ABSTRACT

COLLIER, TIMOTHY SHAWN. The Development of Analytical Strategies for the Quantitative Proteomic Analysis of Biological Systems Impacting Agriculture and Human Health. (Under the direction of Dr. David C. Muddiman).

Protein quantification is an integral part of proteomics science, which seeks to fully characterize the identities, amount, localizations, and functions of all proteins in a biological system under any given set of conditions. Mass spectrometry based proteomics is a robust and continuously developing field with the aim of addressing these goals. Two main strategies exist within the field, termed “top-down” and “bottom-up” for the identification of proteins. A major advantage of top-down mass spectrometric characterization of proteins is that the information of the intact protein is retained over the vastly more common bottom-up approach that uses protease-generated peptides to search genomic databases for protein identification. Detecting the intact mass is instrumental in discerning protein isoforms and combinations of post-translational modifications. The emergence of top-down proteomics has facilitated the identification and characterization of intact protein forms that are not readily apparent in bottom-up studies. Alternatively, bottom-up proteomics commonly yields hundreds of protein identifications compared to tens of identities with the top-down approach.

Concurrent to the development of methods for protein identification have been strategies for quantification including chemical and metabolic incorporation of stable-isotope labels, including stable isotope labeling by amino acids in cell culture (SILAC). Herein, SILAC is employed for the relative quantification of intact proteins in the agriculturally important fungus *Aspergillus flavus* and for the quantification of

proteins in human embryonic stem cells. The extent of isotope label incorporation is found to play a major role in its implementation for top-down quantification, and a mathematical model was developed to facilitate its utilization.

To explore the capability of protein quantification in systems that are not typically amenable to metabolic labeling, SILAC quantification was also compared to label-free quantification using spectral counting. In the bottom-up characterization of human embryonic stem cells, SILAC and spectral counting were found to possess a similar dynamic range of quantification. Spectral counting was also found to identify and quantify more proteins than SILAC, but at the cost of having lower sensitivity to changes in protein expression.

The Development of Analytical Strategies for the Quantitative Proteomic Analysis of Biological Systems Impacting Agriculture and Human Health

by
Timothy Shawn Collier

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Chemistry

Raleigh, North Carolina

2010

APPROVED BY:

_____	_____
David C. Muddiman Professor, Chemistry Committee Chair	Edmond F. Bowden Professor, Chemistry
_____	_____
Morteza G. Khaledi Professor, Chemistry	Gary A. Payne Professor, Plant Pathology

DEDICATION

This dissertation is dedicated to my parents, James B. and Terra N. Collier, my sister, Brandi S. White, brother Dustin B. Collier and their families. Without their unconditional love, support, guidance, friendship, and encouragement this voyage could have never begun, much less have finished. My unending gratitude, thanks, and love extend to this outstanding group of people.

BIOGRAPHY

Timothy Shawn Collier was born on January 24th, 1985 in Wilson, NC and is the son of James and Terra Collier. His interest in science began at an early age and was cultivated throughout his primary and secondary education. His interests were further influenced by his final years of high school at the North Carolina School of Science and Mathematics, leading to pursuit of a B.S. in Chemistry at North Carolina State University. Tim began performing research at the undergraduate level performing laboratory and field based biomolecular research on fossil specimens under the mentorship of Dr. Mary Schweitzer, sparking his interest in protein chemistry and biology. This experience inspired him to pursue his Ph.D. research in biological mass spectrometry under the mentorship of Dr. David Muddiman, who had recently joined the faculty at North Carolina State University.

ACKNOWLEDGMENTS

I would like to acknowledge several people who have been instrumental during my time in graduate school. My advisor, Dr. David C. Muddiman, provided me with encouragement and guidance. I am, and will continue to be, constantly inspired by his drive and energy in the pursuit of science to improve the way we investigate and solve biological questions. I also wish to thank Dr. Adam Hawkridge for his encouragement and the many helpful discussions pertaining to research strategies and experimental processes. I am greatly appreciative of my colleagues in the Muddiman research group for their support and friendships.

I would also like to acknowledge by name the support and companionship from the following individuals, who have each been instrumental in my life while in graduate school: Chris Ion, Nelson Mendoza, Kendra Dougherty, Carrie Langlais, Richard Furr, Lindsay Arndt, T.J. Price, Justine O'Neil, Sandra White, Rachel Williams, and last but not least Boot and Lilly.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF PUBLICATIONS

Peer reviewed manuscripts:

1. Collier, T.S.; Muddiman, D.C.; Analytical Strategies for the Global Quantification of Intact Proteins *Amino Acids,* **2010** In Press

2. Collier, T.S.; Sarkar, P.; Franck, W.L.; Rao, B.M.; Dean, R.A.; Muddiman, D.C.; Direct Comparison of Stable Isotope Labeling by Amino Acids in Cell Culture and Spectral Counting for Quantitative Proteomics *Anal. Chem.,* **2010** published online

3. Collier, T.S.; Sarkar, P.; Rao, B.M.; Muddiman, D.C.; Quantitative Top-Down Proteomics of SILAC Labeled Human Embryonic Stem Cells. *J. Amer. Soc. Mass. Spectrom.* **2010,** 21, 879-889

4. Alcazar, O.; Hawkridge, A.M.; Collier, T.S.; Cousins, S.W.; Bhattacharya, S.K.; Muddiman, D.C.; Marin-Castano, M.E.; Proteomics Characterization of Membrane Blebs in Human Retinal Pigment Epithelium Cells *Mol. Cell. Proteomics*. **2009,** 8, 2201-2211

5. Collier, T.S.; Hawkridge, A.M.; Georgianna, D.R.; Payne, G.A.; Muddiman, D.C.; Top-Down Identification and Quantification of Stable Isotope Labeled Proteins from *Aspergillus flavus* using Online nano-Flow Reversed-Phase Liquid Chromatography Coupled to a LTQ-FTICR Mass Spectrometer *Anal. Chem.* **2008,** 80, 4994-5001

6. Schweitzer, M.H.; Avci, R.; Collier, T.S.; Goodwin, M.B.; Microscopic, Chemical and Molecular Methods for Examining Fossil Preservation *Comptes Rendus Palevol*, **2008**, 7, 159-184

Abstracts for presentation (Oral):

1. Confident Identification and Relative Quantification of Intact Proteins from Human Embryonic Stem Cells using SILAC. The 58th ASMS Conference on Mass Spectrometry and Allied Topics, Salt Lake City, UT, May 2010. T.S. Collier, P. Sarkar, B.M. Rao, and D.C. Muddiman.

2. Unambiguous Proteomic Identification and Relative Quantification of SILAC Labeled Human Embryonic Stem Cells using the Top-Down Approach. Triangle Area Mass Spectrometry Discussion Group Seminar, Research Triangle Park, NC, May 2010. T.S. Collier, P. Sarkar, B.M. Rao, and D.C. Muddiman.

3. Top-Down Identification and Quantification of Proteins from the Prototroph *Aspergillus flavus* using nano-Flow Reversed-Phase LC Directly Coupled to LTQ-FT-ICR Mass Spectrometry. BASF Corporation, Research Triangle Park, NC, April 2009. T. S. Collier, D.R. Georgianna, G.A. Payne and D.C. Muddiman

4. Preparation and Quantitative Intact Protein Analysis of the Fungus *Aspergillus flavus* using SILAC, Accurate Intact Mass and Arginine Counting by RPLC-FT Mass Spectrometry. Southeast Regional Meeting of the American Chemical Society, Greenville, SC, October 2007. T. S. Collier, D.C. Muddiman, D.K. Williams, Jr., A.M. Hawkridge, D.R. Georgianna, and G.A. Payne.

Abstracts for presentation (Poster):

1. Preparation of *Magnaporthe oryzae* Mycelia and Spores for Proteomic Analysis: Protein Extraction, Fractionation, and Identification. 58th ASMS Conference on Mass Spectrometry and Allied Topics, Salt Lake City, UT, May 2010. A.R. Courie, E. Gokce, T.S. Collier, W.L. Franck, Y. Oh, R.A. Dean, D.C. Muddiman

2. Quantitative Proteomic Analysis of Human Embryonic Stem Cells Utilizing Size and Charge-Based Separations and nano-LC-LTQ-FTICR-MS2. Triangle Chromatography Discussion Group Symposium, Raleigh, NC, May 2010. T.S. Collier, P. Sarkar, B.M. Rao, D.C. Muddiman.

3. Quantitative Top-down Proteomic Analysis of SILAC Labeled Human Embryonic Stem Cells. Triangle Area Mass Spectrometry Discussion Group, Research Triangle Park, NC October 2009. T.S. Collier, P. Sarkar, B.M. Rao, D.C. Muddiman

4. Intracellular Signaling During Early Differentiation of Human Embryonic Stem Cells. American Institute of Chemical Engineers Annual Meeting, Nashville, TN 2009. B.M. Rao, P. Sarkar, T.S. Collier, D.C. Muddiman

5. Investigating the Secretome of individual and Co-cultured Thermophiles for Small Putative Proteins of Enzymatic Potential. 57th ASMS Conference on Mass Spectrometry and Allied Topics, Philadelphia, PA, June 2009. G.L. Andrews, D.L. Lewis, S.E. Blumer-Schuette, J. Notey, R.M. Kelly, T.S. Collier, D.C. Muddiman

6. Quantitative Top-Down and Bottom-Up Proteomics of Aflatoxin Producing *Aspergillus flavus* Utilizing Gel Electrophoresis, mRP Chromatography, and Online nano-Flow LC-MS/MS. Triangle Chromatography Discussion Group, Raleigh, NC, May 2009. T.S. Collier, A.M. Hawkridge, D.R. Georgianna, G.A. Payne, D.C. Muddiman

7. Investigating Potential Protein Functions of Small Unknown Reading Frames Secreted by Individual and Co-Cultured Thermophiles. Triangle Chromatography Discussion Croup, Raleigh, NC, May 2009. G.L. Andrews, D.L. Lewis, S.E. Blumer-Schuette, J. Notey, R.M. Kelly, T.S. Collier, D.C. Muddiman

8. Identification and Quantification of Intact Proteins from *A. flavus* using SILAC and nano-Flow LC-LTQ-FT-ICR-MS. 56th Conference on Mass Spectrometry and Allied Topics, Denver, CO, June 2008. T.S. Collier, A.M. Hawkridge, D.R. Georgianna, G.A. Payne, D.C. Muddiman

9. Top-Down Identification and Quantification of Proteins from the Prototroph *Aspergillus flavus* using nano-Flow Reversed Phase LC directly coupled to LTQ-FTICR Mass Spectrometry. Triangle Chromatography Discussion Group Symposium, Raleigh, NC, May 2008. T.S. Collier, A.M. Hawkridge, D.R. Georgianna, G.A. Payne, D.C. Muddiman

10. Protoemic Analysis of Proteins Isolated from Mammalian Bone Using Pressure Cycling Technology. International Human Proteome Organization, Seoul, Korea, 2007. G.B. Smejkal, A.T. Kwan, I. Romanovsky, D.C. Muddiman, M.H. Schweitzer, and T.S. Collier

11. Electrophoretic Analyses of Proteins and Peptides Isolated from Cortical Bone using a Pressure Cycling Technology. United States Human Proteome Organization, 3rd Annual Meeting, Seattle, WA, 2007. G.B. Smejkal, D.C. Muddiman, M.H. Schweitzer, and T.S. Collier

12. Separation and Quantitative Top-Down Analysis of the Fungus *Aspergillus flavus* Using SILAC, Accurate Intact Mass and Arginine Counting by RPLC-FTICR-MS. 55th ASMS Conference on Mass Spectrometry and Allied Topics, Indianapolis, IN, June, 2007. T.S. Collier, A.M. Hawkridge, D.R. Georgianna, G.A. Payne, D.C. Muddiman

CHAPTER 1

Introduction

1.1 High Performance Mass Spectrometry

1.1.1 Principles of Electrospray Ionization

The development of electrospray ionization (ESI) opened the door for the analysis of large biomolecules by mass spectrometry.[1] ESI is a soft ionization technique that enables the conversion of large, otherwise non-volatile biological macromolecules into gas-phase ions suitable for mass spectrometric analysis. These ions are generated through the application of an electric potential to an ESI emitter through which the analyte solution is passed, forming a fine spray of charged droplets which travel towards the mass spectrometer across a potential gradient spanning the distance between the emitter and the entrance to the mass spectrometer. In transit, the droplets evaporate, increasing the density of electrical charge on the droplets' surface. As the surface area of the droplet decreases, it approaches the Rayleigh limit, undergoing Coulombic fission as the charge repulsion force exceeds the surface tension of the droplet.[2] This process produces more, smaller, charged offspring droplets. The mechanism by which the analyte ions are ultimately transferred to the gas phase is described in two models-one by Dole and the other by Iribarne and Thomson. Dole's charged-residue model describes the process as the repeated evaporation and Coulombic fission of offspring droplets until the droplet contains a single molecule, which evaporates and ultimately results in a

gas-phase ion.[3] Iribarne and Thomson proposed the ion desorption model, which states that the electric field at the surface of the droplet provides the work necessary to extract the analyte ion from the droplet and into the gas phase.[4, 5]

There are many variables and conditions that determine the electrospray response of an analyte ion in a given mass spectrum. The nature of analyte ion desorption from a charged droplet is given in **Equation 1.1** for an ion (i) with z number of charges.

$$N_{iz} = 3Ae^{(-\Delta G^{\circ}_{iz}/RT)}(Ni/r)e^{(\Delta r^{*}zQ/4\pi\varepsilon_0 RTr^2)} \qquad \textbf{Equation 1.1}$$

In the above equation, A is a proportionality constant relating analyte amount to surface density and depends on the analyte's surface activity. ΔG°_{iz} is the free energy of solvation, or fugacity. R is the gas constant; T is the temperature, Ni is the number of moles, r is the radius of the droplet, Δr* is the distance required for the analyte to be free of the droplet, z is the number of charges on the analyte, Q is the amount of excess charge on the droplet, and ε_0 is the permittivity of a vacuum. Most variables in this expression are constants that can be controlled in an experimental environment. However, the surface activity (A), and the free energy of solvation, ΔG°_{iz}, are unique to the analyte. The surface activity of an analyte is related to the affinity of the analyte for the surface of a droplet. In the case of electrospray where the solvent is mostly and aqueous/organic mixture, a more hydrophobic analyte will have a higher surface activity. Reversed phase liquid chromatography benefits the

electrospray of complex mixtures by keeping analytes of similar hydrophobicity in the same droplets, reducing charge competition based on surface activity.[6, 7]

1.1.2 Fourier Transform Ion Cyclotron Resonance Mass Spectrometry

After electrospray ionization, a small percentage (<1%) of generated ions enter the inlet of the mass spectrometer.[8] The Muddiman Laboratory is equipped with a hybrid LTQ-FT-ICR mass spectrometer, in which case ions are guided through a series of multipoles and accumulated in a linear ion trap (LTQ) where the ions are either detected by a radial detector or alternatively accumulated and transferred through another set of ion optics to the ICR cell within the bore of a high field superconducting magnet (7 Tesla) for detection. Once the analyte ions have been transferred to the ICR cell (also known as a Penning trap), the analyte ions in the presence of a homogeneous magnetic field assume a naturally occurring circular motion as a result of the Lorentz force (**Figure 1.1)**.[9] This force redirects ions into a circular path when they encounter the perpendicular magnetic field provided by the magnet. An rf voltage applied to

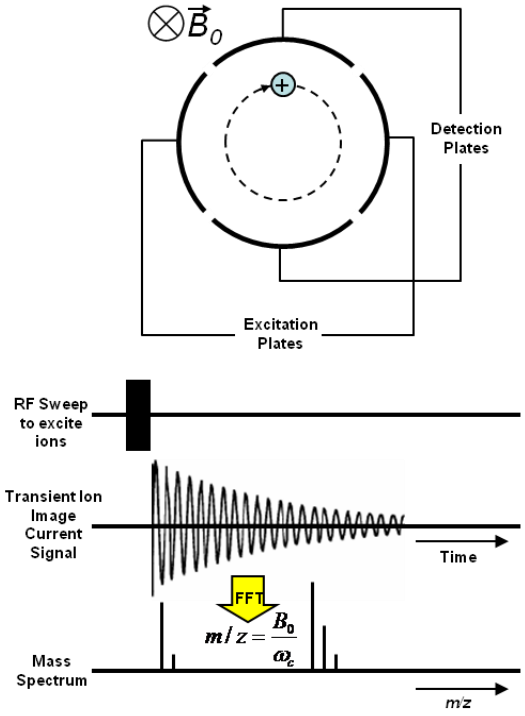

Figure 1.1 Schematic showing ion motion in an ICR cell due to the Lorentz force alongside an overview of the process for detecting signal in an FT-ICR mass spectrometer.

opposing plates of the ICR cell at a given frequency will excite ions with the same cyclotron frequency to a higher radius with the cell and as such an rf frequency sweep will excite ions over the corresponding frequency range into a higher radius.[10]

This sweep also induces phase coherence of ions with the same *m/z*. The higher radius allows the coherent ion clouds to produce an image current on the detector electrodes occupying opposite sides of the ICR cell. This signal appears sinusoidal in the time domain, reflecting the motion of electrons from one detection electrode to the other within the detection circuit. This signal can be Fourier transformed to obtain the specific frequencies of the ion packets, which is easily converted to a *m/z* scale using **Equation 1.2** where the *cyclotron frequency* ω_c is directly proportional to the charge (*q)* and mass (*m*), in a constant magnetic field (B_0).[10]

$$\omega_c = \frac{qB_0}{m} \qquad \textbf{Equation 1.2}$$

The resolving power of an FT-ICR mass spectrometer is given by the strength of the magnetic field, B_0, the acquisition time, $T_{acq'n}$, and the *m/z* of the detected species as described in **Equation 1.3**. The resolving power of a 7 T FT-ICR-MS is routinely 100,000 at *m/z* = 400, and leads to mass measurement accuracy (MMA) routinely < 3 ppm (**Equation 1.4**).

$$R = \frac{m}{\Delta m_{FWHM}} = \frac{1.274 \times 10^7 B_0 T_{acq'n}}{m/z} \qquad \textbf{Equation 1.3}$$

$$MMA(ppm) = \frac{M_{\exp} - M_{theo}}{M_{theo}} \times 10^6 \qquad \textbf{Equation 1.4}$$

The high resolving power afforded by FT-ICR-MS allows for isotopic resolution of analytes, which as a result of the natural abundance of elements, appears as a distribution of various isotopic peaks which increase in complexity as the size of the molecule increases. Each of these isotopic peaks corresponds to a discreet *cyclotron frequency*. As the molecular weight of an analyte increases, the amount of charge on that molecule necessary for detection in a given m/z space also increases. The cyclotron frequencies of isotopes of these larger molecules will be more closely spaced, thus giving rise to isotopic peaks that are more closely spaced in a mass spectrum. In the case of larger molecules, inhomogeneities in the magnetic or electric field, in addition to interactions of ion clouds of the same m/z lead to coalescence of ion packets in the ICR cell and thus their isotopic frequencies, resulting in a large diffuse peak, rather than many sharply resolved isotopic peaks.[11, 12]

The unrivaled mass measurement accuracy, resolving power, dynamic range and sensitivity of FT-ICR mass spectrometry make it the most amenable mass analyzer for the analysis of intact proteins when coupled to the multiple-charging process inherent to ESI.[1, 9, 13-16] Multiple-charging allows for higher analyte frequency in the ICR cell, extending the molecular weight range of the detector. Multiply charged proteins are also easier to fragment and sequence using *MS/MS* techniques.[17, 18]

1.1.3 Orbitrap Mass Analyzers

Until recently, FT-ICR mass analysis surpassed all other mass spectrometric platforms in terms of resolving power and mass accuracy. The development of the orbitrap by Makarov[19] is similar to an ion storage device called a Kingdon trap[20], which has been used to study molecular beams[21] and ion spectroscopy[22, 23] by trapping ions with the use of only electrostatic fields. The development of the orbitrap has provided a means of achieving similar performance to an FT-ICR mass analyzer without the capital and maintenance costs for the support of a high field superconducting magnet.

The orbitrap relies on two electrodes to provide the trapping electrostatic field-an inner spindle electrode surrounded by an outer barrel electrode-and traps ions in three dimensional space.[19, 24] Three different frequencies can be measured within the orbitrap for mass analysis; however, only the frequency associated with ion motion in the z direction as shown in **Figure 1.2** is not associated with kinetic energy and is thus used for mass analysis in commercial instruments. The frequency measured in the orbitrap is inversely proportional to the square root of the

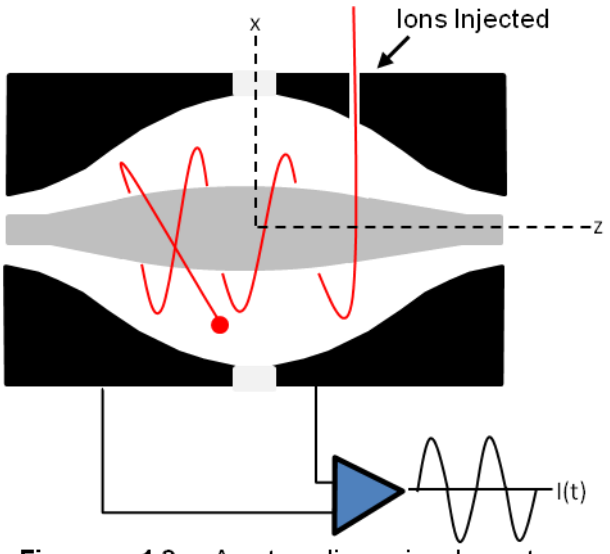

Figure 1.2 A two-dimensional cut-away schematic of the Orbitrap

mass to charge ratio (*m/z*). **Equation 1.5** describes the measured orbitrap frequency where *k* is a constant associated with field curvature.

$$\omega = \sqrt{(z/m)k} \qquad \textbf{Equation 1.5}$$

Detection of ions in the orbitrap is similar to that of an ICR mass analyzer except that there is no excitation event. The oscillation of ions along the z-axis of the orbitrap are detected as a sinusoidal image current and transformed to the frequency domain via fast Fourier transform. From the frequency domain, **Equation 1.5** can be employed to generate mass spectra. The Orbitrap rivals the FT-ICR mass analyzer in resolving power and can routinely obtain measurements with < 2 ppm mass measurement accuracy utilizing lock mass calibration.[25]

1.2 Top-Down and Bottom-Up Proteomics

Two main analytical strategies, termed “bottom-up” and “top-down”, are currently employed for identifying proteins in complex mixtures. Bottom-up proteomics, a more widely implemented strategy relative to top-down, involves submitting a protein or protein mixture (e.g., cell lysate) to enzymatic digestion with a protease (i.e. trypsin, Lys-C, Asp-N, etc.) rendering the sample as a complex mixture of short peptides with defined terminus. The peptides are then separated and analyzed via liquid chromatography (often reversed phase) coupled to electrospray mass spectrometry (ESI-MS) which sequences the peptides using tandem mass spectrometry (MS/MS). Searching the precursor ion mass and sequence

information against a predicted protein database often results in hundreds or thousands of protein identifications.

However, sequence coverage of the original proteins is often low, with many identifications coming from the detection of a single peptide which may correspond to multiple protein forms.[26] These multiple protein forms may have unique biological function, potentially possessing diagnostic, prognostic, and/or therapeutic properties. Proteins such as brain natriuretic peptide, C-reactive protein, and troponin I and T are already regarded biomarkers in their native forms, but specific modifications (e.g. post-translational modifications (PTMs) and unique cleavages) can result in diagnostic markers that are more specifically related to a disease state.[27-32] Knowing the mass of the protein that gave rise to a specific proteolytic peptide would be very useful in determining its identity and whether it is a form related to a pathological stimulus.

Top-down proteomics provides both the intact mass and sequence information for protein identification and, though not employed as often as the bottom-up technique, continues to mature as a method for identifying and characterizing intact proteins in biological systems.[33] The intact protein mass is first obtained by the mass spectrometer, followed by isolation for MS/MS analysis to provide sequence information. The combination of intact mass and sequence data enable the specific form of the protein to be determined, locating PTMs and identifying alternative slicing variants and degradation products.

Top-down proteomics has lagged behind the more widely used bottom-up strategies due, in large part, to the technological challenges of measuring and identifying large intact proteins. These challenges have included time consuming off-line separation of samples[26, 34-36], mass analyzer speed[37], and underdevelopment of database search software tailored to top-down experiments. However, advances in separation technology and methodology[38, 39], the development of faster hybrid mass analyzer technology such as the LTQ-FT mass spectrometer[40], and recent innovations in top-down database searching software (e.g. ProSight PTM[41]) have created the potential for more widespread utilization of top-down proteomics techniques. Electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry (ESI-FT-ICR-MS) is well suited for carrying out top-down proteomics measurements of intact multiply-charged proteins. The combination of high mass accuracy <3 ppm[42, 43], high resolving power, and a multitude of MS/MS dissociation techniques (e.g. CID[44], IRMPD[45], ECD[46], and ETD[47]) provide unprecedented confidence for protein identification and characterization (e.g. localization of PTMs, unique cleavages, and protein variants).[48, 49]

1.3 Analytical Strategies for the Global Quantification of Intact Proteins

Proteomics research is a robust and rapidly evolving research field that seeks to "identify and quantify the structures and biochemical and cellular functions of all proteins in an organism, organ, or organelle, and how they vary in space, time, and physiological state".[50] While these objectives have yet to be fully realized, many

analytical platforms have been developed in pursuit of them. Antibody-based assays such as the enzyme-linked immune-sorbent assay (ELISA) and Western blot are frequently employed in laboratories and clinics, often providing sensitive and specific quantification due to the availability of well characterized antibodies that have been carefully developed to bind to specific protein epitopes corresponding to specific isoforms. However, these assays are often limited in their biological scope, as significant time and cost must be invested in the development of antibody reagents for each protein to be quantified. This renders immune-based techniques impractical for the global quantitative analysis of intact proteins. In this report, we describe strategies that have been developed for the quantitative analysis of intact proteins, including differential gel electrophoresis (DIGE), in which the relative intensity of fluorescent dyed protein spots are quantified within the same gel. We also summarize the direct identification and quantification of intact proteins with mass spectrometry, most notably metabolically labeling strategies in addition to recent developments in chemical labeling and label-free quantification.

1.3.1 Differential Gel Electrophoresis (DIGE)

Two-dimensional gel electrophoresis (2D-GE) is a widely employed means for protein separation, capable of visualizing hundreds of proteins and their isoforms separating first by isoelectric point, then by molecular weight. Using traditional visualization protocols such as colloidal coomassie dye and silver staining, differentially expressed proteins can be compared across replicate gels, although variability across multiple gels negatively affect sensitivity, and these stains lack an

adequate linear dynamic range for detection and relative quantification.[51, 52] Succinctly, the poor reproducibility of 2D-GE made it difficult to accurately report real quantitative change in protein expression. As a means of eliminating inter-gel variability, Ünlü and coworkers introduced differential gel electrophoresis, which loads multiple samples that are differentially labeled with fluorescent tags onto the same gel; thus, the proteins are allowed to electrophoretically co-migrate yet can be detected at different wavelengths and relatively quantified based on fluorescence.[53] An overview of the experimental strategy of DIGE is depicted in **Figure 1.3**. In DIGE, fluorescent labeling of proteins is performed prior to 2D-GE in contrast to traditional total gel staining with fluorescent

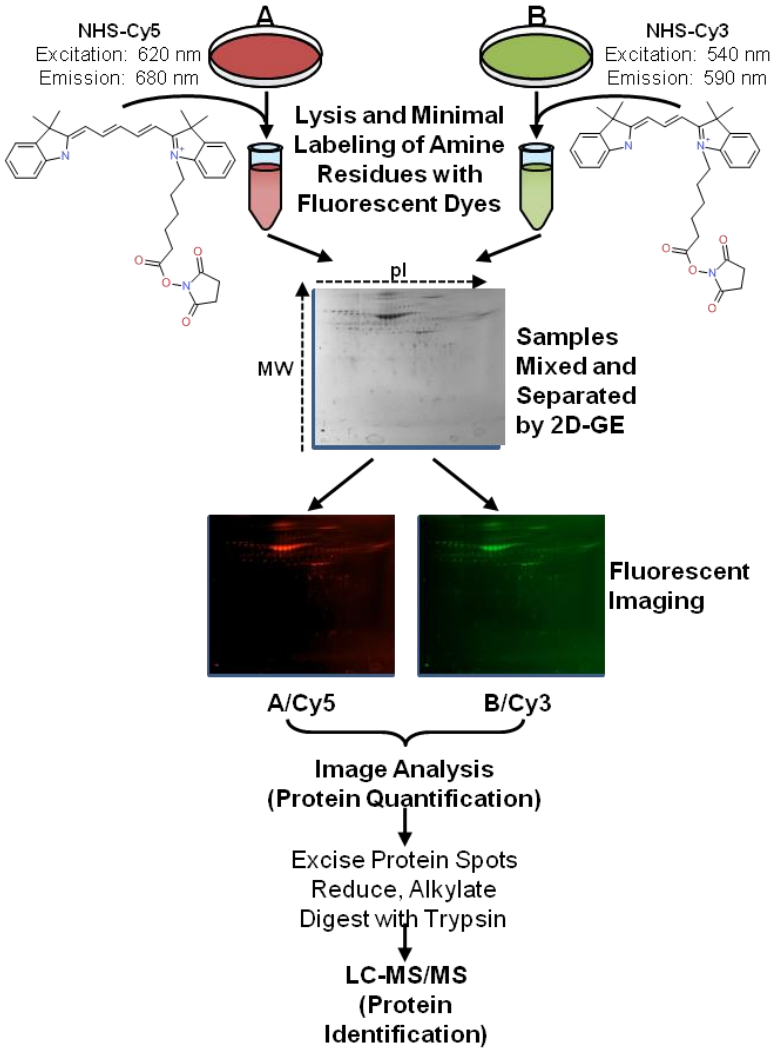

Figure 1.3 Overview of differential gel electrophoresis methodology for the relative quantification of proteins resolved on a two-dimensional electrophoretic gel, in which NHS-Cy dyes are reacted with free amine residues in protein samples and co-eluted on a 2D gel. Spots of interest are excised and identified by bottom-up proteomic analysis.

dyes like Sypro Ruby which is performed after separation. Utilizing fluorescent dyes over the usual coomassie or silver stains provides sensitivity of < 1 fmol over a linear range of detection spanning a dynamic range of 10^4.[54, 55] The most commonly employed reagents for fluorescent labeling in DIGE are synthetic N-hydroxysuccinimidyl (NHS) ester derivatives of cyanine dyes including Cy3 and Cy5, which react with primary amine groups (N-terminal α-amino and lysine ε-amino groups) via nucleophilic substitution.[53] Minimal labeling is often performed (4-8 picomoles of label for each µg of protein), such that < 5% of proteins are labeled. Over-labeling would otherwise result in loss of sample through precipitation. Minimal labeling also reduces interference with downstream mass spectrometric analysis, as a majority of the protein in each spot remains unmodified.[55-57] More recently, iodoacetamide and maleimide derivatives of Cy3 and Cy5 have been utilized for labeling cysteines.[58] As cysteine residues occur less frequently than lysine, the use of cysteine reactive fluorescent tags allows for greater stoichiometric amounts to be labeled without precipitation resulting in increased sensitivity. Greengauz-Roberts et al utilized cysteine reactive dyes for the relative quantification of 1000 proteins in microdissected samples of around 5000 cells, a marked reduction in the amount of sample previously used, and a distinct advantage in the analysis of clinical samples.[59]

While DIGE can potentially allow for the quantification of hundreds of proteins within the same gel, it still presents several challenges. Limitations inherent to 2D-GE are still present in DIGE, including difficulty in characterizing membrane proteins

in addition to proteins possessing very high (>150 kDa) and low (<10 kDa) molecular weight, extreme isoelectric points, and low abundance. Differences in molecular weight between fluorescent labeled proteins and their unlabeled counterparts can also result in differential migration of the protein spots and is most apparent at lower molecular weights[56], causing difficulties for spot picking for MS based identification, as the unlabeled bulk of the protein will not be present at the fluorescing spot. Often, this issue is addressed with the use of a total protein stain such as colloidal coomassie or silver stain. Perhaps one of the greatest pitfalls to DIGE is that although multiple isoforms of a protein may be visualized on the gel, for example multiple combinations of phosphorylation, the specific characteristics of those isoforms are often unobtainable through bottom-up proteomic analysis of the spot. Other variations in DIGE experiments, including fluorescent background[60-62], dye biases[58], and spot volumes[57] demand rigorous statistical analyses to establish thresholds for significant quantitative change with an acceptable number of false positives. Even when limits for identifying significant changes in expression have been put into place, attributing changes in fluorescence to a single protein in a spot that may contain several is problematic and requires further validation with more specific quantification methods.

1.3.2 Mass Spectrometry-Based Quantification

Mass spectrometry based proteomics has led to the development of many platforms that can detect, identify, and quantify up to thousands of proteins from a complex mixture. Two main strategies, termed “bottom-up” and “top-down”, are

currently employed to carry out proteomics analyses. Bottom-up proteomics, a more widely implemented strategy relative to top-down, involves submitting a protein or protein mixture (e.g., cell lysate) to enzymatic digestion rendering the sample as a complex mixture of short peptides with defined N- and C-termini. The peptides are then separated and analyzed via liquid chromatography (often reversed-phase) coupled to electrospray mass spectrometry (ESI-MS) which measures the intact peptide mass and, subsequently, sequences the peptides using tandem mass spectrometry (MS/MS). Searching the precursor ion mass and sequence information against a predicted protein database often results in hundreds or thousands of protein identifications.

