Brigham Young University
BYU ScholarsArchive

All Theses and Dissertations

2007-12-14

Study of the Reproducibility of Proteomics Methods and Variability of Fruit Fly Proteomes.

Thomas Franklin Culwell
Brigham Young University - Provo

Follow this and additional works at: https://scholarsarchive.byu.edu/etd

 Part of the Biochemistry Commons, and the Chemistry Commons

BYU ScholarsArchive Citation

Culwell, Thomas Franklin, "Study of the Reproducibility of Proteomics Methods and Variability of Fruit Fly Proteomes." (2007). *All Theses and Dissertations*. 1232.
https://scholarsarchive.byu.edu/etd/1232

This Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in All Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

STUDY OF THE REPRODUCIBILITY OF PROTEOMICS METHODS AND VARIABILITY OF FRUIT FLY PROTEOMES

by

Thomas Culwell

A thesis submitted to the faculty of

Brigham Young University

in partial fulfillment of the requirements for the degree of

Master of Science

Department of Chemistry and Biochemistry

Brigham Young University

April 2008

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

of a thesis submitted by

Thomas Culwell

This thesis has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

Date	Steven Graves, Chair
Date	Craig Thulin
Date	Allen Buskirk
Date	Duane Jeffery

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the thesis of Thomas Culwell in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

Date	Steven W. Graves Chair, Graduate Committee

Accepted for the Department

Date	David V. Dearden Graduate Coordinator

Accepted for the College

Date	Thomas W. Sederberg, Associate Dean College of Physical and Mathematical Sciences

ABSTRACT

Study of the Reproducibility of Proteomics Methods
And Variability of Fruit Fly Proteomes

Thomas Culwell

Department of Chemistry and Biochemistry

Master of Science

The reliability of biomarker discovery by means of proteomics has been called into question. It was speculated that "background noise" variation resulting from differences in preparation and handling of samples and proteome dynamics may mask subtle, yet important, differences due to the biological condition. Little is understood about complex proteomes and their variability. A critical aspect of proteomic biomarker research that is largely unexplored is the comparative reproducibility of certain methods such as two-dimensional gel electrophoresis and liquid chromatography/mass spectrometry. In particular, with liquid chromatography/mass spectrometry, it is not known whether variability in peptide quantitation is dependent on any of their several

properties such as size, abundance, or hydrophobicity. Such determinations may be critical in properly assessing the value of proteomics data.

The fruit fly *Drosophila melanogaster* was used as a well-controlled multicellular animal model to study the relationship between the background variation and expected changes induced by environmental or genetic factors. The data, gathered by two different proteomics methods, were used to compare and evaluate the reproducibility of the methods. It is reported that there was on average 15 to 18% variability in quantitative measurements of protein abundance using 2-dimensional gel electrophoresis or liquid chromatography/mass spectrometry. Using liquid chromatography/mass spectrometry, peptides with a smaller mass-to-charge ratio were shown to be measured less reproducibly than peptides with a larger ratio. Statistically significant proteomic differences between fly populations could be demonstrated between males and females. In dynamic experiments, less than 0.5% of proteins measured were shown to change after 24 hour starvation of the flies. However, no significant difference in peptide composition could be found for flies fed on a second diet consisting of the standard diet augmented with 10% ethanol.

These results suggest that proteomic variability while evident allowed for biomarker discovery using either method for this model system.

ACKNOWLEDGEMENTS

I sincerely thank Karen Merrell, Zhaoyuan Tony Chen, Chun Wan Jeffrey Lai, Tyler Rencher, Doug Tanner, Jacob Crandall, Yibing Xu, Jason Crellin, Peter Shen, Sarah Warburton, Peter McLean, Brad Reid, David Almond, Nikki Tanabe, Sean Phipps, Landon Costley, Brian Bahr, Bruce Jackson, Janet Fonoimoana, Daniel Simmons, Barry Willardson, and Greg Burton for teaching me and helping me with various responsibilities. I especially thank my committee, consisting of Duane Jeffery, Allen Buskirk, Craig Thulin and Steven Graves, for sharing much of their time and experience when I needed it most. And I thank my family for their love and encouragement.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

2GDE	two-dimensional gel electrophoresis
ACN	acetonitrile
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate
CID	collision induced dissociation
cLC	capillary liquid chromatography
CV	coefficient of variation
DTT	dithiothreitol
IEF	isoelectric focusing
IPG	immobilized pH gradient
LC	liquid chromatography
MS	mass spectrometry
oMALDI	orthogonal matrix assisted laser desorption ionization
PMSF	phenylmethylsulfonyl fluoride
RSD	relative standard deviation
SD	standard deviation
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SELDI	surface enhanced laser desorption ionization
TOF	time-of-flight

CHAPTER 1

INTRODUCTION

Proteomics

Proteomics is the study of diverse collections of proteins found in biological systems, typically fluids, cell fractions, and tissues. In this post-human-genome era, proteomics has become increasingly studied for many of the same reasons genomics was studied, including the potential for high-throughput data gathering and the macroscopic perspective it provides. Proteomics is ambitious and pioneering; it holds great potential for scientific advancement. Whereas most questions in biochemistry are answered through the reduction of systems to progressively smaller components, the big picture provided through proteomics research may lead to a more thorough study of the most relevant proteins, as well as help define interactions and functions of individual proteins in an efficient manner.

Separation and identification of proteins are at the core of proteomics research. Electrophoresis and column chromatography are the most common methods of separating proteins and will be discussed in some detail. Mass spectrometry with subsequent peptide mass fingerprinting is currently the only practical, high-throughput strategy for protein identification and will also be described. The focus of this paper is the evaluation of proteomics as applied to the field of biomarker discovery, although a proteomics approach may be appropriate for other complex problems as well, such as mapping protein interaction networks and characterizing post-translational protein modifications.

Two-Dimensional (2D) Gel Electrophoresis with Subsequent Mass Spectrometric Analysis

2D gel electrophoresis (2DGE) for the separation of complex mixtures of proteins is one of the earliest techniques used in proteomics [1]. In this approach, the proteins are first focused to their respective isoelectric points (pI) along a thin strip of polyacrylamide gel containing a pH gradient. Proteins are denatured with a non-ionic or zwitterionic detergent and urea to improve electrophoretic mobility and limit differences in charge. Ampholytes with a characteristic pKa range are used in the gel to buffer regions at various pH's. A high voltage is applied, and the proteins with any charge migrate toward the electrode of opposite charge. Because pH changes with temperature, the strip is actively cooled as the current heats the gel. As the proteins move through the gradient, they encounter ampholytes that maintain a pH ever closer to their respective pI. The charges on the proteins decrease until eventually the proteins are neutral and are focused at the point where the electric field has no influence.

Whereas ampholytes in solution originally were used to form the pH gradient, immobilized pH gradients (IPG) have been developed for higher resolution and better reproducibility [2]. In IPG, ampholytes are covalently linked to acrylamide and bisacrylamide monomers. Two acrylamide solutions with variable ampholytes, one at each end of the pH range, are poured on plastic strips in a gradient. Rather than depending on a limited number of ampholytes to create a finite number of zones of different pH, a truly linear gradient can be achieved. Also, because the ampholytes are stationary, there is less shift in the gradient between gels. Not all proteins in a sample are

resolved in the first dimension because the pI of some proteins may fall outside the pH boundaries of the IPG strip, and multiple proteins may have nearly the same pI.

This isoelectric focusing is followed by a second electrophoretic separation in an orthogonal dimension based on molecular weight. As is performed in one-dimensional SDS-PAGE, the proteins are given a net negative charge with SDS with all proteins having relatively comparable mass-to-charge ratio. The proteins are run out of the IPG strip and into a gel of higher acrylamide concentration. Smaller proteins have greater mobility and approach the anode more swiftly than the larger proteins. Polyacrylamide gradients may also be employed, with greater polymerization and reduced porosity near the bottom of the gel to more completely separate very large and small proteins in the same gel.

Proteins are stained to allow them to be visualized within the gel. The limits of detection are a drawback to traditional 2DGE. Various stains are employed to detect proteins in the gel. Silver staining was frequently used in early proteomics research because of its high sensitivity, but has since declined in popularity. Fluorescent stains, such Sypro Ruby, are popular in proteomics because they increase detection sensitivity to nanogram levels, have greater dynamic range [3] and achieve lower background because they stain protein more specifically than silver stains [4]. Coomassie Blue stain is popular in many applications, including proteomics, because it is inexpensive and the procedure is simple. This stain adsorbs to the basic amino acids as well as to aromatic residues [5]. Coomassie Blue stain is also nearly as sensitive as fluorescent stains when imaged with an infrared fluorescence imaging system rather than by standard densitometry [4].

Difference Gel Electrophoresis

Difference Gel Electrophoresis (DIGE) is a technique which focuses proteins from two different specimens, each set tagged with a different fluorophore, on a single gel [6]. This technique allows for superlative matching of protein spots while retaining a satisfactory level of sensitivity [7]. Two or three mixtures of proteins are combined with different cyanine dyes, which add only negligible mass and do not alter the isoelectric point. These mixtures are combined and the proteins are resolved together in both dimensions. Each fluorophore emits at a different wavelength when excited by which means the amount of each protein coming from a particular sample can be distinguished.

Enzymatic Digestion

Proteins and peptides separated by the above electrophoretic techniques are often proteolytically digested prior to mass spectrometry for a number of reasons. Proteins located as spots in a gel must be digested if they are to be fully recovered from the gel. Once digested, the fragments are able to diffuse more efficiently out of the gel, whereupon more information about the protein may be attained [8]. Trypsin is the primary enzyme used for in-gel digestion. Trypsin digestion also facilitates identification of proteins using mass spectrometry, which is discussed below. Entire proteins are too large to be sequenced manually, but fragments of 10 or 15 amino acid residues can be sequenced. Also, smaller peptides are more likely to fall within the range of detection of the mass spectrometer because many mass spectrometers only accommodate particles

with m/z values less than 3000. Proteolytic digestion also allows multiple peptides per protein to be analyzed for higher confidence in identification. Most importantly, because trypsin cleaves in a predictable fashion, data acquired from the digested fragments can be matched to masses of theoretical tryptic digests of predicted proteins in genome databases[9]. This process is known as peptide mass fingerprinting and is indispensible for high-throughput protein identification. Peptide mass fingerprinting is discussed in greater detail later in this chapter.

Liquid Chromatography (LC)

Proteins and peptides can be separated based on size, hydrophobicity, or charge by passing a mixture through a column which retains the molecules variably according to the selective chemical or physical property. In the case of charge or hydrophobicity, a buffer gradient can be used to control the rate with which the proteins elute off the column. This is achieved as the solvent is increasingly able to compete for binding sites on the solid phase, or more closely matches the polarity of the bound molecules. High performance liquid chromatography (HPLC) and capillary nano-LC are popular in proteomics because of the small sample size requirements and high speed. More than one chromatographic separation can be used for more complete separation of peptides and proteins [10] and optimized methods have shown to be highly reproducible [11]. Often, two chromatographic steps, exploiting different chemical properties, are used sequentially to more effectively fractionate mixtures. For example, several fractions from a cation exchange can be run separately on a reversed-phase column. These

chromatography techniques are generally used in conjunction with mass spectrometry, often on-line, through electro- or nano-spray ionization interfaces.

Mass Spectrometry (MS)

Mass spectrometry (MS) is a powerful and versatile technique in proteomics. Mass spectrometers detect the mass-to-charge ratio (m/z) of molecular ions based on their inertia in an electric field. Where the charge state of the ion is known, the mass can be obtained with high precision, several orders of magnitude below 1 amu. The most precise mass analyzers take several minutes to read a m/z while others take only a fraction of a second. Large ions can be fragmented inside the instrument, and the masses of the "daughter ions" can be analyzed to provide more information about its original structure. The high precision of mass measurement available through MS makes possible the solution of some of the most formidable challenges of proteomics, such as cataloging post-transcriptional modifications.

Soft Ionization Methods

Mass spectrometry of small molecules has existed since the early 1900's, but large molecules cannot withstand the ionization methods used for small molecules, such as electron impact and fast atom bombardment. Newer methods were developed for ionization of large molecules such as proteins. John Fenn and Koichi Tanaka were awarded the Nobel Prize in Chemistry in 2002 for distinct contributions in the development of soft ionization methods for MS of proteins and peptides. Fenn's

contribution, electrospray ionization (ESI)[12], involves spraying an acidified sample through a positively charged capillary needle at elevated temperatures. As solvent molecules evaporate from the fine aerosol droplets, the protons are left with the biological molecules, and these positively charged molecules are repelled by the needle and drawn to the electric field of the mass analyzer. Uncharged species and solvent are swept away by a curtain gas. Large peptides and proteins can be protonated at several residues and may bear many charges.

Tanaka was the first to modify laser desorption techniques for the ionization of proteins. However, the laser ionization method widely used in proteomics, matrix-assisted laser desorption ionization (MALDI), was conceived in 1987 by Karas and Hillencamp [13]. In MALDI a microliter quantity of the biological sample is co-crystallized on a metal plate along with an organic matrix. Some of the most common matrices used for protein ionization are alpha-cyano-4-hydroxycinnamic acid (CHCA), sinapinic acid, and 2,5-dihydroxybenzoic acid (DHB). A nitrogen UV laser is used to vaporize the sample. The matrix serves to absorb most of the energy, preventing the destruction of the biomolecule, and imparts a positive charge to the biomolecule. The bonds within the protein or peptide remain intact, and the resulting ions are almost exclusively charged +1.

Surface-enhanced laser desorption ionization (SELDI) is a variation of MALDI in which a biomolecule or targeted species is bound to the ionization surface based on chemical affinity of the surface for a given type of functional group or epitope [14]. This can result in greater selectivity. The limit of detection for SELDI is around 30 μg/mL,

and approaches that of some ELISA assays [15]. SELDI holds promise for clinical analysis of serum biomarkers because the influence of irrelevant proteins can often be reduced. MALDI and SELDI are also amenable to high-throughput applications.

As shown later in this chapter, the limit of detection may not be sufficiently low for analysis of some important proteins or peptides, even using SELDI. Additionally, because nearly all of the ions produced by laser desorption are singly charged, the dynamic range of the mass analyzer may severely limit which species can be detected. ESI often has the advantage of multiple charges per ion, such that large proteins beyond the limits of detection in a +1 charge state may be seen in the spectra as +5, +10 or greater charged ions. On the other hand, particularly if separation processes are not adequate, multiple charge envelopes for each protein or peptide may overlap and complicate the spectra, resulting in difficult interpretation.

Mass Analyzers

Quadrupole Mass Detectors

Quadrupoles consist of four parallel rods surrounding a cylindrical space through which ions are propelled. Two poles opposite each other hold a constant charge while the other two carry an alternating current. The electric field produced by the poles can be manipulated to lend a stable trajectory to ions of a particular m/z while other ions are expelled into the surrounding vacuum. Those ions which spiral through the quadrupole are detected directly or channeled into another mass analyzer.

Time-of-Flight (TOF) Mass Detectors

TOF mass analyzers are simpler in principle. Ions are pulsed by a like charge through a tube, and detected at the end. The magnitude of the electric field, the length of the path, and the time of flight are used to obtain the m/z. An ion mirror, or reflectron, is typically used to dramatically improve precision and resolution. It compensates for copies of the same ion having slightly different kinetic energies after the charge pulse. The path of the ions is redirected at an acute angle, focusing the signal as well as increasing the path length, as shown in Figure 1.1 [16]. Multiple mass analyzers are often used in tandem as well as in combination with methods to disrupt peptide bonds, as discussed below.

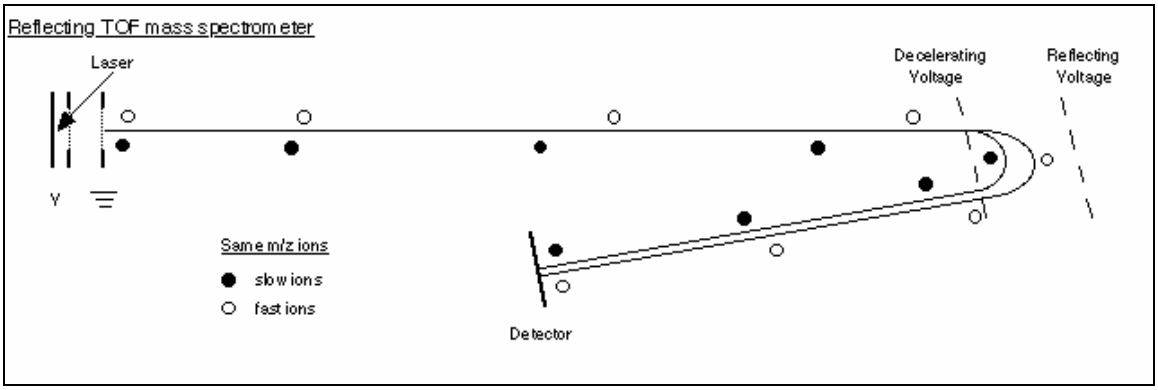


Figure 1.1. Reflecting mass spectrometer diagram. Ions with the same m/z have slightly different kinetic energies immediately after ionization and pulsing. In a reflectron TOF, the ions with higher kinetic energy travel a longer path than those of lower kinetic energy, which causes the ions of the same m/z to arrive at the detector at the same time. The resolution is thus improved. Modified from Lennon [16].

Ion Trap Mass Detectors

Linear and orbital ion traps are new tools which can analyze m/z to high precision but do not allow monitoring of a continuous flow of ions. Ion traps are ideal for many applications in which the volume of sample is limited because very little protein is required. The principles employed in ion traps are similar to those of quadrupoles. However, the ions are sealed in a chamber which rebounds the ions back and forth, or around an orbital path. Ions can selectively be expelled from the chamber based on m/z parameters. This chamber can also serve as a collision cell for fragmentation methods which are discussed below. The unparalleled sensitivity available results in a theoretically unlimited number of tandem MS cycles, which are also discussed below.

Some TOF systems are comparatively inexpensive, but accurate instruments equipped with an ion mirror do not measure ions with as large a m/z as other mass analyzers. Quadrupoles are versatile in that they can be used for multiple reaction monitoring and other more quantitative techniques. Ion traps are ideal for most applications, but cannot monitor separation column output with high time resolution and are much more expensive.

Tandem Mass Spectrometry Using Collision-Induced Dissociation or Electron Transfer Dissociation

Peptide bonds are typically more labile than other covalent bonds within individual amino acids, so a very useful pattern of fragmentation can be generated through selective disruption of these bonds. Ions can be directed into a chamber containing an inert gas, such as argon, at a low pressure. As the ions collide with the gas, covalent bonds are broken, and the mass of those fragments which retain a positive charge can be measured. The process of selecting an ion having a particular m/z and analyzing the spectral patterns of its fragment daughter ions is termed tandem MS, or MS/MS, and the fragmentation technique is known as Collision-Induced Dissociation (CID). Figure 1.2 [17] diagrams how CID is accomplished in a triple quadrupole and in an ion trap. Electron Transfer Dissociation (ETD) is a similar fragmentation technique, but utilizes a free radical anion gas and is able to target different bonds than CID.

In the present study, MALDI and ESI were used as ionization methods. A Q-TOF (triple quadrupole mass analyzer followed by a reflectron TOF) were used for all of the MS and MS/MS work. For MALDI MS/MS, the Q2 region (second quadrupole) of the triple quadrupole serves as a CID chamber. Target ions are selected manually. During LC-MS/MS, Q2 serves as a collision cell for 3-second intervals, with a program feature of the mass spectrometer, information-dependent acquisition (IDA), selecting peaks automatically for CID. These techniques have been in use for almost ten years.

