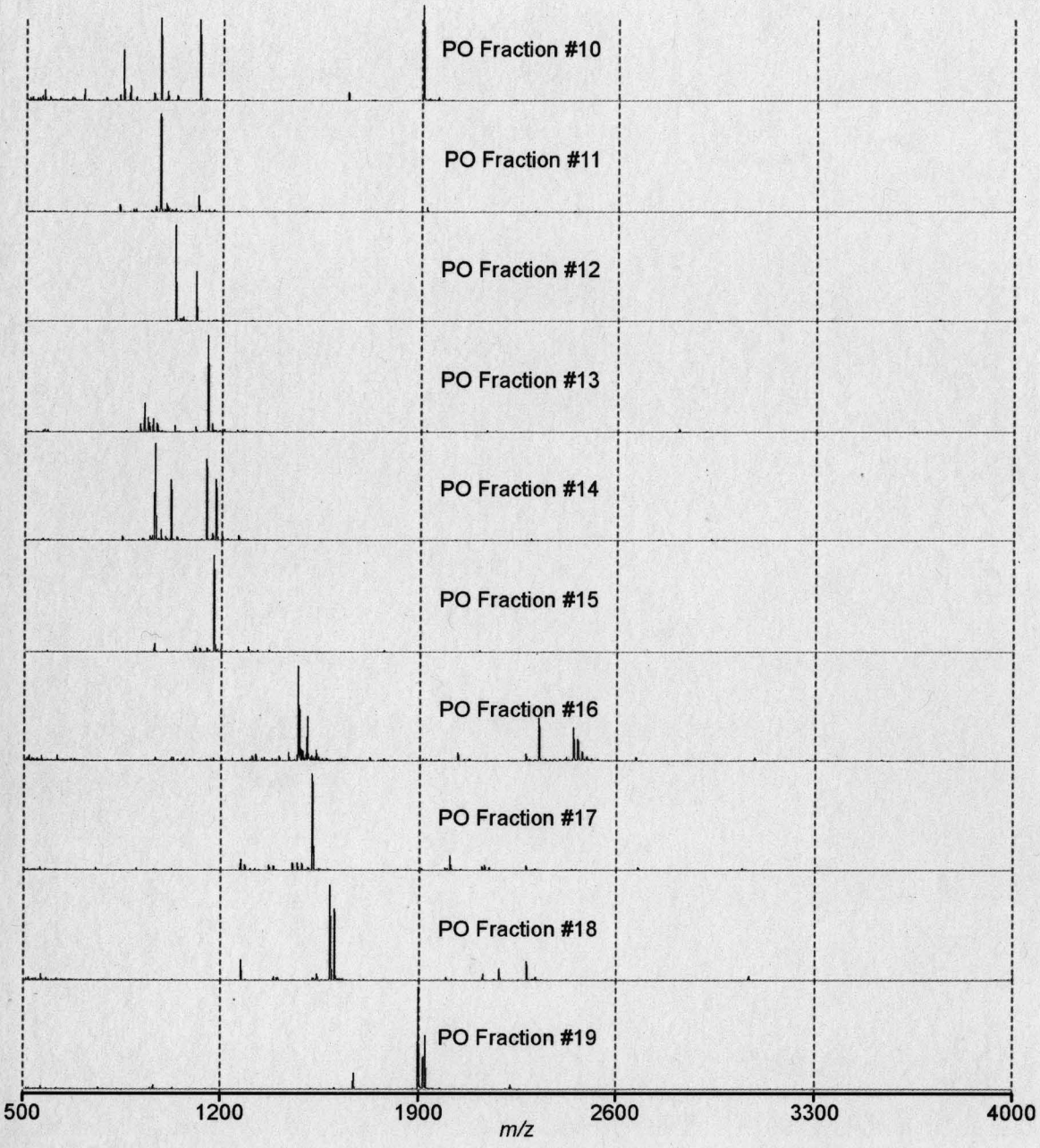

Figure 6.3: Spectra from the PO organ extract fractions 10 through 19 in example of the large number of peptide peaks in many of the fractions' spectra.

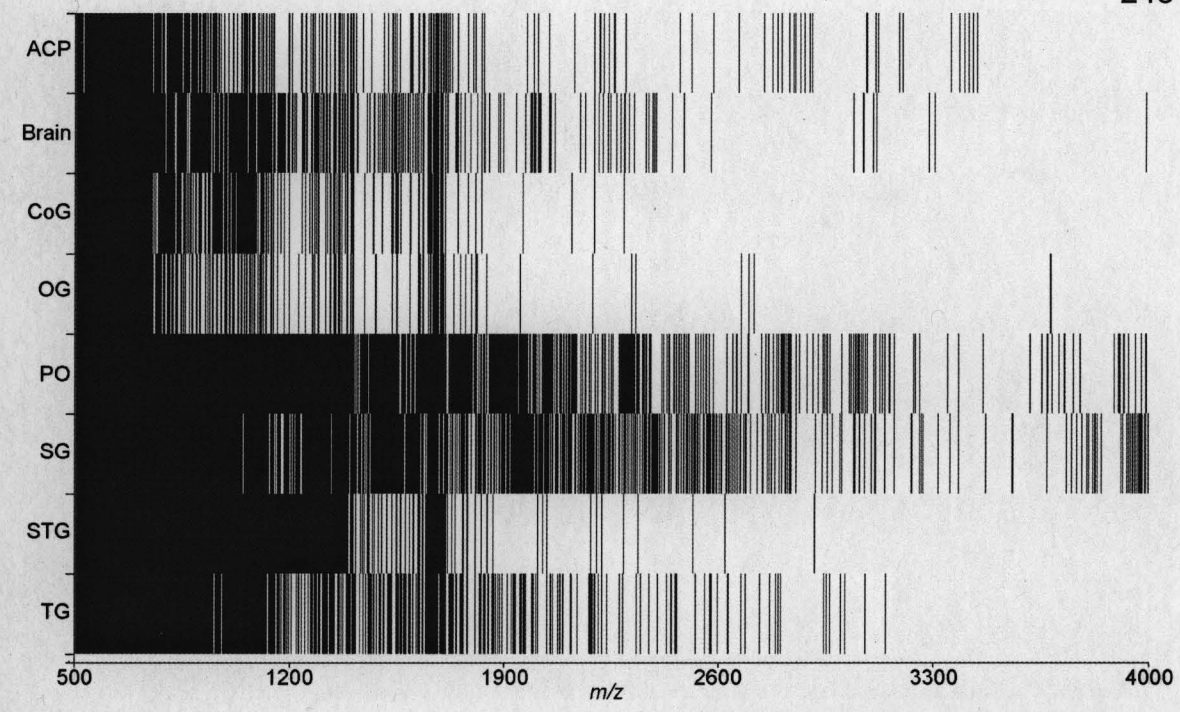

Figure 6.4: The stacked peak lists of all organs. Please note that the ACP and OG organs obviously do not have the number of peaks contained in the other organs.

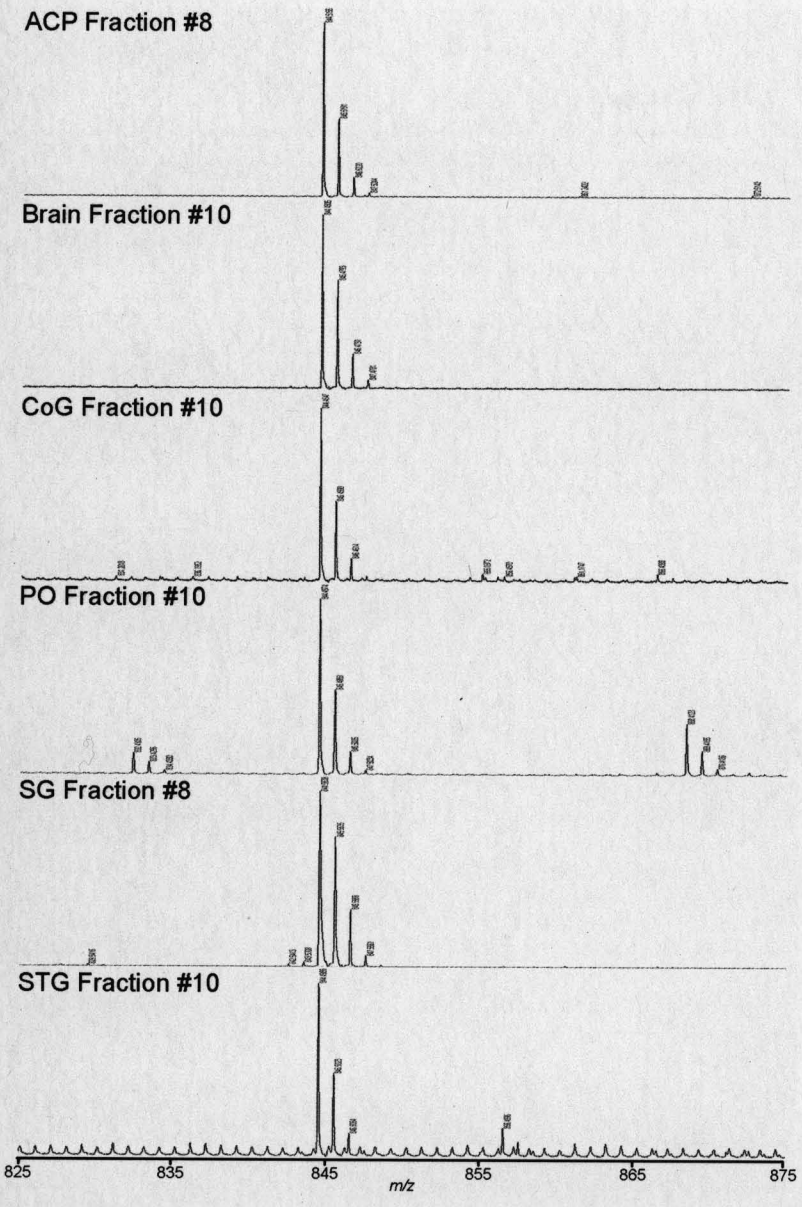

Figure 6.5: An example of the peptide HIGSLYRa found in six of the organ fractions. Note the similarity in fraction number indicating the consistency of the HPLC separation.

Peptide	Theoretical Mass
FLRFamides	
RSFLRFa	824.49
RNFLRFa	851.50
GRNFLRFa	908.52
SKNYLRFa	926.52
NRSFLRFa	938.53
pERNFLRFa	962.53
NRNFLRFa	965.54
GGRNFLRFa	965.54
DRNFLRFa	966.53
APRNFLRFa	1019.59
GNRNFLRFa	1022.56
RDRNFLRFa	1122.63
GYSKNYLRFa	1146.61
APQRNFLRFa	1147.65
AYNRSFLRFa	1172.63
SENRNFLRFa	1181.62
PELDHVFLRFa	1271.69
RYamides	
FVGGSRYa	784.41
FYANRYa	832.41
FYSQRYa	862.42
LFVGGSRYa	897.49
RFVGGSRYa	940.51
SGFYANRYa	976.46
pEGFYSQRYa	1030.47
SSRFVGGSRYa	1114.58
Orcomyotropin	
FDAFTTGFGHS	1186.52
FDAFTTGFGHN	1213.53
PDHs	
NSELINSILGLPKVMNDAa	1927.03
NSELINSILGLSRLMNEAa	1973.05