However, sequence coverage of the originating protein structures is often low, with many identifications coming from the detection of a single peptide which may correspond to multiple protein forms.[63] These multiple protein forms may each have a unique biological function, potentially possessing diagnostic, prognostic, and/or therapeutic value. Proteins such as brain natriuretic peptide, C-reactive protein, and troponin I and T are already regarded biomarkers in their native forms, but specific modifications (e.g. post-translational modifications (PTMs) and unique cleavages) can result in diagnostic markers that are more specifically related to a disease state.[27-30, 32, 64] Knowing the mass of the intact protein, which includes any modifications that have occurred, that gave rise to a specific proteolytic peptide would be very useful in determining the identity of its specific isoform and whether it is related to a pathological stimulus.

Top-down proteomics provides both the intact mass and sequence information for protein identification and, though not employed as often as the bottom-up technique, continues to mature as a method for identifying and characterizing intact proteins in biological systems.[33] The intact protein mass is first obtained by the mass spectrometer, most often an LTQ-FT-ICR[65] or LTQ-Orbitrap mass spectrometer.[66, 67] Intact proteins are then isolated for MS/MS analysis by any of a multitude of dissociation techniques (e.g. CID[44], IRMPD[45], and ECD[46]) to provide sequence information. The combined intact mass and sequence data enable the specific form of the protein to be determined, locating PTMs and identifying alternative slicing variants and degradation products.[68]

1.3.3 Metabolic Labeling for Intact Protein Quantification

Quantitative mass spectrometry most often utilizes the incorporation of stable isotope labels to create an internal standard with similar physical and chemical properties (except for mass) from which reliable quantitative data can be obtained. One of the first attempts at quantifying large, multiply-charged, intact proteins by Gordon et al. established that quantification could be achieved using relative abundances of an analyte and a structural analogue as an internal standard. However, a three amino acid difference in the amino acid sequence between equine and bovine cytochrome c resulted in hydrophobic differences between the analyte and internal standard. The resulting bias in electrospray ionization efficiency produced two different charge state distributions that limited quantification unless the abundances of all charge states were taken into account.[7] This electrospray bias in

intact proteins has also been demonstrated in various post-translationally modified isoforms of histone H4, in which methylated and acetylated lysine and arginine residues impart greater hydrophobicity, and thus has greater ionization efficiency over its unmodified form.[69] Even the smallest structural differences have the potential to greatly influence quantification, and so stable-isotope labeling has been actively pursued for the quantification of intact proteins.

Metabolic labeling of tissues and cell cultures offer the most efficient means of incorporating stable-isotope labels for quantifying intact proteins and have been adapted from protocols for bottom-up analyses to exclude any proteolytic digestion. With metabolic labeling, illustrated in **Figure 1.4**, separate cultures are supplemented with either a light or heavy form of a nutrient (as in $^{15}N/^{14}N$ labeling[70])

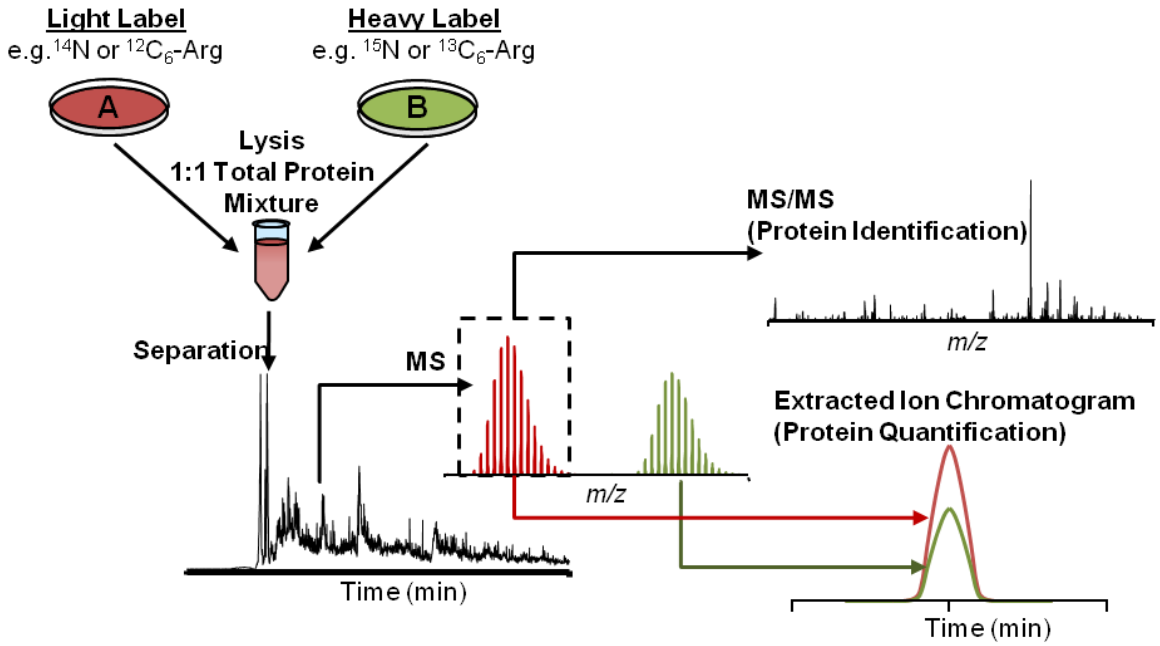

Figure 1.4 General outline of a quantitative top-down proteomics strategy employing metabolic labeling with either $^{15}N/^{14}N$ metabolic labeling or SILAC. Labeled proteins are combined 1:1 then separated, detected, identified and quantified by LC-MS/MS analysis.

or amino acid (as in SILAC[71]), which are combined 1:1 based on total protein concentration and carried through a proteomics workflow including, sample fractionation, LC separation and detection using mass spectrometry followed by bioinformatics analysis where proteins are identified and quantified. Deuterated leucine has been incorporated into model cell cultures, including *E. coli* and *S. cerevisiae*, which facilitated identification of proteins using accurate intact mass and amino acid counting.[72, 73] However, deuterium labeling slightly reduces the hydrophobicity of labeled species relative to hydrogen. The resulting retention time shifts of up to 30 seconds in reversed phase chromatography can hinder quantification, and so deuterated labels are rarely used for top-down quantification.[74, 75] Current intact protein quantification strategies employ stable isotopes that do not hinder liquid chromatography retention times. Originally developed by Chait and coworkers for the peptide level analysis of *S.* cerevisiae, $^{15}N/^{14}N$ labeling was first demonstrated for quantitative proteomics by supplementing minimum culture medium with either natural or stable-isotope labeled ammonium sulfate, resulting in total labeling of every nitrogen atom in the protein structure.[70, 76] Kelleher and co-workers have used $^{15}N/^{14}N$ metabolic labeling to quantify the yeast proteome and recently employed the technique using top-down LC-MS/MS to obtain 50 protein expression ratios.[77, 78]

Stable isotope amino acid labeling with $^{13}C_6$-labeled lysine was introduced by Smith and co-workers for global peptide proteomic analysis as a means of quantifying relative abundances, and also as a means of distinguishing between

near isobaric glutamine and lysine residues.[79] Ong and colleagues later employed a similar strategy utilizing deuterated leucine and eventually $^{13}C_6$-arginine and termed the methodology stable isotope labeling of amino acids in cell culture (SILAC) as another means of quantifying peptides and proteins.[71, 80] Incorporation of ^{13}C and ^{15}N labeled amino acids using the SILAC technique results in co-eluting labeled and unlabeled species and has been widely implemented using isotopically labeled lysine and arginine in conjunction with trypsin digestion for bottom-up proteomics analyses.[80, 81] Waanders and colleagues recently modeled and demonstrated top-down quantification and characterization of the signaling protein Grb2 expressed in *E. coli* using SILAC with $^{13}C_6$-arginine, with accurate quantification at 99% isotope enrichment. The protein was initially purified offline using Ni^{2+} affinity chromatography and analyzed using reversed phase liquid chromatography coupled to an LTQ-Orbitrap mass spectrometer. They also modeled the multiple isotopic envelopes that could result from incomplete incorporation of the heavy isotope amino acid with a 55 kDa protein, but suggested that incomplete labeling could be easily avoided.[82] In addition to relative quantification, SILAC has also been employed for absolute quantification, in which stable isotope labeled recombinant proteins are generated in expressive cell cultures, purified and accurately quantified, and added to an experimental mixture at a known amount.[83]

1.3.4 Other Strategies for Intact Protein Quantification

For global quantification of proteins at the peptide level, there are several types of quantitative strategies in addition to metabolic labeling, including enzymatic,

chemical, and label-free strategies. There are examples of these types of strategies being applied at the intact-protein level, though they have appeared more recently compared to metabolic labeling-based methodologies. Recently, Yan et al have demonstrated the use of the element coded affinity tags for quantification of intact protein standards. In their study lysozyme, ribonuclease A and insulin were reduced and all free thiol groups labeled with the addition of a 10-fold excess of 1,4,7,10-tetraazacyclododecane-1,4,7-triacetic acid-10-maleimidoethylacetamide incorporating either ^{151}Eu of ^{153}Eu and detected using inductively coupled plasma mass spectrometry. While the accurate quantification of tryptic peptides from these proteins was hindered by incomplete digestion, the labeled intact proteins were quantified within 6% with sub-femtomole detection limits.[84]

Label-free quantification strategies present an attractive alternative to metabolic and chemical labeling, which can be time intensive, cost prohibitive, and not amenable to certain types of biological systems (e.g.,, prototrophic organisms, inability for cells to be cultured). Two forms of label-free quantification, utilizing integrated ion abundance or spectral counting, have been most widely applied at the bottom up level. These techniques rely on reproducible instrument methods with a low degree of analytical variability. Muddiman and coworkers established benchmarks of analytical variability for nano-flow liquid chromatography coupled to FT-ICR mass spectrometry including retention time reproducibility, internally and externally calibrated mass measurement accuracy, and ion abundance on the intact species of enriched serum <10 kDa to establish the basis for statistical differentiation

of potential biomarkers between healthy and diseased patients.[85] This strategy, also known as differential mass spectrometry[86], has been recently utilized to quantify intact apolipoproteins in human high-density lipoprotein (HDL) fractions over a 5 fold range of expression in a small patient set as a proof of concept for the utility of this technique to top-down proteomics.[87]

Because top-down identification and quantification methods have largely lagged behind the development of bottom-up techniques, software for the interpretation of top-down has also been slower to develop. However, algorithms have been developed for the determination of accurate mass from high resolution mass spectra of intact proteins (THRASH)[88], and for the prediction of isotopic distributions (ICR-2LS, PNNL, Richland, WA). Kelleher and colleagues have also recently developed software for the interpretation of top-down MS/MS data for database searching.[41] Despite the continuing development of new tools, quantitative top-down methods are largely subject to manual interpretation as the protocols for accurate quantification are still under active development.

The greatest technological challenges to intact protein quantification lie in the separation and detection of intact species, rather in the strategies to quantify them. In gel based techniques such as DIGE, spots representing multiple protein isoforms must be enzymatically digested to be extracted from the gel, making it difficult to fully characterize the originating intact protein. Challenges to the analysis of complex protein mixtures in the past included time consuming off-line separation of samples[34, 35, 89], mass analyzer speed[37], and underdevelopment of database search

software tailored to top-down experiments. However, advances in separation technology and methodology[39, 90-92], the development of faster hybrid mass analyzer technology such as the LTQ-FT mass spectrometer and LTQ-Orbitrap, and recent innovations in top-down database searching software (e.g. ProSight PTM[41]) have created the potential for more widespread utilization of top-down proteomics techniques. Top-down mass spectrometry is still evolving to incorporate quantitative methodologies that have become a commonality with bottom-up analyses. Combined with its powerful capability of molecular characterization, quantitative top-down strategies have the potential to contribute to proteomics on a large scale.

1.4 Quantitative Strategies in Bottom-Up Proteomics

One of the most challenging aspects of proteomic analysis is the relative quantification of proteins once identified. Differential analysis of two-dimensional electrophoretic gels continues to be a widely-used tool for protein separation and quantification[93-97]. However, despite continuous improvements, challenges still exist for the analysis and quantification of proteins that reside at the limits of the gel's analytical capability, whether very high or low molecular weight, extremes in protein isoelectric points, analysis of membrane proteins or proteins of low abundance.

The incorporation of stable isotope labels, both chemically and metabolically, has been widely adopted as a protein quantification strategy as the stable isotope-labeled species possess similar physical and chemical properties as their unlabeled equivalents, but with a different mass. This allows direct comparison of multiple

experimental conditions within the same analytical run. Numerous strategies have been developed for quantification of proteins and peptides using chemical derivatization. Gygi and coworkers developed isotope-coded affinity tags (ICAT) containing a thiol specific reactive group, an isotope labeled linker region containing eight deuteriums (or hydrogens) and a biotin affinity tag[98, 99]. The initial application of ICAT labeling to a complex protein mixture indicated relative standard deviation of peptide area ratios of around 10%, although the potential for greater error has been reported as a result of deuterium labels producing peak retention times that differ between the light and heavy labeled counterparts by as much as 30 seconds[74, 75, 100]. Isobaric tags for relative and absolute quantitation (iTRAQ) is conducive to the direct comparison of four or more experimental samples simultaneously by labeling the N-termini of peptides with a compound containing a reporter group with varying stable isotope (^{13}C, ^{15}N, and ^{18}O) incorporation, a balancing group, and an amine specific reactive group. Tandem mass tags (TMTs) also employ an isobaric chemical labeling strategy for relative quantification. All versions of the tag possess the same total mass, with quantification achieved during MS/MS, in which the reporter groups with varying masses are generated with abundances correlating to the relative abundance of the peptide in each experimental pool[100, 101]. Enzymatically catalyzed ^{18}O-labeling has also been employed, incorporating either ^{16}O or ^{18}O into peptide C-termini during proteolysis to generate light and heavy peptide pairs for relative quantification[102-107].

Stable isotope labels are also introduced to samples for proteomic analysis through metabolic processes including cell and tissue culture, in which growth media is enriched with the label of choice. This strategy was introduced for bacterial proteomics by Oda *et al*.[70]. Deuterated leucine has also been used to label cultures of *E. coli* and *S. cerevisiae* but proved to be less accurate than other labeling methods due to the effect of deuterium on retention time, as discussed *vide* supra[72-75]. $^{15}N/^{14}N$ metabolic labeling has been successfully utilized for quantification of the yeast proteome and has also been successfully utilized for quantification in plant species[108]. Stable isotope labeling by amino acids in cell culture (SILAC) was introduced by Mann and coworkers as a metabolic labeling strategy whereby amino acids containing stable isotopes are supplemented to cell culture growth media to produce peptide analytes that chromatographically co-elute[71, 80]. More recently this strategy has been developed for the *in vivo* labeling of mice[109].

In addition to the stable isotope based methods for protein quantification, label-free strategies continue to be developed. One such strategy uses integrated ion abundance obtained from extracted ion chromatograms to track peptide and protein abundance across experimental samples. Bondarenko and coworkers demonstrated the linear relationship of peptide peak areas over a range of 10 to 1,000 fmol of myoglobin digest spiked into a human plasma sample, with a relative standard deviation near 10%[110, 111]. Another promising method for label-free quantification is spectral counting, where the number of mass spectra contributing to identification of the protein are used as an indication of that protein's abundance[112].

Standard peptides spiked into yeast lysate showed a linear relationship between protein abundance and spectral counts over two orders of magnitude. Additionally, the number of spectral counts can be normalized to protein length allowing for the relative quantification of two different proteins[113]. Overall, these two label-free methods are comparable, with spectral counting reporting to show more sensitivity to changes in protein abundance, and peak area quantification showing slightly higher accuracy in determining relative protein ratios[114].

1.5 Synopsis of Completed Research

The research resented in this dissertation describes the development of quantitative proteomics techniques utilizing both top-down and bottom-up methodologies for the analysis of biological systems impacting agriculture and human health including the filamentous fungus *Aspergillus flavus* and human embryonic stem cells, respectively. **Chapter 2** describes the first application of SILAC labeling with $^{13}C_6$-arginine for the quantification of intact proteins in a complex biological sample, *A. flavus*. Additionally, this chapter addresses the complexities that arise from incomplete isotopic labeling and presents the first version of a mathematical model implemented to address those complexities.

The quantification of intact proteins in human embryonic stem cells is entailed in **Chapter 3** employing SILAC with stable isotope labeled arginine and lysine and achieving label incorporation approaching 99% efficiency. Further improvements are made to the predictive model to account for incomplete labeling and *in vivo*

arginine to proline conversion. This chapter also demonstrates the extent to which the predictive model accounts for the heavy isotopic label.

Chapter 4 describes the comparison of two quantification methodologies, SILAC and spectral counting, employed on a control mixture of normal and SILAC labeled human embryonic stem cell protein extracts using the more high-throughput bottom-up work flow. This comparison explores the capabilities and analytical thresholds for the accurate quantification of proteins in the absence of stable isotope labels, as well as presents comparative information on the number of proteins identified and accurately quantified. This work is expanded upon in **Chapter 5**, in which separate spectral counting and SILAC quantification experiments are performed on differentiating human embryonic stem cells. Here the ability to identify and quantify changes in the relative expression of protein analytes is directly compared given their optimal experimental protocols.

1.6 References

1. Fenn, J.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M., Electrospray Ionization for Mass-Spectrometry of Large Biomolecules. *Science* **1989,** 246, (4926), 64-71.

2. Rayleigh, L., On the equilibrium of liquid conducting masses charged with electricity. *Philos. Mag.* **1882,** 14, 184-186.

3. Dole, M.; Mack, L. L.; Hines, R. L., Molecular Beams of Macroions. *J. Chem. Phys.* **1968,** 49, (5), 2240.

4. Iribarne, J. V.; Thomson, B. A., Evaporation of Small Ions from Charged Droplets. *J. Chem. Phys.* **1976,** 64, (6), 2287-2294.

5. Thomson, B. A.; Iribarne, J. V., Field-Induced Ion Evaporation from Liquid Surfaces at Atmospheric-Pressure. *J. Chem. Phys.* **1979,** 71, (11), 4451-4463.

6. Fenn, J., Ion Formation From Charged Droplets - Roles Of Geometry, Energy, and Time. *J. Am. Soc. Mass Spectrom.* **1993,** 4, (7), 524-535.

7. Gordon, E. F.; Mansoori, B. A.; Carroll, C. F.; Muddiman, D. C., Hydropathic influences on the quantification of equine heart cytochrome c using relative ion abundance measurements by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *J. Mass Spectrom.* **1999,** 34, (10), 1055-1062.

8. Smith, R., New Developments in Biochemical Mass Spectrometry - Electrospray Ionization
Anal. Chem. **1990,** 62, (9), 882-899.

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

10. Comisarow, M. B., Theory of Fourier-Transform Ion-Cyclotron Resonance Mass Spectrometry.2. Signal Modeling for Ion-Cyclotron Resonance. *J. Chem. Phys.* **1978,** 69, (9), 4097-4104.

11. Mitchell, D. W.; Smith, R. D., Cyclotron Motion of 2 Coulombically Interacting Ion Clouds with Implications to Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Phys. Rev. E* **1995,** 52, (4), 4366-4386.

12. Mitchell, D., Prediction of a space charge induced upper molecular mass limit towards achieving unit mass resolution in Fourier transform ion cyclotron resonance mass spectrometry. *J. Mass Spectrom.* **1996,** 31, (7), 771-790.

13. Marshall, A. G., Milestones in Fourier transform ion cyclotron resonance mass spectrometry technique development. *Int. J. Mass. Spectrom.* **2000,** 200, (1-3), 331-356.

14. Grosshans, P. B.; Marshall, A. G., General-Theory of Excitation in Ion-Cyclotron Resonance Mass Spectrometry. *Anal. Chem.* **1991,** 63, (18), 2057-2061.

15. Marshall, A. G.; Hendrickson, C. L., Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Accounts Chem. Res.* **1985,** 18, (10), 316-322.

16. Marshall, A., Fourier transform ion cyclotron resonance detection: principles and experimental configurations. *Int. J. Mass. Spectrom.* **2002,** 215, (1-3), 59-75.

17. Rockwood, A. L.; Busman, M.; Smith, R. D., Coulombic Effects in the Dissociation of Large Highly Charged Ions. *Int. J. Mass. Spectrom.* **1991,** 111, 103-129.

18. Williams, E. R., Proton transfer reactivity of large multiply charged ions. *J. Mass Spectrom.* **1996,** 31, (8), 831-842.

19. Makarov, A., Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. *Anal. Chem.* **2000,** 72, (6), 1156-1172.

20. Kingdon, K. H., A method for the neutralization of electron space charge by positive ionization at very low gas pressures. *Physical Review* **1923,** 21, 408-415.

21. Brooks, P. R.; Herschback, D. R., Kingdon Cage as Molecular Beam Detector. *Rev. Sci. Instrum.* **1964,** 35, 1528-1533.

22. Lewis, R. R., Motion of Ions in the Kingdon Trap. *J. Appl. Phys.* **1982,** 53, 3975-3980.

23. Yang, L. S.; Church, D. A., Confinement of Injected Beam Ions in a Kingdon Trap. *Nucl. Instrum. Meth. B* **1991,** 56-7, 1185-1187.

24. Hardman, M.; Makarov, A. A., Interfacing the Orbitrap Mass Analyzer to an Electrospray Ion Source. *Anal. Chem.* **2003,** 75, 1699-1705

25. Olsen, J. V.; de Godoy, L. M. F.; Li, G. Q.; Macek, B.; Mortensen, P.; Pesch, R.; Makarov, A.; Lange, O.; Horning, S.; Mann, M., Parts per Million Mass Accuracy on an Orbitrap Mass Spectometer via Lock Mass Injection into a C-Trap. *Mol. Cell. Proteomics* **2005,** 4, 2010-2021.

26. Kelleher, N. L.; Lin, H. Y.; Valaskovic, G. A.; Aaserud, D. J.; Fridriksson, E. K.; McLafferty, F. W., Top Down versus Bottom Up Protein Characterization by Tandem High-Resolution Mass Spectrometry. *J. Am. Chem. Soc.* **1999**, (121), 806-812.

27. Hawkridge, A. M.; Heublein, D. M.; Bergen, R. H., III; Cataliotti, A.; Burnett. J. C, Jr.; Muddiman, D. C., Quantitative mass spectral evidence for the absence of circulating brain natriuretic peptide (BNP-32) in severe human heart failure. *Proc. Nat. Acad. Sci. U. S. A.* **2005,** 102, (48), 17442-17447.

28. Das, T.; Sen, A.; Kempf, T.; Pramanik, S. R.; Mandal, C.; Mandal, C., Induction of glycosylation in human C-reactive protein under different pathological conditions. *Biochem. J.* **2003,** 373, 345-355.

29. Das, T.; Mandal, C.; Mandal, C., Variations in binding characteristics of glycosylated human C-reactive proteins in different pathological conditions. *Glycoconjugate J.* **2004,** 20, 537-543.

30. Ardelt, P.; Dorka, P.; Jaquet, K.; Meilmeyer, L. M. G.; Kortke, H.; Korfer, R.; Notohamiprodjo, G., Microanalysis and distribution of cardiac troponin I phospho species in heart areas. *Biol. Chem.* **1998,** 379, (3), 341-347.

31. Katrukha, A. G.; Bereznikova, A. V.; Filatov, V. L.; Esakova, T. V.; Kolosova, O. V.; Pettersson, K.; Lovgren, T.; Bulargina, T. V.; Trifonov, I. R.; Gratsiansky, N. A.; Pulkki, K.; Voipo-Pulkki, L. M.; Gusev, N. B., Degradation of cardiac troponin I: implication for reliable immunodetection. *Clin. Chem.* **1998,** 44, (12), 2433-2440.

32. Labugger, R.; Organ, L.; Collier, C.; Atar, D.; Van Eyk, J. E., Extensive troponin I and T modification detected in serum from patients with acute myocardial infarction. *Circulation* **2000,** 102, (11), 1221-1226.

33. Bogdanov, B.; Smith, R. D., Proteomics by FTICR mass spectrometry: Top down and bottom up. *Mass Spectrom. Rev.* **2005,** 24, (2), 168-200.

34. Kelleher, N. L.; Taylor, S. V.; Grannis, D.; Kinsland, C.; Chiu, H. J.; Begley, T. P.; McLafferty, F. W., Efficient sequence analysis of the six gene products (7-74 kDa) from the Escherichia coli thiamin biosynthetic operon by tandem high-resolution mass spectrometry. *Protein Sci.* **1998,** 7, (8), 1796-1801.

35. Ge, Y.; Lawhorn, B. G.; ElNaggar, M.; Strauss, E.; Park, J. H.; Begley, T. P.; McLafferty, F. W., Top down characterization of larger proteins (45 kDa) by electron capture dissociation mass spectrometry. *J. Am. Chem. Soc.* **2002,** 124, (4), 672-678.

36. Meng, F.; Cargile, B. J.; Patrie, S. M.; Johnson, J. R.; McLoughlin, S. M.; Kelleher, N. L., Processing Complex Mixtures of Intact Proteins for Direct Analysis by Mass Spectrometry. *Anal. Chem.* **2002,** 74, (13), 2923-2929.

37. Siuti, N.; Kelleher, N. L., Decoding protein modifications using top-down mass spectrometry. *Nat. Methods* **2007,** 4, (10), 817-821.

38. Wang, Y.; Balgley, B. M.; Rudnick, P. A.; Lee, C. S., Effects of chromatography conditions on intact protein separations for top-down proteomics. *J. Chromatogr. A.* **2005,** 1073, 35-41.

39. Dillon, T. M.; Bondarenko, P. V.; Ricci, M. S., Development of an analytical reversed-phase high-performance liquid chromatography electrospray ionization mass spectrometry method for characterization of recombinant antibodies. *J. Chromatogr. A* **2004,** 1053, (1053), 299-305.

40. Syka, J. E. P.; Marto, J. A.; Bai, D. L.; Horning, S.; Senko, M. W.; Schwartz, J. C.; Ueberheide, B.; Garcia, B.; Busby, S.; Muratore, T.; Shabanowitz, J.; Hunt, D. F., Novel linear quadrupole ion trap/FT mass spectrometer: Performance characterization and use in the comparative analysis of histone H3 post-translational modifications. *J. Proteome Res.* **2004,** 3, (3), 621-626.

41. Zamdborg, L. L., R. D.; Glowacz, K. J.; Kim, Y.; Viswanathan, V.; Spaulding, I. T.; Early, B. P.; Bluhm, E. J.; Babai, S.; Kelleher, N. L. , ProSight PTM 2.0: improved protein identification and characterization for top down mass spectrometry. *Nucleic Acids Res.* **2007,** 35, W701-W706.

42. Williams, D. K., Jr.; Muddiman, D. C., Parts-per-billion mass measurement accuracy achieved through the combination of multiple linear regression and automatic gain control in a Fourier transform ion cyclotron resonance mass spectrometer. *Anal. Chem.* **2007,** 79, (13), 5058-5063.

43. Nepomuceno, A. I.; Mason, C. J.; Muddiman, D. C.; Bergen, H. R.; Zeldenrust, S. R., Detection of genetic variants of transthyretin by liquid chromatography-dual electrospray ionization Fourier-transform ion-cyclotron-resonance mass spectrometry. *Clin. Chem.* **2004,** 50, (9), 1535-1543.

44. Jennings, K. R., The changing impact of the collision-induced decomposition of ions on mass spectrometry. *Int. J. Mass Spectrom.* **2000,** 200, (1-3), 479-493.

45. Little, D. P.; Speir, J. P.; Senko, M. W.; O'Connor, P. B.; McLafferty, F. W., Infrared Multiphoton Dissociation of Large Multiply-Charged Ions for Biomolecule Sequencing. *Anal. Chem.* **1994,** 66, (18), 2809-2815.

46. Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W., Electron capture dissociation of multiply charged protein cations. A nonergodic process. *J. Am. Chem. Soc.* **1998,** 120, (13), 3265-3266.

47. Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F., Peptide and protein analysis by electron transfer dissociation mass spectrometry. *P. Natl. Acad. Sci. USA* **2004,** 101, (26), 9528-9533.

48. Kelleher, N. L., Top-down proteomics. *Anal. Chem.* **2004,** 76, (11), 196A-203A.

49. Reid, G. E.; McLuckey, S. A., 'Top down' protein characterization via tandem mass spectrometry. *J. Mass Spectrom.* **2002,** 37, (7), 663-675.

50. Kenyon, G. L.; DeMarini, D. M.; Fuchs, E.; Galas, D. J.; Kirsch, J. F.; Leyh, T. S.; Moos, W. H.; Petsko, G. A.; Ringe, D.; Rubin, G. M.; Sheahan, L. C., Defining the mandate of proteomics in the post-genomics era: Workshop report. *Mol. Cell. Proteomics* **2002,** 1, (10), 763-780.

51. Quadroni, M.; James, P., Proteomics and Automation. *Electrophoresis* **1999,** 20, (4-5), 664-677.

52. Patton, W. F., A thousand points of light: The application of fluorescence detection technologies to two-dimensional gel electrophoresis and proteomics *Electrophoresis* **2000,** 21, (6), 1123-1144.

53. 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), 2071-2077.

54. Viswanathan, S.; Unlu, M.; Minden, J. S., Two-dimensional difference gel electrophoresis. *Nat. Protocols* **2006,** 1, (3), 1351-1358.

55. Lilley, K. S.; Razzaq, A.; Dupree, P., Two-dimensional gel electrophoresis: recent advances in sample preparation, detection and quantitaion. *Curr. Opin. Chem. Biol.* **2002,** 6, (1), 46-50.

56. Gharbi, S.; Gaffney, P.; Yang, A.; Zvelebil, M. J.; Cramer, R.; Waterfield, M. D.; Timms, J. F., Evaluation of two-dimensional differential gel electrophoresis for proteomic expression analysis of a model breast cancer cell system. *Mol. Cell. Proteomics* **2002,** 1, (2), 91-98.

57. 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), 377-396.

58. Shaw, J.; Rowlinson, R.; Nickson, J.; Stone, T.; Sweet, A.; Williams, K.; Tonge, R., Evaluation of saturation labelling two-dimensional difference gel electrophoresis fluorescent dyes. *Proteomics* **2003,** 3, (7), 1181-1195.

59. Greengauz-Roberts, O.; Stoppler, H.; Nomura, S.; Yamaguchi, H.; Goldenring, J. R.; Podolsky, R. H.; Lee, J. R.; Dynan, W. S., Saturation labeling with cysteine-reactive cyanine fluorescent dyes provides increased sensitivity for protein expression profiling of laser-microdissected clinical specimens. *Proteomics* **2005,** 5, (7), 1746-1757.

60. Karp, N. A.; Kreil, D. P.; Lilley, K. S., Determining a significant change in protein expression with DeCyder (TM) during a pair-wise comparison using two-dimensional difference gel electrophoresis *Proteomics* **2004,** 4, (5), 1421-1432.

61. Karp, N. A.; Lilley, K. S., Maximising sensitivity for detecting changes in protein expression: Experimental design using minimal CyDyes *Proteomics* **2005,** 5, (12), 3105-3115.

62. Kreil, D. P.; Karp, N. A.; Lilley, K. S., DNA microarray normalization methods can remove bias from differential protein expression analysis of 2D difference gel electrophoresis results *Bioinformatics* **2004,** 20, (13), 2026-2034.

63. Kelleher, N. L.; Lin, H. Y.; Valaskovic, G. A.; Aaserud, D. J.; Fridriksson, E. K.; McLafferty, F. W., Top down versus bottom up protein characterization by tandem high-resolution mass spectrometry. *J. Am. Chem. Soc.* **1999,** 121, (4), 806-812.

64. Katrukha, Degradation of cardiac troponin I: implication for reliable immunodetection. *Clin. Chem.* **1998,** 44, (12), 2433.

65. Syka, J. E. P.; Marto, J. A.; Bai, D. L.; Horning, S.; Senko, M. W.; Schwartz, J. C.; Ueberheide, B.; Gacia, B.; Busby, S.; Muratore, T.; Shabanowitz, J.; Hunt, D. F., Novel linear quadrupole ion trap/FT mass spectrometer: Performance characterization and use in the comparative analysis of histone H3 post-translational modifications. *J. Proteome Res.* **2004,** 3, (3), 621-626.

66. Hu, Q. Z.; Noll, R. J.; Li, H. Y.; Makarov, A.; Hardman, M.; Cooks, R. G., The Orbitrap: a new mass spectrometer. *J. Mass Spectrom.* **2005,** 40, (4), 430-433.

67. Yates, J. R.; Cociorva, D.; Liao, L. J.; Zabrouskov, V., Performance of a linear ion trap-orbitrap hybrid for peptide analysis. *Anal. Chem.* **2006,** 78, (2), 493-500.

68. Du, Y.; Meng, F.; Patrie, S. M.; Miller, L. M.; Kelleher, N. L., Improved Molecular Weight-Based Processing of Intact Proteins for Interrogation by Quadrupole-Enhanced FT MS/MS. *J. Proteome. Res.* **2004**, (3), 801-806.

69. Pesavento, J. J.; Mizzen, C. A.; Kelleher, N. L., Quantitative analysis of modified proteins and their positional isomers by tandem mass spectrometry: Human histone H4 *Anal. Chem.* **2006,** 78, (13), 4271-4280.

70. Oda, Y.; Huang, K.; Cross, F. R.; Cowburn, D.; Chait, B. T., Accurate quantitation of protein expression and site-specific phosphorylation *Proc. Nat. Acad. Sci. U. S. A.* **1999,** 96, (12), 6591-6596.

71. Ong, S.; 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), 376-386.

72. Martinovic, S.; Veenstra, T. D.; Anderson, G. A.; Pasa-Tolic´, L.; Smith, R. D., Selective incorporation of isotopically labeled amino acids for identification of intact proteins on a proteome-wide level. *J. Mass Spectrom.* **2002,** 37, (1), 99-107.

