Prefractionation methods for individual adult fruit fly hemolymph proteomic analysis

Qi Zeng[†], Vitaly Avilov[†], Scott A. Shippy[*†‡]

[†]Department of Chemistry, University of Illinois at Chicago, 845 W Taylor St., SES 4500, M/C 111, Chicago, IL 60607

[‡]Laboratory of Integrative Neuroscience, University of Illinois at Chicago, 840 W Taylor ST., SEL 4311, M/C 067, Chicago, IL 60607, United States.

Email address: sshippy@uic.edu

Phone number: +1312-355-2426.

Abstract: The analysis of blood provides in depth chemical information of physiological states of organisms. Hemolymph (blood) is the fluid in the open circulatory system of *Drosophila melanogaster* that is the medium for molecules regulating a wide variety of physiological activities and signaling between tissues. Adult Drosophila is typically less than 3 mm in length and, as a consequence, the available volume of hemolymph is usually less than 50 nL from individual flies. Proteomic analysis of volume-limited hemolymph is a great challenge for both sample handling and subsequent mass spectrometry characterization of this chemically diverse biological fluid with a wide dynamic range of proteins in concentrations. Less abundant proteins, in particular, could be easily lost during sample preparation or missed by current mass spectrometry methods. This article describes simple and customized RPLC column and IEX columns

to prefractionate volume-limited hemolymph without excessive dilution. Step-gradient elution methods were developed and optimized to enhance the identification of novel proteins from an individual fruit fly hemolymph sample. Fractions from each step gradient was analyzed by an Agilent nano-RPLC chip column and then characterized by high mass resolution and high mass accuracy orbitrap mass spectrometry. As a result, both RPLC (11 proteins) and IEX fractionation approaches (9 proteins) identified more proteins than an unfractionated control approach with higher protein scores, emPAI values and coverage. Furthermore, a significant number of novel proteins were revealed by both RPLC and IEX fractionation methods, which were missed by unfractionated controls. The demonstration of this method establishes a means to deepen proteomic analysis to this commonly used, important biological model system.

Keywords: *Drosophila melanogaster*; reverse phase liquid chromatography; ion exchange chromatography; nanoliter samples; shotgun proteomics.

Abbreviations: RPLC, reverse phase liquid chromatography; HPLC, high performance liquid chromatograpyhy; IEX, ion exchange chromatography; BSA, bovine serum albumin; 2-DE, two-dimensional gel electrophoresis; MudPIT, multidimensional protein identification technology; ACN, acetonitrile; emPAI, exponentially modified protein abundance index; ATP/ADP, adenosine tri-/di-phosphate; GPI, glycosylphosphatidylinositol

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, fly gene correlates have been found for three-quarters of known human disease genes and two-thirds of the fly protein sequences have mammalian homologs. [1,2] The great similarity between fly and human genomes, the low-cost maintenance, the fast generation lifetimes and facile gene manipulation continues to drive many studies with the fly including into the central nervous system, metabolism pathways, and neurodegenerative diseases. [3–5] Hemolymph is a unique biological compartment that is distinct from the whole organism or isolated tissue homogenates. 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. [6] Therefore, the proteomic analysis of fruit fly hemolymph plays an important role in understanding actual protein content in distinct physiological events. [7–10] A novel single fruit fly hemolymph collection method developed previously [11,12] 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. Two-dimensional gel electrophoresis (2-DE) is the most widely adapted prefractionation method for simplification and reduction of the complexity of proteomic

mixtures such as hemolymph. [13] Vierstaete et al. reported the first 2-D database for larval hemolymph proteins, using 2-DE with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. [14] Kalsson et al. investigated larval hemolymph protein related to clotting, using the same set up of instrument as Vierstraete et al. [9] 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 spot overlap. [13] 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. [15] 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. [16] For instance, albumin takes up to 60% of the total proteins in serum [13] and with 2-DE this amount can overlap or cover 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 have led to prefractionation technique development, in order to simplify complex protein mixtures for more accurate and diverse spectra.

To reduce the measurement of 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 class using immobilized Protein A. [17] However, this immunodepletion approach could also lead to the loss of other species of interest which are structurally or functionally related to the depleted protein. Washburn et al. has pioneered and popularized the multidimensional LC fractionation technique, which is known as MudPIT (Multidimentional Protein Identification Technology). [18,19] 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 is 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.