Peptide	Theoretical Mass
Allatostatins A-type	
YAFGLa	569.31
YSFGLa	585.30
GQYAFGLa	754.39
GGAYSFGLa	770.38
DPYAFGLa	781.39
EPYAFGLa	795.40
AGPYSFGLa	810.41
GDPYAFGLa	838.41
AGGAYSFGLa	841.42
PSMYGFGLa	884.43
PDMYGFGLa	898.41
GSGQYAFGLa	898.44
PDMYAFGLa	912.43
pERAYSFGLa	923.47
SSGQYAFGLa	928.45
PRDYAFGLa	937.49
QRAYSFGLa	940.50
pERPYSFGLa	949.49
DRPYSFGLa	953.48
pERTYSFGLa	953.48
APQPYAFGLa	962.51
ERPYSFGLa	967.50
PADLYEFGLa	1023.51
PATDLYAFGLa	1066.56
Allatostatins B-type	
NWNKFQGSWa	1165.55
TSWGKFQGSWa	1182.57
GNWNKFQGSWa	1222.58
NNWSKFQGSWa	1252.59
STNWSSLRSAWa	1293.63
NNNWSKFQGSWa	1366.63
Allatostatin Combos	
DPYAFGLGKRPADL	1519.79
EPYAFGLGKRPATDL	1634.85
DPYAFGLGKRPDMYGFGLa	2002.98
DPYAFGLGKRPDMYAFGLa	2017.00
GSGQYAFGLGKKAGGAYSFGLa	2035.04
DPYAFGLGKRPADLYEFGLa	2128.09
EPYAFGLGKRPATDLYAFGLa	2185.14

Peptide	Theoretical Mass
Orcokinins	
EIDRSGFGFA	1098.52
EIDRSSFGFN	1171.54
NFDEIDRSGF	1199.53
DEIDRSGFGFA	1213.55
NFDEIDRSGFG	1256.55
NFDEIDRSGFA	1270.57
NFDEIDRSSFG	1286.56
NFDEIDRSSFA	1300.58
NFDEIDRSGFGF	1403.62
NFDEIDRSGFGFA	1474.66
NFDEIDRSGFGFV	1502.69
DFDEIDRSGFGFV	1503.68
NFDEIDRSGFGFN	1517.67
NFDEIDRSSFGFV	1532.70
NFDELDRSGFGFH	1540.68
NFDEIDRSSFGFN	1547.68
Truncated CPRPs	
PGGLVHPVE	904.49
RSAQGMGKMEHL	1344.65
RSAQGMGKMERL	1363.69
RSAQGMGKMEHLL	1457.74
RSAQGMGKMEHLLA	1528.77
[13-38] Capr CPRP II	2658.39
Others	
RYLPT	649.37
DFSAWAa	695.31
HIGSLYRa	844.48
pELNFSPGWa	930.45
APSGFLGMRa	934.49
PFCNAFTGCa	956.38
GFKNVEMMTARGFa	1486.73

Table 6.2: List of known crustacean neuropeptides.

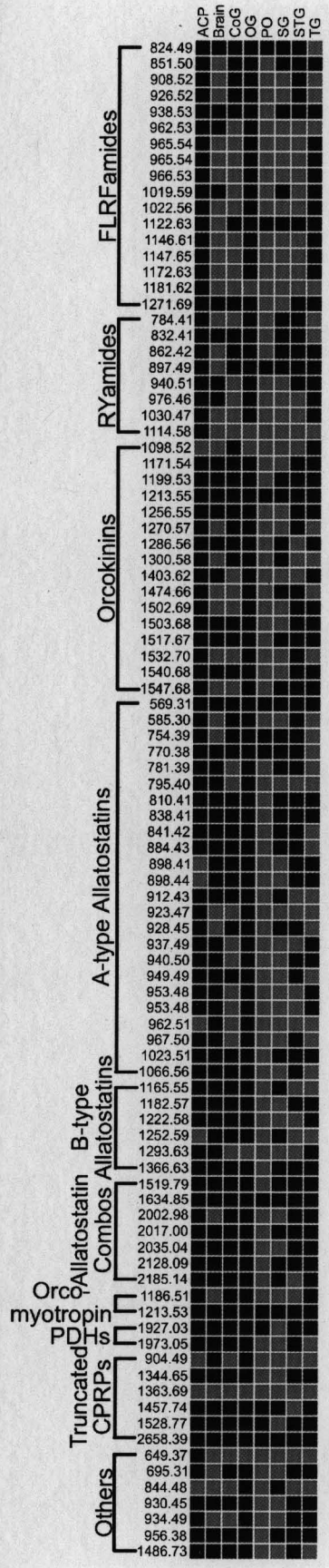

Figure 6.6: The heat map indicating the peptide identifications in each of the organs. This heat map was created after aligning each peak list with the included peaks and using a 100 ppm error window for identifications.

Chapter Seven

Discovery of Prion Disease Biomarkers Using MALDI-FTMS and Bioinformatics

The coauthors for this chapter are Allen Herbst and Judd Aiken from the University of Wisconsin-Madison Department of Animal Health and Biomedical Sciences, Sean McIlwain and David Page of the University of Wisconsin-Madison Department of Biostatistics and Medical Informatics, and my professor, Lingjun Li.

7.1 Abstract

Methods of detecting prion diseases are limited to post-mortem assays for protease- resistant prion protein. By utilizing mass spectrometry based protein profiling and bioinformatics we present a method by which a pre-mortem test can be developed using cerebrospinal fluid (CSF) and/or serum of animals infected with prion diseases. Analysis of CSF from infected/uninfected hamsters is carried out on a matrix assisted laser desorption/ionization – Fourier transform mass spectrometer utilizing a novel pulse sequence for the analysis of multiple masses simultaneously. Following mass spectrometric analysis, data is deisotoped, calibrated, and classified by machine-learning algorithms identifying differentially expressed proteins in infected or control samples. Finally, data is displayed using software developed in house to illustrate differentially expressed features. Using this strategy, we have been able to classify protein profiles from prion-infected animals. These learning algorithms have identified several peaks, which may be indicative of prion infection. Although these peaks have not been sequenced and identified at the present time, they do provide a panel of putative biomarkers that can be used to make more accurate predictions about sample infection status. We continue to adapt and extend these methods using serum in addition to CSF.

7.2 Introduction

Currently, a variety of diagnostic tests for animals infected with prion diseases (or transmissible spongiform encephalopathies) is lacking and those

that do exist do not always produce reliable results. Methods used now can primarily only be performed post-mortem and use microscopic observation of diseased brains and immuno-histochemistry focused on the prion protein itself. Although the immuno tests are fairly good, it would be of great use to have an ante-mortem method of diagnosis that could be performed on a bodily fluid that is relatively easy to obtain. A test of this sort would be of great importance as it would allow suspect animals and people to live while determining infection instead of after death or surgery. Creating such a tool is based on the hypothesis that a molecule, or a set of molecules, is present in infected animals that can serve as a panel of indicative biomarkers. Once these indicators are discovered, a diagnostic tool can be developed that is not only efficient and accurate, but can also be performed on easily attainable samples.

Transmissible spongiform encephalopathies (TSEs) are a unique family of neurodegenerative diseases of the central nervous system that are always fatal. They are caused by the conversion of a normal cellular protein designated PrP^{C} into an abnormal form of the prion protein PrP^{Sc}.[1, 2] Of the various forms of prion disease, those with the most impact upon human and animal health are Creutzfeldt-Jakob disease (CJD), bovine spongiform encephalopathy (BSE), scrapie and chronic wasting disease (CWD). The outbreaks of BSE and variant CJD (vCJD) in the United Kingdom and worldwide, including the United States, has prompted the need for rapid, reliable and inexpensive screening methods that allow TSEs to be identified at both symptomatic and pre-symptomatic stages in live animals. Current diagnostic tests for prion disease have, thus far, focused

on detection of the causal agent of the disease, the abnormal prion protein, or individual proteins that correlate with the neurological disease. These tests are inadequate because they are post-mortem, low throughput and not sufficiently sensitive to detect infection early in the pre-clinical period. Clearly, there is an urgent need for the development of a reliable, sensitive, and specific ante-mortem diagnostic test for the pre-clinical identification of TSE-infected animals or individuals.

Transmissible Spongiform Encephalopathies

The transmissible spongiform encephalopathies (TSEs) are an unusual class of neurological diseases that include scrapie in sheep and goats, bovine spongiform encephalopathy (BSE) in cattle, transmissible mink encephalopathy (TME), and chronic wasting disease (CWD) in elk and deer. Human forms of the disease include Creutzfeldt-Jakob disease (CJD), Gerstmann-Sträussler-Scheinker syndrome, fatal familial insomnia, and kuru.

TSEs are characterized by spongiform degeneration, reactive astrocytosis and prion protein (PrP) accumulation in the central nervous system. An extended incubation period of months to decades is characteristic of this group of inevitably fatal diseases.[3] Clinical features of TSEs vary with host species but, generally, clinically affected animals display pruritus, incoordination and ataxia.[4] TSE-infected animals do not display clinical symptoms throughout the majority of the incubation period despite the presence of high titers of the agent. Although pre-mortem diagnostic tests for TSEs in the preclinical and early clinical stages of

disease are currently being developed and validated, it is not yet possible to reliably identify animals with subclinical infections. TSEs were originally designated as "unconventional" or "slow" viruses based on the inability to identify a conventional virus and the long incubation periods associated with these infections.[5] Subsequent hypotheses were developed that proposed that the TSE agent consisted of an immunologically neutral (modified host) protein protecting a nucleic acid core (i.e., a virion[6]). The extreme resistance of these agents to ionizing and ultra-violet irradiation[7, 8] combined with the inability to isolate a virus or scrapie-specific nucleic acid[9] suggested, however, that TSE agents lacked a nucleic acid genome. These data support the hypothesis that a protein with self-replicating properties could be the TSE agent.[10] Further support for the self-replicating protein hypothesis came from studies that demonstrated the sensitivity of the TSE agent to protease digestion while not being affected by nuclease treatments.[11, 12] These data led to the refinement and renaming of the self-replicating protein hypothesis as the prion hypothesis, which stated, "prions are small proteinaceous particles which are resistant to inactivation by most procedures that modify nucleic acids".[13] TSEs are characterized at the molecular level by the accumulation of abnormal aggregated isoforms (PrP^{Sc}) of the prion protein (PrP^{C}) in affected animals (**Figure 7.1**). PrP^{C} is a 33-35 kDa protein encoded by a single copy gene.[14, 15] During the course of a scrapie infection, PrP^{C} undergoes a post-translational conformational conversion to disease-specific isoforms (PrP^{Sc}) that have increased resistance to proteinase K

digestion. *In vitro* cell culture studies have demonstrated that PrP^{C} is the precursor to PrP^{Sc}.[1, 2]

Prion Diseases

Scrapie is the prototypic TSE disease affecting sheep and goats. Scrapie is endemic in the United States where there are approximately 40 confirmed cases of scrapie per year (http://www.aphis.usda.gov). Bovine spongiform encephalopathy (BSE) was first described in Great Britain in 1986 and has since reached epidemic proportions. BSE is thought to originate by feeding cattle either scrapie-infected sheep protein or cattle protein from a spontaneous case of BSE, and was propagated by the practice of feeding cattle-derived proteins to cattle. BSE has resulted in the confirmed infection of 180,880 cattle in Great Britain since it was first diagnosed in 1985 (http://www.defra.gov.uk/animalh/bse /statistics/bse/worldwide.htm). As of January 1, 2006, clinical cases of BSE have also been reported in twenty-four other countries. A larger number of cattle have been involved in the "over thirty months" scheme (OTMS). In this attempt to further eradicate BSE, all cattle over the age of 30 months are culled and then processed such that they do not enter the food chain. In Great Britain alone, 8,074,600 cattle have been culled in the OTMS. The average incubation period of BSE is approximately 4 to 5 years, with clinical symptoms of head rubbing and aggression.

A new human form of Creutzfeldt-Jakob disease, variant CJD (vCJD), has been linked with exposure to the BSE agent.[16] Since first diagnosed in 1995 to

early January 2006, there have been 153 definite and probable cases of vCJD (Britain's Department of Health monthly CJD statistics webpage: http://www.dh.gov.uk/PublicationsAndStatistics/PressReleases/PressReleasesNotices/fs/en?CONTENT_ID=4126119&chk=joOCEs). vCJD differs from other forms of CJD and is characterized by large kuru-like amyloid plaques in the CNS, early age of onset of clinical symptoms, with a mean age of 27, atypical electroencephalogram readings, and a prolonged clinical stage of the disease.[16] Lesion profile analysis, a method that assesses the severity of spongiform change in defined regions of the brain, has demonstrated that BSE and vCJD produce similar lesion profiles in mice that differ from other forms of CJD.[17] As with the other TSE diseases, there is no treatment for vCJD and it is always fatal.

