ABSTRACT

FRAHM, JENNIFER LEIGH. Understanding the Role of Chemical and Physical Processes as Related to the Quantification of Proteins by ESI-FT-ICR Mass Spectrometry. (Under the direction of David Charles Muddiman).

Since the completion of the Human Genome Project and subsequent development of genomic and proteomic databases, proteomics research has experienced explosive growth. Proteomics is the venture to identify and characterize proteins encoded by genes, including post-translational modifications and alternatively-spliced forms. Electrospray ionization-mass spectrometry has become a key technology for proteomics measurements owing to the afforded sensitivity and molecular specificity of the technique. Moreover, in this dissertation it is shown that the ability of Fourier transform-ion cyclotron resonance mass spectrometry to analyze large numbers of analytes (known as the peak capacity) enables complex mixture analysis. Mass spectrometry-based proteomics measurements to date in large part have focused on the identification of proteins, including post-translational modifications. Increasingly, mass spectrometry based proteomics methods for absolute and relative quantification are emerging. Herein, normalization strategies utilizing synthetic exogenous peptides with masses that do not overlap with endogenous peptides, allowing for the assessment of mass measurement accuracy and retention time reproducibility, are presented. In combination, the internal standards and normalization strategies afford accurate and reproducible protein quantification. Little research has been published to date that explores the fundamental aspects of quantitative proteomics. The work presented in this dissertation examines fundamental aspects, specifically, the role physical (post-excite radius, axial confinement fields, signal decay rate, temperature) and chemical (hydrophobicity) characteristics have on accurate protein quantification by electrospray ionization-Fourier transform-ion cyclotron resonance mass spectrometry. Examination of the instrumental parameters (post-excite radius, excitation voltages, trapping voltages,

excitation waveforms) that lend to confident identification and accurate quantification of proteins in complex mixtures enables future quantitative proteomics measurements. Moreover, the understanding and development of strategies that increase ion abundance (hydrophobicity), and subsequently lower the limit of detection, will facilitate future studies of low abundance proteins.

UNDERSTANDING THE ROLE OF CHEMICAL AND PHYSICAL PROCESSES AS RELATED TO THE QUANTIFICATION OF PROTEINS BY ESI-FT-ICR MASS SPECTROMETRY

by
JENNIFER LEIGH FRAHM

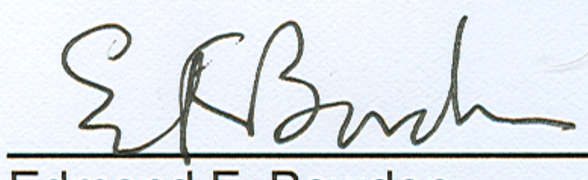
A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

CHEMISTRY

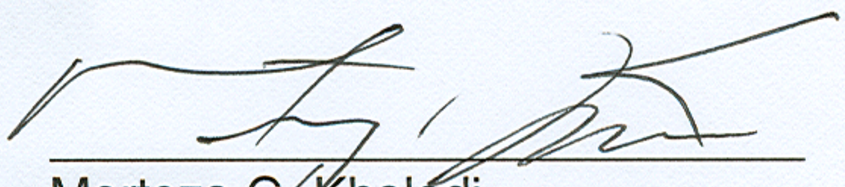
Raleigh, North Carolina

May 10, 2007

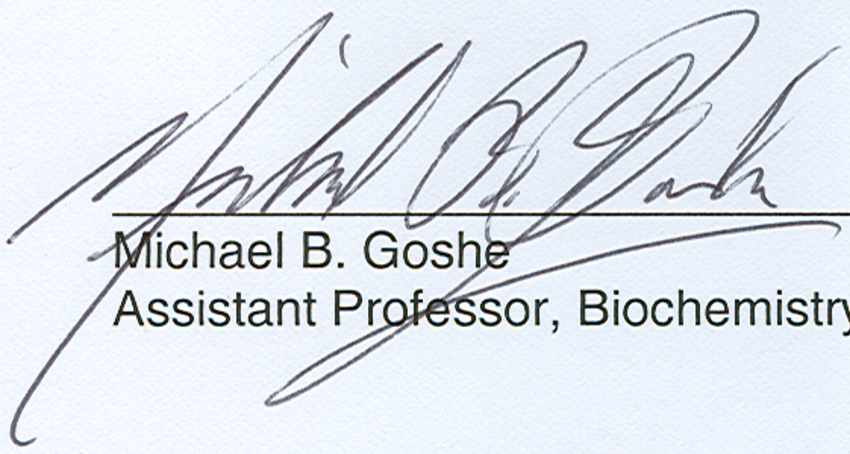
Approved by:



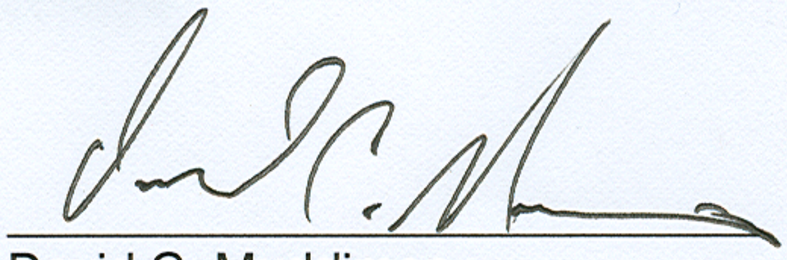
Edmond F. Bowden
Professor, Chemistry



Morteza G. Khaledi
Professor, Chemistry



Michael B. Goshe
Assistant Professor, Biochemistry



David C. Muddiman
Professor, Chemistry
Chair of Advisory Committee

DEDICATION

This dissertation is dedicated to my parents Mr. David and Mrs. Shirley Frahm, and fellow graduate student and friend, Mrs. Jeanne Theis for without their love, encouragement, and understanding I would have never made it this far.

BIOGRAPHY

Jennifer Leigh Frahm was born on April 24, 1979 in Nashville, Tennessee to David and Shirley Frahm. After the birth of her sister, Tiffany Lynn, in 1981 the Frahm family moved to Aurora, Colorado for a short time before relocating to the Naval Air Station in Lemoore, California. Jennifer's other sister, Vanessa Anne, was born in 1983. When Jennifer was in third grade the Frahm family relocated to Waldorf, Maryland. In 1997 she graduated from Westlake High School in Waldorf, Maryland. Her education continued at St. Mary's College of Maryland, where in May 2001 she received her Bachelor of Arts degree in Chemistry. Jennifer began her graduate studies in Analytical Chemistry at Virginia Commonwealth University in the fall of 2001. After a year of studies Jennifer followed her advisor Dr. David Muddiman to the Mayo Clinic College of Medicine, where she finished her Master's research. In May 2004 Jennifer received her Master's of Science in Chemistry from Virginia Commonwealth University, and subsequently began working toward her Ph.D. in Biochemistry and Structural Biology at the Mayo Clinic College of Medicine. Shortly afterward, in the spring of 2005, Jennifer decided to finish her Ph.D. research at North Carolina State University, where her advisor Dr. David Muddiman had accepted a faculty position. Jennifer will begin post-doctoral studies in July 2007 at the University of North Carolina, Chapel Hill, which is funded by a NIH/NIMH National Research Service Award Institution Post-Doctoral Training Fellowship. Her research is centered on the understanding of the effects of nutritional status on the post-translational regulation of glycerol-3-phosphate acyltransferase, a key enzyme in the triacylglycerol synthetic pathway, and therefore is important in obesity and diabetes disease mechanisms.

ACKNOWLEDGEMENTS

The metaphorical graduate school journey towards becoming an independent scientist has a more literal meaning for me, traveling from Virginia to Minnesota and finally to North Carolina for my graduate studies. I owe both my metaphorical and literal graduate school journey to my research advisor, Dr. David Muddiman, who over the years has been an extraordinary mentor and now friend. He has taught me the fundamental skills and knowledge in our field so that I can achieve my goals, but more importantly he has instilled in me the so that I will achieve my goals.

Dr. Adam Hawkridge's help and guidance has been invaluable, as well. He constantly raises his standards for himself, thereby causing me reevaluate my own aspirations. I specifically want to thank him for his suggestion that I consider applying for a postdoctoral fellowship, especially since I never thought that I could compete with other scientists for such an award.

I have been fortunate to work alongside and learn from many other gifted scientists during my graduate career. I would like to acknowledge all former and current members of the Muddiman groups at Virginia Commonwealth University, Mayo Clinic College of Medicine, and North Carolina State University. Special thanks are given to Dr. Allison Null and Mr. Angelito Nepomuceno for teaching me basic laboratory skills and setting the standard during my early years.

I extend thanks to my committee members Dr. Edmond Bowden, Dr. Michael Goshe, Dr. Morteza Khaledi for challenging me with their questions and for their insightful comments and suggestions. I also acknowledge the generous financial support from the Mayo Clinic College of Medicine, North Carolina State University, IonSpec Graduate Student Fellowship in Bioanalytical Chemistry, National Institutes of Health/ National Cancer Institute (R33

CA105295), and the North Carolina Biotechnology Center in conjunction with LEAP Technologies, Inc.

I would also like to acknowledge all the teachers and professors during my educational journey that helped me get to this point today, especially Mr. David Kelsey and Mrs. Susan Gray who helped incite my interest in the sciences. I would like to extend special thanks to Dr. Ruth Ann Armitage for provoking me to pursue research and steering me toward graduate school.

Lastly, I am indebted to all my family and friends for their support. I am ever grateful for my parents Mr. David and Mrs. Shirley Frahm who instilled in me the importance of higher education, and, along with my sisters Ms. Tiffany and Ms. Vanessa Frahm, unwaveringly supported me along the way. Ms. Angela Schlueter has also provided me with tremendous encouragement during these many years that we have known each other, for which I am extremely thankful. I am also thankful for Mrs. Jeanne Theis, who turned grande nonfat lattes into therapeutic sessions making the last few years seem far less tumultuous and who has turned out to be a remarkable colleague and friend.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF REACTION SCHEMES

LIST OF FIGURES

LIST OF PUBLICATIONS

The research presented in this dissertation has resulted in the following publications:

Peer-reviewed manuscripts:

1. **Frahm, J.L.**; Bori, I.D.; Comins, D.L.; Hawkridge, A.M.; Muddiman, D.C. "Achieving Augmented Limits of Detection for Peptides with Hydrophobic Alkyl Tags (ALiPHAT)" *Analytical Chemistry*, accepted as accelerated article.

2. **Frahm, J.L.**; Capo, C.; Muddiman, D.C. "Understanding the Influence of Post-Excite Radius and Axial Confinement on Quantitative Proteomics Measurements using Fourier Transform-Ion Cyclotron Resonance Mass Spectrometry" *Rapid Communications in Mass Spectrometry*, **Vol 21(7)**:1196-1204. 2007.

3. **Frahm, J. L.**; Howard, B. E.; Heber, S.; Muddiman, D. C., Accessible Proteomics Space and Its Implications for Peak Capacity for Zero-, One-, and Two-Dimensional Separations Coupled with TOF and FT-ICR Mass Spectrometry. *J. Mass Spectrometry,* 41(3), 281-288, 2006.

4. **Frahm, J. L**.; Burke, M. J.; Muddiman, D. C., Leveling Response Factors in the Electrospray Ionization Process Using a Heated Capillary Interface. *J. Am. Soc. Mass Spectrom.*, 16, (5), 772-778, 2005.

Abstracts for Presentations

1. **Frahm, J.L.**; Hawkridge, A.M.; Comins, D.L.; Bori, I.D.; Muddiman, D.C. "Augmented Limits of Detection for Peptides with Hydrophobic Alkyl Tags (ALiPHAT)," abstract for **poster**

presentation, *American Society for Mass Spectrometry Annual Conference*, Indianapolis, IN, accepted.

2. **Frahm, J.L.**, Hawkridge, A.M.; Howard, B.E.; Heber, S.; Muddiman, D.C. "Forbidden Zones in Proteomics and Ramifications and Possibilities for Label-Free Proteomics", abstract for **poster presentation**, *American Society for Mass Spectrometry Annual Conference*, Seattle, WA, 2006.

CHAPTER 1

Introduction

1.1 The Alphabet Soup of Inheritance

In this Omics era [1, 2] the transfer of genetic information is ubiquitously accepted, but merely three decades ago Crick's organization of the alphabet of DNA, RNA, and proteins into the central dogma of molecular biology [3] (**Figure 1.1**) was heatedly disputed. The nucleic acids adenosine (A), thymidine (T), guanosine (G), and cytidine (C) comprise the alphabet of DNA when attached to a deoxyribose moiety. Similarly, the alphabet of RNA consists of the same nucleoside (nucleobase attached to a sugar moiety), except thymidine is replaced with uridine and the sugar moiety is a ribose moiety. The DNA and RNA structures are shown in **Figure 1.2**. According to the central dogma of molecular biology, DNA is replicated, and then transcribed into messenger RNA (mRNA). During translation, the protein synthesis machinery reads the mRNA sequence three nucleobases at a time (*i.e.,* codon) starting with the start signal AUG, which codes for methionine. The genetic code, along with their observed frequency in *Homo sapiens* [4], is

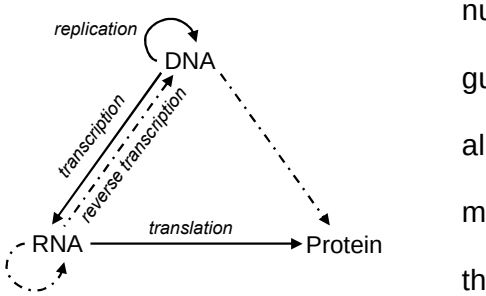


Figure 1.1 Crick's central dogma of molecular biology. Solid arrows indicate general transfer of information that occurs in all cells. Dashed arrows indicate special transfers that occur in some cells. For example, reverse transcription is known to occur in viruses.

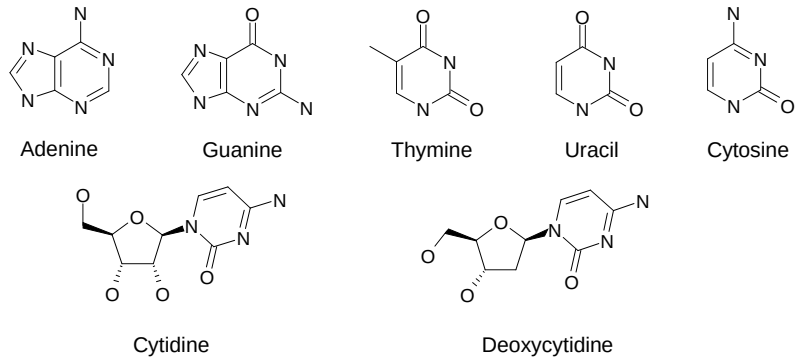


Figure 1.2 Structures of the nucleic acids. Structures for the nucleobases are shown in the top row. In the bottom row are examples of the nucleobase cytosine attached to deoxyribose (as in DNA) and ribose (as in RNA).

shown in **Figure 1.3**.

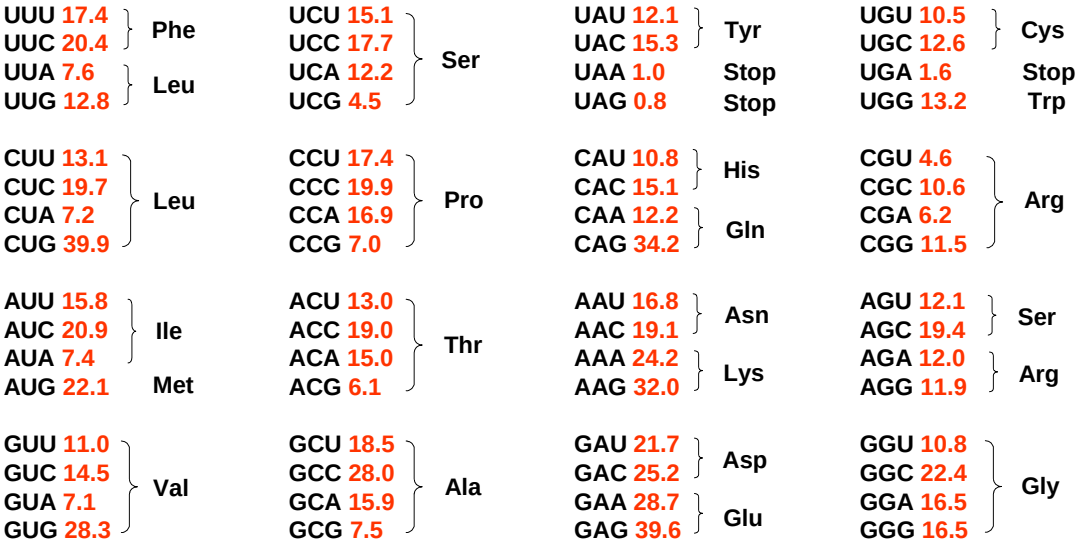


Figure 1.3 Genetic code. The three-nucleotide sequence (codons) and corresponding amino acids are shown. Stop codons are indicated as such. Numbers indicate the frequency per thousand codons in *Homo sapiens.*

1.1.1 Chemical Nature of Amino Acids

Proteins are the functional components of all cells. All proteins are comprised from the same alphabet of amino acids; however, their functions are unarguably very different. For example, the sequence of laminin β1 chain has an unusual 50 amino acid homologous repeat containing 8 cysteines that is repeated 13 times in the β1 chain sequence [5]. These structural elements (disulfide linkages) are necessary to cross link the β1 chain with the other chains (α and γ) giving rise to the cruciform structure of laminin [5] and are important to illicit an immune response [6]. Laminin is a structural component of basement membranes, which support the epithelium [7], and therefore an important protein in tumor biology [8]. A nine amino acid sequence derived from the laminin β1 chain has been shown to inhibit metastasis *in vitro* [8]. In an alternative example, the peptide hormone angiotensin I is a 10 amino acid sequence containing an aspartic acid (negatively-charged) residue at the N-terminus and a lysine (positively-charged) residue at the C-terminus; the polarity of the

termini is key for enzyme recognition/receptor binding [9]. Angiotensin I is an important peptide hormone in the cardiovascular and renal systems [10]. The range of chemical properties exhibited by amino acids allows for such functional diversification. The chemical structures of the 20 amino acids, along with their molecular weights, are shown in **Figure 1.4**.

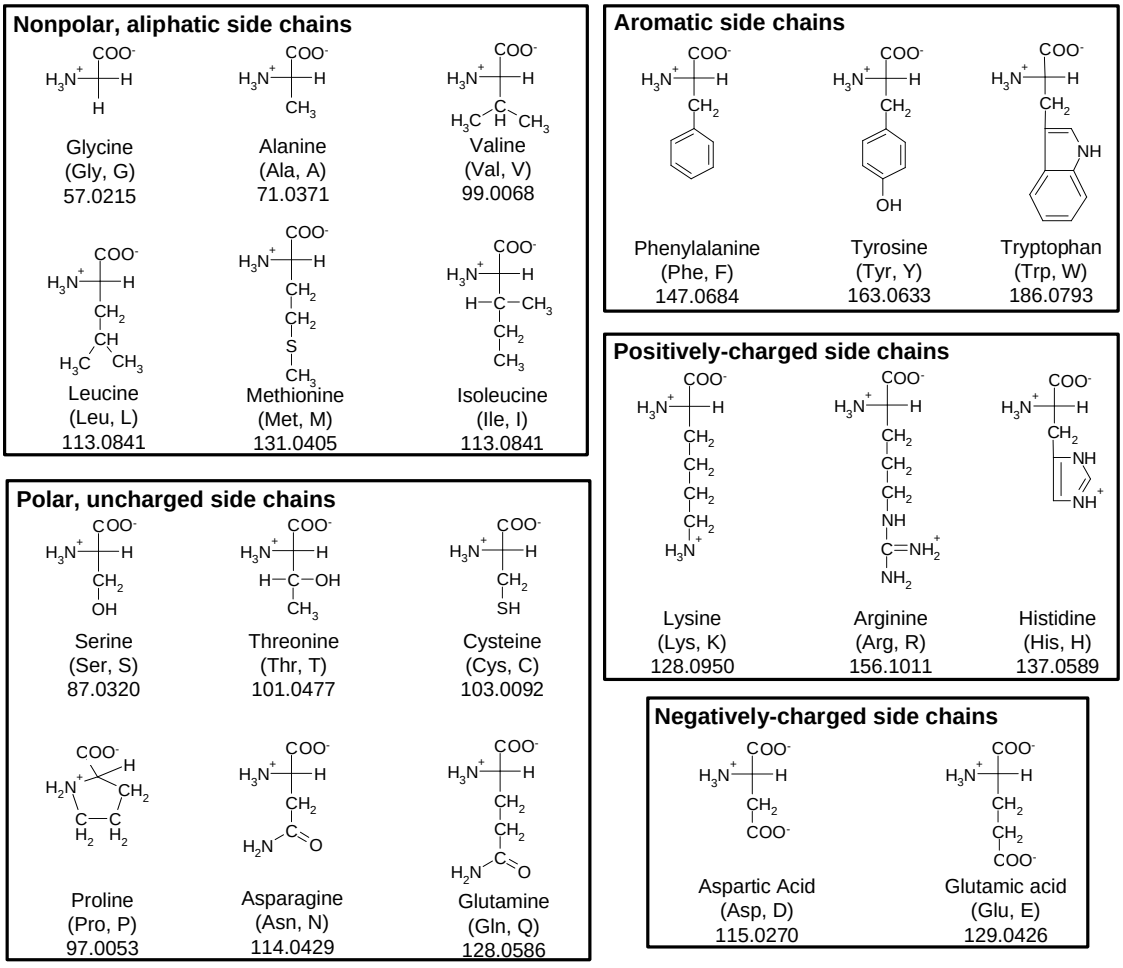


Figure 1.4 Chemical structures of the amino acids

Amino acids are zwitterions and, therefore, have the ability to act as a proton donor or proton acceptor; the acidity is quantified by the pK_a. The pK_as for the C- and N-termini, as well as some of the side chains, of the amino acids are listed in **Table 1.1**. A lower pK_a indicates the molecule is a better proton donor (*i.e.,* stronger acid). Additionally, amino acids have an isoelectric point (pI) because of their amphoteric nature. The isoelectric point

is the point at which the molecule has a net charge of zero; the isoelectric points for the amino acids are shown in **Table 1.1**. Moreover, amino acids can be characterized according to their hydropathy index (**Table 1.1**) [11]. The hydropathy index relates the relative hydrophobic nature of amino acids to one another with the most positive number indicating the most hydrophobic amino acid (isoleucine, 4.5) and the most negative number indicating the most hydrophilic amino acid (arginine, -4.5) [11].

Table 1.1 Chemical properties of the amino acids.

Amino Acid	Monoisotopic Mass[1]	pKa (C-terminus)[2]	pKa (N-terminus)[2]	pKa (Side chain)[2]	pI[2]	Hydropathy (Kyte-Doolittle)[3]	Average % Occurrence[2]
Aspartic Acid (D)	115.0270	1.99	9.90	3.90	2.94	-3.5	5.3
Glutamic Acid (E)	129.0426	2.10	9.47	4.07	3.08	-3.5	6.3
Histidine (H)	137.0589	1.80	9.33	6.04	7.68	-3.2	2.3
Arginine (R)	**156.1011**	**1.82**	**8.99**	**12.48**	**10.74**	**-4.5**	**5.1**
Lysine (K)	**128.0950**	**2.16**	**9.06**	**10.54**	**9.80**	**-3.9**	**5.9**
Serine (S)	87.0320	2.19	9.21		5.70	-0.8	6.8
Threonine (T)	101.0477	2.09	9.10		5.60	-0.7	5.9
Asparagine (N)	114.0429	2.14	8.72		5.43	-3.5	4.3
Glutamine (Q)	128.0586	2.17	9.13		5.65	-3.5	4.3
Tyrosine (Y)	163.0633	2.20	9.21	10.46	6.33	-1.3	3.2
Cysteine (C)	103.0092	1.92	10.70	8.37	5.14	2.5	1.9
Glycine (G)	57.0215	2.35	9.78		6.06	-0.4	7.2
Alanine (A)	71.0371	2.35	9.87		6.11	1.8	7.8
Valine (V)	99.0068	2.29	9.74		6.02	4.2	6.6
Phenylalanine (F)	147.0684	2.33	9.74		6.04	2.8	3.9
Leucine (L)	113.0841	2.32	9.76		6.04	3.8	9.1
Isoleucine (I)	113.0841	2.13	9.28		5.71	4.5	5.3
Proline (P)	97.0053	1.95	10.64		6.30	-1.6	5.2
Tryptophan (W)	186.0793	2.20	9.31		5.76	-0.9	1.4
Mehtionine (M)	131.0405	2.46	9.41		5.93	1.9	2.2

[1] National Institute of Standards and Technology (www.nist.gov)
[2] Voet D., Voet J. G. *Biochemistry*, 2nd ed., 1995, pp 58-59.
[3] Kyte J., Doolittle R.F., *Journal of Molecular Biology*, 1982, 157, 105-132.

1.1.2 Relevance of Proteins to the Understanding and Study of Human Disease

1.1.2.1 Why Study Proteins?

The completion of the Human Genome Project [12, 13] and subsequent availability of genomic and proteomic databases has greatly facilitated proteomics efforts. Proteomics [14], by definition, is the venture to identify, quantify, and structurally and functionally characterize all proteins in a given system (*e.g.,* organism, organelle) and their spatial, temporal,

physiological variations [15]. The analysis of proteins, in addition to genes and mRNA, is important primarily because of the connectivity between proteins and function[16].

Proteomic analysis enables the determination of post-translational modifications, which are covalent modifications added to proteins after translation to modulate a protein's activity; the reader is referred to the following reviews for further information [17-21]. For example, the post-translational modification phosphorylation is a key regulator in a wide range of cellular processes, including cell proliferation, enzyme inactivation/activation, cell signaling, metabolism, immune response, and cell death [22]. Moreover, aberrant phosphorylation has been implicated in a number of disease mechanisms (*e.g.,* cancer [23], Alzheimer's [24], insulin resistance [25]). In eukaryotes, phosphorylation primarily occurs on serine, threonine, and tyrosine residues, and is a dynamic post-translational modification and, therefore, can not be precisely predicted with genomic or proteomic databases (although available databases can predict potential phosphorylation sites from consensus sequences of known phosphorylation sites [26]).

Moreover, proteomics enables three dimensional structural determination of proteins (protein folding), which can not be inferred from mRNA. Although there are databases to predict potential structural motifs from the primary protein sequence, these algorithms simply correlate consensus sequences in the inquiry protein to those of proteins with known structures [27-29]. In reference to human health, structural determination would be important to aid in the understanding of disease mechanisms. For example, the accumulation of misfolded proteins (amyloids) has been implicated in a number of disease mechanisms (amyloidoses) [30]. One of the more commonly known amlyoidoses is Alzheimer's disease, which results from the accumulation of misfolded amyloid beta protein [31].

Lastly, quantitative information of proteins can be obtained from proteomics measurements. In yeast and mammalian cells there is a discrepancy between mRNA abundance and protein expression levels [32, 33]. Additionally, expression levels are

particularly incongruent for low abundance proteins [33]. There are a number of cellular processes that contribute to the discordant expression levels, which are summarized in **Figure 1.5**. Therefore, direct quantification of proteins is necessary in order to accurately determine protein expression levels, which is significant since a number of clinical tests utilize protein levels as prognostic markers of disease [34-38]. One specific example is the measurement of b-type natriuretic peptide (BNP-32), which is used to diagnose congestive heart failure [39]. Collectively, the inherent limitations of genomic and mRNA analysis to predict protein activity (modulated by post-translational modifications), structure, and abundance indicate the necessity of proteomic analyses.