73. Veenstra, T. D.; Martinovic, S.; Anderson, G. A.; Pasa-Tolic´, L.; Smith, R. D., Proteome analysis using selective incorporation of isotopically labeled amino acids. *J. Am. Soc. Mass Spectrom.* **2000,** 11, (1), 78-82.

74. Zhang, R. J.; Sioma, C. S.; Wang, S. H.; Regnier, F. E., Fractionation of isotopically labeled peptides in quantitative proteomics. *Anal. Chem.* **2001,** 73, (21), 5142-5149.

75. Julka, S.; Regnier, F., Quantification in proteomics through stable isotope coding: A review. *J. Proteome Res.* **2004,** 3, (3), 350-363.

76. Washburn, M. P.; Ulaszek, R. R.; Deciu, C.; Schieltz, D. M.; Yates, J. R., Analysis of quantitative proteomic data generated via multidimensional protein identification technology *Anal. Chem.* **2002,** 74, (7), 1650-1657.

77. Parks, B. A.; Jiang, L.; Thomas, P. M.; Wenger, C. D.; Roth, M. J.; Boyne Ii, M. T.; Burke, P. V.; Kwast, K. E.; Kelleher, N. L., Top-down proteomics on a chromatographic time scale using linear ion trap Fourier transform hybrid mass spectrometers. *Anal. Chem.* **2007,** 79, (21), 7984-7991.

78. Du, Y.; Parks, B. A.; Sohn, S.; Kwast, K. E.; Kelleher, N. L., Top-Down Approaches for Measuring Expression Ratios of Intact Yeast Proteins Using Fourier Transform Mass Spectrometry. *Anal. Chem.* **2006,** 78, (3), 686-694.

79. Berger, S. J.; Lee, S.; Anderson, G. A.; Pasa-Tolic, L.; Tolic, N.; Shen, Y.; Zhao, R.; Smith, D. L., High-throughput global peptide proteomic analysis by combining stable isotope amino acid labeling and data-dependent multiplexed-MS/MS *anal. Chem.* **2002,** 74, (19), 4994-5000.

80. Ong, S.; Kratchmarova, I.; Mann, M., Properties of 13C-Substituted Arginine in Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC). *J. Proteome Res.* **2003,** 2, 173-181.

81. de Godoy, L. M. F.; Olsen, J. V.; de Souza, G. A.; Li, G.; Mortensen, P.; Mann, M., Status of complete proteome analysis by mass spectrometry: SILAC labeled yeast as a model system. *Genome Biol.* **2006,** 7, (6), R50-R50.15.

82. Waanders, L. F.; Hanke, S.; Mann, M., Top-down quantitation and characterization of SILAC-labeled proteins. *J. Am. Soc. Mass Spectrom.* **2007,** 18, (11), 2058-2064.

83. Hanke, S.; Besir, H.; Oesterhelt, D.; Mann, M., Absolute SILAC for Accurate Quantitation of Proteins in Complex Mixtures Down to the Attomole Level. *J. Proteome Res.* **2008,** 7, 1118-1130.

84. Yan, X.; Xu, M.; Yang, L. M.; Wang, Q. Q., Absolute Quantification of Intact Proteins via 1,4,7,10-Tetraazacyclododecane-1,4,7-trisacetic acid-10-Maleimidoethylacetamide-Europium Labeling and HPLC Coupled with Species-Unspecific Isotope Dilution ICPMS. *Anal. Chem.* **2010,** 82, (4), 1261-1269.

85. Johnson, K. L.; Mason, C. J.; Muddiman, D. C.; Eckel, J. E., Analysis of the Low Molecular Weight Fraction of Serum by LC-Dual ESI-FT-ICR Mass Spectrometry: Precision of Retention Time, Mass, and Ion Abundance. *Anal. Chem.* **2004,** 76, (17), 5097-5103.

86. Meng, F.; Wiener, M. C.; Sachs, J. R.; Burns, C.; Verma, P.; Paweletz, C. P.; Mazur, M. T.; Deyanova, E. G.; Yates, N. A.; Hendrickson, C. L., Quantitative Analysis of Complex Peptide Mixtures Using FTMS and Differential Mass Spectrometry. *J. Am. Soc. Mass Spectrom* **2007,** 18, 226-233.

87. Mazur, M. T.; Cardasis, H. L.; Spellman, D. S.; Liaw, A.; Yates, N. A.; Hendrickson, R. C., Quantitative analysis of intact apolipoproteins in human HDL by top-down differential mass spectrometry *Proc. Nat. Acad. Sci. U.S.A.* **2010,** 107, (17), 7728-7733.

88. Horn, D. M.; Zubarev, R. A.; McLafferty, F. W., Automated reduction and interpretation of high resolution electrospray mass spectra of large molecules. *J. Am. Soc. Mass Spectrom.* **2000,** 11, (4), 320-332.

89. Meng, F. Y.; Cargile, B. J.; Patrie, S. M.; Johnson, J. R.; McLoughlin, S. M.; Kelleher, N. L., Processing complex mixtures of intact proteins for direct analysis by mass spectrometry. *Anal. Chem.* **2002,** 74, (13), 2923-2929.

90. Wang, Effects of chromatography conditions on intact protein separations for top-down proteomics. *J. Chromatogr.* **2005,** 1073, (1-2), 35.

91. Tran, J. C.; Doucette, A. A., Gel-eluted liquid fraction entrapment electrophoresis: An electrophoretic method for broad molecular weight range proteome separation. *Anal. Chem.* **2008,** 80, (5), 1568-1573.

92. Vellaichamy, A.; Tran, J. C.; Catherman, A. D.; Lee, J. E.; Kellie, J. F.; Sweet, S. M. M.; Zamdborg, L.; Thomas, P. M.; Ahlf, D. R.; Durbin, K. R.; Valaskovic, G. A.; Kelleher, N. L., Size-Sorting Combined with Improved Nanocapillary Liquid Chromatography-Mass Spectrometry for Identification of Intact Proteins up to 80 kDa. *Anal. Chem.* **2010,** 82, (4), 1234-1244.

93. Link, A.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvic, B. M.; Yates, J. R., Direct analysis of protein complexes using mass spectrometry. *Nat. Biotechnol.* **1999,** 17, (7), 676-682.

94. Shevchenko, A.; Jensen, O. N.; Podtelejnikov, A. V.; Sagliocco, F.; Wilm, M.; Vorm, O.; Mortensen, P.; Shevchenko, A.; Boucherie, H.; Mann, M., Linking genome and proteome by mass spectrometry: Large-scale identification of yeast proteins from two dimensional gels. *Proc. Nat. Acad. Sci. U. S. A.* **1996,** 93, (25), 14440-14445.

95. Gygi, S. P.; Rochon, Y.; Franza, B. R.; Aebersold, R., Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* **1999,** 19, (3), 1720-1730.

96. Garrels, J. I.; McLaughlin, C. S.; Warner, J. R.; Futcher, B.; Latter, G. I.; Kobayashi, R.; Schwender, B.; Volpe, T.; Anderson, D. S.; Mesquita-Fuentes, R.; Payne, W. E., Proteome studies of Saccharomyces cerevisiae: Identification and characterization of abundant proteins. *Electrophoresis* **1997,** 18, (8), 1347.

97. Boucherie, H.; Sagliocco, F.; Maillet, I.; Labarre, J.; Perrot, M., Two-dimentional gel protein database of Saccharomyces cerevesiae. *Electrophoresis* **1996,** 17, (11), 1683-99.

98. 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), 994-999.

99. Li, J. X.; Steen, H.; Gygi, S. P., Protein profiling with cleavable isotope-coded affinity tag (cICAT) reagents - The yeast salinity stress response. *Mol. Cell. Proteomics* **2003,** 2, (11), 1198-1204.

100. Ross, P. L.; Huang, Y. L. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Prkayastha, S.; Jahasz, P.; Martin, S.; Bartlet-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J., Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* **2004,** 3, (12), 1154-1169.

101. Thompson, A.; Schafer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Johnstone, R. A. W.; Mohammed, A. K. A.; Hamon, C., Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* **2003,** 75, (8), 1895-904.

102. Desiderio, K. M.; Kai, M., Preparation of Stable Isotope-Incorporated Peptide Internal Standards for Field Desorption Mass-Spectrometry Quantification of Peptides in Biologic Tissue. *Biomed. Mass Spectrom.* **1983,** 10, (8), 471-479.

103. Mirgorodskaya, O. A.; Kozmin, Y. P.; Titov, M. I.; Korner, R.; Sonksen, C. P.; Roepstorff, P., Quantitation of Peptides and Proteins by Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry Using O-18-Labeled Internal Standards. *Rapid Commun. Mass Spectrom.* **2000,** 14, (14), 1226-1232.

104. Yao, X. D.; Afonso, C.; Fenselau, C., Dissection of proteolytic O-18 labeling: Endoprotease-catalyzed O-16-to-O-18 exchange of truncated peptide substrates. *J. Proteome Res.* **2003,** 2, (2), 147-152.

105. Yao, X. D.; Freas, A.; Ramirez, J.; Demirev, P. A.; Fenselau, C., Proteolytic O-18 Labeling for Comparative Proteomics: Model Studies with Two Serotypes of Adenovirus. *Anal. Chem.* **2001,** 73, (13), 2836-2842.

106. Mason, C. J.; Therneau, T. M.; Eckel-Passow, J. E.; Johnson, K. L.; Oberg, A. L.; Olsen, J. E.; Nair, K. S.; Muddiman, D. C.; Bergen, H. R., A method for automatically interpreting mass spectra of 18O-labeled isotopic clusters. *Mol. Cell. Proteomics* **2007,** 6, (2), 305-318.

107. Johnson, K. L.; Muddiman, D. C., A method for calculating O-16/O-18 peptide ion ratios for the relative quantification of proteomes. *J. Am. Soc. Mass Spectrom* **2004,** 15, (4), 437-445.

108. Schaff, J. E.; Mbeunkui, F.; Blackburn, K.; Bird, D. M.; Goshe, M. B., SILIP: a novel stable isotope labeling method for in planta quantitative proteomic analysis. *The Plant Journal`* **2008,** 56, 840-854.

109. Kruger, M.; Moser, M.; Ussar, S.; Thievessen, I.; Luber, C. A.; Forner, F.; Schmidt, S.; Zanivan, S.; Fassler, R.; Mann, M., SILAC Mouse for Quantitative Proteomics Uncovers Kindlin-3 as an Essential Factor for Red Blood Cell Functioin. *Cell* **2008,** 134, 353-364.

110. Chelius, D.; Bondarenko, P. V., Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *J. Proteome Res.* **2002,** 1, (4), 317-323.

111. Bondarenko, P. V.; Chelius, D.; Shaler, T. A., Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography-tandem mass spectrometry. *Anal. Chem.* **2002,** 74, (18), 4741-4749.

112. Wolters, D.; Washburn, M. P.; Yates, J. R., An automated multidimensional protein identification technology for shotgun proteomics. *Anal. Chem.* **2001,** 73, (23), 5683-5690.

113. Zybailov, B.; Mosley, A. L.; Sardiu, M. E.; Coleman, M. K.; Florens, L.; Washburn, M. P., Statistical analysis of membrane proteome expression changes in Saccharomyces cerevisiae. *J. Proteome Res.* **2006,** 5, (9), 2339-2347.

114. Old, W. M.; Meyer-Arendt, K.; Aveline-Wolf, L.; Pierce, K. G.; Mendoza, A.; Sevinsky, J. R.; Resing, K. A.; Ahn, N. G., Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol. Cell. Proteomics* **2005,** 4, (10), 1487-1502.

Chapter 2

Top-Down Identification and Quantification of Stable Isotope Labeled Proteins from *Aspergillus flavus* using nano-LC couples to LTQ-FT-ICR Mass Spectrometry

2.1 Introduction

Mass spectrometry based proteomics is a robust and rapidly developing field of interest seeking to understand "the identities, quantities, structures, and biochemical and cellular functions of all proteins in an organism, organ, or organelle, and how they vary in space, time, and physiological state".[1] While the proteomics field is yet to achieve this ultimate goal, two main strategies, termed "bottom-up" and "top-down", are currently employed to pursue it. Bottom-up proteomics, a more widely implemented strategy relative to top-down, involves submitting a protein or protein mixture (e.g., cell lysate) to enzymatic digestion rendering the sample as a complex mixture of short peptides with defined C-termini. The peptides are then separated and analyzed via liquid chromatography (often reversed phase) coupled to electrospray mass spectrometry (ESI-MS) which sequences the peptides using tandem mass spectrometry (MS/MS). Searching the precursor ion mass and sequence information against a predicted protein database often results in hundreds or thousands of protein identifications.

However, sequence coverage of the original proteins is often low, with many identifications coming from the detection of a single peptide which may correspond to multiple protein forms.[2] These multiple protein forms may have unique biological function, potentially possessing diagnostic, prognostic, and/or therapeutic properties.

Proteins such as brain natriuretic peptide, C-reactive protein, and troponin I and T are already regarded biomarkers in their native forms, but specific modifications (e.g. post-translational modifications (PTMs) and unique cleavages) can result in diagnostic markers that are more specifically related to a disease state.[3-8] Knowing the mass of the protein that gave rise to a specific proteolytic peptide would be very useful in determining its identity and whether it is a form related to a pathological stimulus.

Top-down proteomics provides both the intact mass and sequence information for protein identification and, though not employed as often as the bottom-up technique, continues to mature as a method for identifying and characterizing intact proteins in biological systems.[9] The intact protein mass is first obtained by the mass spectrometer, followed by isolation for MS/MS analysis to provide sequence information. The combination of intact mass and sequence data enable the specific form of the protein to be determined, locating PTMs and identifying alternative slicing variants and degradation products.

Top-down proteomics has lagged behind the more widely used bottom-up strategies due, in large part, to the technological challenges of measuring and identifying large intact proteins. These challenges have included time consuming off-line separation of samples[2, 10-12], mass analyzer speed[13], and underdevelopment of database search software tailored to top-down experiments. However, advances in separation technology and methodology[14, 15], the development of faster hybrid mass analyzer technology such as the LTQ-FT mass spectrometer[16], and recent

innovations in top-down database searching software (e.g. ProSight PTM[17]) have created the potential for more widespread utilization of top-down proteomics techniques. Electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry (ESI-FT-ICR-MS) is well suited for carrying out top-down proteomics measurements of intact multiply-charged proteins. The combination of high mass accuracy <3 ppm[18, 19], high resolving power, and a multitude of MS/MS dissociation techniques (e.g. CID[20], IRMPD[21], and ECD[22]) provide unprecedented confidence for protein identification and characterization (e.g. localization of PTMs, unique cleavages, and protein variants).[23, 24]

In addition to identifying and characterizing protein structures, understanding how they vary in quantity between different states is an integral goal of proteomic science. One of the first attempts at quantifying large, multiply-charged, intact proteins by Gordon et al. established that quantification could be achieved using relative abundances of an analyte and a structural analogue as an internal standard. However, a three amino acid difference in the amino acid sequence between equine and bovine cytochrome c resulted in hydrophobic differences between the analyte and internal standard. The resulting ESI bias produced two different charge state distributions that limited quantification unless the abundances of all charge states were taken into account.[25]

Top-down analyses have also been used in conjunction with stable isotope labeling. Intact protein identification using amino acid counting and accurate mass has been accomplished using isotopic labeling of model cell cultures, including *E.*

coli and *S. cerevisiae,* with deuterated leucine.[26, 27] However, deuterium labeling hinders quantification on a chromatographic scale as elution times vary between labeled (deuterium) and unlabeled analytes.[28, 29] Current intact protein quantification strategies currently employ stable isotopes that do not hinder liquid chromatography retention times. Kelleher and co-workers have used $^{15}N/^{14}N$ metabolic labeling to quantify the yeast proteome[30] and recently employed the technique using top-down LC-MS/MS to identify 22 proteins from *S. cerevisiae*. [31]

Ong and colleagues introduced stable isotope labeling of amino acids in cell culture (SILAC) as another means of quantifying peptides and proteins.[32, 33] Incorporation of ^{13}C and ^{15}N labeled amino acids using the SILAC technique results in co-eluting labeled and unlabeled species and has been widely implemented using isotopically labeled lysine and arginine in bottom-up experiments.[33, 34] Waanders and colleagues recently modeled and demonstrated top-down quantification and characterization of the signaling protein Grb2 expressed in *E. coli* using SILAC, with accurate quantification at 99% isotope enrichment. The protein was initially purified offline using Ni^{2+} affinity chromatography and analyzed using reversed phase liquid chromatography coupled to an LTQ-Orbitrap mass spectrometer. They also modeled the multiple isotopic envelopes that could result from incomplete incorporation of the heavy isotope amino acid with a 55 kDa protein.[35]

In this paper, we have assessed the performance of whole-organism SILAC labeling as a tool for top-down proteomics on the filamentous fungus *Aspergillus flavus* using a $^{13}C_6$-arginine label and nLC-LTQ-FT-ICR-MS. *A. flavus* contaminates

many plant species with the carcinogen aflatoxin resulting in economic losses in the hundreds of millions of dollars annually.[36] In addition to the agricultural impact of *A. flavus*, inhalation of conidia produced by the fungus can cause aspergillosis in patients with compromised immune systems. Secreted proteins of this organism has been previously examined by Medina et al., but were limited at that time by the absence of an annotated genome for the organism from which to identify many of their proteins.[37] A currently available annotated genome sequence for this fungus and the ability to easily culture this organism made it a good model for these studies.

SILAC labeling of this organism provides a powerful tool for measuring the biological effects of temperature, nutrition, and antifungal agents. As an example, this strategy will allow the identification, characterization, and quantification of regulatory proteins that are difficult of study by transcriptional profiling. Many regulatory proteins are constitutively expressed at a low level and are activated by protein localization or post-translational modification. Such regulatory elements are better studied by a proteomic analysis than by genomics.

This study represents our initial efforts to establish a top-down analysis of SILAC labeled *A. flavus*. We establish effective chromatography conditions for intact protein separation and discuss the limits and issues with amino acid incorporation in the culturing process.

2.2 Experimental

2.2.1 Culture of *A. flavus* for SILAC

Stock cultures of *Aspergillus flavus* strain NRRL 3357 were stored at -80 °C in 35% glycerol. Inoculum was prepared by streaking an aliquot of the stock onto potato dextrose agar (PDA) (DIFCO; Lawrence, KS) supplemented with either i) 340 µg/mL $^{12}C_6$-L-arginine (Sigma-Aldrich; St. Louis, MO) or ii) 340 µg/ml $^{13}C_6$-L-arginine (Cambridge Isotopes; Andover, MA). After growth of the culture at 37 °C for 5 days, conidia were dislodged from the cultures with a glass rod and suspended in 0.05% Triton X-100. The conidial suspension was quantified using a hemocytometer. Flasks containing 100 ml of A&M medium[38] supplemented with 340 µg/mL of either arginine $^{12}C_6$ or arginine $^{13}C_6$ were inoculated with conidial suspension to a final concentration of 10^6 conidia/mL. These liquid cultures were grown for 24 h at 28 °C with shaking at 200 rpm allowing sufficient time for the conidia to germinate (around 6 to 7 hours) and grown for more than long enough to sufficiently incorporate the supplemented arginine (approximately 1 to 2 hours per cell cycle). Safety considerations: 1) Sample spores were treated as a potential respiratory hazard and handled in a LABGARD Laminar Flow Biological Safety Cabinet (Nuaire, Inc Minneapolis, MN). 2) Samples were handled with gloves at all time to prevent and contamination from any aflatoxin present in the sample.

2.2.2 Protein Isolation and Sample Preparation

Fungal mycelium was separated from the medium by filtration through miracloth. The retained fungal mat was washed with cold PBS, removed from the

filter, immediately frozen in liquid nitrogen, lyophilized, and stored at -80 °C until extracted for proteins. A 0.2 g sample of lyophilized tissue was ground with a pestle in a mortar containing liquid nitrogen and 0.1 g of 150-200 µm glass beads (Sigma-Aldrich). The resultant fine powder was placed into 1.5 mL microcentrifuge tubes, resuspended with 1 mL of cold PBS, and ground an additional 5 min at 4 °C in a Vortex Genie (Scientific Industries, Bohemia, NY). The homogenate was spun for 5 min at 12,000 × g to pellet large cellular debris. The supernatant was moved to a new tube and spun at 100,000 × g for 45 min to pellet membranes and other debris. The resulting supernatant was submitted to acetone precipitation[39] to remove any residual detergents from culture. The pellet was resuspended in 20 mM monobasic potassium phosphate (pH 7.0) and passed through a 10 kDa MWCO filter. The resulting retentate (total protein) was quantified using a bicinchoninic acid (BCA) assay (BioRad, Hercules, CA) and utilized for SDS-PAGE and LC-MS experiments.

2.2.3 1D SDS-PAGE

Electrophoresis was performed using 12.5% Tris-HCl Criterion Gels (BioRad, Hercules, CA) in 25 mM Tris buffer containing 192 mM L-Glycine and 3.5 mM sodium dodecyl sulfate (SDS). 50 µg of the processed cell lysate (> 10 kDa) was loaded onto the gel and run at 150 V for 60 minutes and stained using 50 mL of Bio Safe Coomassie Blue G-250 (BioRad) according to the manufacturer's protocol.

2.2.4 Online LC-MS

All LC solvents were purchased from Burdick and Jackson (Muskegon, MI). Reversed phase liquid chromatography was performed using a 75 µm i.d. PicoFrit

capillary column (New Objective, Woburn, MA) with a 15 □m emitter tip packed in-house with 5 µm mRP-C18 silica stationary phase (Agilent, Palo Alto, CA). The packed volume had dimensions 75 µm i.d. × 150 mm and was operated at room temperature. Samples (1 µg total protein) were injected using a PAL Autosampler (LEAP Technologies, Carrboro, NC) and over the course of 10 minutes trapped and washed on a custom built mRP-C18 OPTI-PAK trap cartridge (Optimize Technologies; Oregon City, OR) with 100% Mobile Phase A (95/5 water/acetonitrile) at 1 µL/min until a 10 port switching valve (VICI, Houston, TX) was triggered to move the sample in-line with the gradient. Elution was carried out by a Chorus 220 nano-flow pump (CS Analytics, Zwingen, Switzerland) at 250 nL/min with mobile phases containing 95/5 (v/v) (Mobile Phase A) and 5/95 (Mobile Phase B) Water and Acetonitrile, respectively. The ion pairing reagent used was 0.2% formic acid (Sigma Aldrich, St. Louis, MO) in both mobile phases. The LC gradient was held at initial conditions of 10% B for 10 minutes followed by a ramp to 25% B over 5 minutes; 65% B was reached over the next 70 minutes, increased to 95% B in five minutes and held for an additional ten minutes before re-equilibrating at 5% B.

Mass spectrometry measurements were performed on a 7T LTQ-FT Ultra from Thermo Scientific (Brehmen, Germany). The pulse sequence consisted of four events where mass measurements were all performed inside the ICR cell. The pulse sequence provided for a broadband acquisition in profile mode, followed by three data dependent MS/MS measurements. All events used 1 microscan to determine ionization time to reach the target AGC limit of 5×10^5. The resolving

power of all four events was set at $100{,}000_{FWHM}$ at *m/z*=400. MS/MS settings used an isolation width of 7 *m/z* and a normalized collision energy of 18% for a duration of 30 ms. CID was performed the three most abundant *m/z* values from the precursor ion scan in the LTQ followed by product ion detection in the ICR cell. Precursor ions with 1, 2, and 3 charges were excluded from MS/MS and dynamic exclusion for 60 seconds was used to reduce redundant analysis of the same precursors.

2.2.5 Data Analysis

RAW data files were processed manually, obtaining precursor *m/z* values from the MS/MS spectrum header and importing the MS/MS mass values into ProSightPC (Thermo Electron, Waltham MA) using a THRASH[40] algorithm to determine the monoisotopic peak of analyte signals with a S/N ratio of 5:1 or greater. The database was searched initially with absolute mass and biomarker search modes with precursor and fragment ion tolerances of ±5 ppm and again with tolerances of ±1 Da to accommodate sample handling modifications to the mass (e.g. deamidation) or error in monoisotopic peak picking by the THRASH algorithm. Sequences reported from the search were confirmed using arginine counting from the precursor ion spectra. Analysis of SILAC labeled protein distributions and theoretical distributions with varying amounts of arginine labeling was carried out using ICR-2LS (PNNL, Richland, WA) and Microsoft Excel® (Redmond, WA).

2.3 Results and Discussion

Two cultures of *Aspergillus flavus* were grown, one in standard medium and another in which $^{13}C_6$-arginine was substituted for $^{12}C_6$-arginine. After growth,

mycelium from both cultures was lysed and the lysate combined in a 1:1 mixture based on total protein determination. After acetone precipitation the resuspended pellet was filtered through a molecular weight cut-off filter to remove detergents. After processing, the 1:1 lysate mixture of the two cultures was analyzed via nLC-LTQ-FT-ICR-MS.

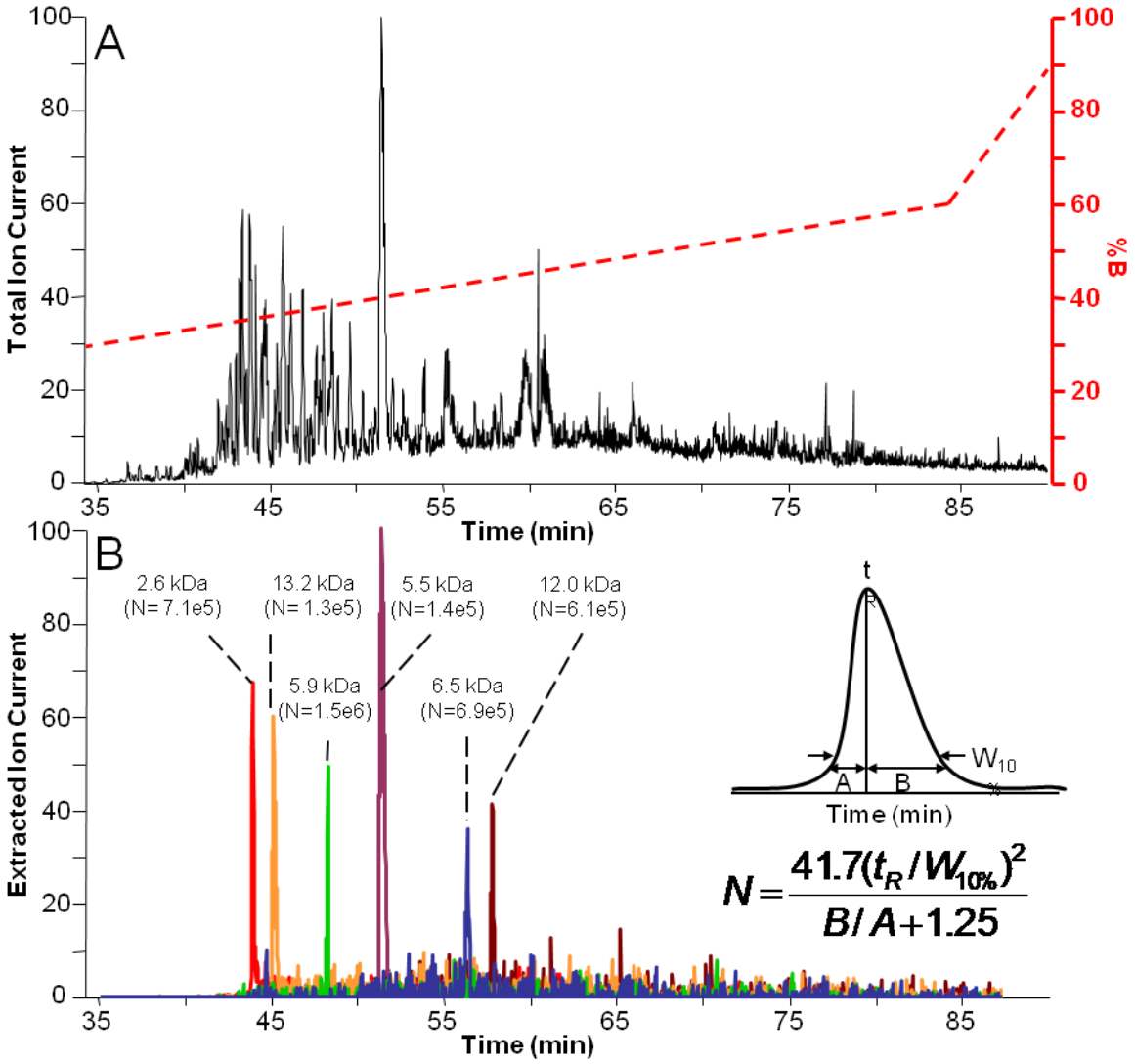

Figure 2.1 Base peak chromatogram **(A)** of reversed-phase separation of whole cell lysates from *A. flavus* and **(B)** Extracted ion chromatograms of peaks selected to analyze column performance using the Foley-Dorsey Equation.

A representative base peak chromatogram showing total ion current of cell lysate samples processed from *A. flavus* is shown in **Figure 2.1A**. The left axis shows the relative total ion current and the right axis shows the % B of the mobile phase. The gradient profile represented by the dashed line spans 30% to 95% B and is superimposed upon the LC elution profile after being adjusted for dwell volume. Proteins producing precursor and product ion spectra eluted from 35 to 85 minutes, spanning most of the chromatographic gradient.

Figure 2.1B shows representative extracted ion chromatograms superimposed on the same chromatographic time scale axis. The separation efficiency of these species was evaluated using the equation shown in the inset and serves as a metric to judge the performance of the separation and set a benchmark against which future separations can be compared. The number of theoretical plates (N) was calculated using the exponentially modified Gaussian peak model developed by Foley and Dorsey[41], which takes account peak asymmetry (B/A) in addition to retention time (t_R) and peak width at 10% height ($W_{10\%}$). This model is well-suited for our investigations due to the asymmetric nature of the extracted chromatographic peaks. These peak shapes suggest non-ideal separation conditions that may include column temperature, ion pairing reagent used, and flow rate. Theoretical plate values ranged from 1.4×10^5 to 1.5×10^6 of the representative peaks having masses from 2.6 to 13.2 kDa with widths at 10% height ranging from 15 to 20 seconds.

The molecular weight distribution of the predicted proteome of *A. flavus* is shown in **Figure 2.2** with a 0.1 kDa bin width. The *A. flavus* proteome database consisted of 12,847 predicted proteins[42]. The predicted proteins range in mass from 1.9 to 858 kDa, but for simplicity the 101 proteins exceeding 100 kDa are not shown. In a single LC-MS experiment, 1,318 unique species were observed spanning 1.9 to 18.7 kDa which are plotted in **Figure 2.2** as a function of molecular weight. The same sample used in the LC-MS experiment was also run on a one-dimensional SDS-polyacrylamide gel showing bands spanning the entire molecular weight range of the gel from 10 to 100 kDa.

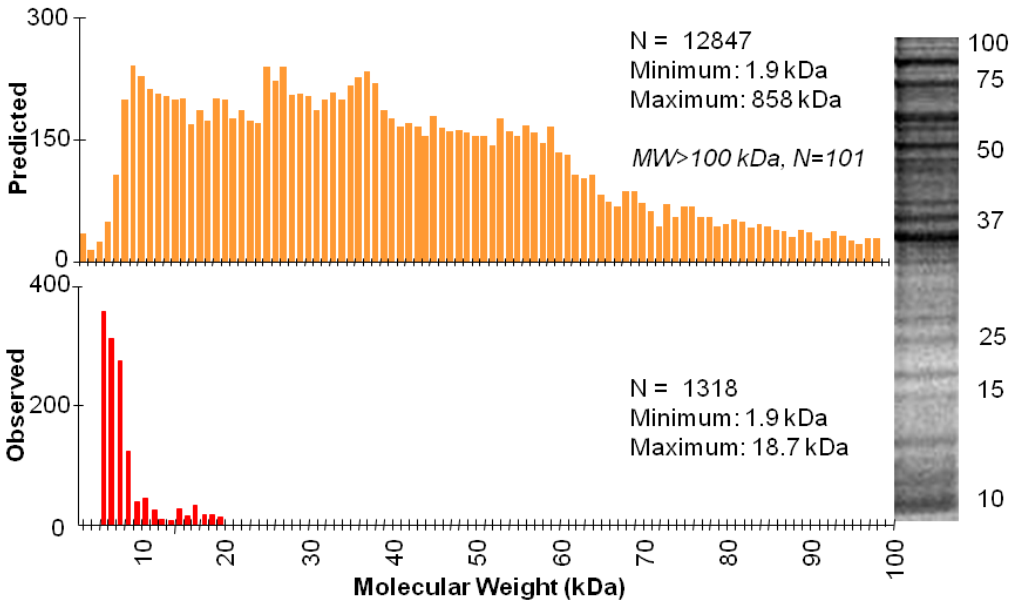

Figure 2.2 Histograms showing the molecular weight distributions in 0.1 kDa bins of the predicted proteome of *A. flavus* (top) and the observed molecular weights from a single LC-MS experiment (bottom). The distribution of molecular weight bands of intact species is visualized on a 1D-gel (right).

Product ion spectra from many of the low molecular weight precursor spectra were identified as protein fragments derived from proteins with significantly larger

molecular weight when searched using ProSightPC. This suggested some *in vivo* processing or biological degradation of the sample from endogenous proteases. The overall low molecular weight distribution may also be attributed to electrospray bias for low molecular weight species when co-eluting with larger analytes, and/or chromatography biases.[14] The addition of other front end fractionation methods to further simplify the sample in addition to tailoring current instrument parameters including the ICR detection time, column temperature, gradient profile, and electrospray voltages may improve detection of larger intact species.