Chronic Wasting Disease (CWD) was first identified in 1967 in captive mule deer[18] and later in captive Rocky Mountain elk.[19] CWD has also been detected in free-ranging populations of mule deer, white-tailed deer and Rocky Mountain elk.[20] The prevalence of CWD in endemic areas in Colorado is about 15%.[21] Although CWD can be transmitted horizontally in captive populations, it is unknown how transmission occurs in wild populations. Since CWD's initial identification, the disease has spread to game farms across the U.S. and Canada and even as far as South Korea. With the spread of CWD within captive animals the risk of accidental infection of wild animals increases. The development of a sensitive pre-clinical diagnostic coupled with mandatory testing could help to prevent the human assisted transmittance of prion disease.

Health and Economic Impact of Prion Disease

Although the species barrier (the resistance to TSE infection by one species following exposure to the TSE agent of another species) provides significant protection from the interspecies transmission of prion disease, the BSE epidemic and the resulting rise in vCJD illustrates the potential impact of prion disease upon human and economic health. The ability of CWD to infect ~25% of cows during the first passage demonstrates a weak species barrier in comparison to the levels of resistance found in other experimental interspecies transmissions where no animals developed disease even after serial passage. The cohabitation of TSE infected deer with beef and dairy cattle creates thousands of natural interspecies transmission experiments. To the extent that reservoirs of prion disease remain, prophylaxis with specific high throughput ante-mortem diagnostic testing to identify and eliminate sources of infectivity will prevent transmissions and help to ensure the safety of food and blood supplies.

Current Diagnostic Testing for Prion Disease and Its Limitations

Currently validated diagnostic tests for prion diseases are all post-mortem, and all involve detection of abnormal accumulation of the prion protein in the central nervous system. The "gold standard" in prion diagnostic testing is immunohistochemistry utilizing anti-prion protein antibodies on the obex region of

the brain.[22] Drawbacks to utilizing immunohistochemistry on the obex include the low throughput of samples and the necessity for post-mortem analysis. Other antibody-based diagnostics, such as the Prionics or Bio-Rad tests, utilize a Western Blot/ELISA approach. These diagnostics take advantage of the protease resistance of the abnormal form of the prion protein. Despite the good specificity of these tests, the sensitivity is so low that animals infected with prion disease cannot be diagnosed until late in the pre-clinical period when sufficient abnormal PrP has accumulated in brain tissue. A further limitation of post-mortem diagnosis is the inability to monitor and track disease progression. The ultimate goal should be the development of pre-clinical diagnostics that can detect disease before potentially contaminated food or blood products enter the market.

Surrogate markers of prion disease have been identified in patients presenting clinical signs of CJD. Among these are the characteristic electroencephalogram pattern observed in CJD cases, the presence of central nervous system-specific protein markers such as 14-3-3, tau, $A\beta_{1-42}$, apolipoprotein E, cystatin C[23-27] and neuron-specific enolase.[25, 28, 29] Despite being secondary to post-mortem immunohistochemistry for diagnosis due to their low specificity and sensitivity, these markers illustrate the biological response to prion infection during clinical stage.

Mass Spectrometry as a Tool for the Discovery of Diagnostic Biomarkers

Mass spectrometry has become an increasingly popular tool for protein biomarker discovery due to its high sensitivity, speed, chemical specificity and capability for complex mixture analysis. At the heart of this rapidly evolving research area is the capability to characterize an ensemble of proteins expressed in a tissue or secreted into body fluids. These new capabilities became possible due to major innovations in ionization methodologies (matrix-assisted laser desorption/ionization[30, 31] and electrospray ionization[32], MALDI and ESI, respectively). With these tools, mass spectrometrists could, for the first time, convert condensed-phase peptides or proteins to intact gas-phase ions for mass measurement. Coupled with the development of various mass analyzers, tandem mass spectrometry (MS/MS) experiments can be conducted to produce peptide ion fragmentation for subsequent *m/z* measurement – a process used to derive amino acid sequence information.

Biostatistics and Biomarker Identification

Because of the large data set that is normally produced when searching for biomarkers, it is usually necessary to employ biostatistics software to fully interpret the data. The learning algorithms that we plan on using are the Naïve-Bayes algorithm, tree-augmented networks, support vector machines, and k-nearest neighbor.[33-35] The Naïve-Bayes classifies depending upon features, but the features are independent of each other (**Figure 7.2A**). Tree-augmented networks are an extension of Naïve-Bayes that learns correlations between

features in a simple manner. Support vector machines map data into high dimensional feature space and then a separating hyper-plane is identified to classify further data (**Figure 7.2B**). Finally, k-nearest neighbor works also by mapping data, but makes a classification based on k points closest to the unknown sample. Each algorithm has its strengths and weaknesses. Finding a stable algorithm that obtains the best accuracy using proteomic feature data requires many cross-validation tests of a given proteomic dataset. By using these and other similar algorithms, it is possible to use large data sets to predict the status of a sample.

7.3 Methods

Sample Preparation

Hamsters were orally inoculated with the 263K prion agent, while an equal number of uninfected animals were kept as controls. CSF was drawn at 15, 18, and 21 weeks after infection from both infected and uninfected animals. Following collection, each sample was digested with trypsin overnight, spotted three times and then each spot was analyzed three times using MALDI-FTMS (see below for more information) resulting in nine spectra being produced for each sample. This replication was necessary to increase the statistical accuracy of our data analysis.

MALDI-FTMS

The instrument used in this project was a matrix assisted laser desorption/ionization – Fourier transform mass spectrometer, or MALDI-FTMS. This instrument works by first spotting the sample onto a target along with a small amount of matrix (a molecule that absorbs UV energy and passes it on to the analytes). The matrix, in our case 2,5-dihydroxybenzoic acid (DHB), is then allowed to cocrystallize with the sample before being placed into the instrument. Then, the sample spot is irradiated using a nitrogen laser emitting light at 337 nm. The DHB absorbs this light and causes the analytes to enter the gas phase and come away from the target (desorb) and be ionized (ionization). This is the process known as MALDI. After being ionized, the ions are carried through a quadrupole into an ion cyclotron resonance cell where all the ions gather in the center in a cloud. Once the ions are in the cell, an RF scan is applied and the ions are excited to a larger radius at their resonant frequency. Immediately following the RF scan, and while the ions are still excited, they are detected and produce a sinusoidal data set. This data is then Fourier transformed to produce mass spectra that are understandable and usable.

Data Analysis

Because the data sets are so large, processing by hand is almost impossible. Thus, we have made use of computers to run learning algorithms in order to identify possible biomarkers in the spectra. In addition to processing data, it is also useful to use computers to visualize the data in a more

understandable manner. Software was designed in-house for the processing and visualization of our unique data. The software identified unique peaks by removing any masses below 500 m/z (in order to suppress matrix ion peaks) and then is deisotoped with peaks lacking isotopic distribution being removed. The data is then clustered into bins 100 ppm wide and then features identified based on whether a peak exists within a bin or not. Using information gain, the top 25 peaks are then chosen, displayed, and used for further classification of data. To classify data, several different machine learning algorithms are used. These include k-nearest neighbor (kNN), Naïve-Bayes, tree-augmented networks (TAN), and support vector machines (SVM) (**Figure 7.2**). The Naïve-Bayes classifies depending upon features, but the features do not depend on each other. The tree-augmented networks are an extension of Naïve-Bayes that learns correlations between features in a simple manner. Support vector machines map data into high dimensional feature space and then a separating hyper-plane is identified to classify further data. Finally, k-nearest neighbor works also by mapping data, but makes a classification based on k points closest to the unknown sample. By using these different classifying learning algorithms, we have been able to attain better than chance prediction of submitted data. While this is not perfect, we are well on our way to our goal of being able to diagnose prion infection in suspect animals without using brain tissue.

Mass Spec Data Miner for Data Analysis

There are many issues to consider when analyzing proteomic mass spectrometry data for classification. Due to the large number of peaks detected for a given sample, we need to reduce raw mass spectrometric data to a simple set of features that are predictive of disease. We also would like to measure the predictive accuracy of our whole method of pre-processing the mass spectra, extracting features, feature selection, and training a classifier on the selected features.

For comparing different approaches to feature extraction and feature selection, we have constructed a software tool called Mass Spec Data Miner (MSDM), which is available at http://www.cs.wisc.edu/~mcilwain/MSDM. MSDM is a modular program that has four main constituents.

The (x,y)-preprocessor vector contains a list of modules that directly affect the *m/z*, intensity data. Examples of these are the deisotoping algorithm and an algorithm for cutting data points that have a *m/z* less than a certain value. The MSDM contains a module that will extract features from the (x,y) data and a feature processor vector that processes the extracted features. The fourth module is the classifier module. Each module has a training method. The (x,y) and the feature processors have a process method, the feature extraction module has a convert method, and the classifier module has a classify method. The processor modules are chained, so that the output of the previous module is passed through as the input of the next module.

The general MSDM algorithm is depicted in **Figure 7.3**. For the training step, the MSDM modules train upon and process the data for the next processor or converter. Each module retains information for processing future data in the same manner. The classifier is then trained on the resulting feature data set. Classification is done by passing the data through the trained modules' process, convert, or classify methods. This allows for cross-validation to be performed across the whole process and for full customizability. New processor, converter, and classifier modules can be easily created, tested, and included into the MSDM.

For processing preliminary mass spectrometry data set collected from prion-infected and uninfected hamster CSF samples, we perform four main steps: (1) Pre-process the *m/z* peak data; (2) Extract features that are (un)common across spectra; (3) Pre-process the feature vector data; (4) Train and perform classification using a number of different classification algorithms and scored their accuracy including nearest neighbor (K = 1, 3, 5, 7), Naïve Bayes, Tree-Augmented Networks (TAN), and Linear SVMs.

7.4 Results

Positive Control

To show that the Mass Spec Data Miner is able to identify potential biomarkers, as well as to determine the status of a sample, a positive control study was carried out. **Figure 7.4** shows the heat map of the positive control

experiment. 24 samples were created from a mixture of 10^{-6} M peptide standards: FMRFa (599 m/z), Angiotensin II (1046 m/z), Vasopressin (1084 m/z), Angiotensin I (1296 m/z), Substance P (1347 m/z), Somatostatin (1637 m/z), and GAHKNYLRFa (synthesized, 1104 m/z). The synthesized peptide was only included in 12 samples and served as a putative differentially regulated biomarker for the Mass Spec Data Miner to identify. Black squares indicate the absence of a peak while blue squares indicate the presence of a peak. Each line represents a collection of 3 independently acquired spectra. Information gain is a statistical metric that measures how well a given peak separates the data according to their specific classification. Information gain values are calculated from the accuracy, precision and recall of a given peak. In this case 1.0, the maximum value, is assigned to the 1104 m/z (boxed in red on the left) because it is the "spiked-in" synthesized peptide designed to be a perfect classifier of the data. The other peptides that made up the spiked mixture have very low information gain values as they are equally represented in both the seven and six standard groups. It is noted that in addition to the peaks corresponding to the peptide standards spiked in the mixture, there are also many other peaks detected in the spectra, which resulted from impurities, matrix peaks and peptide degradation/fragmentation products. After identifying the classifying feature such as 1104 m/z peak, learning algorithms were used to determine the status of the blinded data. As seen in **Table 7.1**, the positive control study demonstrates the utility of our SpecPlot software to identify classifying features and then use them to determine the status of a given sample.

Infected vs. Uninfected

We analyzed 11 infected and 9 uninfected samples from the 18 week timepoint and 8 infected and 9 uninfected from the 21 week timepoint. On average we obtained very good scaling for each spectrum with peak numbers exceeding 1000 for most samples. Unfortunately, both the uninfected (**Figure 7.5 A**) and infected (**Figure 7.5 B**) samples produce very similar sets of peaks. The spectra below are unique as they highlight (double headed arrow) one of the peaks that seems to indicate the presence of prion infection. This peak is not definitive though as it is not always seen in the infected and occasionally appears in the uninfected samples as is illustrated by a visualization of all of the data (**Figure 7.6**).

