Proteomic Analysis of Individual Drosophila Hemolymph

BY

QI ZENG

B. Sc., Renmin University of China, China, 2009

THESIS

Submitted as partial fulfillment of the requirements for the degree of Doctor of Philosophy in Chemistry

in the Graduate College of the

University of Illinois at Chicago, 2014

Chicago, Illinois

Defense committee:

Scott A. Shippy, Chair and Advisor

Timothy Keiderling, Department of Chemistry, UIC

Luke Hanley, Department of Chemistry, UIC

Leslie Fung, Department of Chemistry, UIC

Teresa Vales Orenic, Department of Biological Science, UIC

This thesis is dedicated to my loving parents and husband.

ACKNOWLEDGEMENTS

I would like to give my sincere thanks to Dr. Scott A. Shippy, who is my Ph.D research advisor and guided me through my Ph.D. program for the past five years. I appreciated his patience to teach and lead me along the path to become an independent scientific researcher and keep me fighting to “prove him wrong”.

Dr. David Featherstone and his group members from Department of Biological Science are acknowledged for their help. His group maintained and provided me fruit flies for my research and gave great insights on protein interpretation from the biological function aspect.

I want to thank David Smith, who taught me hemolymph sampling technique and worked with me in my first project. I also thanks Vitaly Avilov who worked with me for CE-MS instrument coupling and protein data analysis. I would like to thank former graduate students in Dr.Shippy’s group, Jeanita Pritess, Sujeewa Piyankarage and current graduate students Marissa Becker and Michael Delamare for helping me improve my presentation and research skills.

At last, I would like to thank Srivani Borra and Geovannie Ojeda-Torres, who are senior students in Dr.Shippy’s group. They helped me tremendously in almost all aspects during my Ph.D. career. I’m grateful for this great friendship. They are the sugar to sweeten my Ph.D. life to be a lemon.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURES **PAGE**

LIST OF TABLES

TABLES **PAGE**

LIST OF EQUATIONS

EQUATIONPAGE

LIST OF ABBREVIATIONS

2-DE	Two-dimensional gel electrophoresis
ACN	Acetonitrile
BSA	Bovin serum albumin
CBQCA	3-(4-carboxybenzoyl) quinoline-2-carboxaldehyde
CE	Capillary electrophoresis
CE-MS	Capillary electrophoresis-mass spectrometry
CSF	Cerebrospinal fluid
DDA	Data dependent analysis
emPAI	Exponentially modified protein abundance index
EOF	Electroosmotic flow
ESI	Electrospray ionization
FTICR-MS	Fourier transform ion cyclotron resonance-mass spectrometry
HPLC	High performance liquid chromatography
i.d.	Inner diameter
IEX	Ion exchange chromatography
LCM	laser capture microdissection
LIF	Laser induced fluorescence
LOD	Limit of detection
LTQ	Linear ion trap
MALDI-TOF-MS	Matrix assisted laser desorption/ ionization time of flight-mass spectrometry

MS	Mass spectrometry
NCBInr	National Center for Biotechnology Information Non-redundant dataset
o.d.	Outer diameter
ODS	Octadecylsilane
OPA	Ortho-phthalaldehyde
RPLC	Reverse phase liquid chromatography
RFU	Relative fluorescence unit
SCX	Strong cation exchange
SEC	Size exclusion chromatography
TFA	Trifluoroacetic acid
TIC	Total ion current
UV	Ultraviolet

SUMMARY

Developing analysis methods and tools to understand the chemical information of the proteome of an organism of limited volume are challenging and vital for modern proteomic studies. *Drosophila melanogaster* (also called fruit fly) is one of the ideal animal models for the study of proteomic composition because of the high degree of genome homology with the human genome. Nevertheless, it has less than 50 nL blood (hemolymph) available for collection due to its small size (about 3 mm). The goal of this thesis is to present sample preparation and analysis methods to improve the identification of proteins in volume-limited biological samples using a series of chromatography and mass spectrometry methods.

A hyphenated nano-reverse phase liquid chromatography chip column-mass spectrometry (nano-RPLC chip column-MS) method was developed to obtain proteomic information of hemolymph from an individual fruit fly. This is the first study to report a qualitative analysis of proteomic composition of hemolymph from individual adult female fruit flies. A microliter-scale protein digestion protocol was also developed to assist the digestion efficiency of limited volume sample. With the improved sample preparation method, six novel proteins were identified for the first time at the translation level. Detection of 13 proteins that are well-known in the literature speaks to the method's validity and demonstrates the ability to reproducibly analyze volume-limited samples from individual fruit flies for protein content.

Further, comprehensive prefractionation methods were developed by fabricating 2-cm long chromatography columns for individual fruit fly hemolymph samples. Detection of lower abundance proteins was enhanced with reverse phase and ion exchange prefractionation methods when followed by high mass resolution and high mass accuracy fourier transform ion cyclotron resonance-mass spectrometry (FT-ICR) and Orbitrap mass spectrometers.

With regard to the sample handling and efficient separation of limited volume samples, capillary electrophoresis (CE) is one of the ideal analysis techniques. Herein, CE was brought on-line connected with a portable ion trap mass spectrometer by developing an interface. The sheathless CE-MS hyphenated instrument, built in house, enables efficient separation and detection of a list of amino acid standards and provides an alternative fast separation and detection method for small molecule analysis. This coupled CE-MS instrument compliments the complicated FT-ICR tandem mass spectrometry identification for larger molecular weight protein samples.

The separation, detection methods and tools in this dissertation are for proteomic study of volume-limited *Drosophila* hemolymph. These tools are quite adaptable to proteomic study of any volume-limited biological sample. These methods enable identification of proteins in complicated biological complexes while minimizing possible sample contamination, excess elution and chemical alteration, which can significantly affect the protein identification quality. The results indicate that these analysis tools promote protein characterization in physiological function study.

1. INTRODUCTION

It's reported that the total number of coding genes in humans would be between 25,000 and 30,000 (Southan, 2004). Nevertheless, the total number of human proteins could be at least 100 times more than coding genes (Righetti et al., 2005). In the current post-genome era, proteomic studies have become the center of attention due to the direct impact of proteome on all physiological activities in an organism, such as membrane trafficking, chemical signaling, and nutrient transportation and so on. However the proteome could be overwhelmingly complicated when considering protein splice variants, cleavage products and numerous post-translational modifications (Southan, 2004). Understanding the function of a specific protein begins with its isolation from biological mixture and identification. However, the protein isolation and identification could be facilitated by the fractionation of a proteomic mixture into discrete groups, to reduce the sample complexity.

An ideal prefractionation method should be easy-to-perform, fast and economical. Frequently, two-dimensional gel electrophoresis (2-DE) is adapted to simplify the complexity of proteomic mixtures by isoelectric point and then, molecule weight. Approximately 30% of proteins in a cell lysate can be visualized using the classical silver-stained 2-DE gel technique, and only a portion of these proteins could be harvested sufficiently for mass spectrometric analysis (Badock et al., 2001). As a powerful analysis tool, mass spectrometry has been intensively employed in protein identification. Nevertheless, the signal suppression caused by certain well ionized species or highly abundant proteins has raised the concern about the efficiency and bias of mass

spectrometry (MS). In addition, it is challenging for MS to cover the wide dynamic range of proteins from 70% by mass to 10^{-14} by molarity or less (Martosella et al., 2005).

Among the wide range of proteins, a significant amount of proteins associated with pathological processes are secreted and transported around the whole body via blood, which is the main transportation vehicle regulating the extracellular fluid balance. Importantly, blood is the most commonly used and easy to obtain material to study complicated physiology of biological organisms. Therefore, development of methods that allows proteomic study of blood is vital to understand any chemical composition change of different developmental stages, species, genders and diseases.

It is unethical and immoral to conduct high-risk or early stage study on humans. Thus, the use of simpler animal models is widely applied for genomic, oncological, and toxicological studies. *Drosophila melanogaster* has served as an important animal model for studying molecular mechanisms underlying intracellular processes, cell-cell interaction, or the pathology of clinical diseases for a decade. Many fly genes show great similarity to human genes. In addition, flies are relatively low-cost to maintain, has short generational time and are relatively easy to manipulate genetically. As a result, the fly has been used in numerous studies of the central nervous system, metabolism pathways, and neurodegenerative diseases (De Gregorio et al., 2001; Driscoll and Gerstbrein, 2003; Shulman et al., 2003).

1.1. Fruit fly

For decades, animal models have been used to investigate fundamental genome and physiology studies of human disease from a wide variety of aspects by mimicking disease mechanisms in animal models. Among various research animal models, for instance, mouse, rat, monkey, dog and others, *Drosophila melanogaster* is a particularly popular animal model. The sequenced genome of the fly has strikingly proved the similarity of genes between humans and distantly related species, such as *Drosophila* (Greenspan and Dierick 2004). Even metabolic and signaling pathways at the cellular level are surprisingly similar. Furthermore, downstream of signaling pathways—behavioral studies have again tightened the link between fruit flies and humans. The behavior similarity with human beings has led the way to study fly circadian rhythms, learning and memory, immune and defense system, sleep disorders and others (Greenspan and Dierick, 2004).

The completion of the *Drosophila* genome sequence permitted the identification of genes in the fly that are homologous to human disease genes. Approximately two-thirds of known human disease genes have a counterpart in *Drosophila*, of which 10% of these genes are involved in neurological diseases (Chien et al., 2002). Given that similarity and simplicity with which the genome can be manipulated, fruit fly has been extensively used for neurodegenerative disease study in human, such as Hungtington's (Shulman et al., 2003), Parkinson's (Xun et al., 2007) and Alzheimer's disease (Iijima-Ando and Iijima, 2010). For instance, there is a *Drosophila* version of a Parkinson's disease gene *parkin*, causing mitochondrial swelling, and muscle and locomotor defects (Driscoll and Gerstbrein, 2003; Shulman et al., 2003). Another gene mutant was created by inserting

the human *alpha-synuclein* gene into the fly brain to manually create the metabolism pathway, in order to decipher this gene's role in Parkinson's disease (Driscoll and Gerstbrein, 2003; Shulman et al., 2003; Zoghbi and Botas, 2002). Furthermore, *Drosophila* has also been employed for drugs abuse studies (Wolf and Heberlein, 2003), such as human alcohol and nicotine addiction (Schafer, 2002). These studies are carried out by mutating genes of the fly. This again shows the feasibility and simplicity of non-human organism gene manipulation for human disease study.

1.2. Sampling techniques for small volume sample

As the start point of a workflow of research experiments, sampling and sample collection is always a challenging and critical step, which determines the quality and sensitivity of detection and analysis that follows. In the realm of volume-limited samples, the sampling method development is even more challenging. Nevertheless, with the development of modern instrument technology, there are a good number of new techniques and tools that have emerged to assist sampling from small volume biological fluid from small anatomically defined regions, for functional physiology studies. Recently, laser capture microdissection (LCM) has become a popular technique due to the ability to excise, ablate and isolate individual tissues and cells from morphologically defined spot sizes of less than 1 μm in diameter without mechanical contact or cross-contamination (Schütze et al., 2007). LCM has been widely employed on fixed cell preparation, for instance, cell smears and histological tissue section and in the clinical study of oncology (Schütze et al., 2007; Gutstein and Morris, 2007; Cheng et al., 2013). However, LCM is

critically sensitive to tissue processing and specimen preservation methods. Significant chemical alteration and molecule variability should be taken into consideration for subsequent proteomic analysis. Microdialysis is a widely used technique for extracellular fluid collection from biological systems. By using a semi-permeable membrane at the tip of a probe, analytes at a sampling site are selectively collected through the semi-permeable membrane. Nevertheless, the excessive dilution at the sampling site from the infusion solution prior to sample collection can actually cause a major loss of sensitivity. Piyankarage et al. reported a novel direct sampling method for small volume hemolymph from *Drosophila melanogaster* larvae and adult fruit flies (Piyankarage, Featherstone, and Shippy, 2012; Piyankarage et al., 2008). This method has been utilized for sampling between 20 nL and 200 nL sample without any dilution or chemical alteration, providing a powerful sampling tool for possible future analysis of nL level biological fluid from any type of biological system. In this dissertation, all sample collection of fruit fly hemolymph was performed using the direct sampling method developed by Dr. Piyankarage.

1.3. Hemolymph analysis of fruit fly

Drosophila melanogaster has an open circulatory system. In this system, the “blood” flows within the body cavity (hemocoel) and makes direct contact with all internal tissues and organs, transporting and exchanging nutrients, metabolites, and hormones. Therefore, a more accurate name for the fly’s blood is hemolymph. The volume of hemolymph varies according to developmental stages (Piyankarage et al., 2008; Piyankarage, Featherstone, and Shippy, 2012). At the larval stage, a normal female fruit fly larva can provide around

200 nL of hemolymph. Whereas, an average adult female fruit fly yields only 50 nL hemolymph if not less. About 90% of the fly hemolymph is plasma. A total of four types of hemocytes make up about 10% of the total hemolymph volume (Vierstraete et al., 2003). These hemocytes are usually involved in clotting, phagocytosis, and encapsulation of foreign bodies (Vierstraete et al., 2003). There are fewer hemocytes in *Drosophila* hemolymph compared to blood cells in human blood. However, the hemolymph contains higher concentrations of inorganic ions, amino acids, and especially proteins (Vierstraete et al., 2003).

There are mainly three types of proteins in the hemolymph classified by their function. They are 1) storage and transport proteins, 2) Enzymes/proteases and 3) structural proteins (Vierstraete et al., 2003).

1) *Storage and transport proteins*. There are a few types of proteins for storage. For example, the calcium binding protein stores a large amount of calcium ions (Vierstraete et al., 2004). One of the main transport proteins in hemolymph is lipophorin, which transports diacylglycerol between the midgut and fat body as a lipid metabolism cycle (Canavoso et al., 2001).

2) *Enzymes/proteases*. Enzymes and inhibitors are involved in nearly all activities of an organism, including digestion, reproduction, and energy synthesis. Specifically, in hemolymph, many proteases and inhibitors play an important role in immunity, since hemolymph is the frontier and only defense system against microbial infection (De Morais Guedes et al., 2005; Vierstraete et al., 2004; Vierstraete et al., 2004).

3) *Structural proteins*. Structural proteins mainly characterize muscle proteins, such as actin and myosin, these cytoskeletal proteins are often in cooperated into movement cells and organs for muscle construction. It is known that muscles controlling wings, legs, and other movement organs have high demand for structural proteins (Vierstraete et al., 2003; Okada and Natori, 1985).

Hemolymph, analogous to blood, is the most important body fluid which is the carrier of proteins, hormones, lipids and nutrients around the fly body. This fluid is also the most common subject to understand the many proteins present in *Drosophila*. Thus, a systematic understanding of protein content from hemolymph could play an important role in deciphering the physiology changes of *Drosophila.*

Vierstrate et al. have pioneered in the field of protein characterization of fruit fly hemolymph in the larval stage. They established a list of hemolymph proteins by SDS-PAGE and MALDI-TOF-MS, using around 100 pooled flies' hemolymph (Vierstraete et al., 2003). Their results showed a significant portion of enzymes identified in larvae hemolymph. Compared to their study, proteins identified in my projects showed a smaller proportion of enzymes, while a larger proportion of storage and transport proteins which are more likely appear and serve various physiological functions in the adult fruit fly hemolymph, while enzyme proteins could be vital in the early stage of larval fruit fly for development. Since then, a few of groups have also reported protein characterization focusing on a specific physiological events in fly, for example, the immune system, the hemolymph clotting (Karlsson et al., 2004; Vierstraete et al., 2004). While the research work presented in this dissertation is a fundamental survey of protein identification in the adult female fly hemolymph, attempting to provide a more detailed mapping of proteins in

the hemolymph. In addition, the innovative study of protein characterization of individual fly has also been developed, which provides a powerful tool for much more customized analysis of specific animals rather than pooled population data.

1.4. Proteomics

The term “proteome”, as an analogue to genome, was first coined by Marc Wilkins in 1996 (Wilkins et al., 1996), and then evolved into a new discipline “proteomics”. Proteomics describes the large-scale systematic study of protein structures and biophysical and chemical properties for understanding of disease pathology, cellular process, and networks at the protein level inside biological systems (Blackstock and Weir, 1999; Anderson and Anderson, 1998; Bantscheff et al., 2007).

In the post-genome era, proteomics has become a main emphasis in the study of biological systems. With DNA strictly contained inside the nucleus and mitochondria, the genome is relatively constant in every cell and over the lifetime. The proteome is completely different. Protein content differs from cell to cell, tissue to tissue and time to time. Proteomes are dynamic and exceedingly specific. Even originating from the same gene, distinct proteins may be expressed in different cell types, or in the same cell but under different conditions. In addition, after initial gene translation, most proteins have to undergo numerous unique post-translation modifications to provide different functions. There are quite a few common post-translation modifications, such as phosphorylation, methylation, acetylation and glycosylation (Olsen et al., 2006; McBride and Silver, 2001; Roth et al., 2010; Choudhary et al., 2009). In contrast to genomics, the study of

proteomics is much more complex and dynamic, reflecting the real status in the biological system. The general workflow of proteomic research is demonstrated in Figure 1.

The original approach to study proteomics that predates the field was presented by O'Farrell and Klose who used two-dimensional polyacrylamide gel electrophoresis to separate proteins by their isoelectric points and molecular weights (Klose, 1975). With modern advanced instruments, most protein separations are performed on HPLC due to its high efficiency and ability to be automated. Proteins are separated by various properties, such as hydrophobicity in reverse phase liquid chromatography (RPLC), molecular size in size exclusion chromatography (SEC), and charges in isoelectric point in chromatography (IEX) and isoeletric focusing (IEF). Vierstraete et al. separated larval hemolymph proteins according to the protein size and isoelectric point, using 2-DE. Kalsson et al. investigated larval hemolymph protein related to clotting, also using 2-DE (Karlsson et al., 2004). Both of these projects required several hundred microliters of pooled larvae. Proteomes are typically complex enough that even after liquid chromatography (LC) separation, an efficient and highly automated detection tool is needed for characterization of protein analytes.

Mass spectrometry is a modern powerful detection method for identification of analytes by their distinct mass-to-charge ratio. The high automation enables MS for high throughput screening. Although the MS spectra interpretation from a 'top-down' proteomics approach of analyzing intact protein ions is simple and easy, the ionization efficiency of an average 20-100 kDa protein is not compatible with the LC separation.

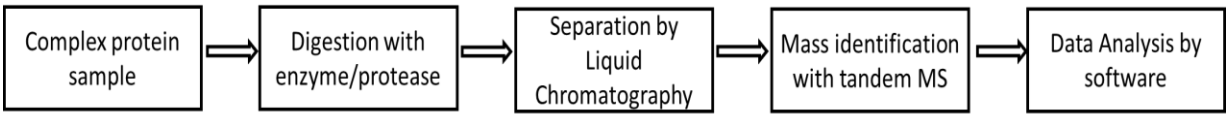


Figure 1 The workflow of proteomics

Instead, a ‘bottom-up’ (also called “shot-gun”) approach, as shown in Figure 1 *The workflow of proteomics* much more popular. In this approach, the collected sample is first digested by trypsin enzyme. Proteins are broken down to around 1 kDa peptide fragments by trypsin. These peptides can then be separated by LC and ionized in MS with ease. In the routine workflow of the shot-gun approach, MS is able to detect peptide fragments (also called “precursor ions”) of a certain protein from a mixture of tryptic peptides.

Figure 2(A) presents a typical total ion current (TIC) chromatogram of a MS survey scan. As shown in Figure 2(B), MS spectra of detected precursor ions can be extracted from any retention time, which corresponds to the TIC chromatogram. This selected precursor ion is allowed to pass to a collision cell filled with helium gas. The precursor ions are accelerated into this collision cell and broken down into fragments (also called as “product ions”). These product ions are subsequently detected (James, 1997). MS/MS spectra are generated for sequencing analysis, as shown in Figure 3.

Along with the rapid development of MS characterization for proteomics, multiple stage mass analysis with computer algorithms have been developed to process the tremendous amount of data generated by MS. The sequencing method used in this dissertation is *De novo* sequencing, which is one of the most classic and popular sequencing approaches. Normally in *de novo* sequencing, a series of product ion peaks observed in the MS/MS spectrum are assigned as b and y ions. Product ion peaks which extend from the amino

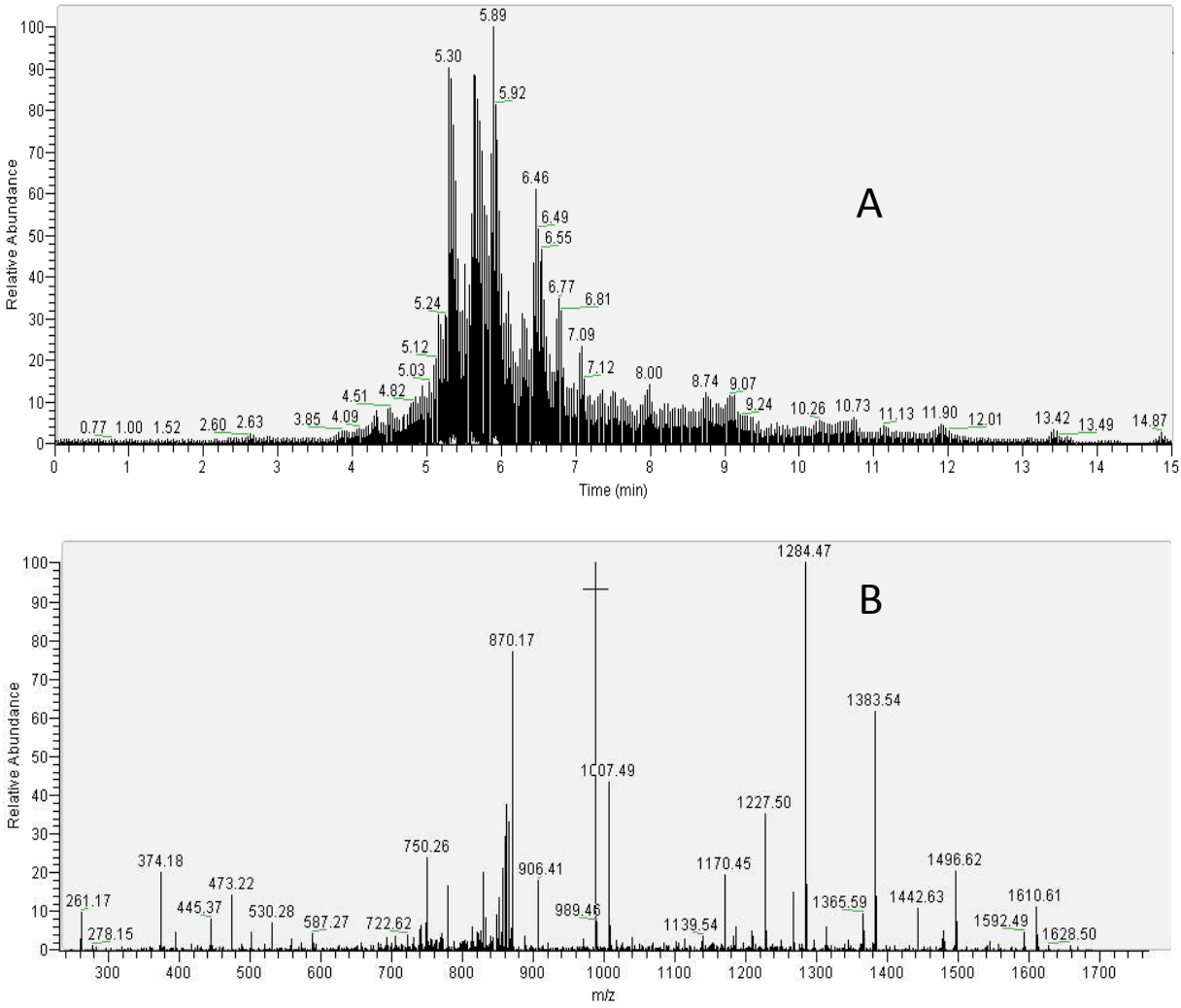
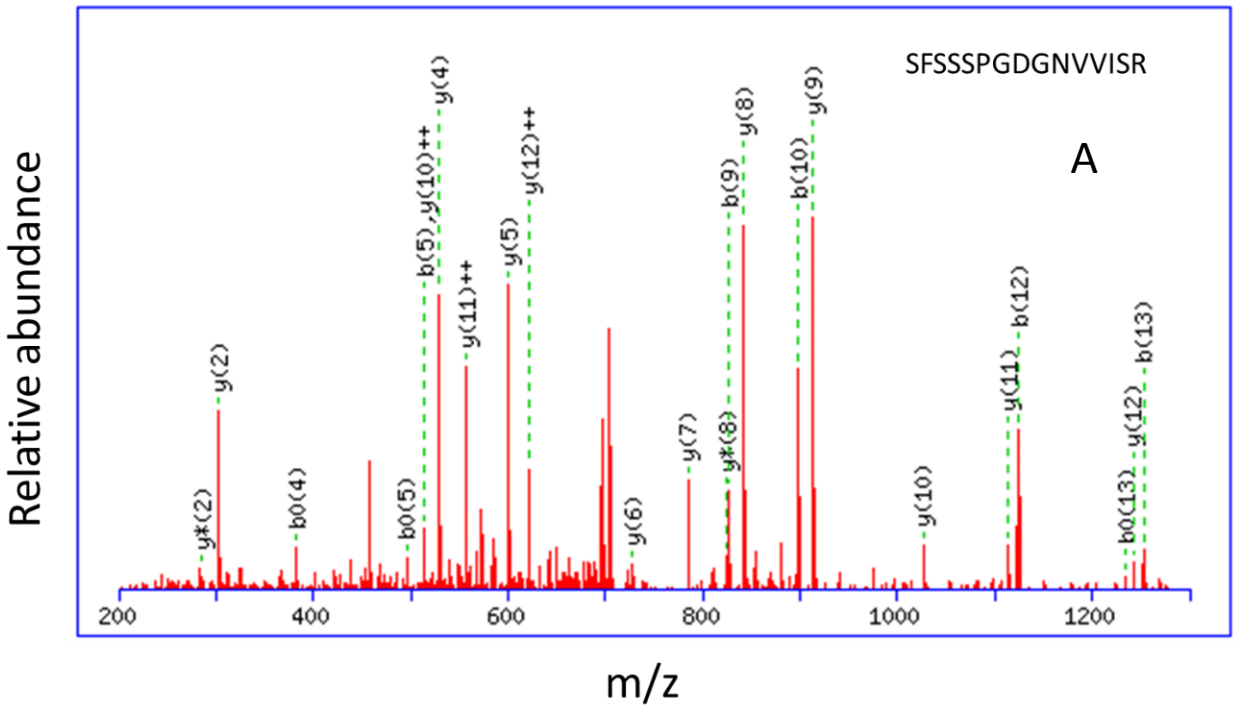


Figure 2 (A)Total ion current chromatogram (TIC) of a single fruit fly hemolymph sample. (B) A MS spectrum corresponds to the above TIC chromatogram at 6.58 min.