The top-down MS analysis of the SILAC labeled cell lysate mixture consisted of a precursor ion mass spectrum followed by three data dependent MS/MS measurements. **Figure 2.3** shows three representative sets of MS and MS/MS spectra of unique SILAC protein pairs with their respective IDs using ProSightPC. The extracted ion chromatogram (XIC) of the monoisotopic masses of the SILAC pair is shown to the left in each example, demonstrating co-elution of the pair. **Figure 2.3A** demonstrates the identification of a peptide sequence corresponding to a fragment of the glycolytic enzyme fructose-bisphosphate aldolase in *A. flavus*. The isotopic envelopes believed to be a SILAC pair from the precursor ion spectrum were confirmed by examining their mass difference and extracted ion chromatograms of the heavy and light monoisotopic masses. The mass difference between the peaks corresponded to the incorporation of two $^{13}C_6$-arginines into the heavy-labeled protein fragment.

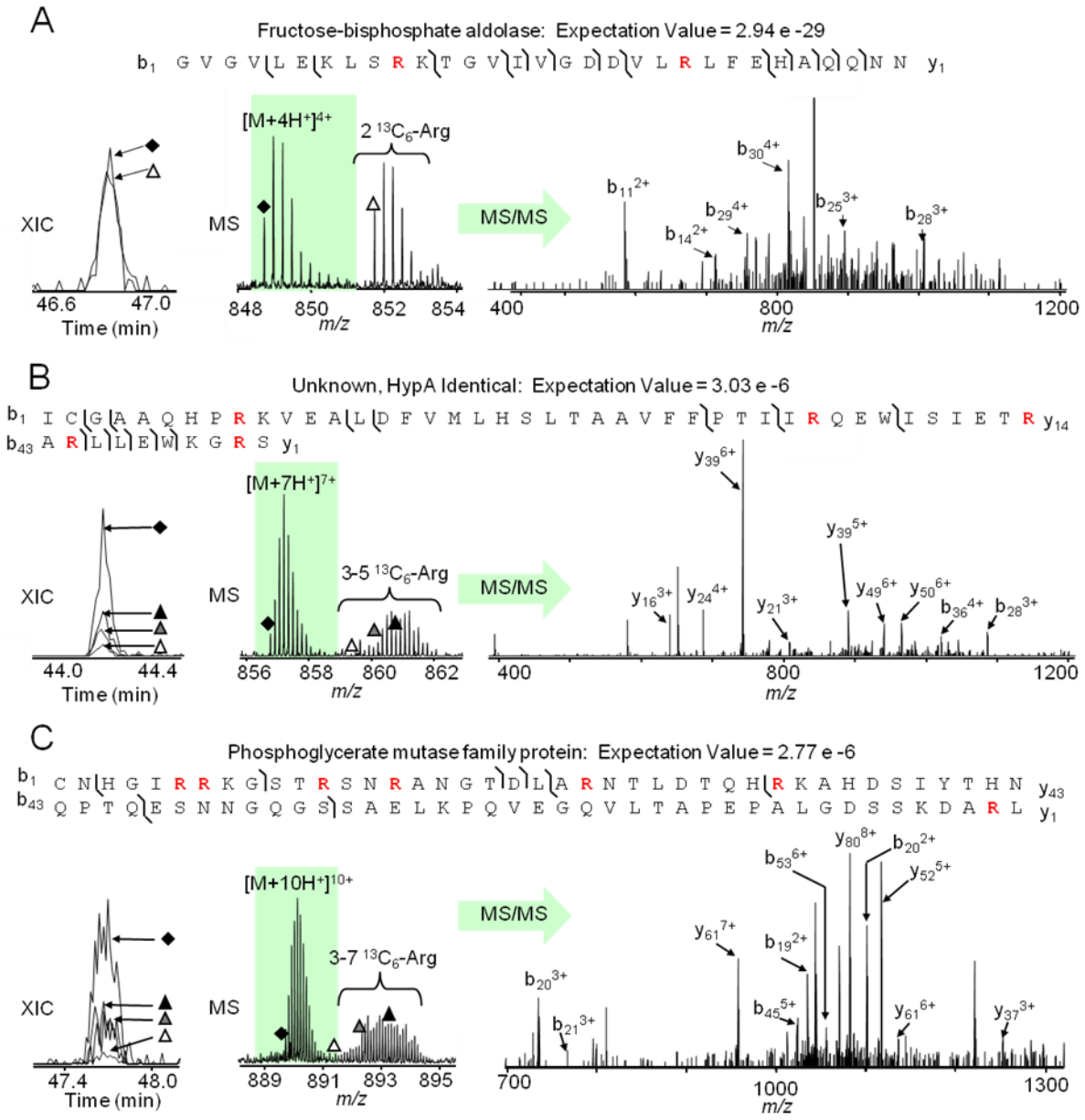

Figure 2.3 Extracted ion chromatograms and MS spectra of SILAC pairs showing the corresponding elution of light and heavy isotopic envelopes of SILAC labeled proteins. Arginine counting and MS/MS spectra of the isolated light peak resulted in the identified sequence identified with confidence reported as an expectation value in ProSightPC.

The lighter of the two peaks had an absolute abundance of 1×10^6 and was isolated for MS/MS resulting in a b- and y-ion series that was searched and identified using ProSightPC. Although the lighter peaks were not exclusively selected for collisional

dissociation, confident identifications only arose from selection of the light peak as the search algorithm did not account for SILAC labeling and the mass difference of labeled arginines was too great to result in successful identification. Seventeen product ions from the MS/MS spectrum matched the predicted sequence giving an expectation value of 1.77×10^{-33}. In addition to the mass accuracy and the MS/MS sequence coverage, the identification of the protein was further confirmed using the number of predicted and observed arginines determined by mass difference of the SILAC pairs divided by the difference in heavy and light arginine (ΔM/6.02013) in the precursor spectrum. Relative quantification of the heavy and light SILAC peak abundances corresponded strongly to a 1:1 ratio.

Figures 2.3B and 2.3C display protein fragments that were identified as HypA identical protein and Phosphoglycerate mutase family protein having absolute abundances of 2×10^6 and 3×10^6, respectively. They were likewise confirmed as SILAC pairs using monoisotopic mass extracted ion chromatograms and identified using MS/MS spectra confirmed with arginine counting as previously described. Quantification was hindered, however, by the complexity of the heavy peak distributions which increased as the number of arginines in the protein increased. Despite the multiple isotopic distributions present in these complex heavy-labeled species, the light and heavy distributions were still regarded as a single SILAC pair. Incomplete incorporation of $^{13}C_6$-arginine into the labeled *A. flavus* cultures due to the presence of endogenously produced arginine is the most likely cause resulting in these complex distributions. Overall, data-dependent MS/MS analysis was

performed over two orders of magnitude of absolute ion abundance ranging from 1 × 10^4 to 3×10^6 and resulted in confident identification of 22 proteins (**Table 2.1)**

Table 2.1 Proteins Identified from *Aspergillus flavus*

Protein Description	Precursor Mass (Da)	MMA (ppm)	p-Score
glutathione peroxidase family protein	2586.3373	0.67	5.87E-51
hypothetical protein	3521.8521	1.49	3.71E-39
Cu,Zn superoxide dismutase SOD1	3481.6271	0.68	7.16E-37
fructose-bisphosphate aldolase, class II	3390.8453	2.34	1.77E-33
U5 snRNP component	5990.1762	0.44	1.15E-30
Cofilin/tropomyosin-type actin-binding protein	4221.0606	0.79	3.23E-28
Peptidyl-prolyl cis-trans isomerase	5320.8437	1.00	3.44E-21
5-methyltetrahydropteroyltriglutamate	2465.2613	1.35	1.52E-20
chaperonin	7254.7827	0.65	1.93E-17
ubiquitin (UbiA)	8521.5737	0.97	3.77E-17
AhpC/TSA family protein	3317.7317	0.55	2.29E-16
oxidoreductase	2510.3505	0.45	1.01E-15
FKBP-type peptidyl-prolyl isomerase	11337.7914	1.24	1.25E-14
malate dehydrogenase, NAD-dependent	3012.6346	0.82	4.22E-14
nucleoside diphosphate kinase	8886.5302	0.75	1.91E-12
Sugar transporter family protein	5851.1426	-1.07	4.27E-12
conserved hypothetical protein	2439.2781	-0.71	4.67E-12
hypothetical protein	4930.3951	2.28	4.09E-10
expressed protein	12400.2759	1.49	1.56E-08
DEAD/DEAH box helicase family protein	6889.7012	0.46	4.53E-08

The performance of SILAC on intact proteins depends greatly on the near-complete incorporation of the desired residue in the cultured organism. **Figure 2.4** demonstrates the probability of complete incorporation of $^{13}C_6$-arginine using **Equation 2.1**, where the abundance of heavy isotope labeled protein, A_H, depends on the $^{13}C_6$-arginine labeling efficiency, E, raised to the exponent of the number of arginine residues to be labeled in the protein, N.

Equation 2.1: $A_H = E^N$

The purity of supplemented $^{13}C_6$-arginine (99.3%) is plotted along with the other efficiency values including the range observed in the top-down analysis of *A. flavus*.

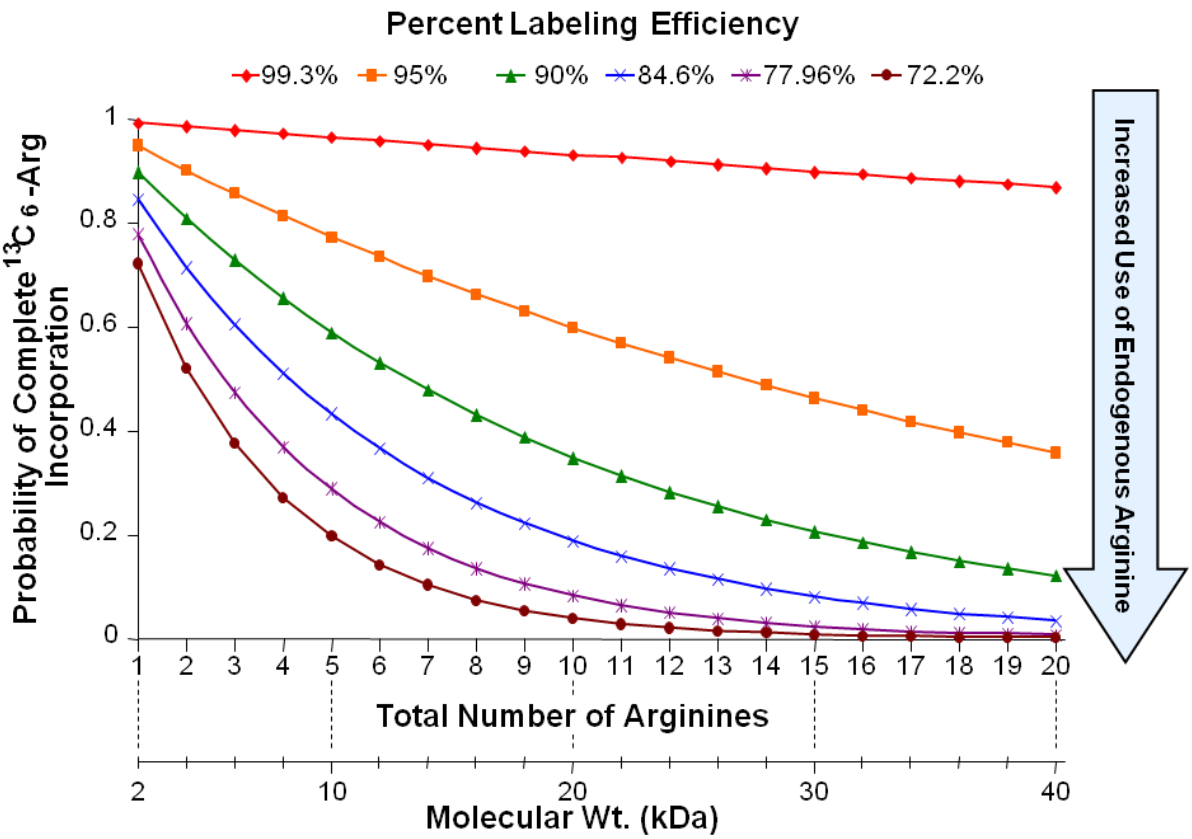

Figure 2.4 Probability of complete heavy isotope incorporation into a given protein based on number of arginines at given labeling efficiencies. Using frequency of arginine in Swiss-Prot Database (~5%), number of Arginine residues are connected to a theoretical molecular weight.

The model illustrates that the probability of complete SILAC labeling decreases with the number of arginines to be labeled in the protein (and proportionally as molecular weight increases). We have observed SILAC labeling efficiency of 72% to 85% in *A. flavus* suggesting incorporation of endogenously produced arginine during translation in cultures supplemented with $^{13}C_6$-arginine. The chance of having a completely labeled 20 kDa protein at 72% $^{13}C_6$-arginine labeling efficiency is 3.7%,

which greatly hampers our ability to quantify any significant amount of the *A. flavus* proteome.

Incorporation of a heavy isotope label is modeled and compared to experimental broadband mass spectra, identified in **Figure 2.3** using accurate intact mass, arginine counting, and MS/MS, by adjusting parameters in **Equation 2.2**, taking into account the total number of arginines in the protein, N, the labeling efficiency, E, and reporting relative abundances, A_n, for the masses of proteins containing n unlabeled arginines.

Equation 2.2: $$A_n = \frac{N!}{n!(N-n)!} E^n (1-E)^{N-n}$$

Altering the labeling efficiency parameter influences the relative abundances of isotopic envelopes corresponding to differential $^{13}C_6$-arginine incorporation. Representative spectra, shown in **Figure 2.5A and B**, which show in greater detail the precursor spectra from **Figure 2.3A and 2.3B**, identify closely with theoretical spectra calculated at 75% and 85% labeling efficiency, respectively, and is shown in the reflected axis below the experimental spectra.

Circles overlaying the experimental spectra show the sum of model isotopic abundances from the different isotopic envelopes for more direct comparison between the model and the experimental data. This range of labeling efficiency agrees with labeling efficiency data obtained from tryptic digest experiments of the same sample performed to analyze relative protein expression at different growth temperatures using a bottom-up approach.[43] The disparity between the purity of

heavy arginine supplemented for organism growth and the amount actually incorporated into the proteome is an issue that can hinder both top-down and bottom-up quantification experiments.

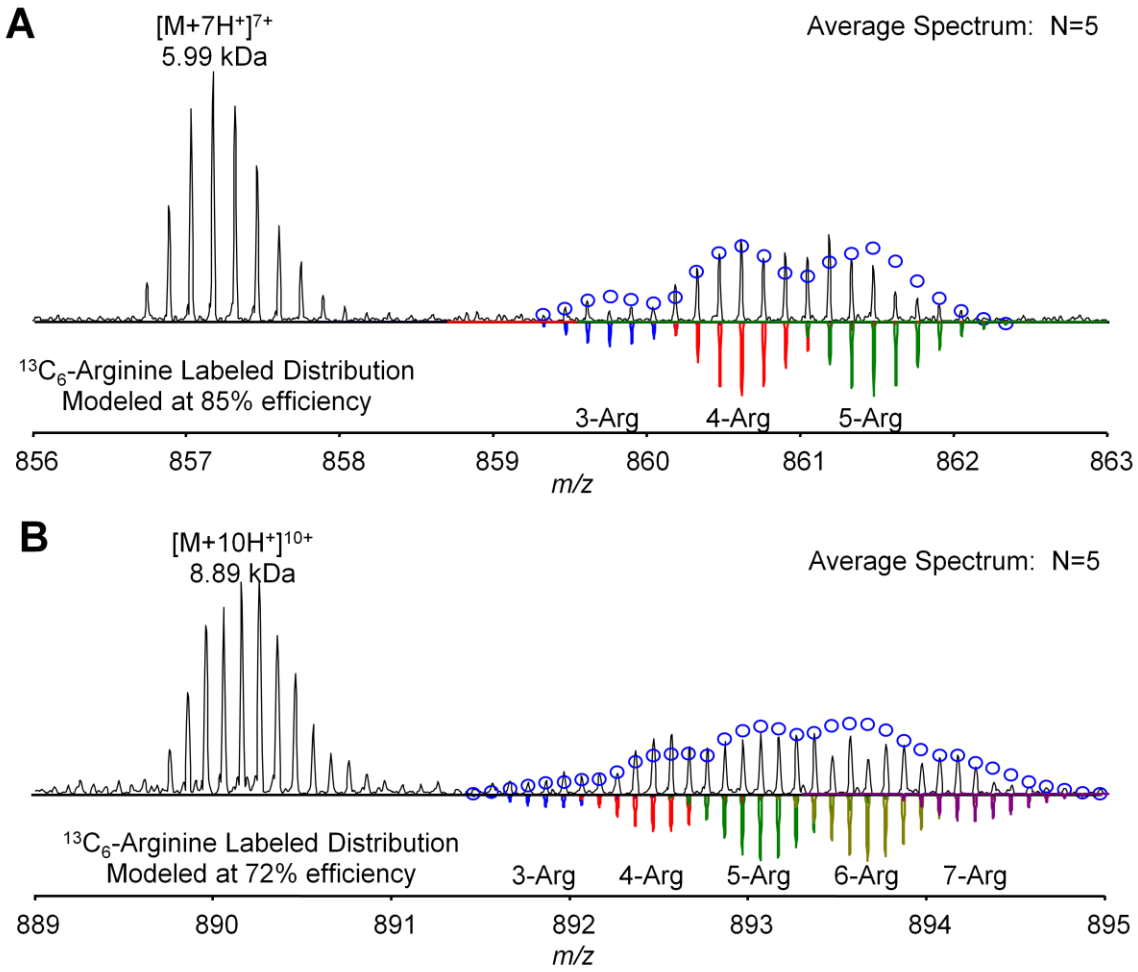

Figure 2.5 Full Scan FTMS spectra of SILAC pairs identified in **Figure 2.3** and modeled heavy labeled isotope distributions using **Equation 2.2** at labeling efficiencies best fitting the raw spectrum distribution. The sum of individual modeled isotopic peaks is overlaid upon the RAW spectra (circles).

However, approaches such as reducing the amount of endogenous arginine incorporated during the culturing process or selecting an amino acid label not produced by the cultured organism could overcome this obstacle.

2.4 Conclusions

Using the SILAC technique, cultures of the filamentous fungus *Aspergillus flavus* were grown under identical conditions incorporating either $^{12}C_6$ or $^{13}C_6$ labeled arginine. Intact proteins from a 1:1 mixture of cell lysates from both cultures were then separated using an mRP stationary phase which produced 15 to 20 second peak widths in a nano-flow reversed phase HPLC coupled to an LTQ-FT-ICR mass spectrometer. In all, 1,318 intact or fragment protein masses were detected corresponding to 659 SILAC pairs, of which 22 were confidently identified. Broadband precursor ion spectra provided accurate intact mass and allowed for arginine counting which aided in the identification of proteins combined with the sequence information from data dependent MS/MS spectra. Quantification of proteins with few arginines was consistent with the 1:1 mixture ratio **(Figure 2.3A)**. With greater numbers of arginine in a protein, increasingly complex heavy isotope spectra were observed, resulting in incorrect quantification. Endogenously produced arginine limited the labeling efficiency of the heavy culture to between 72% and 85% as determined using isotope distribution modeling. We are further developing methods to improve top-down quantification in SILAC experiments and expect to reduce endogenous amino acid incorporation through use of an arginine auxotroph in *A. flavus*.

2.5 References

1. Kenyon, G. L.; DeMarini, D. M.; Fuchs, E.; Galas, D. J.; Kirsch, J. F.; Leyh, T. S.; Moos, W. H.; Petsko, G. A.; Ringe, D.; Rubin, G. M.; Sheahan, L. C. Defining the mandate of proteomics in the post-genomics era: Workshop report. *Mol. Cell. Proteomics* **2002**, *1*, 763-780.

2. Kelleher, N. L.; Lin, H. Y.; Valaskovic, G. A.; Aaserud, D. J.; Fridriksson, E. K.; McLafferty, F. W. Top Down versus Bottom Up Protein Characterization by Tandem High-Resolution Mass Spectrometry. *J. Am. Chem. Soc.* **1999**, 806-812.

3. Hawkridge, A. M.; Heublein, D. M.; Bergen, R. H., III; Cataliotti, A.; Burnett. J. C, Jr.; Muddiman, D. C. Quantitative mass spectral evidence for the absence of circulating brain natriuretic peptide (BNP-32) in severe human heart failure. *P. Natl. Acad. Sci. USA* **2005**, *102*, 17442-17447.

4. Das, T.; Sen, A.; Kempf, T.; Pramanik, S. R.; Mandal, C.; Mandal, C. Induction of glycosylation in human C-reactive protein under different pathological conditions. *Biochem. J.* **2003**, *373*, 345-355.

5. Das, T.; Mandal, C.; Mandal, C. Variations in binding characteristics of glycosylated human C-reactive proteins in different pathological conditions. *Glycoconjugate J.* **2004**, *20*, 537-543.

6. Ardelt, P.; Dorka, P.; Jaquet, K.; Meilmeyer, L. M. G.; Kortke, H.; Korfer, R.; Notohamiprodjo, G. Microanalysis and distribution of cardiac troponin I phospho species in heart areas. *Biol. Chem.* **1998**, *379*, 341-347.

7. Katrukha, A. G.; Bereznikova, A. V.; Filatov, V. L.; Esakova, T. V.; Kolosova, O. V.; Pettersson, K.; Lovgren, T.; Bulargina, T. V.; Trifonov, I. R.; Gratsiansky, N. A.; Pulkki, K.; Voipo-Pulkki, L. M.; Gusev, N. B. Degredation of cardiac troponin I: implication for reliable immunodetection. *Clin. Chem.* **1998**, *44*, 2433-2440.

8. Labugger, R.; Organ, L.; Collier, C.; Atar, D.; Van Eyk, J. E. Extensive troponin I and T modification detected in serum from patients with acute myocardial infarction. *Circulation* **2000**, *102*, 1221-1226.

9. Bogdanov, B.; Smith, R. D. Proteomics by FTICR mass spectrometry: Top down and bottom up. *Mass Spectrom. Rev.* **2005**, *24*, 168-200.

10. Kelleher, N. L.; Taylor, S. V.; Grannis, D.; Kinsland, C.; Chiu, H. J.; Begley, T. P.; McLafferty, F. W. Efficient sequence analysis of six gene products (7-74 kDa) from the Escherichia coli thiamin biosynthetic operon by tandem high-resolution mass spectrometry. *Protein Sci* **1998**, *7*, 1796-1801.

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

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

13. Siuti, N.; Kelleher, N. L. Decoding protein modifications using top-down mass spectrometry. *Nat. Methods* **2007**, *4*, 817-821.

14. Wang, Y.; Balgley, B. M.; Rudnick, P. A.; Lee, C. S. Effects of chromatography conditions on intact protein separations for top-down proteomics. *J. Chromatogr. A* **2005**, *1073*, 35-41.

15. Dillon, T. M.; Bondarenko, P. V.; Ricci, M. S. Development of an analytical reversed-phase high-performance liquid chromatography electrospray ionization mass spectrometry method for characterization of recombinant antibodies. *J. Chromatogr. A* **2004**, *1053*, 299-305.

16. Syka, J. E. P.; Marto, J. A.; Bai, D. L.; Horning, S.; Senko, M. W.; Schwartz, J. C.; Ueberheide, B.; Garcia, B.; Busby, S.; Muratore, T.; Shabanowitz, J.; Hunt, D. F. Novel linear quadrupole ion trap/FT mass spectrometer: Performance characterization and use in the comparative analysis of histone H3 post translational modifications. *Journal of Proteome Research* **2004**, *3*, 621-626.

17. Zamdborg, L. L., R. D.; Glowacz, K. J.; Kim, Y.; Viswanathan, V.; Spaulding, I. T.; Early, B. P.; Bluhm, E. J.; Babai, S.; Kelleher, N. L. ProSight PTM 2.0: improved protein identification and characterization for top down mass spectrometry. *Nucleic Acids Res.* **2007**, *35*, W701-W706.

18. Williams, D. K., Jr.; Muddiman, D. C. Parts-per-billion mass measurement accuracy achieved through the combination of multiple linear regression and automatic gain control in a Fourier transform ion cyclotron resonance mass spectrometer. *Anal. Chem.* **2007**, *79*, 5058-5063.

19. Nepomuceno, A. I.; Mason, C. J.; Muddiman, D. C.; Bergen, H. R.; Zeldenrust, S. R. Detection of genetic variants of transthyretin by liquid chromatography-dual electrospray ionization Fourier-transform ion cyclotron resonance mass spectrometry. *Clin. Chem.* **2004**, *50*, 1535-1543.

20. Jennings, K. R. the changing impact of the collision-induced decomposition of ions on mass spectrometry. *Int. J. Mass Spectrom.* **2000**, *200*, 479-493.

21. Little, D. P.; Speir, J. P.; Senko, M. W.; O'Connor, P. B.; McLafferty, F. W. Infrared multiphoton dissociation of large multiply-charged ions for biomolecule sequencing. *Anal. Chem.* **1994**, *66*, 2809-2815.

22. Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. Electron capture dissociation of multiply charged protein cations. A nonergodic process. *J. Am. Chem. Soc.* **1998**, *120*, 3265-3266.

23. Kelleher, N. L. Top-down proteomics *Anal. Chem.* **2004**, *76*, 196A-203A.

24. Reid, G. E.; McLuckey, S. A. 'Top down' protein characterizaton via tandem mass spectrometry. *J. Mass Spectrom.* **2002**, *37*, 663-675.

25. Gordon, E. F.; Mansoori, B. A.; Carroll, C. F.; Muddiman, D. C. Hydropathic influences on the quantification of equine heart cytochrome c using relative ion abundance measurements by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *J. Mass Spectrom.* **1999**, *34*, 1055-1062.

26. Martinovic, S.; Veenstra, T. D.; Anderson, G. A.; Pasa-Tolic´, L.; Smith, R. D. Selective incorporation of isotopically labeled amino acids for identification of intact proteins on a proteome-wide level. *J. Mass Spectrom.* **2002**, *37*, 99-107.

27. Veenstra, T. D.; Martinovic, S.; Anderson, G. A.; Pasa-Tolic´, L.; Smith, R. D. Proteome analysis using selective incorporation of iotopically labeled amino acids. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 78-82.

28. Zhang, R. J.; Sioma, C. S.; Wang, S. H.; Regnier, F. E. Fractionation of isotopically labeled peptides in quantitative proteomics. *Anal. Chem.* **2001**, *73*, 5142-5149.

29. Julka, S.; Regnier, F. Quantification in proteomics through stable isotope coding: a review. *J. Proteome Res.* **2004**, *3*, 350-363.

30. Du, Y.; Parks, B. A.; Sohn, S.; Kwast, K. E.; Kelleher, N. L. Top-down approaches for measuring expression ratios of intact yeast proteins using Fourier transform mass spectrometry. *Anal. Chem.* **2006**, *78*, 686-694.

31. Parks, B. A.; Jiang, L.; Thomas, P. M.; Wenger, C. D.; Roth, M. J.; Boyne Ii, M. T.; Burke, P. V.; Kwast, K. E.; Kelleher, N. L. Top-down proteomics on a chromatographic time scale using linear ion trap Fourier transform hybrid mass spectrometers. *Anal. Chem.* **2007**, *79*, 7984-7991.

32. Ong, S.; 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**, 376-386.

33. Ong, S.; Kratchmarova, I.; Mann, M. Properties of 13C-substituted arginine in stable isotope labeling by amino acids in cell culture (SILAC). *J. Proteome Res.* **2003**, *2*, 173-181.

34. de Godoy, L. M. F.; Olsen, J. V.; de Souza, G. A.; Li, G.; Mortensen, P.; Mann, M. Status of complete proteome analysis by mass spectrometry: SILAC labeled yeast as a model system. *Genome Biol.* **2006**, *7*, R50-R50.15.

35. Waanders, L. F.; Hanke, S.; Mann, M. Top-down quantitation and characterization of SILAC labeled proteins. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 2058-2064.

36. Hussein, H. S.; Brasel, J. M. Toxicity, metabolism, and impact of mycotoxins on humans and animals.*Toxicology* **2001**, *167*, 101-134.

37. Medina, M. L. H., P. A.; Breci, L.; Fancisco, W. A. Analysis of secreted proteins from Aspergillus flavus. *Proteomics* **2005**, *5*, 3153-3161.

38. Mateles, R. I.; Adye, J. C. Production of aflatoxins in submerged culture. *Appl. Microbiol.* **1965**, *13*, 208-211.

39. Jiang, L.; He, L.; Fountoulakis, M. Comparison of protein precipitation methods for sample preparation prior to proteomic analysis. *J. Chromatogr. A* **2004**, *1023*, 317-320.

40. Horn, D. M.; Zubarev, R. A.; McLafferty, F. W. Automated reduction and interpretation of high resolution electrospray mass spectra of large molecules. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 320-332.

41. Foley, J. P.; Dorsey, J. G. Equations for calculation of chromatographic figures of merit for ideal and skewed peaks. *Anal. Chem.* **1983**, *55*, 730-737.

42. Payne, G. A.; Nierman, W. C.; Wortman, J. R.; Pritchard, B. L.; Brown, D.; Dean, R. A.; Bhatnagar, D.; Cleveland, T. E.; Machida, M.; Yu, J. Whole genome comparison of Aspergillus flavus and A-oryzae. *Med. Mycol.* **2006**, *44*, S9-S11.

43. Georgianna, D. R.; Hawkridge, A. M.; Muddiman, D. C.; Payne, G. A. Temperature-dependent regulation of proteins in Aspergillus flavus: whole organism stable isotope labeling by amino acids. *J. Proteome Res.* **2008**, *7, 2973-2979*

Chapter 3

Quantitative Top-Down Proteomics of SILAC Labeled Human Embryonic Stem Cells

3.1 Introduction

Human embryonic stem cells (hESCs) are pluripotent cells derived from the inner cell mass of blastocyst stage embryos that possess the ability to self-renew or differentiate into any human cell type. This makes hESCs a potentially valuable tool for regenerative medicine, where tissues damaged by disease or injury can be restored via hESC-derived cells. The ability of hESCs to be maintained in cell culture allows the opportunity for investigation into how these cells maintain pluripotency and how they progress through differentiation to various cell types.[1] Exploring the proteome of hESCs and how it changes qualitatively and quantitatively is an integral part of understanding the overall biology of this cell type, and will yield invaluable information toward the development of new medical treatments.

Top-down proteomics has emerged as a powerful technique for protein analysis and is a growing area of research in the mass spectrometry community. The typical experimental strategy for top-down proteomic analysis includes the front-end separation of intact proteins, their detection and fragmentation in a mass spectrometer, and subsequent identification based on sequence information obtained from their resulting MS and MS/MS spectra.[2, 3] Intact proteins present a wide range of physical and chemical properties that challenge front-end separation techniques,[4-6] mass analyzer duty cycle,[7] and protein identification algorithms.

These challenges hinder high throughput analyses, and thus, the top-down approach has lagged the more commonly employed bottom-up approach. However, the emergence and development of new separations technologies[8-11], high resolving power hybrid mass analyzers such as the LTQ-FT-ICR[12] and LTQ-Orbitrap[13, 14], and bioinformatics software such as ProSightPC[15] has greatly improved the capability for addressing the challenges posed by top-down proteomics. Recently, the top-down approach has been utilized by Kelleher and co-workers to examine multiple modifications of human histones, including methlyation and acetylation.[16] Top-Down proteomics has also been used to identify single amino acid polymorphisms and characterize post-translational proteolysis and phosphorylation of affinity purified rat cardiac troponin.[17] Integrated top-down and bottom-up strategies have allowed investigators to take advantage of the top-down approach's ability to identify and characterize multiple post-translational modifications (PTMs) and identify genetic modifications while utilizing bottom-up proteomics to achieve higher throughput.[18]

Quantitative top-down proteomics most often utilizes stable isotope labeling to create an internal standard from which reliable quantitative data may be obtained. Gordon et al. demonstrated with bovine and equine cytochrome c that even small differences-in this case, three amino acids-between a protein and a structural homologue from a different species is enough to produce different charge state distributions that hinder quantification unless the signal from all charge states is taken into account.[19] Deuterated leucine has been incorporated into model cell cultures, including *E. coli* and *S. cerevisiae*,[20, 21] which facilitated identification of

proteins using accurate intact mass and amino acid counting; however, quantification with deuterated labels using reversed-phase chromatography is problematic because elution times decrease for the deuterated analytes.[22, 23] $^{15}N/^{14}N$ metabolic labeling has been successfully utilized the yeast proteome, and moreover was employed on a chromatographic time scale to identify 22 intact proteins from *S. cerevisiae*.[24] Stable isotope labeling by amino acids in cell culture (SILAC) was introduced by Mann and co-workers as another means of quantifying peptides and proteins, whereby stable isotope labeled amino acids are supplemented to cull culture growth media to produce co-eluting labeled and unlabeled protein or peptide analytes. This technique is widely implemented in bottom-up experiments using labeled arginine and lysine residues in conjunction with trypsin proteolysis.[25, 26]

Quantification of intact proteins using SILAC has been computationally modeled and demonstrated in affinity purified proteins expressed with an *E. coli* vector and 99% stable isotope incorporation of a single label was achieved.[27] We have previously described the implementation of SILAC on a chromatographic time scale for protein identification and quantification using $^{13}C_6$-arginine supplemented to the growth media of the filamentous fungus *Aspergillus flavus*, and also demonstrated that incomplete label incorporation in a prototrophic organism results in broad protein isotopic distributions. A mathematical relationship was also described which characterizes the relationship of label incorporation with the observed experimental protein signal.[28]

Mass spectrometry based proteomics is becoming increasingly employed for the analysis of various stem cell populations at increasing depth and detail.[29] Recently, mass spectrometry and electron capture dissociation were used to characterize the phosphoproteome of hESCs[30, 31] and label free quantification has been carried out to characterize 74 different PTM combinations of Histone H4 in hESCs.[32] The iTRAQ method has been used to track the change in mouse embryonic stem cell protein expression during differentiation.[33] Bottom-up quantification using SILAC has been utilized to identify and quantify over 5,000 mouse embryonic stem cell proteins[34] and has been recently demonstrated with hESCs to seek out differentially expressed cell surface markers between pluripotent and differentiating cell populations.[35]

This study represents our initial efforts to establish top-down analysis of SILAC labeled human embryonic stem cells, in which we utilize top-down proteomics to analyze human embryonic stem cells that have been dual labeled with $^{13}C_6$-arginine (Arg6) and $^{13}C_6$, $^{15}N_2$-Lysine (Lys8) with approximately greater than 98% label incorporation as verified by bottom-up proteomics, producing 11 confident protein identifications based on accurate intact mass, MS/MS, and amino acid counting. Furthermore, we characterize the effect of multiple stable isotope labels, minute incorporation of unlabeled amino acids, and arginine-proline conversion on the labeled protein's isotopic distribution.