Software was developed to visualize the large data set by horizontally stacking each spectra. Although this does not necessarily highlight differences between infected and uninfected samples, it can show these peaks once identified and also help to identify outlier samples (**Figure 7.6**). The different colors represent the intensity of the peaks with blue being more intense. While intensity was not utilized as a distinguishing parameter for this study, it would be possible in the future to include this information and thus possibly identify features based on intensity changes as well as presence.

Along with the visualization software that has been developed is an algorithm to pick out the top 25 peaks that are "somewhat" unique between infected and uninfected spectra. These peaks are "somewhat" unique because

none of the peaks were identified in one infection status or the other. This means that the peaks were never always excluded or always present from infected or uninfected sample spectra. The visualization program, SpecPlot, also allows us to compress the data for viewing purposes. By doing this and combining all nine spectra, we are able to combine and view all the data for each sample for the 25 identified peaks creating a "heat map" of peak occurrence (**Figure 7.7**). In this case, the intensity of color is determined by the frequency that the peak appears in those samples' nine separate spectra. By performing these operations, it is now relatively easy to identify peaks that could possibly be used as classifying features.

After identifying peaks that could be used as classifying features, algorithms were designed for diagnosing samples as infected or uninfected. Because features are indefinite (i.e. "somewhat" characteristic), learning algorithms were used to take all features into account when deciding whether a sample was infected or not. Additionally, different algorithms use differing methods of feature classification and therefore it was necessary to try multiple algorithms to find which worked best. To test the algorithms, data that had already been produced was blinded and classified by each algorithm. Most of the algorithms show better than chance abilities in predicting the status of the samples (**Table 7.2**). In the future, better sample processing, before and after MS analysis, could greatly improve the predictive abilities of the algorithms. By trying other pre-processing, extraction, and classification methods, algorithms could be improved greatly.

ROC Curve

To further test the predictive ability of features using two of the best performing learning algorithms, Naïve-Bayes and SVM, we carried out further statistics on each to produce a receiver operator characteristic (ROC) curve and a precision vs. recall curve. To produce the ROC and precision vs. recall curves, we adjusted the number of spectral votes required to consider a sample infected or uninfected. For example, in our original testing, we used a majority vote to determine infection status, i.e. if five out of the nine spectra per sample were classified as infected, the whole sample would be classified infected, and vice versa. In this case, we have adjusted those numbers and have asked what would happen to our predictive ability if we only required four out of the nine to call a sample infected (**Table 7.3**). Interestingly enough, we found that the majority vote (at least 50% positives required) produced the best voting statistics for both the Naïve-Bayes and SVM trials (**Table 7.3**). **Figure 7.8** shows an example of the ROC and precision vs. recall curves produced for this analysis. This figure shows the best results for any of the methods tried and exhibits the closest to optimum curve.

Peptide Identification

One other process that will also be carried out is the sequencing of the unique peaks in order to determine their identity. The four most characteristic peaks are very small in relation to other peaks (**Figure 7.9**). Due to the method

of ion fragmentation (SORI-CID) and suppression from other more prevalent peaks, sequencing will be very difficult. In order to carry this out, pooling and HPLC fractionation will be necessary. Pooling and fractionation will increase the amount of analyte present as well as simplify spectra and help eliminate suppression effects. Once sequenced, it will be possible to create antibodies to the classifying peptides and develop an immuno-based diagnostic tool for prion diseases that does not require brain tissue and can be performed without killing the suspect animal(s).

Future Work

In the future, because any resulting test will need to be as simple as possible for clinical use, it may be necessary to switch to identification of biomarkers from the serum instead of CSF. Obviously, doing this creates many more challenges than it solves, and thus must be done with much forethought. For example, serum contains a very large amount of albumin, which, if not dealt with, will suppress all other peptide signals within a spectrum (Figure 7.10). As a result, it will be necessary to remove the albumin and other highly abundant proteins using one of the immunodepletion columns that is currently commercially available today. Furthermore, by utilizing other known features of the prion protein, i.e. its glycosylation characteristics, it may be possible to identify biomarkers using these features and thus further the research.

7.5 Conclusions

We have demonstrated a method by which biomarkers of prion disease infection can be identified. Furthermore, we have shown that it is possible to diagnose prion infection without using brain tissue and thus allows for the possibility of developing an antemortem test for larger animals (i.e. cattle, cervids, etc.). In all, this research, while having many challenges before it, is not only necessary, but through correct use of today's technology, can be successful, as we have demonstrated here.

7.6 References

1. Caughey, B., Race, R. E., Ernst, D., Buchmeier, M. J., and Chesebro, B. (1989) Prion protein biosynthesis in scrapie-infected and uninfected neuroblastoma cells. *Journal of Virology* 63, 175-181

2. Borchelt, D. R., Scott, M., Taraboulos, A., Stahl, N., and Prusiner, S. B. (1990) Scrapie and cellular prion proteins differ in their kinetics of synthesis and topology in cultured cells. *Journal of Cell Biology* 110, 743-752

3. Hunter, G. D. (1972) Scrapie: a prototype slow infection. *Journal of Infectious Diseases* 125, 427-440

4. Detwiler, L. A. (1992) Scrapie. *Revue Scientifique et Technique* 11, 491-537

5. Sigurdsson, B. (1954) Observations on three slow infections of sheep. *British Veterinary Journal* 110, 341-354

6. Dickinson, A. G., and Outram, G. W. (1979) *The Scrapie Replication-Site Hypothesis and it's Implications for Pathogenesis In Slow Transmissible Diseases of the Nervous System*, Academic Press, New York

7. Alper, T., Haig, D. A., and Clarke, M. C. (1978) The scrapie agent: evidence against its dependence for replication on intrinsic nucleic acid. *Journal of General Virology* 41, 503-516

8. Bellinger-Kawahara, C., Cleaver, J. E., Diener, T. O., and Prusiner, S. B. (1987) Purified scrapie prions resist inactivation by UV irradiation. *Journal of Virology* 61, 159-166

9. Aiken, J. M., and Marsh, R. F. (1990) The search for scrapie agent nucleic acid. *Microbiol Rev* 54, 242-246

10. Griffith, J. S. (1967) Self-replication and scrapie. *Nature* 215, 1043-1044

11. Cho, H. J. (1980) Requirement of a protein component for scrapie infectivity. *Intervirology* 14, 213-216

12. Prusiner, S. B., Cochran, S. P., Downey, D. E., and Groth, D. F. (1981) Determination of scrapie agent titer from incubation period measurements in hamsters. *Advances in Experimental Medicine & Biology* 134, 385-399

13. Prusiner, S. B., Bolton, D. C., Groth, D. F., Bowman, K. A., Cochran, S. P., and McKinley, M. P. (1982) Further purification and characterization of scrapie prions. *Biochemistry* 21, 6942-6950

14. Chesebro, B., Race, R., Wehrly, K., Nishio, J., Bloom, M., Lechner, D., Bergstrom, S., Robbins, K., Mayer, L., and Keith, J. M. (1985) Identification of scrapie prion protein-specific mRNA in scrapie-infected and uninfected brain. *Nature* 315, 331-333

15. Oesch, B., Westaway, D., Walchli, M., McKinley, M. P., Kent, S. B., Aebersold, R., Barry, R. A., Tempst, P., Teplow, D. B., and Hood, L. E. (1985) A cellular gene encodes scrapie PrP 27-30 protein. *Cell* 40, 735-746

16. Will, R. G., Ironside, J. W., Zeidler, M., Cousens, S. N., Estibeiro, K., Alperovitch, A., Poser, S., Pocchiari, M., Hofman, A., and Smith, P. G. (1996) A new variant of Creutzfeldt-Jakob disease in the UK. *Lancet* 347, 921-925

17. Bruce, M. E., Will, R. G., Ironside, J. W., McConnell, I., Drummond, D., Suttie, A., McCardle, L., Chree, A., Hope, J., Birkett, C., Cousens, S., Fraser, H., and Bostock, C. J. (1997) Transmissions to mice indicate that 'new variant' CJD is caused by the BSE agent.[see comment]. *Nature* 389, 498-501

18. Williams, E. S., and Young, S. (1980) Chronic wasting disease of captive mule deer: a spongiform encephalopathy. *Journal of Wildlife Diseases* 16, 89-98

19. Williams, E. S., and Young, S. (1982) Spongiform encephalopathy of Rocky Mountain elk. *Journal of Wildlife Diseases* 18, 465-471

20. Spraker, T. R., Miller, M. W., Williams, E. S., Getzy, D. M., Adrian, W. J., Schoonveld, G. G., Spowart, R. A., O'Rourke, K. I., Miller, J. M., and Merz, P. A. (1997) Spongiform encephalopathy in free-ranging mule deer (Odocoileus hemionus), white-tailed deer (Odocoileus virginianus) and Rocky Mountain elk (Cervus elaphus nelsoni) in northcentral Colorado. *Journal of Wildlife Diseases* 33, 1-6

21. Miller, M. W., Williams, E. S., McCarty, C. W., Spraker, T. R., Kreeger, T. J., Larsen, C. T., and Thorne, E. T. (2000) Epizootiology of chronic wasting disease in free-ranging cervids in Colorado and Wyoming. *Journal of Wildlife Diseases* 36, 676-690

22. Schaller, O., Fatzer, R., Stack, M., Clark, J., Cooley, W., Biffiger, K., Egli, S., Doherr, M., Vandevelde, M., Heim, D., Oesch, B., and Moser, M. (1999) Validation of a western immunoblotting procedure for bovine PrP(Sc) detection and its use as a rapid surveillance method for the diagnosis of bovine spongiform encephalopathy (BSE). *Acta Neuropathologica* 98, 437-443

23. Van Everbroeck, B., Boons, J., and Cras, P. (2005) Cerebrospinal fluid biomarkers in Creutzfeldt-Jakob disease. *Clin Neurol Neurosurg* 107, 355-360

24. Parveen, I., Moorby, J., Allison, G., and Jackman, R. (2005) The use of non-prion biomarkers for the diagnosis of Transmissible Spongiform Encephalopathies in the live animal. *Vet Res* 36, 665-683

25. Otto, M., Wiltfang, J., Cepek, L., Neumann, M., Mollenhauer, B., Steinacker, P., Ciesielczyk, B., Schulz-Schaeffer, W., Kretzschmar, H. A., and Poser, S. (2002) Tau protein and 14-3-3 protein in the differential diagnosis of Creutzfeldt-Jakob disease. *Neurology* 58, 192-197

26. Choe, L. H., Green, A., Knight, R. S., Thompson, E. J., and Lee, K. H. (2002) Apolipoprotein E and other cerebrospinal fluid proteins differentiate ante mortem variant Creutzfeldt-Jakob disease from ante mortem sporadic Creutzfeldt-Jakob disease. *Electrophoresis* 23, 2242-2246

27. Sanchez, J. C., Guillaume, E., Lescuyer, P., Allard, L., Carrette, O., Scherl, A., Burgess, J., Corthals, G. L., Burkhard, P. R., and Hochstrasser, D. F. (2004) Cystatin C as a potential cerebrospinal fluid marker for the diagnosis of Creutzfeldt-Jakob disease. *Proteomics* 4, 2229-2233

28. Burkhard, P. R., Sanchez, J. C., Landis, T., and Hochstrasser, D. F. (2001) CSF detection of the 14-3-3 protein in unselected patients with dementia.[see comment]. *Neurology* 56, 1528-1533

29. Zerr, I., Helmhold, M., Poser, S., Armstrong, V. W., and Weber, T. (1996) Apolipoprotein E phenotype frequency and cerebrospinal fluid concentration are not associated with Creutzfeldt-Jakob disease. *Archives of Neurology* 53, 1233-1238