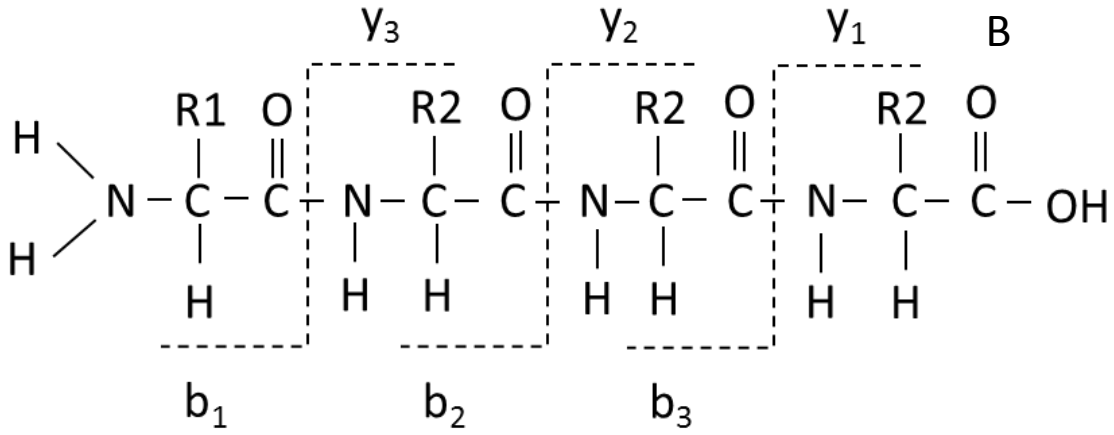


Figure 3 (A) The MS spectrum of tryptic peptide product ion sequenced as SFSSSPGDGNVVISR from hemolymph. (B) A polypeptide sequenced by assigning b and y ions, according to *de novo* sequencing.

terminus are termed as “b ions”, while peaks extend from the carboxyl terminus are name as “y ions”. Normally the difference between two neighbor y ions (or b ions) matches the mass of an amino acid. By calculating the mass difference between the neighbor b ions or y ions, the sequence of amino acids can be established and lead to the sequence of peptides, eventually for the identification of proteins.

Based on the simple conception of *de novo* sequencing, several modern complicated computer algorithms have emerged to solve the massive data interpretation task in proteomics. While this could be performed manually deciphering the sequencing from observed peptides, more often commercial searching engines, such as *Mascot* and *Massmatrix* have been well established, are highly automated, and the protein sequencing result is just a click away. For example, the searching engine software used in this dissertation is *Mascot* (www.matrixscience.com) (Koenig et al., 2008).

1.5. Prefractionation methods for proteins and peptides

Shown in the workflow in Figure 1 is a typical bottom-up proteomics process. Effective identification of proteins in hemolymph require critical and efficient analysis for each step with this volume-limited sample. Common proteomic analysis techniques could be divided into four main categories: 1) spectrophotometry, such as UV/vis, fluorescence; 2) electrophoresis, for instance, CE, SDS-PAGE and IEF; 3) chromatography, like RPLC, SEC and IEX; 4) mass spectrometry, for example, MALDI-TOF, FT-ICR and Orbitrap (Fekete et al., 2012). In this dissertation, the work was mainly focused on the latter two methods that are usually used for bottom-up proteomics.

1.5.1 Reverse Phase Liquid Chromatography (RPLC)

Liquid chromatography methods were adapted and customized for prefractionation for proteomic analysis in this dissertation. The first liquid chromatography was believed to have been performed in the early 1900s, when the Russian botanist Mikhail Tswett separated chlorophyll from plant compounds. Powdered chalk was first packed in a glass cylinder with an outlet tube placed underneath. Liquid sample solution was applied on the top of the packed chalk. Sample was then flushed down by a solvent from the top (Miller, 2005). Components in the sample absorbed on the stationary phase were desorbed and eluted off with different retention times.

In the 1940s, the column of powdered resins was replaced by a strip of paper, known as “Paper Chromatography” (Toennies and Kolb, 1950). Later on, a thin layer of powdered silica coated on a glass plate, was used as the column bed and then evolved into “thin-layer chromatography”, which is still widely used in organic chemistry labs.

The term reverse phase liquid chromatography (RPLC) was first coined by Howard and Martin in 1950 (Howard and Martin, 1950). At the time “normal phase” liquid chromatography, was more common and utilized a polar stationary phase and a non-polar mobile phase, such as hexane. However, Howard and Martin did the opposite. They found out that long-chain fatty acids were separated much better under the condition of a non-polar stationary phase and a polar mobile phase (Howard and Martin, 1950). The most common mobile phase solvent used in RPLC is water. Water brings no other contamination or alteration of analyte properties caused by a high concentration of

organic solvent. The use of environment-friendly substance—water as the solvent—also greatly reduces the consumption of toxic organic solvents.

In RPLC, the alkylated stationary phase is hydrophobic, which has strong affinity for hydrophobic protein analytes in the mobile phase. Hydrophilic proteins have weaker interactions with the stationary phase or directly pass through the column. Hydrophobic analytes can be eluted off the column by decreasing the polarity of the mobile phase, using an organic non-polar solvent. The decreasing polarity of the mobile phase solvent competes to bind the active site on the stationary phase, leaving the hydrophobic analytes unable to bind with the stationary phase, and result in elution from the column.

In this dissertation, acetonitrile was chosen as the organic non-polar solvent to adjust the composition of the mobile phase, due to its low polarity and compatibility with MS. During the elution process, a step-gradient method was used, with acetonitrile content increasing step-wise to elute analytes from the stationary phase. Eventually, all analytes were eluted off the column. Eluted peptides were subsequently detected by fluorometery, or mass spectrometry as desired, due to the compatibility of ACN solvent with these detectors.

Unlike the powdered chalk used with early liquid chromatography, a variety of finely manufactured materials are used, for example, cellulose powder and alumina (Feuerstein et al., 2005; Wincek and Sweat, 1975). One of the most popular and widely used chromatographic column packing materials are microparticulated (2-5 μm) porous silica gel. The bare silica substrates are usually chemically modified with a silane that binds n-alkyl long chain ligands to the surface. There are three commonly used ligands: n-octadecyl (C-18), n-butyl (C-4) and n-octyl (C-8). In the field of small molecules and

peptide separations, octadecylsilane (ODS or C-18) is the packing material of choice. These modified silica beads are able to bear a high mobile phase linear flow rate, with favorable mass transfer properties and fast analysis (Unger, 1990; Henry, 1991). In addition, the pores are specifically manipulated in order to prevent restricted diffusion of the solute and to increase the surface area to volume ratio of the sorbent material for analyte interaction. Thus, octadecylsilane (ODS) was chosen as the column packing material for RPLC separation of fruit fly hemolymph tryptic peptides, as described in Chapter 3. Specifically, C-18 resins were selected for tryptic peptide fractionation in that C-18 is more hydrophobic and retain peptides for a longer time with a better separation, compared to shorter alkyl chain ligands.

1.5.2 Ion Exchange Chromatography (IEX)

Other than separation by hydrophobicity, charges of analytes can be used for an alternative separation mechanism. Ion exchange chromatography was initially developed by Small in 1975 (Small, Stevens, and Bauman, 1975). Ion exchange chromatography separates proteomic mixtures based on electrostatic interaction between charged side chains of a protein and oppositely charged functional groups of

the stationary phase, governed by Coulomb's law, as shown in Equation 1 (Kopaciewicz et al., 1983).

$$F = \frac{q_1 q_2}{dr^2}$$

Equation 1: Where F is the Coulomb force between two charged ions, q_1 and q_2 are point charges of opposite sign, d is the dielectric constant of the medium and r is the distance between these two opposite charge objects.

The absorption-desorption process is governed by isoelectric point (pI) of the analyte protein or peptide and the pH of their environment. When the environment pH is greater than the pI, the analyte will have a net negative charge (q_1), and have a strong electrostatic interaction with anion-exchange stationary phase (q_2). When the environment pH is lower than the pI, the analyte will have a net positive charge, and bind to cation-exchange stationary phase, which is negatively charged. For instance, the strong cation exchange (SCX) stationary phase was used for the prefractionation of *Drosophila* hemolymph tryptic peptides, as described in Chapter 3. Strong cation exchange resins are a good IEX column material to try for an initial separation of an unknown tryptic peptide complex. Because SCX resins are negatively charged in aqueous buffer above pH 2 and most tryptic peptides in acidic pH are characterized by a charge of +2 or above, these peptides are attracted to and subsequently separated by a SCX column from peptides with a net charge of +1, such as trypsin-generated phosphopeptides, C-terminal peptides, or peptides with blocked termini, like peptides with blocked N-terminal free amine group, and peptides from higher charged peptides (Edelmann, 2011). With an increment of pH of the eluent solution, various peptides reach their pI value, when they carry zero net charge. The zero net charged peptides lose the electrostatic interaction with the stationary phase and are eluted from the column accordingly. The size of analytes also plays an important role in this charge-charge interaction. As shown in Equation 1, the Coulomb force is inversely related to the distance r. It indicates the larger charged analytes have weaker interaction with the stationary

phase than the interaction between smaller analytes with the station phase. This concept supports the 'bottom-up' proteomic approach, where proteins are fragmented to peptides prior to LC separation.

Besides a pH step gradient, increasing ion strength of the eluent can be used for a step-by-step desorption of analytes. And eventually even the most positively charged analytes were eluted from where the cation of eluent (usually NaCl) replaces all analytes bound on the stationary phase. SCX chromatography has been a widely applied to proteomic separations, and frequently, combined with RPLC as a prefractionation methods prior to MS characterization. Herein, this dissertation explores the possibility of customizing an exceedingly short SCX column to aid the fractionation of proteomic complex from individual adult female fruit fly hemolymph.

1.5.3 Size exclusion chromatography

Size exclusion chromatography, invented by Lathe and Ruthven in the 1950s (Lathe and Ruthven, 1956), is a widely applied technique for protein separations. The principle underlying SEC is that analytes elute from the column at different retention time, due to their size differences. A molecule, which is below the size of the stationary phase resin pore size, is able to penetrate into every region of the stationary phase. Therefore it is retained on the column for a longer time than a larger molecule that cannot diffuse into a stationary phase pores. Analyte molecules above the size of the stationary phase pore do not diffuse into stationary phase pores, and subsequently elute more quickly from the column (Wang et al., 2010; Lathe and Ruthven, 1956).

As the separation of analytes is solely governed by the size, the stationary phase is treated to have no chemical or physical interaction with analytes. The mobile phase consists of buffer which stabilizes proteins and prevents denaturation or aggregation. The separation mechanism is simple and SEC generally is considered as a low resolution chromatography, because it does not have high enough resolving power to separate analytes with similar molecular size. In this dissertation, a SEC separation was developed prior to MS identification.

1.6. Detection methods for proteins and peptides

1.6.1. Fluorescence detector

Detection is one of the most important steps in protein identification, as shown in Figure 1 *The workflow of proteomics*. UV-Vis and MS are two most commonly used detection methods in proteomic analysis. However, these two universal methods have their own limits: due to their inherent properties, they are not able to differentiate analytes of interests from complex matrix backgrounds, when the matrix possesses similar light absorbance or mass-to-charge properties. In this case, fluorescence stands out for its high sensitivity and high selectivity of analytes from complex matrices.

Fluorescence is one type of luminescence. When an analyte molecule with fluorophores absorbs light or electromagnetic radiation, the valence electron of the molecule becomes excited, and elevated from the ground state to an excited state. When it relaxes back to its ground state, the molecule emits a photon, which is selectively

detected by fluorescence detector. Typical limits of detection (LOD) of fluorescence detectors are able to reach as low as 10^{-9} to 10^{-18} M (Waterval, 2002). Fluorometry can achieve sensitivities 1000-fold higher than UV/vis spectroscopy. Proteins with tryptophan are fluorescent by nature. Proteins without a fluorophore can be derivatized with fluorogenic reagents.

There are a variety of fluorescence derivatizing agents available (Waterval, 2002). Ortho-phthalaldehyde (OPA), dansyl chloride and 3-(4-carboxybenzoyl) quinoline-2-carboxaldehyde (CBQCA) are all popular derivatizing agents that are commercially available. In 1972, Udenfriend et al. first reported fluorescamine as a derivatizing reagent for amino acids and peptides, reaction as shown in Figure 4 (Udenfriend et al., 1972). Fluorescamine specifically binds with primary amine groups which are essential functional group in amino acids, peptides and proteins. The derivatized product is excited around 365 nm and emits fluorescence around 490 nm. The fluorescamine derivatization reaction has several advantages. The derivatization reaction has high quantum yield and is fast, taking less than 1 sec, at room temperature and is pH tolerant from 7 to 9. Fluorescamine itself is not fluorescent, avoiding interference with the product fluorescence (Udenfriend et al., 1972). Fluorescamine was utilized as the

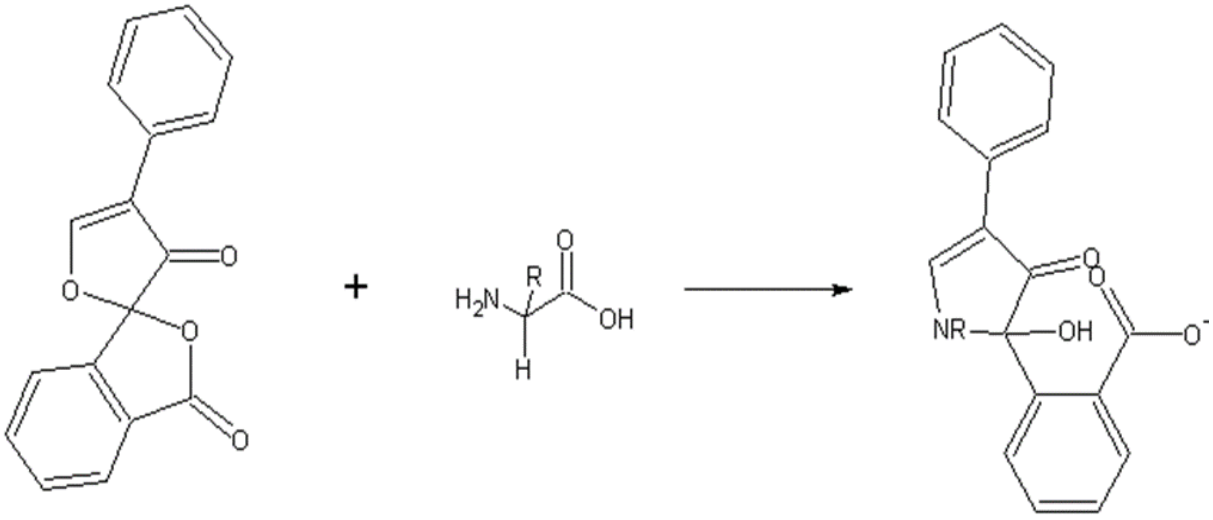


Figure 4 Fluorescamine derivatization reaction.

fluorescence derivatizing reagent for protein study of fruit fly hemolymph in this dissertation. A home-made fluorescence detector was well suited for the limited volume sample following fractionation on the nano-scale liquid chromatography method developed as described in Chapter 3.

1.6.2. MS detection methods for proteomic analysis

MS has become popular because of its robustness and unique chemical identity information by determining the analyte structure. A mass spectrometer detects charged analyte ions by mass-to-charge ratio. With the recent hyphenated technology development, MS has been coupled with gas chromatography, high performance liquid chromatography (HPLC) and so on. These coupled instruments possess high sensitivity and high resolution, which serves as the dominant analysis tool at the final step of proteomic research, as shown in Figure 1.

1.6.2.1 Matrix Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry (MALDI-TOF-MS)

Because of the important role MS plays as the final identification step in the proteomic workflow, as demonstrated in Figure 1, a closer look into the advantages of various MS is needed. The key to meeting the challenge in protein identification by MS is getting protein ions in the gas phase.

In the 1980s, two of the most important ionization and desorption techniques were invented to introduce liquid and solid phase of analytes into gas phase for mass spectrometry. They are electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI).

These two techniques are the most widely used ionization methods for MS. In the application of imaging and solid sample analysis, MALDI has unique advantages, due to its simple and convenient matrix-analyte co-crystallization sample preparation. Furthermore, the singly charged analyte ion peaks that result from MALDI allow interpretation of MALDI MS spectra with ease. In this dissertation, MALDI was used to perform a fast test of the micro-scale protein digestion protocol in Chapter 3.

In order to vaporize macromolecules like proteins and peptides, a matrix, usually an organic acid, is used to co-crystallize with analyte solution on a stainless steel MALDI plate. There has not been an unambiguous explanation of the rationale of MALDI desorption reaction, although several mechanisms have been proposed by different groups (Knochenmuss and Zenobi, 2003; Karas and Krüger, 2003). A generic illustration is as follows: As the laser impinges on the co-crystallized spot, the matrix exhibits a strong resonance absorption and transfers the absorbed energy to analytes, which induces a “uniform and soft desorption” followed by ionization caused by collision with matrix ions (Karas and Krüger, 2003). Once ionized, analyte ions are sent into a mass analyzer, which usually is a time-of-flight (TOF) tube. The TOF tube is basically a 1 m-long vacuum tube. The travel time of ions in the TOF tube is related to ion kinetic energy. The mass-to-charge ratio can be measured according to the equation shown below:

$$\frac{m}{q} = \frac{\sqrt{2U}}{v}$$

Equation 2: U is the voltage in the time-of-flight tube. m is the mass of analyte. q is the charge of analyte. v is the velocity of analyte.

Since MALDI is a soft ionization method, the intact form of the analyte of interest is usually maintained. Proteins and peptides are normally singly charged in MALDI, which greatly simplifies the MS spectrum. Due to the benefits for proteomic characterization, MALDI-TOF-MS was used as a quick identification method for micro-liter volume trypsin digestion of protein mixture in this dissertation.

1.6.2.2 Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (FTICR-MS)

About four decades after the initial accomplishment in employing ion cyclotron resonance (ICR) principle into mass spectrometry, Comisarow and Marshall brought Fourier transform method into ICR mass spectrometry, which eventually evolved into the first Fourier transform mass spectrometry (Schmid et al., 2001; Amster, 1996; Marshall, 1974). Thanks to its high resolution and high mass accuracy, FTMS has been widely adapted in various scientific research fields, for example, small molecules as nucleic acids (Hofstadler, Sannes-Lowery, and Hannis, 2005), and macromolecules, most notably proteins and peptides from mammalian systems (Nittis et al., 2010; Brinkmalm et al., 2012).

This versatile mass analysis tool has four main features:

1) *Superconducting magnet*. With unsatisfactory results with permanent or electromagnets, which provide relatively weak or unstable magnetic fields (less than 2 T) for FTMS, superconducting magnets have become the dominant material for generating strong and stable magnetic field for FTMS. Field strengths between 3 and 9.4 T can be easily achieved and these fields increase the mass resolution of the instrument.

2) *Cylindrical open-ended analyzer cell*. The analyzer cell is claimed as “the heart of FTMS” (Amster, 1996). There are two main types of commercially available analyzer cells, as shown in Figure 5. The cubic cell is the earlier design and is made by 3 pairs of opposing electrode plates. One pair of plates is placed perpendicular to the magnetic field orientation (as illustrated by the dash line arrow in Figure 5). This pair of plates are also called trapping plates, allowing ions entering the cubic cell with a hole in the middle of plates. While trapped in the cubic cell, analyte ions are excited and detected by the other two pairs of electrode plates which are parallel to the magnetic field orientation (Schmid et al., 2001; Amster, 1996).

An alternative design is the cylindrical open-ended cell. The name itself explains the unique cylinder shape of the cylindrical open-ended cell. This specially designed cell shape, allows a better fit into the bore of a superconducting magnet, compared to the cubic cell. Thus the cylindrical open-ended cell has become the most popular analyzer cell in FT-ICR MS (Amster, 1996).

3) *Ultra-high vacuum system*. To avoiding analyte ion collision and loss of ions before detection, an ultra-high vacuum system is necessary for FTMS. Pressures down to 10^{-9} torr need to be assured for FTMS operation (Schmid et al., 2001; Amster, 1996).

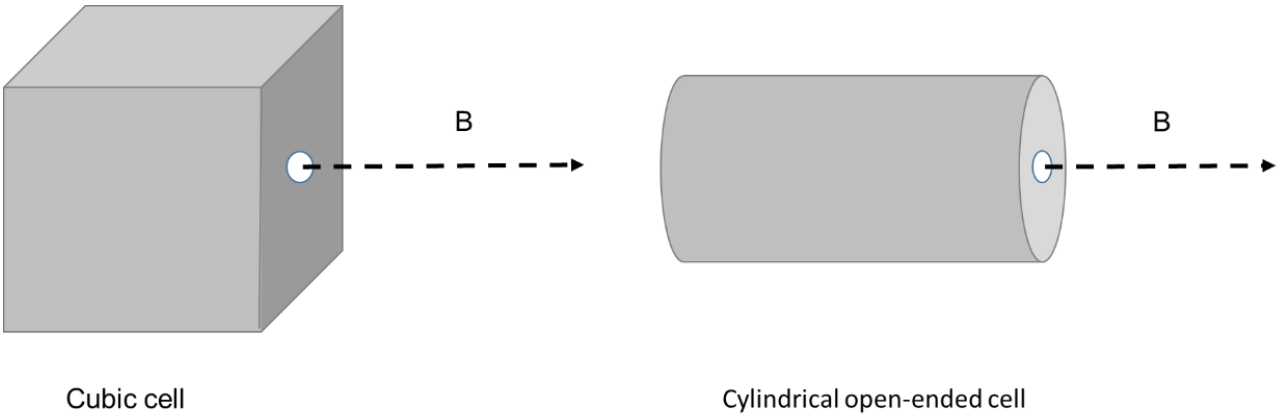


Figure 5 Two main types of analyzer cells in FTICR-MS. The arrows shows the orientation of the magnetic field. B represents the magnetic field strength.

4) *Fourier transform data processing*. The FTMS utilizes the Fourier transform method to convert the frequency signal to mass-to-charge data, as shown in Figure 6.

Ion motions in FTMS are complicated. They can be roughly divided into three types of motions.

i) Trapping motion. Trapping motion describes the ions moving parallel to the magnetic field. Trapping motion is not affected by magnetic field but only controlled by the electric field. The trapping motion follows the generic kinetic energy rule, as demonstrated in Equation 3.

$$E = \frac{1}{2}mv^2$$

Equation 3: Where E is the kinetic energy. m is the mass of analyte ions. And v is the velocity of ions.

ii) Cyclotron motion. Cyclotron motion is the essential part of FTMS. This motion is caused by the Lorentz force, which makes ions circulate in an orbit that is perpendicular to the magnetic field. Ions are separated by their unique cyclotron frequency, which is inversely related to mass-to-charge ratio, as indicated in the Equation 4 (Schmid et al., 2001).

$$F = qB/(2\pi m)$$

Equation 4: Where F is the Lorentz Force. B is the magnetic field strength. q is the electric charge of analyte ion. m is the mass of the analyte.