3.2 Experimental

3.2.1 Culture and SILAC labeling of hESCs

H9 cells were maintained on mouse embryonic fibroblast (MEF) conditioned medium (CM) according to a protocol adapted from Xu *et al* [36]. Briefly, MEF cells were inactivated by Mitomycin C (Sigma Aldrich) treatment and plated at 55,000/cm^2. HESC medium consisting of DMEM/F12 (Invitrogen), 20% Knockout Serum Replacement (Invitrogen), 1mM L-Glutamine (Invitrogen), 0.1mM β-mercaptoethanol, 1% non-essential amino acids (Invitrogen) and 4ng/ml basic Fibroblast Growth Factor (bFGF, Sigma Aldrich) was conditioned for 24h on MEF cells. CM was collected for 7 days, pooled and supplemented with 4 ng/mL bFGF (Sigma Aldrich).

CM for SILAC labeling of hESCs was prepared using protocols previously described.[37] Briefly, SILAC medium consisting of DMEM/F12 without L-Lysine and L-Arginine (Pierce), $^{13}C_6$ L-Arginine (Pierce), $^{13}C_6$ $^{15}N_2$ L-Lysine (Pierce), 20% Knockout Serum Replacement (Invitrogen), 1 mM L-Glutamine (Invitrogen), 0.1 mM β-mercaptoethanol, 1% non-essential amino acids (Invitrogen) and 4 ng/ml basic Fibroblast Growth Factor (bFGF, Sigma Aldrich) was conditioned for 24 h on MEF cells. CM was collected for 7 days, pooled and supplemented with 4 ng/mL bFGF (Sigma Aldrich). It is important to note that the Knockout Serum Replacement does not contain lysine or arginine [38]

3.2.2 Protein Isolation and Sample Preparation

Confluent H9 cells were lysed with 8 M Urea (Sigma Aldrich) and 50 mM Ammonium Bicarbonate (Fisher Scientific) and frozen at -80 °C. Prior to analysis by LC-MS, cell lysates were concentrated and desalted on 5 µm, C_8 stationary phase beads (Michrom Bioresources, Auburn, CA) according to a protocol adapted from Winston and Fitzgerald.[39]

3.2.3 Bottom-Up Sample Preparation

50 µg of Arg6, Lys8-labeled protein from human embryonic stem cells was combined with 2X Laemmli buffer containing β-mercaptoethanol, boiled 5 min and cooled on ice. The sample was loaded and ran on a Criterion 10-20% Tris-HCl gel (BioRad, Hercules, CA) at 200V and visualized with BioSafe coomassie stain. The gel lane was divided into 8 equal fractions, each of which were reduced with DTT, alkylated with iodoacetamide, and digested with trypsin according to a protocol adapted from Shevchenko et al. [40]

3.2.4 Top-Down LC-MS/MS

All LC solvents were purchased from Burdick and Jackson (Muskegon, MI). Reversed-phase nano-flow liquid chromatography was performed on a nanoLC-2D system (Eksigent, Dublin, CA) utilizing a vented column configuration.[41] 10 µL samples containing approximately 1 µg of total protein were injected using the autosampler, desalted on a 75 µm × 50 mm IntegraFrit (New Objective, Woburn, MA) trap packed in-house with mRP-C18 5 µm stationary phase, and eluted onto a PicoFrit capillary column with a 15 µm emitter tip (New Objective, Woburn, MA) also

packed in-house with mRP-C18 stationary phase (Agilent, Palo Alto, CA). The final packed volume of the column had dimensions 75 µm i.d. × 200 mm and was enclosed in a modified Hot-Pocket column heater (ThermoFisher) operating at 80 °C. Mobile phases A and B consisted of 98% water, 2% acetonitrile, and 2% water, 98% acetonitrile respectively, with both containing 0.2% (v/v) formic acid as an ion-pairing reagent. Samples were injected and desalted on the trap at 1.5 µL/min with 2% mobile phase B for a metered injection of 30 µL. After desalting, a 10-port valve (VICI, Houston, TX) switched the trap in line with a nano-flow gradient at 300 nL/min, at which time data collection also commenced. The gradient was held at 2% mobile phase B for 10 minutes, ramped to 25% over 5 minutes, then to 65% B over the following 70 minutes. The gradient then ramped to 95% B in 5 minutes and held for 10 minutes before re-equilibrating at 2% B for 10 minutes.

Mass spectrometric analyses were performed on a 7T LTQ-FT Ultra (ThermoFisher Scientific, San Jose, CA) with a pulse sequence consisting of four events, providing for a broadband acquisition in profile mode followed by three data-dependent MS/MS experiments. All mass measurements were performed inside the ICR cell. All events used 1 microscan to determine ionization time to reach the target AGC limit of 1×10^6. The resolving power of all four events was set at $100{,}000_{FWHM}$ at *m/z*=400. MS/MS settings used an isolation width of 7 *m/z* and a normalized collision energy of 18% for 30 ms. CID was performed on the three most abundant *m/z* values from the precursor ion scan in the LTQ followed by product ion detection in the ICR cell. Precursor ions with 1+, 2+, and 3+ charges were excluded

from MS/MS. Precursors isolated for MS/MS were dynamically excluded for 45 seconds after 3 repeat analyses within a 30 second window to reduce redundant analysis of abundant precursor ions that may dominate the mass spectra for long periods of time.

3.2.5 Top-Down Data Analysis

RAW LC-MS data files were imported into ProSightPC 2.0 (ThermoFisher, San Jose, CA) using a THRASH[5] algorithm to determine the monoisotopic peak of analyte signals with a S/N ratio of 3:1 or greater. Intact precursor and fragment ion masses were searched against the IPI human database (version 3.63) using a search logic tree that submitted all data to an absolute mass search, followed by a biomarker search for all data that did not yield an ID with E value less than 1E-4 in the absolute mass search. Intact protein precursor and fragment ion data were searched initially with ion mass tolerances set at ±10 ppm. An additional search was performed with fragment tolerances of ±1 Da to accommodate sample handling modifications to the mass (e.g., deamidation) or error in monoisotopic peak picking by the THRASH algorithm. Identified protein sequences reported from the searches were confirmed using arginine and lysine counting from the precursor ion spectra. Analysis of SILAC labeled protein distributions and theoretical distributions with varying arginine and lysine labeling efficiencies in addition to different amounts of proline conversion was carried out using ICR-2LS (PNNL, Richland, WA) and Microsoft Excel® (Redmond, WA).

3.2.6 Bottom-Up LC-MS/MS

In-gel digested peptide samples were separated using the same LC system, column configuration and mobile phases described *vide supra*. The stationary phase employed was Magic C18AQ (5 µm particle, 200 Å pore size)(Microm BioResources). 10 µL of sample was injected and desalted at 1.5 µL/min on the trap before switching in line with the gradient at a flow rate of 300 nL/min. The gradient held at 2% B for 5 minutes before ramping to 10% B over 2 minutes, followed by an increase to 50% B over the next 60 minutes. The gradient then ramped to 95% B in 3 minutes, held for 5 minutes, then re-equilibrated at 2% B for 10 minutes.

Peptides were detected on the LTQ-FT Ultra with a pulse sequence containing 4 scan events. A broadband scan was performed in the ICR cell followed by three data-dependent MS/MS events in the ion trap at an isolation width of 2 *m/z* and normalized collision energy of 30% for 30 ms. 1+ charge states were rejected from MS/MS and dynamic exclusion was employed for 45 seconds after 2 repeat analyses in a 30 second window.

3.2.7 Bottom-Up Data Analysis

RAW LC-MS files were processed using MASCOT Distiller (Matrix Science, Boston, MA) to create peak lists in .mgf format to be searched against the IPI Human database using MASCOT(Matrix Science, London, UK).[42] Searches employed a peptide mass tolerance of ±5 ppm and a fragment ion tolerance of ±1 Da. Fixed modification included carboxyamidomethylation of cysteine, with variable

modifications including methionine oxidation and deamidation of asparagines and glutamine residues.

3.3 Results and Discussion

H9 cells cultured in CM and SILAC CM for at least 5 passages were lysed, combined at a 1:1 ratio as determined by Bradford assay, and subjected to an initial sample cleaning procedure with C8 reversed-phase column packing material. The resulting SILAC mixture was analyzed by nano-flow LC-MS/MS as described *vide supra*. **Figure 3.1** shows examples of two proteins identified by top-down LC-MS/MS. **Figure 3.1A** demonstrates the detection of an 8.5 kDa SILAC pair with charge states ranging from 6+ to 12+. The light form of the 7+ charge state was isolated for collision induced dissociation, allowing the protein to be identified as ubiquitin c splice variant using the biomarker search method in ProSightPC. The sequence of the protein, identified with an E-value of 2.94E-19, was then used to calculate a mass difference between the light and heavy protein forms of 80.1798 Da based on the presence of 4 arginine residues and 7 lysine residues. This mass difference translates to an *m/z* difference of 11.4542 at $z = 7$, which strongly correlates with the mass difference between the SILAC pair in the experimental precursor spectrum. This increases the confidence of the protein identification.

Likewise, the principle of making an initial identification from MS and MS/MS data and confirming with mass difference of the SILAC pair is also illustrated in **Figure 3.1B**.

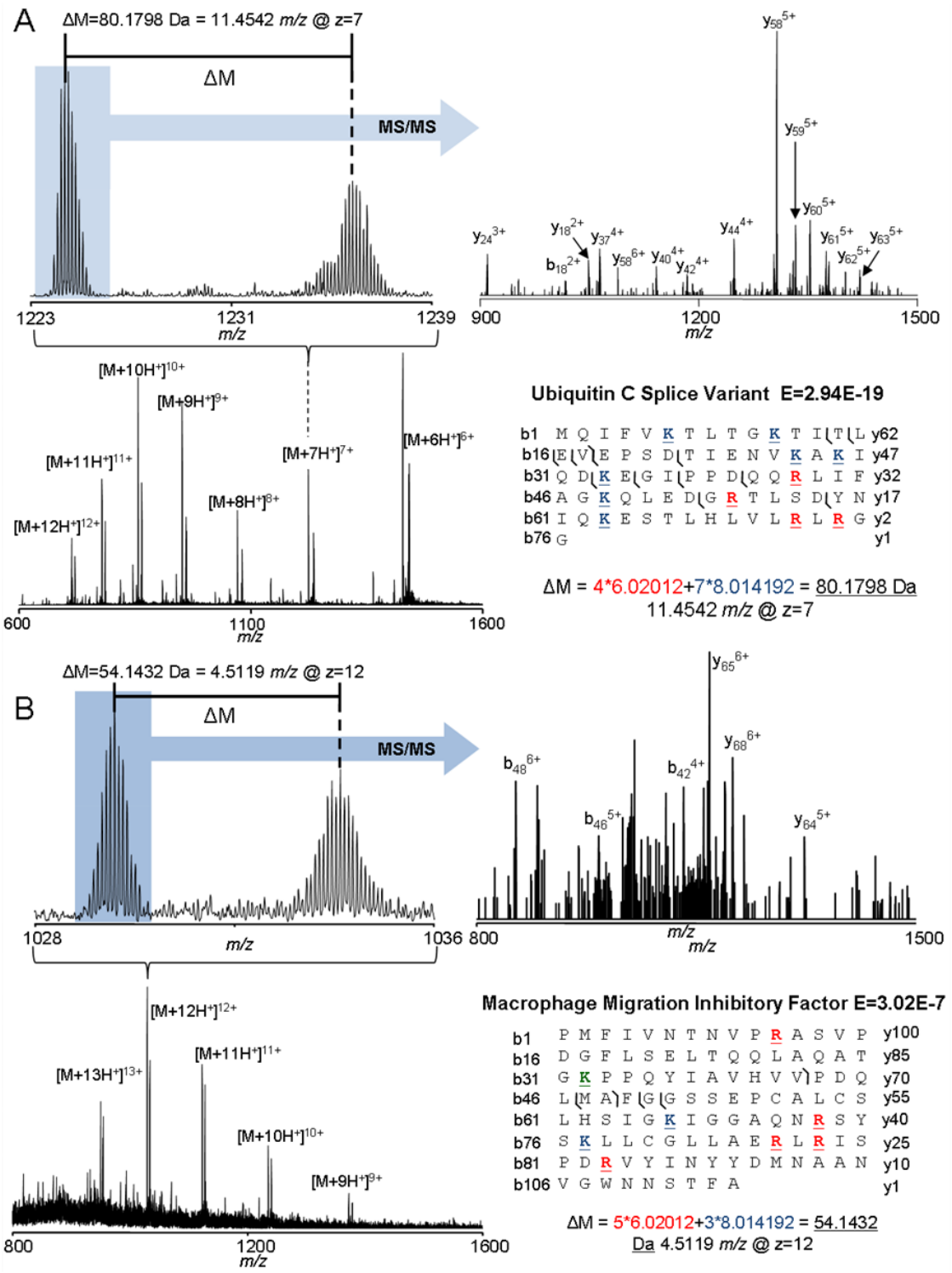

Figure 3.1 Precursor and product ion mass spectra of intact protein SILAC pairs are shown in addition to their identified sequences and E-values as scored by ProSightPC for **(A)** Ubiquitin C Splice Variant and **(B)** Macrophage Migration Inhibitory Factor.

Here, SILAC pairs of a 12.3 kDa protein with charge states ranging from 9+ to 13+ are detected, with the light form of the 12+ charge state being isolated for MS/MS analysis. Searching in absolute mass mode with ProSightPC identified the protein as macrophage migration inhibitory factor with an E-value of 3.02E-7. Although the noisy broadband and MS/MS spectra in addition to only a few identified product ions would normally cast doubt on the confidence of the identification, the calculated mass difference for 5 arginines and 3 lysines of 54.1432 Da between the light and heavy forms given by the predicted sequence again matches the *m/z* difference of 4.5119 at *z*=12 of the experimental precursor ion spectrum, giving additional confidence to the identified sequence. We were similarly able to identify 11 intact proteins from hESC lysate (**Table 3.1**). Seven of the eleven identified proteins were full intact proteins identified by the absolute mass search in ProSightPC, and four were identified by biomarker search, meaning they are truncated forms of intact protein sequences. Four of the proteins identified by top-down analysis are variants of thymosin or prothymosin, of which only one form was identified in a parallel bottom-up study (data not shown). Five of the 11 proteins identified were acetylated at the N-terminus, including Histone H4.

Although heavy and light pairs of SILAC-labeled intact proteins were detected, the ratio of abundance of light to heavy protein initially appeared to deviate from the expected 1:1 ratio. A closer inspection of the isotopic distributions of the heavy labeled proteins, illustrated in the upper left panels of both **Figures 3.1A and 3.1B**, reveals more complexity to the distribution than is present in the light protein

signal, with additional isotopic peaks appearing at both lesser and greater *m/z* values than expected. This complexity is a compounded effect of the incorporation efficiency of the two different stable-isotope labeled amino acids, in addition to contribution from the metabolic conversion of heavy arginine to heavy proline.[37, 43] However, the SILAC ratio calculated as the sum of light isotope abundances divided by the heavy isotope abundances still results in a value of unity. Specifically, the SILAC ratios calculated for ubiquitin c splice variant and macrophage migration inhibitory factor were 1.05 and 0.90, respectively, both falling within the range of variability of SILAC ratios observed in bottom-up SILAC experiments with the same samples (data not shown).

In addition to the ability to confirm protein identifications with the ΔM of the light and heavy forms of the protein observed in the precursor ion spectra, the incorporation of amino acid labels in MS/MS fragment ions of the same protein can be confirmed in cases where both the heavy and light forms of the protein are selected for CID. **Figure 3.2** shows an overlay of the MS/MS spectra of the ubiquitin c splice variant shown also in **Figure 3.1**. Varying differences in *m/z* between the same y and b ion peaks arise from each fragment containing a different number of labeled amino acids. By comparing the difference in mass between those peaks with the mass difference predicted by the assigned sequence of the protein, we can further increase our confidence as to the identity of the protein analyte. For instance, the y_{18}^{2+} ion are expected to contain one lysine and two arginines that should be heavier in the labeled form of the protein by 20.054 Da.

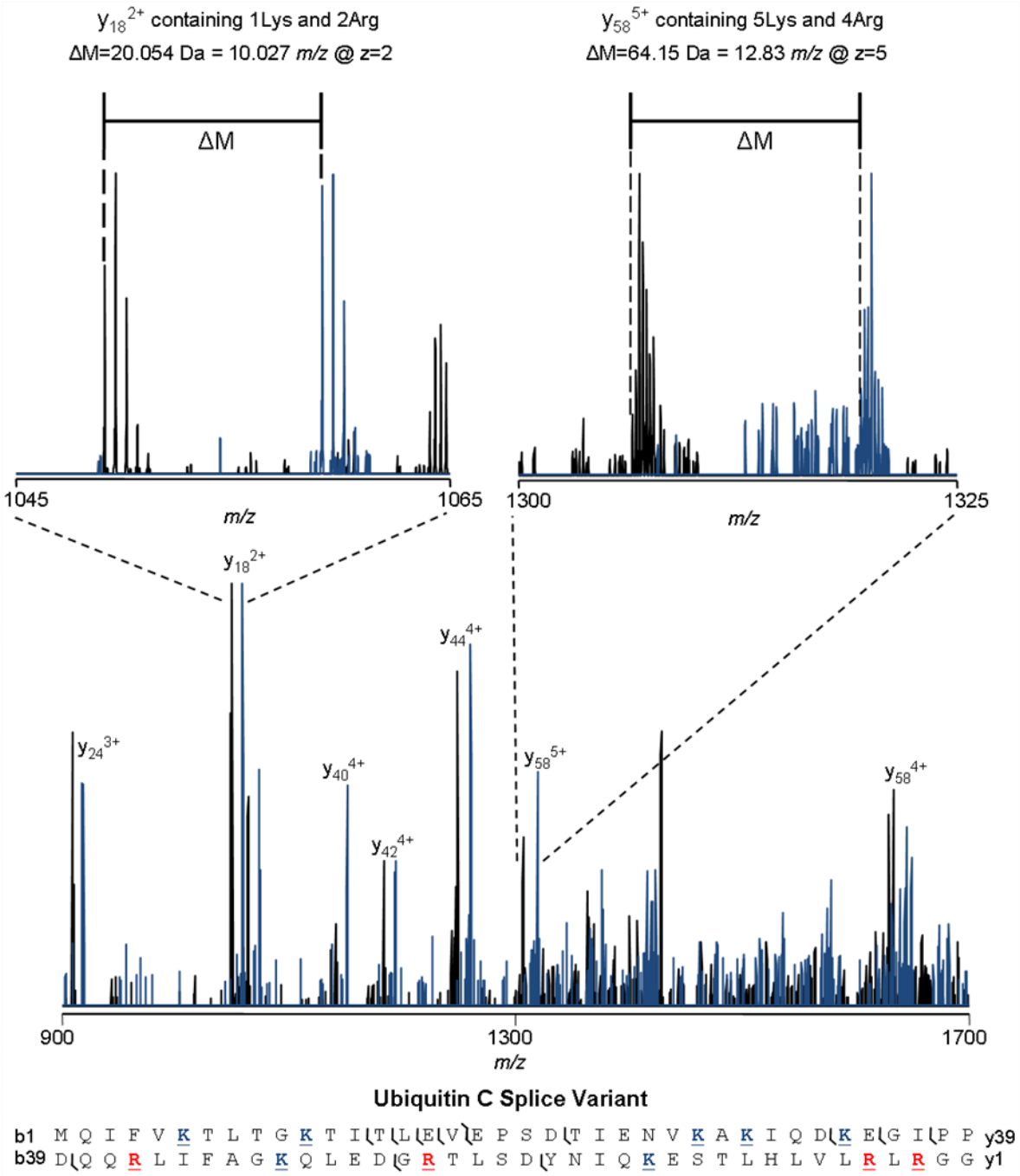

Figure 3.2 The MS/MS spectra of light (black) and heavy (blue) forms of Ubiquitin C Splice Variant are overlaid to demonstrate the confirmation of fragment amino acid sequences based on their differences in mass as correlated to the different numbers of labeling amino acids incorporated.

The *m/z* difference of 10.027 in the experimental spectra (z=2) shown in **Figure 3.1** corresponds closely to the predicted mass difference. Likewise, the light and heavy

forms of the y_{58}^{5+} ion have a predicted mass difference 64.15 Da predicted by the assigned sequence, containing 5 lysines and 4 arginines. The 12.83 *m/z* difference at *z=5* from the overlaid experimental spectra also match with the predicted mass difference, and further confirm the identification of the protein sequence.

Because the stable isotope labeled amino acids were not 100% purely labeled, there exists an increasing statistical probability that one or more light amino acids will be incorporated as the number of amino acids to be labeled in the protein increases. The result of such a probability is an incomplete labeled intact protein, being slightly lower in mass than the completely labeled counterpart. The abundance of these unlabeled forms is modeled in **Equation 3.1** which takes into account the number of amino acids to be labeled in the protein, N, and the incorporation efficiency of the amino acid label, E, and determines the relative abundance of the protein signal arising from n unlabeled residues, A_n.[28]

Equation 3.1: $$A_n = \frac{N!}{n!(N-n)!} E^n (1-E)^{N-n}$$

In order to account for the use of both stable isotope labeled arginine and lysine, in addition to the conversion of heavy arginine to heavy proline, the abundance of the completely labeled isotopic distribution is described by

Equation 3.2:

$$A_H = A_{100\%} - A_{100\%}\left[\left[1-\left[\frac{N_R!}{n_R!(N_R-n_R)!}E_R^{n_R}(1-E_R)^{N_R-n_R}\right]\right] + \left[1-\left[\frac{N_K!}{n_K!(N_K-n_K)!}E_K^{n_K}(1-E_K)^{N_K-n_K}\right]\right] + N_P\frac{\%P}{100}\right]$$

Where A_H the abundance of the completely labeled protein, $A_{100\%}$ is the abundance if labels were incorporated 100% of the time, N_R is the number of arginine residues in the protein, n_R is the number of unlabeled arginine residues in the protein, and E_R is the incorporation efficiency of arginine. N_K, n_K, and E_K represent the same parameters for stable isotope labeled lysine, and N_P and %P is the number of prolines and percent of conversion from arginine to proline, respectively.

In order to test the model described in **Equation 3.2** with the isotopic distributions observed in the top-down analysis of SILAC labeled hESCs, the amino acid incorporation efficiency of both Arg6 and Lys8 were determined from a parallel bottom-up proteomic analysis of just the heavy-labeled cell lysate. **Figure 3.3** shows results for peptides corresponding to proteins identified in **Figure 3.1**, including the ubiquitin c splice variant (**Figures 3.1A, 3.3A, 3.3B**) and macrophage migration inhibitory factor (**Figures 3.1B, 3.3C, 3.3D**). **Figure 3.3A** shows the mass spectrum of the peptide identified by MS/MS as ESTLHLVLR from the ubiquitin c splice variant, including the dominant signal from the form of the peptide containing Arg6, and the much weaker signal from the light form, amplified 50 times in the inset.

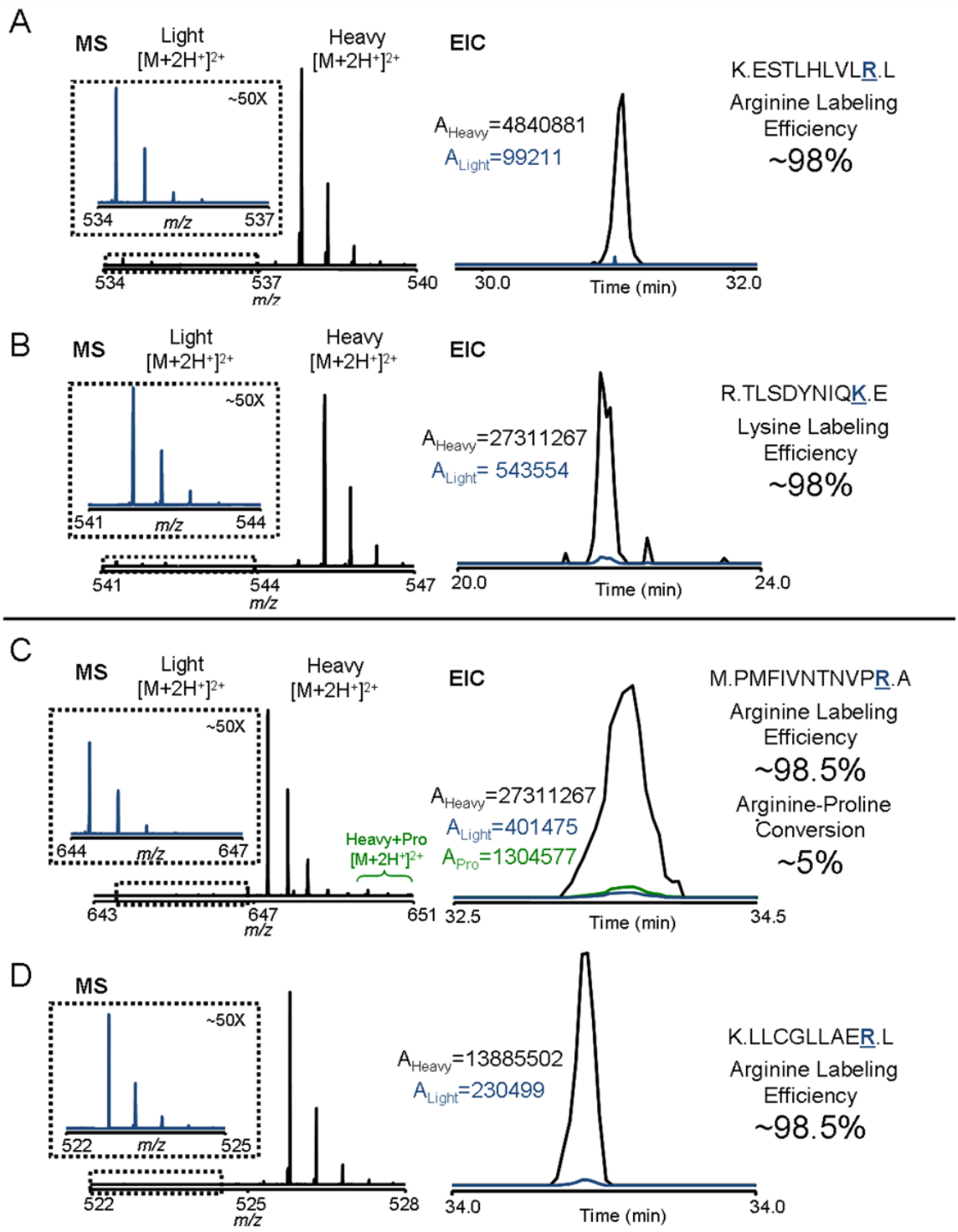

Figure 3.3 MS spectra and extracted ion chromatograms of light and heavy of arginine and lysine terminated peptides are used to calculate the percent incorporation of each amino acid label in to **(A and B)** ubiquitin c splice variant and **(C and D)** macrophage migration inhibitory factor. Proline conversion is also demonstrated **(C)**.

Extracting the ion chromatograms of the monoisotopic masses from both the light and heavy forms and finding the ratio of their peak areas suggests arginine incorporation to be approximately 98%. Similarly, **Figure 3.3B** shows the calculated incorporation of Lys8, also approximately 98%, from the mass spectrum and extracted ion chromatogram of the peptide TLSDYNIQK, which corresponds to the ubiquitin c splice variant as well.

Peptides that correspond to macrophage migration inhibitory factor are illustrated in **Figures 3.3C and 3.3D**. Both were arginine terminated, and yielded Arg6 incorporation values of approximately 98.5 %. The contribution arginine-proline conversion is visible in the mass spectrum of the peptide identified as PMFIVNTNVPR (**Figure 3.3C**) as indicated by the increased abundance of the fifth and sixth isotopic peaks in the labeled peptide signal. The area of the extracted ion chromatogram of the fifth isotopic peak, which is actually the monoisotopic peak of the heavy proline-containing form of the peptide, compared to that of the heavy label, reveals roughly 5% conversion of arginine to proline.

Using the percent incorporation values of Arg6 and Lys8 in addition to the percent conversion of Arg6 to Pro5 obtained from bottom-up proteomic analysis, **Equations 3.1 and 3.2** can be used in combination to plot the predicted abundance of a completely SILAC labeled intact protein signal (using **Equation 3.2**) in addition to signals that arise from having one or more unlabeled lysine or arginine residue and furthermore the signal arising from the incorporation of Pro5 (using **Equation 3.1**). **Figure 3.4** shows the experimental mass spectra of two intact proteins

overlaid upon the inverted theoretical spectra as predicted by **Equations 3.1 and 3.2**.

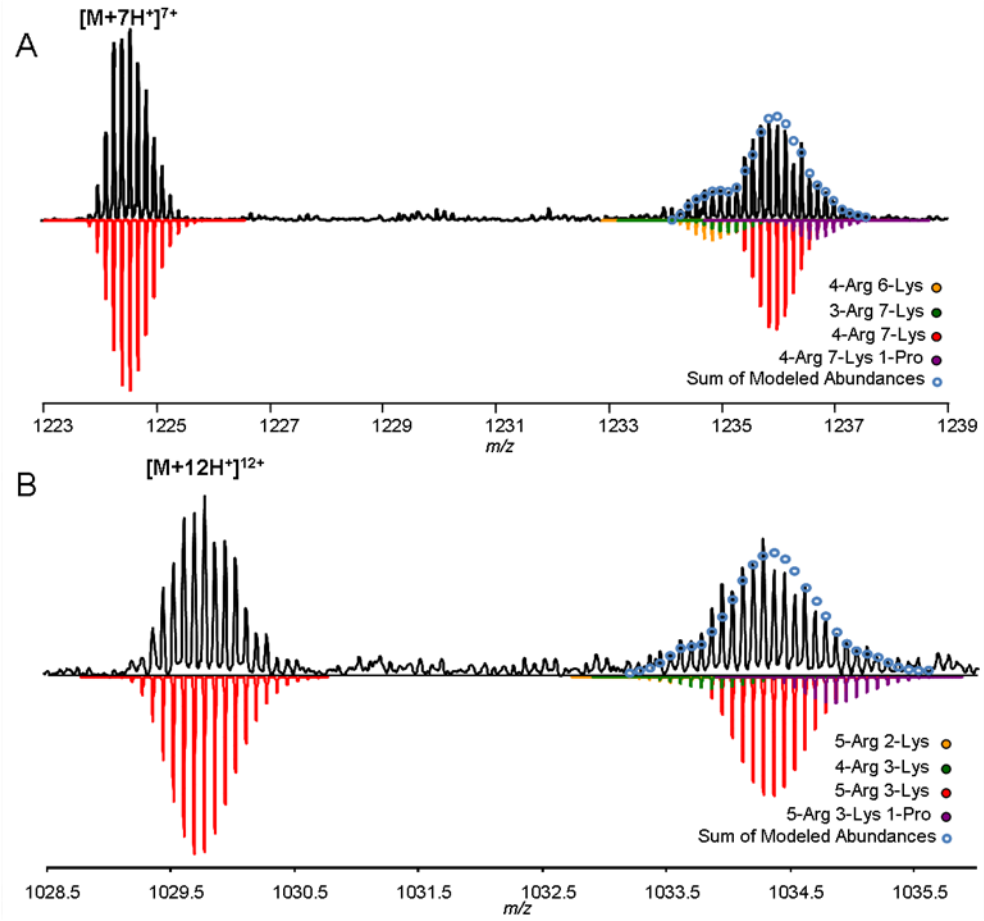

Figure 3.4 Experimental spectra are shown along with model spectra as predicted by Equations 1 and 2. The sum of model isotopic abundances of the inverted model spectra (circles) are superimposed on the experimental spectra of **(A)** Ubiquitin C Splice Variant and **(B)** Macrophage Migration Inhibitory Factor.

The experimental spectrum of the 7+ charge state of the ubiquitin c splice variant is illustrated in **Figure 3.4A**, where the predictive model accounts for 98% incorporation of Lys8 and Arg6 as determined by the peptide level data discussed above and shown in **Figure 3.3A and 3.3B**. 5% conversion of arginine to proline is also taken into account in the predicted spectra. All of the theoretical predicted

spectra are displayed on an inverted axis with the sum of the abundances of each isotopic peak overlaid upon the experimental spectrum, showing a strong correlation between the two. The theoretical isotopic distribution also strongly correlates with the experimental spectrum for the macrophage migration inhibitory factor SILAC pair shown in **Figure 3.3B**.