30. Karas, M., and Hillenkamp, F. (1988) Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* 60, 2299-2301

31. Hillenkamp, F., Karas, M., Beavis, R. C., and Chait, B. T. (1991) Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. *Anal Chem* 63, 1193A-1203A

32. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., and Whitehouse, C. M. (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science* 246, 64-71

33. Mitchell, T. M. (1997) *Machine learning*, MIT Press and McGraw-Hill Companies, Inc.

34. Witten, I. H. F., E. (2000) *Data Mining: Practical machine learning tools with Java implementations*, Morgan Kaufman, San Francisco

35. Platt, J. C. (1999) *Fast training of support vector machines using sequential minimal optimization.*, MIT Press, Cambridge, MA

7.7 Figures

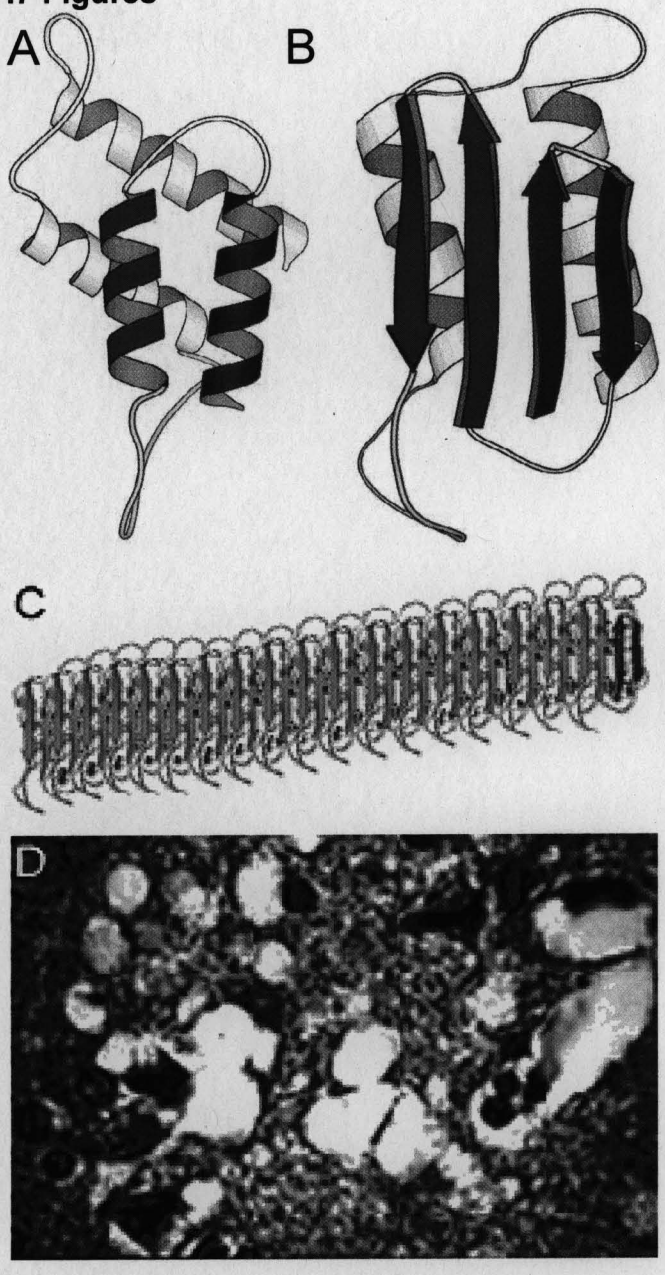

Figure 7.1: **A.** The normal α-helix rich form of the prion protein (PrP^{C}). **B.** The abnormal β-sheet rich form of the prion protein (PrP^{Sc}). **C.** How the PrP^{Sc} aggregates to form amyloid fibrils. **D.** Image of human brain affected by Kuru (another TSE disease similar to CJD) showing the spongiform morphology of clinical prion diseases in the brain.

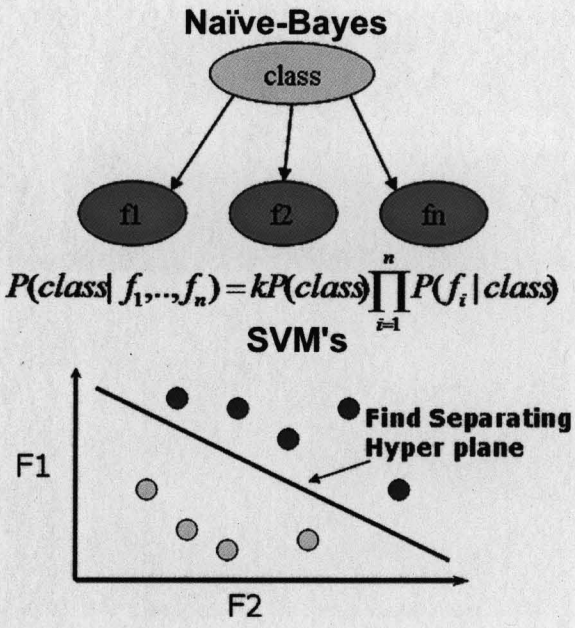

Figure 7.2: An illustration of how the (**A**) Naïve-Bayes and (**B**) Support Vector Machine learning algorithms work.

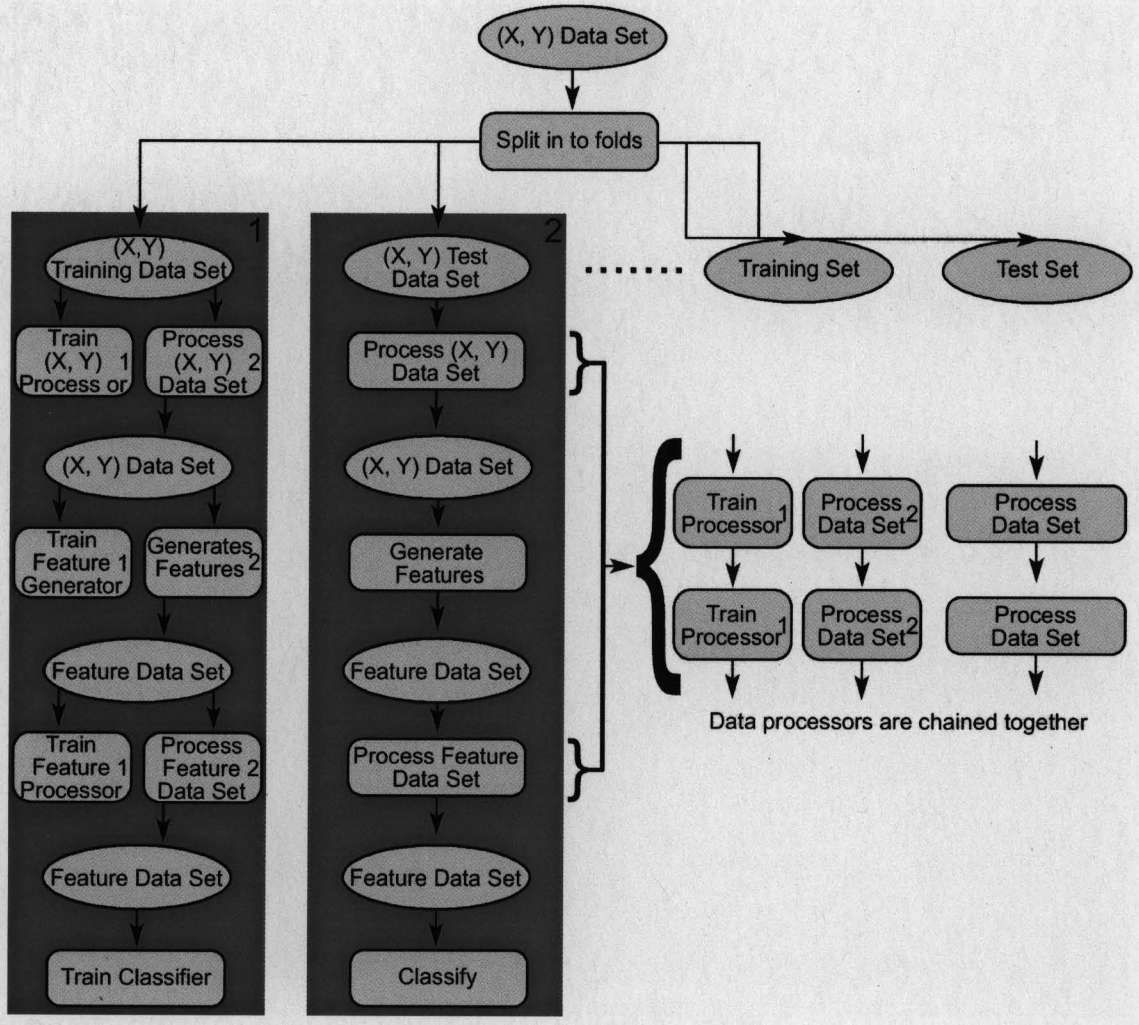

Figure 7.3: Steps taken by the MSDM algorithm.

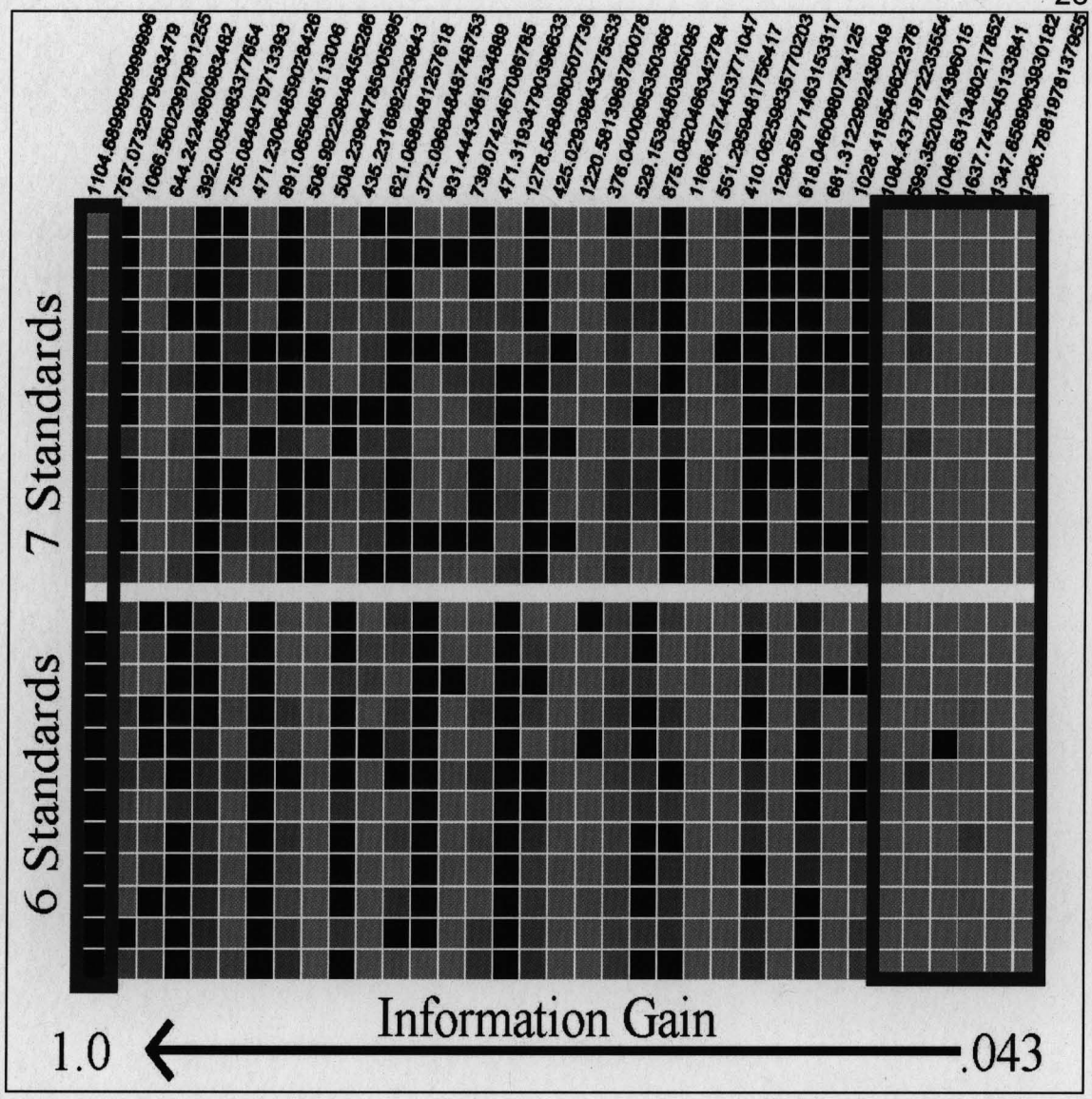

Figure 7.4: Heat Map of positive control experiment.