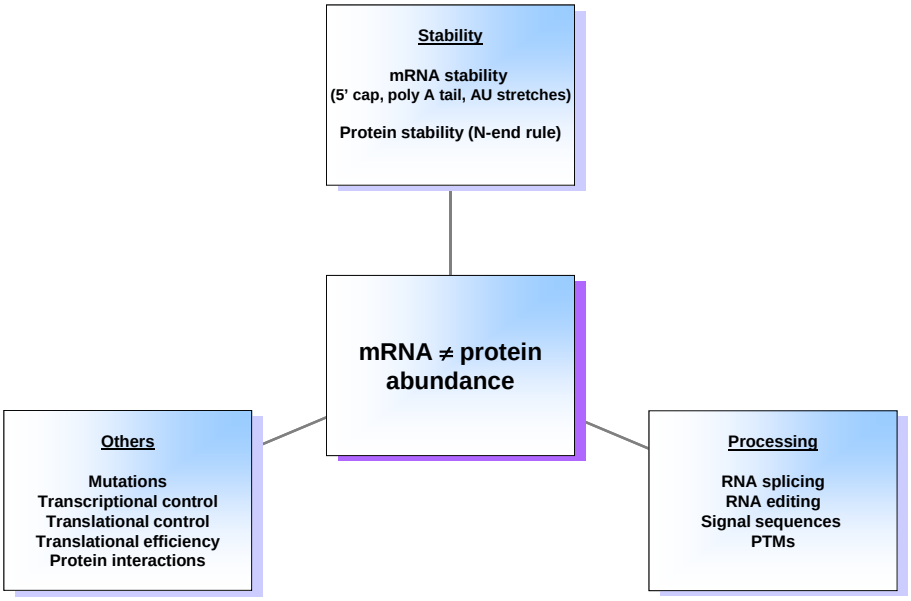


Figure 1.5 Summary of some of the cellular processes that contribute to the poor correlation between mRNA and protein abundance.

1.1.2.2 Role of Mass Spectrometry in Quantitative Proteomics

Early proteomics efforts have focused on the identification and characterization of proteins, as evidenced by the growing numbers of protein databases that include basic information (*e.g.,* primary sequence, molecular weight) and structural information (*e.g.,* tertiary structural motifs, crystallographic structures) [40]. Mass spectrometry has become a key technology for proteomics measurements owing to the afforded sensitivity and

molecular specificity [41, 42]. Additionally, mass spectrometry overcomes many of the limitations of Edman sequencing, a classical protein sequencing method in which the N-terminus of a peptide is reacted with phenylisothiocyanate, subsequently cleaved, and identified based on reversed-phase HPLC retention time of the phenyl thiohydantion derivative [43, 44]. These limitations are primarily due to the inefficiency of the reaction and inability of Edman sequencing to analyze proteins with blocked N-termini or modifications [45].

There are two approaches for protein identification in mass spectrometry proteomics measurements. In top-down proteomics accurate intact masses and fragment ion masses yield protein identifications [46]. Conversely, tryptic peptide masses and subsequent fragment ion masses yield protein identifications in bottom-up proteomics [47]. To date mass spectrometry-based proteomics measurements in large part have been identification studies.

Increasingly, mass spectrometry methods for quantitative proteomics are emerging; reviewed in references [48-51]. Quantitative proteomics measures either absolute (*e.g.,* copy number, concentration determined with a known amount of a standard) or relative (*i.e.,* measured as "fold change" compared to some identified standard) changes in protein concentration. Mass spectrometry is the ideal platform to further the proteomics initiative to the quantitative realm due to its ability to measure an inherent molecular property (specificity), sensitivity, and afforded dynamic range.

1.2 Basic Principles of Electrospray Ionization of Biomolecules

An exponential increase in the success and widespread use of electrospray ionization has occurred within the past two decades. The first studies of electrospray dispersion appeared in the early twentieth century by Zeleny [52, 53]. Zeleny's investigations examined the behavior of a slightly acidic liquid flowing from a pointed-tip capillary with applied positive and negative potentials, and the resulting surface instability at the gas-liquid

interface [52, 53]. Formation of the droplets during this process was later examined and photographed by Magarvey and Outhouse in 1962 [54]. In the late sixties, Dole expanded upon Zeleny's pioneering work by applying the electrospray method described by Hines [55] and Peabody [56], originally used for paint application, to introduce polymers (MW 14, 000 Da) in solution into the gas phase, producing "macroions" [57]. Almost twenty years later, Fenn and coworkers interfaced the ionization process to mass spectrometry for the analysis of large biomolecules and the basis for electrospray ionization, as we know it today, emerged [58, 59]. This indispensable contribution was recognized in 2002 when Fenn was awarded the Nobel Prize in chemistry.

Generally, a high voltage is applied to a capillary as a conducting solution flows through the capillary, causing the formation of a Taylor cone at the capillary tip. The Taylor cone results from the opposing forces between the attraction of the charged liquid to the counter electrode and the surface tension of the liquid. Once the electrostatic repulsion of the charged ions equals the surface tension of the droplet it undergoes a Coulombic explosion, thus producing several smaller droplets containing approximately 15% of the charge and 2% of the mass of the original, larger droplet [60, 61]; this point is known as the Rayleigh limit (*q,* C) and is defined in **Equation 1.1**:

$$q = \sqrt{64\pi^2 \varepsilon_0 \gamma r^3} \qquad (\mathbf{1.1})$$

Equation 1.1 assumes that the droplet is spherical with a radius (*r,* m) and surface tension (γ, N m^{-1}). The gas permittivity constant is denoted by ε_0 (C^2 N^{-1} m^{-2}). Experimental evidence has shown that Coulombic explosions can occur at 70-80% of the Rayleigh limit [60, 61].

The mechanism by which analytes enter the gas phase from the smaller droplets is still under debate [62-65]. There are two proposed mechanisms on how ions may be formed by the electrospray ionization process. The model developed by Dole and coworkers, known as the charged residue model, represents the process as a series of fission events [57, 66].

Solvent evaporation from the droplets leads to an eventual increase in the surface-charge density until the Rayleigh limit (**Equation 1.1**) is reached. Once the Rayleigh limit is achieved, the droplet undergoes a series of Coulombic explosions until a single analyte ion remains with several charges. The analyte molecule enters the gas phase after the remaining solvent evaporates.

In an alternative model, Iribarne and Thomson proposed that after the droplet radii reach a certain size ions (via Coulombic explosions and evaporation) can be directly desorbed from a charged droplet [67]. The driving force for ion emission is the repulsion between charges in the droplet. This model is referred to as the ion evaporation model.

1.3 Fourier Transform-Ion Cyclotron Resonance Mass Spectrometry

After ionization, less than 0.01% of ionized molecules enter the orifice of the mass spectrometer [68]. By a series of voltage changes, the ions are first accumulated in the hexapole and then directed down the analyzer tube into the ICR cell. Once in the ICR cell, charged particles with a directional component of the velocity vector perpendicular to the magnetic field will assume a natural cyclotron motion in a uniform magnetic field [69]. This motion results from the Lorentz force that deflects an ion into a circular orbit when the ion enters the magnetic field moving in a direction perpendicular to the direction of the magnetic field. According to the right hand rule, positive ions will assume a clockwise orbit and negative ions will assume a counterclockwise orbit. The radius of this orbit is the thermal radius, defined in **Equation 1.2**:

$$r_{thermal} = \frac{1}{qB_0}\sqrt{2kmT} \qquad (\mathbf{1.2})$$

The thermal radius of an ion is proportional to the mass (*m,* kg) and temperature (*T,* K), and inversely proportional to the charge (*q,* C) and magnetic field strength (B_0, A/m). Boltzmann's constant is denoted by *k* (m^2 kg s^{-2} K).

The performance of an ICR mass analyzer can be significantly improved upon by changing the magnetic field strength. The highest magnetic field strength commercially available is 15 T; however, few of these systems exist in research facilities. An increase in magnetic field strength corresponds to an increase in several performance characteristics of FT-ICR mass analyzers. Cyclotron frequency and mass resolving power both increase linearly with magnetic field strength [69, 70]. Additionally, the upper mass limit and maximum number of trapped ions that can be stored in the ICR cell before peak coalescence occurs are properties that increase quadratically with increasing magnetic field strength [69, 70]. These relationships between FT-ICR attributes and magnetic field strength demonstrate the attractiveness of higher magnetic fields [70]. Although these features are not measures of performance, the thermal and post-excite radii decrease linearly and the ion translational (kinetic) energy increases quadratically with magnetic field strength [69], as well.

1.3.1 Excitation

Once in the ICR cell the ion cloud needs to be excited in order to detect any appreciable signal; excitation in ICR has two functions 1) to cause phase coherence of ions of the same *m/z* and 2) expand the orbit of the cloud enough to induce a signal. An rf voltage ($V_{p\text{-}p,}$ V) can be applied to the opposing excitation plates (see **Figure 1.6**) for a time t_{excite} (s) and ions with a cyclotron frequency equal to the rf frequency will absorb this energy and expand to a larger orbit; the radius of this orbit is the post-excite radius (**Equation 1.3**).

$$r_{post-excite} = \frac{V_{p-p} t_{excite}}{2 B_0 d} \qquad (\mathbf{1.3})$$

Ions in the cell will then move as a phase-coherent ion packet in proximity to the detection plates. **Equation 1.4** displays the relationship between the instantaneous charge induced on the top detection plate (*Q,* A) and the number of ions (*N*), charge (*q,* C), post-excite radius ($r_{post\text{-}excite}$):

$$Q(t) = \frac{-Nqr_{post-excite}\cos(\omega t)}{d} \qquad (\mathbf{1.4})$$

where *d* is the distance between the detection plates and *cos(ωt)* is the cyclotron frequency [71]. The cyclotron frequency, charge, and distance between the detection plates are constant for a molecule and particular ICR cell geometry, respectively. The number of ions can be changed to some degree by varying the concentration; however, peak coalescence and large frequency shifts occur once the cell nears its capacity threshold, thereby significantly reducing mass resolving power and mass measurement accuracy [72-74]. Induced charge, therefore ion abundance, can be most readily modulated by varying the post-excite radius, corresponding to alterations in excitation voltage.

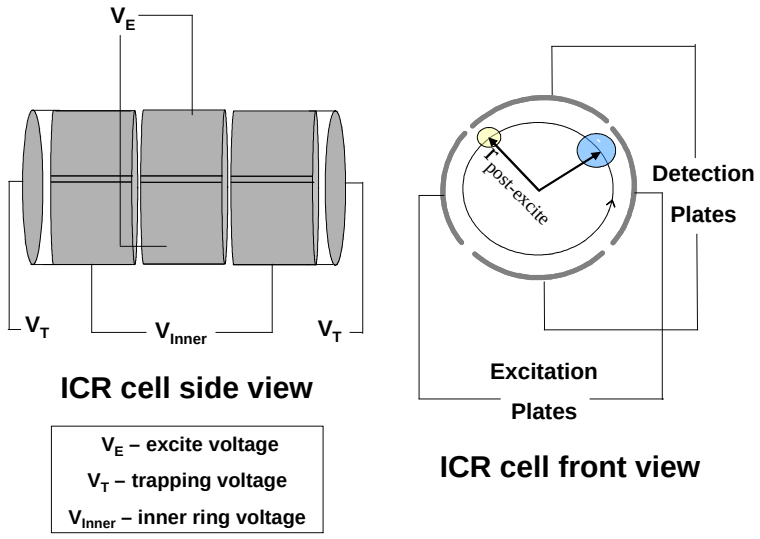


Figure 1.6 ICR cell. Side and front views of the cylindrical ICR cell, indicating applied voltages. The circles represent two different sized ion clouds orbiting at a radius, $r_{post\text{-}excite}$.

1.3.2 Detection

As the ion packet orbits in the ICR cell, an image current is produced resulting from the movement of electrons from one detection plate to the opposing detection plate. This image current appears as a sinusoidal signal as a function of time (*i.e.,* time domain signal), which is Fast Fourier transformed to extract the frequencies of the various ions.

Frequencies can then easily be converted to *m/z* values using **Equation 1.5**, where the cyclotron frequency, ω_c (Hz), is proportional to the charge (*q,* C), magnetic field (B_0, A/m), and the mass (*m*) [69]. One of the remarkable features of Fourier transform lies in the ability to use a broadband excitation to excite and detect ions of various *m/z* values simultaneously (*i.e.,* the multiplex advantage).

$$\omega_c = \frac{qB_0}{m} \qquad (\mathbf{1.5})$$

Signal dampening frequently occurs in ICR and can be quantified with the value τ, which is defined as the time it takes the signal to decay to 1/*e* of the initial amplitude. This occurrence results from the dephasing of ion clouds (loss of cloud coherence), therefore causing a decrease in the detected current (*i.e.* signal amplitude) [75]. It is important to note that pressure/collision dampening from ion-ion and ion-neutral collisions also result in ICR signal dampening [75], however improved vacuum technology has eliminated this dampening mechanism in modern FT-ICR mass spectrometers.

Cloud coherence is notably affected by ion cloud density [75]. Mitchell and Smith reported greater ion cloud densities result in larger rotational frequencies, thus counteracting the shearing forces from magnetic or electric field inhomogeneities that disrupt the cloud [76]. The rotational frequency of the cloud is a consequence of Coulombic interactions between ions of the same *m/z* (*i.e.,* same ion cloud), which cause the cloud to revolve about its symmetry axis [76].

1.3.3 Mass Measurement Accuracy

Mass measurement accuracy is the difference in the observed frequency, ω_{obs}, and theoretical cyclotron frequency. The cyclotron frequency of an ion is reduced by frequency perturbations, which are a consequence of the trapping potential and other ions in the cell (space charge effects) according to the relationship shown in **Equation 1.6**.

$$\omega_{obs} = \frac{qB_0}{m} - \frac{2\alpha V}{a^2 B_0} - \frac{q\delta G_i}{\varepsilon_0 B_0} \qquad \textbf{(1.6)}$$

The magnetron component (second term in **Equation 1.6**) results from the electrostatic trapping potential, which introduces an outwardly-directed force (opposing the Lorentz force) and thus reduces the observed frequency. However, the magnetron component is constant for a particular ICR cell geometry (*a*, diameter of the cell, and α, the cell geometry constant), trapping voltage (*V*), and magnetic field (B_0), therefore this term can be readily corrected for with external calibration. Importantly, the trapping voltage should be the same for the calibration and data acquisition.

Space charge effects (represented by the third term in **Equation 1.6**), which arise from the influence of electric fields produced by the ions on one another, also limit mass measurement accuracy in FT-ICR [77, 78]. The charge (*q*) and ion cloud geometry factor (G_i) affect the magnitude of space charge effects. However, space charge effects are largely a function of the number of ions in the cell (δ). Therefore, a larger number (*i.e.*, higher concentration) will result in a larger space charge effect and significantly reduce mass measurement accuracy. High mass measurement accuracy can be achieved using external calibration if the ion density during external calibration and analyte detection are equal. Until the recent introduction of automatic gain control [79], obtaining reproducible ion populations between experiments was difficult.

The negative consequences of space charge effects on mass measurement accuracy can be reduced by internal calibration. Internal calibration ensures that the analyte and calibrant are detected simultaneously, thus both ion populations will undergo the same frequency perturbations induced by the electrostatic trapping potential and space charge effects in the ICR cell. The density term in **Equation 1.6** is the same for the analyte and internal calibrant.

Improving mass measurement accuracy has long been an area of focus in the Muddiman group. Early experiments centered on internal calibration, and most prominently on the development [80-82] and implementation [83-90, 91] of the dual electrospray source. More recently, external calibration laws have been developed in the Muddiman group [92, 93]. External calibration is greatly facilitated by automatic gain control [79], and is likely to be an area of continued focus. A review of both external and internal calibration strategies for FT-ICR mass spectrometry, including work originating from the Muddiman group and other laboratories, is presented in reference [94].

The unparalleled mass measurement accuracy, mass resolving power, sensitivity, and dynamic range of FT-ICR mass spectrometry make it the paramount mass analyzer technology for biomolecule analysis [69, 95-98]. Moreover, the multiple-charging mechanism inherent to ESI presents itself as a perfect complement to the FT-ICR mass analyzer. Multiply-charging reduces the *m/z* value, increasing the molecular weight range amenable to analysis, and correspondingly a higher frequency. Advantages of higher frequencies are multi-fold. First, mass resolving power is higher at higher frequencies [69]. Second, frequency shifts, therefore mass measurement accuracy in the *m/z* domain, have less of an impact on higher frequencies. Additionally, more charges on a molecule increase the instantaneous charge (*Q*) on the detection plate (**Equation 1.4**). Finally, multiply-charged species are more amenable to fragmentation [99, 100].

1.4 Research Overview

The research presented herein examines the roles of physical processes (instrumental parameters, temperature) and chemical (hydrophobic tagging) on protein quantification by electrospray ionization-Fourier transform-ion cyclotron resonance (ESI-FT-ICR) mass spectrometry. The imparted understanding of the instrumental parameters (post-excite radius, excitation voltages, trapping voltages, excitation waveforms) that lend to

confident identification and accurate quantification of proteins in complex mixtures will enable future experiments in the Muddiman laboratory targeting biomarker discovery and validation. Moreover, the understanding and development of strategies that increase ion abundance, and subsequently lower the limit of detection, will also facilitate future studies of low abundance clinically relevant markers (***e.g.,*** BNP-32).

The large number and wide dynamic range of proteins present in biological fluids [34] warrant technologies capable of surmounting these challenges for global proteomics analyses. As the current trend indicates, more effective global proteomics strategies combine several separation steps in tandem [101-105], thus increasing the overall peak capacity [106] and consequently the number of components that can be analyzed. Complex mixture analysis warrants technologies with higher peak capacities (*i.e.,* the number of equally resolvable components in a defined two dimensional space) [106]. The concept of "peak capacity," originally defined by separation scientists, is developed for TOF and FT-ICR mass analyzers, which are commonly-employed mass spectrometry platforms for proteomics analyses in **Chapter 2**. This study is an objective assessment of these technologies with respect to complex mixture analysis, and ultimately enables the determination of the appropriate mass spectrometry-based platform for future studies.

The large number of proteins present in complex mixtures also necessitates a means to assess retention time reproducibility and mass measurement accuracy. Accurate mass and time tags enable confident protein identifications [107]. Since molecules with similar chemical formulas are comprised of the same atoms, these molecules will have a similar mass excess. The concept that peptides are grouped at particular mass values (owing to similar mass excess) was exploited with the design and utilization of internal standards that fall in the mass regions that are unoccupied by peptides (referred to as mass excess internal standards). In **Chapter 3**, it is shown that use of mass excess internal standards decreases the likelihood of overlap with endogenous peptides in the mass space, thereby allowing

assessment of retention time reproducibility and mass measurement accuracy. Furthermore, several normalization strategies utilizing the mass excess internal standards to enable accurate quantification are presented in **Chapter 3**.

In addition to normalization strategies, the accuracy of ion abundance measurements can be improved upon by adjusting certain instrumental parameters. In a theoretical examination of FT-ICR theory, Gorshkov stated there exists an optimal excite radius [72]. Later, Hawkridge, ***et al.*** experimentally demonstrated the importance of post-excite radius on mass measurement accuracy and ion abundance [87]. Linear improvements in mass measurement accuracy and ion abundance with increasing post-excite radius were observed up to a certain radius, after which the response rapidly decreased [87]. Mass measurement error is minimized and ion abundance is maximized at the optimal post-excite radius. These initial studies utilized a simple system consisting of bradykinin and polymer (polyethylene glycol). The research presented in **Chapter 4** builds upon these initial studies to examine the roles of post-excite radius and axial confinement in quantitative proteomics measurements.

Differences in signal decay constants between two analytes can also affect quantification accuracy. Earlier investigations in the Muddiman laboratory demonstrated the effect of this dampening mechanism on relative quantification using cyclosporin A (CsA) and cyclosporin G (CsG) as a model system [108]. These studies showed a lower concentration of one species (CsA) compared to another species (CsG) resulted in much faster signal dampening (***i.e.,*** small τ value). The less concentrated CsA produced a less dense ion cloud (compared to CsG) with fewer inter-ion Coulombic interactions, consequently resulting in a lower rotational frequency. Hence, the CsA cloud was more susceptible to shearing forces and was readily disrupted, causing the signal to damp rapidly. In order to develop a better understanding of the effect of signal decay on relative quantification and how it relates to proteomics measurements, we studied the signal decay constants from an unlabeled and

stable-isotope labeled peptide in varying ratios (**Chapter 4**). Additionally, previous experiments were performed at a single post-excite radius, which has been shown to affect quantification (*vide supra*). Experiments using a wider range of excitation and trapping voltages is presented in **Chapter 4**. In **Chapter 4** several experiments are presented that examine the role of excitation waveform, post-excite radius, and axial confinement fields on the quantification of proteomes by ESI-FT-ICR mass spectrometry.

A second challenge associated with complex mixture analysis is the wide dynamic range of proteins in complex mixture. Anderson and Anderson report a 10^{10} difference in clinically-relevant concentrations for proteins at extremes of the concentration range [34]. Current LC-MS technologies have dynamic ranges of 10^2-10^4, which undoubtedly is less than the range found in biological samples [34]. One strategy to overcome the dynamic range problem is pre-fractionation, during which high-abundance proteins are removed [83, 102, 109]. Another approach would be to increase the abundance of certain proteins (*i.e.,* protein enrichment) through the use of immunoaffinity purification [110] or special enrichment columns for separation (*e.g.,* IMAC for phosphoproteins [111], lectins for glycosylated proteins [112]).

In addition to the aforementioned strategies, the ion abundance of a low abundant protein can be increased during the electrospray ionization process. The electrospray ionization response of a peptide corresponds to the peptide's ability to escape from the droplet. Two ways to encourage a peptide to leave the droplet would be to 1) increase the rate of droplet desolvation or 2) increase the hydrophobicity of the peptide, thus increasing the peptide's affinity for the droplet surface allowing it to more effectively compete for charge. The effect of temperature (*i.e.,* increasing the electrospray droplet desolvation rate) on the electrospray ionization response is presented in **Chapter 5**. In **Chapter 6**, the role of peptide hydrophobicity on the electrospray ionization response is investigated by comparing ion abundance of an unmodified peptide to the ion abundance of the same peptide with a "hydrophobic tag."

1.5 References

1. Joyce, AR and Palsson, BO. The model organism as a system: Integrating 'omics' data sets. *Nature Reviews Molecular Cell Biology,* 2006. **7**: 198-210.

2. Morel, NM, Holland, JM, van der Greef, J, Marple, EW, Clish, C, Loscalzo, J, et al. Primer on medical genomics part xiv: Introduction to systems biology - a new approach to understanding disease and treatment. *Mayo Clinic Proceedings,* 2004. **79**: 651-658.

3. Crick, F. Central dogma of molecular biology. *Nature,* 1970. **227**: 561-563.

4. Nakamura, Y, Gojobori, T, and Ikemura, T. Codon usage tabulated from international DNA sequence databases: Status for the year 2000. *Nucleic Acids Research,* 2000. **28**: 292-292.

5. Sasaki, M, Kato, S, Kohno, K, Martin, GR, and Yamada, Y. Sequence of the cDNA-encoding the laminin-B1 chain reveals a multidomain protein containing cysteine-rich repeats. *Proceedings of the National Academy of Sciences of the United States of America,* 1987. **84**: 935-939.

6. Timpl, R, Rohde, H, Robey, PG, Rennard, SI, Foidart, JM, and Martin, GR. Laminin - glycoprotein from basement-membranes. *Journal of Biological Chemistry,* 1979. **254**: 9933-9937.

7. Martin, GR and Timpl, R. Laminin and other basement-membrane components. *Annual Review of Cell Biology,* 1987. **3**: 57-85.

8. Iwamoto, Y, Robey, FA, Graf, J, Sasaki, M, Kleinman, HK, Yamada, Y, et al. YIGSR, a synthetic laminin pentapeptide, inhibits experimental metastasis formation. *Science,* 1987. **238**: 1132-1134.

9. Spyroulias, GA, Nikolakopoulou, P, Tzakos, A, Gerothanassis, IP, Magafa, V, Manessi-Zoupa, E, et al. Comparison of the solution structures of angiotensin I & II - implication for structure-function relationship. *European Journal of Biochemistry,* 2003. **270**: 2163-2173.

10. Dzau, VJ. Significance of the vascular renin-angiotensin pathway. *Hypertension,* 1986. **8**: 553-559.

11. Kyte, J and Doolittle, RF. A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology,* 1982. **157**: 105-132.

12. Venter, JC, Adams, MD, Myers, EW, Li, PW, Mural, RJ, Sutton, GG, et al. The sequence of the human genome. *Science,* 2001. **291**: 1304-1351.

13. Lander, ES, Linton, LM, Birren, B, Nusbaum, C, Zody, MC, Baldwin, J, et al. Initial sequencing and analysis of the human genome. *Nature,* 2001. **409**: 860-921.

14. Wasinger, VC, Cordwell, SJ, Cerpapoljak, A, Yan, JX, Gooley, AA, Wilkins, MR, et al. Progress with gene-product mapping of the mollicutes - Mycoplasma-genitalium. *Electrophoresis,* 1995. **16**: 1090-1094.

15. Kenyon, GL, DeMarini, DM, Fuchs, E, Galas, DJ, Kirsch, JF, Leyh, TS, et al. Defining the mandate of proteomics in the post-genomics era: Workshop report. *Molecular & Cellular Proteomics,* 2002. **1**: 763-780.

16. Pandey, A and Mann, M. Proteomics to study genes and genomes. *Nature,* 2000. **405**: 837-846.

17. Mann, M and Jensen, ON. Proteomic analysis of post-translational modifications. *Nature Biotechnology,* 2003. **21**: 255-261.

18. Wilkinson, KD. Ubiquitination and deubiquitination: Targeting of proteins for degradation by the proteasome. *Seminars in Cell & Developmental Biology,* 2000. **11**: 141-148.

19. Johnson, ES. Protein modification by SUMO. *Annual Review of Biochemistry,* 2004. **73**: 355-382.

20. Sinensky, M and Lutz, RJ. The prenylation of proteins. *Bioessays,* 1992. **14**: 25-31.

21. Van den Steen, P, Rudd, PM, Dwek, RA, and Opdenakker, G. Concepts and principles of O-linked glycosylation. *Critical Reviews in Biochemistry and Molecular Biology,* 1998. **33**: 151-208.

22. Hubbard, MJ and Cohen, P. On target with a new mechanism for the regulation of protein-phosphorylation. *Trends in Biochemical Sciences,* 1993. **18**: 172-177.

23. Sherr, CJ. Cancer cell cycles. *Science,* 1996. **274**: 1672-1677.

24. Goedert, M, Spillantini, MG, Cairns, NJ, and Crowther, RA. Tau-proteins of Alzheimer paired helical filaments - abnormal phosphorylation of all 6 brain isoforms. *Neuron,* 1992. **8**: 159-168.

25. Saltiel, AR and Kahn, CR. Insulin signaling and the regulation of glucose and lipid metabolism. *Nature,* 2001. **414**: 799-806.

26. Blom, N, Gammeltoft, S, and Brunak, S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *Journal of Molecular Biology,* 1999. **294**: 1351-62.

27. Guex, N and Peitsch, MC. Swiss-model and the SWISS-PDBviewer: An environment for comparative protein modeling. *Electrophoresis,* 1997. **18**: 2714-2723.

28. Sander, C and Schneider, R. Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins-Structure Function and Genetics,* 1991. **9**: 56-68.