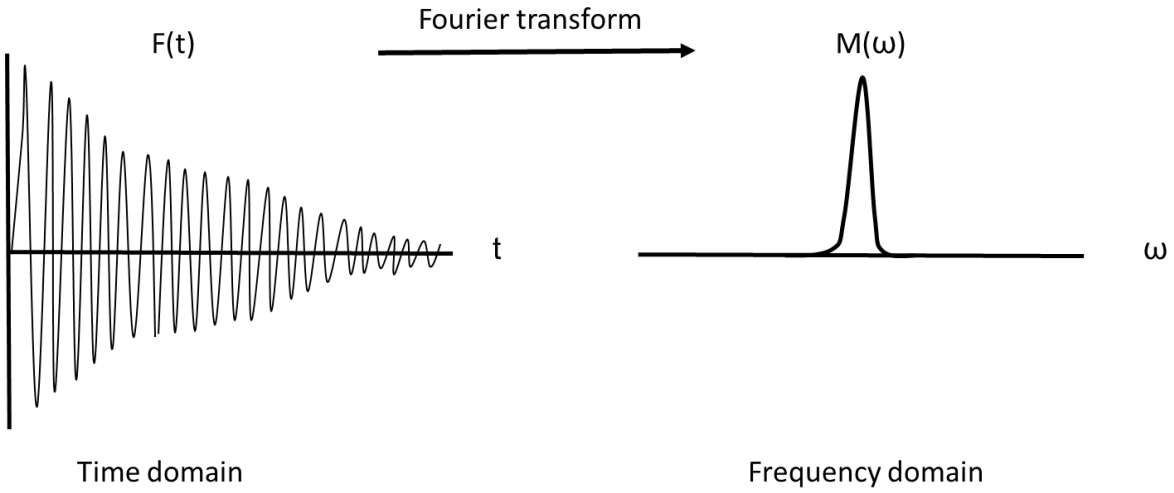


Figure 6 Time-domain ICR signals were converted to frequency-domain magnitude signal by Fourier transform algorithm.

iii) Magnetron motion. Unlike the previous two types of motion, which are only impacted by single type of field, magnetron motion is ruled by a combination of the electric and magnetic fields. This third fundamental ion motion is the reason why FTMS can keep ions stored in an analyzer cell for seconds, minutes or even hours. This special feature of FTMS is distinct from most regular mass spectrometers (Amster, 1996).

In particular, the high resolution and high mass accuracy of FTMS are advantages for proteomic study of limited volumes of complex biological samples as shown in Chapter 3. The mass resolution of FTMS is one of the highest among all mass spectrometers. FTMS can achieve high resolution up to 100,000 while the mass accuracy of FTMS can be less than 5 ppm (Marshall and Hendrickson 2008). Because the high mass accuracy can be employed as an extra constraint in searches of experimental peptide mass against genome sequence data bases and the high mass resolution allows simultaneous analysis of multiple mass similar analytes, FTMS is particularly attractive for proteomic researchers (Heeren et al., 2004). Wilkins's group has identified bacterial proteins directly from whole cells without cell lysis or purification utilizing FTMS coupled with MALDI (Jones et al., 2003). Brinkmalm et al. have characterized several new endogenous amyloid precursor proteins and peptides in human cerebrospinal fluid (CSF) by FTMS using electron capture dissociation to advance understanding the mechanism of Alzheimer's disease pathology (Brinkmalm et al., 2012).

In this dissertation, FTMS coupled with RPLC has been used for identification of proteins in limited volume of individual adult fruit fly hemolymph in Chapter 2. The high mass resolution and high mass accuracy of FTMS helped to identify novel proteins from less than 100 nL fruit fly hemolymph sample complex without sample handling and purification.

1.6.2.3 Orbitrap Mass Spectrometry

Although FT-ICR, is a powerful mass analyzer tool for protein identification as the last step in the 'bottom-up' proteomic approach, as illustrated in Figure 1, it is also extremely expensive to maintain this highly complicated instrument. In 2000, Makarov built the first Orbitrap mass spectrometer, which has a simplified and compact design compared to FT-ICR and demonstrated high throughput screening in MS (Makarov, 2000). The first Orbital trapping (also named as “the Kingdon trap”) was actually implemented by Kingdon in 1923. This orbit trap employs only electrostatic fields for ion trapping, without magnetic field or RF electric fields (Makarov, 2000). Makarov adapted the orbital trapping idea from Kingdon and customized it to be a new type of mass analyzer. As illustrated in Figure 7, the Orbitrap mass analyzer is co-axially assembled with an outer barrel-like electrode and an inner spindle-like electrode (Makarov, 2000). Ions oscillate along the axis of the electrostatic field and travel in a complex spiral pattern. The harmonic oscillation is achieved when the electrostatic attraction towards the central spindle-like electrode is balanced by a centrifugal force caused by the initial tangential velocity of ions, similar to “a satellite on orbit” (Scigelova and Makarov, 2006). The axial frequency of ions oscillation

can be detected as an image current, which can be converted from frequency domain to time domain using a Fourier transform algorithm (Hu et al., 2005). Therefore, the mass-to-charge ratio, which is proportional to axial frequency, can be derived from Equation 5 (Makarov, 2000).

Unlike FT-ICR, the Orbitrap only utilizes the axial frequency of an ion's harmonic oscillation, which is independent of trapping energy and spatial spread of ions. There is increased space charge capacity at higher masses so higher ion fluxes can be trapped and analyzed (Makarov, 2000). Accurately defined electrostatic fields have also elevated resolution up to 100,000, almost to the same level as FT-ICR. In addition to these above advantages, the Orbitrap mass analyzer is very time-efficient, as less than one second is required to obtain a MS spectrum with resolving power of 60,000. Because of these advantages the Orbitrap is a powerful tool for demanding proteomics studies and high throughput screening (Scigelova and Makarov, 2006).

$$\frac{m}{z} = k/\omega^2$$

Equation 5: Here ω is the frequency of ion oscillation along the z-axis. k is the potential (set as a constant) between electrodes.

Orbitrap mass spectrometry has become popular for proteomic applications. Schutzer et al. has been devoted to establishing the proteome of normal human CSF using linear ion trap-Orbitrap MS/MS (LTQ-Orbitrap MS/MS). A total of 2630 proteins in CSF were successfully identified, with 56% classified as CSF-specific (Schutzer et al., 2010).

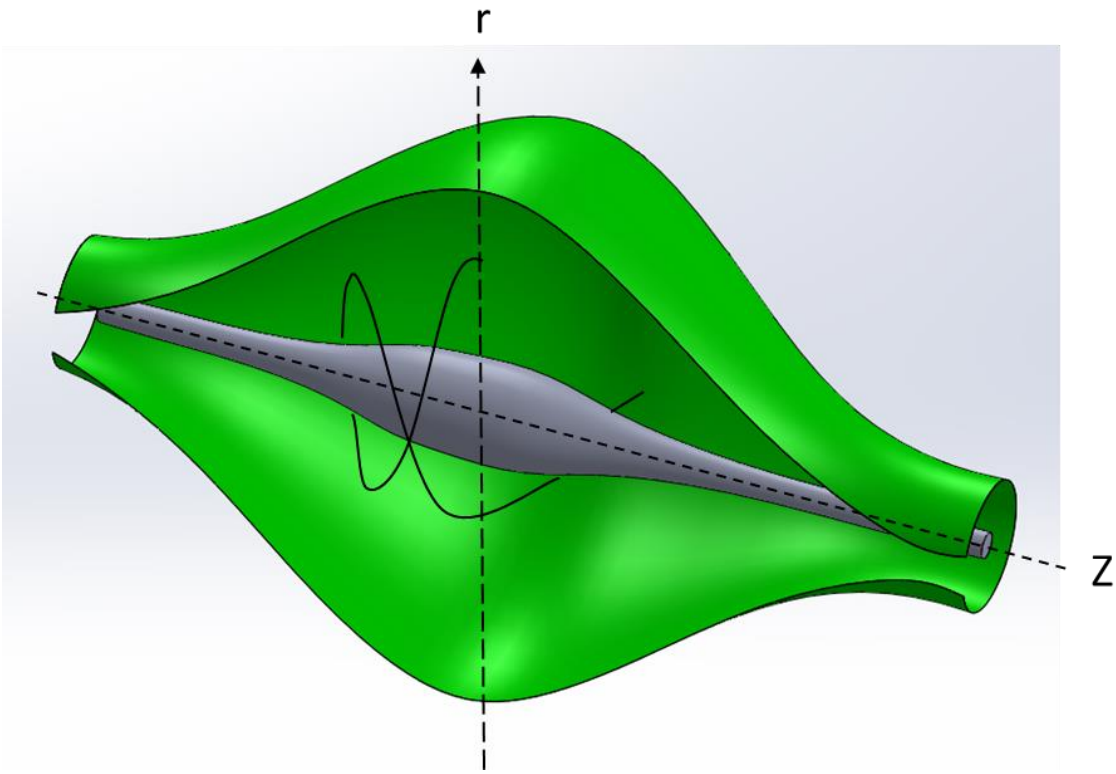


Figure 7 Schematic structure of Orbitrap. The Orbitrap mass analyzer is co-axially assembled with an outer barrel-like electrode and an inner spindle-like electrode. Ions oscillate along the axis of the electrostatic field and travel in a complex spiral pattern.

The Orbitrap mass spectrometer has become a quite versatile and powerful tool in proteomic study. It has greatly decreased the analyzing time and reduced the cost for research. With minimized analysis time, the Orbitrap still meets the high standard of mass resolution and mass accuracy, which is needed for reduction of false positive peptide identification. An Orbitrap MS was used as the mass analyzer for protein and peptide identification in fruit fly hemolymph samples, as presented in Chapter 4 of this dissertation. The Orbitrap MS is coupled with a linear ion trap in the study. High mass resolution and mass accuracy spectra of precursors in the Orbitrap were acquired, while using the linear ion trap as a collision chamber for tandem MS fragmentation and detection.

1.7. Objectives and organization

The overall goal of this dissertation is focused on developing novel proteomic analysis techniques and tools for obtaining qualitative protein information of individual adult fruit fly. Chapter 2 describes an analysis method developed to collect proteomic information from individual fruit fly hemolymph using an Agilent nano-HPLC chip system and then assayed via LTQ-FT MS/MS. From the study, a number of proteins were positively identified, including six proteins that have only been predicted by evidence at the translation level. These findings were submitted as a research article to Journal of Chromatography B in June, 2014. Chapter 3 demonstrates a low pressure, customized RPLC column and an IEX column to prefractionate hemolymph proteins prior to LC-MS/MS analysis for protein identification. The studies reported in this chapter were submitted to Analytical Chemistry in July, 2014. Chapter 4 describes the construction of

an interface to couple CE with mass spectrometer. The interface was fabricated with conductive carbon and brought CE online for direct mass analysis of amino acids by MS. Chapter 5 consists of three preliminary studies that have potential for future progress. The first component presents a measurement of total body fluid of individual fruit fly. The development of a SEC prefractionation method for proteomic analysis of hemolymph is described in the second part of Chapter 5.

2. PROTEIN ANALYSIS OF INDIVIDUAL FRUIT FLY HEMOLYMPH

[submitted to Journal of Chromatography B in June,2014.]

2.1. Introduction

For decades, *Drosophila melanogaster* has been an important animal model for studying molecular mechanisms underlying intracellular processes, cell-cell interaction, or the pathology of clinical diseases. It affords many biological advantages such as an easily manipulated genome, a relatively short lifecycle, and a large body of genome data that is a powerful resource for proteomic methodologies. However, the exceedingly small-size of fruit fly yields biological volumes in the low nanoliter level, which makes sample collection, handling and preparation quite a challenge. Methods to perform proteomic analysis on nL volumes of fluids from the fly would be an enabling tool for fly studies.

The physiology of the fly is relatively simple. It has a short lifecycle that makes maintenance in the laboratory straightforward and economical. With only four chromosomes, its genome system is significantly simpler compared to human (Adams et al., 2000). This simplicity contributes to its power as a model and makes manipulation of the fly genome relatively easy. The fly genome is also homologous to almost three-quarters of the human genome (Adams et al., 2000). Importantly, a number of *Drosophila* genes were found to have human analogues relevant to clinical diseases. This great similarity and relevance to the human genome has prompted close studies of gene expression, metabolites digestion pathways, and proteomic functions in *Drosophila* melanogaster biological system.

Unlike mammals, the fly has an open circulatory system in which hemolymph circulates through the whole body. Hemolymph, similar to blood, transports and exchanges nutrients, metabolites, and antibodies between organs and tissues. There is ongoing active chemical signaling between internal tissues and organ cells by proteins, peptides, and hormones in hemolymph (Grewal, 2012). Many of the secreted proteins in hemolymph serve as the innate immune system for the fly and protect against infections and external invasion (Vierstraete et al., 2004). Thus, hemolymph is an ideal choice for proteomic analysis in order to better understand the effects of particular protein post translation modifications and, more generally, the physiology of disease development in the fly.

A diversity of approaches have been proposed to isolate specific cell populations, tissues and body fluid from small anatomically defined regions, for functional physiology studies. Recently, laser capture microdissection (LCM) has become a popular technique due to the ability to excise, ablate and isolate individual tissues and cells from morphologically defined spot sizes of less than 1 μm in diameter without mechanical contact or cross-contamination (Schütze et al., 2007). LCM has been widely employed on fixed cell preparation, for instance, cell smears and histological tissue section and in the clinical study of oncology (Schütze et al., 2007; Gutstein and Morris, 2007; Cheng et al., 2013). However, LCM is critically sensitive to tissue processing and specimen preservation methods. Significant chemical alteration and molecule variability should be taken into consideration for subsequent proteomic analysis (Zhu et al., 2013). A proteomic profiling of *Drosophila* larval hemolymph clotting has been established by others quantitatively and qualitatively, using a group of 10 animals (Karlsson et al., 2004). The

study of hemolymph coagulation is impressive, but requires tedious and precise sample preparation, in order to harvest and pool microliters of hemolymph solution for SDS-PAGE separation prior to proteomic analysis. To obtain soluble proteomic information of *Drosophila* antenna from a population of flies, Anholt and Williams described an osmotic lysis method of hundreds of antennae in distilled water, followed with homogenization and centrifugation, which required several hours of processing (Anholt and Williams, 2010).

In this Chapter, I present an analysis method to collect proteomic information from individual fruit fly hemolymph which utilizes a unique small-volume sample collection technique developed in our lab previously (Piyankarage et al., 2008; Patterson et al., 2008). A microscale in-solution digestion procedure has been established for efficient low volume hemolymph digestion. Peptides resulting from a tryptic digest of hemolymph were separated using an Agilent nano-HPLC chip system and then assayed via LTQ-FT MS/MS. From the study, a number of proteins were positively identified including six proteins that have only been predicted by evidence at the translation level.

2.2. Experimental section

2.2.1. Chemicals

Deionized water was obtained from a US Filter Purelab Plus purification system (Lowell, MA). Trypsin, Tris hydrochloride, dithiothreitol and iodoacetimide were all purchased from Sigma-Aldrich (St. Louis, MO). Glacial Acetic acid was purchased from

Fisher Scientific (Itasca, IL). Bradford reagent Coomassie blue G-250 solution was from Thermo Scientific (Itasca, IL).

2.2.2. Materials

Fused-silica capillary (360/50-µm o.d. /i.d.) was purchased from Polymicro Technologies (Phoenix, AZ). Tygon tubing (250-µm/20.6-mm o.d. /i.d.) was purchased from Cole-Parmer (Vernon Hills, IL). Insulin injection syringes with a 1-mL volume were from Becton Dickinson Co. (Frankin, NJ). Centrifuge tubes were purchased from USA Scientific (Ocala, FL). The 0.400-mL Spectrosil Far UV quartz window cuvette cell is from Starna Cells (Atascadero, CA).

2.2.3. Fruit fly sample collection

Hemolymph was collected from the Oregon-R strain of *Drosophila melanogaster*, reared on standard cornmeal-agar medium (Piyankarage et al., 2008), and maintained in the UIC Biological Sciences Department. Only female flies were selected and sampled for hemolymph, in order to avoid the gender differences and to access larger hemolymph sample volumes. The sample collection technique was adapted from a protocol described previously (Piyankarage et al., 2008). Hemolymph samples were collected from individual fruit flies after 5-min cold-anesthetization. A single fruit fly was placed on a micro-dissection plate with the head taped to the dish, the wings cut, and the dorsal surface of

the abdomen under view with a dissection microscope. A small incision was made with a microdissection scissors at the second abdominal tergite. The released hemolymph was immediately collected using a probe consisting of a 4 cm-long fused-silica capillary. This sampling probe was held by an electrode holder positioned in a three-dimensional micropositioner to adjust the probe position while collecting hemolymph. Collected hemolymph samples were then transferred to a 250 µL centrifuge tube and left to air-dry with the tube cap open and then stored at -20°C before analysis.

2.2.4. Proteomic sample preparation

2.2.4.1 Development of micro-scale digestion protocol

A volume of 50 nL, 1 µM Bovine Serum Albumin (BSA) was thermally denaturized at 90 °C for 20 min in 2 µL of 0.4 M pH 7.8 stock Tris buffer before adding 2 µL of reducing agent (DTT) and alkylating reagent (iodoacetamide). Next, an aliquot of 0.2 µg/µL trypsin in 0.4 M stock Tris buffer solution was added to the BSA solution, giving a 1:50 protease-to-substrate molar ratio. The digestion was carried out at 37 °C overnight. Trypsin-digested BSA peptides were detected in ABI4700 MALDI-TOF mass spectrometer (Applied Biosystems, Foster City, CA) to confirm digestion efficiency.

2.2.4.2 Digestion and analysis of fruit fly hemolymph sample

With confirmation of digestion efficiency in BSA, the above microliter-scale digestion protocol was applied to each 50 nL fly hemolymph sample. Digested hemolymph samples were frozen at -20 °C until analysis. In order to estimate total protein amount in hemolymph sample regarding instrumental loading restriction, a Bradford assay absorption calibration curve of digested BSA was generated. A series of concentrations (0-30 µg/mL) of BSA tagged with Coomassie blue G-250 were measured at 595 nm by a Cary 300 Bio UV-Visible Spectrophotometer (Agilent Technologies, Santa Clara, CA).

2.2.5. Separation and MS identification

Digested hemolymph samples were separated using an Agilent large capacity nano-HPLC chip system. This nano-HPLC chip is essentially a microfluidic device, with laser-ablated channels designed on a polyimide film (Yin and Killeen, 2007; Yin et al., 2005), The 150mm-long chip integrates columns, connections, and valves together with an ESI-emitter mounted at the end, guiding separated analyte ions into a mass spectrometer for characterization. It has a 160-nL enrichment column for analyte enrichment and desalting. The HPLC column is composed of a 75-µm x 150-mm analytical column packed with ZORBAX C-18 stationary phase resins for highly complicated tryptic digest mixtures.

The analysis was performed with a 0.4 µL/min flow rate and a loading rate of 4 µL/min. The gradient began with 6%B (90% ACN, 0.1 % TFA) which was held for 5 min, then a 30 minute gradient was run from 8% B to 45% B. The column was then washed at

high organic content (80% B) for 2 min, before equilibrated back to initial starting conditions.

Hemolymph proteins were characterized via MS using a Thermo LTQ-FT Ulltra (Waltham, MA). A survey full scan (m/z = 400–2000) at 50,000 resolution was acquired in which the five most intense ions were processed with MS/MS at an isolation width of 2.0 Da, a normalized collision energy of 35 eV, activation Q 0.25 and activation time of 30 ms. During MS/MS process, the collision gas was helium, the tube lens voltage was 150 V, and the capillary voltage was 50 V. The mass spectrometer was operated in positive ion mode with the trap set to data dependent MS/MS acquisition mode (Schilling et al., 2010).

Protein identification was performed via *Mascot* (www.matrixscience.com) sequence searching against NCBInr fruit fly taxonomy database with appropriate parameters setup as follows: trypsin was used as the enzyme; up to two missed cleavages were allowed; carbamidomethylation of cysteine, oxidation of methionine, and deamidation were selected as variable modifications; Peptide tolerance was ± 10 ppm; the MS/MS tolerance was set at ± 0.6 Da.

2.3. Results and Discussion

2.3.1. BSA calibration

A Bradford assay kit was used to generate a calibration curve for protein content determination in hemolymph, following the manufacturer's instruction. The reagent, Coomassie blue G-250 itself, has an absorption maximum at 465 nm.

While the protein-dye complex shifts the absorption maximum to 595 nm. As demonstrated in Figure 8, the concentration of protein in fly hemolymph was subsequently calculated from the calibration curve. With 1000-fold dilution, the volume of sample is still less than 100 µL, which does not fully meet the UV light height for absorbance reading with a regular 1 mL, 1 cm path length cuvette. Thus, a 400-µL quartz cuvette cell with a 1 mm path length was chosen for this study. The concentration of protein in the hemolymph sample was calculated as 0.8 ± 0.3 µg/100 nL. This level is appropriately below the maximal protein loading amount of 10 µg for the Agilent nano-HPLC chip system.

As calculated from the concentration calibration curve in Figure 8, the total concentration of individual *Drosophila* hemolymph protein is around 7.98g/L, about 1/10 compared to reports of 70g/L total protein for human blood (Siegel et al., 1998), Clinically collected blood is a more complicated blood system including blood platelets, leukocytes, erythrocytes, and serum proteins, while there are only four types of hemocytes (Schlenke et al., 2007) in the fly hemolymph. The fly hemolymph may represent a simplified biological matrix solution that can be studied both with respect to small molecule composition (Piyankarage et al., 2008) and, now, protein content.

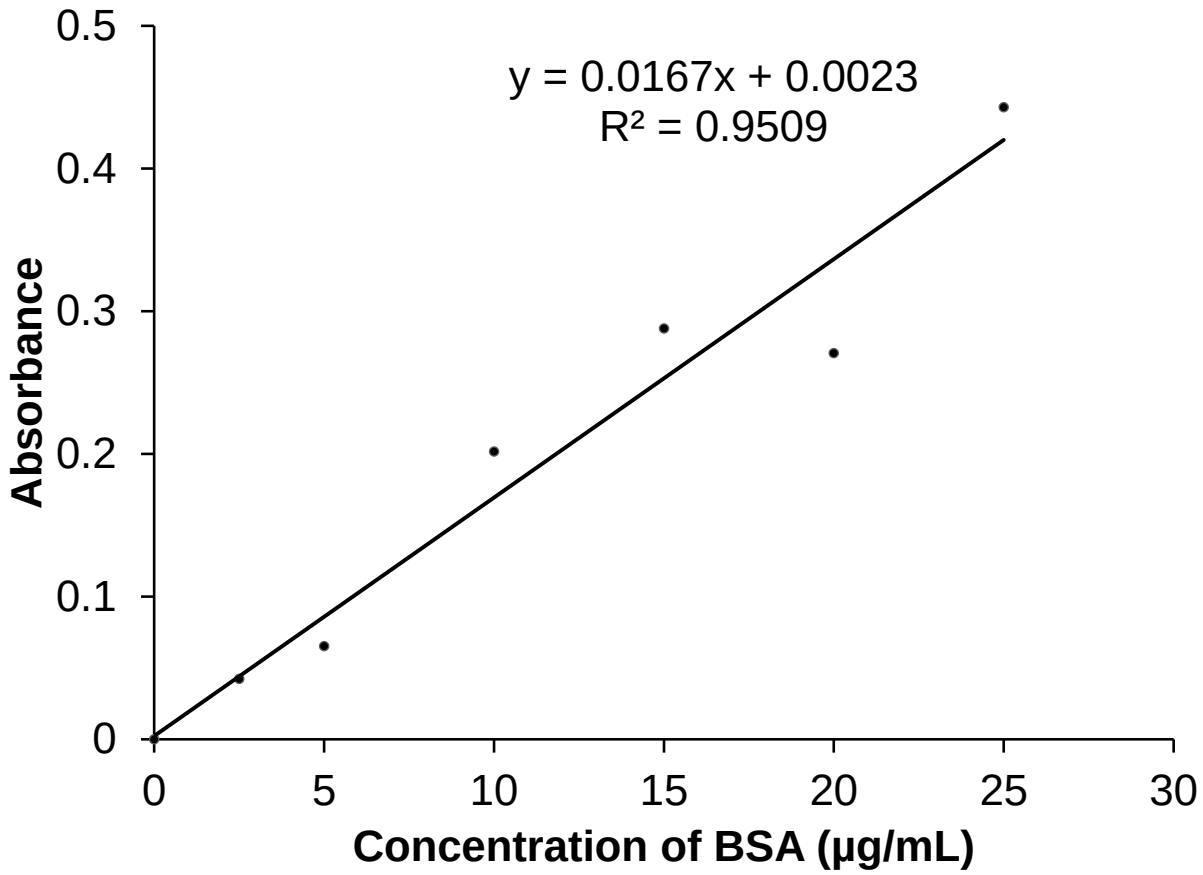


Figure 8 Protein concentration calibration curve of BSA using Bradford assay proteins kit at 595 nm. *D. melanogaster* hemolymph protein concentration was determined as 0.8 ± 0.3 µg/100 nL according to this calibration curve.