Classifier	Train Accuracy	Test Accuracy	Vote Accuracy
k-Nearest Neighbor	97.94 %	95.83 %	100 %
Naïve-Bayes	98.25 %	95.83 %	100 %
Decision Tree	97.22 %	97.22 %	100 %

Table 7.1: Positive control study learning algorithm results.

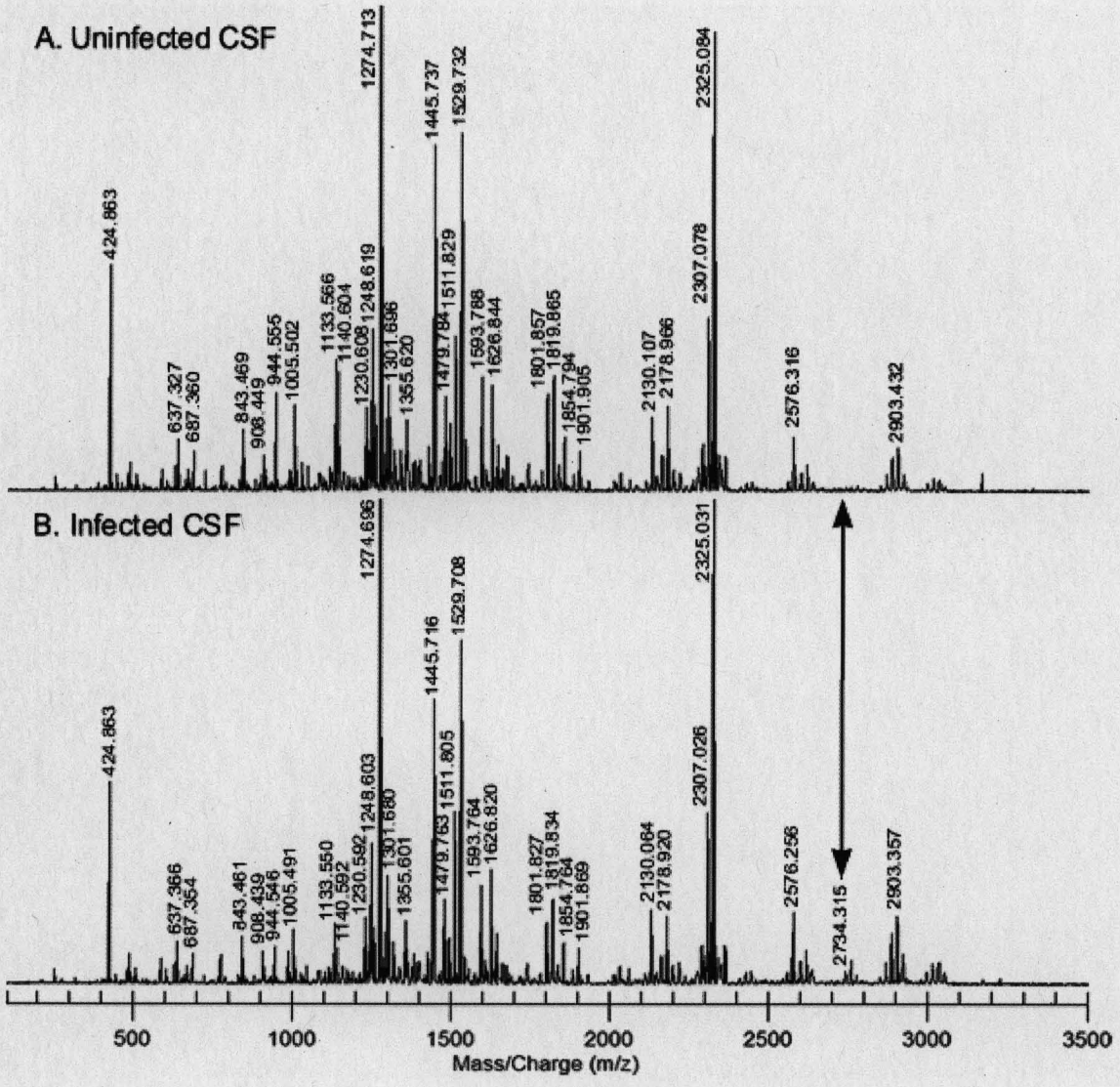

Figure 7.5: Example spectrum from (**A**) uninfected CSF sample, and (**B**) infected CSF sample. Please note the peak at *m/z* 2734.315 that is the only noticeable peak that is different from the infected to uninfected spectra. Although this feature can be differentiated in these two spectra, it is not consistently present in the infected or absent in the uninfected and is therefore can not be used alone as an indicative feature, thus requiring the consideration of other features as well.

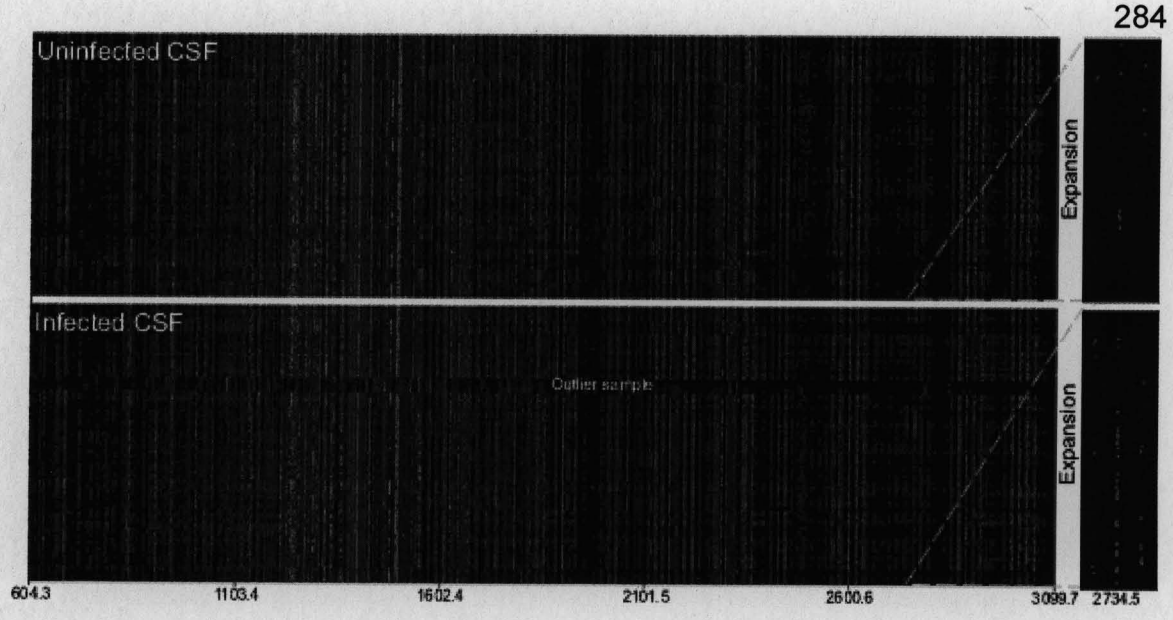

Figure 7.6: Visualization of all hamster CSF spectra stacked using software developed in-house. Colors represent intensity of peaks.

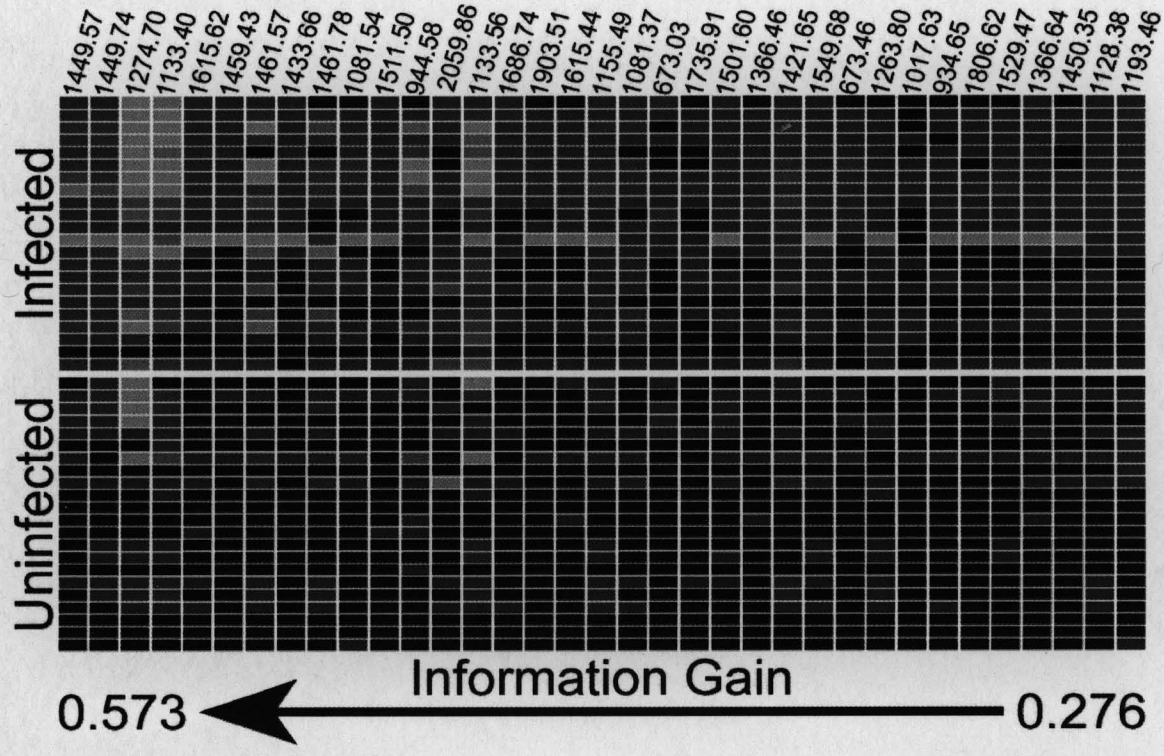

Figure 7.7: Heat map of infected *vs.* uninfected samples compared against the top 25 most indicative peaks. The absence of a peak is represented by black, and the presence of a peak is indicated by blue, with brighter blue showing higher frequency of presence. The information gain arrow indicates the predictive ability of the peaks increasing from right to left.

Classification Algorithm	Accuracy	Vote Accuracy
kNN {k=1,3,5,7,9}	63%	75%
Naïve-Bayes	67%	70%
Tree Augmented Networks	72%	75%
Support Vector Machines	66%	85%

Table 7.2: Classifying accuracy of different learning algorithms. Accuracy is cross validation with one hamster left out while vote accuracy uses all nine spectra per sample for a majority vote classification.