29. Baker, D and Sali, A. Protein structure prediction and structural genomics. *Science,* 2001. **294**: 93-96.

30. Dobson, CM. Protein folding and misfolding. *Nature,* 2003. **426**: 884-890.

31. Selkoe, DJ. Alzheimer's disease: Genes, proteins, and therapy. *Physiological Reviews,* 2001. **81**: 741-766.

32. Tian, Q, Stepaniants, SB, Mao, M, Weng, L, Feetham, MC, Doyle, MJ, et al. Integrated genomic and proteomic analyses of gene expression in mammalian cells. *Molecular & Cellular Proteomics,* 2004. **3**: 960-969.

33. Gygi, SP, Rochon, Y, Franza, BR, and Aebersold, R. Correlation between protein and mRNA abundance in yeast. *Molecular and Cellular Biology,* 1999. **19**: 1720-1730.

34. Anderson, NL and Anderson, NG. The human plasma proteome - history, character, and diagnostic prospects. *Molecular & Cellular Proteomics,* 2002. **1**: 845-867.

35. Partin, AW, Yoo, J, Carter, HB, Pearson, JD, Chan, DW, Epstein, JI, et al. The use of prostate-specific antigen, clinical stage and Gleason score to predict pathological stage in men with localized prostate-cancer. *Journal of Urology,* 1993. **150**: 110-114.

36. Plebani, M and Burlina, A. Biochemical markers of hepatic-fibrosis. *Clinical Biochemistry,* 1991. **24**: 219-239.

37. Duffy, MJ. Clinical uses of tumor markers: A critical review. *Critical Reviews in Clinical Laboratory Sciences,* 2001. **38**: 225-262.

38. Ridker, PM. Clinical application of c-reactive protein for cardiovascular disease detection and prevention. *Circulation,* 2003. **107**: 363-369.

39. Dao, Q, Krishnaswamy, P, Kazanegra, R, Harrison, A, Amirnovin, R, Lenert, L, et al. Utility of b-type natriuretic peptide in the diagnosis of congestive heart failure in an urgent-care setting. *Journal of the American College of Cardiology,* 2001. **37**: 379-385.

40. Bork, P, Dandekar, T, Diaz-Lazcoz, Y, Eisenhaber, F, Huynen, M, and Yuan, YP. Predicting function: From genes to genomes and back. *Journal of Molecular Biology,* 1998. **283**: 707-725.

41. Aebersold, R and Mann, M. Mass spectrometry-based proteomics. *Nature,* 2003. **422**: 198-207.

42. Mann, M, Hendrickson, RC, and Pandey, A. Analysis of proteins and proteomes by mass spectrometry. *Annual Review of Biochemistry,* 2001. **70**: 437-473.

43. Edman, P. On the mechanism of the phenyl isothiocyanate degradation of peptides. *Acta Chemica Scandinavica,* 1956. **10**: 761-768.

44. Edman, P. Method for determination of the amino acid sequence in peptides. *Acta Chemica Scandinavica,* 1950. **4**: 283-293.

45. Biemann, K and Scoble, HA. Characterization by tandem mass-spectrometry of structural modifications in proteins. *Science,* 1987. **237**: 992-998.

46. Mortz, E, O'Connor, PB, Roepstorff, P, Kelleher, NL, Wood, TD, McLafferty, FW, et al. Sequence tag identification of intact proteins by matching tandem mass spectral data against sequence data bases. *Proceedings of the National Academy of Sciences of the United States of America,* 1996. **93**: 8264-8267.

47. Henzel, WJ, Billeci, TM, Stults, JT, Wong, SC, Grimley, C, and Watanabe, C. Identifying proteins from 2-dimensional gels by molecular mass searching of peptide-fragments in protein-sequence databases. *Proceedings of the National Academy of Sciences of the United States of America,* 1993. **90**: 5011-5015.

48. Ong, SE, Foster, LJ, and Mann, M. Mass spectrometric-based approaches in quantitative proteomics. *Methods,* 2003. **29**: 124-130.

49. Ong, SE and Mann, M. Mass spectrometry-based proteomics turns quantitative. *Nature Chemical Biology,* 2005. **1**: 252-262.

50. Bronstrup, M. Absolute quantification strategies in proteomics based on mass spectrometry. *Expert Review of Proteomics,* 2004. **1**: 503-512.

51. Julka, S and Regnier, F. Quantification in proteomics through stable isotope coding: A review. *Journal of Proteome Research,* 2004. **3**: 350-363.

52. Zeleny, J. The electrical discharge from liquid points, and a hydrostatic method of measuring the electric intensity at their surfaces. *Physical Review,* 1914. **3**: 69-91.

53. Zeleny, J. Electrical discharges from pointed conductors. *Physical Review,* 1920. **16**: 102-125.

54. Magarvey, RH and Outhouse, LE. Note on the break-up of a charged liquid jet. *Journal of Fluid Mechanics,* 1962. **13**: 151-157.

55. Hines, RL. Electrostatic atomization and spray painting. *Journal of Applied Physics,* 1966. **37**: 2730-2736.

56. Tilney, R and Peabody, HW. Electrostatic coating processes. *British Journal of Applied Physics,* 1953. **4**: S51-S54.

57. Dole, M, Mack, LL, and Hines, RL. Molecular beams of macroions. *Journal of Chemical Physics,* 1968. **49**: 2240-2249.

58. Fenn, JB, Mann, M, Meng, CK, Wong, SF, and Whitehouse, CM. Electrospray ionization for mass-spectrometry of large biomolecules. *Science,* 1989. **246**: 64-71.

59. Fenn, JB. Ion formation from charged droplets - roles of geometry, energy, and time. *Journal of the American Society for Mass Spectrometry,* 1993. **4**: 524-535.

60. Gomez, A and Tang, KQ. Charge and fission of droplets in electrostatic sprays. *Physics of Fluids,* 1994. **6**: 404-414.

61. Taflin, DC, Ward, TL, and Davis, EJ. Electrified droplet fission and the Rayleigh limit. *Langmuir,* 1989. **5**: 376-384.

62. Kebarle, P and Tang, L. From ions in solution to ions in the gas-phase - the mechanism of electrospray mass-spectrometry. *Analytical Chemistry,* 1993. **65**: A972-a986.

63. de la Mora, JF. Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism. *Analytica Chimica Acta,* 2000. **406**: 93-104.

64. Fenn, JB, Rosell, J, Nohmi, T, Shen, S, and Banks, FJ. *Electrospray ion formation: Desorption versus desertion*, in *Biochemical and biotechnological applications of electrospray ionization mass spectrometry*. 1996, Amer Chemical Soc: Washington. p. 60-80.

65. Cole, RB. Some tenets pertaining to electrospray ionization mass spectrometry. *Journal of Mass Spectrometry,* 2000. **35**: 763-772.

66. Mack, LL, Kralik, P, Rheude, A, and Dole, M. Molecular beams of macroions .2. *Journal of Chemical Physics,* 1970. **52**: 4977-&.

67. Iribarne, JV and Thomson, BA. Evaporation of small ions from charged droplets. *Journal of Chemical Physics,* 1976. **64**: 2287-2294.

68. Smith, RD, Loo, JA, Edmonds, CG, Barinaga, CJ, and Udseth, HR. New developments in biochemical mass-spectrometry - electrospray ionization. *Analytical Chemistry,* 1990. **62**: 882-899.

69. Marshall, AG, Hendrickson, CL, and Jackson, GS. Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrometry Reviews,* 1998. **17**: 1-35.

70. Marshall, AG and Guan, SH. Advantages of high magnetic field for Fourier transform ion cyclotron resonance mass spectrometry. *Rapid Communications in Mass Spectrometry,* 1996. **10**: 1819-1823.

71. Comisarow, MB. Theory of Fourier-transform ion-cyclotron resonance mass-spectroscopy .2. Signal modeling for ion-cyclotron resonance. *Journal of Chemical Physics,* 1978. **69**: 4097-4104.

72. Gorshkov, MV and Nikolaev, EN. Optimal cyclotron radius for high-resolution FT-ICR spectrometry. *International Journal of Mass Spectrometry and Ion Processes,* 1993. **125**: 1-8.

73. Uechi, GT and Dunbar, RC. Space-charge effects on relative peak heights in Fourier transform-ion cyclotron-resonance spectra. *Journal of the American Society for Mass Spectrometry,* 1992. **3**: 734-741.

74. Taylor, PK and Amster, IJ. Space charge effects on mass accuracy for multiply charged ions in ESI-FTICR. *International Journal of Mass Spectrometry,* 2003. **222**: 351-361.

75. de Koning, LJ, Kort, CWF, Pinkse, FA, and Nibbering, NMM. Segmented Fourier-transform and its application to Fourier-transform ion-cyclotron resonance (FT-ICR) mass-spectrometry - ion abundances and mass measurements. *International Journal of Mass Spectrometry and Ion Processes,* 1989. **95**: 71-92.

76. Mitchell, DW and Smith, RD. Cyclotron motion of two coulombically interacting ion clouds with implications to Fourier-transform ion-cyclotron resonance mass-spectrometry. *Physical Review E,* 1995. **52**: 4366-4386.

77. Francl, TJ, Sherman, MG, Hunter, RL, Locke, MJ, Bowers, WD, and McIver, RT. Experimental-determination of the effects of space-charge on ion-cyclotron resonance frequencies. *International Journal of Mass Spectrometry and Ion Processes,* 1983. **54**: 189-199.

78. Jeffries, JB, Barlow, SE, and Dunn, GH. Theory of space-charge shift of ion-cyclotron resonance frequencies. *International Journal of Mass Spectrometry and Ion Processes,* 1983. **54**: 169-187.

79. Syka, JEP, Marto, JA, Bai, DL, Horning, S, Senko, MW, Schwartz, JC, et al. 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.

80. Hannis, JC and Muddiman, DC. A dual electrospray ionization source combined with hexapole accumulation to achieve high mass accuracy of biopolymers in Fourier transform ion cyclotron resonance mass spectrometry. *Journal of the American Society for Mass Spectrometry,* 2000. **11**: 876-883.

81. Nepomuceno, AI, Muddiman, DC, Bergen, HR, Craighead, JR, Burke, MJ, Caskey, PE, et al. Dual electrospray ionization source for confident generation of accurate mass tags using liquid chromatography Fourier transform ion cyclotron resonance mass spectrometry. *Analytical Chemistry,* 2003. **75**: 3411-3418.

82. Burke, MJ, Caskey, PE, and Muddiman, DC. *Dual electrospray ionization source for mass spectrometry*. 2006, Mayo Foundation for Medical Education and Research.

83. Johnson, KL, Mason, CJ, Muddiman, DC, and Eckel, JE. 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. *Analytical Chemistry,* 2004. **76**: 5097-5103.

84. Johnson, KL and Muddiman, DC. A method for calculating O-16/O-18 peptide ion ratios for the relative quantification of proteomes. *Journal of the American Society for Mass Spectrometry,* 2004. **15**: 437-445.

85. Frahm, JL, Mason, CJ, and Muddiman, DC. Utility of accurate monoisotopic mass measurements to confidently identify lambda exonuclease generated single-stranded amplicons containing 7-deaza analogs by electrospray ionization FT-ICR mass spectrometry. *International Journal of Mass Spectrometry,* 2004. **234**: 79-87.

86. Bergen, HR, Vasmatzis, G, Cliby, WA, Johnson, KL, Oberg, AL, and Muddiman, DC. Discovery of ovarian cancer biomarkers in serum using nanoLC electrospray ionization TOF and FT-ICR mass spectrometry. *Disease Markers,* 2003. **19**: 239-249.

87. Hawkridge, AM, Nepomuceno, AI, Lovik, SL, Mason, CJ, and Muddiman, DC. Effect of post-excitation radius on ion abundance, mass measurement accuracy, and isotopic distributions in Fourier transform ion cyclotron resonance mass spectrometry. *Rapid Communications in Mass Spectrometry,* 2005. **19**: 915-918.

88. Hawkridge, AM, Zhou, L, Lee, ML, and Muddiman, DC. Analytical performance of a Venturi device integrated into an electrospray ionization Fourier transform ion cyclotron resonance mass spectrometer for analysis of nucleic acids. *Analytical Chemistry,* 2004. **76**: 4118-4122.

89. Williams, K, Hawkridge, AM, and Muddiman, DC. Sub parts-per-million mass measurement accuracy of intact proteins and product ions achieved using a dual electrospray ionization quadrupole Fourier transform ion cyclotron resonance mass spectrometer. *Journal of the American Society for Mass Spectrometry,* 2007. **18**: 1-7.

90. Nepomuceno, AI, Mason, CJ, Muddiman, DC, Bergen, HR, and Zeldenrust, SR. Detection of genetic variants of transthyretin by liquid chromatography-dual electrospray ionization Fourier-transform ion-cyclotron-resonance mass spectrometry. *Clinical Chemistry,* 2004. **50**: 1535-1543.

91. Null, AP and Muddiman, DC. Determination of a correction to improve mass measurement accuracy of isotopically unresolved polymerase chain reaction amplicons by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Rapid Communications in Mass Spectrometry,* 2003. **17**: 1714-1722.

92. Muddiman, DC and Oberg, AL. Statistical evaluation of internal and external mass calibration laws utilized in Fourier transform ion cyclotron resonance mass spectrometry. *Analytical Chemistry,* 2005. **77**: 2406-2414.

93. Williams, DK and Muddiman, DC. 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. *Analytical Chemistry,* submitted.

94. Zhang, LK, Rempel, D, Pramanik, BN, and Gross, ML. Accurate mass measurements by Fourier transform mass spectrometry. *Mass Spectrometry Reviews,* 2005. **24**: 286-309.

95. Marshall, AG. Milestones in Fourier transform ion cyclotron resonance mass spectrometry technique development. *International Journal of Mass Spectrometry,* 2000. **200**: 331-356.

96. Grosshans, PB and Marshall, AG. General-theory of excitation in ion-cyclotron resonance mass-spectrometry. *Analytical Chemistry,* 1991. **63**: 2057-2061.

97. Marshall, AG. Fourier-transform ion-cyclotron resonance mass-spectrometry. *Accounts of Chemical Research,* 1985. **18**: 316-322.

98. Marshall, AG and Hendrickson, CL. Fourier transform ion cyclotron resonance detection: Principles and experimental configurations. *International Journal of Mass Spectrometry,* 2002. **215**: 59-75.

99. Rockwood, AL, Busman, M, and Smith, RD. Coulombic effects in the dissociation of large highly charged ions. *International Journal of Mass Spectrometry,* 1991. **111**: 103-129.

100. Williams, ER. Proton transfer reactivity of large multiply charged ions. *Journal of Mass Spectrometry,* 1996. **31**: 831-842.

101. Washburn, MP, Wolters, D, and Yates, JR. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnology,* 2001. **19**: 242-247.

102. Wagner, K, Miliotis, T, Marko-Varga, G, Bischoff, R, and Unger, KK. An automated on-line multidimensional HPLC system for protein and peptide mapping with integrated sample preparation. *Analytical Chemistry,* 2002. **74**: 809-820.

103. Issaq, HJ. The role of separation science in proteomics research. *Electrophoresis,* 2001. **22**: 3629-3638.

104. von Horsten, HH. Current strategies and techniques for the analysis of low-abundance proteins and peptides from complex samples. *Current Analytical Chemistry,* 2006. **2**: 129-138.

105. Righetti, PG, Castagna, A, Antonioli, P, and Boschetti, E. Prefractionation techniques in proteome analysis: The mining tools of the third millennium. *Electrophoresis,* 2005. **26**: 297-319.

106. Giddings, JC. *Unified separation science*. 1991, New York: Wiley. 320.

107. Smith, RD, Anderson, GA, Lipton, MS, Pasa-Tolic, L, Shen, YF, Conrads, TP, et al. An accurate mass tag strategy for quantitative and high-throughput proteome measurements. *PROTEOMICS,* 2002. **2**: 513-523.

108. Gordon, EF and Muddiman, DC. Impact of ion cloud densities on the measurement of relative ion abundances in Fourier transform ion cyclotron resonance mass spectrometry: Experimental observations of coulombically induced cyclotron radius perturbations and ion cloud dephasing rates. *Journal of Mass Spectrometry,* 2001. **36**: 195-203.

109. Lecchi, P and Abramson, FP. Analysis of biopolymers by size-exclusion chromatography mass spectrometry. *Journal of Chromatography A,* 1998. **828**: 509-513.

110. Hawkridge, AM, Heublein, DM, Bergen, HR, Cataliotti, A, Burnett, JC, and Muddiman, DC. Quantitative mass spectral evidence for the absence of circulating brain natriuretic peptide (BNP-32) in severe human heart failure. *Proceedings of the National Academy of Sciences of the United States of America,* 2005. **102**: 17442-17447.

111. Andersson, L and Porath, J. Isolation of phosphoproteins by immobilized metal (Fe-3+) affinity-chromatography. *Analytical Biochemistry,* 1986. **154**: 250-254.

112. Ogata, S, Muramatsu, T, and Kobata, A. Fractionation of glycopeptides by affinity column chromatography on concanavalin a-sepharose. *Journal of Biochemistry,* 1975. **78**: 687-696.

CHAPTER 2

Peak Capacity of Zero-, Single-, and Multi-Dimensional Separations Coupled to Time-of-Flight and Fourier Transform-Ion Cyclotron Resonance Mass Analyzers for Bottom-Up and Top-Down Proteomics

2.1 Introduction

Unearthing differences in biological matrices for comparative proteomics presents a formidable challenge that must be overcome with methodologies that afford accurate identification and quantification of proteins in complex mixtures. Mass spectrometry is the paramount platform for global proteomics measurements due its molecular specificity (mass measurement accuracy), detection limits (sensitivity), dynamic range, and peak capacity (mass resolving power). Experiments by Johnson, *et al.* demonstrated better than 3 ppm and 6 ppm mass measurement accuracy for internally and externally calibrated nanoLC-ESI-FT-ICR data, respectively [1, 2]. Furthermore, the introduction of new calibration laws for FT-ICR mass spectrometry data offers potential for improved mass measurement accuracy of complex mixtures using external calibration [3-5]. High mass accuracy measurements distinguish the mass of a particular species from the nominal mass, thus reducing the number of possible amino acid compositions [6]. Therefore, unique protein assignments can be made from high mass accuracy measurements, often in conjunction with HPLC retention time and tandem MS data, with the aid of available databases [7-12]. Moreover, high mass measurement accuracy and mass resolving power facilitate peptide/protein quantification.

High mass accuracy measurements in combination with high mass resolving power also reveal distinct groupings of molecules based on class (e.g., organics, nucleic acids, proteins). It was our aim to use the mass distribution of tryptic peptides/proteins, which is different from other molecule classes, to examine the peak capacity available to Fourier transform-ion cyclotron resonance (FT-ICR) and time-of-flight (TOF) mass spectrometry for

bottom-up and top-down proteomics. In order to define the peak capacity we first explored the distribution of tryptic peptides over the *m/z* range of 500-5000 Da and intact proteins over the *m/z* range 2,500 – 100,000 Da, which permitted us to define the *m/z* values populated by tryptic peptides and intact proteins, respectively. The allowed (populated) proteomics regions, in addition to the mass resolving power, were then used to calculate the peak capacity and peak capacity production rate. These results demonstrate the theoretical capabilities of front-end liquid chromatography coupled with FT-ICR and TOF mass analyzers regarding complex mixture analysis in bottom-up and top-down proteomics.

2.2 Bottom-Up Proteomics

2.2.1 Defining the Bottom-Up Proteomics Mass Space

Perhaps the most recognized utilization of the concept of grouped masses based on chemical composition is the Kendrick mass scale for organic compounds [13]. The Kendrick mass scale sets the mass of the CH_2 radical to exactly 14.0000. Kendrick acknowledged that organic compounds comprised only of CH_2 repeats will have a zero mass defect on the Kendrick mass scale. Moreover, the chemical formulas of organic compounds with a mass defect greater than zero can be identified from tables given that members of the same family (*i.e.*, similar chemical formulas) will have the same mass defect.

There has been discrepancy in the literature with respect to the term “mass defect.” Since the introduction of the relativity theory by Einstein, “mass defect” has been defined as the difference in the sum of the constituents (protons, neutrons, and electrons) and the measured mass of the atom (*i.e.*, exact mass) [14]; the calculation is shown in **Equation 2.1**:

$$mass\,defect = \left(Z \times m_p\right) + \left(\left(A - Z\right) \times m_n\right) + \left(N \times m_e\right) - \left(m\right) \quad \textbf{(2.1)}$$

where Z is the atomic number, m_p represents the mass of a proton, A corresponds to the mass number, m_n is the mass of a neutron, N signifies the number of electrons, m_e is the

mass of an electron, and *m* indicates the exact mass of the molecule. For example, the mass defect of oxygen is calculated according to **Equation 2.2**:

$$mass\,defect = (8\times 1.007276) + ((16-8)\times 1.008665) + \left(8\times 5.485799\times 10^{-4}\right) - (15.994915) = 0.1370\,Da \quad \textbf{(2.2)}$$

The mass defect is also referred to as the "nuclear mass defect" or "mass deficiency," and results from the fact that the mass of an atom is always less than the sum of the masses of the nucleons. This mass difference is due to the energy required to hold the nucleons in the nucleus, known as the nuclear binding energy. A second definition, often adopted by mass spectrometrists, defines mass defect as the difference between the nominal mass (or mass number) and exact mass, sometimes referred to as the "negative mass defect." However, the correct term for this difference, according to IUPAC, is "mass excess." The difference in the mass defect and mass excess is illustrated in **Figure 2.1**.

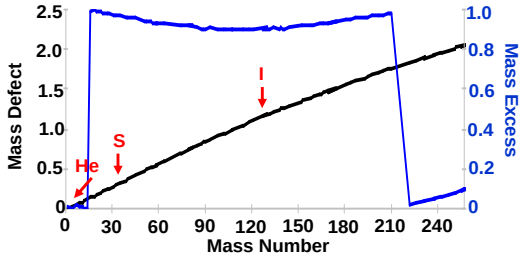


Figure 2.1 Mass defect and mass excess of the elements plotted as a function of mass number.

Similar to organic compounds, tryptic peptides are grouped according to mass excess. These groupings in the mass spectra create allowed (highly populated) and forbidden (unpopulated) regions. The phenomenon of "forbidden zones" in the mass distribution of tryptic peptides has been described as early as 1995 by Mann [15] and since then by several other investigators [6, 16-18].

The mass distribution of singly charged tryptic peptides, which would be generated in a MALDI experiment, was investigated. The source for all peptide data was NCBI's 'nr' (non-redundant) protein sequence database which contains entries from GenBank [19] CDS translations (CoDing Sequences corresponding to the amino acid sequences, including start and stop codons; does not correspond to actual mRNA sequence), PDB [20], SwissProt [21, 22], PIR [23], and PRF. A Java program was written to read in and analyze the data from the

nr.fasta flat file. Non-human proteins were filtered by excluding sequences that did not have the string "*Homo sapiens*" in the description field. In total 119,771 human protein sequences were included in the analysis; 1,993 proteins were excluded due to missing or ambiguous amino acids (as indicated in the database by the presence of the symbols 'X', 'Z', 'B', or 'U' in the sequence). *In-silico* tryptic digestion cleaving after every lysine and arginine residue (complete digestion) resulted in a total 4,596,261 peptide fragments. The mass distribution of singly charged tryptic peptides is shown in **Figure 2.2**; the mass excess (*i.e.*, nominal mass subtracted from exact mass) is plotted as a function of the monoisotopic mass over the mass range 500 to 5000 Da. In **Figure 2.2** there are three populated regions in the mass space, which arise from the similar chemical composition of all peptides (C, H, O, N, S). The unoccupied space encompasses the masses that peptides can not assume (*i.e.*, forbidden zones).

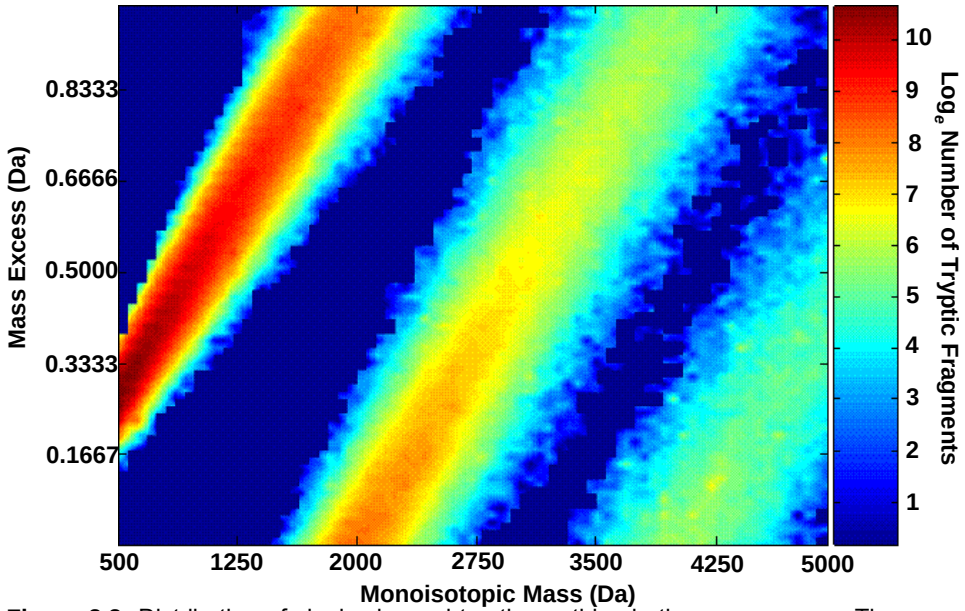


Figure 2.2 Distribution of singly charged tryptic peptides in the mass space. The mass excess of tryptic peptides from NCBI's nr database is plotted as a function of monoisotopic mass for 500-5000 Da. The mass excess was calculated by summing the monoisotopic masses of each amino acid in the sequence and then subtracting the nominal mass, thus resulting in values less than one. Red regions indicate highly populated areas. Blue regions are unpopulated areas or "forbidden zones." The scale on the right indicates the natural logarithm number of fragments for each color. Singly charged tryptic fragments occupy ~54% of the mass space in the region between 500 to 5000 Da.

In order to linearize the data and thus characterize the masses that tryptic peptides can assume, the total mass excess was computed by first determining the mass excess (*i.e.*, nominal mass subtracted from monoisotopic mass) of each amino acid and then summing the mass excess of each amino acid in the sequence. This is different from **Figure 2.2**, for which the mass excess was calculated by totaling the monoisotopic masses of each amino acid in the sequence and then subtracting the nominal mass. A continuous series of mass excess values that do not wrap around at 1.0 results. **Figure 2.3** shows the mass excess plotted as a function of monoisotopic mass. The equation of the line is:

$$\Delta m = 4.7831\times10^{-4}\, m + 0.0279 \qquad (\mathbf{2.3})$$

where Δm is the mass excess and m is the monoisotopic mass. The dotted lines represent the upper and lower confines encompassing roughly 99% of the data.