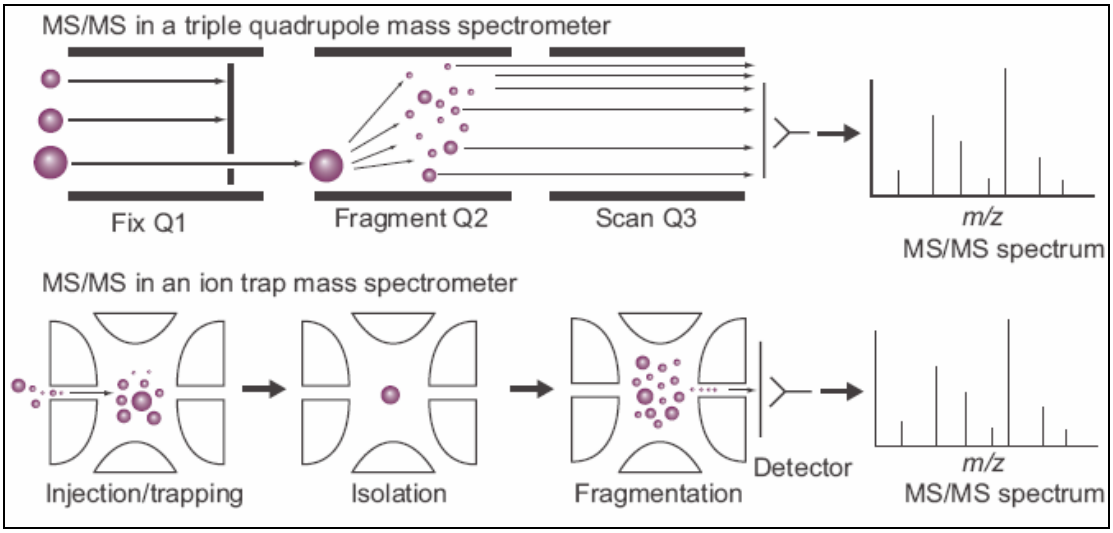


Figure 1.2. Tandem MS in a triple quadrupole or ion trap mass spectrometer using CID. In a triple quadrupole (above), the first quadrupole (Q1) is used to exclude ions outside a narrow range of m/z. Q2 is used as a collision cell, and Q3 is used to generate a mass spectrum of the daughter ions. In an ion trap (below), a particular m/z is isolated and fragmented, and the mass spectrum of fragments is generated, all with only one chamber. Modified from Ahn, et al. [17].

Isotope-Coded Affinity Tagging (ICAT)

Turecek, Aebersold, and others at the University of Washington published a technique, analogous to DIGE, for the relative quantitation of many peptides from two different samples using a single MS run [18]. Protein samples, such as cell lysates or serum, are labeled at cysteine residues with one of two biotin-conjugated tags which differ in mass by 9 amu. The samples are mixed and digested with trypsin. The peptide fragments with tags are isolated by passage through an affinity column (packed with beads linked to avidin). The biotin is cleaved off the tags in an acidic hydrolysis, and the peptides are analyzed, usually by two-dimensional capillary LC-MS and cLC-MS/MS. For two peaks or peak envelopes differing in mass by 9, the sample of origin can be determined. Relative quantitation is achieved by comparing the volume or intensity of peaks which correspond to the same peptide from different samples.

Peptide Mass Fingerprinting

MS/MS data are often used to predict the proteins from which peptides originate. Especially with tryptic digests, theoretical trypsin digestions of the expected protein products of any sequenced genome can be used as a catalogue of possible parent ions. The observed parent mass and the tryptic digest fragmentation pattern are compared with an expected tryptic digest fragmentation pattern of each of these candidates. Probability scores for matching proteins and their sequences are produced in seconds. Database searching algorithms, such as MASCOT or SEQUEST, are able to account for a variety of processing methods, including digestion by a number of enzymes, or the addition of

ICAT tags. Peptide mass fingerprinting and the use of searching algorithms are indispensable for high-throughput proteomic strategies.

Stable Isotopic Labeling with Amino Acids in Cell Culture (SILAC)

The labs of Matthias Mann and Akhilesh Pandey developed a method to simultaneously and quantitatively analyze the proteomes of two samples [19]. Cells are grown in media containing only a heavy isotope of a certain amino acid. Control and SILAC environments can be identical with the exception of the labeled amino acid, which does not change the behavior of the cells. As they grow, virtually all of the protein synthesized will be labeled with the heavy amino acid, and can therefore be distinguished from control in a mass spectrum. This method has potential application in a variety of proteomics and biochemistry problems.

Biomarkers

Biomarker discovery and validation have quickly become an important focus in proteomics. Early detection and diagnosis can often lead to more effective treatment in many conditions, particularly cancers. However, some of these conditions do not manifest symptoms until later stages, or are otherwise difficult to diagnose. Also, current diagnostic methods such as tissue biopsies and MRIs are invasive or expensive. Ideally, a protein or peptide marker from a blood sample could be used to predict the condition or indicate the presence of disease. Biomarkers are used today to help detect some cancers, but sensitivity and specificity are not ideal [20-22].

Landmark Study of Ovarian Cancer Biomarkers

Bioinformatics and statistics are helpful and indeed may be necessary for the discovery and validation of disease markers in biological specimens. With such a complex problem with so many interdisciplinary requirements, many sources of bias can go unrecognized, and poorly understood or misapplied statistics can be misleading. A 2002 paper in Lancet described a set of biomarkers for ovarian cancer which could predict the presence of disease with 95% specificity and 100% sensitivity [23]. In this study, Petricoin and colleagues analyzed 116 serum samples from women, 50 of whom had some stage of ovarian cancer. Computer learning algorithms were used to identify those features of the spectra which best correlated with cancer. Individual spectra were categorized using these patterns and a 94% positive predictive value for these samples was achieved [23].

A few problems with this study were soon identified. There was bias in that variables such as age could also be used to distinguish the groups. The biomarkers did not appear to be relevant to the pathology of the cancer. A second group analyzing the same data found that the same statistical difference could be obtained using peaks too small to represent molecules with biological significance [24]. Also, cases and controls were run on separate days and instrument variability could not be ruled out. Finally, the positive predictive value using the data from these patterns in a randomized, unselected population was much too low to be useful [25-27]. Bias has plagued bioinformatic analysis of spectra and must be controlled scrupulously. A heavy burden of proof rests upon

champions of potential biomarkers, including the demonstration of reproducibility of results, establishment of biological relevance of markers to the condition, and demonstrated exclusion of bias of any sort [28, 29].

Sensitivity in Proteomics and Implications in Biomarker Discovery

Proteins or peptides are not always efficiently detected or identified in proteomic approaches for several reasons, including inadequate sensitivity, poor separations, bias toward abundant species in IDA, or the presence of post-transcriptional modifications which are not included in the genomic map. Most circulating proteins are glycosylated. Sensitivity and dynamic range of 2DGE and MS are of great concern in biomarker discovery. In biological systems, a few proteins are many orders of magnitude more abundant than others, and this broad concentration range is presently impossible to accommodate. To manage this problem, the most abundant proteins are often removed from the sample. The problem of inadequate sensitivity remains. Many interesting proteins and peptides are naturally in low concentrations. Biomarkers whose clinical utility has been established are in much lower concentrations than can be detected by SELDI-TOF or other proteomics methods [30]. This point is illustrated in Table 1.1.

Protein/Peptide	Concentration (pmol/L)	Property
Serum Albumin	600,000,000	Abundant carrier protein
Immunoglobulins	30,000,000	Immune function
C-reactive protein	40,000	Inflammatory marker, acute phase
Apolipoprotein A1	40,000,000	Putative biomarker for ovarian and pancreatic cancer
Inter-alpha-trypsin inhibitor fragment	4,000,000	Putative biomarker for ovarian and pancreatic cancer
Vitamin D-binding protein	10,000,000	Putative biomarker for prostate cancer
Prostate-specific antigen (PSA)	140	Validated prostate cancer biomarker
Alpha-Fetoprotein	150	Validated testicular cancer biomarker
Carcinoembryonic antigen (CEA)	30	Validated colon cancer biomarker

Table 1.1. Comparison of concentrations of serum proteins, relatively new biomarker candidates, and established cancer biomarkers. Adapted from Diamandis [30].

Reproducibility in Proteomics and Implications in Biomarker Discovery

Reproducibility of MS data is imperative to the clinical utility of biomarker candidates. Unfortunately, there are many potential contributions to variability that can compromise reproducibility. These variables include inconsistency in sample collection protocols, specimen processing techniques, storage conditions, and instrument performance [31-33]. Indeed, significant differences in proteomic patterns can be introduced by merely running samples on different days [34]. Reproducibility across laboratories and instruments must be expected of data recommending the use of MS to assay biomarkers [35]. In two somewhat encouraging studies of SELDI-TOF spectra variability, storage of samples for variable amounts of time at -80 °C created no significant differences [36], and reproducibility between labs can be as consistent as that within an individual lab group [37].

Proteomic Variability

An added dimension of difficulty with proteomics is that the proteome is a moving target. Different genes are expressed in different tissues or cells, and genes are regulated to increase or decrease different protein abundances. Some genes are transcribed and translated into multiple proteins. For confidence in quantitation, the protein must be measured directly, given that levels of mRNA do not always correlate with protein concentration [38]. Protein abundance is additionally regulated in several ways including not only transcription, but also translation, covalent modification, rates of secretion into the blood and degradation. Some genes, such as housekeeping genes, are

constitutively expressed and remain at relatively constant levels. Others fluctuate in concentration as the needs of the organism change. For example, when the glucose concentration in the blood increases, i.e. after food ingestion, the concentrations of insulin and glucagon in the blood serum will quickly rise and fall, respectively. GLUT 4 glucose transporter proteins will be upregulated on the surface of hepatocytes. Various other dynamic proteins are affected in response to this single change.

The Hamburger Effect

While much is known about certain individual proteins and which stimuli may trigger changes in their expression, the large-scale dynamics of a proteome have not been thoroughly studied. The crux of this issue is whether actual variability due to biological differences, compounded with variability in sample processing and analysis, may drown out the truly important differences in biomarker signal abundance. It has been speculated that an ordinary environmental factor, e.g., the consumption of a hamburger, could spark radical changes in a proteome [39]. If this were true, inherent variability and background noise from environmental factors would make detection of subtle differences in biomarker concentration practically impossible. This has been termed the "hamburger effect." Although various environmental variables could contribute to the hamburger effect, dietary differences were examined specifically in this thesis.

In a study by Hsieh, et al., the hamburger effect was assessed by comparing SELDI-TOF spectra of blood serum samples taken from fasted and fed human subjects. There were four out of one hundred measured MS peaks shown to be different between

the two groups [31]. The differentially expressed proteins were not identified. The authors mentioned that no significant differences were found between male and female serum. This study, while demonstrating a 4% "hamburger effect," needed to be improved to take into consideration gender discrimination and to provide some biological interpretation of the effect via identification of the proteins.

Before one can account for the hamburger effect, the reproducibility of methods and inherent proteomic variability must be characterized more completely. In order to systematically study environmental factors, the assessment of the proteome of a well-characterized animal could be more informative than that of very limited existing research. Characterization of the variability of such a proteome and analysis of the hamburger effect in such a system may be a realistic goal.

Drosophila melanogaster as a Model System

Drosophila melanogaster is a good model species for the study of proteomics because it is a complex, multicellular organism that has been well characterized (including a completed genomic analysis), has a relatively short life cycle, and is easy to culture. Under ideal conditions, about two weeks are required for maturation from egg to adult, and they live at this stage for about six weeks. Adult females are receptive for fertilization after about eight hours, and may produce hundreds of eggs. The *Drosophila* genome is comprised of about 14,000 genes coded in three sets of autosomal chromosomes and a pair of sex chromosomes. Recently, a comprehensive study of the proteome has identified about half of the proteins and peptides encoded by the genome [40].

Drosophila proteomics may be used to segue into the most complex aspects of human proteomics and systems biology and provide relevant insight into proteomic variability.

Objectives

Chapter 2 is a comparison of the reproducibility of the proteome as observed by LCMS and 2DGE. Several control samples were analyzed using the two methods, and the peaks or spots from one representative sample were compared to the average output for the corresponding peaks or spots. The number of values which were within 10%, 20%, and so on, of the average value were counted and tabulated in order to determine which, if either, of the methods was more reproducible. These data were also compared with previous studies on the variability of other proteomes and methods.

The goal of the third chapter was to test whether there was greater variability in the LCMS method in measuring peptides or proteins as a function of elution time, m/z range, or average intensity. The intensities of selected peaks representing different elution time, intensity, and m/z ranges from the several control spectra were recorded. The coefficient of variation for each peak was obtained, and statistical methods were used to test whether greater variability was associated with any group of each of the parameters. It was found that greater variability existed among peaks with lower m/z than with middle-range or higher m/z. This suggests that greater care should be exercised when working with biomarker candidates of lower m/z.

The fourth and fifth chapters are studies of the hamburger effect. The hypothesis to be tested was whether a hamburger effect is observed in fruit fly populations by 2DGE

or LCMS given two changes in diet, one involving ethanol supplementation and a second imposing short term starvation. Gels and spectra from control and ethanol-fed or starved flies were compared, and significant differences were pursued further in order to identify the proteins whose concentration was affected by the change. The same methods were used to identify proteomic differences between male and female flies and wild-type and mutant flies, in order to verify the utility of the method.

Summary

Proteomics is a relatively new field with potential to augment accuracy and economy of diagnostic methods. 2DGE followed by MS and LC/MS, the two most commonly employed proteomic approaches, were assessed in this research. These techniques are most useful when comparing similar sets of proteins, which is typical of biomarker research. Some of the current challenges in this research are partially due to insufficient dynamic range (sensitivity) and imperfect reproducibility. It is also unknown how much plasticity exists in a complex proteome and how much variability there is between individuals. The level of influence of environment on proteomic variability also has yet to be firmly established. The proteome of *Drosophila melanogaster* may be used as a model to study variability in a complex proteome and to compare the reproducibility of methods in measuring this proteome.

CHAPTER 2

COMPARISON OF THE REPRODUCIBILITY OF 2DGE AND LCMS

Overview

Chapter 2 is a comparison of the reproducibility of the proteome as observed by LCMS and 2DGE. Several control samples were analyzed by either 2DGE or LCMS, and the peaks or spots from one representative sample were compared to the average output for the corresponding peaks or spots for the group. The data were analyzed using statistical methods in order to determine which, if either, of the methods was more reproducible. It was shown that neither method was significantly more reproducible, and that the variability of these methods was comparable to that reported for these and other proteomic methods in other labs.

Introduction

Reproducibility of proteomic methods has been thoroughly evaluated in relatively few contexts. Also, no comprehensive conclusion has yet been established in defining the origins of variability in proteomic measurements. The variability due to biological differences is overshadowed by noise from other sources in some cases. Albrethsen et al. studied the variability of SELDI-TOF of a single sample and between healthy human serum samples [41] and reported the average coefficient of variation for peak intensities ranged from about 13% to 18%. They reported a comparable level of variability between SELDI runs of the same sample and analysis of different samples,

which suggests that most of the measured proteomic variability stems from sample preparation and instrumentation rather than real biological differences between the samples. In another study of MALDI analysis of plasma proteins, intra-individual variation was less than half of that between different subjects, with the average coefficient of variation for proteins from the same individual being less than 10% [42].

Studies of 2DGE variability of various tissue extracts report different levels of reproducibility. A 2002 review by Anderson and Anderson on plasma proteomics suggested that about half of the variation in 2DGE measurements of human plasma proteins was related to real biological differences [43]. One study of human cerebrospinal fluid using DIGE reported a much greater variability between individuals than within the same sample, with a 10% average coefficient of variation for control samples [44]. Nishihara and Champion reported the range of coefficients of variation for 2DGE measurements of *E. coli* proteins from 3 to 33% with an average close to 10% [45]. In contrast, two more recent studies of 2DGE reproducibility using human serum and liver proteomes showed an average coefficient of variation of protein expression of 19% and between 18 and 27% [46, 47].

The question addressed in this chapter is whether there is a significant difference in the variability of molecular species abundance of fruit fly proteins or peptides measured by 2DGE and LCMS, and whether the data are consistent with any previous research. 2DGE and LCMS are fundamental proteomics tools, and a better understanding of their reproducibility will help evaluate the degree to which biological differences affect reproducibility as compared with methodological factors, and also how large a

biological difference would have to be present to overcome the inherent analytical uncertainty. Perhaps the method with least methodological noise will be given greater consideration for clinical applications.

Many methods are used to deplete highly abundant, generally uninformative proteins especially from human serum samples [48-50]. In blood serum, the most abundant proteins are comparatively large. Some of the depletion methods capitalize on the size disparity. Acetonitrile (ACN) precipitation is one such method. ACN denatures large globular proteins because protein hydration spheres are disrupted. As the hydrophobic cores of large proteins are exposed, they aggregate and precipitate out of solution. These large proteins are pelleted by centrifugation. Because the structures of small proteins and peptides are not affected as much, these retain their solubility and remain in solution.

ACN precipitation was used in the preparation of the samples for LCMS. This reduced the complexity of the protein mixture, particularly eliminating the majority of proteins. Many highly abundant proteins, such as actin and other structural proteins, may otherwise have masked signal from less-abundant, small peptides. This depletion method also allowed a different cross-section of the fly proteome to be studied than that of the 2DGE experiments. A disadvantage to this process is that the informative globular proteins are lost. Also, after elimination of the large proteins from the samples, the dynamic range of peptide abundance remains high.

Materials and Methods

All materials were purchased from Fisher Scientific (Fair Lawn, NJ) unless otherwise specified.

Drosophila Culture

Oregon-R wild-type and white-eyed mutant *Drosophila melanogaster* (Carolina Biological Supply, Burlington, NC) were raised on Formula 4-24 Blue (Carolina) prepared according to supplier instructions with 0.5% propionic acid added as a supplemental preservative. An F2 generation of about 30 flies (each sharing the same pair of parents for 2 generations) was transferred to several vials for 24-hour intervals. Although the parent generation may age, the gene pool for each of the new cultures of their offspring is theoretically identical. Thus, genetically homogeneous fly cultures were arranged approximately 24 hours apart in development.

Protein Extraction for 2DGE

Proteins were extracted from adult flies approximately 5 days after maturity. Flies were anesthetized using FlyNap (Carolina) according to manufacturer's protocol, sorted by gender and homogenized under liquid nitrogen in a 7 mL dounce homogenizer (Wheaton, Millville, NJ). Rehydration buffer (15 µL per female or 11 µL per male) composed of 7M urea, 1.5M thiourea, 4% 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS), 1% pH 4-7 IPG buffer (GE Healthcare, Piscataway, NJ), 1mM phenylmethanesulphonyl-fluoride (PMSF) was added to the homogenate. Samples

were kept on ice throughout preparation and used within 6 hours of homogenization. Protein concentration was assayed using a Bradford Protein Assay Kit (Bio-Rad, Hercules, CA).

2D PAGE

For each 2D gel, a volume of 140 µL of rehydration buffer containing 350 µg total protein was rehydrated overnight at room temperature (RT) into a 7 cm linear gradient pH 4-7 Immobiline DryStrips IPG strip (GE). 2DGE was performed as described previously[51]. Briefly, isoelectric focusing was performed at 3500V for 3 hours according to manufacturer's protocol using a Multiphor II system (GE). The strips were equilibrated in a reducing buffer (50mM Tris at pH 6.8 with 8M urea, 30% glycerol, 1% SDS, and 0.5% dithiolthreitol (DTT)) for 10 minutes at room temperature (RT), and again in an alkylating buffer (50mM Tris at pH 6.8 with 8M urea, 30% glycerol, 1% SDS, a few grains of bromophenol blue for visualization on the gel, and 4% iodoacetamide) for 10 minutes at RT. The proteins in the strip were then transferred via SDS-PAGE out of the strip into a 7cm, 10% acrylamide gel at 25 mA per gel for 1 hour. Gels were stained with Coomassie-based GelCode Blue Stain Reagent (Pierce, Rockford, IL) and digitally imaged using the AlphaDigiDoc RT system (Alpha Innotech, San Leandro, CA).

Gel Analysis

The total intensity of each gel was normalized using Imagemaster software (GeneBio, Geneva, Switzerland) to correct for loading differences. Spots were detected in each gel automatically with Imagemaster. Spot detection was verified manually. This was important because many non-authentic spots were eliminated and the recognized shape of some spots was corrected. The spots from each gel were matched together in the software according to shape and location. Pairs were manually verified, and no corrections were made. Imagemaster was used to generate tables listing area, position, intensity and other data about each spot on each gel. These tables were used to compare the abundances of the same protein across multiple gels.