	% of positives required	0%	5%	17%	28%	39%	50%	61%	72%	83%	94%	100%
Naïve-Bayes	Accuracy	0.49	0.70	0.70	0.77	0.77	0.79	0.77	0.77	0.70	0.67	0.51
	Precision	0.49	0.62	0.63	0.70	0.70	0.77	0.79	0.82	0.83	0.89	0
	Recall	1	1	0.90	0.90	0.90	0.81	0.71	0.67	0.48	0.38	0
	Negative Accuracy	0	0.41	0.50	0.64	0.64	0.77	0.82	0.86	0.91	0.95	1
	F-Score	0.66	0.76	0.75	0.79	0.79	0.79	0.75	0.74	0.61	0.53	0
	False Positive Rate	1	0.59	0.50	0.36	0.36	0.23	0.18	0.14	0.09	0.05	0
Linear Support Vector Machine	Accuracy	0.49	0.72	0.74	0.77	0.81	0.88	0.84	0.84	0.77	0.70	0.51
	Precision	0.49	0.64	0.66	0.69	0.74	0.83	0.82	0.85	0.92	1	0
	Recall	1	1	1	0.95	0.95	0.95	0.86	0.81	0.57	0.38	0
	Negative Accuracy	0	0.46	0.50	0.59	0.68	0.82	0.82	0.86	0.95	1	1
	F-Score	0.66	0.78	0.79	0.80	0.83	0.89	0.84	0.83	0.71	0.55	0
	False Positive Rate	1	0.55	0.50	0.41	0.32	0.18	0.18	0.14	0.05	0	0

Table 7.3: Comparison of the predictive ability of Naïve-Bayes and linear Support Vector Machines. Accuracy is a measure of how many of all the predictions are correct (Accuracy = (TP+TN)/(TP+TN+FP+FN)). Precision is a measure of how many of our predicted positives were actually positive (Precision = TP/(TP+FP)). Recall is a measure of how many of the actual positives we predicted (Recall= TP/(TP+FN)). Negative Accuracy is a measure of how many of the actual negatives we predicted correctly (Negative Accuracy = TN/(TN+FP)). F-score is a measure of the algorithm's accuracy (F-score = FP/(FP+TN)). False Positive Rate is a measure of how many false positives are predicted (False Positive Rate = FP/(FP+TN)). Please note the linear support vector machine at the 50% vote provides the highest indicative values for all parameters (red square).

TP = True Positive TN = True Negative

FP = False Positive FN = False Negative

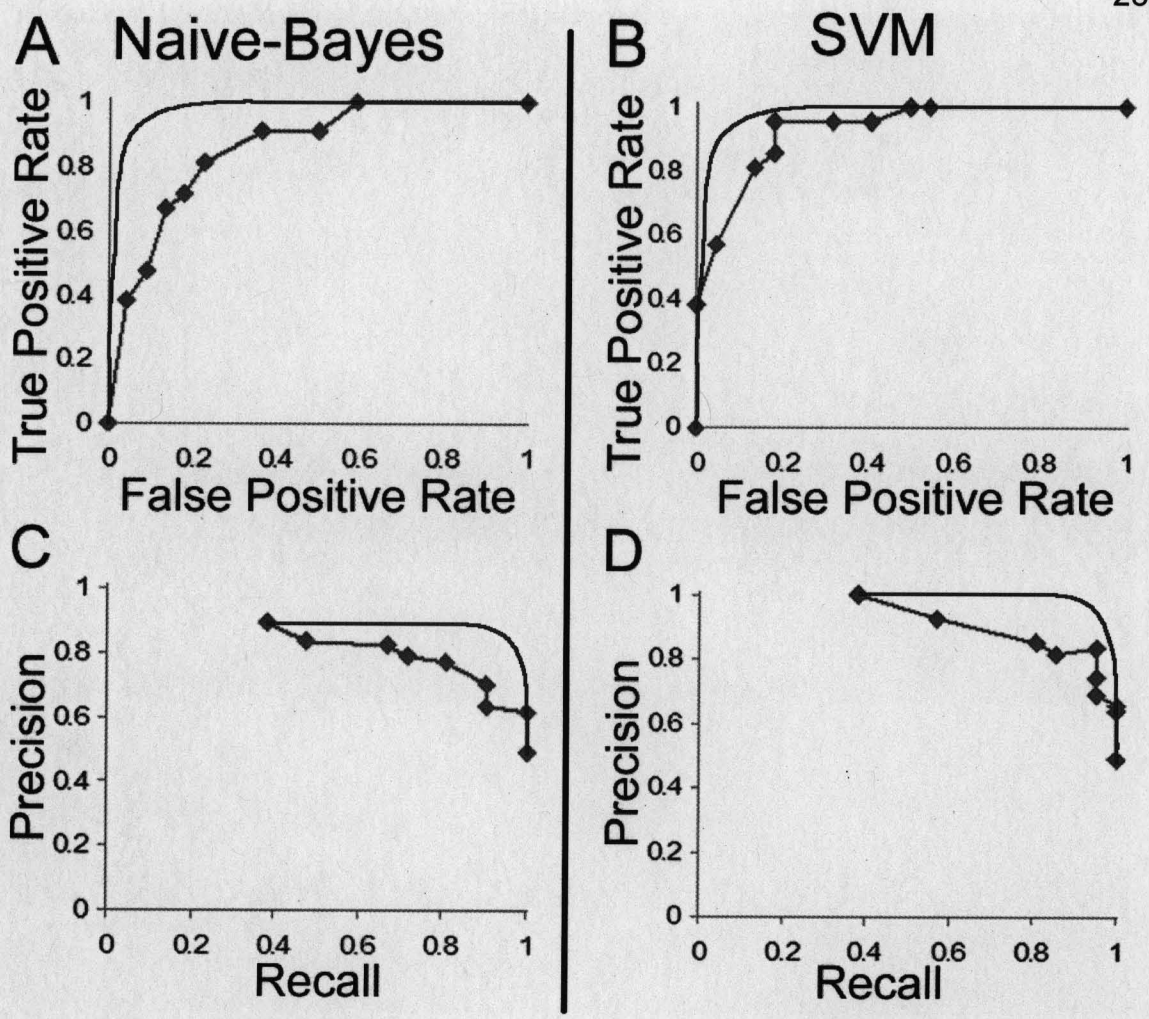

Figure 7.8: **A** and **B**. The receiver operating characteristic (ROC) curve for the Naïve-Bayes and linear support vector machine (**Table 7.3)** with the optimal ROC curve shown in red. **C** and **D**. The precision vs. recall graphs of the same with their optimal curves in red.

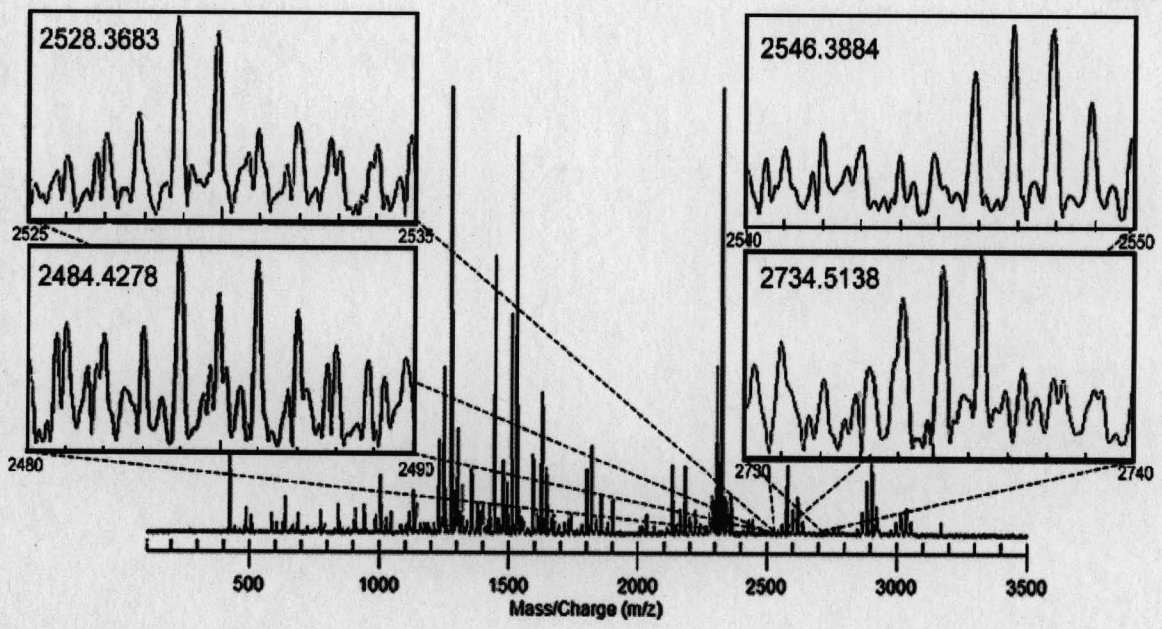

Figure 7.9: Infected CSF spectra with four most diagnostic peaks expanded.

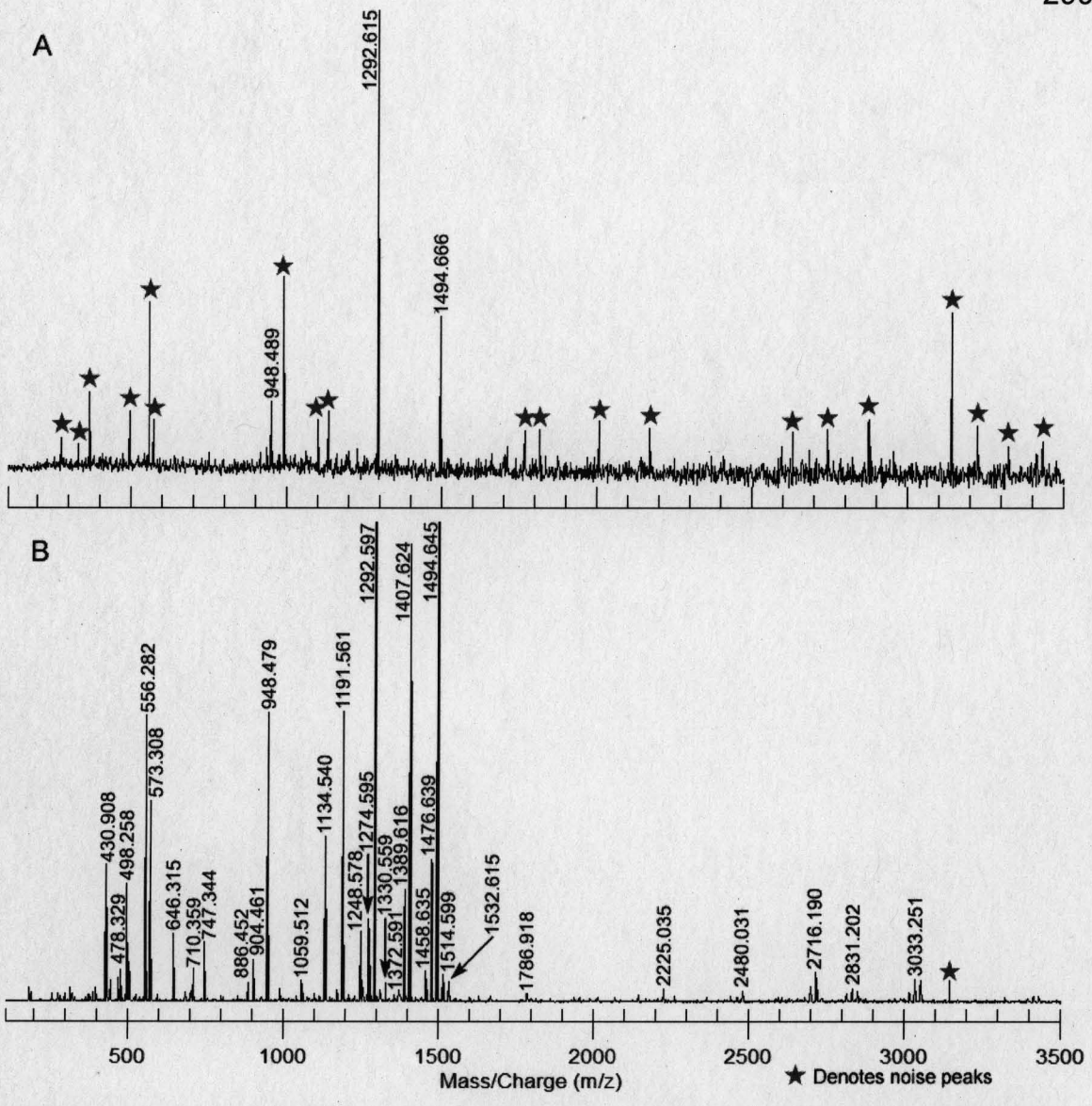

Figure 7.10: Spectra from unprocessed hamster serum (**A**) and spectra from hamster serum processed with albumin depletion column and C_{18} microspin column (**B**).