Herein we introduced 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 improve our previously reported proteomic analysis of fly hemolymph [20] offering a rapid sample

prefractionation, simplifying the complex proteomic sample from adult fruit fly hemolymph, and eventually contributing to better protein identification and low abundance critical protein detection.

2. Experimental section

2.1 Chemicals

Deionized water was obtained from a US Filter Purelab Plus purification system (Lowell, MA). Trypsin, tris hydrochloride, dithiothreitol, iodoacetimide, trifluoroacetic acid and C18 beads were all purchased from Sigma-Aldrich (St. Louis, MO). Glacial acetic acid, fluorescamine and bovine serum albumin was purchased from Fisher Scientific (Itasca, IL). Trifluoroacetic acid was from Sigma-Aldrich (St. Louis, MO). All chemicals reagents were of analytical grade.

2.2 Hemolymph sample preparation.

Hemolymph was harvested from the Oregon-R strain of Drosophila melanogaster maintained in the Featherstone laboratory, Department of Biological Sciences at 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 previous work. [11,12] 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. [20] The < 3 nL volume of tryptic digested hemolymph samples were stored at -20 °C until analysis. All fractionation for proteomic analysis was performed on aliquots from a single fly.

2.3 RPLC fractionation.

Spherical silica C18 beads were aspirated into a 5-cm long tygon tubing by a vacuum created by a syringe connected to the tygon tubing. With C18 beads packed tightly in the 5-cm tygon tube, both ends of the tygon tube was 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 bovine serum albumin (BSA) was chosen as a standard to test the separation efficiency of this customized short column. A 500 nL aliquot of digested 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 flow rate. Three fractions of eluate were collected at each step of 0%, 10%, 20%, 30% and 70% acetonitrile (ACN) in deionized water with 0.1%TFA.

2.4 Strong cation exchange chromatography fractionation.

Finely meshed, hydrogen form Dowex 50WX4 200-400 resin from Alfa Aesar (Ward Hill, MA) was selected for IEX fractionation of tryptic hemolymph proteomic sample. The

column packing follows the same procedure as described for RPLC fractionation. Packed columns were thoroughly rinsed with deionized water, followed up with column equilibration using 20 mM pH 2.58 phosphate buffer. A 500 nL aliquot of digested 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.

2.5 Fluorescence detection.

To test and optimize prefractionation conditions, a fluorimeter, was used to measure the fluorescence intensity of each eluate fraction in a capillary cell. The modified ZetaLIF capillary fluorescence detector (Picometrics, Labège, France) uses a 365 nm diode for excitation and optics for emission centered at 490 nm. The pH of IEX eluate fractions were adjusted with 0.1 M NaOH to pH 7 for derivatization. To determine approximate pH in these small volumes where a pH probe can not be used, larger volumes of eluent buffer and NaOH were mixed and adjusted to appropriate pH. To adjust small volume fractions a proportional amount of NaOH was measured in a 20.6 mm/250 µm o.d. /i.d. tygon tubing and delivered into each IEX eluate fraction.

Each eluate fraction was then derivatized with primary amine reactive fluorescamine by addition of an equivalent volume of 1.5 mg/mL in acetone for 6 min prior to detection. A 10 cm length of (360/50 µm o.d. /i.d.) fused silica capillary was used as the detection

cell. Each eluate was injected into the capillary cell in 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.

2.6 LC-MS analysis.

A 500 nL aliquot of hemolymph was RPLC prefractionated and 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 previous work. [20]

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 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 and solvent A composed of 95% water / 5% ACN / 0.1% formic acid. The gradient began with 6% B (95% ACN, 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 column was set at a 60 min gradient ramping from 8% B to 45% B.

After separation 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, isolation width of 2.0 Da, a normalized collision energy of 35 eV, activation Q 0.25 and activation time of 30 ms. If the charge state was determined to be greater than +1, MS/MS was triggered with minimum signal required (2500). The dynamic exclusion list was restricted to 250 entries with duration of 60 s, 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. Mass spectrometry data was acquired and analyzed by Xcalibur software.