An expression for the width of the line defined by **Equation 2.4** can be deduced by subtracting the equations of the lines for upper and lower boundaries:

$$W(m) = 1.9107\times10^{-4}\, m + 0.0741 \qquad (\mathbf{2.4})$$

where *W(m)* is the width of the line as a function of monoisotopic mass (*m*). The width for

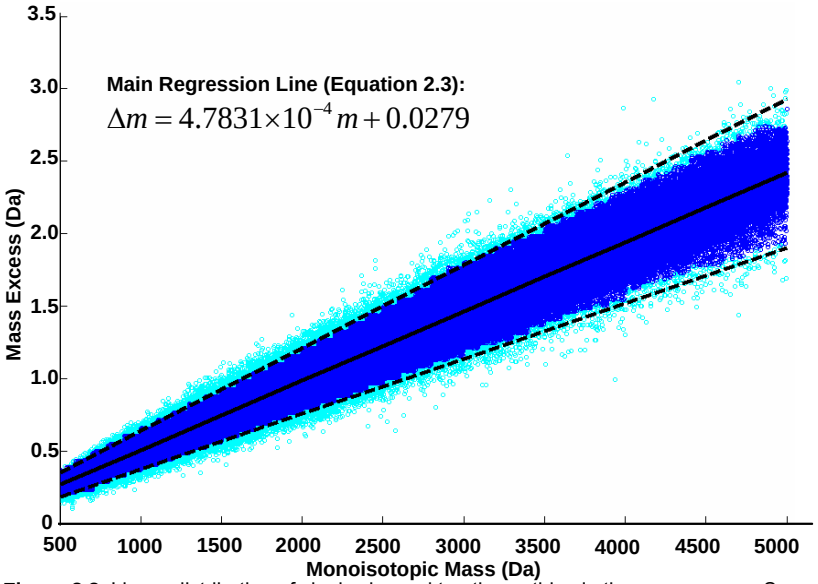


Figure 2.3 Linear distribution of singly charged tryptic peptides in the mass space. Same tryptic peptides shown in **Figure 2.2**, however mass excess is a continuous function. Mass excess in this plot is a summation of mass excesses (i.e., nominal mass subtracted from monisotopic mass) for each amino acid in the sequence, which resulted in values from 0 to 3. The solid line indicates the main regression line, equation shown. The dotted lines contain ~99% of the data; equations of these lines were used to calculate the width equation (**Equation 2.4**).

an average tryptic peptide (monoisotopic mass 1557.1021 Da) is slightly greater than a third of a mass unit.

Both **Equations 2.3** and **2.4** define the most probable mass excess values that peptides can populate for a particular monoisotopic mass. As the monoisotopic mass increases, the width becomes larger; it follows that more probable mass excesses exist for larger masses. **Equations 2.3** and **2.4** are similar to those equations described by Mann [15] and Lehmann *et al.* [16], except that the lines in **Figure 2.3** are not forced through the origin.

The presence of forbidden zones has a number of significant consequences. First, several advantages arise from similar mass excesses of peptides. One such benefit is the straightforward recognition of outliers, which has been exploited for the identification of both glycopeptides [16] and lipids [24]. Deviations from the predicted mass excess (*i.e.,* outside of the calculated width) would suggest possible non-peptide contaminants or modifications (*e.g.,* post-translational modifications, alternative splicing, truncations). Additionally, calibrants can be introduced into unoccupied space, and thus would not overlap with the peptide peaks [25]. Furthermore, innumerable approaches have employed mass tags to shift peptides to avoid chemical noise and distinguish tagged from untagged species (for example see Hall, *et al.* [26]).

Figure 2.4 illustrates how the addition of a chemical tag could shift tagged species from untagged species. The mass excess of peptides modified with a N-terminal tags incorporating either one or four bromines (reported by Hall *et al.* [26]) and unmodified peptides (same peptides in **Figure 2.3**) were plotted. It is apparent from **Figure 2.4** that the peptides modified with the single bromine tag overlap the unmodified peptides; however, peptides modified with the chemical tag containing four bromines do not coincide with the unmodified peptides. Similarly, peptides can be modified in order to expand the allowed regions [27]. However, these strategies are only applicable to the low mass region given that more possible mass excesses exist at higher masses; this is evident from the overlapping of

peptides modified with the chemical tag containing four bromines with unmodified peptides when *m/z* is greater than or equal to 3000 Da in **Figure 2.4**.

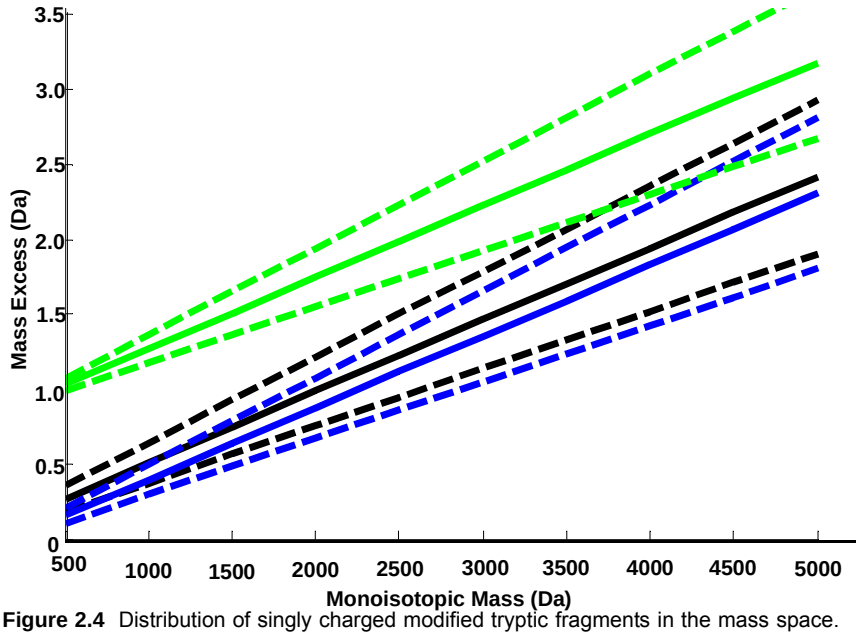


Figure 2.4 Distribution of singly charged modified tryptic fragments in the mass space. Black lines indicate the regression lines determined from the data plotted in **Figure 2.3** (unmodified tryptic peptides). The blue and green lines are regression lines for tryptic peptides that have been N-terminally modified with a "mass defect" tag containing one and four bromines, respectively, reported by Hall *et al.* Roughly 99% of the data is encompassed by the dashed lines for all three groups.

2.2.2 Peak Capacity in FT-ICR and TOF Mass Spectrometry for Bottom-Up Proteomics

There are also negative consequences of similar mass excess among peptides, which can be realized in terms of peak capacity. In chromatography, peak capacity defines the maximum number of equally resolved peaks that can fit into given separation space [28]. Naturally, larger peak capacities are required for more complex mixtures, which contain large numbers of components. Since the total peak capacity is a product of the individual methods for orthogonal methods, the peak capacity can be considerably increased by coupling chromatography methods with mass spectrometry.

In FT-ICR mass spectrometry, resolving power is a function of *m/z* and improves with magnetic field strength (see **Figure 2.5**) [29]. An acquisition time of 1.024 s and 1^+ charge state was assumed for all calculations. Additionally, a mass resolving power of 10,000,

typical of TOF mass analyzers, was plotted for comparison. Specifically for proteomics, the beauty of the exponential decline in resolving power as a function of mass for FT-ICR means that mass resolving power and therefore peak capacity (*vide infra*) are the highest where the majority of peptides are found (*i.e., m/z* 500 – 1500 Da), as shown in **Figure 2.5**.

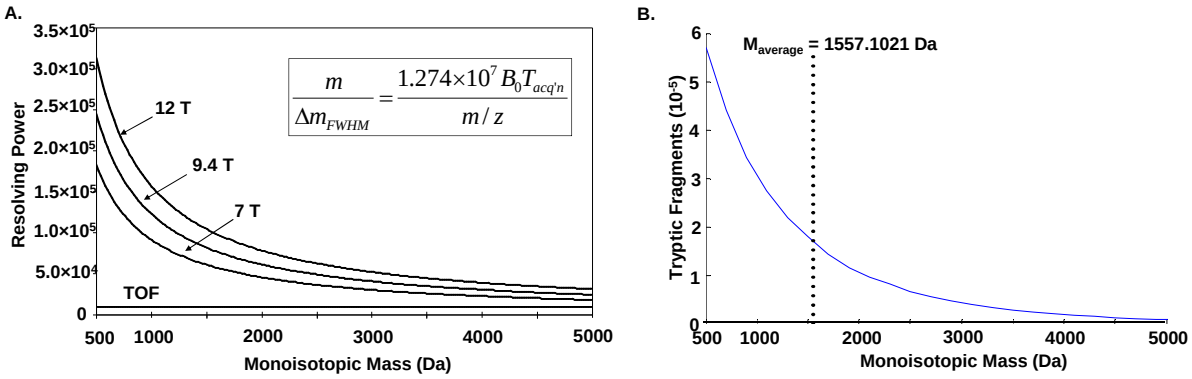


Figure 2.5 (A) Theoretical mass resolving power over the mass range of 500 to 5000 Da. An acquisition time of 1.024 s ($T_{acq'n}$) and a +1 charge state (z) were used to calculate the mass resolving power at the different magnetic field strengths (B_0) afforded to the FT-ICR mass analyzer using the equation described by Marshall [29] and shown in the inset. A constant mass resolving power of 10, 000 was assumed for TOF mass analyzer. **(B)** The number of tryptic fragments per 200 Da bin width from NCBI's nr database is plotted as a function of monoisotopic mass, curved line. The dotted line represents the average molecular weight for tryptic peptides, which is 1557.1021 Da.

Mass resolving power is the ability to distinguish two peaks, m_1 and m_2, some distance Δm apart at a defined percentage of the peak height (**Equation 2.5**); we assumed full width at half maximum (FWHM) for all calculations.

$$RP = \frac{m}{\Delta m_{FWHM}} \qquad (\mathbf{2.5})$$

An equation for the number of peaks that can be resolved in a 1 Da bin (*i.e.*, resolving elements) can be defined by using the definition of mass resolving power (**Equation 2.5**) and setting Δm_{FWHM} to 1 Da:

$$RE = \frac{RP}{m} \qquad (\mathbf{2.6})$$

where *RE* denotes resolving elements, *RP* is the resolving power, and *m* signifies the monoisotopic mass. For example, if we assume a resolving power of 10,000 and a mass of 500, 20 different species could be resolved between mass 500 and 501 Da (*i.e.*, 1 Da bin

width). Since peak capacity is defined as the number of equally resolvable peaks that can occupy a certain space delineated by the separation method [28], the total number of resolving elements in mass spectrometry is analogous to the concept of peak capacity in separation science.

The peak capacity of a mass analyzer will be reduced by the fraction of space consumed by singly charged tryptic peptides; we can think of this as the accessible peak capacity. With this information in mind, equations for peak capacity in FT-ICR and TOF mass analyzers can be derived. The peak capacity (n_c) of mass spectrometry for bottom-up proteomics is related to the number of resolving elements (*RE*) by the following:

$$n_c = RE \times W(m) \qquad (\mathbf{2.7})$$

Equation 2.7 adjusts the number of resolving elements in a 1 Da bin by the fraction of space tryptic peptides can occupy described by the width equation (**Equation 2.4**). **Equation 2.6** can be substituted into **Equation 2.7**, which puts the peak capacity in terms of resolving power for a 1 Da bin:

$$n_c = \frac{RP}{m} \times W(m) \qquad (\mathbf{2.8})$$

Equation 2.8 defines the maximum number of equally resolved peaks in mass spectrometry for tryptic peptides. Additionally, **Equation 2.8** can be extended to any class of molecules with a delimited mass distribution *W(m)*.

If we consider the number of isotopes resulting from the detection of a single tryptic peptide, we can estimate the number of resolved tryptic peptides from the number of resolved peaks. Assuming isotopic resolution, each peptide in the mass spectrum will be comprised of several isotopes. The number of isotopic peaks detected in a mass spectrum with an abundance greater than 5% (n_I) can be predicted from the monoisotopic mass (*m*), as shown in **Equation 2.9**.

$$n_I = 0.001m + 2.0492 \qquad (\mathbf{2.9})$$

This relationship was determined from a plot of the monoisotopic mass versus the number of isotopes in the isotopic distribution (data not shown). Theoretical isotopic distributions were generated using *averagine*-derived chemical formulas [30]. Considering each peptide will produce n_I number of isotopes and considering only singly charged species, if we divide the peak capacity by n_I, the peak capacity will be in terms of the mass resolving power and number of peptides, as shown in **Equation 2.10**.

$$n_c = \frac{\frac{RP}{m} \times W(m)}{n_I} \qquad (\mathbf{2.10})$$

Once the equation for mass resolving power for FT-ICR [29] or a mass resolving power of 10,000 for TOF, the width equation (**Equation 2.4**), and the equation for the number of isotopic peaks (**Equation 2.9**) are substituted into **Equation 2.10**, the resulting peak capacity equations for FT-ICR and TOF mass spectrometry for bottom-up proteomics become **Equation 2.11** and **2.12**, respectively.

$$n_c = \frac{2.4926 \times 10^3 B_0}{1 \times 10^{-3} m^2 + 2.0492 m} + \frac{9.6705 \times 10^5 B_0}{1 \times 10^{-3} m^3 + 2.0492 m^2} \qquad (\mathbf{2.11})$$

$$n_c = \frac{1.9107}{1 \times 10^{-3} m + 2.0492} + \frac{741.2780}{1 \times 10^{-3} m^2 + 2.0492} \qquad (\mathbf{2.12})$$

Equation 2.11 encompasses the equation for mass resolving power in FT-ICR at a particular magnetic field strength, B_0 (Tesla) as described by Marshall [29].

We integrated **Equations 2.11** and **2.12** with respect to monoisotopic mass over the mass range 500-5000 in order to find the area under the curve, which corresponds to the total peak capacity available to these techniques. **Equations 2.13** and **2.14** are the integrated equations for FT-ICR and TOF mass analyzers, respectively:

$$n_C = \left[1910.7 \ln 2.0492 + 1 \times 10^{-3} m\right] + \left[-361.7402 \ln \frac{2.0492 + 1 \times 10^{-3} m}{m}\right] \qquad (\mathbf{2.13})$$

$$n_C = \left[-1216.4032 B_0 \ln \frac{2.0492 + 1\times10^{-3} m}{m} \right] + \left[\frac{-471917.5718 B_0}{m} + 230.2936 B_0 \ln \frac{2.0492 + 1\times10^{-3} m}{m} \right] \quad (\mathbf{2.14})$$

Equation 2.13 can be used to estimate the peak capacity for FT-ICR at any magnetic field strength. Additionally, any mass range can be utilized to compute peak capacity for both FT-ICR and TOF using **Equations 2.13** and **2.14**. To determine the peak capacity over a particular mass range, n_c is calculated at each mass for a particular mass analyzer; the difference yields total n_c.

The total peak capacity determined from **Equation 2.13** over the mass range 500 – 5000 Da is approximately 15,000 for 7 T FT-ICR, 20,000 for 9.4 T FT-ICR, and 25,000 for 12 T FT-ICR mass analyzers. The total peak capacity for a TOF mass analyzer over the mass range 500-5000 Da is 2400. An increase in the magnetic field strength from 7 T to 12 T nearly doubles the total peak capacity. Importantly, a greater than ten-fold improvement in total peak capacity is seen when the TOF mass analyzer is replaced with a 12 T FT-ICR mass spectrometer, which objectively demonstrates the importance of using a high resolving power mass analyzer for complex-mixture analysis. Moreover, unresolved peaks compromise quantification and mass measurement accuracy.

2.2.3 Peak Capacity of 1 D- and 2 D-LC MS for Bottom-Up Proteomics

Regardless of the high peak capacities of FT-ICR and TOF mass spectrometry, direct infusion of complex mixtures is not feasible. The sheer number and wide concentration range of components in biological matrices present several problems for ionization and detection processes, which are primarily due to ion suppression effects during ionization [31-34] and space-charge effects in the ICR cell [35]. Therefore, a front-end separation method must be employed.

In our evaluation of 1 D- and 2 D-LC MS methods we examined two different methods of reversed-phase HPLC; one method utilizes nano-LC column (30 cm × 75 μm

i.d.) [1], while the other employs a longer column (80 cm × 150 μm i.d.) [36]. Increasing the length of a column (*L, cm*) corresponds to an increase in column plate number (*N*), assuming the particle diameter (d_p, *μm*) remains the same; this relationship is shown in **Equation 2.14** [37]:

$$N \approx \frac{3500L}{d_p} \qquad (\mathbf{2.14})$$

The numbers of theoretical plates for the two columns in our investigations are approximately 21,000 and 93,333 for the 30 cm × 75 μm i.d. and 80 cm × 150 μm i.d. columns, correspondingly. Column length primarily influences the resolution (**Equation 2.15**), which can be estimated from the selectivity (determined from the separation factor, $\alpha = k_1/k_2$) and the retention factor (*k*, relative difference in retention time of an analyte compared to an unretained species).

$$R_S = 1/4(\alpha - 1)N^{1/2}\frac{k}{1+k} \qquad (\mathbf{2.15})$$

Equation 2.15 relies on the assumption that the two peaks used to calculate the resolution have similar retention times (peak widths). Resolution, consequently, affects the peak capacity. The experimentally-achieved peak capacity for the Johnson *et al.* method was calculated as 159 using an equation described by Neue to determine the experimental peak capacity (n_c) in gradient elution LC [38], which is likely a generous estimate (**Equation 2.16**):

$$n_c = 1 + \frac{t_g}{(1/n)\sum_1^n w} \qquad (\mathbf{2.16})$$

The gradient time is represented by t_g and *n* is the number of peaks used to determine the peak width (*w*). **Equation 2.16** only holds true when the peak width is approximately the same throughout the gradient separation, as in the case of peptides [38]. The reported peak capacity of 1125 was used for the Shen *et al.* method [36].

Peak capacity is a multiplicative property for orthogonal separations methods [28], therefore we multiplied the peak capacity for each method to obtain the overall peak capacity for the separation method (*i.e.,* reversed-phase LC, strong cation exchange) coupled to mass spectrometry. Peak capacities for the two reversed-phase LC methods were then used to determine the total peak capacity for one-dimensional (1 D) HPLC-MS (RP-HPLC FT-ICR and RP-HPLC TOF MS) and two-dimensional (2 D) HPLC-MS (*i.e.,* strong cation exchange, SCX, RP-HPLC FT-ICR and SCX RP-HPLC TOF MS). We assumed a magnetic field strength of 12 T for all FT-ICR calculations. The numerous variables in chromatography (*e.g.,* stationary phase, column length, gradient) and mass spectrometry (*e.g.,* mass analyzer, acquisition time) make it impossible for an all-inclusive study of the peak capacity of LC-MS methods. Closer examinations of various aspects of separations on peak capacity have been published [39-41]. The methods chosen are intended to illustrate the potential of separations coupled to high-end mass spectrometers for complex mixture analysis. However, one could easily calculate the peak capacity of a separation method using approaches described in the literature and then multiply the separation peak capacity by the FT-ICR or TOF mass spectrometry peak capacity (*vide supra*) to obtain a total peak capacity for peptide analysis by the separation method coupled with FT-ICR or TOF mass spectrometry.

The calculated peak capacities are displayed in a heat map in **Figure 2.6**; more red colors represent higher peak capacities. Peak capacities for direct infusion and RP-HPLC-MS are displayed in the bottom row (labeled with 0 SCX fractions on the y-axis). A two orders of magnitude improvement in peak capacity over direct infusion experiments is attained by coupling RP-HPLC to mass spectrometry. Alternatively, coupling SCX to mass spectrometry improves peak capacity by an order of magnitude over direct infusion analysis. Further augmentation in peak capacity can be achieved by employing SCX, followed by RP-HPLC, and then mass spectrometry (2 D LC-MS). The peak capacity can be approximately

doubled for every 5 SCX fractions. Moreover, RP-HPLC retention times and pI information obtained from SCX fractionation provides two additional analytical constraints [10, 12].

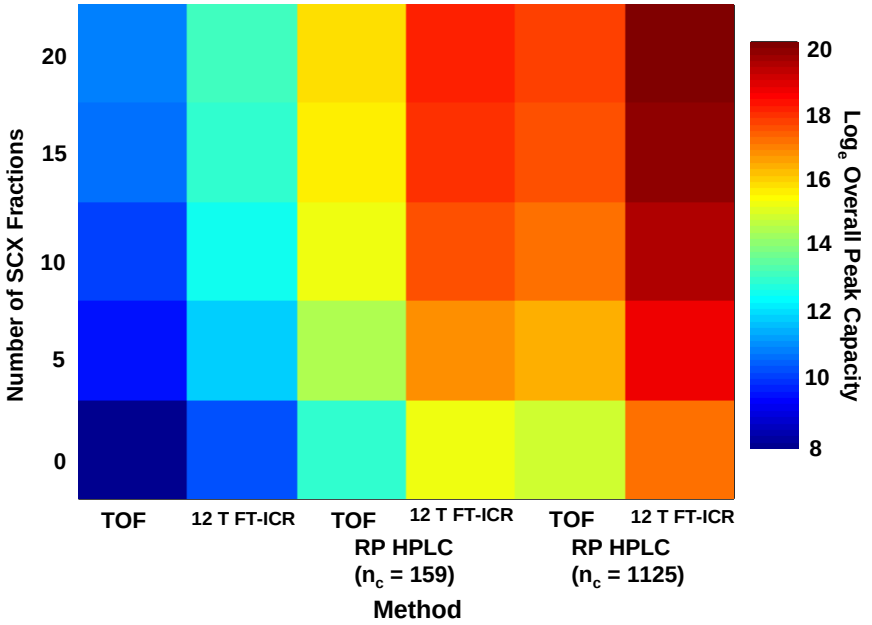


Figure 2.6 Heat map representing total peak capacity for each of the different bottom-up proteomics methods. The first two columns (TOF and 12 T FT-ICR) correspond to direct infusion experiments (y = 0 SCX fractions) and 1 D LC-MS experiments using SCX as the separation tool (y = 5, 10, 15, 20 SCX fractions). The next two columns (RP-HPLC-TOF and RP-HPLC-12 T FT-ICR) are 1 D RP-LC-MS methods utilizing the RP-LC method with a peak capacity of 159 described in the text. As the number of SCX fractions increases from 0 (1 D RP-LC-MS) to 5-20, the peak capacity values are for the 2 D-SCX-RP-LC-MS method. Similarly, the last two columns describe the 1 D-RP-LC-MS (y = 0 SCX fractions) and 2 D-SCX-RP-LC-MS (y= 5, 10, 15, 20 SCX fractions) using the RP-LC method with a peak capacity of 1125 described in the text. The scale on the right indicates the natural log values of the peak capacity.

Owing to the fact the peak capacities for RP-HPLC and SCX used to calculate peak capacities for 1 D-LC MS and 2 D-LC MS for both mass analyzers are the same, the trends described above are valid for both FT-ICR and TOF mass spectrometers. However, an order of magnitude improvement in peak capacity is attained when the LC method(s) is coupled to a FT-ICR mass spectrometer compared to a TOF mass analyzer.

The method with the highest peak capacity is SCX (20 fractions) followed by RP-HPLC separation utilizing a 80 cm column coupled to 12 T FT-ICR MS analysis. However, the practicality of long columns is mitigated by the requirement of pumps that can handle

increased back pressures of 10,000 psi or greater. The relationship between column length (*L, cm*), mobile-phase viscosity (η), flow rate (*F, mL/min*), particle diameter (d_p, *μm*), column inner diameter (d_c, *cm*), and the pressure (*P, psi*) is shown in **Equation 2.17**. Furthermore, gradient run times are several hours long [36]. The trade-off between improved peak capacity and longer analysis times can be realized in terms of the peak capacity production rate.

$$P = \frac{250 L \eta F}{d_p^2 d_c^2} \quad (\mathbf{2.17})$$

2.2.4 Peak Capacity Production Rate for Bottom-Up Proteomics

The peak capacity production rate is the peak capacity relative to the total analysis time. This metric is useful to assess the consequence of increasing analysis time. Increasing the gradient time often improves the peak capacity [37]. However, certain studies, such as differential proteomics studies, necessitate large numbers of samples and repeated measures in order to have the statistical power to discern differences in populations [42, 43]. Therefore, there is a balance between improvements in peak capacity and increases in analysis time that should be considered during method development.

In order to calculate the peak capacity production rate, the calculated peak capacity was divided by the total analysis time [44], which were taken from the reported total gradient times [1, 36]. Additionally, each SCX fraction was assumed to contribute one minute to the total analysis times. Peak capacity production rates for the various methods are shown in **Figure 2.7**. The same general trends described for peak capacity are also applicable to peak capacity production rate.

The differences in the peak capacity production rate for FT-ICR and TOF are a direct result of differences in peak capacity (*i.e.,* resolving power), since the mass analyzers essentially have the same acquisition time. Therefore, any method utilizing FT-ICR mass

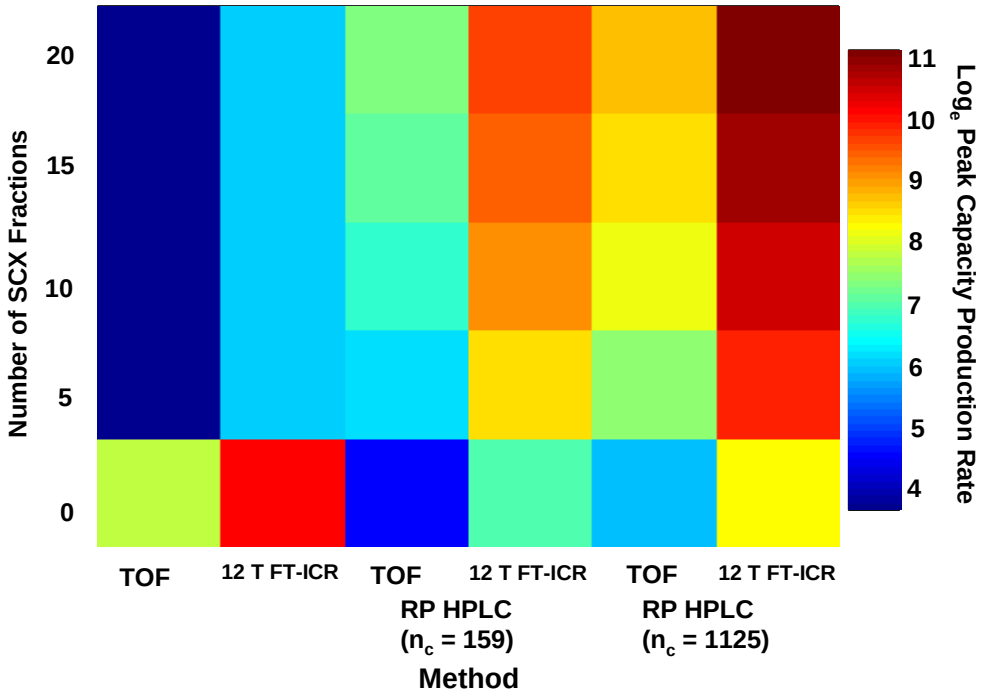


Figure 2.7 Heat map representing the peak capacity production rates for each of the different bottom-up proteomics methods. Notation is the same as in **Figure 2.6**. The scale on the right indicates the natural log values of the peak capacity production rate.

spectrometry analysis will have an order of magnitude higher peak capacity production rate over TOF MS. Accordingly, 10-fold more components can be analyzed by a 1 D or 2 D LC method coupled to FT-ICR mass spectrometry in the same amount of time as a 1 D or 2 D LC method coupled to TOF!

Clearly, the method with the highest peak capacity production rate is 20 SCX fractions followed by RP-HPLC separation with 80 cm column and 12T FT-ICR mass spectrometry. However, when a second separation dimension is employed (SCX fractionation) nano-LC methods begin to compete with 1 D RP-HPLC separations using longer columns. The peak capacity production rate for SCX (5 fractions)-RP-HPLC (n_c = 159)-12 T FT-ICR MS exceeds that of RP-HPLC-(n_c = 1125)-12 T FT-ICR MS, which were calculated to be 4889 and 3967, respectively. In essence, the separation efficiency of methods utilizing nano-LC can be improved by employing SCX fractionation (or another dimension of separation), thus eliminating special instrumentation modifications required to accommodate high back pressures of RP-HPLC separations using longer columns.