Part IV

Conclusion

Chapter Eight

Conclusions, Future Works, and Parting Remarks

8.1 Introduction

Coming in to graduate school, I was not sure where my interests lie other than to say they were in science. One thing I had realized early on was that it was integrative sciences that would lead the way into the future, and thus I was attracted to join the Pharmaceutical Sciences graduate program. I saw this program as heavily involved in both biology and chemistry as well as how this research applied to human issues. Furthermore, I could tell that the department had ties across campus to the many other science departments, thus offering a wide variety of opportunities to new graduate students. Initially, during my time rotating in labs, I was interested in naturally derived products, but found that that research did not quite fit me as I had anticipated. It was not until Lingjun Li joined the school as a new faculty member, and her research on the bioanalysis of neuropeptides in crustaceans using mass spectrometry that I found an area that I was truly interested in and where I could really apply myself. Therefore, I joined her group and have been working in this area, as well as other applicable areas, to work towards completion of my graduate requirements.

One of the things that caught my attention in Lingjun's work was her use of mass spectrometry for sample analysis. Coming from a small liberal arts college that only had a GC-MS at the time I had very little previous exposure to mass spectrometry but had done well in analytical chemistry. Consequently, I was interested in this work and its application to biological problems. As I began working in the lab, I was very quick to realize that I truly was enjoying the work and thus joined the group. The things that appealed to me in Lingjun's work

were the use and application of mass spectrometry, the work with the crustacean model nervous system, and the combination of the two to solve the biological problems addressed in Lingjun's research proposal. It was these issues, the same reasons I had joined the Pharmaceutical Sciences graduate program for, which eventually led me to succeed in Lingjun's group.

8.2 Projects

8.2.1 In-cell accumulation for the analysis of neuropeptides

I began my graduate work in the development of methods for the new MALDI-FTMS that the School of Pharmacy had recently acquired. Because a great deal of Lingjun's work required the use of this instrument, my coworkers and I worked hard to learn about this relatively new technology as well as develop methods that we could directly apply to the work we were doing. Through this process, I learned a great deal not only about MALDI-FTMS, but also about all mass spectrometric methods and how they worked. Additionally, at the same time, I began learning how to dissect the crabs used for our model system (see Chapter two) as well as how to extract peptides from the organs and prepare the samples for analysis. After developing a rather novel method with the MALDI-FTMS in combination with our research on the crustacean nervous system, we published our first paper in the October 2004 issue of Analytical Chemistry on the use of In-Cell Accumulation for the analysis of neuropeptides by MALDI-FTMS (see Chapter Three).

8.2.2 Identification of prion biomarkers using MALDI-FTMS and bioinformatics

While performing this method development, our lab was approached by Allen Herbst and Judd Aiken of the Animal Health and Biomedical Sciences here at the University of Wisconsin-Madison. At the time, they were looking for a collaboration with a bioanalytical mass spectrometry lab to assist them in their work with prion diseases. Specifically, they had hypothesized the existence of a prion disease biomarker that could be used to identify infected animals prior to clinical signs appearing. To address this problem, we began by using the cerebral spinal fluid from hamsters (our model system) infected and uninfected with the prion protein. Obviously, because of the technical challenges involved with the analysis of CSF, we were forced to develop methods for preparing the samples for MS analysis. After creating a set of methods that worked well for our task, we analyzed the samples and created a great amount of data. Because so much data was collected, we were required to use bioinformatics methods of data processing and analysis. Through this need, we began collaborating with Sean McIlwain and David Page of the Department of Biostatistics here at the University of Wisconsin-Madison. To process our data as well as utilize the information gained, we developed a bioinformatics software package that enabled us to manage large data sets and distinguish possible biomarkers from infected and uninfected samples (see Chapter Four). By using these features and training several learning algorithms, we could make predictions about infection that were eventually better than 90 % precise. As a result of this

research, we have not yet published, but plan to in the near future (see Chapter Seven) as well as carrying on with the research towards new and possible methods of biomarker identification. To assist us in the future, we have used the information we have gathered to submit a R21 grant proposal to the NIH in hopes of obtaining funding.

8.2.3 Comparative peptidomics of multiple crustacean species

During the initial stage of method development, I was afforded the opportunity to visit one of our collaborators at the University of Washington-Seattle to learn more dissection and about where many of our samples were originating. While there, I was exposed to the incredible crustacean diversity contained in the marine ecosystem there and began thinking about a way to incorporate that diversity into some of the work we were doing in Wisconsin. As a result of this, Lingjun and I, together with our collaborator in Washington, designed a project whereby we would analyze a peptide rich organ in a variety of species using mass spectrometry. The aim of this research was to examine a variety of species from many of the crustacean species families and determine similarities and differences between them. By understanding the variations and resemblances, we hope to gain some knowledge about how the peptides and peptide families diversified through speciation. To do this, our collaborators collected thoracic ganglia from a variety of species and shipped them to us for analysis. We then extracted the peptides, HPLC fractionated the extracts and analyzed each of the fractions using MALDI-FTMS. As this project carried on,

we did narrow down the scope of the species examined but at the same time increased the impact of the research through the incorporation of bioinformatics to assist in the data processing. I had realized the utility of the tool we had developed in the prion biomarker investigation and through work with Sean McIlwain and David Page, adapted the program to work with this crustacean data and assist in comparing and contrasting the peptidomes of the species surveyed. This was done by creating a heat map of known crustacean peptides contained in each species peak list and comparing it to those identified in other species. After producing a fair amount of data, and coming to conclusions from this work, we have written and submitted a manuscript to the Journal of Proteome Research and are currently awaiting response on regarding publication (see Chapter Five).

8.2.4 Distribution of peptides in organs of C. borealis *using MALDI-TOFTOF and bioinformatics*

While the comparison of the peptidomes of multiple species is informative, we were also interested in the peptide distribution in the nervous system and neurosecretory organs of one species. As *Cancer borealis* is one of the main crustacean model systems in use today, we focused on the elucidation of peptide distribution in this species organs. To do this, we collected tissue from ~25 animals, created extracts from all the organs, and then HPLC separated them. To analyze these fractions, we utilized a MALDI-TOFTOF made available to us through the Biotechnology Center here at the University of Wisconsin-Madison.

We used this instrument for its relative high mass accuracy as well as its high throughput nature and were thus enabled to collect all of our data in a matter of hours (compared to the days/weeks required for MALDI-FTMS analysis). Following data collection, we again used our bioinformatics software program to process the data and identify peptides present in each of the organs' peak lists. This work produced some interesting results and equally remarkable conclusions regarding peptide distribution in the crustacean nervous system and its potential correlation to specific function of peptides (see Chapter 6). As of yet, we have not published this work but plan to do so in the near future.

8.2.5 Other projects

Beyond these projects, I have also been highly involved in many of the other projects in the Li Lab. Of note is the mass spectrometry imaging project and paper that resulted. For this project, I was involved in many aspects of the method development including the mounting of the tissue to the MALDI plate and the application of the matrix to the tissue. This project eventually resulted in a May 2007 publication in the Journal of Proteome Research. Another project I have played a part in the identification of peptidomic differences between fed and unfed crabs. I have primarily been responsible for the dissection of crabs but have also been involved in the process of sample preparation and data processing. Although this work has not yet produced publishable work, I am sure it will in the near future. These are only two of the projects I have directly worked on in the Li Lab, but as a senior lab member I have also taught many newer

students crustacean dissection, operation of the MALDI-FTMS instrument, and consulted with them in the execution of their own projects. Therefore, my role as a senior lab member has given me the opportunity to contribute in many aspects of the research that has been going on in the Li Lab.

8.3 Future Work

While I will be leaving and not continuing this work, I do have some insight in the direction I believe these projects should continue.

8.3.1 Prion biomarker identification

I see many ways that this research could be continued and possibly strengthened. Obviously one way to improve upon what we have already gathered is to begin the process of identifying some of the classifying features. This would mean sequencing the putative peptide biomarkers and then identifying the protein they are derived from. Unfortunately, this will be very challenging as the relative abundance of these features is very low when compared to the majority of other peaks in the samples. To resolve this dynamic range and analyte suppression issue, HPLC separation and multi-dimensional fractionation would be needed as well as sample sizes increased to allow for adequate injection volumes. Another direction that should not be ignored is that of possible glycosylation of biomarkers. It is known that the PrP^{C} itself is glycosylated and that the pattern and degree of glycosylation change when being converted to PrP^{Sc}, so it may be worthwhile to take this into account when

attempting to identify possible biomarkers. Using this knowledge to enrich samples for glycosylated peptides and/or separating the samples based on glycosylation are just two of the many ways this knowledge could be used to our advantage. Already some of this work has begun in our lab and will hopefully greatly benefit the future of this research. In general, while this work has come a long way since we began four years ago, it still has a great deal to go to achieve the aims we originally set out to attain. Fortunately, there are many directions available for this research, all of which hold promise to future investigators.

8.3.2 Multiple crustacean peptidomics

This research has been one of my favorite interests as I have truly enjoyed the contact with the diversity presented by crustaceans in the Puget Sound region of Washington State. While I have performed something never done before, there is so much more that could be done in the future. As one example, my investigation was focused on the peptidomes of crustaceans from one family, but there are several families available for collection if necessary. I believe it would be truly worthwhile to investigate the peptidomes of species from each of the families that can be collected. This would provide even more evidence for how peptide families diversified through speciation. Furthermore, as is evidenced in my discovery of two novel peptides, there is certain to be a plethora of currently unknown peptides with equally unknown functions. Through a broader survey of species, many advances in understanding peptide family diversification and how speciation occurred could be gained.

8.3.3 Crustacean tissue-specific peptide expression and distribution

In the process of performing this research, I have started on what may be one of the most scientifically interesting projects of my graduate career. By analyzing the peptidome of each nervous system or neurohemal organ, I have laid the groundwork for a more thorough understanding of how peptides are distributed among these organs and thus how the potential function of a specific peptide might be involved in the organs' function. One area that needs to be explored is the identification of previously undiscovered peptides in the fractions I have already collected from each of the organs. It may be thought that because *C. borealis* is one of the most popular and well-studied crustaceans for nervous system investigations that all the peptides have been identified, but this is not true. While processing the data from these samples, I was quick to realize that there are a great number of peptides that have never been described before in any crustacean. I am sure that my colleagues will use this information to begin identifying and characterizing these new peptides in the future. Furthermore, another very interesting direction this work could follow is a more definitive peptide map among whole systems, such as the stomatogastric nervous system (STNS). As with the brain and PO, this could be performed by MS imaging of an intact STNS or a thoracic ganglia. I have already attempted this using our MALDI-FTMS, but was unable to produce adequate signal required for this work. In the future, with the development of advanced MS imaging methods though, this work could provide excellent results and further enlighten researchers on

how peptides are distributed among the nervous system. Finally, although we have investigated a majority of the nervous system and neurohemal organs, there are those that we have not surveyed at all. At some point, the inclusion of these organs into such research would extend the information that has already been gathered.

8.4 Parting Remarks

Overall, my time in graduate school has seen me grow as a scientist and a person. I have been so fortunate to have wonderful mentors throughout, especially my professor Lingjun Li. She has fostered in me what it takes to be successful in science and I hope to reflect that as I continue on in my scientific life. As one of the first Li Lab members, I feel I have helped lay some of the groundwork for how this lab operates and succeeds. I am sure that in the future, I will return and find it continuing to produce quality work with high impact and proliferating beyond what I remember. Furthermore, I will look back on my time in this lab as the time when I grew and learned the most and developed into the scientist I am.