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. The absolute abundance of the proteins observed in our sample was found using the exponentially modified protein abundance index (emPAI) as has been described by others.[21]

3. Results and discussion

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 digest was used as a standard protein sample and prepared in the same way as the hemolymph sample to demonstrate sample handling and fractionation by this customized RPLC column. Figure 1 (A) shows the relative fluorescence signal from each fraction corresponding to relative amounts of peptides. The levels show that the standard tryptic BSA peptides were separated efficiently with this customize RPLC column. Notably, 20% ACN and 30% ACN eluted about 36% and 28% of 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 qualitative result shows that there is a separation of the BSA peptides. While the separation is certainly less efficient compared to commercial, high pressure C-18 columns, the columns here are hand-packed with stationary phase at relatively low pressure in <15 min for a significantly lower cost. Notably, each fraction is 2 µL, almost equal to two 0.9 µL column volumes. The baseline separation between

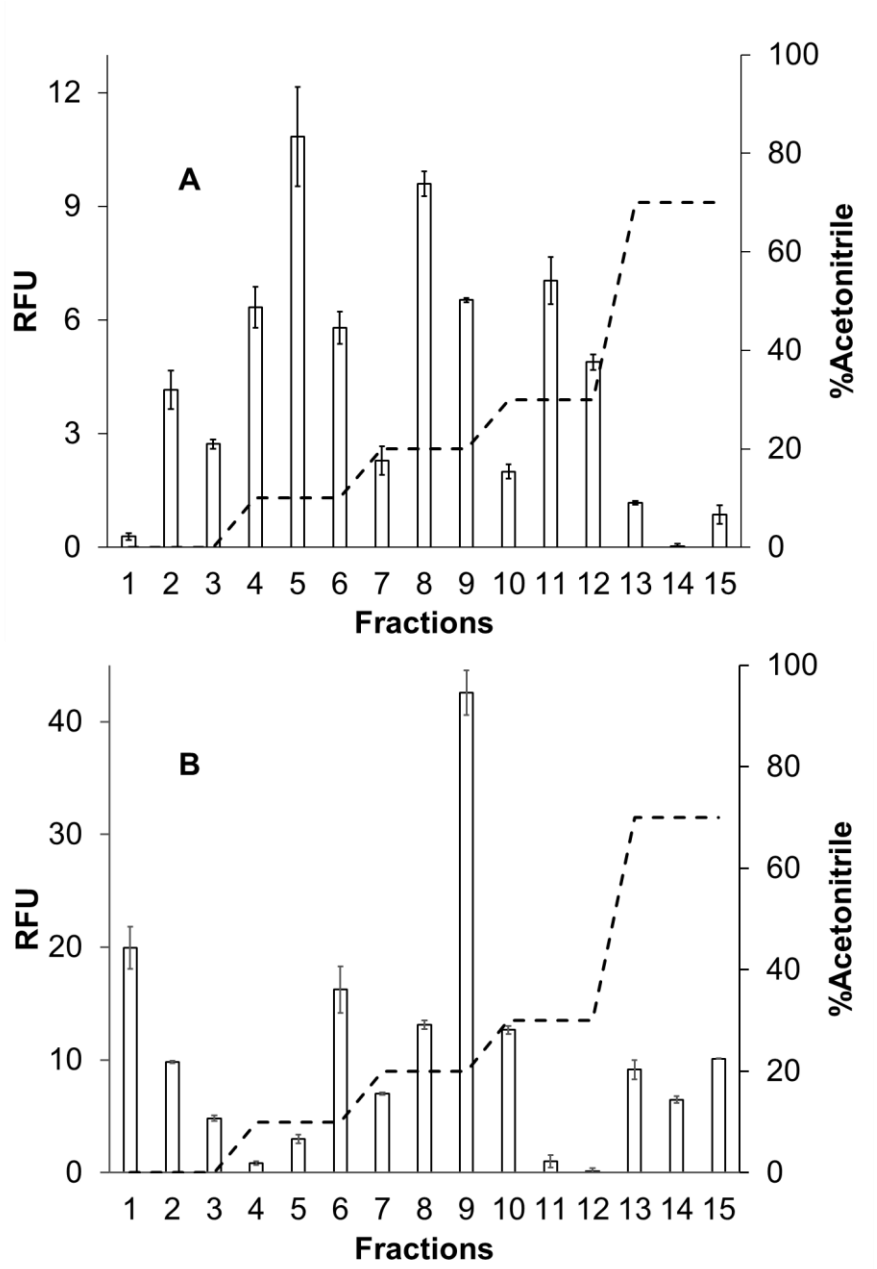

Figure 1. Fractionation chromatogram of tryptic peptides obtained from customized RPLC column. Relative fluorescence from primary amine and peptide content observed from each collected fraction is plotted vs. fraction number. (A) Chromatogram of standard tryptic BSA peptides fractionation using the customized 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 dionized 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 fluoroeter at 365nm excitation/490 nm emission.

eluent polarities suggests that there is both good partition of peptide analytes onto the stationary phase and selective elution with this step gradient elution.