To verify that the enzyme digestion, reduction and amidation methods were sufficient with the hemolymph collected from a fly, 1 µM standard BSA was tested and confirmed by MALDI-TOF-MS. As shown in Figure 9, a list of peptide masses labeled with asterisks were identified belonging to BSA amino acids sequence. Given that analyte ions are commonly singly charged in MALDI, the mass difference between two neighbor analyte ions corresponds to single amino acid mass. The MS spectrum allows a simple and fast means for amino acid sequence interpretation. The major analyte ions listed in Figure 9 were found to match with the National Center for Biotechnology Information Non-redundant dataset (NCBInr) for BSA. These results verify the feasibility and efficiency of the in-solution digestion protocol with a 50 nL initial sample volume. This BSA digestion method was further confirmed by LTQ-FT ultra with a high protein score (data not shown), as a system check prior to hemolymph analysis.

2.3.2. Fruit fly hemolymph separation in HPLC

Digested hemolymph samples were separated on the nano-HPLC chip system. Using a step gradient method, analytes were efficiently separated and eluted out of the analysis column within 15 min, as illustrated in a representative TIC in Figure 10. As noted from the digestion protocol described in experimental section 2.3, the initial nL volume of hemolymph samples were diluted to less than 10 µL total volume, after addition of all digestion reagents, which meets the injection volume restriction (10 µL maximum) of the nano-HPLC chip for analysis in addition to the total protein loading limitation as described in Section 2.3.1.

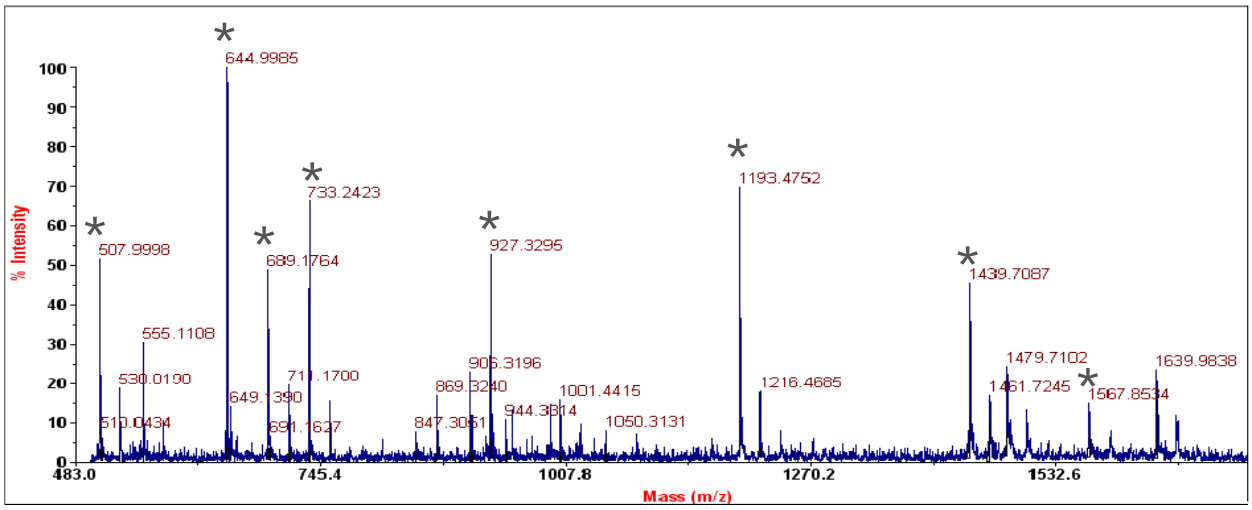


Figure 9 MALDI-TOF-MS spectrum of digested BSA peptides as a verification for standard protein digestion protocol for micro-liter volume of sample. Peaks tagged with asterisks in this MALDI mass spectra represent poly-peptide fragments generated by trypsin digestion, and detected with MALDI-TOF-MS. The MS spectrum demonstrates the feasibility and efficiency of the in-solution digestion protocol at microliter scale and provided guidance for the small volume hemolymph sample digestion.

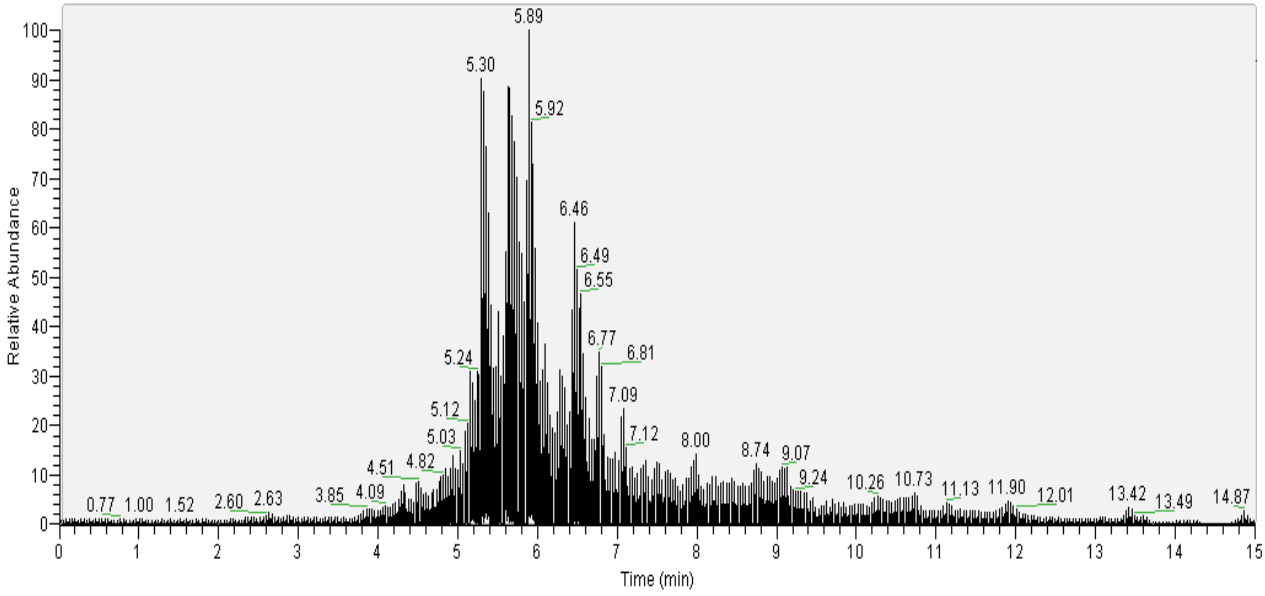


Figure 10 Total ion current chromatogram (TIC) of a single fruit fly hemolymph sample during reverse phase separation on the chip system. Analytes were eluted off column according to step-gradient method within 15 min. A major distribution of analyte peaks were detected between 4 and 10 minutes. Eluents were then characterized by mass spectrometry for protein identification.

2.3.3. Mass spectrometry result of protein identification

A total of 19 proteins were found in three trials with selection parameters (Decoy <10, False discovery rate <5%, Significance threshold $p<0.05$), with tandem MS analysis, as shown in Figure 11. As shown in Table 1, a group of well-known proteins were found in hemolymph sample with less than 100 nL original hemolymph fluid. These proteins have been identified in *Drosophila* hemolymph previously in literature. By confirming the existence of these proteins in fly hemolymph, this low volume analysis method has proved its efficiency and feasibility for proteomic investigations with exceedingly low volumes of biological sample. Furthermore, there are six novel protein hits found which have never been reported at the translation level. This finding serves as a proof-of-concept, indicating the capability of this analysis method in advancing the exploration of proteomics in fruit fly model system.

Based on their functions and in-vivo activities as identified through the *Flybase* and *Uniprot* library, these well-known proteins were divided into five major categories: Storage and transportation proteins, enzymes/proteases, reproduction proteins, signaling proteins and unknown.

2.3.3.1 Storage and transportation proteins

Hemolymph, similar to blood, serves as the only open circulating body fluid inside *Drosophila* and regulates chemical transportation and storage activities for cell-cell signaling and other physiological pathways. Therefore, it is unsurprising to find a large

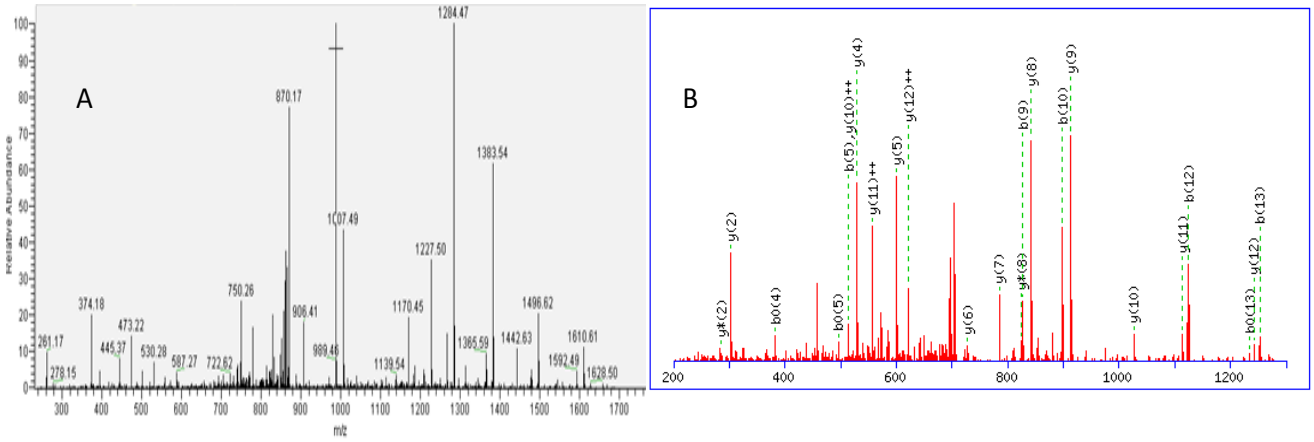


Figure 11 A) MS spectrum corresponds to the TIC chromatogram at 6.58 min in Figure 10. At each point of retention time various eluted analyte ions were detected and differentiated by mass spectrometry according to their mass-to-charge ratios ranging from 200 to 1800 m/z. Data dependent analysis (DDA) method was utilized to select five most abundant ions for tandem MS analysis. B) The MS/MS spectrum of tryptic peptide SFSSSPGDGNVVISR, which is a main fragment of protein CG5791 in *Drosophila melanogaster* hemolymph. De-novo sequencing method was used for proteomic sequencing, by assigning b and y ions as illustrated.

Protein hits	**Protein Score**	**Mass (m/z)**	**Queries matched**	**Sequence coverage (%)**
Retinoid-fatty acid-binding glycoprotein	2289	372414	82	13.3
Yolk protein 1	1415	48682	64	31.7
Ferritin 1 heavy chain homolog	1080	20483	16	28.9
Yolk protein 3	904	46073	19	19.3
Neuropeptide-like precursor 2	539	9406	21	48.8
Yolk protein 2	391	49630	21	17.9
Ferritin 2 light chain homolog	285	25255	8	23.3
RH69634p	272	14156	4	13
Calcium-binding protein	257	21181	10	21.1
Glyceraldehyde-3-phosphate dehydrogenase	225	35294	4	7.2
CG8193	183	79235	7	4.1
RH70154p	150	11484	5	17
Pro-phenol oxidase A1	147	79042	8	12.2
CG31150	109	167737	8	2
RH08557p	64	24095	4	9
CG6409	55	40299	2	3.9
CG5791	47	10202	1	15.3
IP06413p	36	42245	6	3.2
Pif1A isoform C[1]	36	187435	6	2
Pif1A isoform D[1]	36	70618	6	6
Glyceraldehyde-3-phosphate dehydrogenase1, isoform A	34	35328	1	7

Table 1 MS report of protein hits founds in fruit fly hemolymph[1]

[1]Pif1A isoform C and Pif1A isoform D share the same sets of peptides with IP06413p.

portion of proteins related to storage and transportation in hemolymph, as illustrated in Figure 12. Lipids are one of the most vital components for stabilizing and maintaining cell membrane structure and serve as storage for energy. Retinoid-fatty acid-binding glycoprotein, found in this report, is identified as a lipophorin and related to clotting process. This protein has been reported to act as an extracellular transport protein. It primarily transports lipid and lipid-linked morphogens (Tweedie et al., 2009). In addition, another protein hit CG31150 is also reported involved in lipid transportation activity (Tweedie et al., 2009).

Also identified are both Ferritin 1 heavy chain homolog and Ferritin 2 light chain homolog. These are the protein products of the previously discovered *Drosophila* hemoglobin genes (*glob1* and *glob2*) (Burmester et al., 2006) which are homologous to the vertebrate Ferritin heavy chain and light chain. This pair of subunits (25 and 26 kDa) in *Drosophila* has been previously reported (Dunkov and Georgieva, 1999). These ferritin homologs are believed to be able to store significant amount of iron ions.

Calcium ions are one of the most common signal transduction elements. As a metal element, calcium ion cannot be metabolized like other molecules and excess amounts can lead to cell death (Clapham, 1995). Therefore, the intracellular and extracellular equilibrium of calcium ion is strictly regulated via numerous calcium-binding proteins. The calcium binding protein found in our study is believed to belong to a group of calcium-binding proteins, which could either be activated directly by calcium ion, or act as calcium-dependent regulators for other enzymatic events (Kelly, 1990).

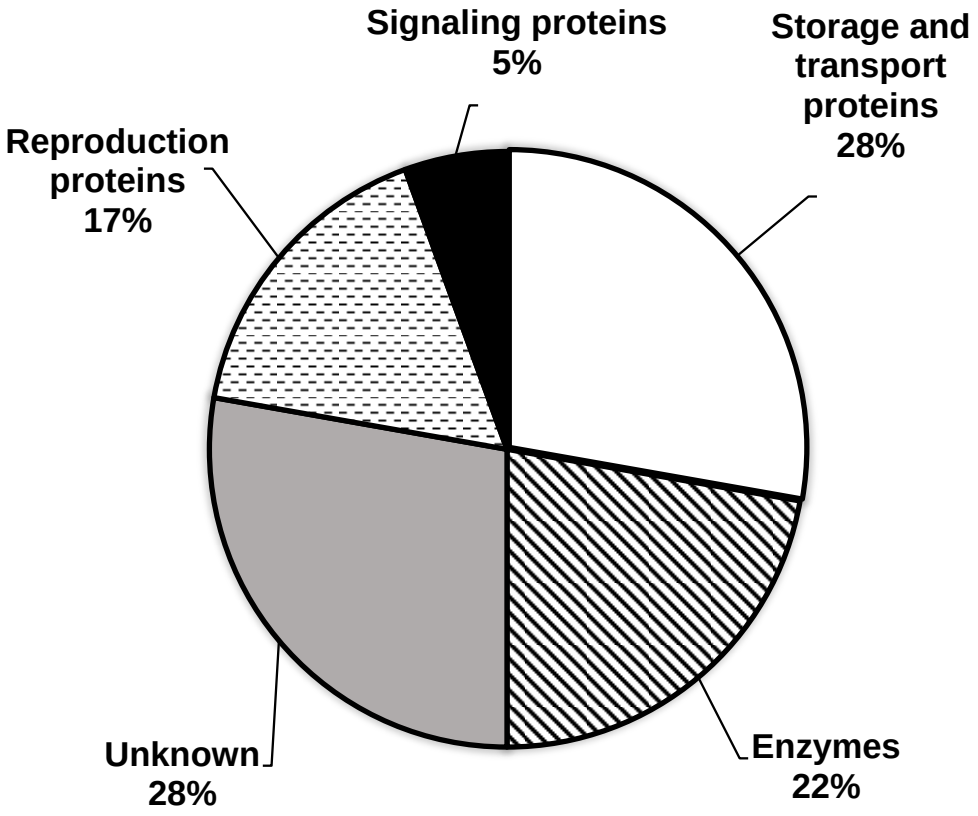


Figure 12 A pie diagram demonstrating the distribution of proteins in *D. melanogaster* hemolymph based on the biological function. Identified hemolymph proteins are categorized into five functional groups based upon their biological activities and functions. Among these proteins, a significant number are expected to play storage and transportation roles in line with the observation that hemolymph is a primary medium for chemical signaling between cells and tissues. Notably, there are an equal number of novel proteins identified in this study.

2.3.3.2 Enzymes/proteases

Glyceraldehyde-3-phosphate dehydrogenase and pro-phenol oxidase A1 are enzymes which were found in *Drosophila* hemolymph. Notably, glyceraldehyde-3-phosphate dehydrogenase is an enzyme that catalyzes the sixth step of glycolysis and thus serves to break down glucose for energy and carbohydrates (Tweedie et al., 2009). Pro-phenol oxidase A1 acts as an oxidoreductase. This enzyme is active in monophenol monooxygenase activity. CG8193 is another protein hit which was indicated to play an important role in monophenol monooxygenase activity as well. Its participation in transport and metabolic biological process has been previously observed (Tweedie et al., 2009). RH69634p is coupled to transmembrane movement of substances and involved in ATPase activity (Tweedie et al., 2009).

2.3.3.3 Reproduction proteins

It was proposed that yolk proteins in *Drosophila* are synthesized in the fat body, released into the hemolymph and then taken up by the oocytes (Barnett et al., 1980; Richard et al., 2001). A series of yolk proteins—yolk proteins 1, 2, and 3—found in this work agrees this hypothesis. Because only female fruit flies were investigated in this study, due to their larger size and relatively larger volume of hemolymph available compared to male flies, it is reasonable that these yolk proteins were observed.

2.3.3.4 Signaling proteins

Neuropeptide-like precursor 2 is a protein, which was reported presence in larval *Drosophila* central nervous system and antenna previously (Anholt and Williams, 2010). The presence of this protein in our study indicates that the neuropeptide-like precursor 2 might be synthesized beyond early developmental stages. While the function of this protein is not yet fully known (Tweedie et al., 2009), there is a hypothesis that the distribution of neuropeptide-like precursor 2 might potentially be related to *Drosophila* innate immunity, which would rationalize its presence in adult hemolymph (Verleyen et al., 2006). The discovery of the protein in our study merits further investigation of this hypothesis.

2.3.4. Identification of novel proteins in fruit fly hemolymph

Among the observed 19 proteins, 6 proteins have only been reported at transcript level, according to searches with *UniProt* and *Flybase*. This report identifies these proteins at translation level in the fly hemolymph. This finding may help to improve and update the database, as well as facilitate the study of these less-known proteins and lead to a more complete understanding of the physiology of the fly model system.

It is somewhat unsurprising that several of these proteins identified for the first time do not have a known function. RH08557p is predicted to involve in oxidation/reduction process (Tweedie et al., 2009). CG6409 is predicted to be involved in GPI anchor biosynthesis, inferred from sequence or structural similarity with the N-

acetylglucosaminyl-phosphatidylinositol biosynthetic protein (or GlcNAc-PI synthesis protein). GlcNAc-PI synthesis protein is necessary for the synthesis of N-acetylglucosaminyl-phosphatidylinositol, the very early intermediate in GPI-anchor biosynthesis (Tweedie et al., 2009). CG5791's biological information is entirely unknown (Tweedie et al., 2009).

Another protein identified here for the first time is IP06413p. IP06413p is suggested to be involved in transcription processes. However, Pif1A isoform C and Pif1A isoform D are two proteins reported to share the same set of peptides with IP06413p with the same MS score in the results reported here. Pif1A isoform C is a protein which has not been identified at either protein, transcript, or even homology levels. Pif1A isoform C is predicted to be involved in sequence-specific DNA binding transcription factor activity (Tweedie et al., 2009). Pif1A isoform D is believed to be involved in protein binding, sequence-specific DNA binding, and sequence-specific DNA binding transcription factor activity. All three of these proteins IP06413p, Pif1A isoform C and Pif1A isoform, can be potential origins of the detected set of peptides in this study. To identify the exact protein which the set of peptides belong to, a further in-depth investigation is required.

Glyceraldehyde-3-phosphate dehydrogenase 1 isoform A, has only been identified at translation level previously as well. It is an isoform of glyceraldehyde-3-phosphate dehydrogenase, which is also found in the result. This isoform has similar function as glyceraldehyde-3-phosphate dehydrogenase. It acts as oxidoreductase during glycolysis.

2.4. Conclusion

In the present project, I showed that the direct sampling method previously developed in our lab previously provides a convenient and fast approach to study individual fly hemolymph proteomics. The microscale digestion protocol has also been verified for a nL volume biosample. Hemolymph proteomic information was obtained from individual fruit flies with a simplified treatment avoiding the time consuming homogenization or sample pooling and excessive solvent dilution. A total of 19 protein hits were found in this study in biological triplicate. Among those proteins, six novel protein hits were found, which were previously reported at transcript level. The characterization of well-known proteins proved the feasibility and efficiency of proteomic analysis of single fruit fly hemolymph in limited volume. The ability to observe the particular and less-known proteins indicates this analysis method by using nano-HPLC chip system promotes a more in-depth proteomic information mining from *Drosophila* individuals, avoiding averaging effects from hemolymph pooled from a large population. This nano-scale analysis method could now facilitate proteomic hemolymph study and lead to a more complete understanding of the physiology of the fly model system.

3. PREFRACTIONATION METHODS FOR INDIVIDUAL FRUIT FLY HEMOLYMPH PROTEOMIC ANALYSIS

[Submitted to the journal of Analytical Chemistry]

3.1. Introduction

Drosophila melanogaster, serves as an important animal model for studying molecular mechanisms underlying intracellular processes, cell-cell interaction, or the pathology of clinical diseases. In fact, three-quarters of known human disease genes have been found acknowledgeable in the genome of fruit flies, and two thirds of fly protein sequences have mammalian homologs (Rubin et al., 2000; Adams et al., 2000). The great similarity between fly and human genomes, the low-cost maintenance, the short generations and the easy gene manipulability continues to lead to many studies with the fly including into the central nervous system, metabolism pathways, and neurodegenerative diseases (De Gregorio et al., 2001; Driscoll and Gerstbrein, 2003; Shulman et al., 2003). The fly has an open circulatory system, in which hemolymph, similar to blood, circulates through the whole body, transporting and exchanging nutrients, metabolites, and antibodies between organs and tissues. There is also ongoing active chemical signaling between internal tissues and organ cells by proteins, peptides, and hormones in hemolymph. Therefore, the proteomic analysis of fruit fly hemolymph, plays an important role in understanding actual protein content in distinct physiological events (Anholt and Williams, 2010; Vierstraete et al., 2004; Karlsson et al., 2004; Takemori and Yamamoto, 2009). A novel single fruit fly hemolymph collection method previously (Piyankarage, Featherstone, and Shippy, 2012; Piyankarage et al., 2008) was adapted here as a sampling tool.

Nevertheless, the overwhelming complexity of the proteome and exceedingly low quantities of hemolymph has presented an extreme challenge to current analysis technologies. 2-DE is the most widely adapted prefractionation method for simplification and reduction of the complexity of proteomic mixtures such as hemolymph (Righetti et al., 2005). Vierstraete et al. reported the first 2-D database for larval hemolymph proteins, using 2-DE with MALDI-TOF MS (Vierstraete et al., 2003). Kalsson et al. investigated larval hemolymph protein related to clotting, using the same set up of instrument as Vierstaete et al. (Karlsson et al., 2004). However, 2-DE sacrifices the separation capacity of highly hydrophobic proteins and proteins with extensively high or low isoelectric points or molecular weights and resolution due to serious spot overlapping. In addition, only 30% proteins of a cell lysate are able to be visualized using the classical silver-stained 2-DE gel approach, and only a portion of these can be harvested sufficiently for mass spectrometric analysis (Badock et al., 2001). As a powerful analysis tool, mass spectrometry has been intensively employed in protein identification, providing an alternative mass separation of protein mixtures by mass-to-charge ratio. Nevertheless, the signal suppression caused by certain well ionized species or highly abundant proteins has increased the concerns of the efficiency and bias of the method, with many proteomic samples there is also a real concern regarding the effective analysis of the wide dynamic range concentration of proteins from 70% to 10^{-12}% or less (Martosella et al., 2005). For instance, albumin takes up to 60% of the total proteins in serum (Righetti et al., 2005) and with 2-DE this amount can overlapped or covered many other protein spots. There may also be serious signal suppression to other underrepresented species with MS, because an overabundance of albumin fragments “blinds” MS. These two analytical challenges

had led to prefractionation technique development, in order to simplify complex protein mixtures and more accurate and diverse spectra.