Furthermore, the addition of an orthogonal separation method provides another analytical constraint for peptide identification. Therefore, in 2 D-LC MS analysis one could use *m/z*, RP-HPLC retention time, and SCX fraction number to facilitate peptide identification.

Moreover, the peak capacity production rate can be improved by decreasing the analysis time, which is undoubtedly due to chromatography (evidenced by the comparison of the peak capacity production rate for direct infusion and 1 D- and 2 D-LC MS in **Figure 2.7**). Significant reductions in chromatography separation times for the separation of large molecules without an accompanying increase in back pressures have been achieved via high-temperature HPLC, research originating in the Horváth laboratory [45-47]. More recently the idea of high temperatures has been applied to 2 D-LC with improved column stability [48, 49].

2.3 Top-Down Proteomics

2.3.1 Defining the Top-Down Proteomics Mass Space

Based on the same concepts developed to describe the mass space occupied by singly charged tryptic peptides (*vide supra*), the mass space occupied by intact proteins can be defined. The mass distribution of singly charged intact proteins is shown in **Figure 2.8**. In **Figure 2.8 A**, which displays the mass distribution of intact proteins over the mass range of 2,500 – 100,000 Da, the highly populated regions (shown in red) are generally in the region less than 15,000 Da; the mass region from 2,500 – 20,000 is enlarged and shown in **Figure 2.8 B**.

Figure 2.9 displays the plot of the mass excess for intact proteins plotted as a function of monoisotopic mass. **Equation 2.18** is the equation for the main regression line.

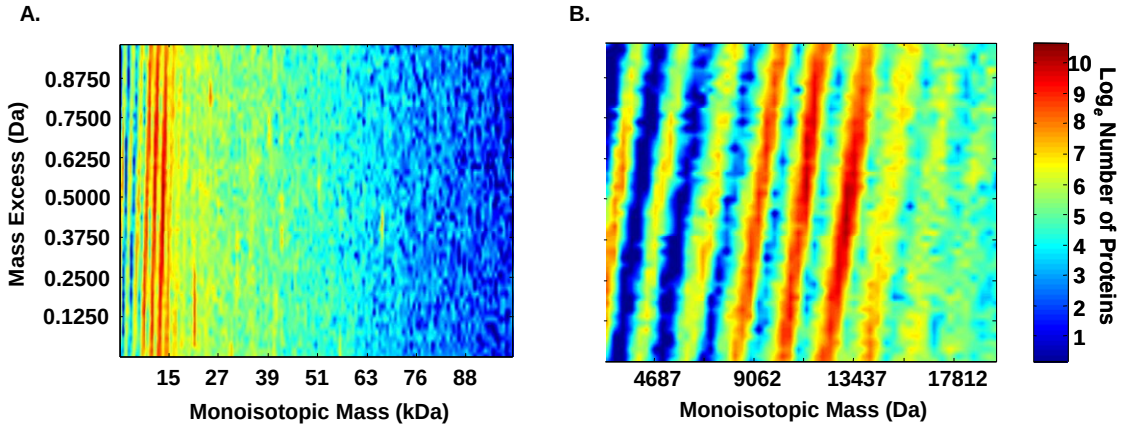


Figure 2.8 Mass excess distribution of intact proteins from NCBI's nr database as a function of monoisotopic mass for the mass region 2,500 - 100,000 Da (**A**) with the region 2,500 – 20, 000 (**B**) enlarged on the right. The mass excess was calculated by summing the monoisotopic masses of each amino acid in the sequence and then subtracting the nominal mass, thus resulting in values less than one. Red regions indicate highly populated areas. Blue regions are unpopulated areas or "forbidden zones." The scale on the right indicates the natural logarithm number of proteins for each color.

$$\Delta m = 5.0702 \times 10^{-4} m - 0.0839 \qquad (\mathbf{2.18})$$

By subtracting the upper and lower boundaries (represented by dashed lines in **Figure 2.9**), which encompass approximately 99% of the data, we obtain the width equation (**Equation 2.19**) for intact proteins:

$$W(m) = 6.2200 \times 10^{-5} m + 0.9176 \qquad (\mathbf{2.19})$$

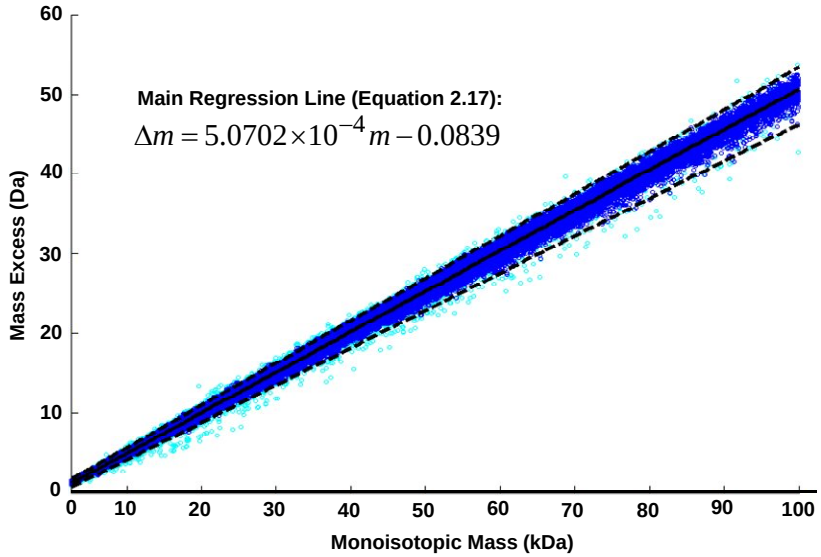


Figure 2.9 Linear distribution of intact proteins in the mass space. Same intact proteins shown in **Figure 2.8**, however mass excess is plotted as a continuous function. Mass excess in this plot is a summation of mass excess (*i.e.,* nominal mass subtracted from monisotopic mass) for each amino acid in the sequence, which resulted in values from 0 to 60. The solid line indicates the main regression line, equation shown. The dotted lines contain ~99% of the data; equations of these lines were used to calculate the width equation (**Equation 2.19**).

2.3.2 Peak Capacity in FT-ICR and TOF Mass Spectrometry for Top-Down Proteomics

In order to determine the peak capacity for top-down proteomics, according to **Equation 2.10**, the relationship between the number of isotopes and monoisotopic mass for intact proteins must first be developed. Following the same procedure outlined in **Section 2.2.2**, *averagine* was used to generate chemical formulas for theoretical proteins with molecular weights of 500 – 100, 000 Da. Subsequently, the number of isotopes with an abundance of greater than 5% was plotted as a function of the monoisotopic mass; the resulting relationship is shown in **Equation 2.20:**

$$n_I = 0.0794m^{0.548} \qquad (\mathbf{2.20})$$

The relationship between the number of isotopes in an isotopic distribution and monoisotopic mass is no longer linear after 10,000 Da, which is a consequence of the greater number of possible combinations of isotopes. The probability of detecting each isotope decreases as the number of isotopic combinations increases; this concept is illustrated in **Figure 2.10**.

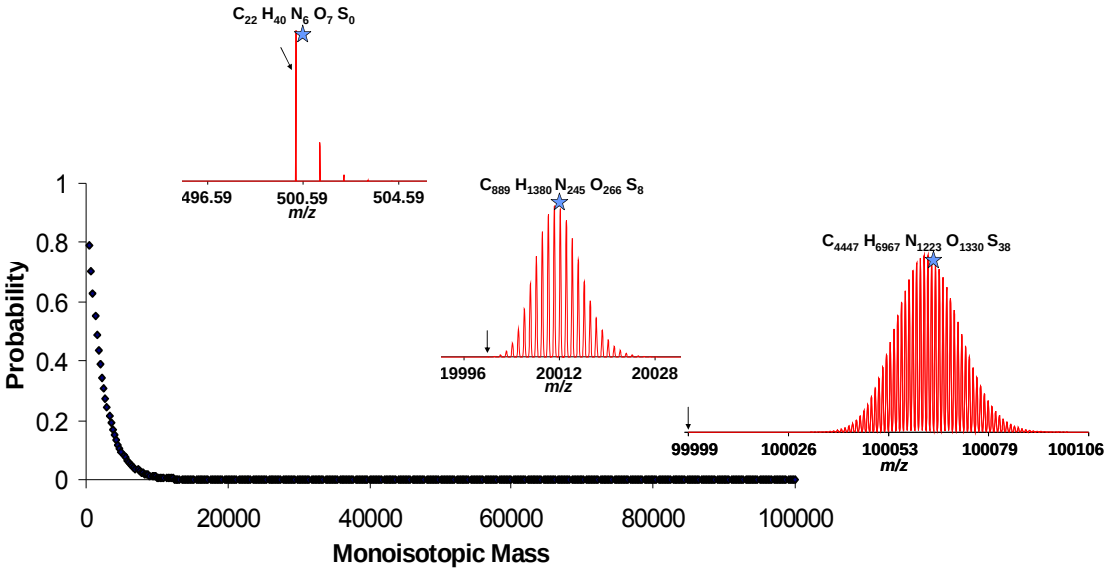


Figure 2.10 Probability of detecting the monoisotopic peak as a function of monoisotopic mass. Representative isotopic distributions for proteins generated using *averagine* are inset. Arrows indicate the *m/z* value of the monoisotopic peak. As the mass increases, there is a larger difference in the monoisotopic mass and the average mass (indicated by the star), which is known as the isotopic shift.

By substituting the equation for the number of isotopes (**Equation 2.20**) into the general peak capacity equation (**Equation 2.10**), the resulting peak capacity equations for FT-ICR and TOF mass spectrometry for top-down proteomics become:

$$n_c = -18638.8792 B_0 m^{-0.548} - 9.739 \times 10^7 B_0 m^{-1.548} \qquad (\mathbf{2.21})$$

$$n_c = 14.2952 m^{0.452} - 210888.23108 m^{-0.548} \qquad (\mathbf{2.22})$$

Thus, the total peak capacity for top-down proteomics afforded to FT-ICR and TOF mass analyzers can be determined by integrating **Equations 2.21** and **2.22** over the mass range of 2,500 – 100,000 Da, which is approximately 5300 (7 T FT-ICR), 7100 (9.4 T FT-ICR), 9100 (12 T FT-ICR) and 4600 (TOF). The dramatic decrease in peak capacity is a direct result of the reduced resolving power at higher *m/z* values (FT-ICR) and the increased number of isotopes in an isotopic distribution, which affects both FT-ICR and TOF mass analyzers.

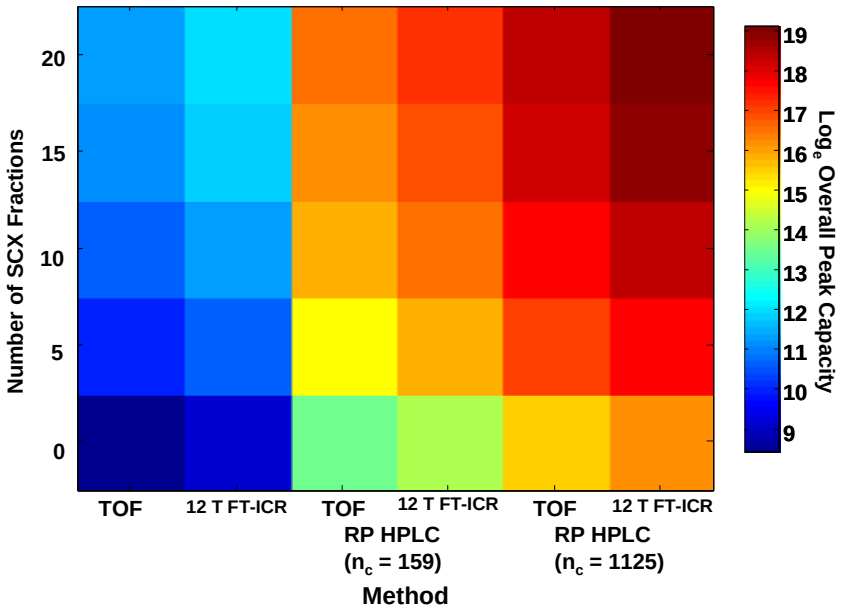


Figure 2.11 Heat map representing total peak capacity for each of the different top-down methods. The first two columns (TOF and 12 T FT-ICR) correspond to direct infusion experiments (y = 0 SCX fractions) and 1 D LC-MS experiments using SCX as the separation tool (y = 5, 10, 15, 20 SCX fractions). The next two columns (RP-HPLC-TOF and RP-HPLC-12 T FT-ICR) are 1 D RP-LC-MS methods utilizing the RP-LC method with a peak capacity of 159 described in the text. As the number of SCX fractions increases from 0 (1 D RP-LC-MS) to 5-20, the peak capacity values represent the 2 D-SCX-RP-LC-MS methods. Similarly, the last two columns describe the 1 D-RP-LC-MS (y = 0 SCX fractions) and 2 D-SCX-RP-LC-MS (y = 5, 10, 15, 20 SCX fractions) using the RP-LC method with a peak capacity of 1125 described in the text. The scale on the right indicates the natural log values of the peak capacity.

2.3.3 Peak Capacity and Peak Capacity Production Rate of 1 D- and 2 D-LC MS for Top-Down Proteomics

The calculated theoretical peak capacity values for top-down proteomics are shown in **Figure 2.11**. These values were calculated as described in **Section 2.2.3**, and follow the same trends observed for bottom-up proteomics. Similarly, the peak capacity production rates for the different top-down proteomics methods are shown in **Figure 2.12**, which also follow the same trend at the peak capacity production rates for bottom-up proteomics.

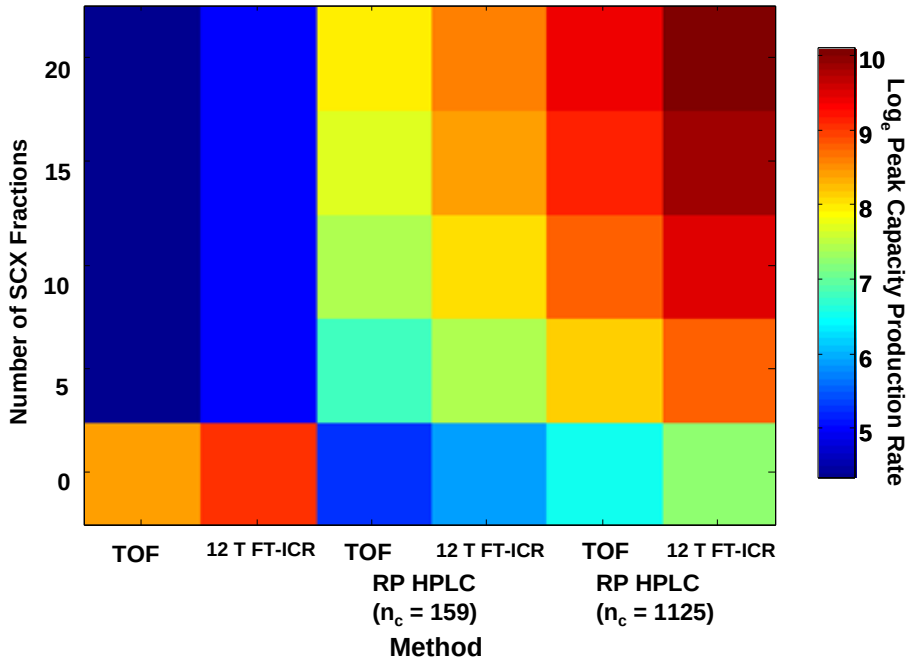


Figure 2.12 Heat map representing the peak capacity production rates for each of the different top-down proteomics methods. Notation is the same as in **Figure 2.11**. The scale on the right indicates the natural log values of the peak capacity production rate.

2.4 Peak Capacity Limitations Due to ICR Cell Charge Capacity

It should be mentioned that the theoretical maximum peak capacity in FT-ICR is restricted by the limit of charges that can be contained in an ICR cell; other investigators have defined the peak capacity in FT-ICR based solely on the charge capacity of the cell [50]. The maximum number of charges that can be contained within the confines of a 12 T cylindrical ICR cell (n_{max}) is 2×10^8 ions [51], calculated using the average mass of a tryptic

fragment. However, peak coalescence from space-charge effects occurs with greater than 10^7 charges in a 12 T ICR cell, and increases with the square of the magnetic field strength [52, 53]. Therefore, the maximum number of charges is reduced to this number for analytical FT-ICR mass spectrometry. Considering that a minimum of ~200 ions are required for detection of a single isotope [54], a single isotopic distribution is estimated to contain approximately 2000 charges per isotopic distribution. This equates to roughly 5000 species as the upper limit before space-charge effects ensue in a 12 T FT-ICR cell. This crude approximation is supported by experimental evidence from the Smith laboratory, who detected 3,000 species in a single mass spectrum [36], and the Marshall laboratory, who detected 8,000 different elemental compositions from crude oil with a 9.4 T system [55].

The value n_{max} increases with the square of the magnetic field strength. When n_{max} exceeds n_c then the peak capacity is fully realized, otherwise the peak capacity is limited by the charge capacity of the ICR cell. At a magnetic field strength of 12 T or greater n_{max} exceeds n_c, meaning that the available peak capacity can be fully realized, which further lends to the argument for higher magnetic fields.

2.5 How Does Multiple Charging Affect Peak Capacity in Mass Spectrometry?

Until now, only singly charged tryptic peptides and proteins have been considered. Once multiply-charged species are considered the percentage of occupied space dramatically increases (see **Figure 2.13**). In **Figure 2.13**, the mass excess was calculated for tryptic peptides by subtracting the nominal mass from the monoisotopic mass of the charge state. The number of possible charges a mass could carry followed these rules: z = 1^+ for masses 500 to <1000, z = 1^+, 2^+, 3^+ for masses ≥1000 to <2000, z = 2^+, 3^+, 4^+ for masses ≥2000 to <3000, z = 2^+, 3^+, 4^+, 5^+ for masses ≥3000 to <4000.

When tryptic fragments are allowed to adopt multiple charges, essentially every mass excess value becomes populated over the mass range (unoccupied space is white in

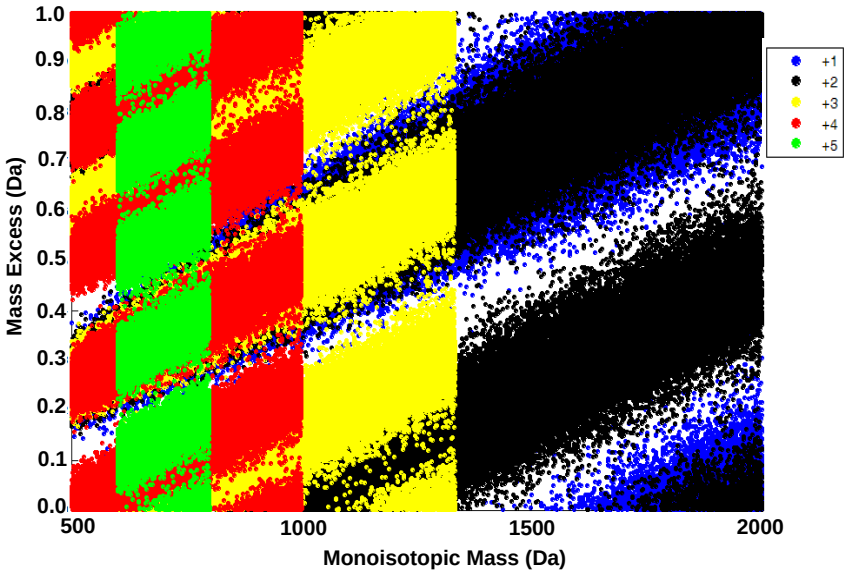


Figure 2.13 Relationship between the monoisotopic mass and the mass excess for multiply-charged ions of the tryptic peptides shown in **Figure 2.2**. Each charge state is plotted in a different color, shown in the legend. The blue points (1+ charge state) have the same distribution as in **Figure 2.2**. White areas represent unoccupied space. Approximately 95% of the mass space from 500 – 5000 Da is occupied by multiply-charged tryptic peptides.

Figure 2.13). For example, a tryptic fragment with a neutral monoisotopic mass of 1999.9840 will have *m/z* values of 1000.9993, 667.6686, and 501.0033 for the 2^+, 3^+, and 4^+ charge states, respectively. The mass excesses of the 2^+, 3^+, and 4^+ charge states fall in regions that are not populated by singly charged peptides (the blue regions in **Figure 2.2**). Consequently, an ionization method that produces multiply-charged proteolytic fragments (*e.g.*, electrospray ionization) will allow greater utilization of the *m/z* space compared to singly charged proteolytic fragments. It follows that the width of the mass distribution (i.e., width equation) will encompass the entire monoisotopic mass vs. mass excess plot. Therefore, the peak capacity of an ESI-FT-ICR bottom-up proteomics experiment becomes defined solely by the mass resolving power and charge-capacity of the ICR cell, since peak capacity is no longer restricted by the occupied space. Thus, the peak capacity in mass spectrometry for multiply-charged species is dramatically improved. The same reasoning is valid for top-down proteomics, as well.

2.6 Conclusions

Although, the peak capacity of top-down proteomics is a smaller number than the peak capacity for bottom-up proteomics it is important to consider the total number of "components" in each experiment. In a bottom-up proteomics experiment the total number of components is the total number of tryptic fragments, which is 4,596,261 (considering all proteins in the *nr* database). Whereas in a top-down proteomics experiment the total number of components is the number of intact proteins (119,771). A TOF mass analyzer affords a maximum of 0.05% and 4% proteome coverage for bottom-up and top-down proteomics, respectively. Proteome coverage is improved to 0.54% and 8% for bottom-up and top-down proteomics when a 12 T FT-ICR mass analyzer is employed. For identification the discrepancy in proteome coverage between bottom-up and top-down proteomics is not a problem, since an unique protein identification can be made if the identity of 4 contiguous amino acids contained within at least an 11-amino acid sequence are known and the accurate mass (within 10 ppm) is achieved [56]. However, the reduced proteome coverage of bottom-up proteomics becomes a problem for the detection of post-translational modifications. Since only a small portion of the total number of tryptic fragments can potentially be detected, many post-translational modifications will go undetected. Therefore, the overall peak capacity afforded to bottom-up and top-down proteomics can not be used to determine which approach one should take in complex mixture analysis. However, the difference in proteome coverage between the two mass analyzers is important.

The overall peak capacity can be utilized to determine the appropriate mass analyzer and separation methods. The complexity of biological matrices studied in proteomics experiments warrants technologies that can handle both the number and large dynamic range of the components. Liquid chromatography separations coupled with mass spectrometry presents itself as the ideal platform to handle complex-mixture analysis,

especially FT-ICR MS analysis. Inherent with the sensitivity, mass resolving power, mass measurement accuracy, and dynamic range of FT-ICR mass spectrometry is a high peak capacity, which offers an order of magnitude higher peak capacity over TOF. Moreover, when coupled to RP-HPLC, the peak capacity is improved 100-fold. Utilization of a second, orthogonal separation step further improves peak capacity, making complex mixture analysis attainable.

2.7 References

1. Johnson, KL, Mason, CJ, Muddiman, DC, and Eckel, JE. 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. *Analytical Chemistry,* 2004. **76**: 5097-5103.

2. Nepomuceno, AI, Muddiman, DC, Bergen, HR, Craighead, JR, Burke, MJ, Caskey, PE, et al. Dual electrospray ionization source for confident generation of accurate mass tags using liquid chromatography Fourier transform ion cyclotron resonance mass spectrometry. *Analytical Chemistry,* 2003. **75**: 3411-3418.

3. Muddiman, DC and Oberg, AL. Statistical evaluation of internal and external mass calibration laws utilized in Fourier transform ion cyclotron resonance mass spectrometry. *Analytical Chemistry,* 2005. **77**: 2406-2414.

4. Wong, RL and Amster, IJ. Sub part-per-million mass accuracy by using stepwise-external calibration in Fourier transform ion cyclotron resonance mass spectrometry. *Journal of the American Society for Mass Spectrometry,* 2006. **17**: 1681-1691.

5. Williams, DK and Muddiman, DC. 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. *Analytical Chemistry,* submitted.

6. Zubarev, RA, Hakansson, P, and Sundqvist, B. Accuracy requirements for peptide characterization by monoisotopic molecular mass measurements. *Analytical Chemistry,* 1996. **68**: 4060-4063.

7. Jensen, ON, Podtelejnikov, AV, and Mann, M. Identification of the components of simple protein mixtures by high accuracy peptide mass mapping and database searching. *Analytical Chemistry,* 1997. **69**: 4741-4750.

8. Clauser, KR, Baker, P, and Burlingame, AL. Role of accurate mass measurement (+/- 10 ppm) in protein identification strategies employing MS or MS/MS and database searching. *Analytical Chemistry,* 1999. **71**: 2871-2882.

9. Conrads, TP, Anderson, GA, Veenstra, TD, Pasa-Tolic, L, and Smith, RD. Utility of accurate mass tags for proteome-wide protein identification. *Analytical Chemistry,* 2000. **72**: 3349-3354.

10. Winnik, WM. Continuous ph/salt gradient and peptide score for strong cation exchange chromatography in 2D-nano-LC/MS/MS peptide identification for proteomics. *Analytical Chemistry,* 2005. **77**: 4991-4998.

11. Eng, JK, Mccormack, AL, and Yates, JR. An approach to correlate tandem mass-spectral data of peptides with amino-acid-sequences in a protein database. *Journal of the American Society for Mass Spectrometry,* 1994. **5**: 976-989.

12. Petritis, K, Kangas, LJ, Ferguson, PL, Anderson, GA, Pasa-Tolic, L, Lipton, MS, et al. Use of artificial neural networks for the accurate prediction of peptide liquid

chromatography elution times in proteome analyses. *Analytical Chemistry,* 2003. **75**: 1039-1048.

13. Kendrick, E. A mass scale based on CH2=14.0000 for high resolution mass spectrometry of organic compounds. *Analytical Chemistry,* 1963. **35**: 2146-&.

14. Inczedy, J, Lengyel, T, and Ure, AM. *IUPAC compendium of analytical nomenclature: The orange book*. 3rd ed. 1998, Oxford: Blackwell Scientific Publications. 964.

15. Mann, M. *Useful tables of possible and probable peptide masses*. in *The 43rd ASMS Conference on Mass Spectrometry and Allied Topics*. 1995. Atlanta, GA: ASMS.

16. Lehmann, WD, Bohne, A, and von der Lieth, CW. The information encrypted in accurate peptide masses - Improved protein identification and assistance in glycopeptide identification and characterization. *Journal of Mass Spectrometry,* 2000. **35**: 1335-1341.

17. Zhang, X, Hines, W, Adamec, J, Asara, JM, Naylor, S, and Regnier, FE. An automated method for the analysis of stable isotope labeling data in proteomics. *Journal of the American Society for Mass Spectrometry,* 2005. **16**: 1181-1191.

18. Juhasz, P and Martin, SA. The utility of nonspecific proteases in the characterization of glycoproteins by high-resolution time-of-flight mass spectrometry. *International Journal of Mass Spectrometry,* 1997. **169**: 217-230.

19. Benson, DA, Karsch-Mizrachi, I, Lipman, DJ, Ostell, J, and Wheeler, DL. Genbank: Update. *Nucleic Acids Research,* 2004. **32**: D23-D26.

20. Berman, HM, Westbrook, J, Feng, Z, Gilliland, G, Bhat, TN, Weissig, H, et al. The protein data bank. *Nucleic Acids Research,* 2000. **28**: 235-242.

21. Boeckmann, B, Bairoch, A, Apweiler, R, Blatter, MC, Estreicher, A, Gasteiger, E, et al. The SWISS-prot protein knowledgebase and its supplement TREMBL in 2003. *Nucleic Acids Research,* 2003. **31**: 365-370.