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 1 (B), the fluorescence levels show hemolymph peptides were fractionated using the same column set up and elution step gradient. The chromatographic behavior, Figure 1 (B) shows that there is a good spread of peptide components over the gradient steps. While the fractionation is sufficient for further analysis, more optimization could be performed to fine tune hemolymph peptides of interest. For instance, a relatively larger group of peptides were eluted and collected in fraction 9, the last fraction collected with 30% ACN mobile phase. The difference from the BSA peptides separation could be attributed to the greater variation of protein components in hemolymph versus BSA, a single protein.

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 proven 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, we demonstrate a

strong cation exchange column fabricated for fractionation of a 500 nL aliquot of digested fruit fly hemolymph. Again, BSA tryptic peptides were used as a standard to show the fractionation capability of this short, in house constructed IEX column.

As illustrated in Figure 2 (A), the step gradient 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 increase 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 2 (B). By

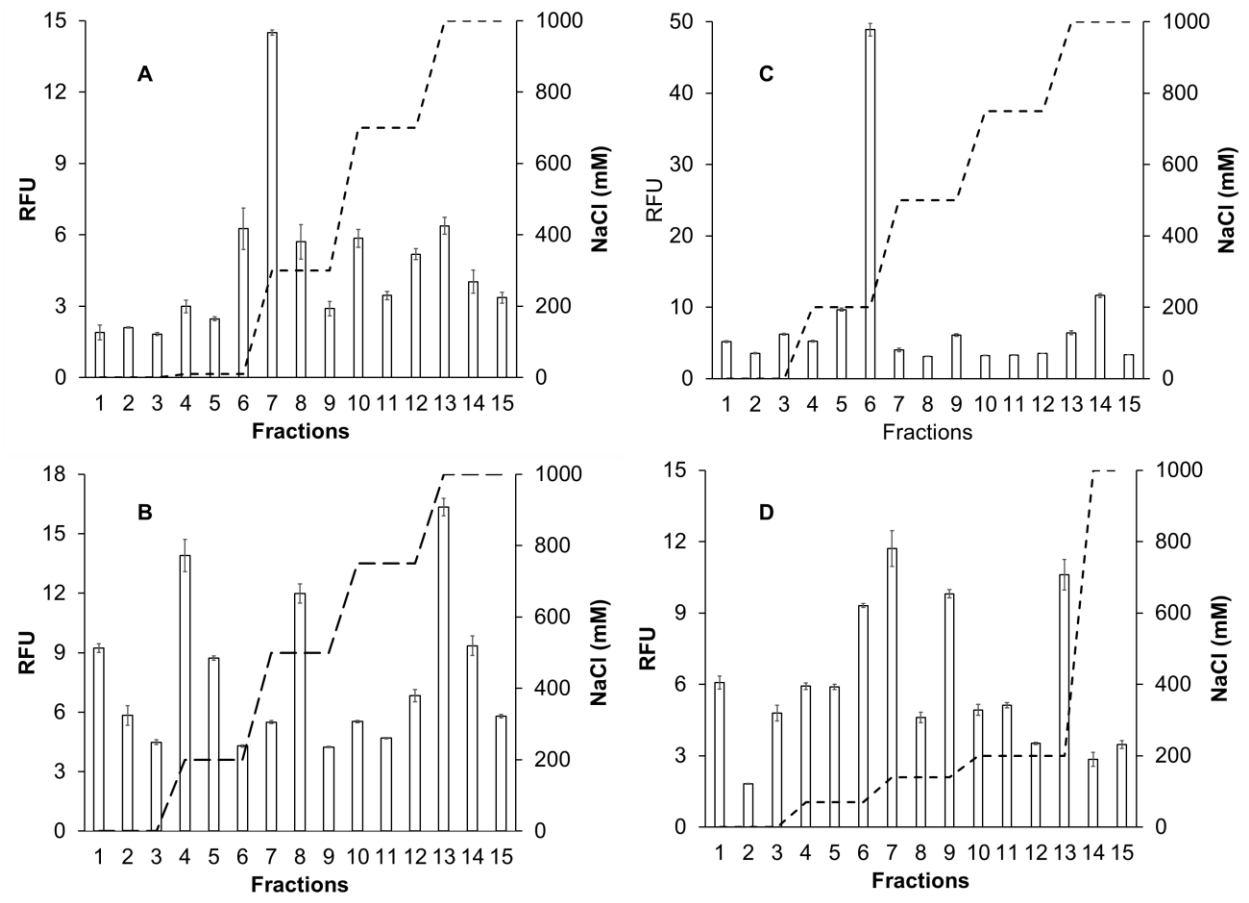

Figure 2. Optimization of separation condition for complex peptide mixture in IEX chromatography. The stepwise composition of eluent NaCl is noted as the dash line related to the values marker on the right side of the y-axis. Fluorescence signal corresponding to primary amine and peptide content is plotted vs. fraction number. (A) Fractionation of BSA tryptic peptides with step gradient set up as: 0 M-10 mM-300 mM-700 mM NaCl-1 M NaCl. The acute increment between 10 mM and 300 mM NaCl caused by 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 NaCl-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 M-70 mM-140 mM- 200 mM-1 M NaCl was used. Hemolymph peptides were fractionated and collected over several separate fractions.

adding 200 mM NaCl and 500 mM NaCl steps, the fluorescence observed from each fraction has been better resolved across multiple fractions, as seen in Figure 2 (B).