To overcome the defect caused by high-abundance protein species and to profile relatively less abundant proteins, numerous strategies based on affinity chromatography have been developed. One of the most notable and widely applied is immunodepletion to selectively bind and remove known high-abundance proteins. Lindmark et al. has removed serum antibodies of immunoglobulin G (IgG) class using immobilized Protein A (Lindmark, Biriell, and Sjoquist, 1981). However, this immunodepletion approach could be problematic for the discovery of other species of interest which are functionally related to the depleted protein which also get removed. Washburn et al. has pioneered and popularized the multidimensional LC fractionation technique, which is known as MudPIT (Multidimentional Protein Identification Technology) (Washburn, Wolters, and Yates, 2001; Fonslow et al., 2011). This method involves two orthogonal separation methods of an ion-exchange column and a reversed-phase column coupled with a mass spectrometer, in order to improve chromatographic peak capacity prior to MS identification. Nevertheless, this 2D or even 3D separation increases experimental time exponentially. Experiments requiring 100 hours of instrument time for each sample, are a major drawback with respect to cost and high throughput screening proteomic analysis. Furthermore, the significant variance of gradient eluent buffer from different separation column systems can alter the characteristics of the original sample, in addition to sample dilution for relatively small volume sample.

In this chapter I introduce a customized low pressure and rapid prefractionation reversed-phase and ion exchange chromatographic columns for low volume protein

sample prior to nano-HPLC chip and Orbitrap MS/MS analysis. These methods offer a rapid sample prefractionation, simplify the complex proteomic sample from adult fruit fly hemolymph, and eventually contribute to better protein identification and low abundance critical protein detection.

3.2. Experimental section

3.2.1. Chemicals

Deionized water was obtained from a US Filter Purelab Plus purification system (Lowell, MA). Trypsin, Tris hydrochloride, dithiothreitol and iodoacetimide were all purchased from Sigma-Aldrich (St. Louis, MO). Glacial acetic acid was purchased from Fisher Scientific (Itasca, IL). Trifluoroacetic acid was from Sigma-Aldrich (St. Louis, MO). All chemicals reagents were of analytical grade.

3.2.2. Hemolymph sample preparation

Hemolymph was harvested from the Oregon-R strain of *Drosophila melanogaster* maintained in the Laboratory of Integrative Neuroscience, Department of Biological Sciences in UIC. Only female flies were chosen for sampling, to eliminate the gender variation and to access larger hemolymph volumes. The fly hemolymph samples were collected using a direct abdomen sampling technique, as described in Chapter 2 (Piyankarage, Featherstone, and Shippy, 2012; Piyankarage et al., 2008). In brief, the

hemolymph was released from a small incision on the fly abdomen and immediately collected by a 4 cm-long fused-silica capillary probe. The around 50 nL collected hemolymph samples were subsequently reduced, alkylated, and then digested at 37°C overnight, following a previously established micro-scale digestion protocol, as shown in Chapter 2. Tryptic digested hemolymph samples were stored at -20 °C until analysis.

3.2.3. RPLC fractionation

Spherical silica C18 beads were aspirated into a 5-cm long tygon tubing by vacuum created by syringe pulling on the other end of the tygon tube. With C18 beads packed tightly in the 5-cm tygon tube, both ends of the tygon tube were trimmed to make a 2-cm long evenly packed C-18 column, and a 0.5-cm long (360/50-µm o.d./i.d.) fused silica capillary was inserted at the end of the tygon tubing column as a weir and an exit for elution collection. Prior to hemolymph separation, tryptic digested BSA was chosen as a standard to test the separation efficiency of this customized short column. Hemolymph sample was loaded onto the column by a 1-mL insulin injection syringe using a PHD 2000 syringe pump (Harvard Apparatus, Holliston, MA) at flow rate 0.5 µL/min, and then eluted according to step-gradient elution at the same nanoscale flow rate. Three fractions of eluate was collected at each step of 0%, 10%, 20%, 30% and 70% acetonitrile in distilled water with 0.1% TFA.

3.2.4. Strong cation exchange chromatography fractionation

Dowex 50WX4 200-400, which is a fine meshed resin in hydrogen form, was selected for IEX fractionation of tryptic hemolymph proteomic sample. The column packing follows the same procedure as described in Section 3.2.3. Packed columns were thoroughly rinsed with deionized water, followed up with column equilibration using 20 mM pH 2.58 phosphate buffer. Hemolymph sample was loaded onto the column, and then eluted by a step gradient of increasing pH and concentration of sodium chloride in 20 mM phosphate buffer. The steps in the eluent gradient are 0 M NaCl at pH 2.58, 70 mM NaCl at pH 5.23, 140 mM NaCl at pH 6.34, 200 mM NaCl at pH 7.58, 1 M NaCl at pH 9.20. Eluate fractions were collected as described in RPLC fractionation method.

3.2.5. Fluorescence detection

To test and optimize the prefractionation conditions, a fluorimeter, was used to measure the fluorescence intensity of each eluate fraction in a capillary cell. The fluorescence detector uses a 365 nm laser for excitation and emission centered at 490 nm. The pH of IEX eluate fractions were adjusted with 0.1 M NaOH to pH of 7 for devivatization. To determine approximate pH in these small volumes, a small volume of NaOH was proportionally measured by 20.6-mm/250-μm o.d. /i.d. tygon tubing and delivered into each IEX eluate fraction to adjust the pH. Each eluate fraction was then derivatized with fluorescamine (1.5 mg/mL in acetone) by tagging the primary amine group for 6 min prior to detection. A 10 cm (360/50-μm o.d. /i.d.) fused silica capillary was used as the detection cell. Each eluate was injected into the capillary cell for triplicate measurements of fluorescence intensity. The order of deionized water, HCl, deionized

water, NaOH and deionized water was followed to rinse the capillary cell before each eluate fraction injection.

3.2.6. MS Detection

With developed prefractionation methods, hemolymph fractions were further separated by an Agilent large capacity nano-HPLC chip system (Santa Clara, CA) prior to MS identification. This nano-HPLC chip system is composed of a 75-µm x 150-mm analytical column packed with ZORBAX C-18 stationary phase resins for highly complicated tryptic digest mixtures, as described in Chapter 2.

As a comparison to protein identification efficiency by the above nano-HPLC chip system, a nanoLC column (Zorbax 300SB C18; 3.5 µm, 150 mm X 75 µm) (Agilent, Santa Clara, CA) with a peptide trap (C18 PepMap 100, 5 µm, 100 Å, 300 µm ID X 5 µm) (Thermo Scientific, Waltham, MA) was used as an alternative separation method after the RPLC fractionation. Both nano-HPLC chip system and nanoLC column were installed in a Dionex Ultimate 3000 HPLC instrument. (Thermo Scientific, Waltham, MA) The separation analysis on the nano-HPLC chip system was performed with a 0.4 µL/min flow rate, with a loading rate of 4 µL/min. The gradient began with 6% B, in which B solvent is composed of 5% H_2O, 95% CAN and 0.1 % Formic acid, held for 5 min, then the 30 min gradient ramping from 8% B to 45% B, the column was then washed at high organic, 80% B, for 2 min, before equilibrated back to initial starting conditions. The analysis on the nanoLC colum was set as 60 min gradient ramping from 8% B to 45% B.

The setup of the MS follows the description in Chapter 2. After separated on the nano-HPLC chip column, each fraction was analyzed by a Velos Pro Obitrap MS/MS (Thermo Scientific, Waltham, MA). The Agilent chip spray voltage was 1.9 kV and the capillary temperature was set at 200 °C. During MS/MS process, the collision gas is helium, the tube lens voltage is 150 V and the capillary voltage is 50 V. A survey full scan (m/z = 400-2000) at 30,000 resolution, was acquired in which the five most intense ions, if the charge state was determined to be greater than +1, MS/MS was triggered with minimum signal required (2500), isolation width of 2.0 Da, a normalized collision energy of 35 eV, activation Q 0.25 and activation time of 30ms. The dynamic exclusion list was restricted to 250 entries with duration of 60s, to minimize redundant spectral acquisitions. The mass spectrometer is operated in positive ion mode with the trap set to data dependent MS/MS acquisition mode. MS data as acquired and analyzed by Xcalibur software.

3.2.7. Data Analysis

Protein identification was performed via *Mascot* (www.matrixscience.com) sequence searching against NCBInr fruit fly taxonomy database with appropriate parameters setup as follows: Trypsin was used as the enzyme; up to two missed cleavages were allowed; carbamidomethylation of cysteine, oxidation of methionine, and deamidation were selected as variable modifications; Peptide tolerance was ± 10 ppm; the MS/MS tolerance was set at ± 0.6 Da. UniprotKB and Flybase were two libraries used for detailed bioinformation interpretation of proteins found in this study.

3.3. Results and Discussion

3.3.1. Protein fractionation with customized RPLC column

Due to the 50 nL of hemolymph sample available, a customized short RPLC column was fabricated, in order to achieve an effective fractionation of the complicated biological sample, prior to the MS characterization. BSA was used as a standard protein and prepared the same way as the hemolymph sample to demonstrate sample handling and fractionation by this customized RPLC column. Figure 13 (A) shows that standard tryptic BSA peptides were separated efficiently with this customize RPLC column. Notably, 20% ACN and 30% ACN eluted about 36% and 28% of the total amount of tryptic BSA peptides, respectively. Within these two gradient steps, a baseline separation was obtained as indicated by fractions 1, 7, 10 and 14. This result can be compared and matched with classical commercial C-18 columns for peptide separation. In addition, each fraction is 2 µL, almost equal to two 0.9 µL column volumes. The baseline separation suggests that there good partition of peptide analytes onto the stationary phase and selective elution by step gradient eluent.

With the proof-of-concept experiment conducted with BSA tryptic peptide mixture, this RPLC column was used for fractionation with fly hemolymph tryptic peptides. As shown in Figure 13 (B), the hemolymph peptide complex was fractionated using the same column set up and step gradients for elution. With respect to chromatographic behavior, Figure 13 (B) shows a significant portion of hemolymph peptides were eluted

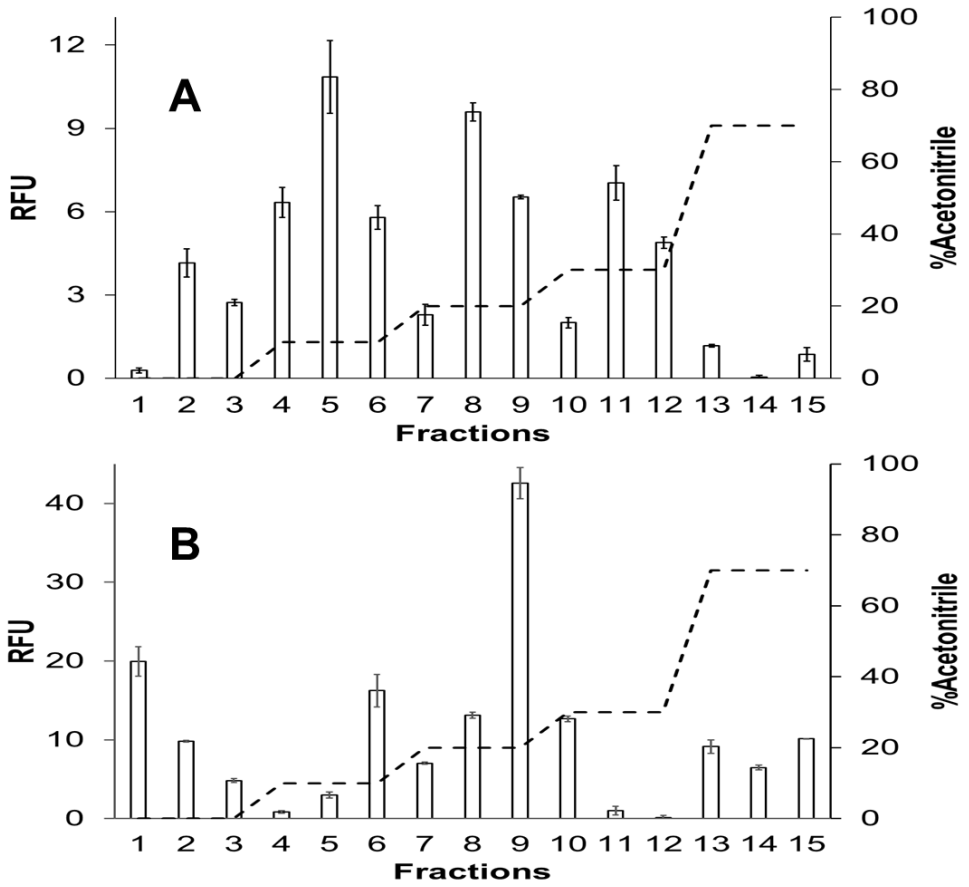


Figure 13 Fractionation chromatogram of tryptic peptides obtained from customized RPLC column.

(A) Chromatogram of standard tryptic BSA peptides fractionation using the customized RPLC column. (B) Chromatogram of tryptic peptides of fruit fly hemolymph protein sample separated by the customized RPLC column. Sample injection volume: 500 nL BSA (1 µM), 500 nL hemolymph. The flow rate was 0.5 µL/min. Step gradient: 0% ACN, 10% ACN, 20% ACN, 30% ACN, and 70% ACN in distilled water with 0.1% TFA (as shown by the dash line). During each step gradient condition, three 2-µL vials of eluate fractions were collected and detected by fluorometer at 365 nm excitation/490 nm emission.

and collected in fraction 9, which is the last fraction collected with 30% ACN mobile phase. The difference from the BSA peptide separation could be attributed to the variation of protein components in hemolymph versus BSA, a single protein.

3.3.2. Optimization of separation condition for customized IEX column

As one of the most frequently used chromatographic methods for protein and peptide separation, ion exchange chromatography has been proved to be a robust and simple prefractionation technique in proteomics, providing high resolving power and high analyte capacity for a wide range of applications. Protein behavior in ion exchange chromatography is ruled by the relationship between the pH of the mobile phase solvent, the pK of proteins and individual amino acid charges. Herein, I demonstrate a strong cation exchange column fabricated for fractionation of fruit fly hemolymph in as small volume as 50 nL. Again, BSA tryptic peptides were used as a standard to show the fractionation capability of this customized exceedingly short IEX column.

As illustrated in Figure 14 (A), the step gradients used for the separation of BSA tryptic peptides covered a wide range of ionic strengths. However, an expected baseline separation of the peptide complex was not achieved, instead, a significant fraction of BSA peptides were eluted and collected in fraction 7, which is the first fraction collected with 300 mM NaCl eluent. Furthermore, the acute increment of NaCl from 10 mM to 300 mM NaCl appears to lead to a majority of the elution. Therefore, a more evenly distributed step gradient arrangement was carried out as shown in Figure 14 (B). By adding 200 mM NaCl and 500 mM NaCl steps, fraction peaks have been better resolved into the whole series of fractions, as seen in Figure 14(B).

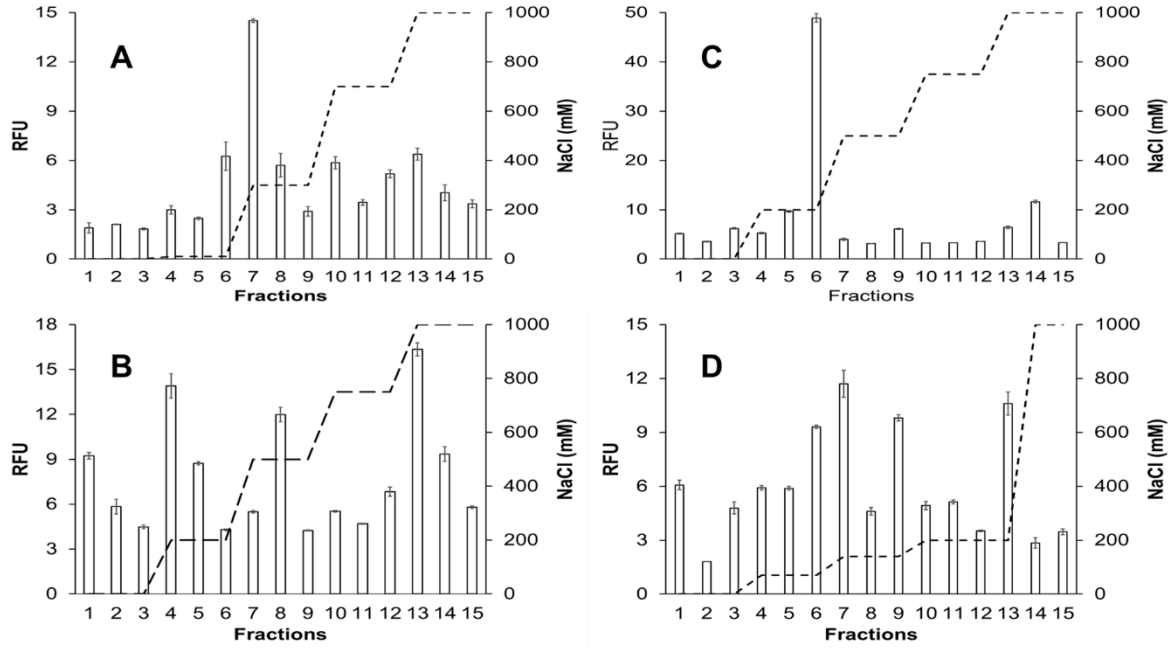


Figure 14 Optimization of separation condition for complex peptide mixture in IEX chromatography. The stepwise composition of eluent NaCl is noted as the dash line related to the values marked on the right side of the y-axis. (A) Fractionation of BSA tryptic peptides with step gradient set up as: 0 M-10 mM-300 mM-700 m-1 M NaCl. The acute increment between 10 mM and 300 mM NaCl caused the majority of analytes to elute in fraction 7, which indicates an inefficient fractionation according to their unique isoelectric points. (B) An even distribution of the concentration of NaCl was made by 0 M-200 mM-500 mM-750 mM-1 M NaCl. Analytes were better separated and peaks were resolved compared to (A). (C) Fractionation of fruit fly hemolymph tryptic peptide mixture using the condition of (B). The majority of analytes were eluted and collected in fraction 6. (D) An optimized condition with 0-70 mM-140 mM-200 mM-1 M NaCl was used. Hemolymph peptides were fractionated and collected over several separate fractions.

The improved step gradient elution was subsequently employed to fractionation of a fly hemolymph sample. From Figure 14(C), it is observed that the fractionation of hemolymph tryptic peptides was far from satisfying. Most of hemolymph tryptic peptides were eluted and collected in fraction 6, which is the last fraction collected with 200 mM NaCl. A further distribution of NaCl concentration steps below 200 mM were used to optimize the fractionation of hemolymph specifically. As illustrated in Figure 14 (D), by adding two extra steps of 70 and 140 mM NaCl, a reasonable fractionation of complex hemolymph tryptic peptides mix was achieved. As expected, the majority of peptides was eluted off the column within the 6 fractions collected under 70 mM NaCl and 140 mM NaCl steps. The optimization of step gradient provided the fractionation of complex fruit fly hemolymph sample, which hopefully will simplify the peptide composition in each fraction, and therefore enhance the MS characterization capability for less abundant protein identification in limited volume of hemolymph sample.

To be noted, both RPLC and IEX prefractionation experiments take less than 3 hours. However, a typical MudPIT can take as much as 100 hours for a sample. The time and cost efficiency enables the proposed RPLC and IEX prefractionation methods to be adapted to any biological proteomic sample complex with ease. Because these prefractionation methods are offline from the mass spectrometer, this should allow for increased efficiency of use of the time with the mass spectrometer compared to online fractionation arrangements.

3.3.3. MS results

Individual fly hemolymph samples were assayed for protein content following the RPLC and IEX prefractionation. A control experiment with the same set up was carried out without any prefractionation prior to nano-HPLC chip column separation and MS identification. The protein results of these three methods were compared and analyzed to determine the efficiency of the prefractionation. The proteins scores, sequence coverage and the relative abundance value emPAI provide factors to evaluate the prefractionation efficiency. Furthermore, the distribution of proteins found in three methods was charted in the Venn diagram, as shown in Figure 15 Venn diagram of proteins identified to compare between fractionation methods.The enhancement from prefractionation is verified by the identification of novel proteins in this study.

3.3.4. MS characterization of proteins in hemolymph by various fractionation methods

In order to assess the efficacy of RPLC and IEX prefractionation in enhancing protein identification, a comparison of 6 proteins found by all three methods were compared. Since individually analyzed fractions contain partial fragments from an intact protein, the MS identification could be underestimated due to insufficient fragments from the same protein passing the threshold protein score. Combining the MS results of all individually analyzed fractions therefore could significantly increase the protein identification of the combined LC-MS/MS approach (Martosella et al., 2005).

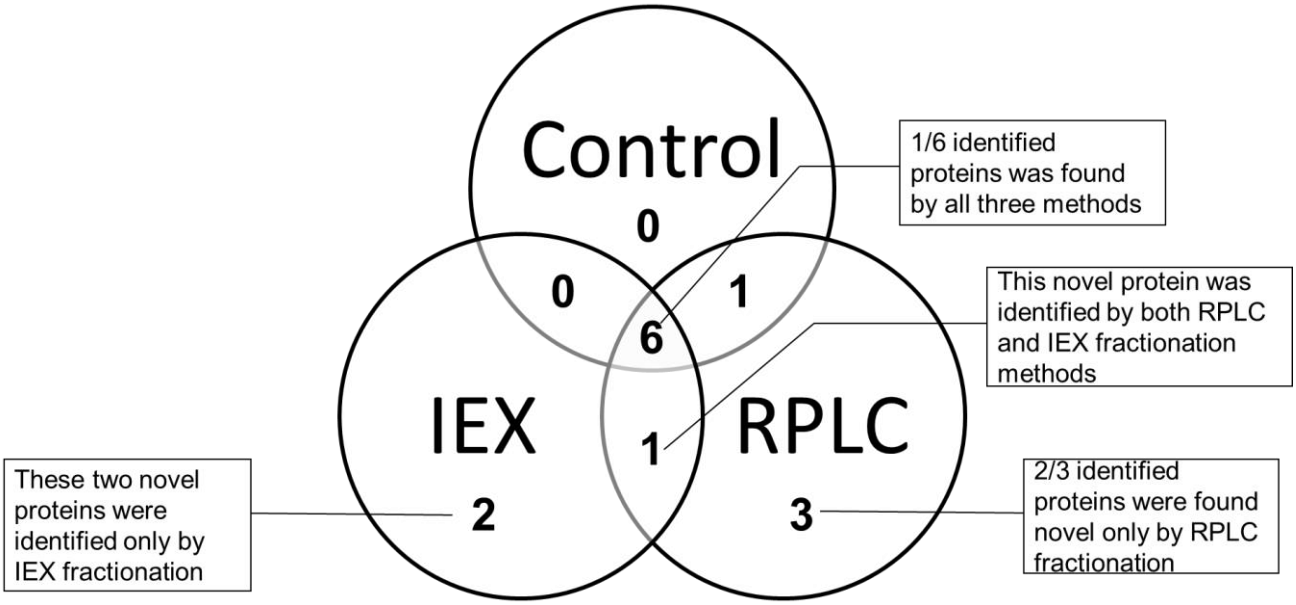


Figure 15 Venn diagram of proteins identified to compare between fractionation methods.

Table 2 illustrates proteins that are enriched by RPLC and IEX fractionation. The degree of enrichment can be indicated by an increase in the protein score, sequence coverage and relative abundance data associated to the exponentially modified protein abundance index value (emPAI). The emPAI represents approximate, label-free, relative quantitation of the proteins in a mixture based on protein coverage by the peptide matches in a database search result. Yolk proteins are dominant proteins in the hemolymph sample. This result is not surprising because these samples are collected exclusively from non-virgin adult female flies. While their presence is reasonable, there are other, less well known, less abundant proteins in hemolymph that are more interesting analytical targets.

RPLC fractionation succeeded in reducing the abundance of these three yolk proteins, as indicated by decreased emPAI. At the same time, there was an increased emPAI and sequence coverage of the other three, less well known proteins, retinoid and fatty acid-binding glycoprotein, neuropeptide-like protein and transferrin precursor, compared to non prefractionated control. Further, these results confirm the selective enrichment effect by RPLC fractionation method as IEX fractionation did not decrease the detection of highly abundant yolk proteins. Nevertheless, IEX improved the MS identification of neuropeptide-like precursor with a significant increase of emPAI value compared to control. Notably, the identification efficacy of retinoid-fatty acid binding glycoprotein, neuropeptide-like precursor and transferrin precursor were all significantly

Protein name	Protein[a] (Mr)	Control (unfractionated) Score[b]	emPAI value[c]	Sequence coverage	RPLC fractionation Score[b]	emPAI value[c]	Sequence coverage	IEX fractionation Score[b]	emPAI value[c]	Sequence coverage
Yolk protein 1	48682	777	1.09	27%	258	0.81	33%	584	1.81	28%
Yolk protein 2	49630	392	0.92	20%	170	0.63	31%	508	1.31	32%
Yolk protein 3	46073	388	0.48	14%	140	0.37	20%	464	0.73	17%
Retinoid- and fatty acid-binding glycoprotein	372414	293	0.07	2%	396	0.12	5%	379	0.07	3%
Neuropeptide-like precursor 2	9406	287	1.92	29%	216	3.17	34%	363	3.17	29%
Transferrin precursor	71765	75	0.11	4%	79	0.16	6%	65	0.11	6%

Table 2. Protein identification by all three methods, and enrichment by RPLC and IEX fractionation methods. [a]protein norminal mass (Mr). [b]Protein score from *Mascot*. [c]Exponentially Modified Protein Abundance Index represents approximate, label-free, relative quantitation of the proteins in a mixture based on protein coverage by the peptide matches in a database search result.

improved as shown by the protein scores, emPAI and sequence coverage with either prefractionation method, compared to control method.