22. Gasteiger, E, Gattiker, A, Hoogland, C, Ivanyi, I, Appel, RD, and Bairoch, A. Expasy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research,* 2003. **31**: 3784-3788.

23. Wu, CH, Yeh, LSL, Huang, HZ, Arminski, L, Castro-Alvear, J, Chen, YX, et al. The protein information resource. *Nucleic Acids Research,* 2003. **31**: 345-347.

24. Jones, JJ, Stump, MJ, Fleming, RC, Lay, JO, and Wilkins, CL. Strategies and data analysis techniques for lipid and phospholipid chemistry elucidation by intact cell MALDI-FTMS. *Journal of the American Society for Mass Spectrometry,* 2004. **15**: 1665-1674.

25. Brock, A, Mason, DE, Kuan, FC, Ficarro, SB, and Peters, EC. *Orthogonal mass space calibrants for MALDI-FT-ICRMS applications in proteomics*. in *53rd ASMS Conference on Mass Spectrometry and Allied Topics*. 2005. San Antonio, TX.

26. Hall, MP, Ashrafi, S, Obegi, I, Petesch, R, Peterson, JN, and Schneider, LV. 'mass defect' tags for biomolecular mass spectrometry. *Journal of Mass Spectrometry,* 2003. **38**: 809-816.

27. Matthiesen, R, Bauw, G, and Welinder, KG. Use of performic acid oxidation to expand the mass distribution of tryptic peptides. *Analytical Chemistry,* 2004. **76**: 6848-52.

28. Giddings, JC. *Unified separation science*. 1991, New York: Wiley. 320.

29. Marshall, AG, Hendrickson, CL, and Jackson, GS. Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrometry Reviews,* 1998. **17**: 1-35.

30. Senko, MW, Beu, SC, and McLafferty, FW. Determination of monoisotopic masses and ion populations for large biomolecules from resolved isotopic distributions. *Journal of the American Society for Mass Spectrometry,* 1995. **6**: 229-233.

31. Krause, E, Wenschuh, H, and Jungblut, PR. The dominance of arginine-containing peptides in MALDI-derived tryptic mass fingerprints of proteins. *Analytical Chemistry,* 1999. **71**: 4160-4165.

32. King, R, Bonfiglio, R, Fernandez-Metzler, C, Miller-Stein, C, and Olah, T. Mechanistic investigation of ionization suppression in electrospray ionization. *Journal of the American Society for Mass Spectrometry,* 2000. **11**: 942-950.

33. Annesley, TM. Ion suppression in mass spectrometry. *Clinical Chemistry,* 2003. **49**: 1041-1044.

34. Pan, P and McLuckey, SA. The effect of small cations on the positive electrospray responses of proteins at low ph. *Analytical Chemistry,* 2003. **75**: 5468-5474.

35. Ledford, EB, Rempel, DL, and Gross, ML. Space-charge effects in Fourier-transform mass-spectrometry - Mass calibration. *Analytical Chemistry,* 1984. **56**: 2744-2748.

36. Shen, YF, Zhao, R, Belov, ME, Conrads, TP, Anderson, GA, Tang, KQ, et al. Packed capillary reversed-phase liquid chromatography with high-performance electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry for proteomics. *Analytical Chemistry,* 2001. **73**: 1766-1775.

37. Snyder, LR, Kirkland, JJ, and Glajch, JL. *Practical HPLC method development*. 2nd ed. 1997, New York: John Wiley & Sons. xxvi, 765 p.

38. Neue, UD. Theory of peak capacity in gradient elution. *Journal of Chromatography A,* 2005. **1079**: 153-161.

39. Rozing, G, Nagele, E, Horth, P, Vollmer, M, Moritz, R, Glatz, B, et al. Instrumentation for advanced micro separations in pharmaceutical analysis and proteomics. *Journal of Biochemical and Biophysical Methods,* 2004. **60**: 233-263.

40. Gilar, M, Daly, AE, Kele, M, Neue, UD, and Gebler, JC. Implications of column peak capacity on the separation of complex peptide mixtures in single- and two-dimensional high-performance liquid chromatography. *Journal of Chromatography A,* 2004. **1061**: 183-92.

41. Dolan, JW, Snyder, LR, Wolcott, RG, Haber, P, Baczek, T, Kaliszan, R, et al. Reversed-phase liquid chromatographic separation of complex samples by optimizing temperature and gradient time III. Improving the accuracy of computer simulation. *Journal of Chromatography A,* 1999. **857**: 41-68.

42. Bergen, HR, Vasmatzis, G, Cliby, WA, Johnson, KL, Oberg, AL, and Muddiman, DC. Discovery of ovarian cancer biomarkers in serum using nanoLC electrospray ionization TOF and FT-ICR mass spectrometry. *Disease Markers,* 2003. **19**: 239-249.

43. McDonald, WH and Yates, JR. Shotgun proteomics and biomarker discovery. *Disease Markers,* 2002. **18**: 99-105.

44. Grall, A, Leonard, C, and Sacks, R. Peak capacity, peak-capacity production rate, and boiling point resolution for temperature-programmed GC with very high programming rates. *Analytical Chemistry,* 2000. **72**: 591-598.

45. Kalghatgi, K and Horvath, C. Rapid analysis of proteins and peptides by reversed-phase chromatography. *Journal of Chromatography,* 1987. **398**: 335-339.

46. Antia, FD and Horvath, C. High-performance liquid-chromatography at elevated-temperatures - examination of conditions for the rapid separation of large molecules. *Journal of Chromatography,* 1988. **435**: 1-15.

47. Chen, H and Horvath, C. High-speed high-performance liquid-chromatography of peptides and proteins. *Journal of Chromatography A,* 1995. **705**: 3-20.

48. Yang, XQ, Ma, LJ, and Carr, PW. High temperature fast chromatography of proteins using a silica-based stationary phase with greatly enhanced low ph stability. *Journal of Chromatography A,* 2005. **1079**: 213-220.

49. Stoll, DR and Carr, PW. Fast, comprehensive two-dimensional HPLC separation of tryptic peptides based on high-temperature HPLC. *Journal of the American Chemical Society,* 2005. **127**: 5034-5035.

50. Smith, RD, Anderson, GA, Lipton, MS, Pasa-Tolic, L, Shen, YF, Conrads, TP, et al. An accurate mass tag strategy for quantitative and high-throughput proteome measurements. *PROTEOMICS,* 2002. **2**: 513-523.

51. Wineland, DJ, Bollinger, JJ, Itano, WM, and Prestage, JD. Angular-momentum of trapped atomic particles. *Journal of the Optical Society of America B-Optical Physics,* 1985. **2**: 1721-1730.

52. Bollinger, JJ, Heinzen, DJ, Moore, FL, Itano, WM, Wineland, DJ, and Dubin, DHE. Electrostatic modes of ion-trap plasmas. *Physical Review A,* 1993. **48**: 525-545.

53. Mitchell, DW and Smith, RD. Cyclotron motion of two coulombically interacting ion clouds with implications to Fourier-transform ion-cyclotron resonance mass-spectrometry. *Physical Review E,* 1995. **52**: 4366-4386.

54. Limbach, PA, Grosshans, PB, and Marshall, AG. Experimental-determination of the number of trapped ions, detection limit, and dynamic-range in Fourier-transform ion-cyclotron resonance mass-spectrometry. *Analytical Chemistry,* 1993. **65**: 135-140.

55. Hughey, CA, Rodgers, RP, and Marshall, AG. Resolution of 11 000 compositionally distinct components in a single electrospray ionization Fourier transform ion cyclotron resonance mass spectrum of crude oil. *Analytical Chemistry,* 2002. **74**: 4145-4149.

56. Hunt, DF, Yates, JR, Shabanowitz, J, Winston, S, and Hauer, CR. Protein sequencing by tandem mass-spectrometry. *Proceedings of the National Academy of Sciences of the United States of America,* 1986. **83**: 6233-6237.

CHAPTER 3

Mass Excess Labeled Peptides for the Assessment of Mass Measurement Accuracy and Retention Time Reproducibility and their Application to Label-Free Proteomics

3.1 Introduction

Proteomics [1], by definition, is the venture to identify, quantify, and characterize the structure and function of all proteins in given system (*e.g.,* organism, organelle) including spatial, temporal, physiological variations [2]. Early proteomics efforts have focused on the identification and structural and functional characterization of proteins, as evidenced by the growing numbers of protein databases that include basic information (*e.g.,* primary sequence, molecular weight) and structural information (*e.g.,* tertiary structural motifs, crystallographic structures) [3]. Mass spectrometry is the ideal platform to further the quantitative proteomics objectives due to its ability to measure an inherent molecular property (specificity), sensitivity, and afforded dynamic range.

Quantitative proteomics measures either absolute (*e.g.,* copy number, concentration determined with a known amount of a standard) or relative (*i.e.,* measured as "fold change" compared to some identified standard) changes in protein concentration. Absolute quantification methods involve the addition of a known quantity of a stable-isotope labeled synthetic peptide as the internal standard. One of the first absolute quantification methods employed a stable-isotope standard of methionine enkephalin, a morphine-like neuropeptide [4]. The relative ion currents of the natural and stable-isotope labeled methionine enkephalin were compared to determine absolute quantification [4]. This report demonstrated two important features necessary for quantitative mass spectrometry methods; first, molecular specificity, which was not afforded by radioimmunoassay methods (the current methodology), and second, improved sensitivity of electrospray ionization over fast-atom bombardment and other ionization technologies [4]. Additionally, this study utilized off-line

reversed-phase fractionation of pituitary tissue for targeted absolute quantification. In one of the first demonstrations of a global, absolute quantification method Hunt, *et al.* demonstrated on-line LC-ESI-MS analysis of peptides bound to MHC HLA-A2.1, which is important for immunological response [5]. Absolute quantification was achieved with the comparison of ion currents from endogenous peptides to several synthetic peptides added at known concentrations [5]. Although this method targeted MHC-bound peptides, it is a "global" quantification method because all of the peptides identified were quantified. Absolute quantification methods are generally applicable only to targeted proteomics analyses, and even then with limited scope, since the development of an appropriate internal standard can be time-consuming and expensive [6, 7].

The limitations of absolute quantification methods led to the development of relative quantification methods, which rely on a comparison of abundances of a "light" (natural) to a "heavy" (stable-isotope label) derivative of the peptide/protein of interest for both global and targeted proteomics measurements. Current methods for relative quantitative proteomics focus on global labeling strategies. Chemical tagging is an effective means to universally introduce a stable isotope label. The repertoire of chemical tagging strategies for relative quantification in proteomics is ever-expanding; some strategies are briefly described below. One can use alkylating reagents [8] to target specific amino acids for chemical labeling, including ICAT™ reagent [9], "solid-phase ICAT" [10], cleavable ICAT [11], and IDBEST™ (Target Discovery) [12]. All of these reagents target cysteine residues, which have a frequency of 1.9% in human proteins [13]. The lower occurrence of cysteine, compared to other amino acids, can be exploited to separate cysteine-containing proteins from the remaining "chemical noise." For example, the solid-phase ICAT reagent contains a glass bead that can covalently bind to a solid phase surface, while the non-cysteinyl peptides can be washed away. However, labeling strategies that target a particular amino acid bias proteomic analyses to only proteins that contain the targeted and labeled amino acid.

Therefore, both the frequency of the amino acid in the organism and the reaction efficiency can affect the results.

A more systematic, global labeling strategy is to target primary amine groups (*i.e.*, N-termini and lysine side chains), which are present on every peptide [14-16]. A unique N-terminal labeling approach is the iTRAQ strategy because it allows for quantification of up to four different populations at the same time [17]. Additionally, all peptides have a C-terminal end; therefore, a labeling strategy that targets C-termini would also provide global labeling. Guanidination of C-terminal lysine for relative quantification is an example of a C-terminal tagging strategy [18]. The numerous chemical tagging approaches for quantitative proteomics are the subject of several reviews [19, 20].

Chemical tagging strategies are one means of globally introducing a stable-isotope label. Other methods utilize biological or enzymatic incorporation of the label. Biological incorporation was first demonstrated with ^{15}N labeling of plant cultures that were fed ^{15}N-labeled ammonium or nitrate [21]. As the organism synthesizes new proteins, the stable-isotope label is incorporated allowing for downstream quantification. This idea was later extended to bacterial and mammalian cultures, and biological incorporation of ^{15}N was achieved with the use of modified media [22]. Disadvantages of this method, including the expense for higher organisms and the requirement of sequence knowledge to discern mass shifts, led to the development of stable-isotope labeling of amino acids in cell culture (SILAC) [23]. In the SILAC method, stable-isotope labeled amino acids are added to the media, while the unlabeled counterparts are removed from the media [23]. Alternatively, the stable isotope label ^{18}O can be enzymatically incorporated from $H_2{}^{18}O$ water by trypsin [24-26]. While all peptides cleaved with trypsin are labeled, the complex data analysis, owing to composite isotope patterns, precludes general applicability of this method [27, 28].

In contrast, label-free proteomics methods utilize a comparison of the absolute ion abundance between samples to determine relative quantification [29, 30]. No stable-isotope

labeling of endogenous peptides (as in chemical tagging strategies) or their exogenous analogs (as done with internal standards that are spiked into samples) is required for label-free proteomics methods, thereby enabling comparisons between datasets collected on different days and even from different laboratories. Furthermore, the analysis of populations not amenable to *in vitro* expression systems (*e.g.,* human) is possible. Application of this approach has yielded new candidate biomarkers for ovarian cancer in the Muddiman laboratory [31].

Mass measurement accuracy and retention time reproducibility are key components in a label-free proteomics experiment, which lend to confident protein identifications. The Smith laboratory demonstrated this concept with the accurate mass and time tag approach [32, 33]. Accurate mass in our laboratory is achieved utilizing a dualESI source [34-36], by the implementation of recently-published calibration laws [37], or automatic gain control [38]. Earlier studies from the Muddiman laboratory demonstrate that retention time reproducibility of ± 54 s (confidence interval of the mean at the 95%) for 95% of 497 species detected in the low molecular weight fraction of serum can be obtained with a split-flow nanoLC system [30]. More recently, we have extensively evaluated the use of splitless-flow nanoLC and dramatically improved the retention time reproducibility to ± 6.3 s ($CIM_{95\%}$) for 19 myoglobin tryptic fragments detected in at least three of 24 injections [39]. Ultrafiltrates of human plasma and serum, complex mixtures, were demonstrated to have an analytical reproducibility on the same system of ± 22 and ± 18.5 s ($CIM_{95\%}$), respectively, for more than 300 species detected in all three replicates. Notably, we did not invoke chromatographic alignment in either of these investigations.

Peptides are quantified based on their ion abundance relative to one another in label-free proteomics. Abundance-normalization procedures reduce bias from systematic errors that are a consequence of sample preparation, experimentation, and instrumentation [40], thus increasing the likelihood that observed variations in ion abundance reflect true

biological differences (*i.e.,* improved accuracy). Generally, normalization is applied globally (to the entire dataset) or locally (to a subset of the data). This should not be confused with global and local normalization, which refer to the features used to determine the normalization factor [41]. In global normalization all the features are used to determine the normalization factor, whereas in local normalization only a subset of the features is utilized to determine the normalization factor [41]. Locally-applied normalization strategies are more common for microarray data [42]; to date only one report of locally-applied normalization for mass spectrometry data has been published [40]. Locally-applied normalization may be useful for comparisons of datasets generated from different laboratories in the future especially as quantitative proteomic databases are created. Several globally-applied normalization strategies for ion abundance have been published to date. These strategies are briefly discussed below.

One of the more widely-used normalization strategies is spectral counting, which is based on the linear relationship between protein abundance and the number of MS/MS mass spectra observed during data dependent acquisition [43]. Spectral counting is the sum of MS/MS spectra for a given protein [43]. The primary advantage of spectral counting is that there is no need for common peptides between LC-MS runs. However, the accuracy is highly dependent on the number of spectral counts and signal-to-noise ratios, therefore, larger differences in abundance ratios and signal-to-noise ratios are more accurate [43-45]. Old, *et al.* found that four or more spectral counts lend to accurate quantification [44]. In a comparison of spectral counting versus signal intensity for protein quantification, spectral counting was found to be more sensitive demonstrated by the fact that 68% of proteins were detected to have significant changes in expression versus only 39% determined by signal intensity; however, signal intensity is more accurate of the two methods [44].

In another comparison study Xia, *et al.* assessed four different normalization methods, including signal intensity at the peptide level, signal intensity at the protein level,

spectral counting at the peptide level, and spectral counting at the protein level [46]. All the proteolytic fragments that map to the same protein are included in the measure for the peptide level methods. Only the proteolytic fragments that map to single proteins (*i.e.,* unique peptides) are utilized in the protein level methods. Normalization is achieved to equalize the total ion abundance or the sum of spectral counts between the two samples for the signal intensity or spectral count methods, respectively [46]. The authors observed better correlation coefficients (indicating better reproducibility) for the spectral counting methods, except when the spectral counts were low (*i.e.,* less than 10) [46]. Furthermore, data normalized using protein level spectral counting was more precise [46], which has been reported by other groups, as well [47].

Natural log transformations of spectral counts and ion abundances have been also used during normalization procedures [40, 48, 49]. The advantage of the natural log transformation is that it generally produces normally-distributed data, thus enabling statistical analysis [48]. Furthermore, a natural log transformation reduces the influence of very abundant species on regression analysis [40]. Moreover, Zybailov, *et al.* normalized the spectral count to the length of the protein before natural log transformation to minimize the bias toward more abundant, which are typically larger, proteins that is typically observed [48].

More recently, Callister, *et al.* applied four different normalization procedures typically used for microarray data to LC-MS data, specifically central tendency normalization, linear regression, locally-weighted regression, and quantile normalization [40]. The mean was used to normalize peptide abundance ratios in the central tendency normalization strategy [40]. Linear regression normalization assumes that the systematic bias is linearly dependent on the magnitude of peptide abundance, and so the authors used the linear least squares regression equation from a plot of the ratio versus abundance to calculate the normalized peptide ratio [40]. Similarly, linear regression analysis was performed on subsets of the data using two dimensional data smoothing and the LOWESS algorithm in SigmaPlot to

normalize the data with the locally-weighted regression normalization procedure [40]. This procedure assumes that the systematic bias is nonlinearly dependent on the magnitude of peptide abundances [40]. Lastly, the authors evaluated a quantile-normalization procedure is a non-parametric means to normalize abundances based on rank [40]. This procedure has been implemented previously in another laboratory [50].

Post-translational modifications can affect normalization procedures. Zybailov, *et al.* noted that large standard deviations occurred when the peptide had been post-translationally modified [45]. Another group has reported a normalization strategy for phosphopeptides that normalizes the abundance of phosphopeptides to unphosphorylated peptides derived from the same protein [51]. This, however, assumes that the decrease in the abundance of an unmodified peptide results in a corresponding increase in the phosphorylated peptide, which may not necessarily be the case. In an attempt to account for variations in ionization and detection efficiency the authors introduce a "flyability ratio," which is dependent on sample preparation, spray conditions and instrument parameters [51].

One of the inherent limitations of label-free proteomics is the necessity of a peptide(s) that can be used to assess run-to-run variability of retention times and ion abundances, and therefore must be present in every run. Ideally, the peptide(s) would be a universally-applicable set of exogenous peptides with similar retention times compared to the peptides of interest and would not overlap in the mass space with endogenous peptides, ensuring that the mass and abundance information is not compromised. Since exogenous peptides are not subject to biological variation, they are preferred over endogenous peptides. There have been a few reports of procedures that utilize ion abundance(s) of one or more exogenous peptides to normalize the ion abundances of endogenous peptides found in human serum [52, 53].

Based on previous work in which we defined the mass space occupied by tryptic peptides [54] (**Chapter 2**), we designed and synthesized four internal standards containing

3,5-diiodotyrosine residues that have masses in the forbidden zones. Herein, we report the design of these iodopeptide standards and demonstrate their applicability to label-free proteomics to assess the retention time reproducibility and mass measurement accuracy. Furthermore, we delineate a local normalization strategy based on abundance-matching that may be more accurate than a global normalization strategy in complex mixtures containing proteins with a wide range of concentrations.

3.2 Experimental

3.2.1 Reagents

Apomyoglobin, trypsin, formic acid, guanidine HCl, and ammonium bicarbonate were purchased from Sigma-Aldrich (St. Louis, MO). HPLC-grade Burdick and Jackson brand water and acetonitrile were acquired thru VWR International (West Chester, PA). Fmoc-3,5-diiodo-D-tyrosine was obtained from AnaSpec (San Jose, CA). Reagents were used as received unless otherwise noted.

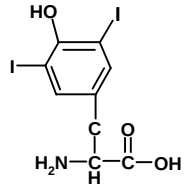
3.2.2 Tryptic Digest of Myoglobin

310 µL of 6 M guanidine HCl was added to 1 mg of apomyoglobin, and incubated for 5 minutes at 37°C until completely dissolved. Then, the solution was heated for 30 minutes at 50°C. After the solution cooled to room temperature, 1.7 times the total volume (527 µL) of water and 0.3 times the total volume (93 µL) of 10 X buffer (1 M ammonium bicarbonate) were added. The pH was between 8.0 - 8.5. Tryspin (1 µg/µL) was added to achieve a final enzyme:substrate ratio of 1:25 (w/w). The solution was vortexed, microcentrufuged for 10 s at 3,000 × g, and incubated for 3 hours at 37°C. In order to check the completeness of the digestion, a 10 µL aliquot was injected onto a LC-UV (Shimadzu, Columbia, MD) and separated with a 10 minute gradient proceeding from 2 - 95% B in 7.5 minutes (mobile phase A: 98/2/0.1% and B: 2/98/0.1% water/acetonitrile/formic acid) on a macrobullet

packed with Magic C18Aq, 5 µm, 200 Å particles (Michrom BioResources, Auburn, CA). The wavelength was monitored at 205 nm.

3.2.3 Iodopeptide Mixtures

All peptides were synthesized at the Mayo Peptide Synthesis Facility (Rochester, MN). The sequences of the four iodopeptides are 1) (I_2Y)GK, 2) (I_2Y)GK(I_2Y)G, 3) (I_2Y)SR(I_2Y)GSYGSSI, and 4) (I_2Y)SR(I_2Y)GSYGSSIGSY, which will be referred to as I1, I2, I3, and I4, respectively. The abbreviation (I_2Y) denotes a 3,5-diiodotyrosine residue (**Figure 3.1**). For an initial assessment of the retention time reproducibility, the four iodopeptides were combined in a 1 µM (total concentration) equimolar mixture. Since I1 was not retained during these experiments, it was not included in subsequent experiments. **Table 3.1** shows the on-column concentrations of the three iodopeptides and myoglobin digest used in experiments to assess the retention time reproducibility of the iodopeptides in a complex mixture and normalization strategies.



Monoisotopic mass = 414.8566
Mass excess = 0.8566

Figure 3.1 3,5-diiodotyrosine (I_2Y)

Table 3.1 On-column concentration of each of the components in the iodopeptide and myoglobin digest mixture.

On-Column Concentration (pmol)				
Total	**I2**	**I3**	**I4**	**Digest**
10	3.2	2.7	2.1	2
4	1.28	1.08	0.84	0.8
2	0.64	0.54	0.42	0.4
1	0.32	0.27	0.21	0.2

3.2.4 nanoLiquid Chromatography-Mass Spectrometry

All reversed-phase nanoLC separations were performed on 10 cm long by 75 µm i.d. IntegraFit column (New Objective, Woburn, MA) packed in house with Magic C18Aq , 5 µm, 200 Å particles (Michrom BioResources, Auburn, CA). A C8 pre-column (Optimize Technologies, Oregon City, OR) was used to desalt and concentrate samples prior to

reversed-phase separation. The flow rate was maintained at 500 nL/min with CS Analytics Chorus 220 pumps (LEAP Technologies, Carrboro, NC). Gradient elution proceeded from 98% A (98% H_2O/ 2% acetonitrile/ 0.1% formic acid by volume) and 2% B (95% acetonitrile/ 5% H_2O/ 0.1% formic acid by volume) to 35% B in 17 minutes.

Mass spectra were acquired utilizing a Q-FT-ICR mass spectrometer (IonSpec FTMS systems, Varian, Inc., Lake Forest, CA) with an actively-shielded 7 T superconducting magnet (Cryomagnetics, Oak Ridge, TN) and a Z-spray ESI source (Waters Corp, Milford, MA). All spectra were acquired with 512 k points and an ADC rate of 1 MHz. Samples were sprayed from a 30 μm i.d. PicoTip Emitter (New Objective, Woburn, MA) with a 2000 V potential applied.

3.2.5 Data Analysis

Selected ion chromatograms were viewed with FTDoc Viewer (part of the OMEGA 8 software package) from which we obtained the retention time. The mass spectra corresponding to the eluted peak apex were processed in IonSpec's OMEGA 8 software with a Blackman window function applied and the 512 k word transient was zero-filled once prior to Fast Fourier transform. All data were externally calibrated with polypropylene-glycol (average MW 1000).

3.2.5.1 Five-Point Global Normalization

Using the average absolute ion abundance of the 1^+ and 2^+ charge states of three iodopeptides ($\overline{A}_{ITs}$) that were detected in the first replicate (*N1*) of the 4 pmol concentration (total concentration on-column of the iodopeptides and myoglobin fragments) and the

spectrum of interest, we calculated a normalization factor (**Equation 3.1**).

$$NF = \frac{\overline{A}_{ITs,(4,N1)}}{\overline{A}_{ITs,(C,N)}} \qquad (\mathbf{3.1})$$

In the equation *C* indicates the concentration and *N* signifies the replicate of the spectrum of interest that is being normalized. The normalization factor (*NF*) was then used to adjust the absolute ion abundances of the myoglobin fragments (*F*) as shown in **Equation 3.2**.

$$Normalized\ Abundance = A_{F,(C,N)} * NF \qquad (\mathbf{3.2})$$

3.2.5.2 Abundance-Matched Normalization

Using the absolute ion abundance of the iodopeptide (A_{IT}) in the first replicate (*N1*) of the 4 pmol concentration (total concentration on-column of the iodopeptides and myoglobin fragments) and the same iodopeptide in subsequent spectra, we calculated a normalization factor (**Equation 3.3**).

$$NF = \frac{A_{IT,(4,N1)}}{A_{IT,(C,N)}} \qquad (\mathbf{3.3})$$

This was repeated for every iodopeptide in each concentration (*C*) and replicate (*N*). For each replicate, there were 5 normalization factors corresponding to the 5 iodopeptide species. Then, we compared the abundance of a myoglobin fragment (*F*) to the abundances of the iodopeptides from the same LC-run. The normalization factor from the abundance of the iodopeptide closest to the abundance of the myoglobin fragment to be normalized ($A_{F,(C,N)}$) was chosen for normalization (**Equation 3.4**).

$$Normalized\ Abundance = A_{F,(C,N)} * NF \qquad (\mathbf{3.4})$$

3.3 Results and Discussion

3.3.1 Design of Iodopeptide Internal Standards

In order to reduce the probability that the chosen exogenous peptide internal standards would overlap with endogenous tryptic peptides in the mass space, we needed to incorporate an element that was not found in peptides. The most obvious choice is a halogen owing to the large mass excess compared to the elements comprising amino acids. Therefore, based on the commercial availability of synthetic amino acids, we designed a set of four peptides that incorporated a 3, 5-diiodotyrosine amino acid and that spanned the range of *m/z* space, which could be used in future studies using internal calibration with the dual electrospray source [35, 36]. These sequences were designed such that the monoisotopic mass of the singly-charged peptide would fall in the forbidden zone. Furthermore, the sequences were designed to have similar amino acid compositions so that the electrospray ionization responses and retention times would be similar, since it has been previously shown that more hydrophobic species have larger electrospray ionization responses [55, 56] and are retained longer on reversed-phase columns [57]. The four sequences, along with their distributions in the mass space are shown in **Figure 3.2**. Each iodopeptide has a monoisotopic mass that falls in the forbidden zone (blue regions in **Figure 3.2**).