The improved step gradient elution was subsequently employed to fractionation of a fly hemolymph sample. From Figure 2 (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 in the 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 2 (D), by adding two eluent 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 the elution step gradient provided the fractionation of complex fruit fly hemolymph sample. The spread of peptides through multiple fractions will presumably simplify the peptide composition in each fraction, and therefore enhance the MS characterization capability for less abundant and/or novel protein identification in this limited volume of hemolymph sample.

To be noted, these fractionations are not as robust as separations performed with commercially prepared columns. The in house packed, low pressure operated, single-use columns presented here spread peptide components across the collected fractions for LC-MS analysis and display reasonable run-to-run reproducibility. While not strongly

reproducible, both RPLC and IEX prefractionation experiments presented here take less than 3 hours. The time and cost efficiency enables the proposed RPLC and IEX prefractionation methods to be adapted to any biological proteomic sample complex with ease. These prefractionation methods are offline from the mass spectrometer. The offline approach allows for both optimization of the fractionation for different samples and for increased efficiency in use of mass spectrometer compared to online fractionation arrangements.

3.3 MS results.

Aliquots from an individual fly hemolymph sample were assayed for protein content following the RPLC and IEX prefractionation as described above. For comparison a control sample was also assayed without prefractionation. Protein content was determined following a 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 protein scores, sequence coverage and the relative abundance emPAI value provide factors to evaluate the prefractionation efficiency. The distribution of proteins found in three methods was charted in the Venn diagram, as shown in Figure 3 to compare between fractionation methods. The enhancement from prefractionation is verified by comparison of proteins discovered by all methods and the identification of novel proteins in this study.