Shown in Figure 15 is the Venn diagram to compare protein identification between the RPLC, IEX fractionation methods and the control. In total, 7 proteins were identified in the control, while there are 11 proteins identified following RPLC, and 9 proteins identified after IEX prefractionation. This result suggests the prefractionation plays an important role in enhancing protein identification from these chemical complex but, volume-limited hemolymph samples. From Figure 15, there are 6 proteins identified by all three methods. These proteins are yolk protein 1, yolk protein 2, yolk protein 3, neuropeptide-like precursor 2, retinoid and fatty acid-binding glycoprotein, and transferrin precursor. A previous report of a population of fly larvae found mostly proteases in the larval hemolymph and did not report those six proteins (Vierstraete et al., 2003). That could be explained by the difference between adult and larval flies. For example, for larvae reproductive activities are not paramount, even for the female larvae. Yolk proteins are not expected to present in larval stage. Among the commonly found proteins reported here, transferrin precursor is a novel protein that is secreted extracellularly and is associated with cellular iron ion homeostasis, iron ion transport and olfactory behavior (Consortium, 2014). The preponderance of these secreted proteins suggests that they are dominant in the hemolymph sample. This result is interesting because any hemocyte tissue present (10%) is not leading to a major protein hit observed by all techniques tested.

Other than the common proteins shared with control and IEX methods, RPLC prefractionation led to the characterization of unique proteins, such as Bellwether, Ferritin 2 light chain homolog, and RH70154P. Notably, there is only evidence at the transcript

level RH70154P and prediction of for Ferritin 2 light chain homolog, while our results demonstrate the expected protein products. Bellwether is also known as ATP synthase subunit alpha and belongs to the ATPase alpha/beta chains family. As a mitochondrial membrane ATP synthase, Bellwether is involved in the production of ATP from ADP. Because ATP synthase occurs inside the mitochondria within cells, it appears that the observation of Bellwether in hemolymph is originally from the hemocyte component of our sample. Further study is needed to confirm this idea. The existence of Ferritin 2 light chain homolog is currently at predicted level, which means the expression of this Ferritin 2 light chain homolog was only predicted from its genome. In this study the identification of peptide fragments of this protein was able to prove the existence of this protein, which confirms this gene-predicted product. It is interesting that it is found in adult hemolymph and is suggestive of the physiological role of aerobic respiration for the adult. Ferritin 2 light chain homolog is believed to bind to ferric iron in both extracellular region and intracellular ferritin complex. This protein is also inferred associated with hemolymph coagulation in the fly (Tweedie et al., 2009). RH70154P's molecular function and biological process is not known.

There are two novel proteins that were identified for the first time by IEX fractionation. CG6409 and Imaginal disc growth factor 4. CG6409 is a protein with evidence at the mRNA transcript level and is predicted to be involved in GPI anchor biosynthesis, inferred from sequence or structural similarity with the N-acetylglucosaminyl-phosphatidylinositol biosynthetic protein (or GlcNAc-PI synthesis protein). GlcNAc-PI synthesis protein serves as glycosyltransferase and is necessary for the synthesis of N-acetylglucosaminyl-phosphatidylinositol, the very early intermediate in

GPI-anchor biosynthesis (Tweedie et al., 2009). Imaginal disc growth factor 4 was also found as a protein here for the first time. The protein was identified to be secreted in hemolymph. It is probably transported to target tissues via hemolymph (Kawamura et al., 1999). Imaginal disc growth factor 4 was reported to be primarily expressed in yolk cells and fat body. While it is expressed throughout development stages, but with a much stronger expression during larval stages (Kawamura et al., 1999). Previous literature reported the identification of imaginal disc growth factor 2 in larvae. The imaginal disc growth factor 2 protein belongs to the same protein family of imaginal disc growth factor 4. The finding of imaginal disc growth factor 4 might reveal the different protein expression of the same family of gene. Further studies are needed for confirmation. Imaginal disc growth factor 4 cooperates with insulin-like peptides to stimulate the proliferation, polarization and motility of imaginal disc cells. This protein may act by stabilizing the binding of insulin-like peptides to its receptor through a simultaneous interaction with both molecules to form a multiprotein signaling complex (Kawamura et al., 1999). Further investigation is needed for confirmation of this hypothesis.

3.3.1. Effect of separation on unfractionated samples prior to MS

With the success of prefractionation methods to reduce the complexity of protein samples, a further experiment was performed to determine whether the LC separation step prior to MS identification could also provide an enhancement in protein measurement. A traditional nanoLC column was used for separation of digested but not prefractionated hemolymph prior to MS characterization to compare to the nano-HPLC chip system.

Listed in Table 3 and Table 4 are proteins found by nano-HPLC chip system and nanoLC column. Proteins from each fraction are categorized and listed. These results show that the two different separation columns showed a few similarities, but also significant differences. Both separation systems were able to identify neuropeptide-like precursor 2 in the early fractions. There are six proteins identified by both separation systems. However, there are also significant differences between these two separation systems. The nanoLC column was able to identify 18 proteins, while the nano-HPLC chip system identified 11 proteins. In addition, the nano LC column helped to characterized seven novel proteins with evidence at transcription level in database library, while the nano-HPLC chip system has only found five novel proteins. These differences could be attributed by the configuration of the separation column and time. While the length of the columns are both 150 mm, the nanoLC separation was performed over 60 min. This is double the time required by the separation with the nano-HPLC system. The separation appears to be a main factor affecting the final protein identification. This suggests that further fractionation would lead to an even deeper probing of the hemolymph proteome.

Fraction 1 (0% ACN)	Fraction 2 (10%ACN)	Fraction 3 (20%ACN)	Fraction 4 (30%ACN)	Fraction 5 (70%ACN)
bellwether	yolk protein 1	yolk protein 1	yolk protein 1	NONE
Neuropeptide-like presursor2	Neuropeptide-like presursor2	Neuropeptide-like presursor2	Lp06392p	
	yolk protein 2	Retinoid-and fatty acid-binding glycoprotein	yolk protein 2	
	Retinoid-and fatty acid-binding glycoprotein	yolk protein 2	Neuropeptide-like presursor2	
			Transferrin precursor	

Table 3. A list of proteins identified in various fractions by RPLC column with nano-HPLC chip system.

Fraction 1 (0% ACN)	Fraction 2 (10%ACN)	Fraction 3 (20%ACN)	Fraction 4 (30%ACN)	Fraction 5 (70%ACN)
Neuropeptide-like precursor 2	ATP synthase beta subunit	ATP synthase beta subunit	ATP synthase beta subunit	NONE
	Bellwether	Aldolase alpha	Alcohol dehydrogenase	
	yolk protein	Yolk protein 3	Yolk protein 1	
	Retinoid-and fatty acid-binding glycoprotein	yolk protein 2	Bellwether	
	skpA associated protein, isoform A	TA 01815p	LP03138p	
		Glyceraldehyde-3-phosphate dehydrogenase (Gadph-1) protein(EC 1.2.1.12)		
		Bellwether		
		Cu-Zn superoxide dismutase		
		Pyruvate kinase		
		Pyruvate kinase, isoform B		

Table 4. A list of proteins identified in various fractions by RPLC column with nanoLC column.

3.4. Conclusion

In summary, we present a method using a customized, low pressure, RPLC column and IEX column for prefractionation of individual adult female fruit fly hemolyph, for proteomic analysis. The short columns appear well suited for prefractionation of limited volume hemolymph samples and achieved an enhancement of less abundant protein identification. A list of novel proteins were successfully identified by both RPLC and IEX fractionation methods that were missed when the sample was unfractionated. The ability to enhance less abundant and less well known proteins of interest indicates that these prefractionation approaches promote an in-depth understanding of the protein composition in limited-volume biological complex. Potentially, this simplified prefractionation method followed with the nano-HPLC chip system coupled with Orbitrap MS/MS shows promise in serving as an alternative strategy to balance time consuming proteomic procedures with a need to characterize protein composition.

4. AMINO ACIDS ANALYSIS BY SHEATHLESS CE-MS

4.1. Introduction

Capillary electrophoresis is a well-established technique since the 1980s. As opposed to liquid chromatography, CE utilizes an open fused silica capillary for separation, which eliminates the eddy diffusion and mass transfer; therefore, these particular characteristics benefit the separation with faster separation, high separation efficiency, small sample size and limited solvent consumption (Brocke and Bayer, 2001). The volume of sample required for an efficient CE separation is at the nanoliter level. This low volume compatibility matches well with the limited sample volume available from fruit fly studied in this dissertation. The combination of CE with MS enables a powerful detection method for chemical identify information of limited volume sample. The first sheathless CE-MS interface was by Olivares et al., who presented a study of five quaternary ammonium salt mixture to demonstrate high separation efficiency and high mass resolution with this hyphenated CE-MS (Olivares et al., 1987). They covered the CE capillary outlet with a thin metal tube as the substitute of the terminal electrode in a conventional CE setup. Analytes traveled through the fused silica capillary driven by electroosmotic flow, and were emitted into the gas phase as electrospray. These charged analytes ions later were detected by mass spectrometry.

Over the past two decades, the basic idea of CE-MS has been kept the same. However, the most critical part of CE-MS—the interface between CE and MS, has been updated continuously. Various designs and modification of the interfaces were proposed, for instance liquid junction interface, a porous capillary tip, or inserting electrode into the

capillary (Moini, 2007; Cao and Moini, 1997; Garcia and Henion, 1992). Overall, the development of CE-MS interface has been focused on two main directions: 1) a sheath-flow interface and 2) a sheathless interface.

1) Sheath-flow interface was first developed by adding a flow of liquid in the coaxial tube outside the separation capillary, as shown in Figure 16. This make-up liquid mixes with the CE eluate at the tip of the CE capillary, acting as the terminal electrode to close the electrical circuit for CE. Mass spectrometry has low tolerance to high-boiling point salt content (Maxwell and Chen, 2008; Brocke, Nicholson, and Bayer, 2001). By properly modifying the composition of this make-up liquid, the volatility of the CE electrolyte can be increased to be suitable for mass spectrometry. However, this approach also brings an unavoidable issue. As illustrated in Figure 16 *A) sheath-flow schematic diagram in CE-MS*the mix of make-up liquid with the CE buffer excessively dilutes the analyte bands and results in band broadening, which eventually sacrifices the MS detection sensitivity (Brocke, Nicholson, and Bayer, 2001).

2) In order to elevate the sensitivity of detection for CE-MS, the sheathless interface was developed. The exit tip of CE capillary was tapered to as thin as 10μm (i.d.) serving as the ESI spray tip. The outside of the CE capillary tip is coated with various conductive materials, ranging from metal to graphite and even conductive polymers (Chang and Her, 2000; Kele et al., 2005; Chen and Her, 2003). In this way, the coated exit tip serves as the terminal electrode, without adding any extra fluid to the CE eluent, as shown in Figure 16. The exit tip wall is thin enough to allow the electric field needed to drive electrolytes towards the tip for emission to form the Taylor cone before entering the mass spectrometer.

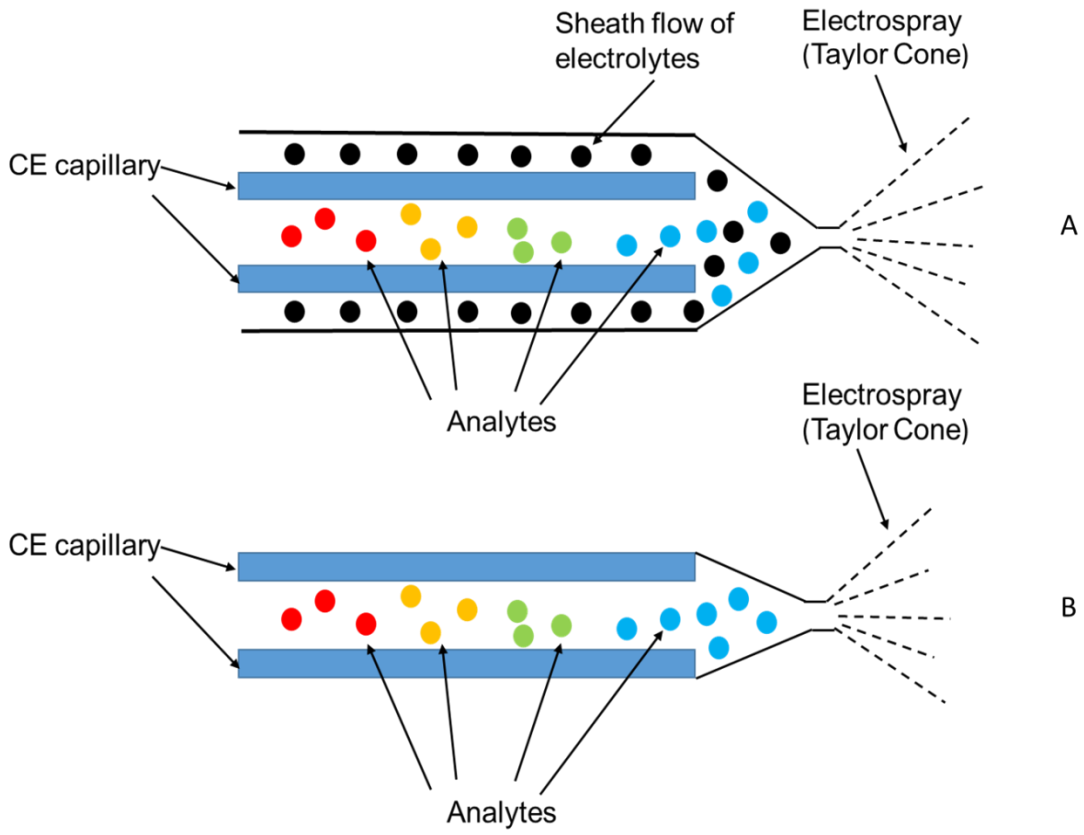


Figure 16 A) sheath-flow schematic diagram in CE-MS. B) sheathless interface diagram in CE-MS.

In this project, a sheathless interface for CE-MS was made by coating the tapered capillary tip with conductive graphite, which is much cheaper than gold or silver coating, serves a longer lifetime, and provides electric stability as good as other coating approaches. Common biologically relevant amino acids glutamate, glycine, and arginine were well separated and identified by the CE-MS interface described in this Chapter.

4.2. Experimental section

4.2.1 Chemicals and Materials

Deionized water was obtained from a US Filter Purelab Plus purification system (Lowell, MA). Amino acid standards of arginine, glycine, glutamate, glutathione were all purchased from Sigma-Aldrich (St. Louis, MO). HPLC grade methanol and ammonium acetate were from Sigma-Aldrich (St. Louis, MO). Conductive carbon was purchased from SPI supplies (West Chester, PA). Fused-silica capillary (360/50-μm o.d. /i.d.) was purchased from Polymicro Technologies (Phoenix, AZ). Microtorch was from Blazer products (Farmingdale, NY).

4.2.2 Fabrication of a conductive carbon coated capillary tip for electrospray

A 70 cm (360/50-μm o.d. /i.d.) fused silica capillary was positioned on a three-dimensional manipulator platform with a section of exposed capillary, as illustrated in Figure 17.

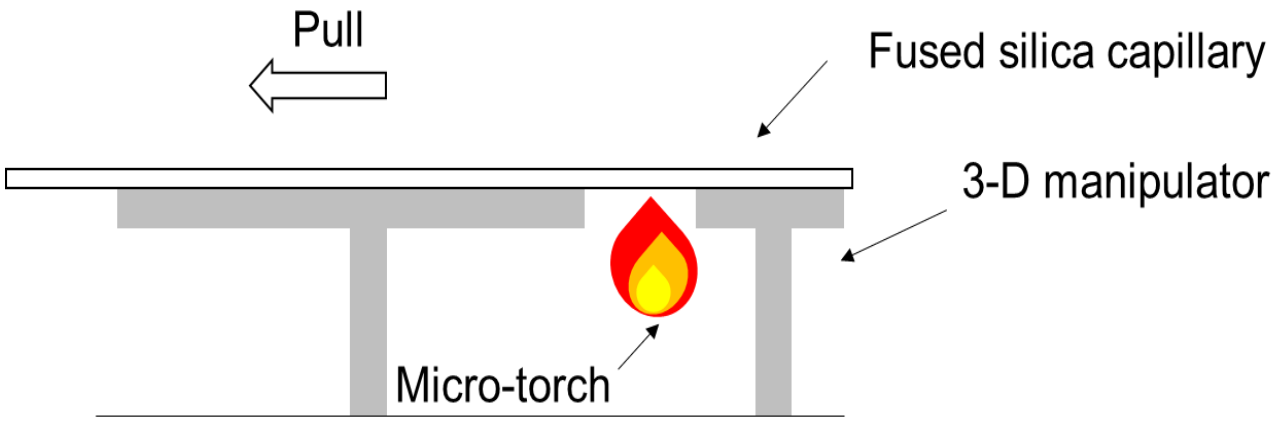


Figure 17 Schematic illustration of electrospray tip fabrication.

A piezomicrotorch (Blazer products, Farmingdale, NY) was placed at the gap, with the flame point at the same height as the mounted capillary. The exposed capillary was heated and upon reaching the melting stage, the longer side of the capillary was immediately pulled away from the gap using the adjustment dial on the manipulator. The dimension of the fabricated capillary tip can reach as low as 50/10-μm o.d. /i.d. as visualized under a dissection scope in comparison to a USAF resolution test target (Thorlabs, Jessup, MD). Any small glass pieces at the tip were removed using a ceramic capillary cutter to create a smooth opening, which was visually verified under a dissecting microscope before use.

A 3-cm long stainless tube was placed coaxially 1-cm away from the emitter tip, to protect the fragile pulled capillary when connected to the ESI power supply by an alligator tip. The emitter tip with the stainless tube was uniformly coated with conductive carbon while nitrogen gas was flowing through the capillary to prevent the tip from being clogged by conductive carbon.

4.2.3 Configuration of CE-ESI-MS instrumentation

As shown in Figure 18, the configuration of CE/ESI-MS instrument was achieved by coupling a home-built CE with a commercial MT Explorer 50 portable ion trap mass spectrometer from MassTech (Columbia, MD). The in-house CE is essentially assembled with a CZE 1000R high voltage power supply (Spellman, Plainview, NY) and a platinum electrode in a buffer vial inside a polyester box to avoid electric shock.

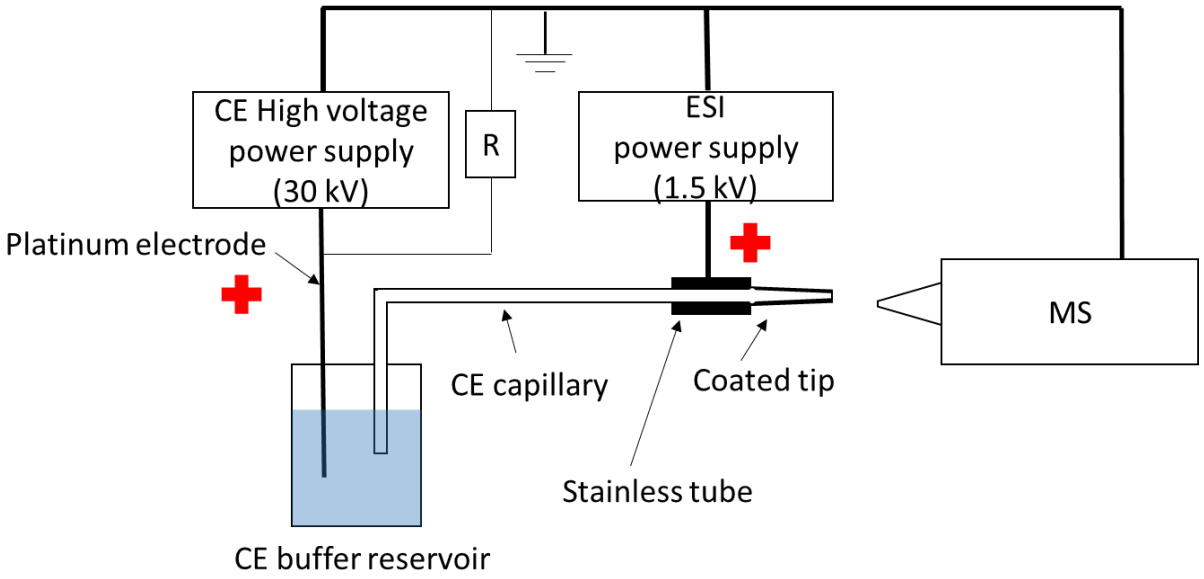


Figure 18 Scheme diagram of CE-MS with conductive carbon coated capillary tip as the interface. R is a resistor to restrict the current on the CE capillary.

CE separation was carried out by applying 27 kV on the cathode side of the CE capillary and 1.5 kV applied at the electrospray tip for electroionization of elelctroosmotic flow from the CE separation capillary. An Analytica ESI power supply (Branford, CT) was connected with the electrospray emitter. A 25 kΩ resistor is inserted into the CE circuit to restrict the current on the CE capillary, and to protect ESI power supply.

The electrospray tip was mounted on a three-dimensional manipulator and aligned with the MS inlet with 1-2 mm horizontal separation distance. The tip was also placed about 2 mm higher than the MS inlet (to account for gravity). The heated transfer capillary was set at a temperature of 150°C. Samples were introduced hydrodynamically into the CE capillary for 30 seconds. Ammonium acetate is a volatile solvent for MS. Herein, a 20 mM ammonium acetate buffer with 10% methanol was adjusted to pH 5.28 by HCl. Then this buffer was used as CE run buffer for separation, and assisted in the formation of the electrospray Taylor cone. In between CE runs, the CE capillary was washed with 0.1 M NaOH and deionized water followed by CE run buffer to avoid any carry-over. A full MS survey scan was set from 35 to 200 m/z. CE electrophoregrams and MS spectra data files were acquired by MT 20, which is embedded in the MTE 50 mass spectrometer, and then analyzed by Chromexplorer software (MassTech, Columbia, MD). The threshold of signal-to-noise ratio was set at 3.

4.3. Results and Discussion

4.3.1. Stability and durability of electrospray by conductive carbon coated capillary tip

One of the prerequisites for successful CE-MS experiments is a reliable and stable electrospray from the CE capillary tip at the MS inlet. The stability of electrospray is vital to subsequent MS identification. The CE eluent needs to be nebulized or dispersed into a fine aerosol and eventually ionized due to extensive solvent evaporation. Charged analyte ions entering the MS are selected by the mass analyzer for detection by mass-to-charge ratio.

Various dimensions of fused silica capillary were tested for feasibility and durability. The commercially available 20/90-µm o.d. /i.d. capillary is as thin as hair and extremely fragile. Capillaries of this dimension could not withstand the pull force required in pulling a tapered capillary tip. Capillaries with 75/360-µm o.d./i.d. were able to endure the pulling required for a tapered tip to a 10/50-µm o.d./i.d. However, the large capillary diameter increased the CE eluent flow rate which did not provide enough time for evaporation of solvent needed for effective electrospray ionization. Fortunately, 50/3600-µm o.d./i.d. capillary was found ideal both for forming the tapered the capillary tip and for generating appropriate flow rates for stable electrospray ionization.

Another experimental variable that had to be developed was the CE run buffer which is compatible with MS. Volatile solvents with organic compounds are required to assist the formation of an ideal electrospray. A run buffer consisting of 20 mM ammonium acetate with 10% methanol at pH 3.45 compatible with MS was tested, which is expected to aid the electrospray by providing sufficient proton transfer. However, there appeared to be severe suppression of electroosmotic flow (EOF) in this acidic buffer and the separation time increased beyond one hour. The time alone is unacceptable for a CE and causes problems in generating stable electrospray due to

low flow rate. Therefore, a higher pH buffer with ammonium acetate and 10% methanol at pH 5.28 were used as the CE run buffer. As shown in Figure 19, a total ion current (TIC) chromatogram of this CE run buffer generated an ion count at 10^6 which is high enough for electrospray ionization. Notably, this high ion count level was reasonably stable for more than 70 min, which provides sufficient time for CE-MS analysis.