Comparing the heavy isotopic distribution of **Figure 3.4A** to that of **Figure 3.4B**, the statistical relationships of the model become more apparent. For instance, the maximum abundance of the heavy form of the ubiquitin c splice variant is just over half that of the light form, whereas the heavy form of the macrophage migration inhibitory factor is approximately 60 % that of the light form. The abundance of the SILAC labeled protein signal to its unlabeled counterpart is related to the number of labeling amino acids present for which imperfect incorporation subtracts from the abundance of the completely labeled form. This is evident in that the ubiquitin c splice variant has 11 total arginine and lysine residues for labeling versus 8 residues for macrophage migration inhibitory factor. The greater number of labeling amino acids in ubiquitin increases the probability of one being unlabeled, leading to a more pronounced “shoulder” of incompletely labeled protein in **Figure 3.4A** compared with **Figure 3.4B.** Additionally, 2 proline residues are present for potential incorporation of heavy proline in ubiquitin c splice variant and 8 proline residues are present in macrophage migration inhibitory factor, which in turn leads to a slightly more pronounced “tail” from the heavy proline protein signal in the latter relative to the former.

The complex distribution of stable isotope labeled intact proteins warrants careful interpretation for accurate quantification. This concept is illustrated in **Figure 3.5** where the base 2 logarithm of the light to heavy ratio of proteins in a SILAC pair are plotted in relation to the number of labeled amino acids in a given protein.

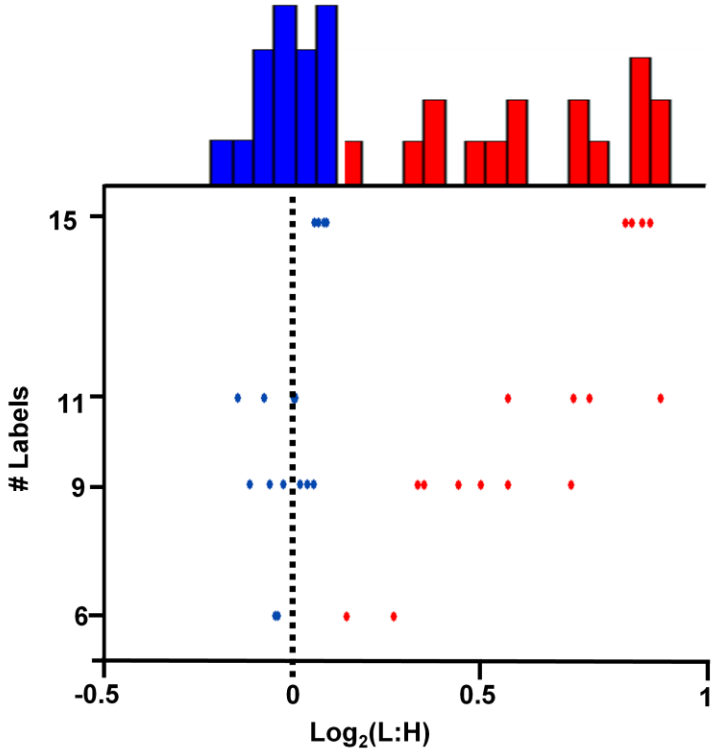

Figure 3.5 Uncorrected (RED) and corrected (BLUE) distribution of intact protein SILAC ratios using the model in **Equation 3.2** as related to the number of labels present in each labeled protein.

The red data points represent the ratios obtained when only taking into account the *m/z* space occupied the isotopic distribution of a completely labeled heavy protein. As the number of amino acid labels in the protein increase, the Log_2(L:H) deviates from zero with increasing magnitude. This deviation is a result of the increasing statistical probabilities of incorporating an unlabeled amino acid as the number of

labels increase. To correct for such a deviation one must account for the resulting expanded isotopic envelope using the model described in **Equations 3.1 and 3.2**. The successful implementation of this model on quantification is demonstrated by the blue data points in **Figure 3.5**, in which Log_2(L:H) values are tightly clustered about zero, representing a 1:1 protein abundance ratio.

3.4 Conclusions

Quantifying intact proteins using the SILAC method remains a challenging task, even with the relatively high incorporation of stable isotope labeled lysine and arginine < 98%. While the use of dual labeling strategies is beneficial in the case of bottom-up quantification by SILAC, even the small amount of light amino acid incorporated in the place of the stable isotope label is compounded. This effect is magnified further when a labeling amino acid is converted by the biological system into a different residue, as is the case with arginine-proline conversion. Many of these effects can be taken into account easily with quantification software with bottom-up experiments, but remains unaddressed from the top-down perspective. Still, multiple intact proteins were confidently identified, including multiple intact forms of proteins for which only a single form was identified from bottom-up analysis. It would be beneficial in future studies involving intact proteins for the use of a single, highly purified, essential amino acid label to be utilized for SILAC labeling. Stable isotope labeled lysine would be a strong candidate for future use, as it occurs frequently enough in the proteome to be present in most proteins, is a limiting amino

acid, and is not converted biologically into other amino acids. The result of this labeling system would be less complex heavy protein isotope distributions for easier intact protein quantification, giving researchers the ability to thoroughly examine the intact proteomes of human embryonic stem cells in addition other biological systems.

3.5 References

1. Vallier, L.; Pedersen, R. A., Human Embryonic Stem Cells: An In Vitro Model to Study Mechanisms Controlling Pluripotency in Early Mammalian Development. *Stem Cell Rev.* **2005,** 1, 119-130.

2. Kelleher, N. L., Top-down proteomics. *Anal. Chem.* **2004,** 76, (11), 196A-203A.

3. Bogdanov, B.; Smith, R. D., Proteomics by FTICR mass spectrometry: Top down and bottom up. *Mass Spectrom. Rev.* **2005,** 24, (2), 168-200.

4. Kelleher, N. L.; Taylor, S. V.; Grannis, D.; Kinsland, C.; Chiu, H. J.; Begley, T. P.; McLafferty, F. W., Efficient sequence analysis of the six gene products (7-74 kDa) from the Escherichia coli thiamin biosynthetic operon by tandem high-resolution mass spectrometry. *Protein Sci.* **1998,** 7, (8), 1796-1801.

5. Ge, Y.; Lawhorn, B. G.; ElNaggar, M.; Strauss, E.; Park, J. H.; Begley, T. P.; McLafferty, F. W., Top down characterization of larger proteins (45 kDa) by electron capture dissociation mass spectrometry. *J. Am. Chem. Soc.* **2002,** 124, (4), 672-678.

6. Meng, F. Y.; Du, Y.; Miller, L. M.; Patrie, S. M.; Robinson, D. E.; Kelleher, N. L., Molecular-level description of proteins from Saccharomyces cerevisiae using quadrupole FT hybrid mass spectrometry for top down proteomics. *Anal. Chem.* **2004,** 76, (10), 2852-2858.

7. Siuti, N.; Kelleher, N. L., Decoding protein modifications using top-down mass spectrometry. *Nat. Methods* **2007,** 4, (10), 817-821.

8. Dillon, T. M.; Bondarenko, P. V.; Ricci, M. S., Development of an analytical reversed-phase high-performance liquid chromatography electrospray ionization mass spectrometry method for characterization of recombinant antibodies. *J. Chromatogr. A* **2004,** 1053, (1053), 299-305.

9. Tran, J. C.; Doucette, A. A., Multiplexed Size Separation of Intact Proteins in Solution Phase for Mass Spectrometry. *Anal. Chem.* **2009,** 81, (15), 6201-6209.

10. Tran, J. C.; Doucette, A. A., Gel-eluted liquid fraction entrapment electrophoresis: An electrophoretic method for broad molecular weight range proteome separation. *Anal. Chem.* **2008,** 80, (5), 1568-1573.

11. Wang, Y.; Balgley, B. M.; Rudnick, P. A.; Lee, C. S., Effects of chromatography conditions on intact protein separations for top-down proteomics. *J. Chromatogr. A.* **2005,** 1073, 35-41.

12. Syka, J. E. P.; Marto, J. A.; Bai, D. L.; Horning, S.; Senko, M. W.; Schwartz, J. C.; Ueberheide, B.; Garcia, B.; Busby, S.; Muratore, T.; Shabanowitz, J.; Hunt, D. F., Novel linear quadrupole ion trap/FT mass spectrometer: Performance characterization and use in the comparative analysis of histone H3 post-translational modifications. *J. Proteome Res.* **2004,** 3, (3), 621-626.

13. Hu, Q. Z.; Noll, R. J.; Li, H. Y.; Makarov, A.; Hardman, M.; Cooks, R. G., The Orbitrap: a new mass spectrometer. *J. Mass Spectrom.* **2005,** 40, (4), 430-433.

14. Yates, J. R.; Cociorva, D.; Liao, L. J.; Zabrouskov, V., Performance of a linear ion trap-orbitrap hybrid for peptide analysis. *Anal. Chem.* **2006,** 78, (2), 493-500.

15. Zamdborg, L. L., R. D.; Glowacz, K. J.; Kim, Y.; Viswanathan, V.; Spaulding, I. T.; Early, B. P.; Bluhm, E. J.; Babai, S.; Kelleher, N. L. , ProSight PTM 2.0: improved protein identification and characterization for top down mass spectrometry. *Nucleic Acids Res.* **2007,** 35, W701-W706.

16. Pesavento, J. J.; Bullock, C. R.; LeDuc, R. D.; Mizzen, C. A.; Kelleher, N. L., Combinatorial Modification of Human Histone H4 Quantitated by Two-dimensional Liquid Chromatography Coupled with Top-Down Mass Spectrometry. *J. Biol. Chem.* **2008,** 283, (22), 1427-1437.

17. Solis, R. S.; Ge, Y.; Walker, J. W., Single Amino Acid Sequence Polymorphisms in Rat Cardiac Troponin Revealed by Top-Down Tandem Mass Spectrometry. *J. Muscle Res. Cell Motil.* **2008,** 29, 203-212.

18. Wu, S.; Lourette, N. M.; Tolic, N.; Zhao, R.; Robinson, E. W.; Tolmachev, A. V.; Smith, R. D.; Pasa-Tolic, L., An Integrated Top-Down and Bottom-Up Strategy for Broadly Characterizing Protein Isoforms and Modifications. *J. Proteome Res.* **2009,** 8, 1347-1357.

19. Gordon, E. F.; Mansoori, B. A.; Carroll, C. F.; Muddiman, D. C., Hydropathic influences on the quantification of equine heart cytochrome c using relative ion abundance measurements by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *J. Mass Spectrom.* **1999,** 34, (10), 1055-1062.

20. Martinovic, S.; Veenstra, T. D.; Anderson, G. A.; Pasa-Tolic´, L.; Smith, R. D., Selective incorporation of isotopically labeled amino acids for identification of intact proteins on a proteome-wide level. *J. Mass Spectrom.* **2002,** 37, (1), 99-107.

21. Veenstra, T. D.; Martinovic, S.; Anderson, G. A.; Pasa-Tolic´, L.; Smith, R. D., Proteome analysis using selective incorporation of isotopically labeled amino acids. *J. Am. Soc. Mass Spectrom.* **2000,** 11, (1), 78-82.

22. Julka, S.; Regnier, F., Quantification in proteomics through stable isotope coding: A review. *J. Proteome Res.* **2004,** 3, (3), 350-363.

23. Zhang, R. J.; Sioma, C. S.; Wang, S. H.; Regnier, F. E., Fractionation of isotopically labeled peptides in quantitative proteomics. *Anal. Chem.* **2001,** 73, (21), 5142-5149.

24. Du, Y.; Parks, B. A.; Sohn, S.; Kwast, K. E.; Kelleher, N. L., Top-Down Approaches for Measuring Expression Ratios of Intact Yeast Proteins Using Fourier Transform Mass Spectrometry. *Anal. Chem.* **2006,** 78, (3), 686-694.

25. Ong, S.; 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), 376-386.

26. Ong, S.; Kratchmarova, I.; Mann, M., Properties of 13C-Substituted Arginine in Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC). *J. Proteome Res.* **2003,** 2, 173-181.

27. Waanders, L. F.; Hanke, S.; Mann, M., Top-down quantitation and characterization of SILAC-labeled proteins. *J. Am. Soc. Mass Spectrom.* **2007,** 18, (11), 2058-2064.

28. Collier, T. S.; Hawkridge, A. M.; Georgianna, D. R.; Payne, G. A.; Muddiman, D. C., Top-down identification and quantification of stable isotope labeled proteins from Aspergillus flavus using online nano-flow reversed-phase liquid chromatography coupled to a LTQ-FTICR mass spectrometer. *Anal. Chem.* **2008,** 80, (13), 4994-5001.

29. Choi, M. Y.; An, Y. J.; Kim, S. H.; Roh, S. H.; Ju, H. K.; Hon, S. S.; Park, J. H.; Cho, K. J.; Cho, D. W.; Kwon, S. W., Mass Spectrometry Based Proteomic Analysis of Human Stem Cells: A Brief Review. *Exp. Mol. Med* **2007,** 39, (6), 690-695.

30. Swaney, D.; Wenger, C. D.; Thomson, J. A.; Coon, J. J., Human Embryonic Stem Cell Phosphoproteome Revealed by Electron Transfer Dissociation Tandem Mass Spectrometry. *Proc. Nat. Acad. Sci. U.S.A.* **2009,** 106, (4), 995-1000.

31. Van Hoof, D.; Munoz, J.; Braam, S. R.; Pinske, M. W.; Linding, R.; Heck, A. J.; Mummery, C. L.; Krijgsveld, J., Phosphorylation dynamics during early differentiation of human embryonic stem cells. *Cell Stem Cell* **2009,** 5, 214-226.

32. Phanstiel, D.; Brumbaugh, J.; Berggren, W. T.; Conard, K.; Feng, X.; Levenstein, M. E.; McAlister, G. C.; Thomson, J. A.; Coon, J. J., Mass Spectrometry Identifies and Quantifies 74 Unique Histone H4 Isoforms in Differentiating Human Embryonic Stem Cells. *Proc. Nat. Acad. Sci. U.S.A.* **2008,** 105, (11), 4093-4098.

33. Williamson, A. J. K.; Smith, D. L.; Blinco, D.; Unwin, R. D.; Pearson, S.; Wilson, C.; Miller, C.; Lancashire, L.; Lacaud, G.; Kouskoff, V.; Wheeton, A. D., Quantitative Proteomics Analysis Demonstrates Post-transcriptional Regulation of Embryonic Stem Cell Differentiation to Hamatopoiesis. *Mol. Cell. Proteomics* **2008,** 7, (3), 459-472.

34. Graumann, J.; Hubner, N. C.; Kim, J. B.; Ko, K.; Moser, M.; Kumar, C.; Cox, J.; Scholer, H.; Mann, M., Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC) and Proteome Quantitation of Mouse Embryonic Stem Cells to a Depth of 5,111 Proteins. *Mol. Cell. Proteomics* **2008,** 7, (4), 672-683.

35. Prokhorova, T. A.; Rigbolt, K. T.; Johansen, P. T.; Henningsen, J.; Kratchmarova, I.; Kassem, M.; Blagoev, B., Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC) and Quantitative Comparison of the Membrane Proteomes of Self-renewing and Differentiating Human Embryonic Stem Cells. *Mol. Cell. Proteomics* **2009,** 8, 959-970.

36. Xu, C.; Inokuma, M. S.; Denham, J.; Golds, K.; Kundu, P.; Gold, J. D.; Carpenter, M. K., Feeder-free growth of undifferentiated human embryonic stem cells. *Nat Biotechnol* **2001,** 19, (10), 971-4.

37. Van Hoof, D.; Pinkse, M. W.; Oostwaard, D. W.; Mummery, C. L.; Heck, A. J.; Krijgsveld, J., an experimental correction for arginine to proline conversion artifacts in SILAC-based quantitative proteomics. *Nature Methods* **2007,** 4, 677-678.

38. Price, P. J.; Goldsborough, M. D.; Tilkin, M. L. Embryonic stem cell serum replacement. International patent application WO98/30679, 1998.

39. Winston, R. L.; Fitzgerald, M. C., Concentration and Desalting of Protein Samples for Mass Spectrometry Analysis. *Anal. Biochem.* **1998,** 262, 83-85.

40. Shevchenko, A.; Jensen, O. N.; Podtelejnikov, A. V.; Sagliocco, F.; Wilm, M.; Vorm, O.; Mortensen, P.; Shevchenko, A.; Boucherie, H.; Mann, M., Linking genome and proteome by mass spectrometry: Large-scale identification of yeast proteins from two dimensional gels. *Proc. Nat. Acad. Sci. U. S. A.* **1996,** 93, (25), 14440-14445.

41. Andrews, G. L.; Shuford, C. M.; Burnett, J. C.; Hawkridge, A. M.; Muddiman, D. C., Coupling of a vented column with splitless nanoRPLC-ESI-MS for the improved separation and detection of brain natriuretic peptide-32 and its proteolytic peptides. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* **2009,** 877, (10), 948-954.

42. Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S., Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999,** 20, (18), 3551-3567.

43. Bendall, S. C.; Hughes, C.; Stewart, M. H.; Doble, B.; Bhatia, M.; Lajoie, G. A., Prevention of Amino Acid Conversion in SILAC Experiments with Embryonic Stem Cells. *Mol. Cell. Proteomics* **2008,** 7, (9), 1587-1597.

Chapter 4

Direct Comparison of Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC) and Spectral Counting for Quantitative Proteomics

4.1 Introduction

Mass spectrometry-based proteomics, which seeks to identify, quantify, and determine the function of all the proteins in a given biological system (whether organelle, organ, or organism) as they vary over time, space, and physiological state, is a robust and rapidly developing area of research which must adapt and develop novel analytical strategies to address the dramatic variation across dissimilar biological systems[1]. One of the most challenging aspects of proteomic analysis is the relative quantification of proteins once identified. Differential analysis of two-dimensional electrophoretic gels continues to be a widely-used tool for protein separation and quantification[2-6]. However, despite continuous improvements, challenges still exist for the analysis and quantification of proteins that reside at the limits of the gel's analytical capability, whether very high or low molecular weight, extremes in protein isoelectric points, or proteins of low abundance.

The incorporation of stable isotope labels, both chemically and metabolically, has been widely adopted as a protein quantification strategy as the stable isotope-labeled species possess similar physical and chemical properties as their unlabeled equivalents, but with a different mass. This allows direct comparison of multiple experimental conditions within the same analytical run. Numerous strategies have been developed for quantification of proteins and peptides using chemical derivatization. Gygi and coworkers developed isotope-coded affinity tags (ICAT)

containing a thiol specific reactive group, an isotope labeled linker region containing eight deuteriums (or hydrogens) and a biotin affinity tag[7, 8]. The initial application of ICAT labeling to a complex protein mixture indicated relative standard deviation of peptide area ratios of around 10%, although the potential for greater error has been reported as a result of deuterium labels producing peak retention times that differ between the light and heavy labeled counterparts by as much as 30 seconds[9-11]. Isobaric tags for relative and absolute quantitation (iTRAQ) is conducive to the direct comparison of four or more experimental samples simultaneously by labeling the N-termini of peptides with a compound containing a reporter group with varying stable isotope (^{13}C, ^{15}N, and ^{18}O) incorporation, a balancing group, and an amine specific reactive group. Tandem mass tags (TMTs) also employ an isobaric chemical labeling strategy for relative quantification. All versions of the tag possess the same total mass, with quantification achieved during MS/MS, in which the reporter groups with varying masses are generated with abundances correlating to the relative abundance of the peptide in each experimental pool[9, 12]. Enzymatically catalyzed ^{18}O-labeling has also been employed, incorporating either ^{16}O or ^{18}O into peptide C-termini during proteolysis to generate light and heavy peptide pairs for relative quantification[13-18].

Stable isotope labels are also introduced to samples for proteomic analysis through metabolic processes including cell and tissue culture, in which growth media is enriched with the label of choice. This strategy was introduced for bacterial proteomics by Oda *et al*.[19]. Deuterated leucine has also been used to label cultures

of *E. coli* and *S. cerevisiae* but proved to be less accurate than other labeling methods due to the effect of deuterium on retention time, as discussed *vide* supra[10, 11, 20, 21]. $^{15}N/^{14}N$ metabolic labeling has been successfully utilized for quantification of the yeast proteome and has also been successfully utilized for quantification in plant species[22]. Stable isotope labeling by amino acids in cell culture (SILAC) was introduced by Mann and coworkers as a metabolic labeling strategy whereby amino acids containing stable isotopes are supplemented to cell culture growth media to produce peptide analytes that chromatographically co-elute[23, 24]. More recently this strategy has been developed for the *in vivo* labeling of mice[25].

In addition to the stable isotope based methods for protein quantification, label-free strategies continue to be developed. One such strategy uses integrated ion abundance obtained from extracted ion chromatograms to track peptide and protein abundance across experimental samples. Bondarenko and coworkers demonstrated the linear relationship of peptide peak areas over a range of 10 to 1,000 fmol of myoglobin digest spiked into a human plasma sample, with a relative standard deviation near 10%[26, 27]. Another promising method for label-free quantification is spectral counting, where the number of mass spectra contributing to identification of the protein are used as an indication of that protein's abundance[28]. Standard peptides spiked into yeast lysate showed a linear relationship between protein abundance and spectral counts over two orders of magnitude. Additionally, the number of spectral counts can be normalized to protein length allowing for the relative quantification of two different proteins[29]. Overall, these two label-free

methods are comparable, with spectral counting reporting to show more sensitivity to changes in protein abundance, and peak area quantification showing slightly higher accuracy in determining relative protein ratios[30]. A recent comparative study by Stevenson et al. found the statistical processing associated with spectral counting to be more amenable to high throughput quantification relative to ion peak integration[31].

The quantitative proteomics strategy employed in any given experiment is dependent on a number of factors. In many cases, biological systems of interest may not be capable of metabolic labeling, and chemical isotopic labeling may be cost and/or time prohibitive. The rice blast fungus *Magnaporthe oryzae* is a potent agricultural pathogen with impacts in the billions of dollars *per annum* affecting a dietary staple of half the planet's human population[32]. However, understanding how the fungus infects and destroys rice from a proteomic perspective is challenging at best using a stable isotope labeling strategy such as SILAC. This is primarily a result of the fact that fungi such as *M. oryzae* and the previously studied *Aspergillus flavus* are prototrophic organisms capable of synthesizing their own amino acids regardless of any that may be supplemented to their growth media[33, 34]. Also, clinical samples and most animal based samples cannot employ metabolic label-based quantification strategies. Studies of this nature stand to directly benefit from label-free quantification strategies and a more thorough understanding of how label-free techniques compare to stable-isotope facilitated methodologies is warranted. In order to compare SILAC and spectral counting as methods of quantification, we

have employed human embryonic stem cells (hESCs), which we are able to readily label using SILAC.

Few examples exist in the literature for the direct comparison of stable isotope labeling and label-free quantification. Hendrickson *et al.* reported on the comparison of $^{15}N/^{14}N$ with spectral counting in their study of two microbial species and suggest a lesser degree of sensitivity in detecting a two-fold expression change with spectral counting[35]. In a comparison of iTRAQ and label-free quantification using spectral counting, spectral counting was found to identify many more proteins with greater sequence coverage when compared to iTRAQ, with both techniques possessing similar quantification capabilities[36]. Recently, Parker *et al.* used spectral-counting to analyze SILAC data sets as a means of demonstrating the ability to use a SILAC experimental design to combine different experimental states, but not require high mass accuracy mass spectrometers for data collection, using the spectral counts for quantification of light and heavy peptides separately[37]. In this report, we use a similar strategy to evaluate and directly compare the performance of SILAC and label-free quantification using spectral counting on a 1:1 mixture of SILAC labeled human embryonic stem cells. Within this context, we evaluate the number of proteins identified at a 1% false discovery rate, the confidence intervals of quantification, as well as assessing the constraints of each quantification strategy.

4.2 **Experimental**

4.2.1 Cell Culture and Sample Preparation

H9 human embryonic stem cells were maintained on mouse embryonic fibroblast (MEF) conditioned medium according to a protocol adapted from Xu et. Al.[38] Briefly, MEF cells were inactivated by mitomycin c (Sigma Aldrich, St. Louis, MO, USA) treatment and plated at 55,000 cells/cm^2. HESC medium consisting of DMEM/F12 (Invitrogen, Carlsbad, CA, USA), 20% knockout serum replacement (Invitrogen), 1 mM L-glutamine (Invitrogen), 0.1 mM β-mercaptoethanol, 1% nonessential amino acids (Invitrogen), and 4 ng/mL basic fibroblast growth factor (bFGF; Sigma Aldrich) was conditioned for 24 hours on MEF cells. The conditioned media was collected for 7 days, pooled, and supplemented with an additional 4 ng/mL bFGF (Sigma Aldrich). Conditioned media for SILAC labeling of hESCs was prepared using protocols previously described[39]. Briefly, SILAC medium consisted of DMEM/F12 without L-lysine and L-arginine, $^{13}C_6$-L-arginine (Pierce, Rockford, IL, USA), $^{13}C_6$, $^{15}N_2$-L-lysine (Pierce), 20% knockout serum replacement (not containing lysine or arginine), 1 mM L-glutamine, 0.1mM β-mercaptoethanol, 1% nonessential amino acids and 4 ng/mL bFGF, which was conditioned over MEF cells. Stable isotope labeled arginine and lysine incorporation of 98% and 98.5%, respectively, was achieved as previously reported [40].

Once confluent, hESC cells were lysed with 8 M Urea (Sigma Aldrich) and 50 mM ammonium bicarbonate (Fisher Scientific, Hampton, NH, USA) and frozen at -80 °C until needed. Unlabeled and SILAC labeled hESC lysate were combined at a

ratio of 1:1 according to total protein concentration as determined by Bradford assay using a BSA standard curve.

4.2.2 GELFrEE Fractionation

200 µg of total protein from a 1:1 mixture of unlabeled and SILAC labeled hESC lysate was fractionated using a mid-mass range gel cartridge on a GELFrEE[41, 42] 8100 Fractionator (Protein Discovery, Knoxville, TN, USA) according to the manufacturer's instructions. Twelve 150 µL fractions were recovered at different times along the course of the fractionation. SDS was removed by methanol/water/chloroform precipitation using a protocol described by Wessel and Flugge[43]. Protein pellets were redissolved in 100 µL of 100 mM ammonium bicarbonate pH 8.0 for further processing.

4.2.3 Protein Digestion

In-solution reduction, alkylation, and digestion were performed on each GELFrEE fraction as follows. Cys-cys disulfide linkages were reduced through the addition of 10 µL of 500 mM dithiothreitol to each fraction and incubation at 56 °C for one hour with occasional vortexing before removing and allowing to cool to room temperature. The free thiols were then carbamidomethlyated by adding 15 µL of 500 mM iodoacetamide to each sample and incubating for one hour in the dark at room temperature. Digestion was performed for 18 hours at 37 °C in the presence of 1.5 µg of proteomics grade trypsin in each fraction. Digestion was halted with the addition of 10 µL of formic acid to each sample. The resulting peptide fractions were then stored at -80 °C prior to LC-MS/MS analysis.

4.2.4 Nano-flow Liquid Chromatography

Reversed phase liquid chromatography was performed on an Eksigent nanoLC-2D system (Dublin, CA) running under a vented-column configuration as previously described[44]. All LC solvents were purchased from Burdick and Jackson (Muskegon, MI, USA). Briefly, mobile phase A and B consisted of 2% and 98% acetonitrile in water, respectively. Each mobile phase was supplemented with 0.2% (vol/vol) formic acid as an ion pairing reagent. Four microliters of sample were injected onto a 100 µm ID IntegraFrit trap column packed in-house to ~3 cm with Magic C18AQ (5 µm particle, 200 Å pore size) stationary phase (Michrom Bioresources, Auburn, CA, USA). The analytical column consisted of a 75 µm ID capillary coupled to a 15 µm PicoFrit tip packed in-house to ~20 cm with Magic C18AQ stationary phase. Peptide samples were trapped and desalted at 1.5 µL/min before switching in-line with the gradient flow rate at 300 nL/min using a 10-port valve (VICI, Houston, TX, USA). The gradient remained at 2% B for 5 min prior to increasing to 10% B over 2 min. The gradient then ramped to 50% B over the next 60 min, and then increased to 95% B over 3 min, holding for 5 min, then re-equilibrated at 2% B for 10 min. The total time of the chromatographic run was 85 minutes. IntegraFrit and PicoFrit columns were from New Objective (Woburn, MA, USA).

4.2.5 LTQ-FTICR Mass Spectrometry

Peptides eluting from the analytical column were directly introduced via electrospray ionization to a hybrid LTQ-FT[45] Ultra equipped with a 7T

superconducting magnet (Thermo Fisher Scientific, San Jose, CA). Mass calibration was performed as specified by the manufacturer. Peptides were detected using a pulse sequence containing four scan events. A precursor ion scan was performed in the ICR cell to identify the 3 most abundant candidates for MS/MS analysis with an AGC of $1x10^6$ (maximum injection time of 1 s) and resolving power of $100,000_{FWHM}$ at 400 m/z. The three data-dependent MS/MS events were carried out in the ion trap with an AGC of $1x10^4$, an isolation width of 2 m/z and normalized collision energy of 30% for 30 ms. The isolation Q was set to 0.25. 1+ charge states were rejected from MS/MS analysis and dynamic exclusion was employed for 45 seconds after 2 repeat analyses in a 30 second window. A total of 12 samples were analyzed with 3 technical replicates each for a total of 36 .RAW data files.

4.2.6 Data Processing

.RAW LC-MS/MS files were processed using MASCOT Distiller (Matrix Science, Boston, MA, USA) to generate peak lists in .mgf format for database searching using MASCOT[46] server (Matrix Science). For SILAC quantification, the data files were searched with carbamidomethylation of cysteine as a fixed modification and with variable modifications including glutamine and asparagines deamidation, methionine oxidation, $^{13}C_6$-arginine, and $^{13}C_6$, $^{15}N_2$-lysine. For label-free quantification using spectral counting, database searches were designed to only look for light peptides or to only look for heavy peptides. This was accomplished by searching with the same modifications as above, with the exception searching $^{13}C_6$-arginine, and $^{13}C_6$, $^{15}N_2$-lysine as fixed modifications for heavy peptides, or not at all

for light peptides. All searches were performed against the IPI human database (v3.63) containing both target and reverse protein sequences for determination of protein false discovery rate with a precursor ion tolerance of ±5 ppm and a fragment ion tolerance of ±1 Da. SILAC quantification was performed within the quantification module of MASCOT Distiller, while label-free quantification using spectral counts was performed by importing MASCOT results files into ProteoIQ (BioInquire, Athens, GA, USA).

4.3 Results and Discussion

The use of the a metabolic labeling strategy such as SILAC was compared directly with label-free quantification strategy of spectral counting by searching the same LC-MS .RAW files in different ways appropriate to the corresponding quantitative strategy. The overall experimental strategy for this study is summarized in **Figure 4.1**, in which human embryonic stem cells were cultured in a light media containing $^{12}C_6$-arginine and $^{12}C_6$, $^{14}N_2$-lysine and a heavy media containing $^{13}C_6$-arginine and $^{13}C_6$, $^{15}N_2$-lysine. The hESC lysate was combined 1:1 based on total protein concentration as determined by Bradford assay, followed by GELFrEE fractionation and nanoLC-MS/MS analysis on an LTQ-FT-ICR mass spectrometer. The resulting .RAW files were processed using MASCOT distiller to generate .mgf peak list files, which were subsequently searched using MASCOT server. In the case of label-free quantification using spectral counting, light and heavy peaks were searched separately and treated as separate samples when quantified in ProteoIQ. For

SILAC quantification, the heavy and light data were searched together, with SILAC quantification performed within the MASCOT distiller software platform.

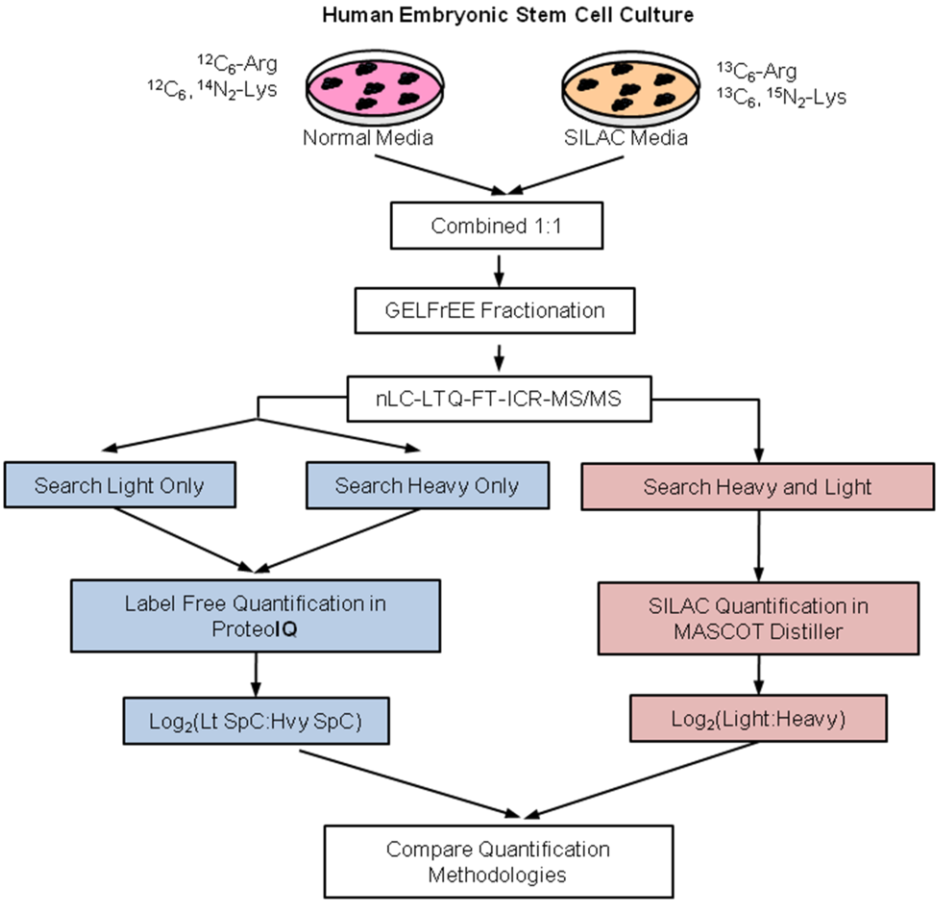

Figure 4.1 Outline of the experimental strategy for the direct comparison of SILAC and label-free quantification using spectral counting. An identical set of LC-MS/MS data from SILAC labeled human embryonic stem cells was analyzed using the two separate quantification methodologies.