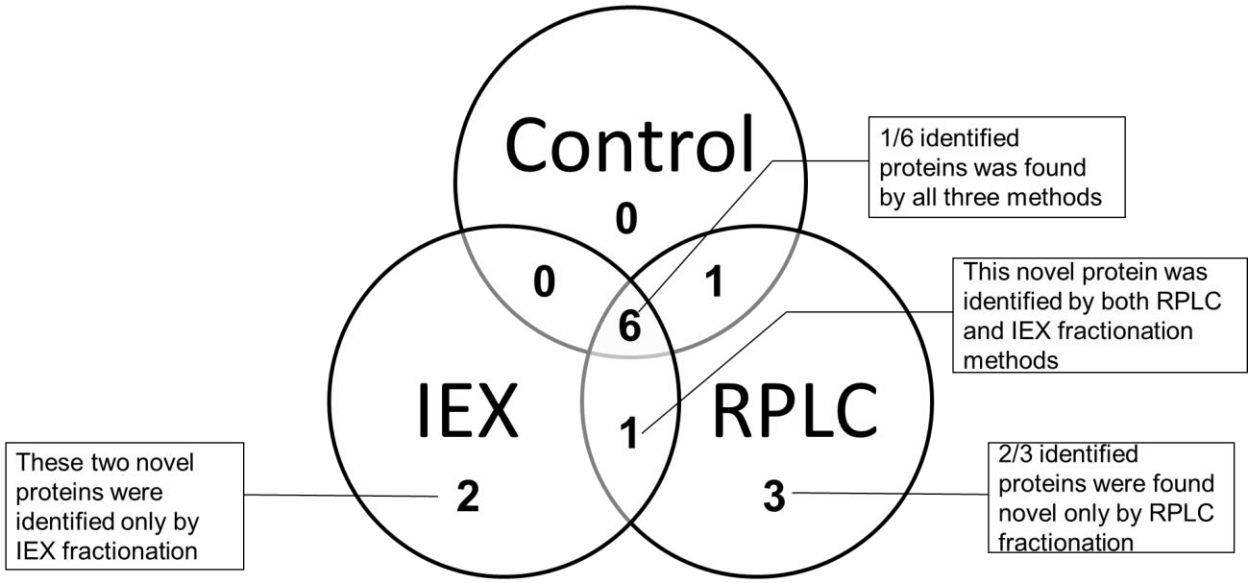

Figure 3. Venn diagram of protein identification to compare between fractionation methods.

3.4 MS characterization of proteins in hemolymph after fractionation.

A comparison of 6 proteins found by the two prefractionation methods and control was done in order to assess the efficacy of RPLC and IEX prefractionation in enhancing protein identification. Table 1 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 as scored with 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, [22,23] 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,

Table 1. Common proteins identifed by RPLC and IEX fractionation and control methods.

Protein name	Protein[a] (Mr)	Control (unfractionated)			RPLC fractionation			IEX fractionation		
		Score[b]	emPAI value[c]	Sequence coverage	Score[b]	emPAI value[c]	Sequence coverage	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%

[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.

neuropeptide-like precursor and transferrin precursor were all significantly improved as shown by the protein scores, emPAI and sequence coverage with either prefractionation method, compared to control method.

Shown in Figure 3 is the Venn diagram to compare protein identification between the RPLC, IEX fractionation methods and the control with 500 nL aliquots of a digested fly hemolymph sample. In total, ~2900 peptides are identified by each method; 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 chemically complex but, volume-limited hemolymph samples. From Figure 3, there are 6 proteins identified by all three methods showing good reproducibility of analysis for aliquots of this digested hemolymph from a single fly. The common 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 the 6 proteins found here. [14] This difference might be explained by the fact that the adult and larval flies likely express different hemolymph proteomes. In particular, the discovery of yolk proteins are not expected for larvae as this is related to reproductive activities that are not paramount for that stage of the lifecycle. 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. [24] The preponderance of these secreted proteins suggests that they are dominant in

the hemolymph sample. The ability to study secreted extracellular proteins is a particular feature of this method compared to tissue homogenates. Further, even though hemocyte tissue present (~10%) in hemolymph [14] and there are some hemocyte-derived proteins observed by individual sample preparations, there is not a major protein hit related to hemocytes observed by all sample preparation methods 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 of RH70154P and a gene prediction of ferritin 2 light chain homolog [24], 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. [24] Because ATP synthase occurs inside the mitochondria within cells, it appears that the observation of bellwether in hemolymph is originally from hemocyte component of our sample. Further study is needed to confirm this idea. The existence of Ferritin 2 light chain homolog is currently predicted from the sequence fly genome. In this study the identification of peptide fragments of this protein was able to prove the existence of this Ferritin 2 light chain and 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 to be associated with hemolymph coagulation in the fly. [25] The discovery

of RH70154P is important because the molecular function is not known. With the ability to measure this protein, studies can now be designed to determine its physiological roles.

There are two novel proteins that were identified for the first time by IEX fractionation. They are CG6409 and Imaginal disc growth factor 4. CG6409 was discovered with a Mascot protein score of 49 and 5% of sequence coverage. This is a protein with evidence at the mRNA transcript level and is predicted to be involved in glycosylphosphatidylinositol (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. [24] Imaginal disc growth factor 4 was also found as a protein here for the first time with a Mascot protein score of 52 and sequence coverage of 4%. The protein was identified to be secreted in hemolymph [24] and is probably transported to target tissues via hemolymph. [26] Imaginal disc growth factor 4 is reported to be primarily expressed in yolk cells and the fat body. While it is expressed throughout development stages, it is expected to have higher expression during larval stages. [26] Previous literature reported the identification of imaginal disc growth factor 2 in larvae. The imaginal disc growth factor 2 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 disk 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. [26] Further investigation is needed for confirmation of this hypothesis.