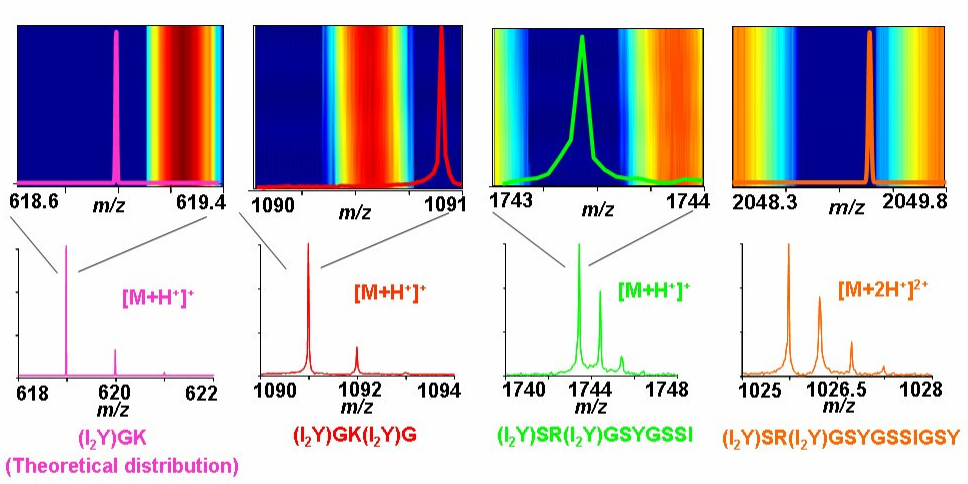


Figure 3.2 Mass spectra and sequences of the four iodopeptides. Above the acquired mass spectra are enlargements of the monoisotopic mass of the singly charged species overlaid with the distributions of all tryptic peptides from NCBI nr database. Iodopeptides were designed to appear in forbidden zones (blue regions), which are unoccupied by tryptic peptides.

3.3.2 Retention Time Reproducibility of the Iodopeptides in Simple and Complex Mixtures

In our initial evaluation of the iodopeptides for retention time reproducibility assessment, we injected a simple mixture containing just the iodopeptides. The average retention time reproducibility is within 5 seconds (confidence interval of the mean, 95% confidence level). In **Figure 3.3**, the total ion chromatograms for the five replicates of the iodopeptide mixture are shown. The inset table displays the average retention time.

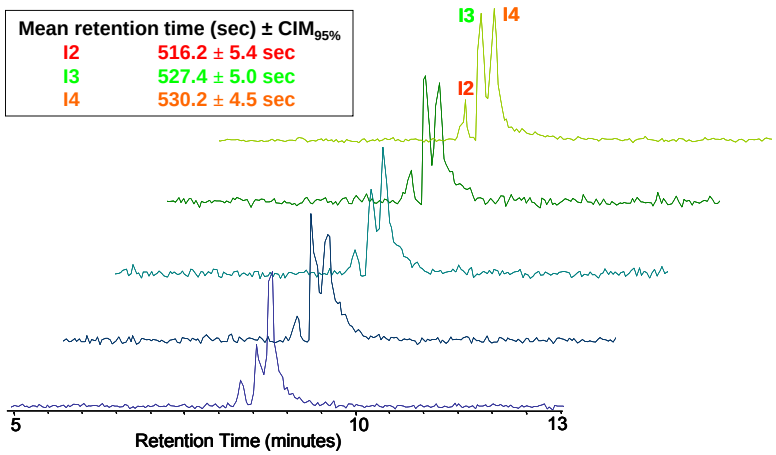


Figure 3.3 nanoLC chromatographs of iodopeptide mix. Table displays retention time reproducibility for the iodopeptides 5 nanoLC-ESI-FT-ICR runs.

After this initial evaluation, the three detected iodopeptides were mixed in varying concentrations with a myoglobin tryptic digest to increase the complexity of the sample. The retention time reproducibility determined from these experiments indicate the potential of the iodopeptides as internal standards for label-free proteomics. Furthermore, the myoglobin tryptic fragments present in the mixture allowed for the assessment of different normalization strategies aiming to improve quantification accuracy (*vide infra*). Five replicate injections for each of four different concentrations (*i.e.,* 10 pmol, 4 pmol, 2 pmol, 1 pmol on column), for a total of 20 injections, were made in random order. **Figure 3.4** displays

Table 3.2 Retention time reproducibility, reported as the mean retention time ± confidence interval of the mean (95% confidence level) for each iodopeptide and myoglobin fragment over the 20 LC-MS runs.

Species	t_R (sec) ± $CIM_{95\%}$
I2	478.5 ± 2.0
I3	482.8 ± 2.0
I4	483.0 ± 1.9
Fragment 80-97	371.0 ± 33.8
Fragment 81-97	400.8 ± 30.0
Fragment 18-32	457.0 ± 2.0
Fragment 120-134	448.2 ± 1.6
Fragment 65-78	457.0 ± 25.5
Fragment 33-43	450.8 ± 1.6
Fragment 147-154	458.9 ± 1.7
Fragment 141-148	459.8 ± 2.2
Fragment 135-140	475.7 ± 1.9
Fragment 149-154	464.8 ± 1.6

the chromatogram for one of the replicates, along with representative mass spectra for two myoglobin fragments and one of the iodopeptides. **Table 3.2** contains the mean retention time of the species listed over the 20 LC-MS analyses. Generally, the retention time reproducibility is within 2 seconds. The tryptic fragments that are outside of the 2 second reproducibility window are low abundant species (approximately 10-fold lower abundance than other fragments). Furthermore, these tryptic fragments are weakly retained by the column and consequently elute with the solvent front, which contributes to the poor reproducibility.

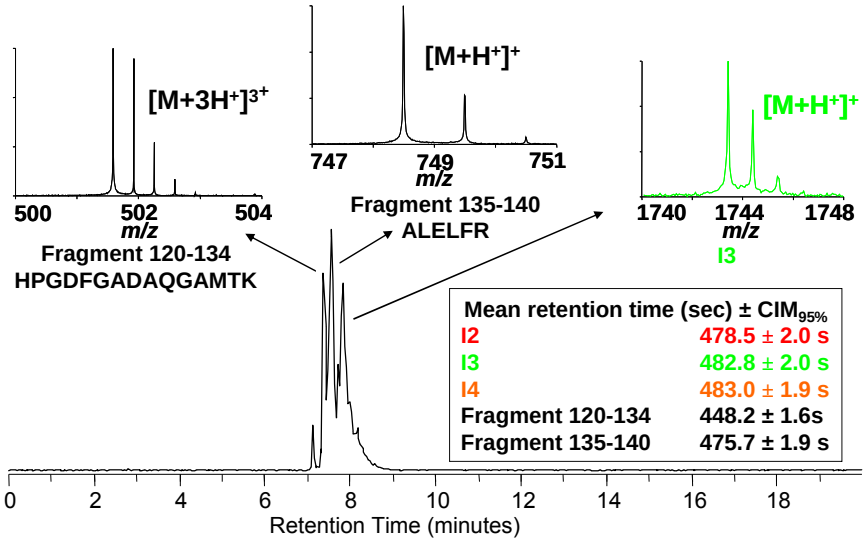


Figure 3.4 nanoLC chromatogram of iodopeptides mixed with a myoglobin digest. Mass spectra for two myoglobin tryptic fragments and iodopeptide 3 are shown. Table displays retention time reproducibility for the iodopeptides and two tryptic fragments over 20 nanoLC-ESI-FT-ICR runs.

3.3.3 Ion Abundance 5-Point Global Normalization

The high mass measurement accuracy afforded to FT-ICR mass spectrometry is a consequence of the mass analyzer's ability to precisely measure an ion's frequency. There are modest x-axis variation FT-ICR measurements, which can be corrected with the implementation of various calibration laws. Conversely, ion abundances (a y-axis measurement) vary considerably due to a variety of factors, including the nature of a

dynamic ionization process (*i.e.,* electrospray ionization), sample handling error, robotic handling error. An accurate quantification method necessitates precise y-axis measurements, or at least a means to account for the variability. Normalization strategies moderate y-axis variations via a comparison of abundances of an internal standard in an initial mass spectrum to subsequent mass spectra. The internal standard does not change, therefore its ionization efficiency remains the same and variations in ion abundance will be due to some external process (*vide supra*).

Global normalization was the first strategy investigated owing to the simplicity of the procedure. Briefly, the average absolute ion abundance of the three iodopeptides in the first replicate of the 4 pmol concentration (total concentration on-column of the three iodopeptides and myoglobin fragments) was used to normalize absolute ion abundances to the other spectra. The percent mean relative error in abundance was calculated using the average absolute ion abundance of the iodopeptides in the first replicate of the 4 pmol concentration as the accepted value. The average percent relative error in abundance for the data prior to and after normalization is shown in **Table 3.3** for each of the total on-column concentrations. These values indicate that global normalization effectively reduces y-axis variation over a 10-fold difference in concentration.

Table 3.3 Average percent relative error in abundances of the iodopeptides determined from 5 replicate injections of each concentration onto a nanoLC-ESI-FT-ICR mass spectrometer for untreated and globally-normalized data.

	I2 1^+		I2 2^+		I3 1^+		I3 2^+		I4	
	Untreated	Global	Untreated	Global	Untreated	Global	Untreated	Global	Untreated	Global
10 pmol	379.7	14.6	242.8	28.2	1242.4	157.0	444.5	13.8	1180.7	169.7
4 pmol	25.2	6.5	35.6	6.0	19.0	19.0	20.1	8.2	64.7	32.8
2 pmol	66.5	10.8	66.5	14.8	58.4	46.2	70.4	16.5	24.4	180.7
1 pmol	78.6	23.5	74.9	13.0	68.8	87.3	77.6	20.7	26.9	165.8

One of the primary disadvantages of a normalization method that utilizes an average value to calculate a normalization factor is that extreme values will skew the average, thus affecting the normalized abundance. This is evident by the large percent relative errors of

I4. The untreated abundances result in less error than the normalized values, which demonstrates that normalization over corrects the abundance.

3.3.4 Abundance-Matched Normalization

Previous work has shown that relative quantification is affected by the ion abundance of the internal standard compared to the ion abundance of the analyte [58]. A strategy that normalizes species to an iodopeptide internal standard with similar abundance could compensate for the adverse effects resulting from disparities between absolute ion abundances of various components in a complex mixture. Therefore, in order to compensate for the effect a disparity between absolute ion abundances of various components in a complex mixture has on quantification, we evaluated a normalization strategy that was based on the absolute abundance of the tryptic fragment compared to the abundances of the iodopeptides. The absolute ion abundance for each iodopeptide in the first replicate of the 4 pmol concentration (total concentration on-column of the iodopeptides and myoglobin fragments) and the corresponding iodopeptide in each subsequent LC run was used to calculate a normalization factor. Then, the normalization factor for the iodopeptide closest in abundance to the myoglobin fragment of interest was employed for normalization, illustrated in **Figure 3.5**.

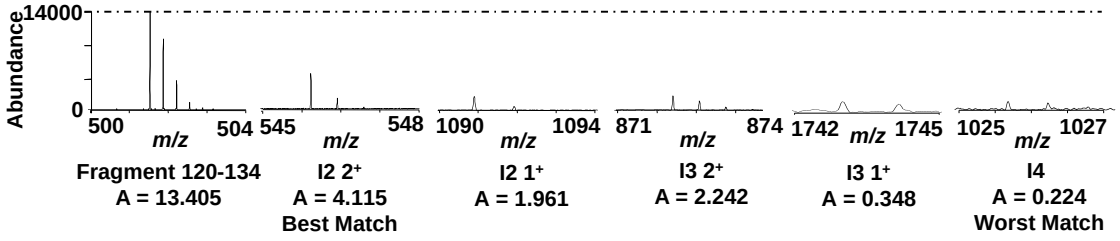


Figure 3.5 Illustration of the abundance-matching normalization procedure. Mass spectra of the iodopeptides and a myoglobin tryptic fragment are shown. The 2^+ charge state of iodopeptide 2 and iodopeptide 4 are the best and worst match in abundance, respectively, with the myoglobin fragment.

Abundance-matched normalization yielded mean relative errors less than 54% (**Table 3.4**), which is essentially the same percent relative error compared to the global normalization strategy. However, the mean percent relative error averaged from all concentrations is misleading because the best-match abundance normalization results in a larger percent relative error for the 10 pmol concentration, which skews the average across all concentrations. If each concentration is considered independently, the best-matched abundance normalization does as well as (4 pmol) and better (2 pmol and 1 pmol) than global normalization. Moreover, global normalization may perform better, since the normalization factor is an average versus a single measure. Further experiments are warranted to examine the utility of abundance-matched normalization for more complex mixtures with a wider range of concentrations.

Table 3.4 Average percent relative error in untreated and normalized abundances for the 22 myoglobin tryptic fragments from 5 replicate injections onto a nanoLC-ESI-FT-ICR mass spectrometer of each concentration indicated.

Mean Percent Relative Error				
	Untreated	**Global**	**Best Match**	**Worst Match**
10 pmol	264.9	55.8	67.6	65.5
4 pmol	45.2	36.6	37.8	40.1
2 pmol	59.2	64.9	57.7	69.5
1 pmol	69.9	55.7	52.4	78.9
Average	**109.8**	**53.3**	**53.9**	**63.5**

In order to verify the improvement in percent relative error for the abundance-matched normalization procedure was due abundance-matching and not the normalization procedure per se, myoglobin fragments were also normalized to the iodopeptide with the greatest disparity in abundance. While the worst-match abundance normalization improved the mean percent relative error over untreated data, it did not improve the percent relative error as well as global normalization or the best-match abundance normalization strategy.

Another way to assess the improvement in ion abundances after normalization is linear regression. Ideally, a plot of the abundances from each replicate versus replicate 1 (our comparison standard) should be linear. If a perfect correlation (slope = 1) exists we can assume that the same species will produce the same ion abundance time after time, thus lending to accurate quantification. By normalizing the abundances we are trying to reduce the variability in ion abundance resulting from systematic errors, and therefore forcing the correlation to approach 1. A natural logarithm transformation of the data reduces the weight

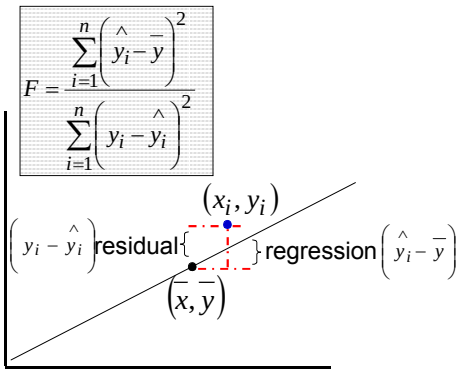


Figure 3.6 Illustration of the F-statistic. The F-statistic is the ratio of the regression sum of squares to the residual sum of squares (inset equation). Residuals are the difference in the predicted value $\hat{y}$ from the actual value y. Regression measures the difference in the predicted value $\hat{y}$ from the average $\bar{y}$.

of more abundant species (reduces positive skew), so that the data follow an approximately normal distribution consequently affording statistical evaluation. The F statistic is a goodness-of-fit measure and is the ratio of the regression sum of squares to the residual sum of squares; this concept is demonstrated in **Figure 3.6**. Plots of the natural logarithm-transformed abundances for replicate 2 versus replicate 1 for the untreated data and data normalized with the three normalization strategies are shown in **Figure 3.7**. Normalization improves the correlation between the abundances (better R^2 value), which is evident from the reduced y-axis variability (smaller confidence interval of the mean). Furthermore, the F statistic also improves with

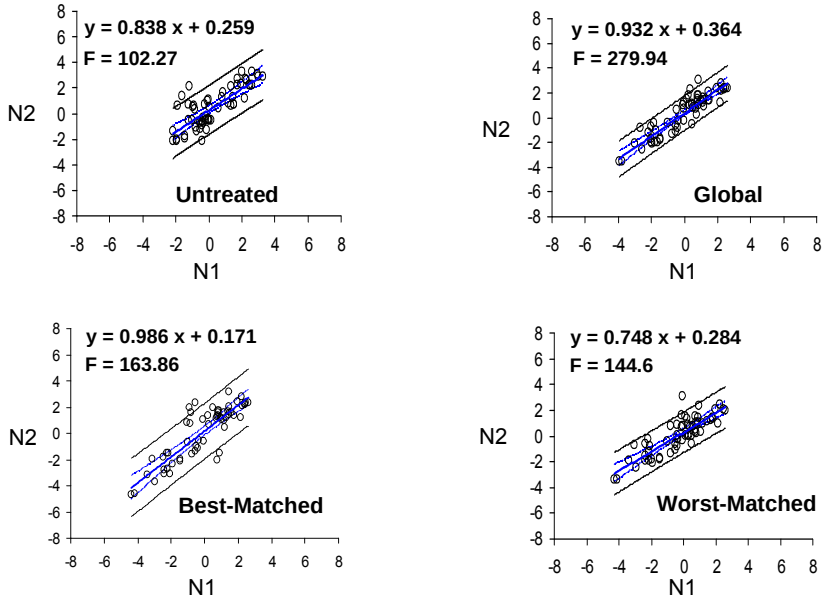


Figure 3.7 Correlation of abundances between replicates 1 (N1) and 2 (N2). Plot of the natural logarithm of the absolute ion abundances for each myoglobin species in the second replicate (N2) versus the first replicate (N1). The abundances for all concentrations are included in each plot. Additionally, the F statistic, indicating goodness-of-fit, is shown. Note the abundances are statistically significantly correlated in all plots at the alpha level of 0.05 ($F_{(1,57,0.95)}$ = 4.01). The blue lines indicate the regression line and the black outer lines are the 95% confidence intervals of the mean. Importantly, normalization reduces the width of the confidence intervals of the mean, thereby improving the correlation in abundances between replicates.

normalization; a larger F statistics indicate a better fit. The remaining replicates follow the same trends (data not shown).

3.4 Conclusions and Future Experiments

Utilization of iodopeptides reduces the chance of overlap with tryptic peptides in the *m/z* domain, thereby allowing accurate mass and retention time assessment in label-free proteomics. Our experimental set-up affords retention-time reproducibility generally within 2 seconds. Global or abundance-matched absolute ion abundance of the iodopeptides can be employed to normalize the abundance of endogenous peptides, consequently facilitating label-free applications. Abundance-matched and globally- normalized data yields relative errors ~54% for 22 different myoglobin species spanning a ten-fold range of concentration over 20 replicate injections. Furthermore, normalization based on the worst match in abundance produced larger percent relative errors <64%, which is further evidence that relative quantification is affected by differences in ion abundance. These preliminary results clearly indicate the potential of the approach for label-free proteomics. We are currently implementing this approach for both top-down and bottom-up proteomics applications.

Future studies are planned to examine the retention time reproducibility of the iodopeptides in the low-molecular weight fraction of serum (MW ≤ 10,000 Da), which is a more complex mixture. Furthermore, the wide range of concentrations of the proteins present in serum will allow us to test the different normalization strategies over a much wider concentration range. In these studies the concentrations of the four iodopeptides will be varied to match the abundances of the proteins present. We will also utilize automatic gain control and external calibration laws to assess mass measurement accuracy of the proteins/peptides present in the low molecular weight fraction of serum. In these planned experiments the mass excess of the iodopeptides is crucial in order to assess the run to run

variability of retention times and ion abundances in these complex mixtures since the numerous endogenous peptides will not mask the iodopeptides.

3.5 References

1. Wasinger, VC, Cordwell, SJ, Cerpapoljak, A, Yan, JX, Gooley, AA, Wilkins, MR, et al. Progress with gene-product mapping of the Mollicutes - Mycoplasma-genitalium. *Electrophoresis,* 1995. **16**: 1090-1094.

2. Kenyon, GL, DeMarini, DM, Fuchs, E, Galas, DJ, Kirsch, JF, Leyh, TS, et al. Defining the mandate of proteomics in the post-genomics era: Workshop report. *Molecular & Cellular Proteomics,* 2002. **1**: 763-780.

3. Bork, P, Dandekar, T, Diaz-Lazcoz, Y, Eisenhaber, F, Huynen, M, and Yuan, YP. Predicting function: From genes to genomes and back. *Journal of Molecular Biology,* 1998. **283**: 707-725.

4. Dass, C, Kusmierz, JJ, Desiderio, DM, Jarvis, SA, and Green, BN. Electrospray mass-spectrometry for the analysis of opioid-peptides and for the quantification of endogenous methionine enkephalin and beta-endorphin. *Journal of the American Society for Mass Spectrometry,* 1991. **2**: 149-156.

5. Hunt, DF, Henderson, RA, Shabanowitz, J, Sakaguchi, K, Michel, H, Sevilir, N, et al. Characterization of peptides bound to the class-I MHC molecule HLA-A2.1 by mass-spectrometry. *Science,* 1992. **255**: 1261-1263.

6. Ong, SE and Mann, M. Mass spectrometry-based proteomics turns quantitative. *Nature Chemical Biology,* 2005. **1**: 252-262.

7. Bronstrup, M. Absolute quantification strategies in proteomics based on mass spectrometry. *Expert Review of Proteomics,* 2004. **1**: 503-512.

8. Sechi, S and Chait, BT. Modification of cysteine residues by alkylation. A tool in peptide mapping and protein identification. *Analytical Chemistry,* 1998. **70**: 5150-5158.

9. Gygi, SP, Rist, B, Gerber, SA, Turecek, F, Gelb, MH, and Aebersold, R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnology,* 1999. **17**: 994-999.

10. Zhou, HL, Ranish, JA, Watts, JD, and Aebersold, R. Quantitative proteome analysis by solid-phase isotope tagging and mass spectrometry. *Nature Biotechnology,* 2002. **20**: 512-515.

11. Li, J, Steen, H, and Gygi, SP. Protein profiling with cleavable isotope-coded affinity tag (cICAT) reagents: The yeast salinity stress response. *Molecular & Cellular Proteomics,* 2003. **2**: 1198-1204.

12. Hall, MP and Schneider, LV. Isotope-differentiated binding energy shift tags (IDBEST(tm)) for improved targeted biomarker discovery and validation. *Expert Review of Proteomics,* 2004. **1**: 421-431.

13. Doolittle, RF. *Redundancies in protein sequences*, in *Prediction of protein structure and the principles of protein conformation*, Fasman, GD, Editor. 1989, Plenum Press: New York. p. 599-623.

14. Geng, MH, Ji, JY, and Regnier, FE. Signature-peptide approach to detecting proteins in complex mixtures*. Journal of Chromatography A,* 2000. **870**: 295-313.

15. Chakraborty, A and Regnier, FE. Global internal standard technology for comparative proteomics*. Journal of Chromatography A,* 2002. **949**: 173-184.

16. Mason, DE and Liebler, DC. Quantitative analysis of modified proteins by LC-MS/MS of peptides labeled with phenyl isocyanate*. Journal of Proteome Research,* 2003. **2**: 265-272.

17. Ross, PL, Huang, YN, Marchese, JN, Williamson, B, Parker, K, Hattan, S, et al. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents*. Molecular & Cellular Proteomics,* 2004. **3**: 1154-1169.

18. Cagney, G and Emili, A. De novo peptide sequencing and quantitative profiling of complex protein mixtures using mass-coded abundance tagging*. Nature Biotechnology,* 2002. **20**: 163-170.

19. Leitner, A and Lindner, W. Current chemical tagging strategies for proteome analysis by mass spectrometry*. Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences,* 2004. **813**: 1-26.

20. Regnier, FE and Julka, S. Primary amine coding as a path to comparative proteomics*. PROTEOMICS,* 2006. **6**: 3968-3979.

21. Schripsema, J, Peltenburglooman, A, Erkelens, C, and Verpoorte, R. Nitrogen-metabolism in cultures of Tabernaemontana-divaricata*. Phytochemistry,* 1991. **30**: 3951-3954.

22. Conrads, TP, Alving, K, Veenstra, TD, Belov, ME, Anderson, GA, Anderson, DJ, et al. Quantitative analysis of bacterial and mammalian proteomes using a combination of cysteine affinity tags and N-15-metabolic labeling*. Analytical Chemistry,* 2001. **73**: 2132-2139.

23. Ong, SE, Blagoev, B, Kratchmarova, I, Kristensen, DB, Steen, H, Pandey, A, et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics*. Molecular & Cellular Proteomics,* 2002. **1**: 376-386.

24. Desiderio, DM and Kai, M. Preparation of stable isotope-incorporated peptide internal standards for field desorption mass-spectrometry quantification of peptides in biologic tissue*. Biomedical Mass Spectrometry,* 1983. **10**: 471-479.

25. Mirgorodskaya, OA, Kozmin, YP, Titov, MI, Korner, R, Sonksen, CP, and Roepstorff, P. Quantitation of peptides and proteins by matrix-assisted laser desorption/ionization mass spectrometry using O-18-labeled internal standards*. Rapid Communications in Mass Spectrometry,* 2000. **14**: 1226-1232.

26. Reynolds, KJ, Yao, XD, and Fenselau, C. Proteolytic O-18 labeling for comparative proteomics: Evaluation of endoprotease Glu-C as the catalytic agent. *Journal of Proteome Research,* 2002. **1**: 27-33.

27. Johnson, KL and Muddiman, DC. A method for calculating O-16/O-18 peptide ion ratios for the relative quantification of proteomes. *Journal of the American Society for Mass Spectrometry,* 2004. **15**: 437-445.

28. Mason, CJ, Therneau, TM, Eckel-Passow, JE, Johnson, KL, Oberg, AL, Olson, JE, et al. A method for automatically interpreting mass spectra of 180-labeled isotopic clusters. *Molecular & Cellular Proteomics,* 2007. **6**: 305-318.

29. Wiener, MC, Sachs, JR, Deyanova, EG, and Yates, NA. Differential mass spectrometry: A label-free LC-MS method for finding significant differences in complex peptide and protein mixtures. *Analytical Chemistry,* 2004. **76**: 6085-6096.

30. Johnson, KL, Mason, CJ, Muddiman, DC, and Eckel, JE. 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. *Analytical Chemistry,* 2004. **76**: 5097-5103.

31. Bergen, HR, Vasmatzis, G, Cliby, WA, Johnson, KL, Oberg, AL, and Muddiman, DC. Discovery of ovarian cancer biomarkers in serum using nanoLC electrospray ionization TOF and FT-ICR mass spectrometry. *Disease Markers,* 2003. **19**: 239-249.

32. Smith, RD, Anderson, GA, Lipton, MS, Pasa-Tolic, L, Shen, YF, Conrads, TP, et al. An accurate mass tag strategy for quantitative and high-throughput proteome measurements. *PROTEOMICS,* 2002. **2**: 513-523.

33. Conrads, TP, Anderson, GA, Veenstra, TD, Pasa-Tolic, L, and Smith, RD. Utility of accurate mass tags for proteome-wide protein identification. *Analytical Chemistry,* 2000. **72**: 3349-3354.

34. Hannis, JC and Muddiman, DC. A dual electrospray ionization source combined with hexapole accumulation to achieve high mass accuracy of biopolymers in Fourier transform ion cyclotron resonance mass spectrometry. *Journal of the American Society for Mass Spectrometry,* 2000. **11**: 876-883.

35. Nepomuceno, AI, Muddiman, DC, Bergen, HR, Craighead, JR, Burke, MJ, Caskey, PE, et al. Dual electrospray ionization source for confident generation of accurate mass tags using liquid chromatography Fourier transform ion cyclotron resonance mass spectrometry. *Analytical Chemistry,* 2003. **75**: 3411-3418.