4.3.2. Detection efficiency of CE-MS with amino acids and peptide standards

With a reliable and stable electrospray ionization by the sheathless conductive carbon coated CE capillary, the detection efficiency of this sheathless CE-MS was examined with amino acid and peptide standards. As demonstrated in Figure 20(A), 2.5 µM glutamate and arginine standards were resolved and detected in CE-MS. The mass electropherogram at m/z 146.4 and 174.4 represents the molecular ions of glutamate and arginine, respectively. Angiotensin I is a polypeptide with monoisotopic mass 1296.7 Da. A singly charged angiotension I was not detected in Figure 20(B). Nevertheless, the doubly charged form of angiotensin I was identified at m/z 648.1.The mass peaks at m/z 612.0 and 281.3 are probably fragments from angiotensin I.

A mixture of 1 mM four amino acids glutathione, arginine, glycine and glutamate, were separated by this sheathless CE-MS within 13 min. Figure 21 shows the CE separation electropherogram extracted from the TIC chromatogram of these four amino acids. Glutathione and arginine were well resolved and separated. However, glycine and glutamate peaks have similar retention time, according to the electropherogram. However, with the aid of MS spectra, these two amino acids can be distinguished from

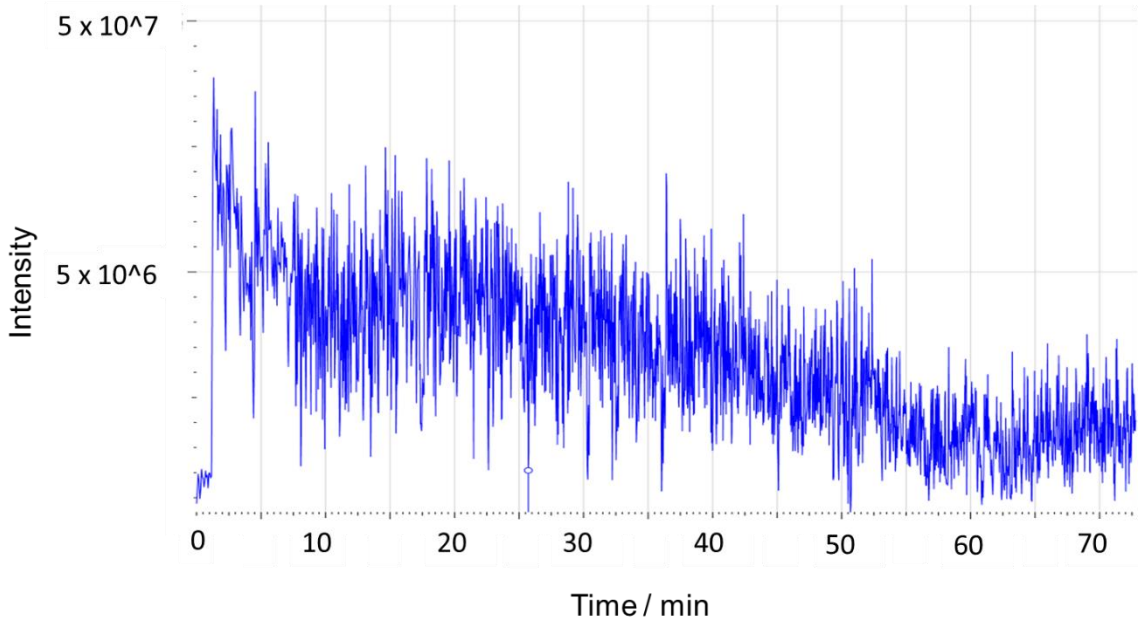


Figure 19 Stability of electrospray ionization by conductive carbon coated capillary tip. The total ion count of the 20 mM ammonium acetate with 10 % methanol at pH 5.28 is observed stable at the magnitude of 10^6 for more than 70 min.

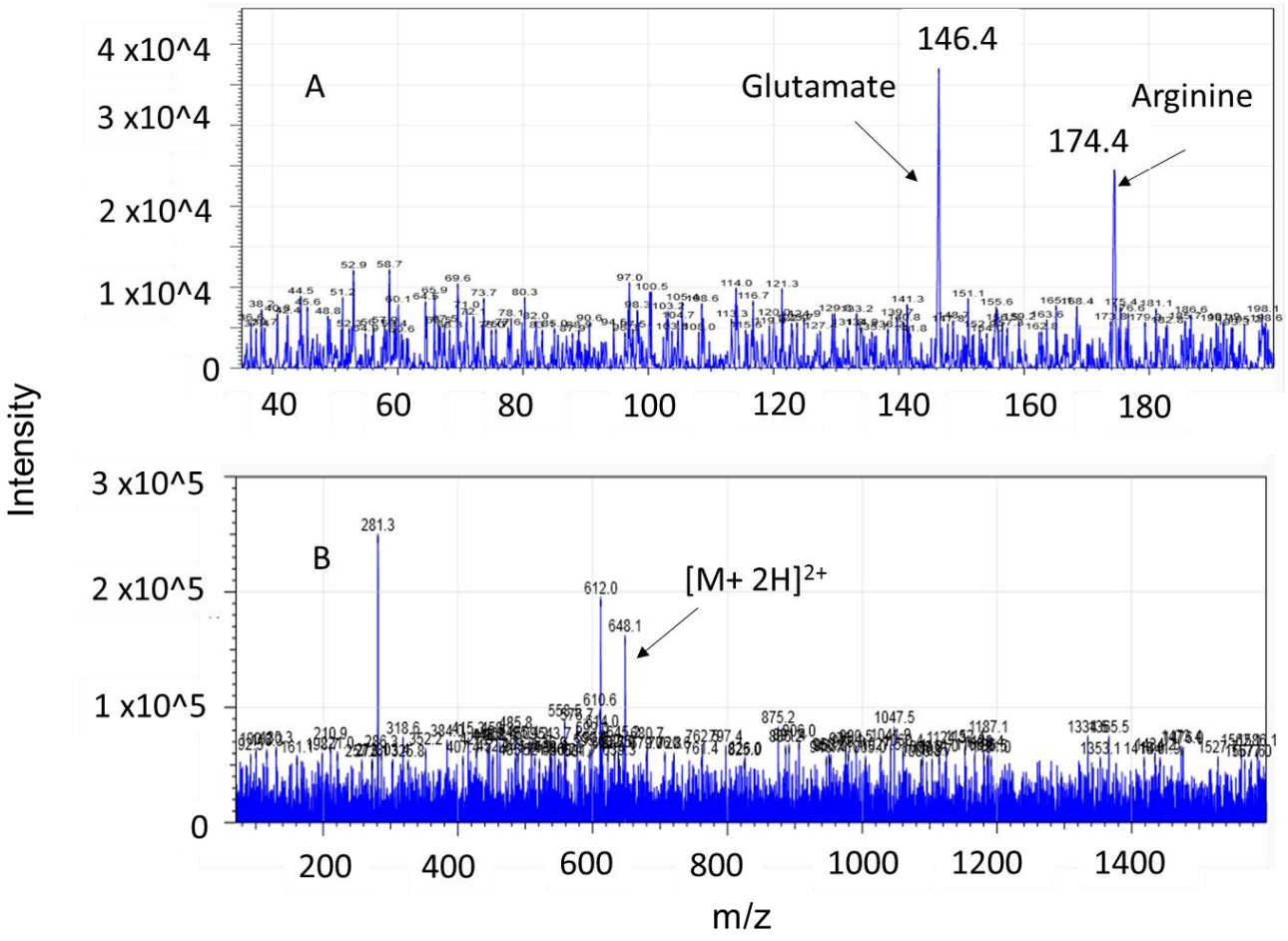


Figure 20 (A) MS spectrum of amino acids: 2.5 µM glutamate at 146.4 m/z and arginine at 174.4 m/z. (B) MS spectrum of peptide standard 40 µM angiotensin I in the form of $[M+ 2H]^{2+}$.

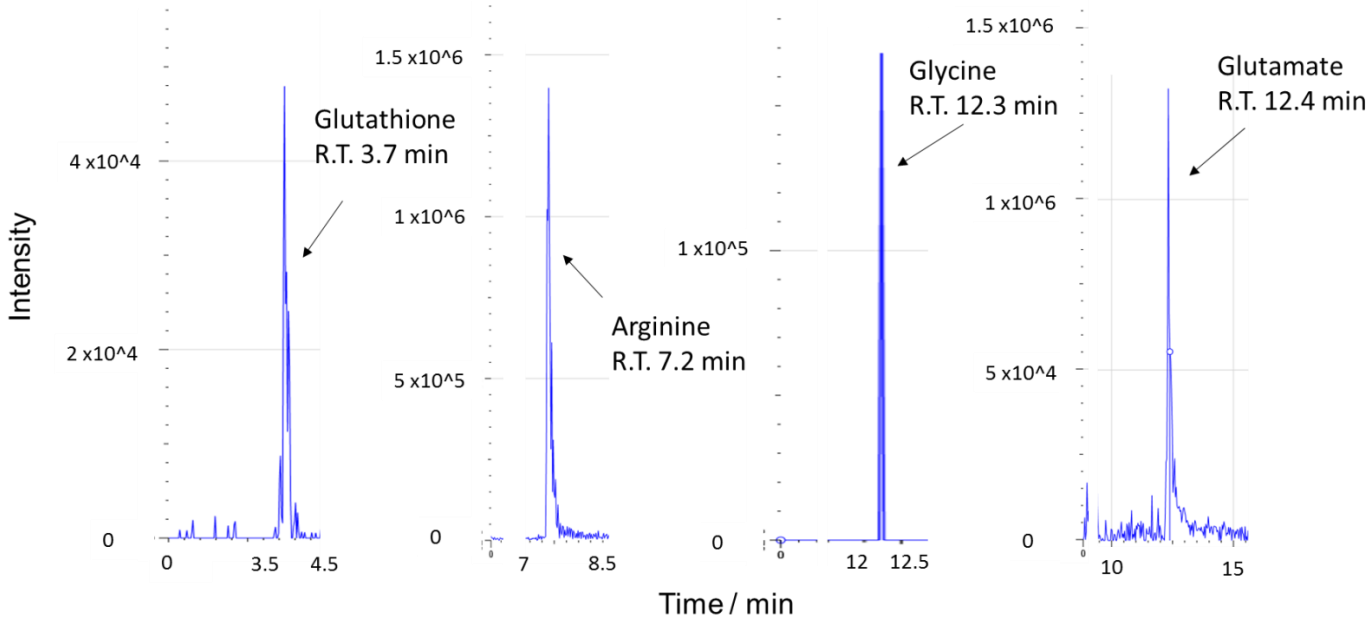


Figure 21 CE electropherogram of 4 amino acids by CE-MS. Each amino acid electropherogram was extracted from the TIC chromatogram of a CE-MS run.

each other. This sheathless CE-MS provides an alternative detection method compared to fluorescence. As described in Chapter 2 and Chapter 3, the detection was forced to be off-line due to the extra fluorescence derivatization reaction. In addition, the derivatization is sensitive to pH.

It is difficult to measure pH accurately in limited volume samples and there is a risk of significant dilution. In this study, a conductive carbon coated capillary tip was developed to couple CE with MS. A list of amino acids was separated and detected on-line by this coupled CE-MS instrument without any extra time and effort consuming derivatization procedure.

4.4. Conclusion

In this chapter, an inexpensive and simple sheathless electrospray interface was developed by fabricating a conductive carbon coated capillary for the electrospray ionization. With optimized conditions of CE separation, a stable and reliable electrospray was achieved for more than an hour. The CE run buffer was also optimized, and 20 mM ammonium acetate with 10 % methanol at pH 5.28 was selected as an acceptable separation buffer in CE. Amino acids and peptide standards were well resolved and separated by this sheathless CE-MS. Future application of this instrumentation will be performed with a complex biological sample, for instance, fruit fly hemolymph, human tear is in consideration using this sheathless CE-MS method.

5. FUTURE DIRECTIONS AND SUPPLEMENTARY

5.1. Introduction

Overall, the goal of this thesis is focused on developing analysis methods to understand chemical information, especially proteomic composition in limited volume hemolymph samples. *Drosophila melanogaster* is an ideal research tool, due to its great similarity to human genes, the low-cost maintenance, the fast reproducibility and the easy gene manipulability. In Chapter 2 and 3, the development of efficient LC-MS analysis and prefractionation methods were described. In addition, Chapter 4 describes development of a sheathless flow CE-MS instrument, which could be used for analysis of amino acids in fruit fly hemolymph. These methods are a significant advance for the study of protein level composition of limited volume biological sample, as well as the capability of enhancing identification of low-abundance proteins and novel proteins. Nevertheless, a wider investigation of analysis methods should be taken into account to demonstrate alternative and improved methods available as analysis tools for this and other volume-limited biological samples. Therefore, a couple of techniques are discussed in this chapter to understand and interpret the chemical information from *Drosophila* hemolymph. The first part of this Chapter is devoted for determination of total body fluid in individual fruit fly. The second part of this Chapter describes an alternative chromatographic approach—size exclusion chromatography, for prefractionation of limited-volume, complex proteomic samples. Preliminary result of fractionation of hemolymph indicates the potentially full development of this chromatographic method. The last part of this Chapter was focused on the optimization of CE separation buffer condition for resolving amino acids, which

otherwise might not be well separated in CE. This part work supplements research done in Chapter 4, and hopefully can improve peak capacity in CE-MS.

5.2. Determination of total body fluid in individual fruit fly

5.2.1. Sample collection of total body fluid

The Oregon-R strain of *Drosophila melanogaster* was employed for the investigation of total body fluid in individual adult fruit fly. The flies were reared on standard cornmeal-agar medium and maintained in UIC Biological Science Department. Both female and male adult flies were collected. Female flies are larger in size compared to male ones as female flies possess larger amount of fluid due to their egg production role in reproduction. Each group of flies were cold anesthetized in a 10-mL tube in freezer for 5 min, and then transferred to a microcentrifuge tube. The net weight of each group of flies was measured after the anesthetization. Then each group of flies were lyophilized overnight prior to a second measurement of the weight of the desiccated flies.

5.2.2. Result and Discussion

As shown in Table 5, three replicates were measured for determination of body fluid weight in the average adult fruit fly. On average, there is 69.93 ± 0.25 % of body fluid in individual fly, which is comparable to values in a report of flies subjected to severe dehydration (Albers and Bradley, 2004). With a closer look into the gender difference, it is found that male flies have significantly less volume of body fluid compared to their

female counterparts. Exact volume solved from a system of two variable equations from sample 2 and 3 is presented in Table 5 Determination of total body fluid in individual adult fruit fly. It can be explained from the body size difference of both genders. Male flies are normally in smaller size than female ones. When the female flies are about to lay eggs, the body size increases dramatically, in order to prepare and store large amount of nutrients for reproduction. The amount of body fluid would likely also increase to adapt for this physiological change. With this preliminary study of the total body fluid of *Drosophila*, it appears that female flies possess almost ten times larger volume of body fluid compared to males. Therefore, adult female flies are exclusively selected as research target for hemolymph sampling and collection in this dissertation. Even though female flies provide around 1 µL of body fluid in total, the amount of hemolymph available is extensively limited. Hemolymph is only the extracellular fluid circulating through the open circulatory cavity in the fly. There is a significant amount of intracellular fluid contained inside cells, and could not be considered as part of hemolymph. Therefore, the amount of hemolymph available is much smaller than 1 µL. In fact, a maximum of 50 nL of hemolymph was able to collect from individual adult female fruit fly, as reported by previous studies in our lab (Piyankarage, Featherstone, and Shippy, 2012; Piyankarage et al., 2008).

Sample #.	# of flies	Weight of fluid in all flies (mg)	Average weight of fluid in single fly (mg) (♂&♀)	Gross weight of all flies (mg)	Percentage of fluid in fly (%)
1	16 (8 ♂ & 8 ♀)	11.69	0.73	16.66	70.2
2	16 (8 ♂ & 8 ♀)	12.04	0.75	17.23	69.9
3	14 (8 ♂ & 6 ♀)	9.09	0.65	13.04	69.7

Average vol. of fluid from single fly (nL) (♂&♀)	Vol. of fluid from single MALE fly ♂ (nL)	Vol. of fluid from single FEMALE fly ♀ (nL)
730	161.25	1300
750		
650		

Table 5 Determination of total body fluid in individual adult fruit fly

5.3. Prefractionation of fruit fly hemolymph by size exclusion chromatography

Molecular size is one of the important characteristics that plays a vital role in proteomic separation, due to its simplicity. For example, the most common separation technique used in biology science is SDS-PAGE, in which size difference affects the length analytes migrate through the gel. However, gel separation involves a tedious procedure and large amount of buffer, which could excessively dilute the sample and lost the possible vital chemical information. Therefore, a microliter-scale size exclusion chromatography could be helpful in prefractionation of proteomic sample, to improve the identification of proteins, without the excessive dilution and information loss concern.

5.3.1. Fabrication of size exclusion column for prefractionation

Prefractionation method using size exclusion chromatography was developed similar to reversed-phase and ion-exchange chromatography described in Chapter 3. In brief, Sephadex G-25 resins were used as column packing material, since this resin is suitable for fractionation of analytes with molecular weight between 1000 and 5000 g/mol, which is the range of hemolymph peptides molecular mass after trypsin digestion. A detailed procedure of hemolymph collection from individual adult female fruit fly has been developed in our lab and elaborated in detail in Chapter 2.

The Sephadex G-25 resins were aspirated into a 5-cm long tygon tubing by vacuum made by syringe attached to the end of the tygon tubing. With Sephadex G-25 resins packed tightly in the 5-cm tygon tube, both ends of the tygon tube was trimmed to

make a 2-cm long evenly packed size exclusion column. Finally, a 0.5-cm long (360/50-µm o.d. /i.d.) fused silica capillary was inserted at the end of the tygon tubing column. Prior to hemolymph separation, tryptic digested BSA was employed as a standard to test the separation efficiency of this customized column. Hemolymph sample was loaded onto the column by a 1-mL insulin injection syringes using a Harvard syringe pump at flow rate 0.5 µL/min, and then eluted using isocratic elution. The eluent was 20 mM phosphate buffer at pH 7.24. A series of 2-µL fractions of eluate were collected. Each eluate fraction was then derivatized with fluorescamine (1.5 mg/mL in acetone) for 6 min prior to detection. Due to the limited volume of each fraction, a 10 cm (360/50-µm o.d. /i.d.) fused silica capillary was used as the detection cell. Each eluate was injected into the capillary cell for triplicate measurements of fluorescence intensity. The order of deionized water, HCl, deionized water, NaOH and deionized water was followed to rinse the capillary cell before each eluate fraction injection.

5.3.2. Result and Discussion

The fractionation of SEC follows the principle, in which, the analyte retention time vary due to the size difference. Since a molecule below the size of the resin pore size is able to penetrate into every region of the stationary phase, it is retained on the stationary phase for a longer time than a larger molecule that cannot diffuse into stationary phase pores. An analyte molecule above the size of the stationary phase pore is too large to penetrate into every region of the stationary phase pore system, and subsequently spend less retention time on the stationary phase.

An isocratic elution method was used in this study, other than step-gradient elution method used in Chapter 2 for reversed-phase and ion-exchange chromatography. This is because the role of mobile phase in SEC is merely to carry analytes through the column. From Figure 22(A), it is observed that 13 fractions were collected from this customized size exclusion column. This fractionation chromatogram suggests that most of hemolymph peptides were eluted out and collected in fraction 1. This observation indicates that most analytes passed directly through the column, without entering the pores of the stationary phase. This results implies that many or most tryptic peptide analytes are fairly large molecules which could not penetrate into the stationary phase pores. Because peptide fragments were cleaved by enzyme trypsin, one approach to improving this results is to adjust the substance-to-protease ratio. In Figure 22(A), a 50:1 ratio was applied for protein-to-trypsin, and resulted in cleavage coverage of proteins that did not lead to efficient SEC fractionation. Therefore, a 20:1 ratio was adapted. As shown in Figure 22(B), a significant decrease of signal intensity of fraction 1 was observed, which is about 5 times less than fraction 1 in Figure 22(A). In addition, fraction 2, 9, 10, and 13 reported relatively high intensity fractions over the whole series of fractions, which present a reasonable baseline separation of complex hemolymph peptide mixture.

Notably, Figure 22 (B) does require more fractions to collect all analytes compared to fractionation in Figure 22 (A). It’s exactly what would be expected, because a sufficient amount of trypsin promotes a much more thorough cleavage of peptide fragments and therefore, generates a larger quantity of small polypeptide fragments, rather than several large

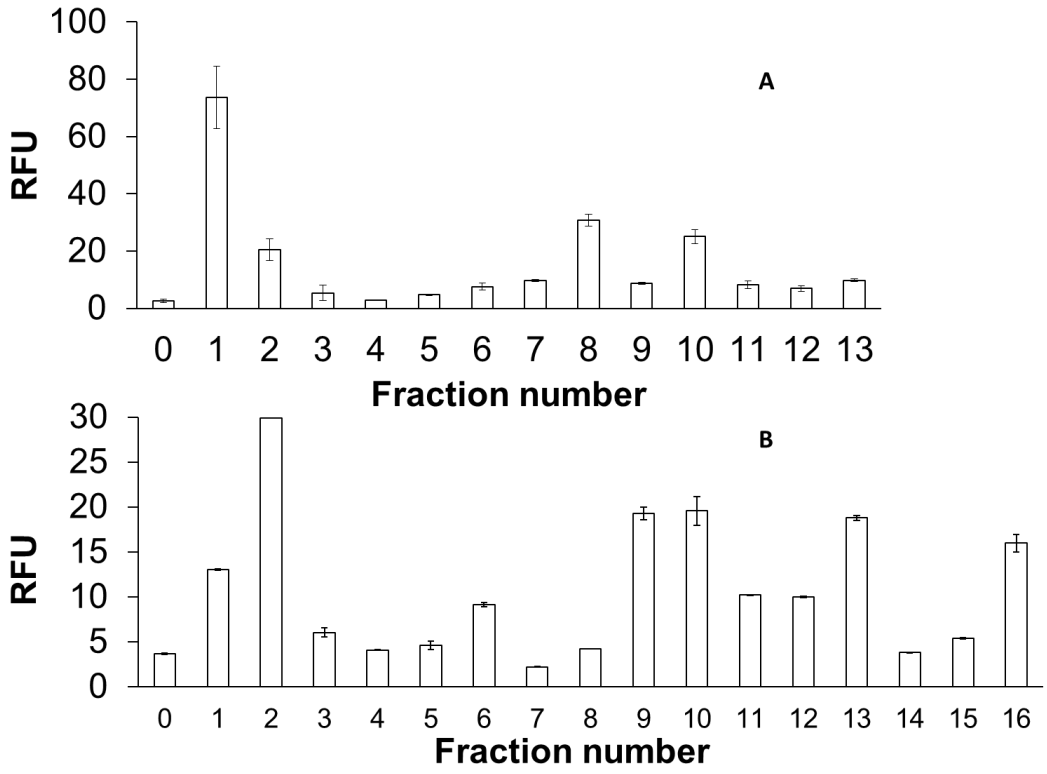


Figure 22 Fractionation chromatogram of tryptic hemolymph peptides from size exclusion column. A) Inefficient SEX fractionation of hemolymph sample with a 50:1 ratio applied for protein-to-trypsin. B) An improved SEX fractionation of hemolymph sample with a 20:1 ratio applied for protein-to-trypsin.

fragments of undigested proteins. With a more thoroughly digested peptide mixture, the LC fractions are composed of more small size peptides, which could likely promote more efficient analysis by mass spectrometry to aid in the identification of low abundance and novel proteins.

5.4. Summary and conclusions

Overall, the research described in this dissertation is focused on developing analysis tools for volume-limited complicated biological sample for proteomic study. The novel LC-MS analysis method enables separation and identification of 19 proteins from less than 100 nL hemolymph from individual adult fruit fly. This method also helps to identify novel proteins in this limited volume of sample. With the novel prefractionation methods developed, a simpler and efficient reduction of protein complexity aids the identification of less abundance proteins without being blinded by large abundant proteins. The coupling technique developed for coupling CE and MS, along with preliminary analysis approaches in chapter 5 demonstrates alternative tools for proteomic analysis, which complements the two approaches in Chapter 2 and 3 and have the potential to play an important role in understanding chemical information of fundamental physiology issues, from the proteomics aspect.