3.5 Effect of separation on prefractionated 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 a modification of the LC separation step prior to MS analysis could also provide an enhancement in protein identification. A traditional nano LC column was substituted for the nano-HPLC chip system separation in the separation of another 500 nL aliquot of 60-fold digested and RPLC fractionated hemolymph sample. The protein identification data from this nano LC column system is compared to the results with the nano-HPLC chip separation of the RPLC fractionated sample described earlier.

Listed in Table 2 are proteins found by nano-HPLC chip system and a nano LC column. Proteins identified are listed according to the eluent gradient fraction. Additionally, because individually analyzed fractions contain partial fragments from an intact protein, the MS identification could be underestimated due to insufficient fragments in the same sample fraction from the same protein passing the threshold protein score. The MS results from all fractions of a single hemolymph aliquot were merged to determine any

Table 2. Comparison of proteins identified in various fractions by the nano-HPLC chip system and nanoLC column.

	Fraction 1 (0% ACN)	Fraction 2 (10% ACN)	Fraction 3 (20% ACN)	Fraction 4 (30% ACN)	Fraction 5 (70% ACN)	Identified proteins from merged fractions
Chip system	bellwether Neuropeptide-like presursor2	yolk protein 1 Neuropeptide-like presursor2 yolk protein 2 Retinoid-and fatty acid-binding glycoprotein	yolk protein 1 Neuropeptide-like presursor2 Retinoid-and fatty acid-binding glycoprotein yolk protein 2	yolk protein 1 Lp06392p yolk protein 2 Neuropeptide-like presursor2 Transferrin precursor	NONE	1. Bellwether 2. Retinoid-and fatty acid-binding glycoprotein 3. Neuropeptide-like presursor2 4. Yolk protein 1 5. Ferritin 2 light chain homolog† 6. TA01815P* 7. Yolk protein 2 8. Transferrin precursor* 9. Ferritin 1 heavy chain homolog 10. RH70154P* 11. LP06392P*
	Fraction 1 (0% ACN)	**Fraction 2 (10% ACN)**	**Fraction 3 (20% ACN)**	**Fraction 4 (30% ACN)**	**Fraction 5 (70% ACN)**	**Identified proteins from merged fractions**
NanoLC column	Neuropeptide-like precursor 2	ATP synthase beta subunit Bellwether yolk protein 2 Retinoid-and fatty acid-binding glycoprotein SkpA associated protein, isoform A	ATP synthase beta subunit Aldolase alpha Yolk protein 3 yolk protein 2 TA 01815p Glyceraldehyde-3-phosphate dehydrogenase (Gadph-1) protein(EC 1.2.1.12) Bellwether Cu-Zn superoxide dismutase Pyruvate kinase Pyruvate kinase, isoform B	ATP synthase beta subunit Alcohol dehydrogenase Yolk protein 1 Bellwether LP03138p	NONE	1. ATP synthase beta subunit* 2. TA 01815p* 3. Aldolase, isoform E 4. Yolk protein 3 5. Glyceraldehyde-3-phosphate dehydrogenase (Gradph-1) protein (EC 1.2.1.12)* 6. Neuropeptide-like precursor 2 7. Cu-Zn superoxide dismutase 8. Retinoid-and fatty acids-binding glycoprotein 9. Yolk protein 1 10. SkpA associated protein, isoform A* 11. Bellwether 12. Yolk protein 2 13. Aldolase alpha 14. Pyruvate kinase* 15. Pyruvate kinase isoform B* 16. LP03138p 17. Alcohol

*Proteins identified with previous evidence at the transcript level. †Protein identified as predicted by genomic analysis.

increase the protein identification of the combined LC-MS/MS approach. [16] The proteins identified from fraction file merging are also listed in Table 2. 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 and there are 6 proteins identified by both separation systems. However, there are also significant differences between these two separation systems. As shown in the last column in Table 2, by merging fraction files from 1 to 5, the nanoLC column led to the identification of 18 proteins, while the nano-HPLC chip system identified 11 proteins. In addition, the nano LC column helped to characterize 7 novel proteins with evidence at the transcription level in the UniprotKB database, while the nano-HPLC chip system has only found 5 novel proteins. These differences could be attributed to the configuration of the separation column and analysis time. While the length of the columns are both 150 mm and the stationary phase material is similar, the chip-based versus capillary column nature of the separation may contribute some of the observed differences. More clearly, the nano LC separation was twice as long as the nano-HPLC Chip at 60 min. This extended separation with the nano LC appears to be a main factor affecting protein identification. Although a direct comparison between equivalent separation systems isolate separation time effects, these results suggests that further fractionation of this digested 50 nL sample from a single fly could lead to an even deeper probing of the hemolymph proteome.