The software was also used to produce a synthetic, or composite, gel by averaging the shape and intensity of all of the spots in five representative gels. Wild-type male fly protein extracts were used for each of the gels combined and averaged into the synthetic gel. The data from the synthetic gel as well as each representative single gel were transferred to Excel (Microsoft, Redmond, WA) for comparison. The values representing the amount of protein in each spot in a representative gel were divided by the corresponding spot in the synthetic gel. These data were distributed normally, and were compared with LCMS variability data using a T-test as described below.

LCMS Sample Preparation

Groups of 25 to 30 flies were homogenized as before. Protein was extracted into 400 µL phosphate-buffered saline (PBS) with 1mM freshly added phenylmethanesulphonyl-fluoride (PMSF). ACN precipitation was performed as described previously [52]. Briefly, two volumes of ACN were added to the protein extract. The mixture was vortexed for 5 seconds, and incubated at room temperature for 30 minutes. The precipitate was pelleted in an IEC Micromax RF centrifuge (Thermo Fisher Scientific, Waltham, MA) at 14,000 rpm for 10 minutes. The supernatant was evaporated in a Centrivap concentrator (Labconco, Kansas City, MO) at 30° C to near-dryness (about 5 hours), and protein was resuspended to a total volume of 30µL with ultrapure water. Samples were stored at -80° C for no longer than 2 weeks before proteomic analysis. Samples were diluted to a final concentration of 0.2 µg/µL protein in 50% formic acid just prior to LC/MS analysis.

Liquid Chromatography/Mass Spectrometry

Online reversed-phase capillary LC/MS was performed with the LC Packings UltiMate HPLC system (Dionex, Sunnyvale, CA) and a Q-star Pulsar i triple quadrupole/TOF mass spectrometer (Applied Biosystems) in electrospray ionization mode. Guard columns and capillary columns were packed with SelfPack Poros 10 R1 (Applied Biosystems), which is a C4-like resin used mostly for resolution of digested peptides. One µg total protein (5µL) was loaded onto the column by the autosampler robot. A gradient from 95% aqueous to 95% ACN was applied over 45 minutes at 5

μL/min to elute the peptides from the column beginning with the least hydrophobic and ending with the most hydrophobic. For the first 30 minutes, the rate of ACN% increase was 1.5/min, and for the last 10 minutes it was 4.5/min because most of the peptides are eluted by 50% ACN. At the end of this gradient, 95% aqueous solvent was passed through the column for 5 minutes to reequilibrate the column. The mass spectrometer was programmed to acquire MS and IDA data throughout the entire LC gradient.

LCMS Spectra Analysis

A comparable method was used for analysis of variability in LCMS spectra as was used for 2DGE. Because of limitations in the software and for practicality, only 36 peaks were compared rather than the hundreds of peaks that could be found in each spectrum. The selection of the peaks, described in Chapter 3, was designed to represent equally peptides of different size, hydrophobicity, and abundance. For each peak, the intensity from one representative sample was compared to the average intensity for that peak from all ten control spectra. The quotients of these values were distributed normally around 1, and were used for T-test comparison with the data from 2DGE as described below. Each of these values was subtracted from 1 and multiplied by 100 to find the percent difference. Percent differences were grouped in a table as before with values between 0 to 10% difference, 10 to 20% difference, etc.

Statistical Comparison of 2DGE and LCMS Variability

In order to determine whether one method was more consistent in its measurement of abundance, the values obtained in these experiments were compared using statistical methods. T-tests can be used to determine whether the means of two populations are significantly different. The value of spot or peak intensity from the single gel or spectrum was divided by that of the corresponding average for that spot or peak giving the ratios which were normally distributed centering around 1. These groups of ratios were compared using the 2-tailed homoscedastic T-test function in Excel. The percent differences between average and individual peak intensities or spot abundances were compiled into tables as another form of comparison.

Results

2DGE

Approximately 300 proteins were visible in each gel, with higher resolution and protein density on the more basic side of the gel and between 40 and 80 kD. A representative 2D gel with molecular weight markers and pH gradient is shown in Figure 2.1. The synthetic gel produced from 5 real gels was used as a control, or average. The spots of the synthetic gel were converted from black to red, and the representative gel for comparison was colored green, as shown in Figure 2.2, using Adobe Photoshop software. These images were overlaid, shown in Figure 2.3, to illustrate the quality of the synthetic gel and reproducibility of the 2D gels.

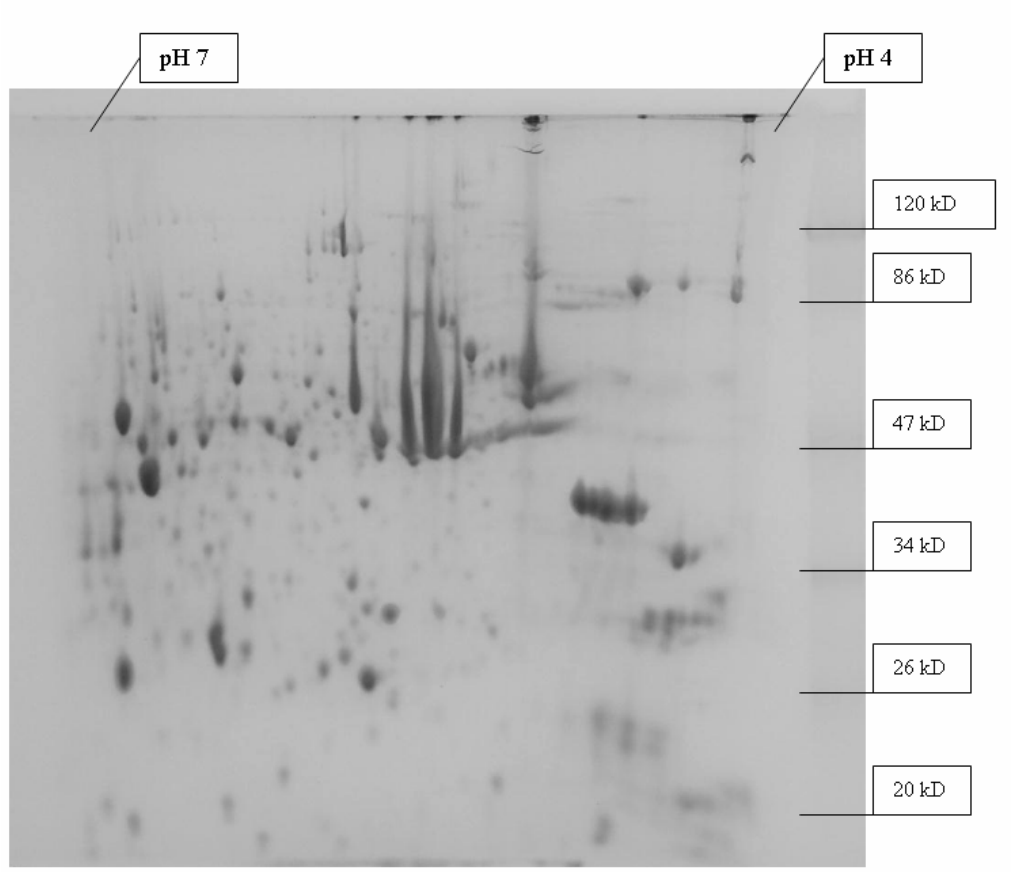


Figure 2.1. Representative 2D gel of fruit fly homogenate protein extract. Molecular weight markers and pH boundaries are labeled.

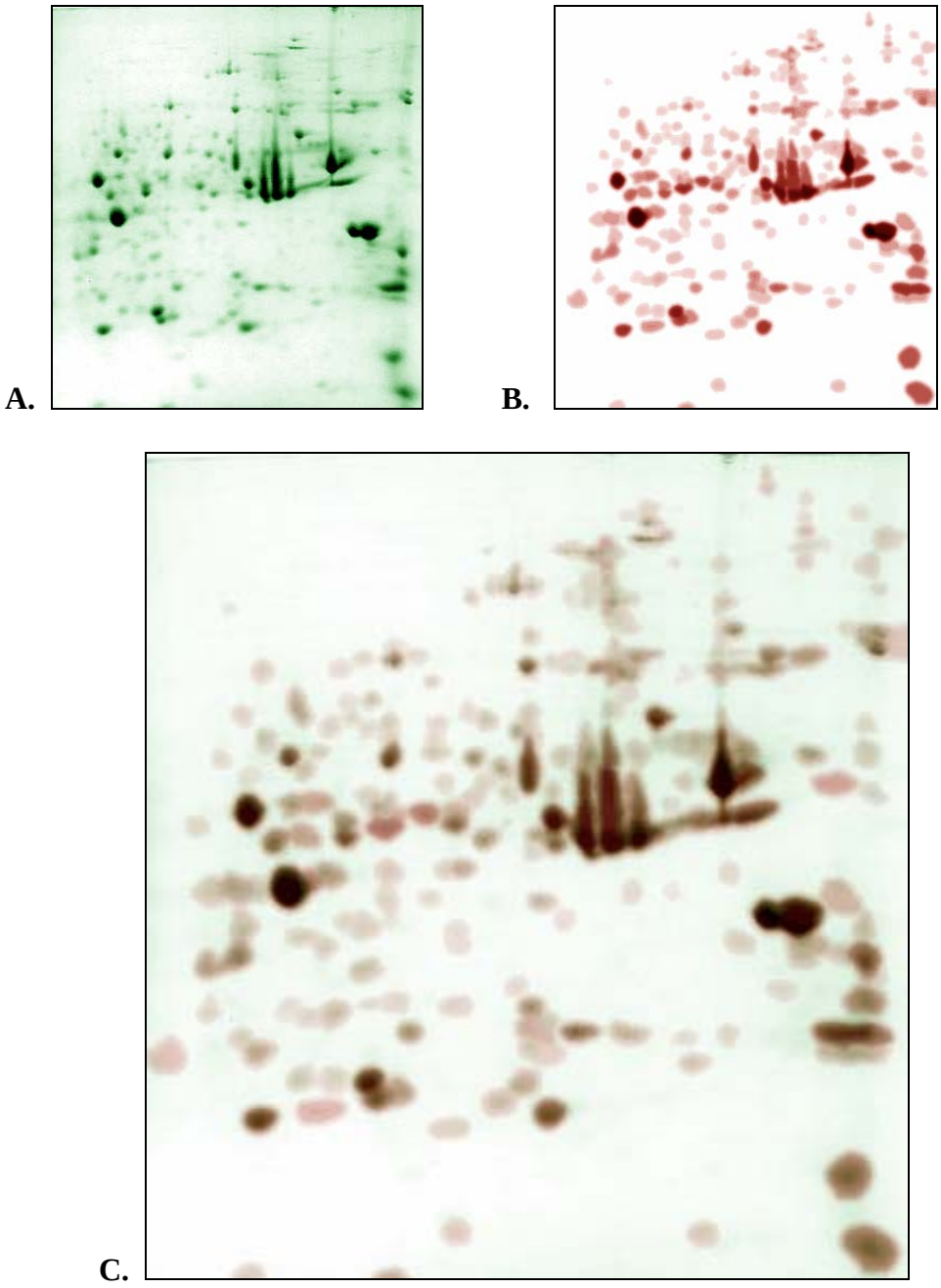


Figure 2.2. Visual display of synthetic gel average compared with individual gel. A representative gel was colored green (A) and the synthetic gel was colored red (B). Synthetic and representative 2D PAGE images are superimposed (C) to show the congruity of the proteomic profiles.

Only spots matched in at least 3 individual gels were included in the synthetic gel. The synthetic gel had a total of 256 spots. Of the 256 total spots in the synthetic gel, 77% were matched to all five gels, 15% were matched to only four gels, and 8% had matching spots in only three gels. 196 of these matched with the representative gel chosen for comparison. The protein abundance in each spot in the representative and synthetic gels is given in Appendix 1. These data along with the data from LCMS were used in the T-test described below. The values representing the amount of protein in each spot in the representative gel were divided by their counterpart from the synthetic gel. These data, representing the amount of variability in 2DGE and analysis of protein abundance using this software, are given in Table 2.1.

Difference in protein abundance	Number of spots	Percent of spots
0-10%	69	35
10-20%	58	30
20-30%	24	17
30-50%	27	14
50-75%	6	3
75-100%	2	1
>100%	0	0

Formula: |synthetic – real| / (synthetic) * 100 = % difference in protein abundance

Table 2.1. Variability of a cross-section of the wild-type fruit fly proteome as measured by 2DGE. The difference in the values between spots in the representative gel and corresponding spots in the synthetic gel was divided by the value of the same spot in the synthetic gel. This formula is given below the table. The spots were grouped according to percentage difference, and the number of spots and percentage of the total spots in each group are given.

LCMS Variability Compared with 2DGE

The disparity of peak intensity between a representative mass spectrum and the mean peak values from 10 spectra for the 36 peaks was determined. The results were grouped according to percentage differences, and Table 2.2 shows the number of peaks falling into each category as well as the 2DGE data from Table 2.1 for comparison. Two-thirds of the peaks were less than 20% different from the mean values, much like the results from 2DGE. The distributions appeared to be similar between 2DGE and LCMS.

The mean difference percent for the 2DGE data was 18.5, compared to 14.8 for the LCMS data. The raw data, given in Appendices 1 and 2, were analyzed statistically to compare the mean differences of the 2DGE and LCMS data. Student's T test was applied because the data sets are homoscedastic and distributed normally. This yielded a p-value of 0.182 with n = 36 for the LCMS data and n = 196 for the 2DGE data. This result fails to confirm that there is a significant difference between the populations. Thus, the variability in measurement of fruit fly protein or peptide abundance using 2DGE is not significantly different from that of LCMS.

LC/MS				2DGE		
Differences in peak intensity				Differences in spot intensity		
	Number of peaks	% of peaks			Number of spots	% of spots
0-10%	13	36		0-10%	69	35
10-20%	15	42		10-20%	58	30
20-30%	4	11		20-30%	34	17
30-50%	3	8		30-50%	27	14
50-75%	1	3		50-75%	6	3
75-100%	0	0		75-100%	2	1
>100%	0	0		>100%	0	0

Formula: |average – individual| / (average) * 100 = % difference in protein abundance

Table 2.2. Variability of the proteome as measured by LCMS compared to that of 2DGE. The intensity of 36 peaks representing different elution times and m/z windows in 10 different spectra were collected. The average intensity for each peak was calculated. The intensity of each of the 36 peaks from a single spectrum was compared to the average intensity value, and the formula used for evaluation of the variability of 2DGE was employed. Table 2.1 was included to facilitate comparison. A T-test of the distributions of the individual-to-average ratios (given in Appendices 1 and 2) gives a p-value of 0.182, meaning that there is not a statistically significant difference in the means of these data sets.

Discussion

The objective of this study was to characterize the reproducibility of 2DGE and LCMS of the soluble proteome of the fruit fly, and to determine which method introduces the least variability. The data in this study suggest that neither of the two methods is significantly more reproducible than the other. Also, both mean difference values fall within two percent of the range of mean differences for SELDI-TOF reported by Albrethsen et al. [41] and the 2DGE data of Zhang et al. [47]. It could be that all proteomics methods currently employed share similar reproducibility.

The non-biological sources of variability in 2DGE and LCMS are not precisely identified. It is likely that inconsistent resolution in the gels and problems with the imaging system or the analysis software contributed to the variability of the 2DGE data. One possible source of variability for the LCMS data is inconsistent resolution or sensitivity of the mass spectrometer. When MS resolution and sensitivity are greater, the peaks are sharper and taller, while low resolution creates shorter and broader peaks having the same area under the peak. Because the peak height, and not the area, of each peak was used throughout the experiments, and because sensitivity and resolution fluctuate slightly, this could account for some of the observed variability.

Another possible source of variability in ESI-MS is the possibility of multiple charge states for the same molecule. If the distribution between charge states differs from run to run, it would increase variability. Multiply charged species dominated the spectra, but any trend of greater variability among species with more charge was not

immediately obvious.

Another factor to be considered with the 2DGE data is the possibility of multiple proteins contributing to the intensity of a single spot. 2DGE resolution of proteins depends on the size of the gel, and because reproducibility was considered to be more important than resolution in this study, some proteins were not resolved absolutely. Indeed, one excised spot was analyzed by MALDI-MS for protein identification, and matched two proteins with a high degree of confidence. This is noted in Chapter 4. Also, it is clear that resolution of proteins with pI near 4 and 5 was not as good as that of proteins with a pI closer to 7. The possibility of more than one protein contributing to the intensity and percent volume of some spots adds complexity to the problem of describing the variation among many individual proteins using 2DGE.

The information-dependent data acquisition (IDA) feature was used to select for the most abundant peaks for CID and MS/MS, but no positive identifications of proteins were obtained from the LCMS studies. This may be due to poor fragmentation or an excessively complex sample. The studies which are most successful in positive identification of proteins use different methods, including trypsin digestion and more separation steps, and routinely identify hundreds of proteins [40, 53, 54].

Conclusion

The variability of a cross-section of the *Drosophila melanogaster* proteome as measured by both 2DGE and LCMS was characterized. The level of variability is comparable, although LCMS appears to be somewhat more reproducible. One difference between the methods is that many more species were detected in LCMS than in 2DGE. There were several peaks in the LCMS spectra which represented the same species in different charge states. Conversely, more than one spot in the gels could represent the same protein with phosphorylation or other modifications that alter the pI. In this case, LCMS is also more sensitive than 2DGE.

CHAPTER 3

ANALYSIS OF THE VARIABILITY OF LCMS AS A FUNCTION OF DIFFERENT ELUTION TIMES, PEAK INTENSITIES, AND MASS-TO-CHARGE RATIOS

Overview

This study tested whether there was greater variability in the LCMS method used in the previous chapter in measuring features of different elution time, m/z range, or peak intensity. The intensities of random peaks of representative elution time, intensity, and m/z from ten control spectra were recorded. The coefficient of variation for each peak was obtained, and statistical methods were used to test whether greater variability was associated with any group within each of the parameters. It is shown that a greater variability was associated with peaks having lower m/z than with middle-range and higher m/z.

Introduction

Understanding the degree to which a scientific measurement is reproducible is important for establishing confidence in the measurement. In many proteomic or other protein survey studies such as immunoblots, controls or standard proteins of predictable behavior are used to validate the measurement of the targeted protein. In proteomics, and particularly with LCMS, an internal standard protein of known concentration and molecular mass can be spiked into the samples in order to normalize for loading differences, and calibrate for mass accuracy. However, in the second chapter it was

reported that there was an average of 19% variability in measuring a given peak using LCMS. Another concern is that the standard, having one elution profile and molecular mass, may behave differently than other species elsewhere in the MS run.

This chapter addresses whether there is greater or lesser variability in the measured peak intensities of proteins having various elution times, m/z values, or peak intensity ranges. This may be important in identifying some regions of MS spectra that may be more or less reliable than others. No bias in variability due to elution time, peptide size, or relative abundance was anticipated.

Materials and Methods

Drosophila culture, protein extraction, LCMS sample preparation, and LCMS were performed as described in Chapter 2.

Selection of Peaks

Manual selection and measurement of a small number of peaks was chosen over measurement of all peaks by the Analyst software for two reasons. First, Analyst software sometimes fails to pick prominent peaks, and detects peaks which are not real. Second, the quantitation which accompanies the peaks using the volume below the peak is often inaccurate. Although more data would have been generated, the quality may have been poor. The peaks chosen for this study were checked manually for authenticity and peak height was used rather than peak area in an effort to avoid these problems.