6. REFERENCES

M. D. Adams, S. E. Celniker, R. A. Holt, C. A. Evans, J. D. Gocayne, P. G. Amanatides, S. E. Scherer, P. W. Li, R. A. Hoskins, R. F. Galle, R. A. George, S. E. Lewis, S. Richards, M. Ashburner, S. N. Henderson, G. G. Sutton, J. R. Wortman, M. D. Yandell, Q. Zhang, L. X. Chen, R. C. Brandon, Y. C. Rogers, R. G. Blazej, M. Champe, B. D. Pfeiffer, K. H. Wan, C. Doyle, E. G. Baxter, G. Helt, C. R. Nelson, G. L. G. Miklos, J. F. Abril, A. Agbayani, H. An, C. Andrews-Pfannkoch, D. Baldwin, R. M. Ballew, A. Basu, J. Baxendale, L. Bayraktaroglu, E. M. Beasley, K. Y. Beeson, P. V. Benos, B. P. Berman, D. Bhandari, S. Bolshakov, D. Borkova, M. R. Botchan, J. Bouck, P. Brokstein, P. Brottier, K. C. Burtis, D. A. Busam, H. Butler, E. Cadieu, A. Center, I. Chandra, J. M. Cherry, S. Cawley, C. Dahlke, L. B. Davenport, P. Davies, B. de Pablos, A. Delcher, Z. Deng, A. D. Mays, I. Dew, S. M. Dietz, K. Dodson, L. E. Doup, M. Downes, S. Dugan-Rocha, B. C. Dunkov, P. Dunn, K. J. Durbin, C. C. Evangelista, C. Ferraz, S. Ferriera, W. Fleischmann, C. Fosler, A. E. Gabrielian, N. S. Garg, W. M. Gelbart, K. Glasser, A. Glodek, F. Gong, J. H. Gorrell, Z. Gu, P. Guan, M. Harris, N. L. Harris, D. Harvey, T. J. Heiman, J. R. Hernandez, J. Houck, D. Hostin, K. A. Houston, T. J. Howland, M. Wei, C. Ibegwam, M. Jalali, F. Kalush, G. H. Karpen, Z. Ke, J. A. Kenniso, K. A. Ketchum, B. E. Kimmel, C. D. Kodira, C. Kraft, S. Kravitz, D. Kulp, Z. Lai, P. Lasko, Y. Lei, A. A. Levitsky, J. Li, Z. Li, Y. Liang, X. Lin, X. Liu, B. Mattei, T. C. McIntosh, M. P. McLeod, D. McPherson, G. Merkulov, N. V. Milshina, C. Mobarry, J. Morris, A. Moshre, S. M. Mount, M. Moy, B. Murphy, L. Murphy, D. M. Muzny, D. L. Nelson, D. R. Nelson, K. A. Nelson, K. Nixon, D. R. Nusskern, J. M. Pacleb, M. Palazzolo, G. S. Pittman, S. Pan, J. Pollard, V. Puri, M. G. Reese, K. Reinert, K. Remington, R. D. C. Saunders, F. Scheeler, H. Shen, B. C. Shue, I. Siden-Kiamos, M. Simpson, M. P. Skupski, T. Smith, E. Spier, A. C. Spradling, M. Stapleton, R. Strong, E. Sun, R. Svirskas, C. Tector, R. Turner, E. Venter, A. H. Wang, X. Wang, Z. Wang, D. A. Wassarman, G. M. Weinstock, J. Weissenbach, S. M. Williams, T. Woodage, K. C. Worley, D. Wu, S. Yang, Q. A. Yao, J. Ye, R. Yeh, J. S. Zaveri, M. Zhan, G. Zhang, Q. Zhao, L. Zheng, X. Zheng, F. N. Zhong, W. Zhong, X. Zhou, S. Zhu, X. Zhu, H. O. Smith, R. A. Gibbs, E. W. Myers, G. M. Rubin, J. C. Venter, Science. 287 (2000) 2185–2195.

Albers, M. A. & Bradley, T. Osmotic regulation in adult Drosophila melanogaster during dehydration and rehydration. J. Exp. Biol. 207, 2313–2321 (2004).

Amster, I. Fourier transform mass spectrometry. J. Mass Spectrom. 31, 1325–1337 (1996).

Anderson, N. L. & Anderson, N. G. Proteome and proteomics: new technologies, new concepts, and new words. Electrophoresis. 19, 1853–1861 (1998).

Anholt, R. R. H. & Williams, T. I. The soluble proteome of the Drosophila antenna. Chem. Senses 35, 21–30 (2010).

Badock, V., Steinhusen, U., Bommert, K. & Otto, A. Prefractionation of protein samples for proteome analysis using reversed-phase high-performance liquid chromatography. Electrophoresis 22, 2856–2864 (2001).

Bantscheff, M., Schirle, M., Sweetman, G., Rick, J. & Kuster, B. Quantitative mass spectrometry in proteomics: a critical review. Anal. Bioanal. Chem. 389, 1017–1031 (2007).

Barnett, T., Pachl, C., Gergen, J. P. & Wensink, P. C. The Isolation and Characterization Yolk Protein Genes of Drosophila. Cell 21, 729–738 (1980).

Blackstock, W. P. & Weir, M. P. Proteomics: quantitative and physical mapping of cellular proteins. Trends Biotechnol. 17, 121–127 (1999).

Brinkmalm, G., Portelius, E., Öhrfelt, A., Mattsson, N., Persson, R., Gustavsson, M. K., Vite, C. H., Gobom, J., Månsson, J., Nilsson, J., Halim, A., Larson, G., Rüetschi, U., Zetterberg, H., Blennow, K., Brinkmalm, A., An online nano-LC-ESI-FTICR-MS method for comprehensive characterization of endogenous fragments from amyloid β and amyloid precursor protein in human and cat cerebrospinal fluid. J. Mass Spectrom. 47, 591–603 (2012).

Brocke, A. Von, Nicholson, G. & Bayer, E. Review Recent advances in capillary electrophoresis / electrospray-mass spectrometry. Electrophoresis 10, 1251–1266 (2001).

Burmester, T., Storf, J., Hasenjäger, A., Klawitter, S. & Hankeln, T. The hemoglobin genes of Drosophila. FEBS J. 273, 468–480 (2006).

Canavoso, E., Jouni, Z. E., Karnas, K. J., James, E. & Wells, M. A. Fat etabolism in insects. Annu. Rev. Nutr. 23–46 (2001).

Cao, P. & Moini, M. A Novel Sheathless Interface for Capillary. Am. scociety mass Spectrom. 8, 561–564 (1997).

Cheng, L., Zhang, S., Maclennan, G. T., Williamson, S. R., Davidson, D. D., Wang, M., Jones, T. D., Lopez-beltran, A., Montironi, R.. Laser-assisted Microdissection in Translational Research. Appl. Immunohistochem Mol Morphol. 21, 31–47 (2013).

Chien, S., Reiter, L. T., Bier, E. & Gribskov, M. Homophila: human disease gene cognates in Drosophila. Nucleic Acids Res. 30, 149–151 (2002).

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

Clapham, D. E. Calcium Signaling Review. Cell 80, 259–268 (1995).

Comisarow, M. B. & Marshall, A. G. Fourier transform ion cyclotron resonance. *Chem. Phys. Lett.* **25,** 282–283 (1974).

Consortium, T. U. Activities at the Universal Protein Resource (UniProt). Nucleic Acids Res. 42, D191–8 (2014).

De Gregorio, E., Spellman, P. T., Rubin, G. M. & Lemaitre, B. Genome-wide analysis of the Drosophila immune response by using oligonucleotide microarrays. Proc.Natl.Acad.Sci.U.S.A 98, 12590–12595 (2001).

De Morais Guedes, S., Vitorino, R., Domingues, R., Tomer, K., Correia, A. J. F., Amado, F., Domingues, P., Proteomics of immune-challenged Drosophila melanogaster larvae hemolymph. Biochem. Biophys. Res. Commun. 328, 106–115 (2005).

Driscoll, M. & Gerstbrein, B. Dying for a cause: invertebrate genetics takes on human neurodegeneration. Nat. Rev. Genet. 4, 181–194 (2003).

Dunkov, B. C. & Georgieva, T. Organization of the ferritin genes in Drosophila melanogaster. DNA Cell Biol. 18, 937–944 (1999).

Edelmann, M. J. Strong cation exchange chromatography in analysis of posttranslational modifications: innovations and perspectives. J. Biomed. Biotechnol. 2011, 1–7 (2011).

Feuerstein, I., Morandell, S., Stecher, G., Huck, C. W., Stasyk, T., Huang, H., Huber, L. A., Bonn, G. K., Phosphoproteomic analysis using immobilized metal ion affinity chromatography on the basis of cellulose powder. Proteomics 5, 46–54 (2005).

Fonslow, B. R., Carvalho, P. C., Academia, K., Freeby, S., Xu, T., Nakorchevsky, A., Paulus, A., Yates, J. R., Improvements in proteomic metrics of low abundance proteins through proteome equalization using ProteoMiner prior to MudPIT. J. Proteome Res. 10, 3690–3700 (2011).

Garcia, F. & Henion, J. Gel-filled capillary electrophoresis/mass spectrometry using a liquid junction ion spray interface. Anal. Chem. 64, 985–990 (1992).

Greenspan, R. J. & Dierick, H. A. “Am not I a fly like thee?” From genes in fruit flies to behavior in humans. Hum. Mol. Genet. 13, R267–R273 (2004).

Grewal, S. S. Controlling animal growth and body size - does fruit fly physiology point the way? F1000 Biol. Rep. 4, 1–6 (2012).

Gutstein, H. B. & Morris, J. S. Laser capture sampling and analytical issues in proteomics. Expert Rev. Proteomics 4, 627–637 (2007).

Heeren, R. M. A., Kleinnijenhuis, A. J., McDonnell, L. A. & Mize, T. H. A mini-review of mass spectrometry using high-performance FTICR-MS methods. Anal. Bioanal. Chem. 378, 1048–1058 (2004).

Hofstadler, S. A., Sannes-Lowery, K. A. & Hannis, J. C. Analysis of nucleic acids by FTICR MS. Mass Spectrom. Rev. 24, 265–285 (2005).

Howard, G. A. & Martin, a J. P. The separation of the C12-C18 fatty acids by reversed-phase partition chromatography. Biochem. J. 46, 532–8 (1950).

Hu, Q., Noll, R. J., Li, H., Makarov, A., Hardman, M., Graham Cooks, R., The Orbitrap: a new mass spectrometer. J. Mass Spectrom. 40, 430–443 (2005).

Iijima-Ando, K. & Iijima, K. Transgenic Drosophila models of Alzheimer's disease and tauopathies. Brain Struct. Funct. 214, 245–262 (2010).

James, P. Protein identification in the post-genome era: the rapid rise of proteomics. Q. Rev. Biophys. 30, 279–331 (1997).

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

Karas, M. & Krüger, R. Ion formation in MALDI: the cluster ionization mechanism. Chem. Rev. 103, 427–440 (2003).

Karlsson, C. Korayem, A. M., Scherfer, C., Loseva, O., Dushay, M. S., Theopold, U., Proteomic analysis of the Drosophila larval hemolymph clot. J. Biol. Chem. 279, 52033–52041 (2004).

Kawamura, K., Shibata, T., Saget, O., Peel, D. & Bryant, P. J. A new family of growth factors produced by the fat body and active on Drosophila imaginal disc cells. Development 126, 211–219 (1999).

Kelly, L. E. Purification and properties of a 23 kDa Ca2(+)-binding protein from Drosophila melanogaster. Biochem. J. 271, 661–666 (1990).

Klose, J. & Embryonal-pharmakologie, A. Protein Mapping by Combined Isoelectrie Focusing and Electrophoresis of Mouse Tissues. Humangenetik 26, 231–243 (1975).

Knochenmuss, R. & Zenobi, R. MALDI ionization: the role of in-plume processes. Chem. Rev. 103, 441–452 (2003).

Koenig, T., Menze, B. H., Kirchner, M., Monigatti, F., Parker, K. C., Patterson, T., Steen, J. J., Hamprecht, F. A., Steen, H., Robust Prediction of the MASCOT Score for an

Improved Quality Assessment in Mass Spectrometric Proteomics research articles. J. Proteome Res. 3708–3717 (2008).

Kopaciewicz, W., Rounds, M. A., Fausnaugh, J. & Regnier, F. E. Chromatography, 266 (1983) 3-2 I. J. Chromatogr. 266, 3–21 (1983).

Lathe, G. H. & Ruthven, C. R. The separation of substances and estimation of their relative molecular sizes by the use of colums of starch in water. Biochem. J. 62, 665–74 (1956).

Lindmark, R., Biriell, C. & Sjoquist, J. Quantitation of Specific IgG Antibodies in Rabbits by a Solid-phase Radioimmunoassay with 125I-Protein A from Staphylococcus aureus. Scand. J. Immunol. 14, 409–420 (1981).

Makarov, A. Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. Anal. Chem. 72, 1156–1162 (2000).

Marshall, A. G. & Hendrickson, C. L. High-resolution mass spectrometers. Annu. Rev. Anal. Chem. 1, 579–599 (2008).

Martosella, J., Zolotarjova, N., Liu, H., Nicol, G. & Boyes, B. E. Prefractionation of Immunodepleted Human Serum Proteins to Enhance Mass Spectrometry Identification of Lower-Abundant Proteins research articles. J. Proteome Res. 4, 1522–1537 (2005).

Maxwell, E. J. & Chen, D. D. Y. Twenty years of interface development for capillary electrophoresis-electrospray ionization-mass spectrometry. Anal. Chim. Acta 627, 25–33 (2008).

McBride, A. E. & Silver, P. A. State of the arg: protein methylation at arginine comes of age. Cell 106, 5–8 (2001).

Miller, J. Chromatography: concepts and contrast. 543 (2005).

Moini, M. Simplifying CE-MS operation. 2. Interfacing low-flow separation techniques to mass spectrometry using a porous tip. Anal. Chem. 79, 4241–4246 (2007).

Nittis, T., Guittat, L., LeDuc, R. D., Dao, B., Duxin, J. P., Rohrs, H., Townsend, R. R., Stewart, S. A., Revealing novel telomere proteins using in vivo cross-linking, tandem affinity purification, and label-free quantitative LC-FTICR-MS. Mol. Cell. Proteomics 9, 1144–1156 (2010).

Okada, M. & Natori, S. Primary Structure of Sarcotoxin I, an Antibacterial Protein Induced in the Hemolymph. J. Biol. Chem. 260, 7174–7177 (1985).

Olivares, J., Nguyen, N., Yonker, C. & Smith, R. On-Line Mass Spectrometric Detection for Capillary Zone Electrophoresis. Anal. Chem. 59, 1230–1232 (1987).

Olsen, J. V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P., Mann, M., Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell 127, 635–648 (2006).

Patterson, E. E., Piyankarage, S. C., Myasein, K. T., Pulido, J. S., Dundervill, R. F., Hatfield, R. M., Shippy, S.A., A high-efficiency sample introduction system for capillary electrophoresis analysis of amino acids from dynamic samples and static dialyzed human vitreous samples. Anal. Bioanal. Chem. 392, 409–416 (2008).

Piyankarage, S. C., Augustin, H., Grosjean, Y., Featherstone, D. E. & Shippy, S. A. Hemolymph amino acid analysis of individual Drosophila larvae. Anal. Chem. 80, 1201–1207 (2008).

Piyankarage, S. C., Featherstone, D. E. & Shippy, S. A. Nanoliter hemolymph sampling and analysis of individual adult Drosophila melanogaster. Anal. Chem. 84, 4460–4466 (2012).

Richard, D. S., Gilbert, M., Crum, B., Hollinshead, D. M., Schelble, S.,Scheswohl, D., Yolk protein endocytosis by oocytes in Drosophila melanogaster: immunofluorescent localization of clathrin, adaptin and the yolk protein receptor. J. Insect Physiol. 47, 715–723 (2001).

Righetti, P. G., Castagna, A., Antonioli, P. & Boschetti, E. Prefractionation techniques in proteome analysis: the mining tools of the third millennium. Electrophoresis 26, 297–319 (2005).

Roberts, D. B., Wolfe, J. & Akam, M. E. The development profiles of two major haemolymph proteins from drosophila melanogaster. J. Insect Physiol. 23, 871–878 (1977).

Roth, J., Zuber, C., Park, S., Jang, I., Lee, Y., Kysela, K. G., Le Fourn, V., Santimaria, R., Guhl, B., Cho, J. W., Protein N-glycosylation, protein folding, and protein quality control. Mol. Cells 30, 497–506 (2010).

Rubin, G. M., Hong, L., Brokstein, P., Evans-Holm, M., Frise, E., Stapleton, M., Harvey, D. A., A Drosophila Complementary DNA Resource. Science 287, 2222–2224 (2000).

Schafer, W. R. Genetic analysis of nicotinic signaling in worms and flies. J. Neurobiol. 53, 535–541 (2002).

Schilling, A. B., Crot, C., Helseth, D. L., Xu, H. & Davis, R. G. Conversion of an Agilent Chip Cube system for the analysis of proteomics samples using a LTQ-FT Ultra mass spectrometer. 58th ASMS Conf. (2010).

Schlenke, T. A, Morales, J., Govind, S. & Clark, A. G. Contrasting infection strategies in generalist and specialist wasp parasitoids of Drosophila melanogaster. PLoS Pathog. 3, 1486–1501 (2007).

Schmid, D. G., Grosche, P., Bandel, H. & Jung, G. FTICR-mass spectrometry for high-resolution analysis in combinatorial chemistry. Biotechnol. Bioeng. 71, 149–161 (2001).

Schütze, K., Niyaz, Y., Stich, M. & Buchstaller, A. Noncontact laser microdissection and catapulting for pure sample capture. Methods Cell Biol. 82, 649–673 (2007).

Schutzer, S. E.,Liu, T., Natelson, B. H., Angel, T. E., Schepmoes, A. A., Purvine, S. O., Hixson, K. K., Lipton, M. S., Camp, D. G., Coyle, P. K., Smith, R. D., Bergquist, J., Establishing the proteome of normal human cerebrospinal fluid. PLoS One 5, 123–124 (2010).

Scigelova, M. & Makarov, A. Orbitrap mass analyzer--overview and applications in proteomics. Pract. Proteomics 1, 16–21 (2006).

Shulman, J. M., Shulman, L. M., Weiner, W. J. & Feany, M. B. From fruit fly to bedside: translating lessons from Drosophila models of neurodegenerative disease. Curr. Opin. Neurol. 16, 443–449 (2003).

Siegel, G., Agranoff, B., Albers, R. W., Fisher, S. & Uhler, M. Basic Neurochemistry: Molecular, Cellular and Medical Aspects. 673 (Lippincott Williams & Wilkins, 1998).

Small, H., Stevens, T. S. & Bauman, W. C. Novel ion exchange chromatographic method using conductimetric detection. Anal. Chem. 47, 1801–1809 (1975).

Southan, C. Has the yo-yo stopped? An assessment of human protein-coding gene number. Proteomics 4, 1712–1726 (2004).

Takemori, N. & Yamamoto, M.T., Proteome mapping of the Drosophila melanogaster male reproductive system. Proteomics 9, 2484–2493 (2009).

Toennies, G. & Kolb, J. Techniques and Reagents for Paper Chromatography. Anal. Chem. 23, 4–7 (1950).

Tweedie, S., Ashburner, M., Falls, K., Leyland, P., McQuilton, P., Marygold, S., Millburn, G., Osumi-Sutherland, D., Schroeder, A., Seal, R., Zhang, H., FlyBase: enhancing

Drosophila Gene Ontology annotations. Nucleic Acids Res. 37, D555–D559 (2009).

Udenfriend, S., Stein, S., Böhlen, P., Dairman, W., Leimgruber, W., Weigele, M., Fluorescamine: a reagent for assay of amino acids, peptides, proteins, and primary amines in the picomole range. Science 178, 871–2 (1972).

Verleyen, P., Baggerman, G., D'Hertog, W., Vierstraete, E., Husson, S. J., Schoofs, L., Identification of new immune induced molecules in the haemolymph of Drosophila melanogaster by 2D-nanoLC MS/MS. J. Insect Physiol. 52, 379–388 (2006).

Vierstraete, E., Cerstiaens, A., Baggerman, G., Van den Bergh, G., De Loof, A., Schoofs, L. Proteomics in Drosophila melanogaster: first 2D database of larval hemolymph proteins. Biochem. Biophys. Res. Commun. 304, 831–838 (2003).

Vierstraete, E., Verleyen, P., Baggerman, G., D'Hertog, W., Van den Bergh, G., Arckens, L., De Loof, A., Schoofs, L., A proteomic approach for the analysis of instantly released wound and immune proteins in Drosophila melanogaster hemolymph. Proc. Natl. Acad. Sci. U. S. A. 101, 470–475 (2004).

Wang, Y., Teraoka, I., Hansen, F. Y., Peters, G. H. & Hassager, O. A Theoretical Study of the Separation Principle in Size Exclusion Chromatography. Macromolecules 43, 1651–1659 (2010).

Washburn, M. P., Wolters, D. & Yates, J. R. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat. Biotechnol. 19, 242–247 (2001).

Waterval, J. C. M. Review Derivatization trends in capillary electrophoresis : An update. Electrophoresis 23, 3922–3933 (2002).

Wilkins, M. R., Pasquali, C., Appel, R. D., Ou, K., Golaz, O., Sanchez, J., Yan, J. X., Gooley, A. A., Hughes, G., Humphery-Smith, I., Williams, K. L., Hochstrasser, D. F.,From Proteins to Proteomes: Large Scale Protein Identification by Two-Dimensional Electrophoresis and Arnino Acid Analysis. Bio/Technology 14, 61–65 (1996).

Wincek, T. & Sweat, F. The assay of adenyl cyclase by sequential dowex-alumina chromatography. Anal. Biochem. 64, 631–635 (1975).

Wolf, F. W. & Heberlein, U. Invertebrate models of drug abuse. J. Neurobiol. 54, 161–178 (2003).

Xun, Z., Sowell, A., Kaufman, T. C. & Clemmer, D. E. Protein Expression in a Drosophila Model of Parkinson ' s Disease research articles. J. Proteome Res. 348–357 (2007).

Yin, H. & Killeen, K. The fundamental aspects and applications of Agilent HPLC-Chip. J. Sep. Sci. 30, 1427–1434 (2007).

Yin, H., Killeen, K., Brennen, R., Sobek, D., Werlich, M., Van de Goor, T., Microfluidic chip for peptide analysis with an integrated HPLC column, sample enrichment column, and nanoelectrospray tip. Anal. Chem. 77, 527–533 (2005).

Zhu, J., Nie, S., Wu, J. & Lubman, D. M. Target proteomic profiling of frozen pancreatic CD24+ adenocarcinoma tissues by immuno-laser capture microdissection and nano-LC-MS/MS. J. Proteome Res. 12, 2791–804 (2013).

Zoghbi, H. Y. & Botas, J. Mouse and fly models of neurodegeneration. Trends Genet. 18, 463–471 (2002).

7. APPENDIX

Publication permission need to be added here.

8. VITA

NAME: Qi Zeng

EDUCATION:

- Ph.D., Analytical Chemistry, 2014, University of Illinois at Chicago.

Thesis title: Proteomic Analysis of Individual Drosophila Hemolymph

- B.Sc., Chemistry, 2009, Renmin University of China.

RESEARCH EXPERIENCE:

- Developed a conductive carbon coated fused silica capillary tip for electrospray ionization, built the interface of CE and MS.
- Developed liquid chromatography methods for small volume (nL) hemolymph protein analysis, using RPLC, SEC and IEX material.
- Developed HPLC-MS method for protein identification of single fruit fly hemolymph (blood) in nano-liter volume.
- Experience with LC, GC, MS softwares: Empower, Masshunter, Xcalibur, Data Series Explorer 4000, Scaffold, Mascot, Massmatrix, and LabView.
- Experience with HPLC (Waters, Agilent), GC-MS (Varian-Agilent), UV-vis spectrometer, Nano-HPLC (Ultimate 3000), MALDI-TOF MS/MS (ABI-4700), Fluorometer(Ocean Optics), LTQ-FTMS/MS (Thermo Finigan), Atomic absortion/ Atominc emission (Perkin-Elmer), Orbitrap LTQ MS/MS (Verlos Pro), SDS-PAGE (Bio-Rad), Capillary Electrophoresis (CE), FTIR.

PUBLICATIONS:

Proteomic analysis of individual fruit fly hemolymph. **Q. Zeng**, D. Smith, S. Shippy, J. Chromatogr. B. Submitted.

Prefractionation methods for individual adult fruit fly hemo-lymph proteomic analysis. **Q. Zeng**, V. Avilov, S. Shippy, Anal. Chem. Submitted.

AWARDS:

2013: UIC Provost's & Deiss awards for graduate research.

PRESENTATIONS:

2013: Nano-liter sampling technique applied on rat, mice and fruit fly. CBC symposium, Chicago

2012: Preptide and protein analysis of single fruit fly hemolymph sample. Pittcon Conference, Orlando.

TEACHING EXPERIENCE:

Teaching Assistant, 2009-2014

- UIC Undergraduate General Chemistry Courses 101, 112 and 114.
- Analytical Chemistry 222
- Inorganic Chemistry 314
- Advanced Instrumental Analysis 421.

PERSONAL AFFILIATIONS:

Membership in American Chemistry Society
Membership in Analytical Chemistry Division, American Chemistry Society
Membership in Society for Applied Spectroscopy