4. Conclusions

In summary, we presented a method using a customized, low pressure, RPLC column and IEX column for prefractionation of aliquots of digested hemolyph from an individual

adult female fruit fly for proteomic analysis. These short columns appear well suited for prefractionation of limited-volume hemolymph samples and led to a consistent identification from separate aliquots with an unfractionated control. The prefractionation also achieved identification of twice as many, presumably less abundant, proteins. The ability to access a deeper analysis of the hemolymph proteome with simple and inexpensive RPLC and IEX fractionation will allow more facile proteomic studies of this important animal model system. The ability to further enhance the proteins identification with a longer separation condition suggests even deeper analysis is possible. While it may be considered a relatively low number of protein hits of 23 proteins across all sample preparation methods, an exceedingly low 50 nL sample volume from an individual fly all analyses. Pooling may be used to increase protein hits at a cost of losing individual fly characteristics. The analysis of hemolymph as distinguished from other potential fly tissues provided data regarding the extracellular protein content with few proteins related to hemocytes or whole tissues. Potentially, this simplified prefractionation method followed with the nano-HPLC chip system or a capillary-based nano LC 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.

5. Acknowledgements

This work was supported in part by a UIC Provost Award for Graduate Research (QZ). We would like to thank the Research Resources Center/ Mass spectrometry,

Metabolomics and Proteomics Facility at UIC for access to Orbitrap Velos Pro and for bioinformatics software. The facility was established in part by a grant from The Searle Funds at the Chicago Community Trust to the Chicago Biomedical Consortium. We also thank Dr. David E. Featherstone for use of the dissection microscope and all fruit fly stocks used in this work.

6. References

[1] G.M. Rubin, A Drosophila Complementary DNA Resource, Science. 287 (2000) 2222–2224.

[2] M.D. Adams, The Genome Sequence of Drosophila melanogaster, Science. 287 (2000) 2185–2195.

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

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

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

[6] S.S. Grewal, Controlling animal growth and body size - does fruit fly physiology point the way?, F1000 Biol. Rep. 4 (2012) 12.

[7] R.R.H. Anholt, T.I. Williams, The soluble proteome of the Drosophila antenna.,

Chem. Senses. 35 (2010) 21–30.

[8] E. Vierstraete, P. Verleyen, G. Baggerman, W. D'Hertog, G. Van den Bergh, L. Arckens, et al., 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 (2004) 470–475.

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

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

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

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

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

[14] E. Vierstraete, A. Cerstiaens, G. Baggerman, G. Van den Bergh, A. De Loof, L. Schoofs, Proteomics in Drosophila melanogaster: first 2D database of larval

hemolymph proteins., Biochem. Biophys. Res. Commun. 304 (2003) 831–838.

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

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

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

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

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

[20] Q. Zeng, D.J. Smith, S.A. Shippy, Proteomic analysis of individual fruit fly hemolymph., J. Chromatogr. B. Analyt. Technol. Biomed. Life Sci. 981-982 (2015) 33–39.

[21] Y. Ishihama, Y. Oda, T. Tabata, T. Sato, T. Nagasu, J. Rappsilber, et al.,

Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein., Mol. Cell. Proteomics. 4 (2005) 1265–1272.

[22] M. Bownes, B.A. Hodson, Mutant fs(1) 1163 of Drosophila melanogaster alters yolk protein secretion from the fat body, Mol. Gen. Genet. 180 (1980) 411–418.

[23] M. Bownes, B.D. Hames, Genetic analysis of vitellogenesis in Drosophila melanogaster: the identification of a temperature-sensitive mutation affecting one of the yolk proteins., J. Embryol. Exp. Morphol. 47 (1978) 111–120.

[24] T.U. Consortium, Update on activities at the Universal Protein Resource (UniProt) in 2013., Nucleic Acids Res. 41 (2013) D43–47.

[25] S. Tweedie, M. Ashburner, K. Falls, P. Leyland, P. McQuilton, S. Marygold, et al., FlyBase: enhancing Drosophila Gene Ontology annotations., Nucleic Acids Res. 37 (2009) D555–559.

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