From these separate quantification methods, proteins that were identified in both quantification strategies could be utilized for direct methodological comparison.

A summary of the results of label-free and SILAC quantification of hESC proteins in an experimental control are shown in **Table 4.1**. As both techniques were performed on the same set of raw data, it is expected that the results would be

similar, with the exception of the number of proteins quantified. Label-free quantification using spectral counting yielded 885 quantified proteins out of 932 identified at 1% FDR, meaning each protein has at least one spectral count from both heavy and light peptides. However, confident quantification required more than a single spectral count (*vide infra)*, and of the 885 proteins quantified, only 304 of those had more than 5 SpC with which to more accurately quantify. By comparison, the quantification module in MASCOT Distiller gave rise to 450 quantified proteins out of the 998 identified with 1% FDR. It should be noted that while more than 450 proteins had peptides quantified by SILAC, proteins had to have at least 2 accurately quantified peptides to be reported by the quantification module.

Table 4.1 Overall results of Label Free and SILAC Experimental Strategies

	Label Free	*SILAC*
Number of Peptides	15,168	16,594
Number of Unique Peptides	4,435	4,863
Proteins Identified (1% FDR)	932	998
Proteins Quantified	885 (*304 w/ >5 SpC*)	450

While similar numbers of peptides, unique peptides, identified proteins at 1% false discovery rate, and proteins quantified are reported for each quantification strategy, it is important to note that these results are unique to the experimental design of a direct comparison. For instance, a two-condition labeling experiment consists of two different samples being combined in a single LC-MS experiment, so

in a typical 1 µg column load, only 500 ng is injected from either experimental condition. A label-free experiment would consist of separate LC-MS runs, allowing the maximum column load for each experimental condition, or twice the capacity of a two-condition labeling experiment. The increased sample load afforded by a label-free workflow would thus be expected to produce an increase in detected peptides and at the very least an increase in spectral counts allowing more accurate quantification. For the purpose of directly comparing SILAC quantification and label-free quantification using spectral counting, spectral counts of light and heavy members of stable isotope-labeled pairs are counted as opposed to using the spectral counts of two separate runs, eliminating run-to-run variability that would otherwise complicate the comparison of quantification strategies. Thus, the total protein in this label-free experiment is half that of a typical experiment employing two separate LC-MS runs per comparison.

An example of the two quantification methodologies employed in this study is shown in **Figure 4.2** for the quantification of ribosomal protein S10 from hESC lysate. Two different peptides, AEAGAGSATEFQFR^{2+} and IAIYELLFK^{2+}, contributed to the protein's identification using MASCOT server, the former being present in two out of three technical replicate runs of the same GELFrEE fraction. SILAC quantification was performed by MASCOT Distiller by obtaining the extracted ion chromatograms of the light and heavy labeled members of each peptide SILAC pair and determining the ratios of the light to heavy peak area.

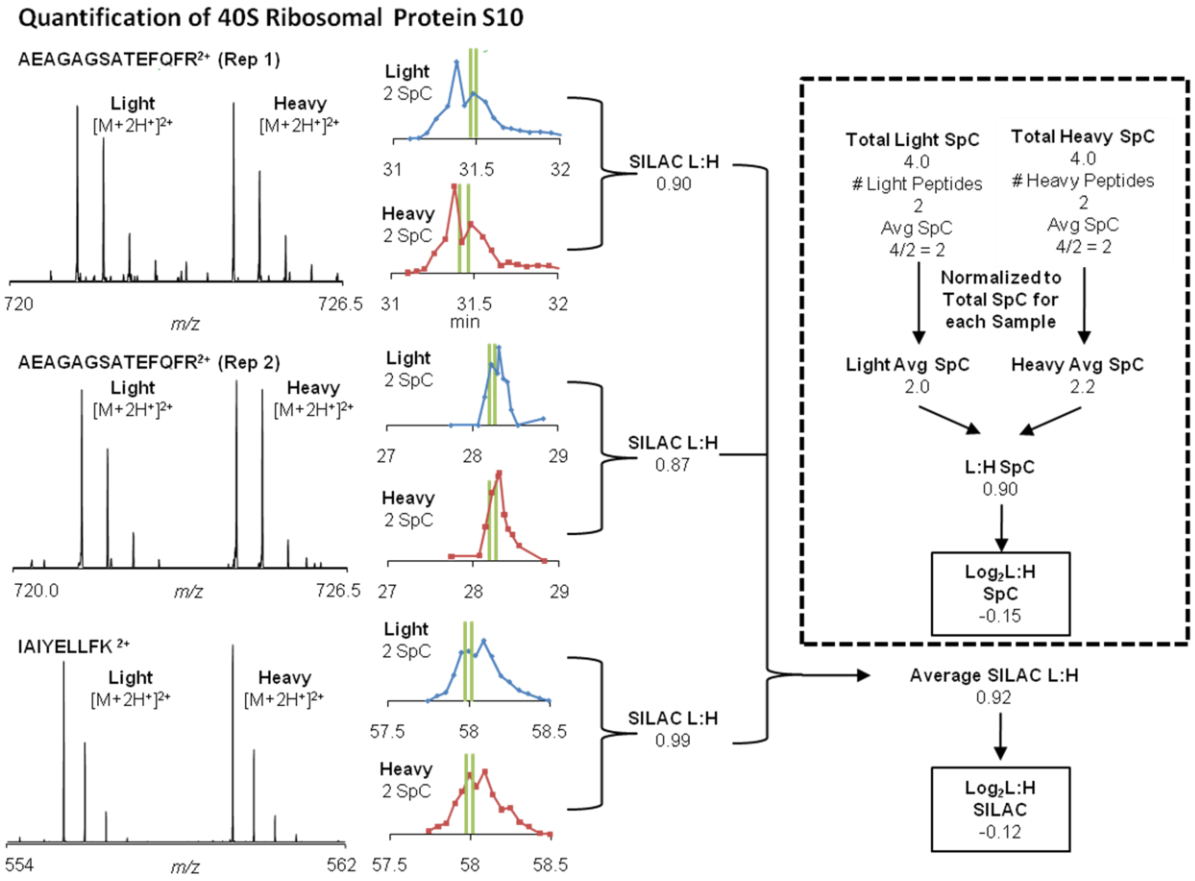

Figure 4.2 The relative quantification of 40S ribosomal protein S10 from SILAC labeled hESCs using both SILAC quantification by averaging the area of the extracted ion chromatograms from each peptide pair and spectral counting of peptides over technical replicates. Individual spectral counts are indicated with vertical green bars in each extracted ion chromatogram.

The protein SILAC ratio is then determined by the average of the peptide SILAC ratios; in this case peptide SILAC ratios of 0.90, 0.87, and 0.99 are averaged to give a protein SILAC ratio of 0.92. To quantify the protein using spectral counting requires a few simple calculations. First, the spectral counts of each peptide are averaged over its appearance in all technical replicates. For example, the four spectral counts for the light condition of AEAGAGSATEFQFR^{2+} are averaged for its two replicates giving rise to two spectral counts for that peptide in the light

experimental condition. The same calculation for IAIYELLFK^{2+} gives two spectral counts as well for its detection in one out of three technical replicates. The sum of the spectral counts from the constituent peptides of ribosomal protein S10 are thus calculated (4 in this case) for both the light and heavy experimental conditions. The average spectral counts per peptide are then calculated by dividing the total number of spectral counts determined *vide supra* by the number of different peptides (2 in this case) to give 2 average spectral counts for both the light and heavy labeled peptides. To account for any deviation in run-to-run reproducibility, the average spectral counts are then normalized to the total number of spectral counts for each experimental condition. Searching the same data set for light and heavy peaks separately yielded 10% more total light spectral counts than total heavy spectral counts for the entire experiment, and so the heavy average spectral count is adjusted upward by 10%. Taking the ratio of the light to heavy in this case gives a ratio of 0.90, which closely approximates the value determined using SILAC quantification.

By quantifying proteins identified and quantified using both SILAC and spectral counting methodologies, we are able to directly compare the variation of each method. **Figure 4.3A** depicts the distribution on a logarithmic base 2 scale of quantified proteins present in both quantification protocols described *vide supra*. In all, 317 proteins were quantified in both the label-free and SILAC results. SILAC quantification gave rise to an average Log_2(L:H) of -0.16 with a 95% confidence interval of extending from -.44 to 0.12, while the same 317 proteins quantified by

spectral counting gave a similar average Log_2(L:H) of -0.13, but with a 95% confidence interval that spans from -0.76 to 1.02, a nearly two-fold change in the negative direction to a two-fold change of expression in the positive (Log_2(L:H)=1 or -1). The principle cause of such a wide confidence interval arises from proteins with a very low number of spectral counts contributing to their quantification.

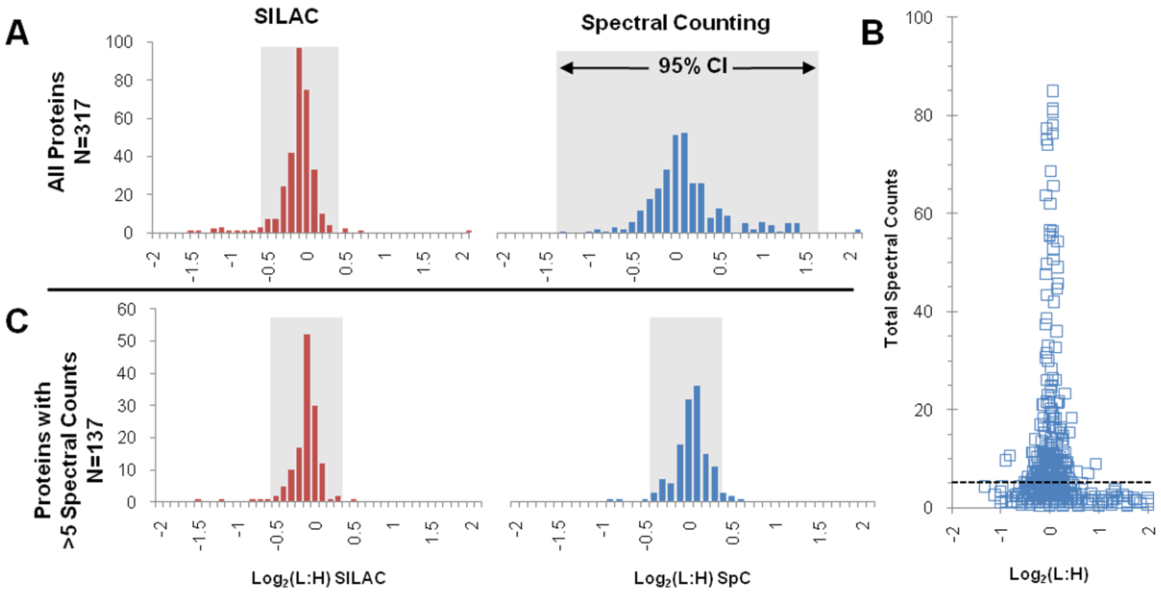

Figure 4.3 (A)The analytical variation of quantification by either SILAC or spectral counting of all proteins identified by both quantification strategies, (B) the distribution of L:H ratios determined by spectral counting plotted against the total number of spectral counts indicates the necessity for a cut-off of 5 SpC to achieve accurate quantification and (C) the analytical variation of each quantification strategy after the 5 SpC threshold is applied to proteins quantified by each strategy.

This distribution of spectral-counting based L:H ratios as a function of the number of spectral counts is shown in **Figure 4.3B**, where it is clearly evident that the majority of variation occurs at lower abundance as determined by number of spectral counts. Establishing a conservative cut-off value of 5 spectral counts needed for more accurate quantification successfully reduces the confidence interval of quantification using spectral counts to a level comparable to SILAC quantification whose

confidence interval remained relatively unchanged relative to no spectral count cut-off (**Figure 4.3C**). Our established spectral count cut-off is slightly more conservative but agrees with the 4 spectral count cut-off employed by Old *et al.* for human erythroleukemia cell cultures[30]. It must be noted that while establishing a 5 spectral count threshold for quantification resulted in a narrower confidence interval, only 137 of the 317 proteins identified and quantified using both quantification methods met the threshold. 304 of 885 proteins quantified by spectral counts alone met the 5 spectral count threshold as well (see **Table 1**). Challenges in the quantification of low abundant proteins are readily apparent in establishing a 5 spectral count threshold. However, as discussed above, a typical label-free experiment consisting of two separate runs and thus a higher sample loading capacity would result in more quantified proteins overall.

The correlation of SILAC quantification and spectral counting can be seen in the plot shown in **Figure 4.4**, in which each point represents a protein identified and quantified using both methodologies. Within a quantitative control experiment, in which all proteins should be present at a 1:1 ratio, a perfect correlation between these two methodologies would have all the data points clustered tightly about the origin. Taking into account all 317 quantified proteins regardless of the number of spectral counts one can observe an diffuse distribution of the data points. As the number of spectral counts for the quantified proteins increases from >5 spectral counts (green squares) to >20 spectral counts (red triangles), the points begin to cluster about the origin, illustrating again that the analytical variability of

quantification decreases with increasing numbers of spectral counts as one would expect, and thus correlation between the two different quantitative techniques would also increase.

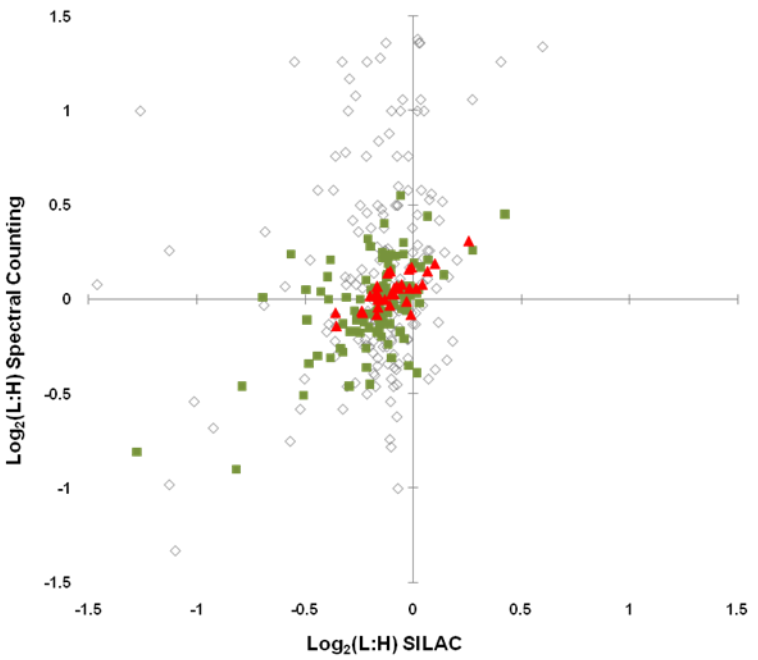

Figure 4.4 Correlation plot of quantified proteins plotted against the Log_2(L:H) of either SILAC (x-axis) or spectral counting (y-axis). Open diamonds indicate 1-5 spectral counts, green squares indicate 5-20 spectral counts, and red triangles depict proteins with more than 20 spectral counts.

4.4 Conclusion

The quantification of proteins is one of the paramount objectives of proteomics and a number of strategies have been developed to address its challenges, including chemical derivitization and metabolic labeling with stable isotope-labeled reagents and label-free methods such as peak area quantification and spectral counting. Investigators must consider many factors when choosing a quantitative strategy for their particular biological system, as some options may be incapable for efficient implementation. This is the case for SILAC labeling of fungal organisms such as *M. oryzae*, which are capable of synthesizing all of their amino acid constituents. Label-free methodologies such as spectral counting are an attractive alternative for such a dilemma, although to date the capabilities of SILAC and spectral counting had yet to be directly compared. Using human embryonic

stem cells as a model system that could be readily quantified by both SILAC and spectral counting, we were able to establish similar capabilities in protein identification by analyzing an identical data set with each strategy. With respect to protein quantification, spectral counting was inherently able to quantify more proteins (885) than SILAC (450), although less accurately unless a 5 spectral count limit was established for protein quantification, reducing the number of proteins quantified by spectral counting to 340. In a normal experimental setting, a label-free strategy allows for double the total protein amount to be analyzed using spectral counting compared to SILAC. Additionally, samples analyzed by label-free typically possess half the relative complexity of a SILAC sample as a result of single peptide precursors in label-free versus light and heavy pairs in SILAC. Future experiments will seek to evaluate the potential of these strategies, including their comparative sensitivity to changes in protein abundance and dynamic ranges of quantification. As a result of these inherent experimental properties, the potential exists for label-free quantification using spectral counting to quantify more proteins than a given SILAC experiment.

4.5 References

1. Kenyon, G. L.; DeMarini, D. M.; Fuchs, E.; Galas, D. J.; Kirsch, J. F.; Leyh, T. S.; Moos, W. H.; Petsko, G. A.; Ringe, D.; Rubin, G. M.; Sheahan, L. C. Defining the mandate of proteomics in the post-genomics era: workshop report. *Mol. Cell. Proteomics* **2002**, *1*, 763-780.

2. Link, A.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvic, B. M.; Yates, J. R. Direct analysis of protein complexes using mass spectrometry. *Nat. Biotechnol.* **1999**, *17*, 676-682.

3. Shevchenko, A.; Jensen, O. N.; Podtelejnikov, A. V.; Sagliocco, F.; Wilm, M.; Vorm, O.; Mortensen, P.; Shevchenko, A.; Boucherie, H.; Mann, M. Linking genome and proteome by mass spectrometry: large scale identification of yeast proteins from two dimensional gels. *Proc. Nat. Acad. Sci. U. S. A.* **1996**, *93*, 14440-14445.

4. Gygi, S. P.; Rochon, Y.; Franza, B. R.; Aebersold, R. Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* **1999**, *19*, 1720-1730.

5. Garrels, J. I.; McLaughlin, C. S.; Warner, J. R.; Futcher, B.; Latter, G. I.; Kobayashi, R.; Schwender, B.; Volpe, T.; Anderson, D. S.; Mesquita-Fuentes, R.; Payne, W. E. Proteome studies of Saccharomyces cerevisiae: Identification and characterization of abundant proteins. *Electrophoresis* **1997**, *18*, 1347.

6. Boucherie, H.; Sagliocco, F.; Maillet, I.; Labarre, J.; Perrot, M. Two-dimensional gel protein database of Saccharomyces cerevisiae *Electrophoresis* **1996**, *17*, 1683-1699.

7. 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*, 994-999.

8. Li, J. X.; Steen, H.; Gygi, S. P. Protein profiling with cleavable isotope-coded affinity tag (cICAT) reagents - The yeast salinity stress response. *Mol. Cell. Proteomics* **2003**, *2*, 1198-1204.

9. Ross, P. L.; Huang, Y. L. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Prkayastha, S.; Jahasz, P.; Martin, S.; Bartlet-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* **2004**, *3*, 1154-1169.

10. Zhang, R. J.; Sioma, C. S.; Wang, S. H.; Regnier, F. E. Fractionation of isotopically labeled peptides in quantitative proteomics. *Anal. Chem.* **2001**, *73*, 5142-5149.

11. Julka, S.; Regnier, F. Quantification in proteomics through stable isotope coding: A review. *J. Proteome Res.* **2004**, *3*, 350-363.

12. Thompson, A.; Schafer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Johnstone, R. A. W.; Mohammed, A. K. A.; Hamon, C. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* **2003**, *75*, 1895-1904.

13. Desiderio, K. M.; Kai, M. Preparation of stable isotope-incorporated peptide standards for field desorption mass-spectrometry quantification of peptides in biologic tissue. *Biomed. Mass Spectrom.* **1983**, *10*, 471-479.

14. Mirgorodskaya, O. A.; Kozmin, Y. P.; Titov, M. I.; Korner, R.; Sonksen, C. P.; Roepstorff, P. Quantitation of peptides and proteins by matrix-assisted laser desorption/ionization mass spectrometry using O-18-labeled internal standards. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 1226-1232.

15. Yao, X. D.; Afonso, C.; Fenselau, C. *J.* Dissection of proteolytic O-18 labeling: endoproteinase-catalyzed O-16-to-O-18 exchange of truncated peptide substrates. *Proteome Res.* **2003**, *2*, 147-152.

16. Yao, X. D.; Freas, A.; Ramirez, J.; Demirev, P. A.; Fenselau, C. Proteolytic O-18 labeling for comparative proteomics: Model studies with two serotypes of adenovirus. *Anal. Chem.* **2001**, *73*, 2836-2842.

17. Mason, C. J.; Therneau, T. M.; Eckel-Passow, J. E.; Johnson, K. L.; Oberg, A. L.; Olsen, J. E.; Nair, K. S.; Muddiman, D. C.; Bergen, H. R. A method for automatically interpreting mass spectra of 18O-labeled isotopic clusters. *Mol. Cell. Proteomics* **2007**, *6*, 305-318.

18. Johnson, K. L.; Muddiman, D. C. A method for calculating O-16/O-18 peptide ion ratios for the relative quantification of proteomes. *J. Am. Soc. Mass Spectrom* **2004**, *15*, 437-445.

19. Oda, Y.; Huang, K.; Cross, F. R.; Cowburn, D.; Chait, B. T. Accurate Quantitation of protein expression and site-specific phosphorylation. *Proc. Nat. Acad. Sci. U. S. A.* **1999**, *96*, 6591-6596.

20. Martinovic, S.; Veenstra, T. D.; Anderson, G. A.; Pasa-Tolic´, L.; Smith, R. D. Selective incorporation of isotopically labeled amino acids for identification of intact proteins on a proteome-wide level. *J. Mass Spectrom.* **2002**, *37*, 99-107.

21. Veenstra, T. D.; Martinovic, S.; Anderson, G. A.; Pasa-Tolic´, L.; Smith, R. D. Proteome analysis using selective incorporation of isotopically labeled amino acids. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 78-82.

22. Schaff, J. E.; Mbeunkui, F.; Blackburn, K.; Bird, D. M.; Goshe, M. B. SILIP: a novel stable isotope labeling method for in planta quantitative proteomic analysis. *The Plant Journal*` **2008**, *56*, 840-854.

23. Ong, S.; 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**, 376-386.

24. Ong, S.; Kratchmarova, I.; Mann, M. Properties of 13C-substituted arginine in stable isotope labeling by amino acids in cell culture (SILAC). *J. Proteome Res.* **2003**, *2*, 173-181.

25. Kruger, M.; Moser, M.; Ussar, S.; Thievessen, I.; Luber, C. A.; Forner, F.; Schmidt, S.; Zanivan, S.; Fassler, R.; Mann, M. SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function. *Cell* **2008**, *134*, 353-364.

26. Chelius, D.; Bondarenko, P. V. Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *J. Proteome Res.* **2002**, *1*, 317-323.

27. Bondarenko, P. V.; Chelius, D.; Shaler, T. A. Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography-tandem mass spectrometry. *Anal. Chem.* **2002**, *74*, 4741-4749.

28. Wolters, D.; Washburn, M. P.; Yates, J. R. An automated multidimensional protein identification technology for shotgun proteomics. *Anal. Chem.* **2001**, *73*, 5683-5690.

29. Zybailov, B.; Mosley, A. L.; Sardiu, M. E.; Coleman, M. K.; Florens, L.; Washburn, M. P. Statistical analysis of membrane proteome expression changes in Saccharomyces cerevisiae. *J. Proteome Res.* **2006**, *5*, 2339-2347.

30. Old, W. M.; Meyer-Arendt, K.; Aveline-Wolf, L.; Pierce, K. G.; Mendoza, A.; Sevinsky, J. R.; Resing, K. A.; Ahn, N. G. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol. Cell. Proteomics* **2005**, *4*, 1487-1502.

31. Stevenson, S. E.; Chu, Y.; Ozias-Atkins, P.; Thelen, J. J. Validation of gel-free, label-free quantitative proteomics approaches: applications for seed allergen profiling. *J. of Proteomics* **2009**, *72*, 555-566.

32. Wilson, R. A.; Talbot, N. J. Under pressure: investigating the biology of plant infection by Magnaporthe oryzae. *Nat. Rev. Microbiol.* **2009**, *7*, 185-195.

33. Collier, T. S.; Hawkridge, A. M.; Georgianna, D. R.; Payne, G. A.; Muddiman, D. C. Top-down identification and quantification of stable isotope labeled proteins from Aspergillus flavus using online nano-flow reversed-phase chromatography coupled to a LTQ-FTICR mass spectrometer. *Anal. Chem.* **2008**, *80*, 4994-5001.

34. Georgianna, D. R.; Hawkridge, A. M.; Muddiman, D. C.; Payne, G. A. Temperature-dependent regulation of proteins in Aspergillus flavus: Whole organism stable isotope labeling by amino acids. *J. Proteome Res.* **2008**, *7*, 2973-2979.

35. Hendrickson, E. L.; Xia, Q. W.; Wang, T. S.; Leigh, J. A.; Hackett, M. Comparison of spectral counting and metabolic stable isotope labeling for use with quantitative microbial proteomics. *Analyst* **2006**, *131*, 1335-1341.

36. Patel, W. J.; Thalassinos, K.; Slade, S. E.; Connolly, J. B.; Crombie, A.; Murrell, J. C.; Scrivens, J. H. A comparison of labeling and label-free mass spectrometry based proteomics approaches. *J. Proteome Res.* **2009**, *8*, 3752-3759.

37. Parker, S. J.; Halligan, B. D.; Greene, A. S. Quantitative analysis of SILAC data sets using spectral counting. *Proteomics* **2010**, *10*, 1408-1415.

38. Xu, C.; Inokuma, M. S.; Denham, J.; Golds, K.; Kundu, P.; Gold, J. D.; Carpenter, M. K. Feeder-free growth of undifferentiated human embryonic stem cells. *Nat. Biotechnol.* **2001**, *19*, 971-974.

39. Van Hoof, D.; Pinkse, M. W.; Oostwaard, D. W.; Mummery, C. L.; Heck, A. J.; Krijgsveld, J. An experimental correction for arginine to proline conversion artifacts in SILAC-based quantitative proteomics. *Nat. Methods* **2007**, *4*, 677-678.

40. Collier, T. S.; Sarkar, P.; Rao, B.; Muddiman, D. C. Quantitative top-down proteomics of SILAC labeled human embryonic stem cells. *J. Am. Soc. Mass Spectrom* **2010**, *21*, 879-889.

41. Tran, J. C.; Doucette, A. A. Gel-eluted liquid fraction entrapment electrophoresis: An electrophoretic method for broad molecular weight range proteome separation. *Analytical chemistry* **2008**, *80*, 1568-1573.

42. Tran, J. C.; Doucette, A. A. Multiplexed size separation of intact proteins in solution phase for mass spectrometry. *Anal. Chem.* **2009**, *81*, 6201-6209.

43. Wessel, D.; Flugge, U. I. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal. Biochem.* **1984**, *138*, 141-143.

44. Andrews, G. L.; Shuford, C. M.; Burnett, J. C.; Hawkridge, A. M.; Muddiman, D. C. Coupling of a vented column with splitless nanoRPLC-ESI-MS for the improved separation and detection of brain natriuretic peptide-32 and its proteolytic peptides. *J. Chrom. B* **2009**, *877*, 948-954.

45. Syka, J. E. P.; Marto, J. A.; Bai, D. L.; Horning, S.; Senko, M. W.; Schwartz, J. C.; Ueberheide, B.; Gacia, B.; Busby, S.; Muratore, T.; Shabanowitz, J.; Hunt, D. F. Novel linear quadrupole ion trap/FT mass spectrometry: performance characterization and use in the comparative analysis of histone H# post-translational modifications. *J. Proteome Res.* **2004**, *3*, 621-626.

46. Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999**, *20*, 3551-3567.

Chapter 5

Comparison of Stable Isotope Labeling by Amino Acids in Cell Culture and Spectral Counting for the Quantification of Differentiating Human Embryonic Stem Cells

5.1 Introduction

Mass spectrometry-based proteomics is an increasingly robust and rapidly developing area of research which must adapt and develop novel analytical strategies to address the dramatic variation across dissimilar biological systems[1]. After identification, quantitation of proteins remains one of the greatest challenges to proteomic analysis. Differential analysis of two-dimensional electrophoretic gels continues to be a widely-used tool for protein separation and quantification[2-6]. However, despite continuous improvements, challenges still exist for the analysis and quantification of proteins that reside at the limits of the gel's analytical capability, whether very high or low molecular weight, extremes in protein isoelectric points, membrane proteins, or proteins of low abundance.

A widely employed strategy for protein quantification has been the chemical or metabolic incorporation of stable isotope labels. One of the chief advantages to this approach is that stable isotope-labeled species possess similar physical and chemical properties as their unlabeled analogues, but with a different mass that can be differentiated and quantified via mass spectrometry. This allows direct comparison of multiple experimental conditions within the same analytical experiment, thus reducing run to run variability. Numerous strategies have been developed for quantification of proteins and peptides using chemical derivatization.

Gygi and coworkers developed isotope-coded affinity tags (ICAT) containing a thiol specific reactive group, an isotope labeled linker region containing eight deuteriums (or hydrogens) and a biotin affinity tag[7, 8]. One of the first applications of ICAT labeling to a complex protein mixture indicated 10 % relative standard deviation of peptide area ratios, although retention time differences between protium and hydrogen labeled species, approaching 30 seconds, have the potential for greater error. [9-11] Isobaric tags for relative and absolute quantitation (iTRAQ) is amenable to the direct comparison of up to eight experimental samples simultaneously by labeling the N-termini of peptides with a compound containing a reporter group with varying stable isotope (^{13}C, ^{15}N, and ^{18}O) incorporation, a balancing group, and an amine specific reactive group. Tandem mass tags (TMTs) also employ an isobaric chemical labeling strategy for relative quantification. All versions of the tag have the same total mass, with quantification achieved during MS/MS, in which the reporter groups with varying mass are generated with abundances correlating to the relative expression of the peptide in each experimental sample[9, 12]. Enzymatically catalyzed ^{18}O-labeling has also been employed, incorporating either ^{16}O or ^{18}O into peptide C-termini during proteolysis giving rise to peptide pairs that can be relatively quantified.[13-18]

Metabolic processes can also be employed for the incorporation of stable isotope labels by enriching cell or tissue culture media with light and heavy versions of the label of choice. This strategy was introduced for bacterial proteomics by Oda *et al*.[19]. *E. coli* and *S. cerevisiae* have been labeled with deuterated leucine but

proved to be less accurate than other labeling methods due to the effect of the deuterium label on retention time, as discussed *vide* supra.[10, 11, 20, 21] $^{15}N/^{14}N$ metabolic labeling has been successfully employed for the quantitative proteomic analysis of yeast and has also been successfully utilized for quantification in plant species[22]. Mann and coworkers introduced stable isotope labeling by amino acids in cell culture (SILAC) as a metabolic labeling strategy whereby amino acids containing stable isotopes are supplemented to cell culture growth media to produce peptide analytes that chromatographically co-elute[23, 24]. More recently this strategy has been developed for the *in vivo* labeling of mice[25].

In addition to the stable isotope based methods for protein quantification, label-free strategies continue to be developed. One strategy utilizes integrated ion abundance obtained from extracted ion chromatograms to track peptide and protein abundance across experimental samples. Bondarenko and coworkers demonstrated the linear relationship of peptide peak areas over a range of 10 to 1,000 fmol of myoglobin digest spiked into a human plasma sample, with a relative standard deviation near 10%.[26, 27] Another method for label-free quantification is spectral counting, where the number of mass spectra that contribute to the identification of a given protein are used as an indication of that protein's abundance[28]. Standard peptides spiked into yeast lysate showed a linear relationship between protein abundance and spectral counts over two orders of magnitude. Also, two different proteins can be relatively quantified by normalizing the spectral counts for each protein to its number of amino acids.[29] Overall,

integrated ion abundance and spectral counting are comparable, with spectral counting reporting to show more sensitivity to changes in protein abundance, and peak area quantification showing slightly higher accuracy in determining relative protein ratios[30]. A recent \ study by Stevenson et al. compared the two strategies and found the statistical processing associated with spectral counting to be more amenable to high throughput quantification relative to itegrated ion abundance[31].

The quantitative proteomics strategy employed in any given experiment is dependent on a number of factors. In many cases, biological systems of interest may not be capable of metabolic labeling, and chemical isotopic labeling may be cost and/or time prohibitive. Proteotrophic organisms such as fungi are an example of biological systems that are less amenable to SILAC, as their more flexible metabolisms allow them to produce their own amino acids from raw nutrients, regardless of the amount of label supplemented.[32, 33] Also, clinical samples and most animal based samples cannot employ metabolic label-based quantification strategies. Studies of this sort stand to directly benefit from label-free quantification strategies and a more thorough understanding of how label-free techniques compare to stable-isotope facilitated methodologies is warranted.