Three different elution time windows were chosen near the beginning, middle, or

end of the region of the spectra in which most of the peptides had eluted. A representative MS spectrum was examined at each window and a peak list was generated automatically using 5 as the lower threshold of intensity. The peaks were sorted by m/z, and divided into three numerically equal groups representing low, middle, and high m/z ranges. It was found that the boundary between the low and middle m/z range fell at approximately 800, and the boundary between middle and high at ~1000 for each of the three elution time windows. Peaks were then randomly selected for analysis to provide equal representation for each of the 3 representative elution time windows and m/z ranges. Each peak was assigned a number, and four were chosen randomly, using a pseudo-random number generator from the three different m/z ranges (500-800, 800-1000, and 1000-1400), in three different elution time windows (about 8-10, 14-16, and 21-23 minutes into the LC gradient) giving a total of 36 peaks. Thus, each elution time window and m/z range was represented equally and the total number of peaks was small enough to allow practical, manual analysis.

Normalization of Peak Intensities

Normalization of the data compensated for differences in sample loading or instrument sensitivity. The difference between the intensity of each peak and the average from all ten samples was determined. These were converted to percentages so that each peak was weighted equally. The 36 percentages for each sample were averaged. This average as a decimal was subtracted from 1 to create a normalization factor particular to each of the ten samples. Each intensity value was multiplied by the normalization factor

for its respective sample. Normalization factors ranged from 0.85 to 1.16.

Statistical Analysis of Variability Among Control Samples

The normalized intensity of each of the 36 random peaks in 10 control MS runs was recorded on an Excel spreadsheet. The coefficient of variation for each peak was obtained. Because it was not assumed that the coefficients of variation were normally distributed, a non-parametric test was used to determine statistical significance of differences between the groups. The Kruskal-Wallis test [55] was performed using R (R Foundation for Statistical Computing, Vienna, Austria) comparing the three elution times and ranges of m/z as the parameters. This test returns a p-value describing the difference between the means of more than two populations of any distribution.

Results

The raw data, including the intensity values of each of the 36 peaks in 10 spectra, and the mean, standard of deviation (SD), and the coefficient of variation (CV) for each peak are given in Appendix 3. Table 3.1 gives the mean CV and SD for each of the 36 peaks. Bar graphs showing the mean CV for the peaks grouped according to elution time, m/z range, or average intensity are given in Figure 3.1, Figure 3.2, and Figure 3.3, respectively. The results of the Kruskal-Wallis test were 0.160, 0.00407, and 0.685 for the comparisons among the 3 parameters, respectively. These results suggest that, in this study, peak intensities of peptides below 800 m/z have a greater CV than do peptides with m/z greater than 800.

Elution Time	CV	SD
14 - 16 Minutes (Early)	0.222	0.103
20 - 22 Minutes (Intermediate)	0.161	0.077
27 - 29 Minutes (Late)	0.185	0.064
m/z Range	CV	SD
500 - 800 (Lower)	0.260*	0.104
800 - 1000 (Intermediate)	0.167	0.048
1000 - 1400 (Higher)	0.142	0.036
Average Intensity	CV	SD
1 - 9 (Lower)	0.197	0.074
10 - 18 (Intermediate)	0.199	0.106
> 18 (Higher)	0.173	0.075

Table 3.1. Mean coefficient of variation (CV) and standard deviation (SD) for 36 peaks in control LCMS spectra. The same peaks are grouped three different ways, according to elution time, m/z range, and average intensity, with n = 12 determinations for each of the nine subgroups. Lower m/z range peaks (asterisk), have a significantly higher CV than intermediate and higher m/z peptides according to the Kruskal-Wallis test.

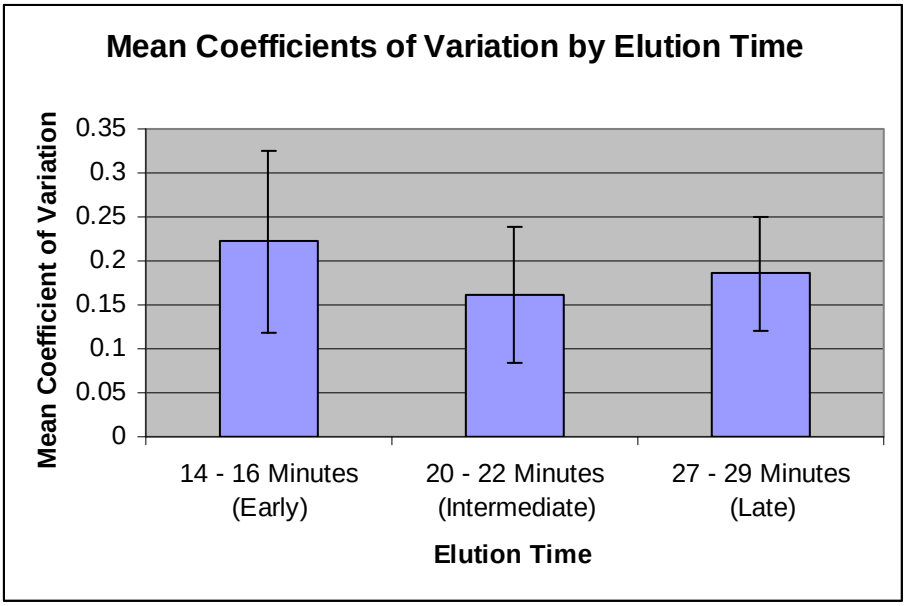


Figure 3.1. Bar graph showing mean CV +/- SD for measured intensity of peptides from 10 spectra at three elution times, n = 12. No significant difference between mean CV within these groups was found (Kruskal-Wallis, p = 0.160).

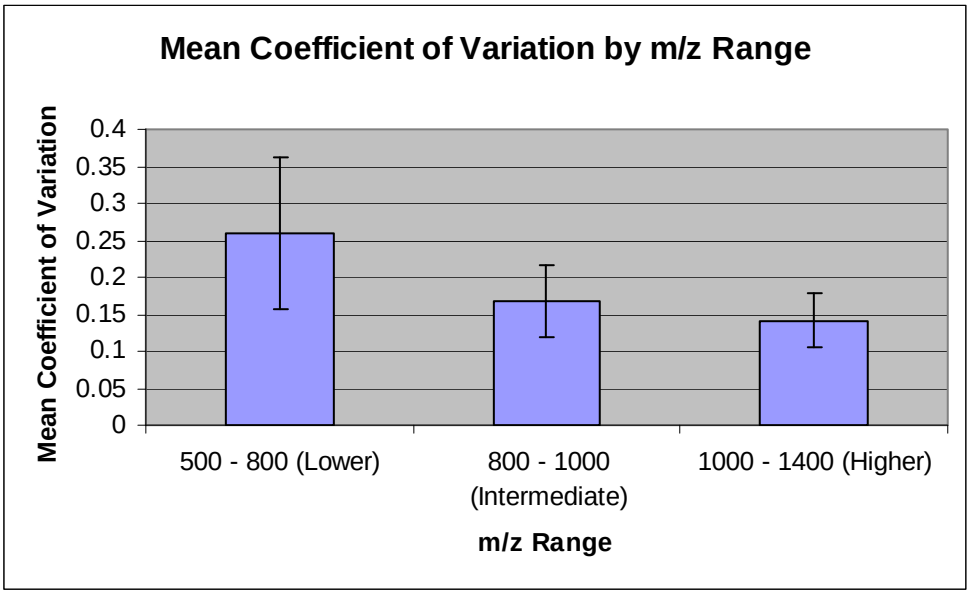


Figure 3.2. Bar graph showing mean CV +/- SD for measured intensity of peptides from 10 spectra in three ranges of m/z, n = 12. Measured intensities of peptides in a lower m/z range, below 800, were found to have a greater CV than intermediate and higher m/z peptides (Kruskal-Wallis, p = 0.00407).

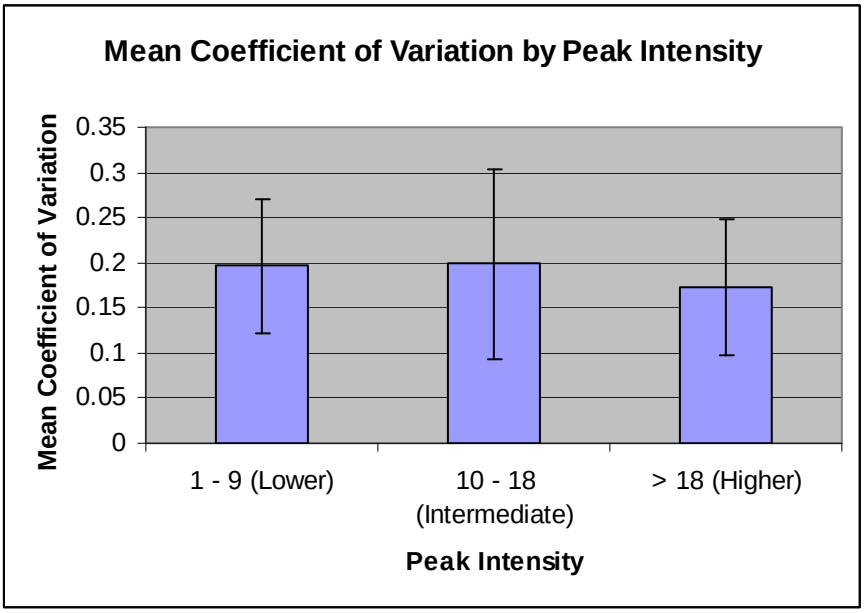


Figure 3.3. Bar graph showing mean CV +/- SD for measured intensity of peptides from 10 spectra with three ranges of average intensity, n = 12. No significant difference between mean CV within these groups was found (Kruskal-Wallis, p = 0.685).

Discussion

The null hypothesis was that there would be no significant difference between the mean coefficient of variation of peptides of different intensity, elution time, or m/z range. However, the findings of this study suggest that MS measurements of abundance of peptides of lower m/z are less reproducible than those of higher m/z. It should be noted that non-proteinaceous polymers and nitrogen-rich species are more often detected in the lower m/z range than in a higher m/z range, so perhaps some of the peaks in this category were not peptides. One possible reason low m/z peaks are more variable is that there is greater biological variation in the abundances of smaller peptide species. Another explanation could be that the mass analyzer is less-consistent at the lower range of the mass spectrum.

The manner in which protein abundance was measured may play a part in the reason lower m/z peptides show greater variation. Resolution of the mass spectrometer is defined as the width of the peak at maximum height divided by the m/z. If resolution is consistent throughout the spectrum, those peaks with lower m/z are sharper than those at higher m/z. The peptide abundance was measured using the height of the peaks rather than the area underneath the peaks, and the disparity between the heights of sharper peaks is a greater percentage than that of wider peaks of comparable area and variability. The area under the peaks was also analyzed for comparison, but due to inaccuracy of the software in assigning peak areas the coefficients of variation were roughly double those of the data comparing peak height (data not shown).

Finally, although the trend was not statistically significant, it is likely that peaks of lower intensity are measured less-reproducibly than those of higher intensity. A reason for this is that, due to the inherent level of noise in the mass spectra, peaks of lower intensity will be affected more proportionally. A study including peaks with average intensities below 5 as a fourth category could show a statistically significant trend. However, the noise level makes measurement of some of these peaks difficult.

While overall reproducibility of LCMS spectra in these experiments is moderately good, it is not uniform for every peak or range of peaks.

Conclusion

In this LCMS study of fruit fly peptides and low molecular weight proteins, it is demonstrated that species in a lower m/z range are measured less consistently than those of intermediate and higher m/z. Interesting questions are raised by this result involving the reason for this disparity. The practical implications of this finding are that potential biological differences will be harder to observe in peptide expression measured by LCMS involving peaks in the low m/z range. Some real differences may be lost. A similar study of a larger number of peaks would be useful in confirming these results, and more precisely defining the average variation of peaks of different ranges of m/z.

CHAPTER 4

OBSERVATION OF PROTEOMIC DIFFERENCES BETWEEN POPULATIONS OF FRUIT FLIES, AND EVALUATION OF THE HAMBURGER EFFECT USING 2DGE

Overview

2DGE was used to evaluate the hamburger effect of the most abundant proteins in the fruit fly proteome. Proteomic differences between male and female flies and wild-type and mutant flies were studied in order to confirm the ability of this method to detect fixed genotypic proteomic differences between genders or phenotypes. Those proteins found to be expressed differently, as well as the most abundant proteins, were identified by MALDI-MS and peptide mass fingerprinting. The diets of some populations were supplemented with ethanol to assess whether measurable dynamic changes in the proteomes occurred. Extreme sample handling conditions were also tested for impact on the proteomic profile measured by 2DGE. No significant changes in the proteome were found after subjection to freeze-thaw cycles. Significant proteomic differences were demonstrated between male and female flies as well as between wild-type and white-eyed flies. No significant proteomic difference was shown between control flies and those given ethanol.

Introduction

One potential source of biological variability in proteomic assessment is the hamburger effect, described on page 19. Despite the potential importance of this concept, it has been little studied. Should it exist and be extensively seen it would have a profoundly confounding effect on studies of biomarker discovery. The serum proteome of an individual is known to undergo significant changes as part of normal development and physiology [43], but little is known about the short-term dynamics of the proteome.

Challenges in studying the hamburger effect include especially the limited sensitivity of 2DGE and potential limits in method or instrument reproducibility, discussed in Chapter 2. Also, sampling tissue multiple times from the same subject under different circumstances is not possible. The latter difficulty can be circumvented using an animal model, such as fruit flies. Some advantages of using fruit flies as a model include that they are well-studied, complex multicellular organisms and easy to culture. *Drosophila* can have a high fecundity and mature within about ten days.

Two methods, LCMS and 2DGE, were used to study the hamburger effect. The LCMS work is described in Chapter 5. For the 2DGE study, 7 cm width gels were chosen over larger (18 to 24 cm) gels because smaller gels are more reproducible, less expensive, and require less time for electrophoresis. The resolution is greater in a larger gel, but spot alignment is more difficult. Reproducibility was chosen as a higher priority so that the gel analysis software would match the spots accurately. MALDI-MS (rather than ESI) was chosen for identification of the more abundant proteins because it is more efficient for purified proteins, which in this case are already resolved by 2DGE.

Materials and Methods

Proteomic Comparison of Male and Female Flies

In order to confirm that the sensitivity and reproducibility of these methods were adequate for detection of fixed genotypic proteomic differences, protein extractions of phenotypically different populations were first compared. Protein from five wild-type male and five wild-type female populations of 15 to 25 individuals were analyzed by 2DGE. *Drosophila* culture, protein extraction, sample preparation, and 2DGE were performed as in Chapter 2. Imagemaster software was again used to identify spots and quantify intensity. Differences were considered significant if a spot was absent in one group of gels, and detected in the other gels with a percent of total protein abundance, combining all spots from each gel together, of at least 0.2.

Proteomic Comparison of Wild-type and White-eyed Mutant Flies

For the same reasons mentioned in the previous section, protein extractions from three white-eyed mutant male populations were compared with those of the wild-type males using 2DGE. *Drosophila* culture, protein extraction, sample preparation, and 2DGE were performed as in Chapter 2. Those proteins found to be different between wild-type and white-eyed mutant proteomes were identified by in-gel protein digestion, MALDI-MS and MS/MS, and peptide mass fingerprinting.

Study of the Hamburger Effect

The diets of some populations were manipulated to test whether a hamburger effect on the fruit fly proteome may be observed by 2DGE. Twelve and 24 hours prior to protein extraction, populations were introduced into new vials prepared with Formula 4-24 Blue as before (0% ethanol) or supplemented with 5% or 10% ethanol. Male flies were isolated, and the same methods were used for protein extraction, sample preparation and 2DGE. Each experiment was repeated so that 12 gels were produced in total.

To address the concern of the effect of sample storage and handling, one protein extraction was divided into two parts. Prior to preparation for 2DGE, one part was frozen under liquid nitrogen and thawed in a water bath at 70 °C ten times over a period of ten minutes. The other part was stored on ice for these ten minutes. These two samples were then analyzed by 2DGE using the same methods.

Protein Identification

Some of the proteins found to be different between male and female proteomes or between wild-type and white-eyed mutants were identified by in-gel protein digestion, MALDI-MS and MS/MS, and peptide mass fingerprinting. In-gel digestion with trypsin (Promega, Madison, WI) was performed as described by Shevchenko et al. [8]. Briefly, the small pieces of gel containing the protein were equilibrated with 50 mM ammonium bicarbonate/50% acetonitrile at RT for 30 minutes followed by a second equilibration with 100 mM ammonium bicarbonate at RT for 30 minutes. The pieces were then dehydrated by equilibration with 50 mM ammonium bicarbonate/50% acetonitrile at RT

for 30 minutes, followed by aspiration of the buffer and lyophilization to completely dry the pieces. The gel pieces were rehydrated with about 15 µL of a solution containing trypsin (20 µg/mL, Promega, Madison, WI) according to manufacturer's protocol, with an additional 20 µL of 25 mM ammonium bicarbonate. The digestion proceeded at 37 °C overnight, and the reaction was quenched with 1 µL of concentrated formic acid.

The pieces of gel were sonicated using a FS 30 water bath sonicator (Fisher Scientific, Pittsburgh, PA) for 20 minutes at RT to expedite diffusion of the peptides out of the gel. ZipTip columns (Millipore, Billerica, MA) were used to desalt the specimen according to manufacturer's procol. Small volumes of sample were passed through a tip prepared with a peptide affinity resin. The column was washed to eliminate much of the salt in the sample. The peptides were eluted in a low-salt, acidic buffer. The peptides were thereby purified and concentrated. Peptide in 1 µL of sample was co-crystallized with 1 µL α-cyano-4-hydroxycinnamic acid (CHCA) for at least 20 minutes in the dark.

MALDI mass spectrometry and MS/MS were performed as previously described [56] using an oMALDI ion source and QSTAR Pulsar I quadrupole TOF mass analyzer (Applied Biosystems, Foster City, CA). Three or four abundant, monoisotopic parent ions with m/z between 1000 and 2000 were selected manually and submitted to CID. The laser power level was set to 7.6 with a pulse rate of 20 Hz. CID collision energy was shifted between 50 and 105 in order to produce ions evenly throughout the range of m/z from 80 to the size of the chosen peak.

The MASCOT search engine (Matrix Science, London, UK) was used to screen masses of daughter peptides against the NCBI database for peptide matches allowing

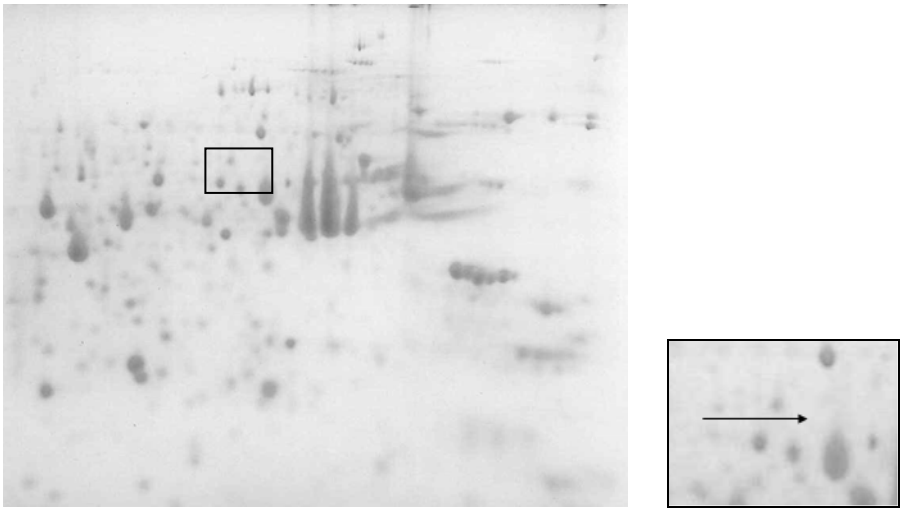
protein identification. For each spot, at least two MS/MS spectra were matched to the same protein with a mowse score greater than 50. Mowse scoring is an effective way to calculate the probability that a spectra matches a given peptide [57]. In this case, a mowse score of 50 represents a p-value of 0.05. Peptide matches with such scores are considered statistically significant.

Results

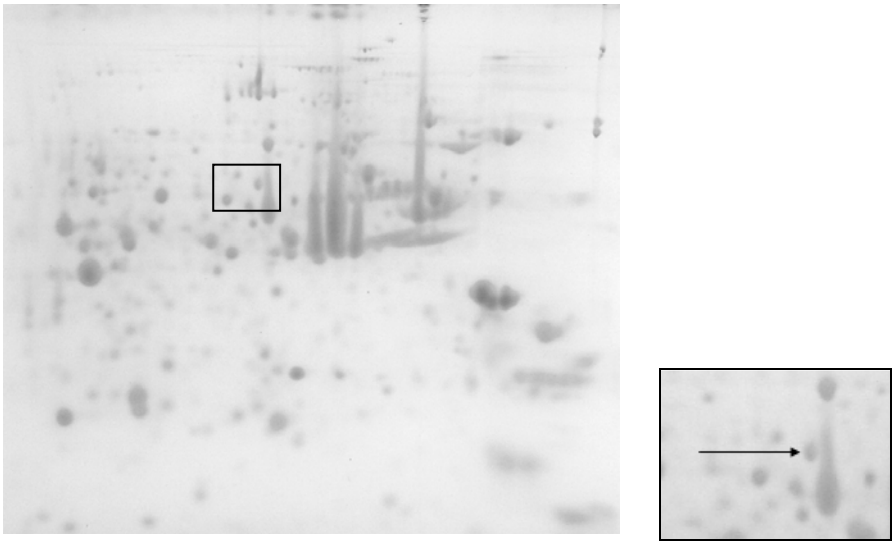
Proteomic Differences by Gender and Phenotype

Differences between male and female wild-type flies as well as wild-type and white-eyed mutants were observed. Some of these differences are shown in Figures 4.1 and 4.2 respectively. Spots found to be different between these populations were identified by MALDI-MS and MS/MS and peptide mass fingerprinting. The protein absent in white mutants matched both CG7445-PA and Flightin, both of whose function is unknown. Two proteins absent in wild-type male flies but present in females were identified as yolk protein 1 and CG16985. Yolk protein 1 is known to be involved in vitellogenesis and sex differentiation, while the function of CG16985 is unknown.

About twenty other, highly abundant proteins were identified to illustrate the range of function of this fraction of the proteome. A figure showing the position of each of these proteins in the gel is given in Figure 4.3, followed by Table 4.1 giving any available information about the roles of these proteins according to a widely recognized database of fruit fly proteins (FlyBase; http://flybase.bio.indiana.edu/).

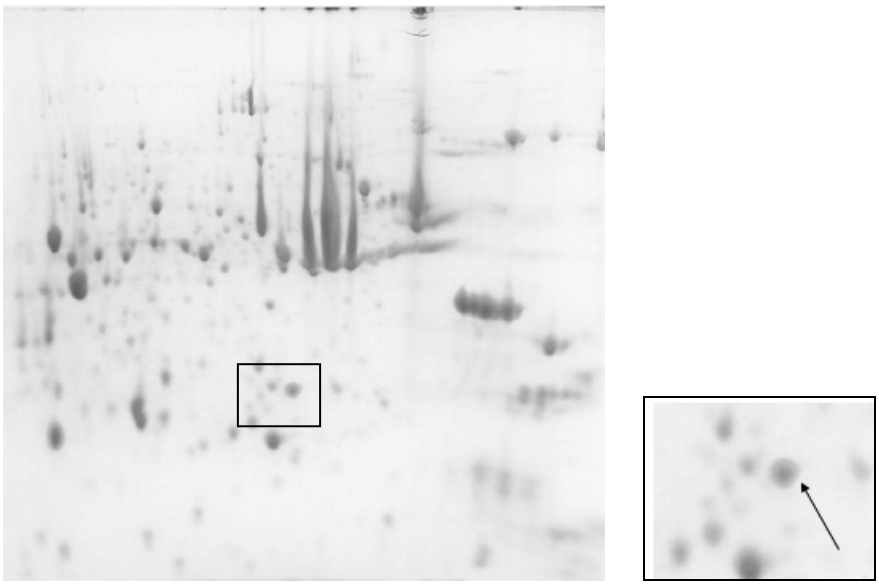


Male

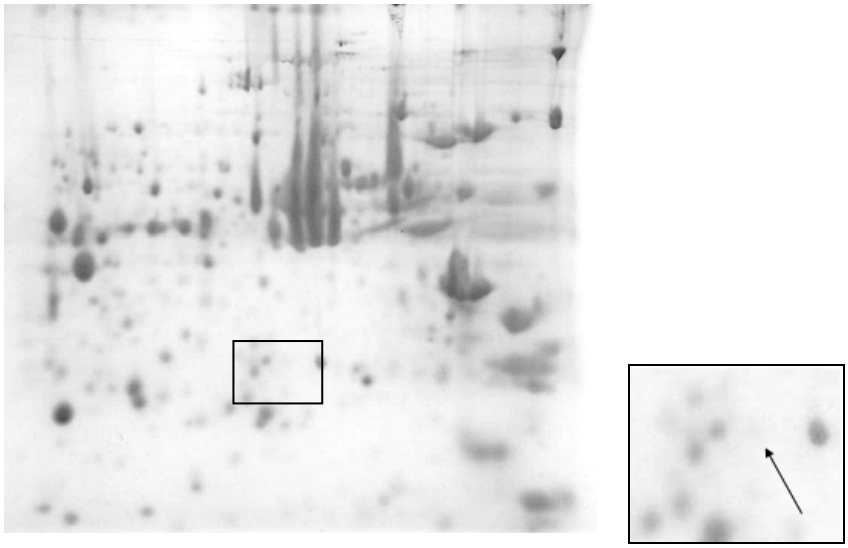


Female

Figure 4.1. Proteomic difference in wild-type male (above) and female (below) flies demonstrated by 2D PAGE. The boxed areas are enlarged to the right of each gel to highlight differences. The protein present in females and absent in males indicated by the arrow is CG16985.



Wild-type



White-eyed mutant

Figure 4.2. Proteomic difference in wild-type (above) and white-eyed (below) flies demonstrated by 2D PAGE. The boxed areas are enlarged to the right of each gel to highlight differences. The arrowed protein, present in wild-type and absent in the mutant, is CG7445-PA.

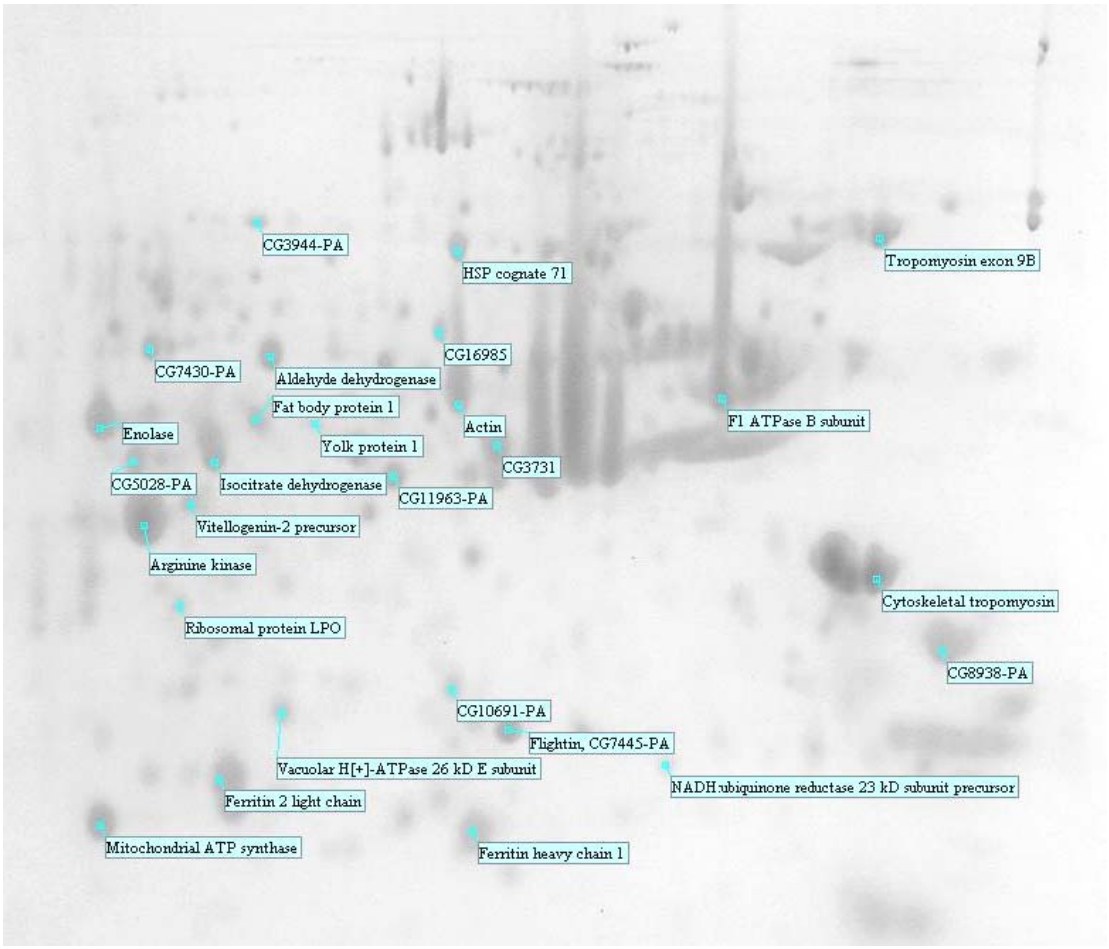


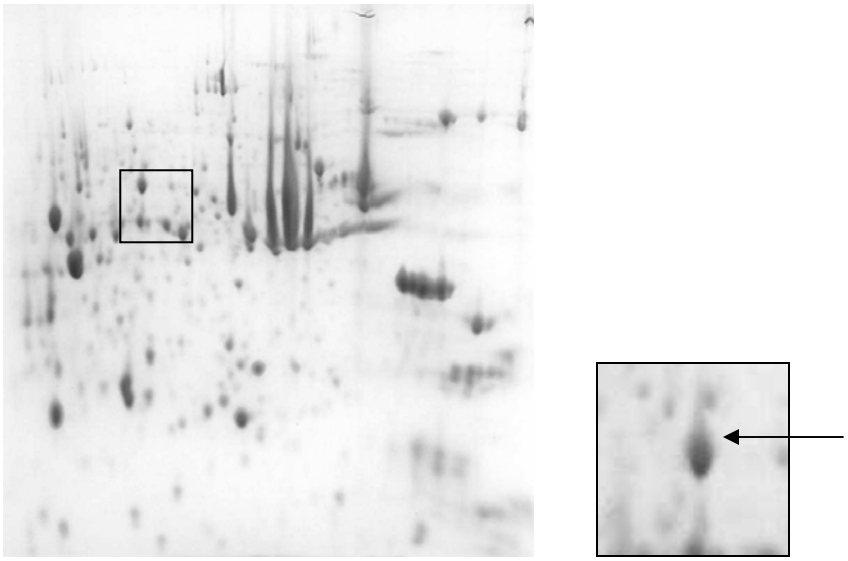
Figure 4.3. 2D gel of wild-type female fly proteins, with all identified proteins labeled. The known function of each protein is given in Table 4.1.

Protein	function
Actin	structural
Aldehyde dehydrogenase	aldehyde dehydrogenase
Arginine kinase	arginine kinase
CG16985	unknown
CG10691-PA	unknown
CG11963-PA	succinate-CoA ligase
CG3731	mitochondrial processing peptidase
CG3944-PA	NADH dehydrogenase
CG5028-PA	isocitrate dehydrogenase
CG7430-PA	dihydrolipoyl dehydrogenase
CG8938-PA	glutathione transferase
Cytoskeletal tropomyosin	actin binding
Enolase	phosphopyruvate hydratase
F1 ATPase B subunit	hydrogen-exporting ATPase
Fat body protein 1	protein, oxygen transporter
Ferritin 2 light chain	iron binding
Ferritin heavy chain 1	iron ion homeostasis
Flightin	unknown
HSP cognate 71	unfolded protein binding
Isocitrate dehydrogenase	Isocitrate dehydrogenase
Mitochondrial ATP synthase	hydrogen-exporting ATPase
NADH:ubiquinone reductase	NADH dehydrogenase
Ribosomal protein LPO	DNA lyase, ribosome structure
Tropomyosin exon 9B	actin binding
Vacuolar H[+]- ATPase	hydrogen-exporting ATPase
Vitellogenin-2 precursor	structural, catalytic
Yolk protein 1	sex differentiation and vitellogenesis

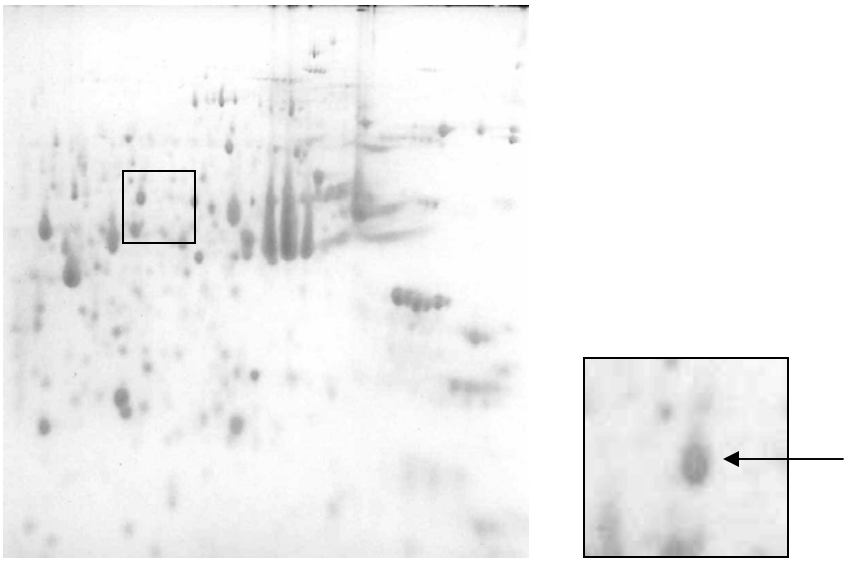
Table 4.1. Proteins identified by MS and peptide mass fingerprinting, and their respective function according to FlyBase at Indiana University.

The Hamburger Effect Studied by 2DGE

No proteomic differences were found between control male fly populations compared with those fed with 5% and 10% ethanol for either 12 or 24 hours prior to protein extraction. Figure 4.4 shows a 2D gel from a control population and that of a population fed with 10% ethanol for 24 hours. One of the proteins known to be involved in ethanol metabolism, aldehyde dehydrogenase [58], was among those proteins identified by MALDI-MS. Aldehyde dehydrogenase catalyzes the conversion of acetaldehyde to acetic acid in the second step of the conversion of ethanol to acetyl Co-A. This protein is highlighted in Figure 4.4. The expression of this protein in the control and ethanol-fed flies was analyzed specifically because of its role in ethanol metabolism. Figure 4.5 shows that there was no significant difference in its expression as measured by 2DGE. No difference was observed between the sample subjected to freeze/thaw cycles and its control (data not shown).



Control diet



Ethanol diet

Figure 4.4. 2D gels of protein extracted from control and ethanol-fed male fly populations. The control fly gel is above, and that of the ethanol-fed fly proteins is below. Aldehyde dehydrogenase is indicated by the arrow in the enlarged rectangle to the right of each gel.

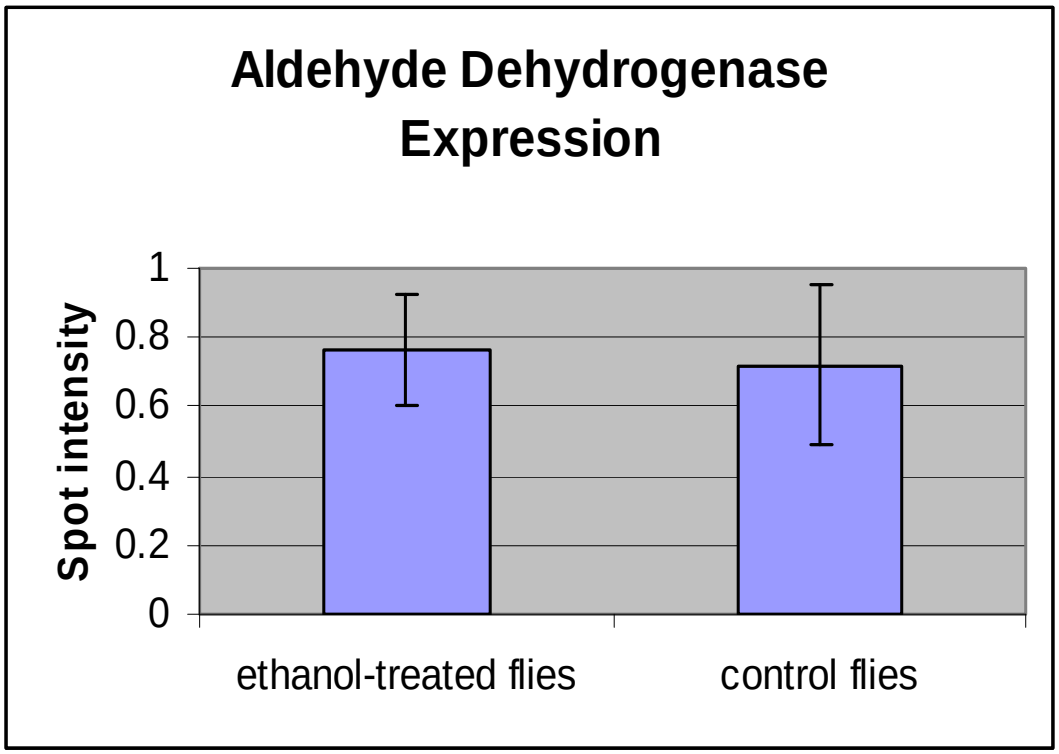


Figure 4.5. Aldehyde dehydrogenase abundance in control vs ethanol-fed fruit flies +/- 1 standard deviation; n = 4 for each group.

Discussion

The 2DGE studies demonstrated gender- and other fixed, gene-based proteomic differences, but failed to detect dynamic differences caused by one type of change in diet. The methods, software and reproducibility in this study were sufficient to visualize differences between the genders and between wild-type and white-eyed mutant flies. Had a more sensitive stain been used, perhaps more differences could have been found between these populations known to be different. Also, given that background variability in this method is close to 20% (see Chapter 2), subtle differences in protein abundance may be obscured.

The *white* gene, CG2759, which is responsible for pigmentation of the eye, was not located on the gels. This protein may be present and simply non-functional in the mutant populations. There may be a number of genetic differences between the two groups which could account for the reported proteomic difference.

With respect to the hamburger effect, the degree of sensitivity of these methods must be considered. Only the most abundant proteins were detected. Several of the most abundant proteins which were identified are housekeeping or structural. The abundance of these and other highly abundant proteins is not likely to change significantly over a short time. Although no hamburger effect was detected using these methods, given the level of variability of the methods, a significant hamburger effect particularly among less-abundant proteins cannot be ruled out.

The Albrethsen and Hsieh studies both reported that no difference in protein abundance as measured by SELDI was observed between male and female human serum

using their approach [31, 41]. The current study did show significant differences of the male and female fly proteome using 2DGE. Reasons why differences in gender were noted in this study and not in the others could be because of the larger number of molecular species analyzed in these studies than were analyzed in the others. The approach used here also undoubtedly samples a different fraction of the proteome than does SELDI-MS. It appears human serum as sampled by SELDI-MS does not contain as many abundant peptides which are gender-specific as whole fly homogenate as measured by 2DGE.

Although fruit fly proteomes were studied for maximum control of the environment of the animal as well as unlimited sample availability, human tissues would be more clinically relevant to biomarker discovery. It should be recognized that the hamburger effect on protein abundance in a specific compartment or tissue such as serum may be very different from the effect on a whole organism. The blood of a mammal, for instance, is more proximally linked to the diet than are perhaps most other tissues. Mammalian serum or plasma proteomes may therefore be more sensitive to a hamburger effect than whole fruit fly proteomes or parts of such proteomes.

Conclusion

The 2DGE methods are sufficient to detect differences between male and female fly proteins and between wild-type and white-eyed mutant fly proteins. A dietary variable was used to evaluate whether changes in the fruit fly proteome due to a specific change in the diet could be observed by 2DGE. No differences due to this change were documented. Freeze-thaw treatment did not seem to affect the proteome sampled here.