Few examples exist in the literature for the direct comparison of stable isotope labeling and label-free quantification. Hendrickson *et al.* reported on the comparison of $^{15}N/^{14}N$ with spectral counting in their study of two microbial species and suggest a lower sensitivity to the detection of a two-fold expression change with spectral counting[34]. In a comparison of iTRAQ and label-free quantification using

spectral counting, spectral counting was found to identify many more proteins with greater sequence coverage when compared to iTRAQ, with both techniques possessing similar quantification capabilities[35]. Recently, Parker *et al.* used spectral-counting to analyze SILAC data sets as a means of demonstrating the ability to use a SILAC experimental design to combine different experimental states, but not require high mass accuracy mass spectrometers for data collection, using the spectral counts for the separate quantification of light and heavy peptides.[36] Previously, we reported similar strategy to evaluate and directly compare the performance of SILAC and label-free quantification using spectral counting on a 1:1 mixture of SILAC labeled human embryonic stem cells. We established a similar capacity for protein quantification on that data set, and although spectral counting was ostensibly able to quantify more proteins, the need for at least 5 spectral counts to accurately quantify proteins hindered its performance relative to SILAC. However, we also noted that a typical label-free experiment would allow for twice the amount of protein to be analyzed per experimental condition.[37] Herein, we describe the comparison of SILAC and spectral counting for the relative quantification of separate data sets collected from human embryonic stem cells directed to differentiate toward the trophectoderm tissue type. We observe the number of peptides and proteins identified and quantified over three different time points.

5.2 Experimental

5.2.1 Cell Culture

H9 human embryonic stem cells were maintained on mouse embryonic fibroblast (MEF) conditioned medium according to a protocol adapted from Xu et. al.[38] Briefly, MEF cells were inactivated by mitomycin c (Sigma Aldrich, St. Louis, MO, USA) treatment and plated at 55,000 cells/cm^2. HESC medium consisting of DMEM/F12 (Invitrogen, Carlsbad, CA, USA), 20% knockout serum replacement (Invitrogen), 1 mM L-glutamine (Invitrogen), 0.1 mM β-mercaptoethanol, 1% nonessential amino acids (Invitrogen), and 4 ng/mL basic fibroblast growth factor (bFGF; Sigma Aldrich) was conditioned for 24 hours on MEF cells. The conditioned media was collected for 7 days, pooled, and supplemented with an additional 4 ng/mL bFGF (Sigma Aldrich). Conditioned media for SILAC labeling of hESCs was prepared using protocols previously described. Briefly, SILAC medium consisted of DMEM/F12 without L-lysine and L-arginine, $^{13}C_6$-L-arginine (Pierce, Rockford, IL, USA), $^{13}C_6$, $^{15}N_2$-L-lysine (Pierce), 20% knockout serum replacement (not containing lysine or arginine), 1 mM L-glutamine, 0.1mM β-mercaptoethanol, 1% nonessential amino acids and 4 ng/mL bFGF, which was conditioned over MEF cells. Stable isotope labeled arginine and lysine incorporation of 98 % and 98.5%, respectively, was achieved.

5.2.2 Trophectoderm Differentiation

H9 cells on conditioned medium were treated with SB431542 (Sigma Aldrich) at 25μM concentration for 6 days, starting two days after passage. Flow cytometry and sub-cellular fractionation was carried out as described below. Cells were further passaged into conditioned medium containing 25μM SB431542 for another 6 days. Flow cytometry and sub-cellular fractionation was carried out again.

5.2.3 Flow Cytometry

Cells were dissociated with 0.25% Trypsin/EDTA (Sigma Aldrich). Trypsin was subsequently inactivated by adding Fetal Bovine Serum (Sigma Aldrich). Cells were fixed with 2% para-formaldehyde (Fisher Scientific, Houston, TX, USA) for 30 minutes on ice. Cells were permeabilized for 15 minutes at room temperature using Saponin Permeabilization Buffer (SPB): 1 mg/ml Saponin (Sigma Aldrich), 1% BSA (Sigma Aldrich) and Dulbecco's PBS without Ca^{2+} or Mg^{2+} (Sigma Aldrich). Cells were incubated with the OCT4 antibody (Cell Signaling, Danvers, MA, USA) for 1 hr at room temperature at a 1:50 dilution in SPB. Cells were centrifuged and incubated with FITC conjugated Goat anti-Rabbit IgG (Sigma Aldrich) for 1 hr on ice at a dilution of 1:500 in SPB. Finally, cells were centrifuged and resuspended in FACS buffer: 0.1% BSA (J. T Baker, Phillipsburg, NJ, USA) in PBS, and analyzed.

5.2.4 Subcellular Fractionation

The Sucrose Buffer 1 (SB1) comprised of 250mM Sucrose (Fisher Scientific), 25mM potassium chloride (Fisher Scientific), 5mM magnesium chloride (Fisher Scientific), 10mM triethanolamine (Sigma Aldrich), 10mM acetic acid (ACROS, NJ, USA), 2.5mM sodium pyrophosphate (Fisher Scientific), 1mM β-glycerophosphate disodium salt (MP Biomedicals, Solon, OH, USA), 1mM sodium orthovanadate (Fisher Scientific), c∅mplete mini® protease inhibitor cocktail tablets (Roche, Indianapolis, IN, USA) and phosphatase inhibitor cocktails I and II (Sigma Aldrich). The pH was adjusted to 7.6 using either triethanolamine and/or Acetic acid. The Sucrose Buffer 2 (SB2) comprised of 1.42M Sucrose and the rest as above. The Sucrose Buffer 3 (SB3) comprised of 2 M Sucrose and the rest as above.

3ml of ice-cold SB1 was added to a 10cm dish and cells were scraped off using a cell scraper. Cells were homogenized with eight vigorous strokes of a tight pestle in a pre-cooled Dounce Homogenizer. The lysate was centrifuged for 10min at 1000g at 4°C in a swinging bucket rotor, using Eppendorf 5804R (Eppendorf, Westbury, NY, USA). The supernatant was collected in an ultracentrifuge tube and the pellet was resuspended in 2ml of SB1 and centrifuged again for 10min at 1000g at 4°C. This supernatant was discarded and the pellet was resuspended in 3 ml SB2 slowly to avoid clump formation. The resuspended fraction was homogenized again with eight vigorous strokes of a tight pestle in a pre-cooled Dounce Homogenizer, to mechanically separate nuclei from membrane sheets. This homogenate was layered carefully over a 300μl cushion of SB3 in an ultracentrifuge tube, and was covered

with a 300μl cushion of SB1, forming a discontinuous sucrose gradient. The tube was then ultra-centrifuged, along-side the supernatant collected previously from the first centrifugation, at 40,000 rpm for 80min at 4°C in an MLS-50 swinging bucket rotor in a TL-100 Ultracentrifuge (Beckman Coulter, Brea, CA, USA). The cytoplasmic fraction was collected from the supernatant tube as the supernatant and the pellet was discarded. The membrane fraction was recovered from the SB1-SB2 interface of the sucrose gradient tube, mixed with 1ml SB1 and centrifuged for 10min at 3000g at 4°C in a swinging bucket rotor, using Eppendorf 5804R. The supernatant was discarded and the pellet retained as membrane fraction. The pellet below the S3 cushion in the ultracentrifuge tube was retained as the nuclear fraction.

5.2.5 Sample Solubilization

The nuclear and membrane fractions were solubilized in 8 M urea (Sigma Aldrich) and 50 mM ammonium bicarbonate (Fisher Scientific) with c∅mplete mini® protease inhibitor cocktail tablets (Roche, Indianapolis, IN, USA) and phosphatase inhibitor cocktails I and II (Sigma Aldrich). The lysate for the nuclear sample was passed vigorously through a 20G needle to shear the chromatin, and reduce the viscosity appreciably. Total protein amounts were determined via Bradford assay. The nuclear, membrane and cytoplasmic fractions were frozen at -80 °C until needed.

5.2.6 Protein Fractionation and In-Gel Digestion

For SILAC quantification, mixtures were prepared by combining 25 µg of unlabeled cytoplasmic proteins from either differentiation time point (6 or 12 days) with an equal amount of the stable isotope labeled hESC control for a total of 50 µg. For quantification using spectral counting, 50 µg of either the hESC control, 6 day differentiated, or 12 day differentiated cytoplasmic protein samples were run in their own gel lanes. Samples were combined with an equal volume of 2X Laemmli buffer containing β-mercaptoethanol(BioRad, Hercules, CA), boiled for 5 minutes and cooled on ice. Samples were then loaded and run on a Criterion 10-20% Tris-HCl Gel (BioRad) at a constant 200 V and visualized using BioSafe Coomassie Stain (BioRad). Each gel lane was cut into 12 slices which were subsequently reduced with dithiothreitol (DTT), alkylated with iodoacetamide, and digested with proteomics grade trypsin according to a protocol adapted from Shevchenko et al.[39]

5.2.7 LC-MS/MS Analysis

In-gel digested peptide samples were separated using an Eksigent 1D+ nano-LC system (Eksigent, Dublin, CA) utilizing a vented column configuration.[40] The stationary phase employed was Magic C18AQ (5 µm particle, 200 Å pore size)(Microm BioResources) and was packed to a length of 5 cm in a 75 µm i.d. IntegraFrit capillary (New Objective, Woburn, MA) for the trap column, and to 15 cm in a 75 µm i.d. PicoFrit capillary (New Objective) with a 15 µm electrospray emitter

for the analytical column. All LC solvents were purchased from Burdick and Jackson (Muskegon, MI). Mobile phase A consisted of 98% water, 2% acetonitrile, and 0.2% formic acid. Mobile phase B was composed of 2% water, 98% acetonitrile, and 0.2% formic acid. 8 µL of sample was injected and desalted at 1.5 µL/min on the trap before switching in line with the gradient at a flow rate of 350 nL/min. The gradient held at 2% B for 5 minutes before ramping to 10% B over 2 minutes, followed by an increase to 50% B over the next 120 minutes. The gradient then ramped to 95% B in 3 minutes, held for 5 minutes, then re-equilibrated at 2% B for 10 minutes.

Peptides were detected on an LTQ-Orbitrap XL with a pulse sequence containing 9 scan events. A broadband scan was performed in the Orbitrap analyzer with 60,000 resolving power at *m/z*=400 followed by eight data-dependent MS/MS events in the ion trap employing collision induced dissociation with an isolation width of 2 *m/z* and normalized collision energy of 35% for 30 ms. Monoisotopic precursor selection was enabled and 1+ charge states were rejected from MS/MS. Dynamic exclusion was employed for 180 seconds. Automatic gain control was employed and set to 8×10^3 and 1×10^6 ions in the LTQ and the Orbitrap respectively. The maximum allowed ionization time was 80 ms for the LTQ and 500 ms for the Orbitrap.[41]

5.2.8 Data Analysis

.RAW LC-MS/MS files were processed using MASCOT Distiller (Matrix Science, Boston, MA, USA) to generate peak lists in .mgf format for database searching using MASCOT[42] server (Matrix Science). For SILAC quantification, the data files were searched with carbamidomethylation of cysteine as a fixed modification and with variable modifications including glutamine and asparagine deamidation, methionine oxidation, $^{13}C_6$-arginine, and $^{13}C_6$, $^{15}N_2$-lysine. For label-free quantification using spectral counting, the SILAC labeled hESC control was searched with $^{13}C_6$-arginine, and $^{13}C_6$, $^{15}N_2$-lysine as fixed modifications in addition to the other fixed and variable modifications stated above. Searches corresponding to the samples of differentiation time points did not have to take into account the SILAC modifications. All searches were performed against the NCBI human database containing both target and reverse protein sequences for determination of protein false discovery rate with a precursor ion tolerance of ±5 ppm and a fragment ion tolerance of ±0.6 Da. Label-free quantification using spectral counts was performed using ProteoIQ (BioInquire, Athens, GA, USA) at a 1% false discovery rate (FDR). SILAC quantification was performed using the quantitation module in MASCOT Distiller (Matrix Science, Boston, MA).

5.3 Results and Discussion

Stable isotope labeling by amino acids in cell culture (SILAC) was compared to label-free quantification using spectral counting for the relative quantification of

cytoplasmic proteins expressed during the differentiation of human embryonic stem cells toward trophectoderm over the course of 0, 6 and 12 day time points. Analysis of SILAC samples was performed with single technical replicate analysis of two SILAC mixtures run on a one-dimensional gel- one consisting of protein extracted on day 6 grown on normal media mixed with the 0 day hESC control which was SILAC labeled with $^{13}C_6$-arginine and $^{13}C_6$, $^{15}N_2$-lysine, and another consisting of day 12 grown on normal media mixed with the SILAC labeled 0 day control. The overall experimental strategy is illustrated in **Figure 5.1**

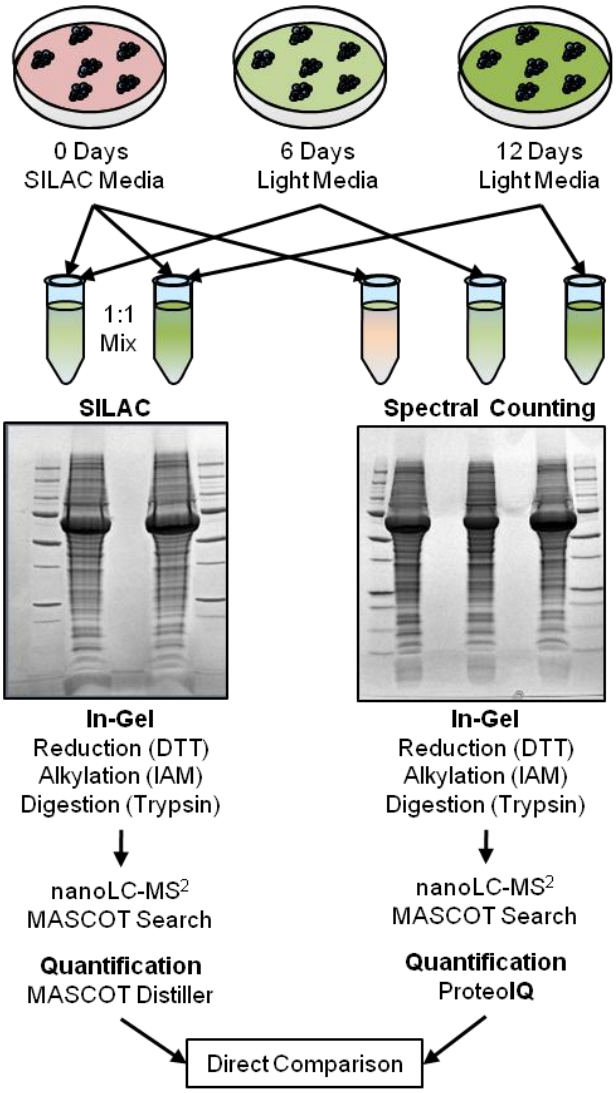

Figure 5.1 Overall experimentall strategy for the direct comparison of SILAC and Spectral Counting for the relative quantification of proteins from differentiating hESCs

The results from both the identification and quantification of proteins using both strategies are summarized in **Table 5.1**. The number of proteins identified in each time point is relatively consistent across time points with quantification

strategies, with the number of proteins identified in all three time points being the lowest number as it means that the protein is present in each observed sample.

Table 5.1 Overall results of quantification using spectral counting and SILAC.

		-Spectral Counting-		-SILAC-
		> 0 SpC	≥ 5 SpC	All
Protein Groups Identified (Quantified)	0 Days	2351 (N/A)	2028 (N/A)	-
	6 Days	2637 (2383)	2204 (1738)	1112 (626)
	12 Days	2509 (2275)	1969 (1653)	1292 (691)
	All Time Points	2180 (2180)	1551 (1551)	1174 (380)
Number of Peptides	All Time Points	185,471		27,672
Number of Unique Peptides	All Time Points	24,554		9,687
Avg. Sequence Coverage	All Time Points	31.59%		24.66%

The number of quantified proteins also steadily decreases as time increases. This is a reasonable conclusion as the proteins are quantified relative to the 0 day control, and fewer proteins are predicted to be in common between time points that are further spaced. As expected, the number of proteins identified using a spectral counting methodology versus SILAC is greater. This is most likely a direct result of the amount of protein loaded from each experimental condition. In spectral counting, protein from each time point was loaded individually. When utilizing SILAC quantification, the same amount of total protein is analyzed in a run, but as an equal mixture of two experimental conditions. While SILAC allows for intra-run quantification, it presents a more complex sample to the mass spectrometer and taxes the duty-cycle of the instrument more so than the alternative strategy, with the caveat that greater care must be taken to ensure run-to-run reproducibility and multiple technical replicates to ensure accurate quantification. The advantage of

increased protein load is further illustrated in the number of peptides identified by each quantification strategy. The three technical replicates of the spectral counting approach identified 185,471 peptides, or roughly 62,000 per technical replicate, which is more twice the 27,672 peptides identified in a single SILAC technical replicate. This is further reflected in the number of unique protein sequences – 24,551 to 9,687, respectively. The result is not only more proteins identified using the spectral counting methodology, but also an increase in the average sequence coverage from 24.66% to 31.59%.

The correlation of the two measurement techniques was explored by determining proteins that were identified in both quantification strategies and in all time points. All 380 of the proteins identified in all time points by the SILAC strategy were also identified and quantified with more than 5 SpC using spectral counting. Their correlation is plotted in **Figure 5.2**. In our previous analysis of a 1:1 control data set, a correlation plot yielded little insight into the nature of how the two quantitative techniques were related. Since all proteins were predicted to be expressed equally, all data points were clustered about the origin of the plot (within a twofold change of expression) with the

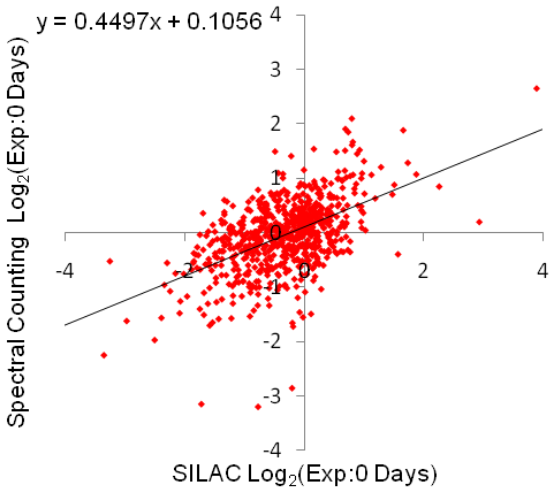

Figure 5.2 Correlation plot of quantitative data from SILAC and Spectral Counting. The trend line with a slope less < 1 indicates increased sensitivity to change in protein expression compared to spectral counting.

distribution being mainly attributed to analytical variability (data not shown). **Figure 5.2** shows a direct positive correlation between SILAC and label-free quantification using spectral counting with quantitative values spanning up to a 16-fold relative change in either direction. The diffuse distribution of data points along the correlation can be attributed to analytical and run-to-run variability of each quantification technique. Establishing a linear regression through the correlation plot allows insight into the sensitivity of each quantitation strategy to changes in protein expression. The reported slope of 0.44 indicates label-free quantification using spectral counting is less sensitive than SILAC. This can be explained by the nature of how proteins are quantified in each method. Quantification using spectral counting is determined by the number of mass spectra identifying a given protein. In our particular experimental configuration, that means that a limited number of peptides from a protein are continuously competing with other peptides within a complex mixture for identification by MS/MS from a limited number of isolation and fragmentation events. For SILAC quantification, a peptide for a protein need only be isolated and fragmented a single time to achieve identification, with measurements contributing to quantification coming from the abundance of the precursor ion in each broadband spectrum obtained while that peptide elutes. Thus, while label-free quantification is shown to have the capacity to identify and quantify more proteins due to its increased loading capacity and lower sample complexity, sensitivity to changes in expression is sacrificed.

The decreased sensitivity to changes in relative expression is further demonstrated in **Figure 5.3**, in which the quantification of three proteins with varying

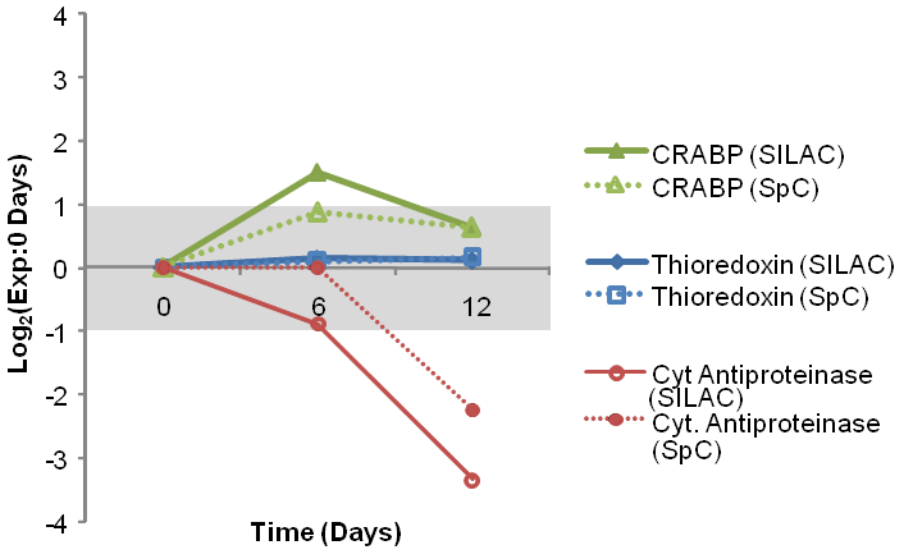

Figure 5.3 Comparison of Spectral Counting (SpC) and SILAC for three cytoplasmic proteins exhibiting different levels of change in relative expression during hESC differentiation. The grey box represents a 2-fold change in expression generally regarded as the threshold for biological significance.

levels of quantitative change are plotted using both quantitative strategies. Thioredoxin exhibits no significant change in expression throughout the differentiation of human embryonic stem cells to the trophectoderm, and the quantitative values from both SILAC and spectral counting reflect this, with the lines corresponding to each method being barely distinguishable. Cellular retinoic acid binding protein 2 (CRABP) shows moderate changes in its relative expression. At day 6, the SILAC method indicates a Log_2(6 day:0 day) value of 1.50, well beyond the threshold generally regarded as biological change(Log_2=1), before decreasing to 0.65 at day 12. However, the same measurement performed with spectral counting does not exceed the threshold for significant change, with values at day 6 and 12 of

0.88 and 0.64, respectively. The quantification of cytoplasmic antiproteinase 3 also demonstrates that even with an increased magnitude of change in protein relative expression SILAC quantification shows an increased response.

5.4 Conclusions

We have examined the performance of two quantitative proteomics strategies on their abilities to identify and quantify proteins expressed over three different time points during the differentiation of human embryonic stem cells toward a trophectoderm tissue type. As hypothesized, the increased sample load capacity coupled with the relatively less complex mixture presented to the mass analyzer during the label free quantification strategy allows for the identification and quantification of more proteins with greater sequence coverage relative to the SILAC technique. The two quantitative strategies also positively correlate, reporting similar quantitative values for identical proteins at the same time points over an approximately 16-fold upward or downward change of expression, well beyond the two-fold threshold generally regarded for biologically significant change. We have also described a decreased sensitivity to quantitative change in spectral counting versus SILAC, and should be taken into account by investigators considering .spectral counting as an alternative strategy to protein quantification in systems that are not amenable to metabolic labeling.

5.5 References

1. Kenyon, G. L.; DeMarini, D. M.; Fuchs, E.; Galas, D. J.; Kirsch, J. F.; Leyh, T. S.; Moos, W. H.; Petsko, G. A.; Ringe, D.; Rubin, G. M.; Sheahan, L. C., Defining the mandate of proteomics in the post-genomics era: Workshop report. *Mol. Cell. Proteomics* **2002**, 1, (10), 763-780.

2. Link, A.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvic, B. M.; Yates, J. R., Direct analysis of protein complexes using mass spectrometry. *Nat. Biotechnol.* **1999**, 17, (7), 676-682.

3. Shevchenko, A.; Jensen, O. N.; Podtelejnikov, A. V.; Sagliocco, F.; Wilm, M.; Vorm, O.; Mortensen, P.; Shevchenko, A.; Boucherie, H.; Mann, M., Linking genome and proteome by mass spectrometry: Large-scale identification of yeast proteins from two dimensional gels. *Proc. Nat. Acad. Sci. U. S. A.* **1996**, 93, (25), 14440-14445.

4. Gygi, S. P.; Rochon, Y.; Franza, B. R.; Aebersold, R., Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* **1999**, 19, (3), 1720-1730.

5. Garrels, J. I.; McLaughlin, C. S.; Warner, J. R.; Futcher, B.; Latter, G. I.; Kobayashi, R.; Schwender, B.; Volpe, T.; Anderson, D. S.; Mesquita-Fuentes, R.; Payne, W. E., Proteome studies of Saccharomyces cerevisiae: Identification and characterization of abundant proteins. *Electrophoresis* **1997**, 18, (8), 1347.

6. Boucherie, H.; Sagliocco, F.; Maillet, I.; Labarre, J.; Perrot, M., Two-dimentional gel protein database of Saccharomyces cerevesiae. *Electrophoresis* **1996**, 17, (11), 1683-99.

7. 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), 994-999.

8. Li, J. X.; Steen, H.; Gygi, S. P., Protein profiling with cleavable isotope-coded affinity tag (cICAT) reagents - The yeast salinity stress response. *Mol. Cell. Proteomics* **2003**, 2, (11), 1198-1204.

9. Ross, P. L.; Huang, Y. L. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Prkayastha, S.; Jahasz, P.; Martin, S.; Bartlet-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J., Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* **2004**, 3, (12), 1154-1169.

10. Zhang, R. J.; Sioma, C. S.; Wang, S. H.; Regnier, F. E., Fractionation of isotopically labeled peptides in quantitative proteomics. *Anal. Chem.* **2001**, 73, (21), 5142-5149.

11. Julka, S.; Regnier, F., Quantification in proteomics through stable isotope coding: A review. *J. Proteome Res.* **2004**, 3, (3), 350-363.

12. Thompson, A.; Schafer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Johnstone, R. A. W.; Mohammed, A. K. A.; Hamon, C., Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* **2003**, 75, (8), 1895-904.

13. Desiderio, K. M.; Kai, M., Preparation of Stable Isotope-Incorporated Peptide Internal Standards for Field Desorption Mass-Spectrometry Quantification of Peptides in Biologic Tissue. *Biomed. Mass Spectrom.* **1983**, 10, (8), 471-479.

14. Mirgorodskaya, O. A.; Kozmin, Y. P.; Titov, M. I.; Korner, R.; Sonksen, C. P.; Roepstorff, P., Quantitation of Peptides and Proteins by Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry Using O-18-Labeled Internal Standards. *Rapid Commun. Mass Spectrom.* **2000**, 14, (14), 1226-1232.

15. Yao, X. D.; Afonso, C.; Fenselau, C., Dissection of proteolytic O-18 labeling: Endoprotease-catalyzed O-16-to-O-18 exchange of truncated peptide substrates. *J. Proteome Res.* **2003**, 2, (2), 147-152.

16. Yao, X. D.; Freas, A.; Ramirez, J.; Demirev, P. A.; Fenselau, C., Proteolytic O-18 Labeling for Comparative Proteomics: Model Studies with Two Serotypes of Adenovirus. *Anal. Chem.* **2001**, 73, (13), 2836-2842.

17. Mason, C. J.; Therneau, T. M.; Eckel-Passow, J. E.; Johnson, K. L.; Oberg, A. L.; Olsen, J. E.; Nair, K. S.; Muddiman, D. C.; Bergen, H. R., A method for automatically interpreting mass spectra of 18O-labeled isotopic clusters. *Mol. Cell. Proteomics* **2007**, 6, (2), 305-318.

18. Johnson, K. L.; Muddiman, D. C., A method for calculating O-16/O-18 peptide ion ratios for the relative quantification of proteomes. *J. Am. Soc. Mass Spectrom* **2004**, 15, (4), 437-445.

19. Oda, Y.; Huang, K.; Cross, F. R.; Cowburn, D.; Chait, B. T., Accurate quantitation of protein expression and site-specific phosphorylation *Proc. Nat. Acad. Sci. U. S. A.* **1999**, 96, (12), 6591-6596.

20. Martinovic, S.; Veenstra, T. D.; Anderson, G. A.; Pasa-Tolic´, L.; Smith, R. D., Selective incorporation of isotopically labeled amino acids for identification of intact proteins on a proteome-wide level. *J. Mass Spectrom.* **2002**, 37, (1), 99-107.

21. Veenstra, T. D.; Martinovic, S.; Anderson, G. A.; Pasa-Tolic´, L.; Smith, R. D., Proteome analysis using selective incorporation of isotopically labeled amino acids. *J. Am. Soc. Mass Spectrom.* **2000**, 11, (1), 78-82.

22. Schaff, J. E.; Mbeunkui, F.; Blackburn, K.; Bird, D. M.; Goshe, M. B., SILIP: a novel stable isotope labeling method for in planta quantitative proteomic analysis. *The Plant Journal*` **2008**, 56, 840-854.

23. Ong, S.; 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), 376-386.

24. Ong, S.; Kratchmarova, I.; Mann, M., Properties of 13C-Substituted Arginine in Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC). *J. Proteome Res.* **2003**, 2, 173-181.

25. Kruger, M.; Moser, M.; Ussar, S.; Thievessen, I.; Luber, C. A.; Forner, F.; Schmidt, S.; Zanivan, S.; Fassler, R.; Mann, M., SILAC Mouse for Quantitative Proteomics Uncovers Kindlin-3 as an Essential Factor for Red Blood Cell Functioin. *Cell* **2008**, 134, 353-364.

26. Chelius, D.; Bondarenko, P. V., Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *J. Proteome Res.* **2002**, 1, (4), 317-323.

27. Bondarenko, P. V.; Chelius, D.; Shaler, T. A., Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography-tandem mass spectrometry. *Anal. Chem.* **2002**, 74, (18), 4741-4749.

28. Wolters, D.; Washburn, M. P.; Yates, J. R., An automated multidimensional protein identification technology for shotgun proteomics. *Anal. Chem.* **2001**, 73, (23), 5683-5690.

29. Zybailov, B.; Mosley, A. L.; Sardiu, M. E.; Coleman, M. K.; Florens, L.; Washburn, M. P., Statistical analysis of membrane proteome expression changes in Saccharomyces cerevisiae. *J. Proteome Res.* **2006**, 5, (9), 2339-2347.

30. Old, W. M.; Meyer-Arendt, K.; Aveline-Wolf, L.; Pierce, K. G.; Mendoza, A.; Sevinsky, J. R.; Resing, K. A.; Ahn, N. G., Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol. Cell. Proteomics* **2005**, 4, (10), 1487-1502.

31. Stevenson, S. E.; Chu, Y.; Ozias-Atkins, P.; Thelen, J. J., Validation of gel-free, label-free quantitative proteomics approaches: Applications for seed allergen profiling. *J. of Proteomics* **2009**, 72, (3), 555-566.

32. Collier, T. S.; Hawkridge, A. M.; Georgianna, D. R.; Payne, G. A.; Muddiman, D. C., Top-down identification and quantification of stable isotope labeled proteins from Aspergillus flavus using online nano-flow reversed-phase liquid chromatography coupled to a LTQ-FTICR mass spectrometer. *Anal. Chem.* **2008**, 80, (13), 4994-5001.

33. Georgianna, D. R.; Hawkridge, A. M.; Muddiman, D. C.; Payne, G. A., Temperature-dependent regulation of proteins in Aspergillus flavus: Whole organism stable isotope labeling by amino acids
J. Proteome Res. **2008**, 7, (7), 2973-2979.

34. Hendrickson, E. L.; Xia, Q. W.; Wang, T. S.; Leigh, J. A.; Hackett, M., Comparison of spectral counting and metabolic stable isotope labeling for use with quantitative microbial proteomics. *Analyst* **2006**, 131, 1335-1341.

35. Patel, W. J.; Thalassinos, K.; Slade, S. E.; Connolly, J. B.; Crombie, A.; Murrell, J. C.; Scrivens, J. H., A comparison of labeling and label-free mass spectrometry-based proteomics approaches. *J. Proteome Res.* **2009**, 8, 3752-3759.

36. Parker, S. J.; Halligan, B. D.; Greene, A. S., Quantitative analysis of SILAC data sets using spectral counting. *Proteomics* **2010**, 10, (7), 1408-1415.

37. Collier, T. S.; Sarkar, P.; Franck, W. L.; Dean, R. A.; Rao, B. M.; Muddiman, D. C., Direct Comparison of Stable Isotope Labeling by Amino Acids in Cell Culture and Spectral Counting for Quantitative Proteomics. *Anal. Chem.* **2010**, In Press.

38. Xu, C.; Inokuma, M. S.; Denham, J.; Golds, K.; Kundu, P.; Gold, J. D.; Carpenter, M. K., Feeder-Free Growth of Undifferentiated Human Embryonic Stem Cells. *Nat. Biotechnol.* **2001**, 19, (10), 971-974.

39. Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M., Mass spectrometric sequencing of proteins from silver stained polyacrylamide gels. *Anal. Chem.* **1996**, 68, (5), 850-858.

40. Andrews, G. L.; Shuford, C. M.; Burnett, J. C.; Hawkridge, A. M.; Muddiman, D. C., Coupling of a vented column with splitless nanoRPLC-ESI-MS for the improved separation and detection of brain natriuretic peptide-32 and its proteolytic peptides. *J. of Chrom. B* **2009**, 877, (10), 948-954.

41. Hu, Q. Z.; Noll, R. J.; Li, H. Y.; Makarov, A.; Hardman, M.; Cooks, R. G., The Orbitrap: a new mass spectrometer. *J. Mass Spectrom.* **2005**, 40, (4), 430-433.

42. Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S., Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999**, 20, (18), 3551-3567.