While the hamburger effect may be significant, the results of these methods did not provide evidence of its causing major distortions in the fruit fly proteome sampled.

CHAPTER 5

OBSERVATION OF PROTEOMIC DIFFERENCES BETWEEN POPULATIONS OF FRUIT FLIES, AND EVALUATION OF THE HAMBURGER EFFECT USING LCMS

Overview

As a follow-up to Chapter 4, LCMS was used to assess the hamburger effect in the fruit fly proteome. LCMS spectra of protein extracts from wild-type flies given an ethanol diet or subjected to starvation were compared to control spectra. No differences were found between the spectra of ethanol-fed flies and control flies, consistent with 2DGE findings. Quantitative differences between starved and control flies were found. The hamburger effect under starvation conditions was small, representing only a small percent of the proteome sampled by this method. Three peaks (representing two peptides) out of about one thousand analyzed peaks were different between the starved and fed populations, suggesting that the hamburger effect is modest in this system and does not markedly compromise proteomic biomarker discovery.

Introduction

In Chapter 4, 2DGE was unsuccessful in detecting a hamburger effect among fruit fly proteins due to ethanol supplementation using this approach. Using LCMS as an alternative proteomics platform may yield different results. LCMS was used to compare proteomes undergoing two dietary maneuvers. In one set of experiments the ethanol experiments were repeated. In a second set of experiments wild-type flies underwent 24

hours of food deprivation. Flies fed ethanol or starved were compared to control populations. Preparation methods for LCMS were different from 2DGE, so a different cross-section of the proteome was analyzed by LCMS than by 2DGE. Most proteins are removed by ACN precipitation. Owing to the greater sensitivity of the method and depletion of large proteins, peptides and small proteins were targeted by LCMS. There was evidence of greater variability among less-abundant proteins than among highly abundant proteins [59]. Analysis of another cross-section of the proteome using LCMS would then serve to strengthen the conclusions of the previous study, or provide evidence of a significant hamburger effect in the fruit fly.

Materials and Methods

Protein extractions, sample preparation, and LCMS were performed as described in Chapter 2. *Drosophila* culture was also performed as described in Chapter 2 except where specified otherwise.

Processing of Data for Comparison of Spectra

In order to best compare the spectra visually, it was necessary to break the data into manageable sized parts. A spectrum compiling two minutes of data from the total cLC run was empirically determined to be sufficiently simplified for analysis and comparison.

Reference peaks present in the chromatogram were chosen to normalize two-minute windows with respect to elution time. Six reference peaks were chosen which

met the following criteria: each had signal at least 5 times the level of noise, could be extracted from the spectrum as a fully resolved peak with a precise time of elution, and had an average elution time roughly two minutes from the neighboring two reference peaks. The elution time of each reference peak for each sample was recorded. Two-minute windows of each spectrum were created around each reference peak which could be used for comparison. There were roughly 180 peaks with intensity greater than or equal to twice the level of the noise in each window. Using the 6 reference peaks, a total of about 1000 peaks could be manually screened for differences between ethanol fed and control or between starved and control fly samples.

Although the LC gradient was applied over 40 minutes, the six two-minute windows of elution included most of the useful data. The first 5 to 6 minutes of the LC gradient as assessed by the total ion chromatogram contained little or no eluted material and represent the void volume of the column and any solvent delay from the LC pumps to the column and the mass spectrometer. As the gradient shifts toward 50% organic solvent composition, the total ion chromatogram count reaches into the hundreds of thousands of ions. The last 8 to 10 minutes of the gradient, where the mobile phase is of high organic solvent concentration is a wash step to purge the column of accumulated contaminants. This highly hydrophobic material does not represent chromatographic separation and was not included in analysis.

Comparison of Male and Female Fly LCMS Spectra

To verify that the LCMS methods are capable of identification of fixed, genotypic proteomic differences, male and female fly populations were compared by LCMS. Peptides from eight populations of male and eight populations of female flies were analyzed by LCMS. Three windows of spectra (as outlined previously) were overlaid using Analyst QS 1.1, with spectra from each gender represented by a different color. This allowed spectra from each gender to be distinguished visually. The spectra were searched for peak intensities appearing to be different between the populations. Student's T-test analyses were used to assess statistical significance of those peaks that appeared to be present in markedly different amounts.

Study of Hamburger Effect in Fruit Fly Populations Via LCMS

Populations of flies were given modified diets as in Chapter 4 to determine whether a hamburger effect was found using the LCMS approach. In addition to control flies, other populations were temporarily starved as a means of imposing an extreme dietary change. Control fly populations were fed the standard diet and transferred to another vial 24 hours prior to protein extraction, whereas starved flies or ethanol-fed flies were transferred to empty vials (starved) or vials prepared with 10% of the volume of water replaced with ethanol as in Chapter 4. Both control fly samples and alternate diet fly samples were included in the same run and they were ordered so that control specimens were run between starved samples to avoid bias from any drifting of the sensitivity of the instrument or other time-dependent trends.

Analyst QS 1.1 software was used to overlay spectra as before. Control specimens were denoted by one color and the comparison group was given a second color allowing individual and group intensities to be observed. Student's T test was applied to evaluate the statistical significance of those peaks that looked different, using Excel software.

Results

More than 20 marked quantitative differences between the male and female fly proteomes were found by comparison of the MS spectra. There were at least 10 peptides present only in female proteomes and at least 10 peptides specific to males within only 3 2-minute windows. This search was not meant to be exhaustive, but to serve as a positive control in showing that obvious proteomic differences may be observed between male and female flies using these methods. Three representative isotope peak families showing clear differences are shown in Figure 5.1 as examples. Two of these three peaks represent the same peptide. These peptides because of their sizes have not lent themselves to straightforward peptide identification and are under investigation. The P values of T-tests comparing the intensities of the representative peaks were 0.0003 and 0.0002 for the peaks at 627 and 1090, respectively, with n = 4.

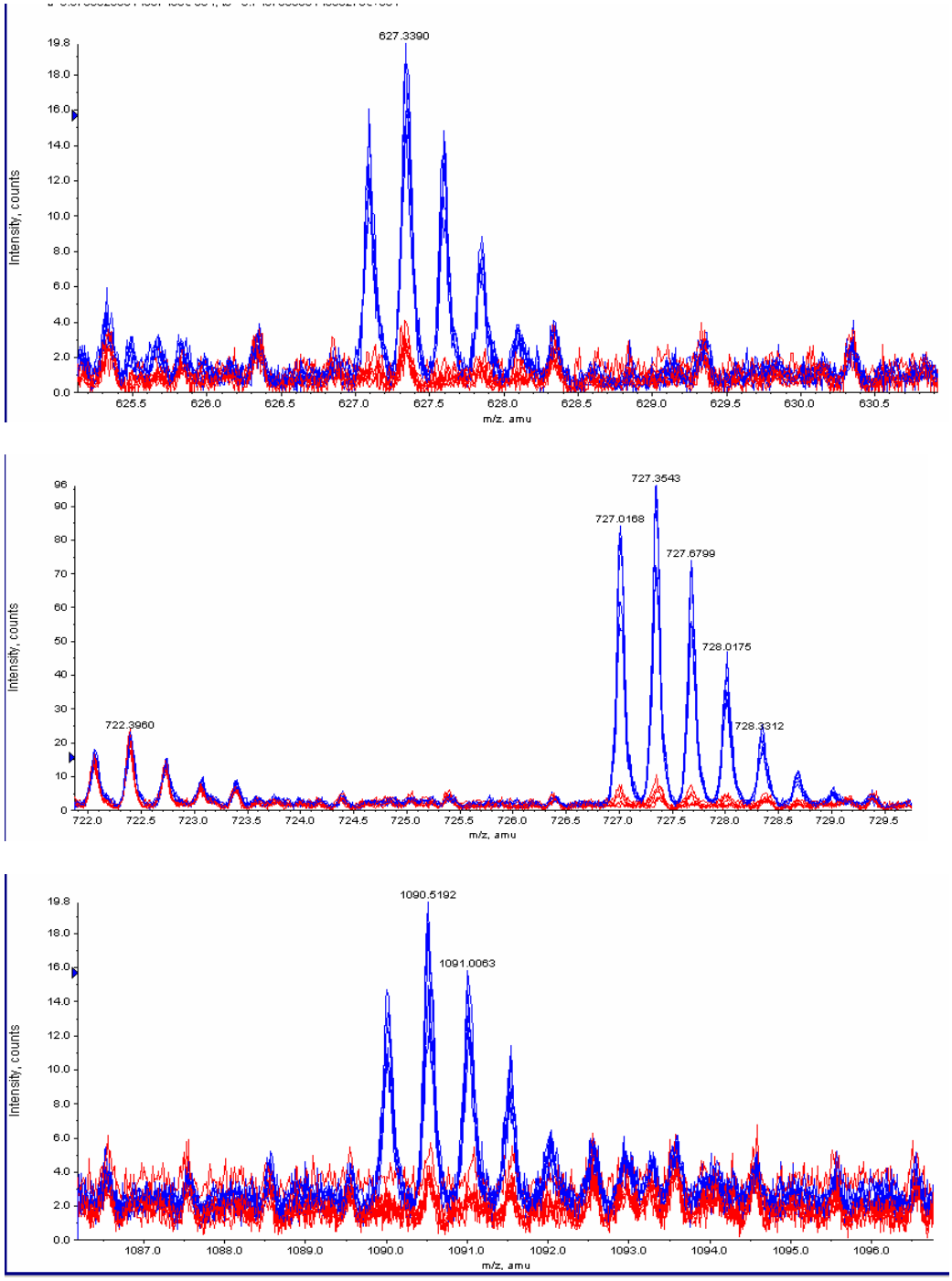


Figure 5.1. Difference between male fly peptides (blue) and female fly peptides (red) observed at 627, 727, and 1090 m/z, with monoisotopic masses of 2504, 2178, and 2178, respectively. P-values were 0.003, 0.003, and 0.004 respectively with n = 4.

In another set of experiments testing the hamburger effect, ethanol was added to the diet. However, this particular dietary change did not appear to change the proteome quantitatively as observed by LCMS (data not shown). As before, about 1000 peaks were observable in LCMS spectra, which had been color coded by group and overlaid manually. Although a few species appeared to be quantitatively lower in the ethanol-treated group, analysis of intensity of each of these peaks followed by statistical analysis demonstrated no significant differences (data not shown).

In the final set of experiments, a second dietary intervention was used. In response to this diet change, differences were observed between spectra from starved flies compared with control flies. At least two peptides were found to be significantly reduced in starved flies. Overlaid spectra of control and starved male flies, coded by color, are shown in Figures 5.2 and 5.3. The P values were 5.3E-7 and 6.0E-4 for the peaks at 852.5 and 880.5, respectively, comparing ten spectra from each group of flies. Figure 5.4 and Table 5.1 serve to illustrate these differences.

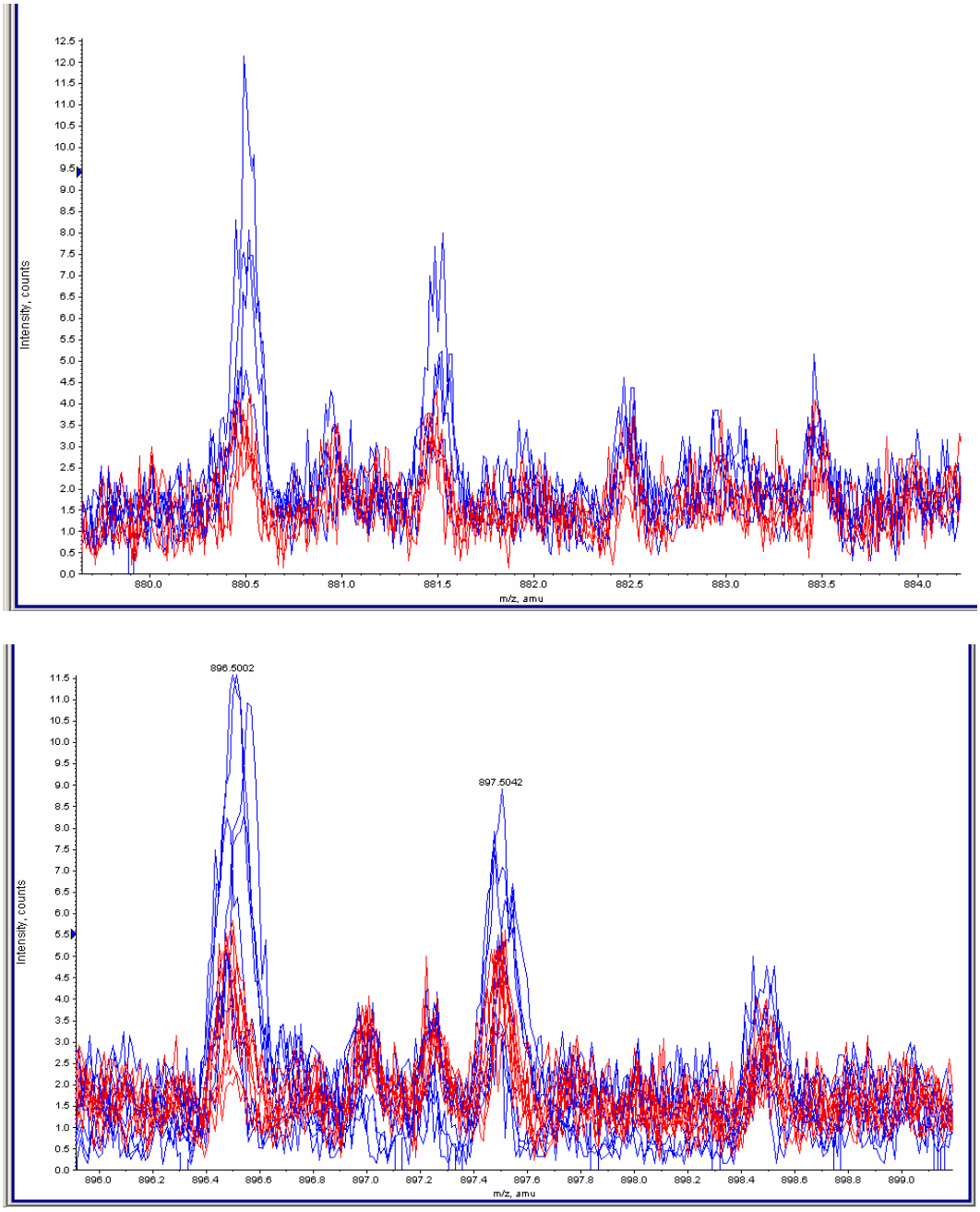


Figure 5.2. Difference in isotopic peak families between control fly peptide (blue) and starved fly peptide (red) observed at 880 m/z (above) and 896 m/z (below) with monoisotopic masses of 879.5 and 895.5, respectively. These isotope peak families most likely represent the same peptide.

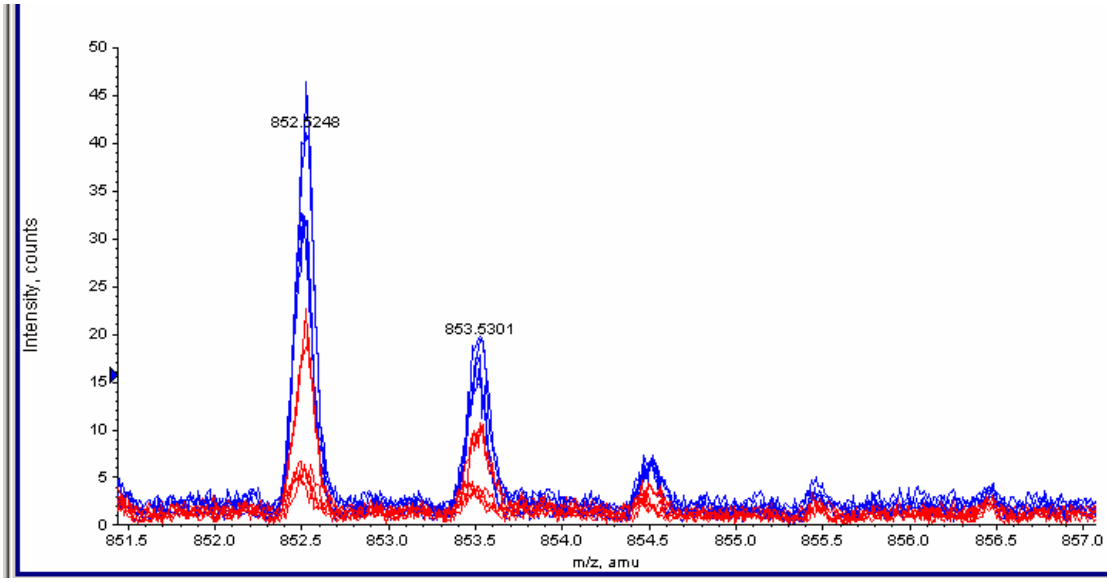
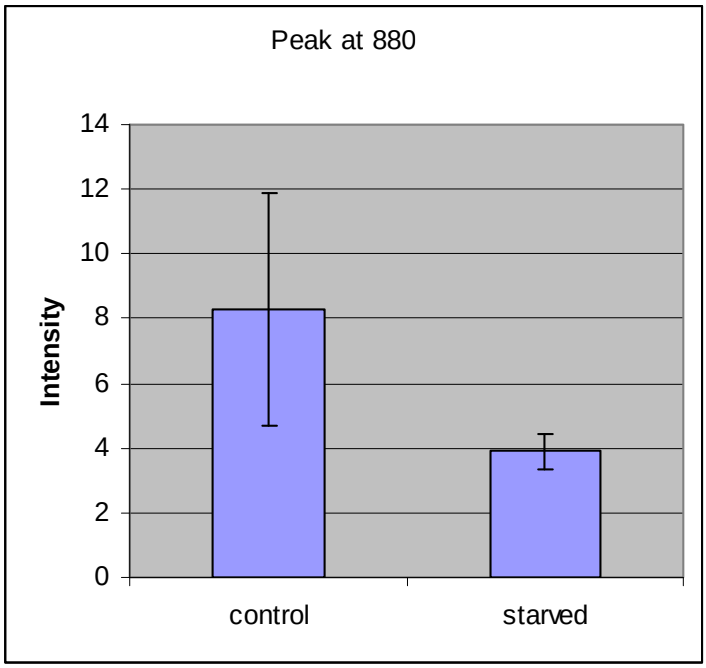


Figure 5.3. Difference in isotopic peak families between control fly peptide (blue) and starved fly peptide (red) observed at 852 m/z, monoisotopic mass = 851.5.



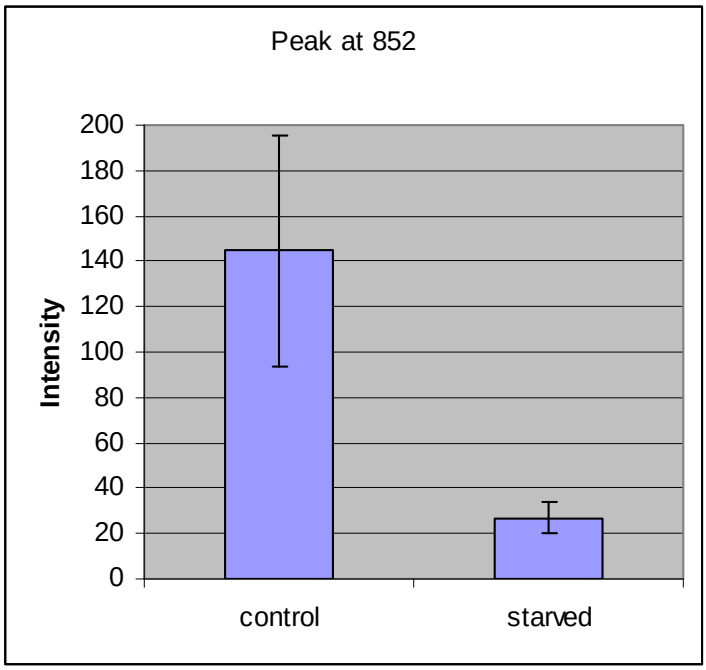


Figure 5.4. Bar graphs showing differences in intensity of peaks at 880 (above) and 852 (below) for fed control flies and starved flies +/- s.d.; n = 10 for each group. P values were both statistically significant, and are given in Table 5.1.

Control	Int 852	Int 880		Starved	Int 852	Int 880
1	101	6		1	14.2	3
2	127	15.2		2	27.6	3.77
3	85.8	12.2		3	32.2	3.85
4	91.5	8.08		4	18	4.46
5	191	4.92		5	37.7	4.08
6	191	4.85		6	32.3	4.77
7	154	6.92		7	23.8	3.17
8	93.7	4.77		8	24.8	3.54
9	188	8.58		9	27.5	4.23
10	223	11.3		10	32.5	4.08
Mean	144.6	8.282		Mean	27.06	3.895
SD	51.10	3.573		SD	7.152	0.551
RSD	0.353	0.431		RSD	0.264	0.142
T test	5.28E-07	0.0006				
81% decrease in average intensity of peak at 852						
53% decrease in avg int of peak at 880						

Table 5.1. Mean intensity, standard deviation (SD) and relative standard deviation (RSD) for peaks at 880 and 852 for control fed flies and starved flies with P values of 0.0006 and 0.0000005 respectively. Statistical differences were significant between the two groups. Starvation for 24 hours on average decreased the peptide at 880 m/z by 53% and the peptide at 852 m/z by 81%.

Discussion

Ethanol treatment for 24 hours did not cause any significant or obvious differences in the few hundred small proteins and peptides analyzed in this study by LCMS. However, starvation was shown to cause significant changes in the fruit fly proteome, including significant reductions in at least two molecular species. These differences were visually obvious. Had a more thorough evaluation of each peak been conducted, perhaps more species would have been found to be significantly but less dramatically different. Some of the limitations mentioned earlier also apply, e.g., relatively small differences in expression of peptides may be obscured given the level of background variability of the methods.

Abundance of one peptide with m/z 852 was decreased, and the second peptide with m/z signals at 880 and 896 was decreased. The peaks at m/z 880 and 896 are likely to be related because mass difference of 16 amu often represents oxidation. The identity of these peptides proved complicated and is currently under investigation.

Conclusions

Three of the approximately 1000 viewed molecular species in the spectra were shown to be significantly different in the starved fly populations. This study suggests that the influence of even an extreme diet change on the variability of the abundant proteome or peptidome is modest, and that the abundant proteome of a whole organism is generally unchanged by short-term maneuvers and yet consistent enough to allow detection of modest to substantial modification. This suggests that this proteomic

approach can be used for biomarker discovery with minimal interference of short-term "environmental" changes such as diet.

CHAPTER 6

CRITIQUE OF CURRENT EXPERIMENTS AND FUTURE EXPERIMENTS

The research presented in this thesis was meant to compare the overall reproducibility of LCMS and 2DGE, to evaluate factors that may contribute to variability of LCMS spectra, and to assess the hamburger effect in a portion of the fruit fly proteome via 2DGE and LCMS. There are several limitations to the approach currently used, and some alternative or additional experiments which may be considered to build on this work.

Limitations

Large amounts of data are typical of proteomic analysis. Several thousands of proteins can be detected and interrogated in a single study [54, 60]. Our proteomic sampling was not quite on this order for the present study because only a small cross-section of the proteome was explored. Also, due to logistical limitations, specimens available and experiments performed were fewer than would be desired. Perhaps the smaller differences could have been detected and confirmed by statistics had more proteins been detected and analyzed or more replicates run. After looking at the hamburger effect by LCMS in starved fruit flies, only two of a thousand proteins were shown to be different. With a larger sample of proteins measured, the proteome would be better represented, and a more complete characterization of the hamburger effect in this system could be performed.

Perhaps a more significant limitation of current work is the inherent insensitivity of the methods. Only relatively abundant, soluble proteins were analyzed. Both 2DGE and LCMS fail to detect most of the proteins in the proteome. Although data in Chapter 3 do not establish a link between protein abundance and variability, some lower-abundance proteins, and perhaps those that are less-soluble, may change more drastically than those measured by our approach in this study. If more of the variability of the proteome occurs among these proteins, the hamburger effect may be more influential than reported.

The proteomes of different organisms are as different as the organisms themselves. Proteomics work with tissue and organelles is also different than proteomics of entire organisms. It is not certain that the hamburger effect in many proteomes of interest, such as the human proteome, or that of human serum, is the same as suggested by these data gathered from *Drosophila* proteins. Although comparable estimates of variability of other proteomic methods for other proteomes have been reported in Chapter 2 and are similar to these results, it would be valuable to duplicate the methods from this study with a different, perhaps more clinically relevant proteome, e.g., human blood serum. Bias could be reduced if the same equipment and reagents were used across studies.

Future work

Future work to study the reproducibility of proteomic methods could include identification of many or all of the randomly selected proteins used in this study. Protein identification in a previous methods reproducibility study gave no recognized trends in variability of human liver proteins [47]. If the identities of the proteins in this study were available, the data sets could be more easily compared, which might provide information on the compatibility of proteomics data across species. Additionally, increased variability of certain groups of proteins, e.g., liver enzymes, may be noticed.

Other important follow-up studies could involve, as mentioned before, a larger sample size of randomly selected proteins from 2DGE and LCMS grouped into more diverse categories. Proteins could be grouped according to molecular function, amino acid composition, or localization in the cell. This would involve a significant amount of work in identification and categorization of the proteins, but may be fruitful in determining which types of proteins are more susceptible to variability.

The identities of the proteins affected by starvation experiments remain unknown. This information certainly would add value to the study of this dietary maneuver to determine whether there is a logical connection between the physiological stimulus and the affected proteins. If a connection between starvation and the reduced protein is not immediately obvious, it could be interpreted that the hamburger effect may be indirect and complex. If the stimulus and response are related as part of a metabolic pathway or counter-regulatory pathway already understood, it would support the performance of the methods and imply a direct model for the hamburger effect.

Appendix 1. Data used in analysis of 2DGE reproducibility. The values representing protein abundance in 196 spots matched between a single gel and a synthetic gel average were normalized and recorded as listed below. The ratio for each pair, used in the T-test, as well as percent difference, used to construct Table 2.1, is also given.

	Protein Abundance			
Spot	Representative Gel	Synthetic Gel	Ratio (used in T test)	Percent Difference
1	0.2220	0.2011	1.1036	10.36%
2	0.0534	0.0709	0.7539	24.61%
3	0.1732	0.2603	0.6655	33.45%
4	0.6953	1.2259	0.5671	43.29%
5	0.4321	0.3559	1.2140	21.40%
6	0.2361	0.2819	0.8375	16.25%
7	0.2903	0.3942	0.7365	26.35%
8	0.0110	0.0113	0.9746	2.54%
9	0.2313	0.2283	1.0130	1.30%
10	0.4521	0.5942	0.7609	23.91%
11	0.2371	0.2428	0.9764	2.36%
12	1.9646	2.7342	0.7185	28.15%
13	0.2823	0.2902	0.9728	2.72%
14	0.2973	0.2849	1.0436	4.36%
15	1.9290	2.3102	0.8350	16.50%
16	1.2086	1.1574	1.0442	4.42%
17	1.3099	0.9083	1.4422	44.22%
18	0.3409	0.4210	0.8098	19.02%
19	0.9039	0.8069	1.1202	12.02%
20	0.5553	0.6431	0.8635	13.65%
21	0.4063	0.5231	0.7768	22.32%
22	1.4531	1.0392	1.3983	39.83%
23	0.2174	0.2353	0.9236	7.64%
24	0.2016	0.1789	1.1266	12.66%
25	0.6590	0.8144	0.8092	19.08%
26	0.4031	0.3664	1.1002	10.02%
27	0.2075	0.2637	0.7869	21.31%
28	0.0900	0.0749	1.2019	20.19%

29	0.1562	0.1411	1.1069	10.69%
30	0.1321	0.1322	0.9993	0.07%
31	0.3261	0.4137	0.7883	21.17%
32	0.6068	0.5902	1.0281	2.81%
33	0.3787	0.4162	0.9100	9.00%
34	3.5636	3.7350	0.9541	4.59%
35	0.1949	0.1874	1.0398	3.98%
36	0.1273	0.1235	1.0309	3.09%
37	0.1265	0.1500	0.8436	15.64%
38	0.3579	0.2703	1.3241	32.41%
39	0.3907	0.3855	1.0136	1.36%
40	0.1112	0.1110	1.0011	0.11%
41	0.1066	0.1094	0.9746	2.54%
42	0.2206	0.2788	0.7912	20.88%
43	0.1451	0.1468	0.9886	1.14%
44	1.1008	1.2447	0.8844	11.56%
45	0.1368	0.1404	0.9746	2.54%
46	0.2744	0.3208	0.8553	14.47%
47	0.4763	0.4719	1.0092	0.92%
48	0.2037	0.1621	1.2562	25.62%
49	1.0903	1.6964	0.6427	35.73%
50	0.6320	0.7064	0.8947	10.53%
51	0.1336	0.1146	1.1654	16.54%
52	0.3798	0.3281	1.1575	15.75%
53	0.2439	0.2118	1.1516	15.16%
54	0.1777	0.1976	0.8996	10.04%
55	0.2951	0.2442	1.2085	20.85%
56	0.3366	0.3954	0.8512	14.88%
57	0.1968	0.1757	1.1202	12.02%
58	2.0537	1.6267	1.2625	26.25%
59	0.1015	0.0770	1.3181	31.81%
60	0.0534	0.0548	0.9746	2.54%
61	2.1684	2.4589	0.8819	11.81%
62	3.0484	3.1328	0.9731	2.69%
63	0.2232	0.2148	1.0387	3.87%
64	0.1524	0.1492	1.0212	2.12%
65	0.0938	0.1324	0.7085	29.15%
66	0.1250	0.0761	1.6423	64.23%
67	0.1831	0.2104	0.8703	12.97%
68	0.3260	0.2827	1.1531	15.31%
69	0.5109	0.3950	1.2932	29.32%
70	3.9379	3.7265	1.0567	5.67%

71	0.2341	0.1224	1.9126	91.26%
72	0.5275	0.6763	0.7799	22.01%
73	0.2563	0.2294	1.1172	11.72%
74	0.0972	0.0825	1.1774	17.74%
75	2.3111	2.6373	0.8763	12.37%
76	0.5275	0.7783	0.6778	32.22%
77	0.7198	0.6186	1.1636	16.36%
78	0.3030	0.4260	0.7113	28.87%
79	0.5777	0.5873	0.9836	1.64%
80	2.4254	2.4760	0.9796	2.04%
81	2.2238	2.4373	0.9124	8.76%
82	0.5021	0.3277	1.5320	53.20%
83	0.2815	0.2288	1.2299	22.99%
84	0.5065	0.5897	0.8588	14.12%
85	0.4420	0.5149	0.8584	14.16%
86	0.8518	1.0494	0.8117	18.83%
87	0.9400	0.7163	1.3122	31.22%
88	1.4313	1.9356	0.7395	26.05%
89	2.0099	2.5133	0.7997	20.03%
90	1.4409	1.0467	1.3766	37.66%
91	0.7725	0.8900	0.8680	13.20%
92	1.9208	2.1403	0.8974	10.26%
93	0.2559	0.1916	1.3353	33.53%
94	0.1630	0.2662	0.6123	38.77%
95	0.0707	0.0573	1.2331	23.31%
96	0.4589	0.2957	1.5517	55.17%
97	2.2686	1.4993	1.5131	51.31%
98	0.1710	0.2221	0.7699	23.01%
99	0.1031	0.1264	0.8160	18.40%
100	0.1257	0.1480	0.8497	15.03%
101	0.1229	0.0858	1.4332	43.32%
102	0.1060	0.0920	1.1516	15.16%
103	0.1296	0.0935	1.3864	38.64%
104	0.1238	0.1080	1.1463	14.63%
105	1.2201	1.8106	0.6739	32.61%
106	0.1540	0.1689	0.9115	8.85%
107	0.1956	0.2141	0.9138	8.62%
108	1.7811	1.4888	1.1963	19.63%
109	0.0936	0.0938	0.9980	0.20%
110	3.9814	2.3526	1.6924	69.24%
111	1.8827	2.4965	0.7542	24.58%
112	0.2610	0.3235	0.8070	19.30%

113	0.6891	0.6982	0.9870	1.30%
114	0.2692	0.4389	0.6134	38.66%
115	0.1980	0.2031	0.9746	2.54%
116	0.5491	0.6574	0.8353	16.47%
117	0.4445	0.6146	0.7231	27.69%
118	0.1473	0.1593	0.9246	7.54%
119	0.1177	0.1237	0.9510	4.90%
120	0.2322	0.2809	0.8266	17.34%
121	0.4459	0.3416	1.3052	30.52%
122	0.3230	0.3706	0.8715	12.85%
123	0.0809	0.1271	0.6365	36.35%
124	0.0859	0.0531	1.6185	61.85%
125	0.1529	0.2085	0.7333	26.67%
126	0.3318	0.4285	0.7744	22.56%
127	0.1757	0.2928	0.6002	39.98%
128	0.2031	0.2226	0.9125	8.75%
129	0.0848	0.0770	1.1008	10.08%
130	0.1212	0.1794	0.6754	32.46%
131	0.1166	0.0989	1.1786	17.86%
132	0.0942	0.0966	0.9746	2.54%
133	0.1098	0.1512	0.7260	27.40%
134	0.0846	0.0624	1.3558	35.58%
135	0.1619	0.1420	1.1399	13.99%
136	0.1067	0.0852	1.2516	25.16%
137	0.7830	0.8210	0.9538	4.62%
138	0.1663	0.1118	1.4871	48.71%
139	0.2771	0.2458	1.1270	12.70%
140	0.1071	0.1054	1.0159	1.59%
141	0.0718	0.0757	0.9491	5.09%
142	0.1238	0.1536	0.8060	19.40%
143	0.2681	0.2244	1.1948	19.48%
144	0.4176	0.4285	0.9746	2.54%
145	0.1785	0.1195	1.4931	49.31%
146	0.2636	0.2795	0.9434	5.66%
147	0.0803	0.0824	0.9746	2.54%
148	0.3888	0.2430	1.5997	59.97%
149	0.1377	0.1250	1.1022	10.22%
150	0.3767	0.3757	1.0027	0.27%
151	0.0741	0.0970	0.7641	23.59%
152	0.2828	0.2744	1.0307	3.07%
153	0.0920	0.1472	0.6253	37.47%
154	0.2964	0.2449	1.2102	21.02%

155	0.1185	0.1216	0.9746	2.54%
156	0.0909	0.0755	1.2040	20.40%
157	0.0989	0.1425	0.6938	30.62%
158	0.4347	0.4322	1.0057	0.57%
159	0.0673	0.0691	0.9746	2.54%
160	0.4520	0.3249	1.3914	39.14%
161	0.4019	0.5400	0.7444	25.56%
162	0.1884	0.2229	0.8452	15.48%
163	0.0970	0.1203	0.8067	19.33%
164	0.1926	0.2993	0.6435	35.65%
165	0.0786	0.0442	1.7756	77.56%
166	0.3895	0.3826	1.0179	1.79%
167	0.1688	0.1731	0.9746	2.54%
168	0.0368	0.0401	0.9178	8.22%
169	0.1720	0.1336	1.2871	28.71%
170	0.1326	0.1608	0.8245	17.55%
171	0.0780	0.0686	1.1384	13.84%
172	0.0704	0.0835	0.8427	15.73%
173	0.0175	0.0179	0.9746	2.54%
174	0.0941	0.1031	0.9120	8.80%
175	0.0457	0.0491	0.9298	7.02%
176	0.1117	0.1146	0.9746	2.54%
177	0.2425	0.2674	0.9072	9.28%
178	0.0801	0.1147	0.6983	30.17%
179	0.0665	0.0840	0.7919	20.81%
180	0.1251	0.1131	1.1064	10.64%
181	0.1001	0.1150	0.8700	13.00%
182	0.0950	0.1153	0.8239	17.61%
183	0.0318	0.0327	0.9746	2.54%
184	0.0154	0.0158	0.9746	2.54%
185	0.1049	0.1219	0.8607	13.93%
186	0.1997	0.1957	1.0202	2.02%
187	0.0485	0.0493	0.9847	1.53%
188	0.0556	0.0633	0.8789	12.11%
189	0.0226	0.0298	0.7577	24.23%
190	0.0873	0.0880	0.9914	0.86%
191	0.0410	0.0421	0.9746	2.54%
192	0.0605	0.1159	0.5224	47.76%
193	0.0937	0.0921	1.0180	1.80%
194	0.1274	0.1181	1.0781	7.81%
195	0.1586	0.1346	1.1785	17.85%
196	0.0954	0.1132	0.8427	15.73%

Appendix 2. Data used in analysis of LCMS reproducibility. For each of the 36 peaks, the average peak height from 10 spectra was compared to the height of the corresponding peak from a representative sample. The ratio for each pair, used in the T-test, as well as percent difference, used to construct Table 2.2, is also given.

	Peak Intensity			
Peak	Representative spectrum	Average	Ratio (used in T test)	Percent Difference
1	7.54	7.975	0.945	5.45%
2	41.7	63.24	0.659	34.06%
3	9.98	8.914	1.12	11.96%
4	8.56	14.74	0.581	41.91%
5	4.96	11.1	0.447	55.32%
6	18.9	17.09	1.106	10.59%
7	5.66	6.409	0.883	11.69%
8	42.3	33.88	1.249	24.85%
9	9.2	7.621	1.207	20.72%
10	5.74	4.627	1.241	24.05%
11	9.58	8.043	1.191	19.11%
12	7.39	6.267	1.179	17.92%
13	7.31	8.378	0.873	12.75%
14	14.1	15.49	0.91	8.97%
15	31	38	0.816	18.42%
16	11.7	10.35	1.13	13.03%
17	7.7	9.397	0.819	18.06%
18	5.81	5.455	1.065	6.51%
19	19.9	15.15	1.314	31.35%
20	16.9	19.48	0.868	13.24%
21	15.4	14.91	1.033	3.29%
22	56.5	45.12	1.252	25.22%
23	20.2	18.22	1.109	10.88%
24	20.7	20.11	1.029	2.93%
25	12.5	13.47	0.928	7.20%
26	12.8	15.65	0.818	18.21%

27	23.3	20.32	1.147	14.67%
28	23	23.98	0.959	4.09%
29	9.9	11.16	0.887	11.27%
30	26.6	25.25	1.053	5.35%
31	6.13	5.908	1.038	3.76%
32	7.79	7.581	1.028	2.76%
33	49.2	44.13	1.115	11.49%
34	14.7	14.19	1.036	3.59%
35	84.3	84.32	1	0.02%
36	12.9	13.8	0.935	6.49%

Appendix 3. Data used in evaluation of LCMS reproducibility. Intensities of 36 selected representative peaks of different m/z range and elution time for 10 spectra, and the coefficient of variation (CV) for each data set.

	Early Elution							
	Lower m/z				**Intermediate m/z**			
	Peak1	*Peak2*	*Peak3*	*Peak4*	*Peak5*	*Peak6*	*Peak7*	*Peak8*
Sample1	5.37	72.9	5.29	16.4	11.3	14.2	6.16	38.2
Sample2	6.29	96.7	4.81	22.5	10.4	16.3	5.02	40.5
Sample3	6.06	93.6	7.58	24.5	11.4	16.6	6.57	31.9
Sample4	7.21	81.7	9.37	21.2	8.94	17.6	7.14	26.5
Sample5	7.54	41.7	9.98	8.56	4.96	18.9	5.66	42.3
Sample6	8.77	32	9.93	8.95	6.82	21	6.74	32.5
Sample7	10.1	48.8	8.96	14.3	7.9	16.7	5.54	31.2
Sample8	11	35.2	9.58	9.49	16.8	20.1	7.95	32
Sample9	10	61.4	14.4	8.17	14.1	13.6	6.4	28.2
Sample10	7.41	68.4	9.24	13.3	18.4	15.9	6.91	35.5
CV	**0.24**	**0.37**	**0.3**	**0.42**	**0.39**	**0.14**	**0.13**	**0.15**
	Middle Elution							
	Lower m/z				**Intermediate m/z**			
	Peak13	*Peak14*	*Peak15*	*Peak16*	*Peak17*	*Peak18*	*Peak19*	*Peak20*
Sample1	7.1	10.1	50.1	11.5	9.56	5.29	19.3	18
Sample2	7.57	12.1	41	11.4	11.9	4.44	19.4	19.3
Sample3	10.2	13.7	39.9	9.53	10.2	5.13	13.5	18.3
Sample4	9.81	15.6	38.7	8.49	9.46	4.98	12.2	18.8
Sample5	7.31	14.1	31	11.7	7.7	5.81	19.9	16.9
Sample6	8.59	11.9	28.7	10.1	7.8	7.27	14	19.3
Sample7	10.3	15.7	43.8	10.3	8.42	5.88	13.4	22
Sample8	6.99	17.5	27.8	9.49	8.93	6.02	13.4	21.2
Sample9	7	22.7	36.2	10.2	11.5	5.15	12	19
Sample10	8.91	21.5	42.8	10.8	8.5	4.58	14.4	22
CV	**0.16**	**0.26**	**0.19**	**0.1**	**0.15**	**0.15**	**0.21**	**0.09**

	Late Elution							
	Lower m/z				**Intermediate m/z**			
	Peak25	*Peak26*	*Peak27*	*Peak28*	*Peak29*	*Peak30*	*Peak31*	*Peak32*
Sample1	14.1	15.9	22	30.2	15.2	25	6.8	7.98
Sample2	14.2	16.2	13.5	26	12.8	24.1	4.37	7.44
Sample3	12.9	17.1	18.5	26.2	11.1	28.4	6.21	6.64
Sample4	13.3	13.8	16.1	25.8	10.2	24.5	5.73	6.81
Sample5	12.5	12.8	23.3	23	9.9	26.6	6.13	7.79
Sample6	13.4	13.6	27.1	20.6	9.04	25	6.19	9.57
Sample7	16.5	18	28.1	23.5	11	25.6	6.23	8.42
Sample8	12	19	17.3	16.9	8.64	23.5	6.4	6.6
Sample9	13.9	16.6	20.5	25.2	12.1	20.1	5.45	6.4
Sample10	11.9	13.5	16.8	22.4	11.6	29.7	5.57	8.16
CV	**0.1**	**0.14**	**0.24**	**0.15**	**0.17**	**0.1**	**0.11**	**0.13**
	Early Elution				Middle Elution			
	Higher m/z				**Higher m/z**			
	Peak9	*Peak10*	*Peak11*	*Peak12*	*Peak21*	*Peak22*	*Peak23*	*Peak24*
Sample1	6.24	3.47	6.95	6.16	15.2	51.5	9.08	15.8
Sample2	5.87	4.44	10.3	6.33	20.2	48.6	17.1	17.1
Sample3	5.86	4.26	7.6	5.93	14.5	38.5	18.2	16.9
Sample4	6.61	4.38	8.02	5.39	16.8	50.2	21.9	18.5
Sample5	9.2	5.74	9.58	7.39	15.4	56.5	20.2	20.7
Sample6	5.67	6.29	9.3	8.42	13	45.3	20.5	24
Sample7	5.18	4.39	7.28	4.83	13.7	38.3	20.1	22.3
Sample8	7.18	5.82	6.31	8.35	11	37.6	22.5	26.4
Sample9	12.5	4.23	7.18	6.54	14.6	39	16.9	21.3
Sample10	11.9	3.25	7.91	3.33	14.7	45.7	15.7	18.1
CV	**0.35**	**0.22**	**0.16**	**0.25**	**0.16**	**0.15**	**0.21**	**0.17**
	Late Elution							
	Higher m/z							
	Peak33	*Peak34*	*Peak35*	*Peak36*				
Sample1	52	15.2	96	17.2				
Sample2	49.9	10.3	76.6	12.9				
Sample3	38.5	14.8	84.5	15.6				
Sample4	47.1	14.5	93.7	16.2				
Sample5	49.2	14.7	84.3	12.9				
Sample6	44	15.2	92.5	14.4				
Sample7	33.7	15.4	90.7	13.9				
Sample8	37	17.1	88.3	12.7				
Sample9	44.2	13.1	63.9	9.76				
Sample10	45.7	11.6	72.7	12.4				
CV	**0.14**	**0.14**	**0.12**	**0.16**				

LITERATURE CITED

1. O'Farrell, P. H., High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* **1975,** 250, (10), 4007-21.
2. Gorg, A.; Postel, W.; Gunther, S., The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* **1988,** 9, (9), 531-46.
3. Lopez, M. F.; Berggren, K.; Chernokalskaya, E.; Lazarev, A.; Robinson, M.; Patton, W. F., A comparison of silver stain and SYPRO Ruby Protein Gel Stain with respect to protein detection in two-dimensional gels and identification by peptide mass profiling. *Electrophoresis* **2000,** 21, (17), 3673-83.
4. Harris, L. R.; Churchward, M. A.; Butt, R. H.; Coorssen, J. R., Assessing detection methods for gel-based proteomic analyses. *J Proteome Res* **2007,** 6, (4), 1418-25.
5. Compton, S. J.; Jones, C. G., Mechanism of dye response and interference in the Bradford protein assay. *Anal Biochem* **1985,** 151, (2), 369-74.
6. Unlu, M.; Morgan, M. E.; Minden, J. S., Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* **1997,** 18, (11), 2071-7.
7. Tonge, R.; Shaw, J.; Middleton, B.; Rowlinson, R.; Rayner, S.; Young, J.; Pognan, F.; Hawkins, E.; Currie, I.; Davison, M., Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. *Proteomics* **2001,** 1, (3), 377-96.
8. Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M., Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem* **1996,** 68, (5), 850-8.
9. Henzel, W. J.; Billeci, T. M.; Stults, J. T.; Wong, S. C.; Grimley, C.; Watanabe, C., Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc Natl Acad Sci U S A* **1993,** 90, (11), 5011-5.
10. Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R., 3rd, Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol* **1999,** 17, (7), 676-82.
11. Wolters, D. A.; Washburn, M. P.; Yates, J. R., 3rd, An automated multidimensional protein identification technology for shotgun proteomics. *Anal Chem* **2001,** 73, (23), 5683-90.
12. Whitehouse, C. M.; Dreyer, R. N.; Yamashita, M.; Fenn, J. B., Electrospray interface for liquid chromatographs and mass spectrometers. *Anal Chem* **1985,** 57, (3), 675-9.
13. Karas, M.; Hillenkamp, F., Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* **1988,** 60, (20), 2299-301.
14. Engwegen, J. Y.; Gast, M. C.; Schellens, J. H.; Beijnen, J. H., Clinical proteomics: searching for better tumour markers with SELDI-TOF mass spectrometry. *Trends Pharmacol Sci* **2006,** 27, (5), 251-9.

15. Tolson, J.; Bogumil, R.; Brunst, E.; Beck, H.; Elsner, R.; Humeny, A.; Kratzin, H.; Deeg, M.; Kuczyk, M.; Mueller, G. A.; Mueller, C. A.; Flad, T., Serum protein profiling by SELDI mass spectrometry: detection of multiple variants of serum amyloid alpha in renal cancer patients. *Lab Invest* **2004,** 84, (7), 845-56.
16. Lennon, J. J. Matrix Assisted Laser Desorption Ionization Time-of-flight Mass Spectrometry. http://www.abrf.org/ABRFNews/1997/June1997/jun97lennon.html (September 28, 2007),
17. Ahn, N. G.; Shabb, J. B.; Old, W. M.; Resing, K. A., Achieving in-depth proteomics profiling by mass spectrometry. *ACS Chem Biol* **2007,** 2, (1), 39-52.
18. Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R., Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* **1999,** 17, (10), 994-9.
19. Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M., Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **2002,** 1, (5), 376-86.
20. Sjovall, K.; Nilsson, B.; Einhorn, N., The significance of serum CA 125 elevation in malignant and nonmalignant diseases. *Gynecol Oncol* **2002,** 85, (1), 175-8.
21. Charalabopoulos, K.; Karakosta, A.; Bablekos, G.; Golias, C.; Charalabopoulos, A.; Tsanou, E.; Peschos, D.; Zoganas, L.; Batistatou, A., CEA levels in serum and BAL in patients suffering from lung cancer: Correlation with individuals presenting benign lung lesions and healthy volunteers. *Med Oncol* **2007,** 24, (2), 219-25.
22. Thompson, I. M.; Pauler, D. K.; Goodman, P. J.; Tangen, C. M.; Lucia, M. S.; Parnes, H. L.; Minasian, L. M.; Ford, L. G.; Lippman, S. M.; Crawford, E. D.; Crowley, J. J.; Coltman, C. A., Jr., Prevalence of prostate cancer among men with a prostate-specific antigen level < or =4.0 ng per milliliter. *N Engl J Med* **2004,** 350, (22), 2239-46.
23. Petricoin, E. F.; Ardekani, A. M.; Hitt, B. A.; Levine, P. J.; Fusaro, V. A.; Steinberg, S. M.; Mills, G. B.; Simone, C.; Fishman, D. A.; Kohn, E. C.; Liotta, L. A., Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* **2002,** 359, (9306), 572-7.
24. Sorace, J. M.; Zhan, M., A data review and re-assessment of ovarian cancer serum proteomic profiling. *BMC Bioinformatics* **2003,** 4, 24.
25. Elwood, M., Proteomic patterns in serum and identification of ovarian cancer. *Lancet* **2002,** 360, (9327), 170; author reply 170-1.
26. Pearl, D. C., Proteomic patterns in serum and identification of ovarian cancer. *Lancet* **2002,** 360, (9327), 169-70; author reply 170-1.
27. Rockhill, B., Proteomic patterns in serum and identification of ovarian cancer. *Lancet* **2002,** 360, (9327), 169; author reply 170-1.
28. Diamandis, E. P., Serum proteomic profiling by matrix-assisted laser desorption-ionization time-of-flight mass spectrometry for cancer diagnosis: next steps. *Cancer Res* **2006,** 66, (11), 5540-1.
29. Ransohoff, D. F., Lessons from controversy: ovarian cancer screening and serum proteomics. *J Natl Cancer Inst* **2005,** 97, (4), 315-9.

30. Diamandis, E. P., Analysis of serum proteomic patterns for early cancer diagnosis: drawing attention to potential problems. *J Natl Cancer Inst* **2004,** 96, (5), 353-6.
31. Hsieh, S. Y.; Chen, R. K.; Pan, Y. H.; Lee, H. L., Systematical evaluation of the effects of sample collection procedures on low-molecular-weight serum/plasma proteome profiling. *Proteomics* **2006,** 6, (10), 3189-98.
32. Banks, R. E.; Stanley, A. J.; Cairns, D. A.; Barrett, J. H.; Clarke, P.; Thompson, D.; Selby, P. J., Influences of blood sample processing on low-molecular-weight proteome identified by surface-enhanced laser desorption/ionization mass spectrometry. *Clin Chem* **2005,** 51, (9), 1637-49.
33. Pilny, R.; Bouchal, P.; Borilova, S.; Ceskova, P.; Zaloudik, J.; Vyzula, R.; Vojtesek, B.; Valik, D., Surface-enhanced laser desorption ionization/time-of-flight mass spectrometry reveals significant artifacts in serum obtained from clot activator-containing collection devices. *Clin Chem* **2006,** 52, (11), 2115-6.
34. Karsan, A.; Eigl, B. J.; Flibotte, S.; Gelmon, K.; Switzer, P.; Hassell, P.; Harrison, D.; Law, J.; Hayes, M.; Stillwell, M.; Xiao, Z.; Conrads, T. P.; Veenstra, T., Analytical and preanalytical biases in serum proteomic pattern analysis for breast cancer diagnosis. *Clin Chem* **2005,** 51, (8), 1525-8.
35. Baggerly, K. A.; Morris, J. S.; Edmonson, S. R.; Coombes, K. R., Signal in noise: evaluating reported reproducibility of serum proteomic tests for ovarian cancer. *J Natl Cancer Inst* **2005,** 97, (4), 307-9.
36. Traum, A. Z.; Wells, M. P.; Aivado, M.; Libermann, T. A.; Ramoni, M. F.; Schachter, A. D., SELDI-TOF MS of quadruplicate urine and serum samples to evaluate changes related to storage conditions. *Proteomics* **2006,** 6, (5), 1676-80.
37. Semmes, O. J.; Feng, Z.; Adam, B. L.; Banez, L. L.; Bigbee, W. L.; Campos, D.; Cazares, L. H.; Chan, D. W.; Grizzle, W. E.; Izbicka, E.; Kagan, J.; Malik, G.; McLerran, D.; Moul, J. W.; Partin, A.; Prasanna, P.; Rosenzweig, J.; Sokoll, L. J.; Srivastava, S.; Thompson, I.; Welsh, M. J.; White, N.; Winget, M.; Yasui, Y.; Zhang, Z.; Zhu, L., Evaluation of serum protein profiling by surface-enhanced laser desorption/ionization time-of-flight mass spectrometry for the detection of prostate cancer: I. Assessment of platform reproducibility. *Clin Chem* **2005,** 51, (1), 102-12.
38. Gygi, S. P.; Rochon, Y.; Franza, B. R.; Aebersold, R., Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol* **1999,** 19, (3), 1720-30.
39. Garber, K., Debate rages over proteomic patterns. *J Natl Cancer Inst* **2004,** 96, (11), 816-8.
40. Brunner, E.; Ahrens, C. H.; Mohanty, S.; Baetschmann, H.; Loevenich, S.; Potthast, F.; Deutsch, E. W.; Panse, C.; de Lichtenberg, U.; Rinner, O.; Lee, H.; Pedrioli, P. G.; Malmstrom, J.; Koehler, K.; Schrimpf, S.; Krijgsveld, J.; Kregenow, F.; Heck, A. J.; Hafen, E.; Schlapbach, R.; Aebersold, R., A high-quality catalog of the Drosophila melanogaster proteome. *Nat Biotechnol* **2007,** 25, (5), 576-83.
41. Albrethsen, J.; Bogebo, R.; Olsen, J.; Raskov, H.; Gammeltoft, S., Preanalytical and analytical variation of surface-enhanced laser desorption-ionization time-of-flight mass spectrometry of human serum. *Clin Chem Lab Med* **2006,** 44, (10), 1243-52.

42. Nelsestuen, G. L.; Zhang, Y.; Martinez, M. B.; Key, N. S.; Jilma, B.; Verneris, M.; Sinaiko, A.; Kasthuri, R. S., Plasma protein profiling: unique and stable features of individuals. *Proteomics* **2005,** 5, (15), 4012-24.
43. Anderson, N. L.; Anderson, N. G., The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics* **2002,** 1, (11), 845-67.
44. Hu, Y.; Malone, J. P.; Fagan, A. M.; Townsend, R. R.; Holtzman, D. M., Comparative proteomic analysis of intra- and interindividual variation in human cerebrospinal fluid. *Mol Cell Proteomics* **2005,** 4, (12), 2000-9.
45. Nishihara, J. C.; Champion, K. M., Quantitative evaluation of proteins in one- and two-dimensional polyacrylamide gels using a fluorescent stain. *Electrophoresis* **2002,** 23, (14), 2203-15.
46. Molloy, M. P.; Brzezinski, E. E.; Hang, J.; McDowell, M. T.; VanBogelen, R. A., Overcoming technical variation and biological variation in quantitative proteomics. *Proteomics* **2003,** 3, (10), 1912-9.
47. Zhang, X. G., Yuanbiao; Song, Yanping; Sun, Wei; Yu, Chaohui; Zhao, Xiaohang; Wang, Hongyang; Jiang, Hongchi; Li, Youming; Qian, Xiaohong; Jiang, Ying; He, Fuchu, Proteomic analysis of individual variation in normal livers of human beings using difference gel electrophoresis. *Proteomics* **2006,** 6, (19), 5260 - 5268.
48. Pieper, R.; Su, Q.; Gatlin, C. L.; Huang, S. T.; Anderson, N. L.; Steiner, S., Multi-component immunoaffinity subtraction chromatography: an innovative step towards a comprehensive survey of the human plasma proteome. *Proteomics* **2003,** 3, (4), 422-32.
49. Jmeian, Y.; El Rassi, Z., Tandem affinity monolithic microcolumns with immobilized protein A, protein G', and antibodies for depletion of high abundance proteins from serum samples: integrated microcolumn-based fluidic system for simultaneous depletion and tryptic digestion. *J Proteome Res* **2007,** 6, (3), 947-54.
50. Ettorre, A.; Rosli, C.; Silacci, M.; Brack, S.; McCombie, G.; Knochenmuss, R.; Elia, G.; Neri, D., Recombinant antibodies for the depletion of abundant proteins from human serum. *Proteomics* **2006,** 6, (16), 4496-505.
51. Warburton, S.; Southwick, K.; Hardman, R. M.; Secrest, A. M.; Grow, R. K.; Xin, H.; Woolley, A. T.; Burton, G. F.; Thulin, C. D., Examining the proteins of functional retinal lipofuscin using proteomic analysis as a guide for understanding its origin. *Mol Vis* **2005,** 11, 1122-34.
52. Merrell, K.; Southwick, K.; Graves, S. W.; Esplin, M. S.; Lewis, N. E.; Thulin, C. D., Analysis of low-abundance, low-molecular-weight serum proteins using mass spectrometry. *J Biomol Tech* **2004,** 15, (4), 238-48.
53. Shi, R.; Kumar, C.; Zougman, A.; Zhang, Y.; Podtelejnikov, A.; Cox, J.; Wisniewski, J. R.; Mann, M., Analysis of the mouse liver proteome using advanced mass spectrometry. *J Proteome Res* **2007,** 6, (8), 2963-72.
54. Miyamoto, M.; Yoshida, Y.; Taguchi, I.; Nagasaka, Y.; Tasaki, M.; Zhang, Y.; Xu, B.; Nameta, M.; Sezaki, H.; Cuellar, L. M.; Osawa, T.; Morishita, H.; Sekiyama, S.; Yaoita, E.; Kimura, K.; Yamamoto, T., In-depth proteomic profiling of the normal human kidney glomerulus using two-dimensional protein prefractionation in combination with liquid chromatography-tandem mass spectrometry. *J Proteome Res* **2007,** 6, (9), 3680-90.
55. Kruskal, W. H., Wallis, W. Allen, Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association* **1952,** 47, (260), 583-621.

56. Chen, Z.; Southwick, K.; Thulin, C. D., Initial analysis of the phosphoproteome of Chinese hamster ovary cells using electrophoresis. *J Biomol Tech* **2004,** 15, (4), 249-56.
57. Pappin, D. J.; Hojrup, P.; Bleasby, A. J., Rapid identification of proteins by peptide-mass fingerprinting. *Curr Biol* **1993,** 3, (6), 327-32.
58. Fry, J. D., Bahnck, C. M., Mikucki, M., Phadnis, N., Slattery, W. C., Dietary ethanol mediates selection on aldehyde dehydrogenase activity in Drosophila melanogaster. *Integrative and Comparative Biology* **2004,** 44, 275-283.
59. Taraszka, J. A.; Gao, X.; Valentine, S. J.; Sowell, R. A.; Koeniger, S. L.; Miller, D. F.; Kaufman, T. C.; Clemmer, D. E., Proteome profiling for assessing diversity: analysis of individual heads of Drosophila melanogaster using LC-ion mobility-MS. *J Proteome Res* **2005,** 4, (4), 1238-47.
60. Hojlund, K.; Yi, Z.; Hwang, H.; Bowen, B.; Lefort, N.; Flynn, C. R.; Langlais, P.; Weintraub, S. T.; Mandarino, L. J., Characterization of the human skeletal muscle proteome by one-dimensional Gel electrophoresis and HPLC-ESI-MS/MS. *Mol Cell Proteomics* **2007**.