36. Burke, MJ, Caskey, PE, and Muddiman, DC. *Dual electrospray ionization source for mass spectrometery*. 2006, Mayo Foundation for Medical Education and Research.

37. Muddiman, DC and Oberg, AL. Statistical evaluation of internal and external mass calibration laws utilized in Fourier transform ion cyclotron resonance mass spectrometry. *Analytical Chemistry,* 2005. **77**: 2406-2414.

38. Syka, JEP, Marto, JA, Bai, DL, Horning, S, Senko, MW, Schwartz, JC, et al. 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.

39. Mason, CJ, Johnson, KL, and Muddiman, DC. Reproducibility of retention time using a splitless nanoLC coupled to an ESI-FTICR mass spectrometer. *Journal of Biomolecular Techniques,* 2005. **16**: 414-22.

40. Callister, SJ, Barry, RC, Adkins, JN, Johnson, ET, Qian, WJ, Webb-Robertson, BJM, et al. Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics. *Journal of Proteome Research,* 2006. **5**: 277-286.

41. Listgarten, J and Emili, A. Statistical and computational methods for comparative proteomic profiling using liquid chromatography-tandem mass spectrometry. *Molecular & Cellular Proteomics,* 2005. **4**: 419-434.

42. Quackenbush, J. Microarray data normalization and transformation. *Nature Genetics,* 2002. **32**: 496-501.

43. Liu, HB, Sadygov, RG, and Yates, JR. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Analytical Chemistry,* 2004. **76**: 4193-4201.

44. Old, WM, Meyer-Arendt, K, Aveline-Wolf, L, Pierce, KG, Mendoza, A, Sevinsky, JR, et al. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Molecular & Cellular Proteomics,* 2005. **4**: 1487-1502.

45. Zybailov, B, Coleman, MK, Florens, L, and Washburn, MP. Correlation of relative abundance ratios derived from peptide ion chromatograms and spectrum counting for quantitative proteomic analysis using stable isotope labeling. *Analytical Chemistry,* 2005. **77**: 6218-6224.

46. Xia, QW, Wang, TS, Park, Y, Lamont, RJ, and Hackett, M. Differential quantitative proteomics of Porphyromonas gingivalis by linear ion trap mass spectrometry: Non-label methods comparison, Q-values and LOWESS curve fitting. *International Journal of Mass Spectrometry,* 2007. **259**: 105-116.

47. Zhang, B, VerBerkmoes, NC, Langston, MA, Uberbacher, E, Hettich, RL, and Samatova, NF. Detecting differential and correlated protein expression in label-free shotgun proteomics. *Journal of Proteome Research,* 2006. **5**: 2909-2918.

48. Zybailov, B, Mosley, AL, Sardiu, ME, Coleman, MK, Florens, L, and Washburn, MP. Statistical analysis of membrane proteome expression changes in Saccharomyces cerevisiae. *Journal of Proteome Research,* 2006. **5**: 2339-2347.

49. Wang, GH, Wu, WW, Zeng, WH, Chou, CL, and Shen, RF. Label-free protein quantification using LC-coupled ion trap or FT mass spectrometry: Reproducibility,

linearity, and application with complex proteomes. *Journal of Proteome Research,* 2006. **5**: 1214-1223.

50. Higgs, RE, Knierman, MD, Gelfanova, V, Butler, JP, and Hale, JE. Comprehensive label-free method for the relative quantification of proteins from biological samples. *Journal of Proteome Research,* 2005. **4**: 1442-1450.

51. Steen, H, Jebanathirajah, JA, Springer, M, and Kirschner, MW. Stable isotope-free relative and absolute quantitation of protein phosphorylation stoichiometry by MS. *Proceedings of the National Academy of Sciences of the United States of America,* 2005. **102**: 3948-3953.

52. Chelius, D and Bondarenko, PV. Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *Journal of Proteome Research,* 2002. **1**: 317-323.

53. Wang, WX, Zhou, HH, Lin, H, Roy, S, Shaler, TA, Hill, LR, et al. Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. *Analytical Chemistry,* 2003. **75**: 4818-4826.

54. Frahm, JL, Howard, BE, Heber, S, and Muddiman, DC. Accessible proteomics space and its implications for peak capacity for zero-, one- and two-dimensional separations coupled with FT-ICR and TOF mass spectrometry. *Journal of Mass Spectrometry,* 2006. **41**: 281-288.

55. Null, AP, Nepomuceno, AI, and Muddiman, DC. Implications of hydrophobicity and free energy of solvation for characterization of nucleic acids by electrospray ionization mass spectrometry. *Analytical Chemistry,* 2003. **75**: 1331-1339.

56. Fenn, JB. Ion formation from charged droplets - roles of geometry, energy, and time. *Journal of the American Society for Mass Spectrometry,* 1993. **4**: 524-535.

57. Snyder, LR, Kirkland, JJ, and Glajch, JL. *Practical HPLC method development*. 2nd ed. 1997, New York: John Wiley & Sons. xxvi, 765 p.

58. Gordon, EF and Muddiman, DC. Impact of ion cloud densities on the measurement of relative ion abundances in Fourier transform ion cyclotron resonance mass spectrometry: Experimental observations of coulombically induced cyclotron radius perturbations and ion cloud dephasing rates. *Journal of Mass Spectrometry,* 2001. **36**: 195-203.

CHAPTER 4

Influence of Post-Excite Radius and Axial Confinement on Quantitative Proteomics Measurements

4.1 Introduction

Quantitative proteomics has experienced a paradigm shift with the inception of relative quantification of proteomes, initially demonstrated with ICAT labeling [1], and now includes a vast repertoire of methods [2-4]. Relative quantification methods (including label free methods [5, 6]) for proteomics overcome many of the limitations of absolute quantification, namely the identification and synthesis of an appropriate stable-isotope-labeled internal standard [7-10]. However, absolute quantification remains the method of choice for targeted proteomics analyses (*e.g.,* biomarker validation) [9, 11].

The unparalleled mass measurement accuracy, mass resolving power, sensitivity, and dynamic range of FT-ICR mass spectrometry make it one of the leading mass analyzer technologies for biomolecule analysis and quantification [12-16]. Furthermore, coupling FT-ICR with separations dramatically increases the peak capacity of the method, thus enabling complex mixture analysis [17, 18]. Research in the area of quantitative proteomics by FT-ICR mass spectrometry has primarily focused on the development of new methods and improvement of existing methods. Only a handful of studies have explored the fundamental issues surrounding relative and absolute quantification of proteomes by FT-ICR mass spectrometry, including the role of various instrument parameters on ion abundance ratios. Some of the more relevant instrumental parameters to quantitative proteomics measurements and the corresponding fundamental FT-ICR studies are described below.

FT-ICR mass analyzers are unique in that acquisition is comprised of two events, excitation and detection. Excitation in ICR has two functions 1) to cause phase coherence

of ions of the same *m/z* and 2) expand the orbit of the cloud enough to induce a signal [15]. The radius of the expanded orbit is the post-excite radius. As the phase-coherent ion packet orbits near the detection plates, the instantaneous charge on one detection plate can be determined from **Equation 4.1** [19]:

$$Q(t) = \frac{-Nqr_{post-excite}\cos(\omega t)}{d} \qquad (\mathbf{4.1})$$

The instantaneous charge on a detection plate (Q) is a function of the number of ions (N), charge of the ion (q), post-excite radius ($r_{post\text{-}excite}$), cyclotron frequency (ω) at time t, and the distance between the detection plates (d) of the ICR cell. In a single acquisition the cyclotron frequency, charge, and distance between the detection plates are constant for a molecule and particular ICR cell geometry, respectively. The number of ions can be changed to some degree by varying the concentration or accumulation time; however, peak coalescence and large frequency shifts occur once the cell nears its capacity threshold, thereby significantly reducing mass resolving power and mass measurement accuracy [20-22]. Induced charge, therefore ion abundance, can be most readily modulated by varying the post-excite radius, corresponding to alterations in excitation voltage [15].

Early studies present a largely theoretical examination of the optimal post-excite radius in FT-ICR measurements [20, 23]. More recently, Hawkridge, *et al.* experimentally examined the relationship between post-excite radius and total ion abundance [24]. The authors also explored the role of post-excite radius on mass measurement accuracy for internally-calibrated data and isotopic distributions, both of which influence identification [24]. The goal of these studies was to identify the optimal post-excite radius for the instrument platform that would lend to accurate measurements.

In addition to excitation voltage, the post-excite radius can also be influenced by ion cloud interactions. In a system with two unequal ion populations, the more abundant species can perturb the less abundant species to a larger post-excite radius, causing

artificially higher abundances (*i.e.,* increased image current) because the cloud is closer to the detection plates [25]. Uechi and Dunbar made an analogous observation for two fragment ions from n-butylbenzene with similar *m/z* values (small differences in cyclotron frequencies) [21]. Smaller, if any, radial perturbations were observed for two species with very different *m/z* values (large differences in cyclotron frequencies), therefore the authors concluded the radial perturbation of the system with close *m/z* values was a result of space-charge effects during excitation [21]. Consequently, the optimal excite voltage for a quantitative measurement maximizes ion abundance and reduces space charge effects.

An applied electrostatic trapping potential is necessary to reduce z-axial motion and confine ions within the ICR cell; however, the resulting nonlinear electric field, which is a consequence of the finite ICR cell geometry, can affect signal amplitude [15, 26]. Additionally, the parabolic shape of the electric field lines from the trapping potential requires higher trapping voltages to axially confine ions excited to a larger post-excite radius [27-29].

The instrumental parameters chosen for detection, the second event in a FT-ICR acquisition, also influence quantitative measurements. Early investigations revealed that accurate quantification with FT-ICR mass spectrometry can be realized as long as the second isotopic beat is detected [30, 31]. In a study by Gordon, *et al.* of cyclosporin A (CsA) and cyclosporin G (CsG), the authors reported coefficients of variation within ~10% over the concentration range of 0 – 250 ng/ml CsA with a constant concentration of 200 ng/ml CsG were achieved with the detection of only 10 isotopic beats [31]. Moreover, the relative standard deviation of the standard curve (CsA/CsG abundance ratio) was not significantly improved with the detection of more isotopic beats. This result has important implications for quantification, especially for LC-FT-ICR experiments during which shorter acquisition times are necessary to ensure adequate chromatographic peak profiling.

During longer acquisition times signal decay becomes more apparent, which results from a loss of ion cloud coherence, therefore causing a decrease in the detected current

(*i.e.,* signal amplitude) [32]. Signal decay in FT-ICR is quantified with the signal decay rate, τ, and is defined as the time required for the signal amplitude to decay to its initial value divided by the base of the natural logarithm (*e*) [32]. Pressure dampening from ion-neutral collisions also result in ICR signal decay [32]; however, improved vacuum technology has virtually eliminated this decay mechanism in modern FT-ICR mass spectrometers.

The loss of cloud coherence, resulting in signal decay, is notably affected by ion cloud density [32]. Less dense ion clouds (lower concentration species) generally have fast signal decay rates (*i.e.,* small τ values), which has been attributed to a reduced rotational frequency [32-34]. The rotational frequency of an ion cloud is a consequence of Coulombic interactions between ions of the same *m/z* (*i.e.,* same ion cloud), causing the cloud to rotate about its symmetry axis [34]. Peurrung and Kouzes reported greater ion cloud densities result in larger rotational frequencies, thus counteracting the shearing forces from magnetic or electric field inhomogeneities acting to disrupt the cloud [34]. More recently, Kaiser and Bruce demonstrated cloud coherence can be increased with electron promoted ion coherence (EPIC), during which a high density electron beam is utilized to alter electric fields in the ICR cell [35].

Gordon, *et al.* demonstrated that signal decay can adversely affect quantification if the signal decay rates for two species appreciably differs [25]. In such a system, the species with a faster signal decay rate will appear less abundant than the actual concentration over the acquisition period ($T_{acq'n}$) due to the signal averaging aspect of FT. The effect of signal decay on the accuracy of ion abundance in FT-ICR mass spectrometry has been the focus of several investigations [25, 32, 33, 36-38].

One strategy to overcome the adverse effects of signal decay on ion abundance accuracy is to decrease the acquisition time. However, this strategy compromises mass resolving power [39], and therefore mass measurement accuracy. As an alternative approach,

our group and others have suggested Fourier transforming part of the time domain signal, before complete signal decay, to improve ion abundance accuracy for quantitative measurements [25, 32, 33, 36]. Recently this strategy was implemented by Kelleher's group to determine expression ratios of intact yeast proteins [40]. In their two-pronged approach, they used the full time domain to achieve accurate molecular weight information for identification and truncated time domains for quantitative information [40].

These early studies of FT-ICR mass spectrometry provide strong evidence that the accuracy of ion abundance measurements, and therefore quantification, depends on experimental conditions (*e.g.,* post-excite radius, trapping voltage, acquisition time, and signal decay). However, improvements in instrument technology and the analysis of larger molecules in increasingly complex mixtures warrant not only a revisit to some of these earlier studies, but a more comprehensive examination of the influence of various instrument parameters on quantitative (absolute and relative) measurements in proteomics. Furthermore, the scope of current fundamental quantitative proteomics studies needs to be extended to include several different parameters, so the correlation between these instrumental values and quantification can be explored. To this end, we present a systematic evaluation of the optimal instrumental parameters including excitation voltage, trapping voltage, acquisition time, and type of excitation waveform for the accurate quantification of biological molecules by electrospray ionization FT-ICR mass spectrometry. The use of a stable-isotope-labeled and unlabeled-phenyl isocyanate derivatized peptide allows us to ascribe the effects of FT-ICR mass spectrometry on quantification, thus eliminating the contribution of ionization differences to ion abundance. To adequately assess the multiple parameters in the large dataset, we develop a multiplicative quality factor that encompasses the total ion abundance, as well as the accuracy and the precision of abundance ratios. This assessment allows facile determination of optimal instrument parameters for quantitative measurements. The evaluation imparted herein demonstrates

the complexity of quantitative measurements by FT-ICR mass spectrometry and the necessity of fundamentally understanding these issues.

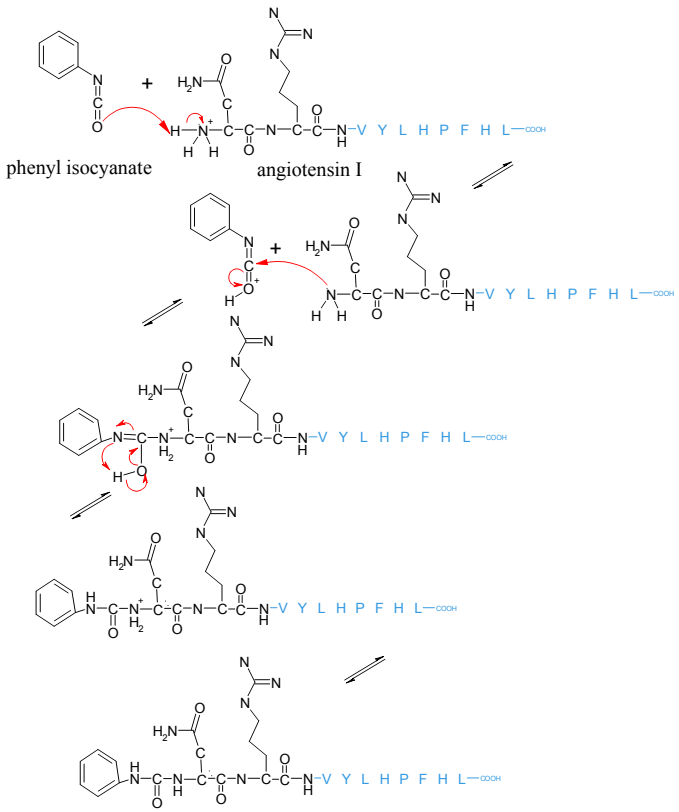
4.2 Experimental

4.2.1 Reagents

Phenyl isocyanate (PIC) (≥99%) and the peptide angiotensin I (sequence Asp-Arg-Val-Tyr-Ile-His-Pro-Phe-His-Leu) were purchased from Sigma-Aldrich (St. Louis, MO) and used as received. ^{13}C-labeled PIC (99 atom% ^{13}C) was obtained from Isotec, Inc. (Miamisburg, OH) and used without further purification. HPLC-grade Burdick and Jackson water and acetonitrile were acquired from VWR International (Suwanee, GA).

4.2.2 Phenyl Isocyanate Labeling

The following procedure is modified from the PIC-labeling protocol reported by Mason and Liebler [41]. First, a 1:1 PIC/*PIC solution was made by mixing 6 µL of 0.1 M PIC in acetonitrile and 6 µL of 0.1 M *PIC in acetonitrile. Then 6 µL of the pre-mixed 0.1 M PIC/*PIC was added to 200 µL of 0.1 mg/ml angiotensin I in Tris buffer (10mM, pH 8.00). The solution was then incubated at 37°C for 30 minutes. After the time elapsed, 1 µL of formic acid was added to quench the reaction. We followed the same procedure for the 1:50 PIC/*PIC; however, in the first step 2 µL of 0.1 M PIC was mixed with 100 µL of 0.1 M *PIC. The proposed reaction mechanism is shown in **Scheme 4.1**.



Scheme 4.1 Derivatization mechanism of angiotensin I with phenyl isocyanate. The chemical structure of the first two residues, asparagine and arginine, are shown. The single letter abbreviation for the remaining angiotensin I sequence is shown in blue.

4.2.3 Liquid Chromatography-Ultraviolet Absorption Analysis

The actual concentration of the phenyl isocyanate solution was determined by LC-UV (Shimadzu, Columbia, MD) analysis prior to derivatization with the peptide. To accomplish this, a linear gradient of mobile phase A (98/2 water/acetonitrile containing 0.1% formic acid) and mobile phase B (98/2 acetonitrile/water containing 0.1% formic acid) was performed from 2% B initially to 95% B at 7.5 minutes, with a flow rate of 0.5 ml/min. Solvent conditions returned to 2% B at 9 minutes, followed by a 10 minute column re-equilibration period prior to the next injection. Independently, 10 µL of a 1:100 dilution of each PIC stock solution with water (1 mM PIC) was injected on a macrobullet packed with Magic C18Aq, 5 µm, 200 Å particles (Michrom BioResources, Auburn, CA). The wavelength

was monitored at 205 nm (Shimadzu SPD-20A, Columbia, MD) and the average integrated peak area for the eluted PIC species from three separate chromatographic runs was used to calculate the actual PIC/*PIC ratio. All experimentally determined ratios obtained from mass spectrometry were compared to this "accepted" value.

4.2.4 Mass Spectrometry

Mass spectra were acquired utilizing a modified IonSpec FT-ICR mass spectrometer (IonSpec FTMS systems, Varian, Inc., Lake Forest, CA) with an actively-shielded 9.4 T superconducting magnet (Cryomagnetics, Oak Ridge, TN) and Analytica of Branford (Branford, CT) electrospray source. Samples (total concentration 1μM) were electrosprayed from a 15 μm i.d. PicoTip Emitter (New Objective, Woburn, MA) at a flow rate of 0.5 μL/min in a 50:50 water/acetonitrile with 0.1% formic acid electrospray solution. A 1800 V potential was applied to the capillary during infusion of samples.

Two excitation waveforms were used to excite and detect ions in the ICR cell: chirp excitation waveform and a stored waveform inverse Fourier transform (SWIFT) waveform [42]. Chirp excitation was accomplished with an RF sweep from 300 to 2500 *m/z* in 4 ms; the amplitude of excite voltages were randomly varied from 20 – 120 V in 10 V increments over the course of experiments. An arbitrary waveform was designed for the SWIFT excitation waveform such that full power was applied over the *m/z* range 300 – 2500 with excite voltages varying from 90 – 390 V in 30 V increments over the experimental course. A digital-to-analog converter sampling rate of 16,000 kHz, 256 k samples, and a filter setting of 5 (8,000 kHz cutoff), which produced a waveform period of ~16 ms, were chosen based on a previous evaluation of this waveform (unpublished observations). The trapping voltages for both excitation waveforms were the same, and ranged from 1.0 – 3.0 V in increments of 0.4 V. The range of excitation and trapping voltages was determined empirically from ESI-FT-ICR mass spectrometry of the 1:1 PIC/*PIC-angiotensin I solution prior to data

acquisition. All spectra are single-acquisition spectra acquired with 4096 k points and digitized at a rate of 2 MHz. Four replicate spectra for each combination of excitation voltage, trapping voltage, and excitation waveform were randomly acquired in the same day; in total 552 spectra were acquired. Additionally, two consecutive spectra were acquired for each set of parameters and only the second acquisition was saved for analysis to ensure voltage changes were realized.

4.2.5 Data Analysis

4.2.5.1 Post-acquisition Signal Processing to Determine the Signal Decay Rate

ICR-2LS was used to calculate the value of the signal decay rate (τ) [43]. The full time domain signal corresponding to the monoisotopic peak of the PIC- or *PIC-angiotensin I was individually isolated during signal processing, and subsequently used to calculate the signal decay rate for that species. More specifically, the initial amplitude of the time domain signal was divided by e^1 (~2.71828) and the time corresponding to this y-axis value was taken to be τ.

4.2.5.2 Ion Abundance Determinations

All mass spectra were processed in IonSpec's OMEGA 8 software with a Blackman window function applied. The 4096 k word transient was zero-filled once, 2048 k word transient was zero-filled twice, and the 1024 k word transient was zero-filled three times, prior to Fast Fourier transform. This processing kept the FFT size constant, and allowed us to simulate acquisition times of 2, 1, and 0.5 seconds. All data were externally calibrated with polypropylene-glycol (average MW 1000). Relative ion abundances of PIC-angiotensin I/*PIC-angiotensin I obtained from mass spectra were compared to the abundance ratio determined from LC-UV-vis analysis. The accuracy and precision of the relative

abundances are the calculated average relative error and coefficient of variation, respectively.

4.3 Results and Discussion

4.3.1 ESI-FT-ICR Mass Spectrometry of PIC-/*PIC-Angiotensin I

Unlabeled phenyl isocyanate (PIC) and stable-isotope-labeled phenyl isocyanate (*PIC) were combined in a volumetric ratio of 1:1 and 1:50, and then allowed to react with angiotensin I; the reaction scheme is shown in **Figure 4.1 A**. After derivatization, the 1:1 and 1:50 PIC/*PIC-angiotensin I mixtures (total concentration 1μM) were analyzed by ESI-FT-ICR mass spectrometry. Several FT-ICR parameters, including a wide range of excitation and trapping voltages, two different excitation waveforms (chirp and SWIFT [42]), and three different acquisition times were systematically varied for each acquisition. Representative spectra for the 1:1 and 1:50 PIC/*PIC-angiotensin I mixtures are shown in **Figure 4.1 B** and **4.1 C**, respectively. An enlargement of the mass spectrum showing the

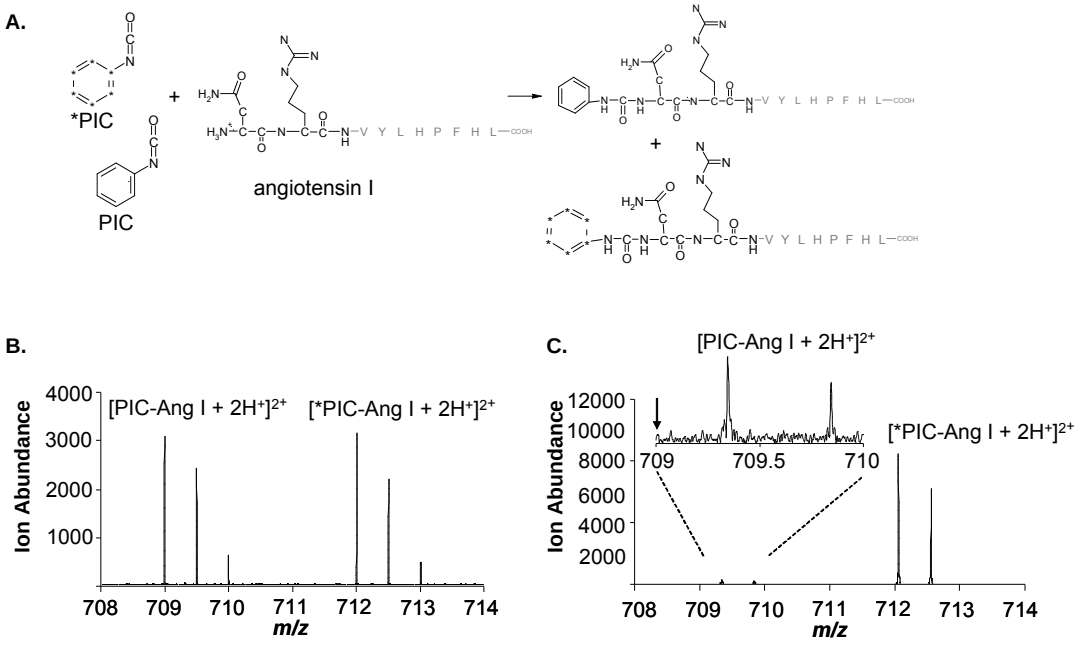


Figure 4.1 (A) Reaction of unlabeled- and stable-isotope labeled phenyl isocyanate (PIC) with angiotensin I. **(B)** Mass spectrum of the 1:1 PIC/*PIC-angiotensin I mixture (total concentration 1 μM) acquired using the SWIFT excitation waveform with an excitation voltage of 210 V and a trapping voltage of 2.2 V. **(C)** Mass spectrum of the 1:50 PIC/*PIC-angiotensin I mixture (total concentration 1 μM). An enlargement of the region showing PIC-angiotensin I isotopic distribution is inset; the arrow points to the *m/z* value of the PIC-angiotensin I in the 1:1 mixture. Note the frequency shift.

PIC-angiotensin I from the 1:50 mixture is shown (inset **Figure 4.1 C**). The PIC-angiotensin I in the 1:50 mixture is shifted to a higher *m/z* compared to the same species in the 1:1 mixture, which we attribute to space-charge induced frequency perturbations [22, 25, 44, 45].

4.3.2 Signal Decay Rate

To extend the investigations of de Koning, *et al.* and Gordon, *et al.* [25, 32], we examined the role of acquisition time on the signal decay rate (τ), calculated as described in **Section 4.2.5.1**, for both the unlabeled and labeled species over a wide range of experimental conditions. The ratio of signal decay rates for the PIC- to *PIC-angiotensin I were then compared for each set of conditions. In order to adequately assess and compare the different experimental parameters, we developed a multiplicative quality factor (MQF), which incorporates the total ion abundance (A_T), average relative error and coefficient of variation of the ion abundance ratio of PIC-/*PIC-angiotensin I. The MQF combines these three analytical figures of merit into one value, thus simplifying the dataset. To calculate the MQF, first the A_T for each spectrum was normalized to the spectrum with the largest A_T ($A_{T,max}$), irrespective of the experimental parameters (*i.e.,* across all spectra). Both the average relative error and coefficient of variation for the abundance ratio of the monoisotopic peak for PIC- to *PIC-angiotensin I, calculated for each set of parameters, were then subtracted from one. Consequently, the best value that can be achieved (0% average relative error, 0% coefficient of variation) becomes 1. These three values were multiplied to calculate the MQF, shown in **Equation 4.2**; optimal experimental conditions for quantification will have a MQF value approaching 1.

$$MQF = \frac{A_T}{A_{T,\max}} \times \left(1 - \frac{\left(\frac{A_{PIC}}{A_{*PIC}} \right)_{MS} - \left(\frac{A_{PIC}}{A_{*PIC}} \right)_{LC-UV}}{\left(\frac{A_{PIC}}{A_{*PIC}} \right)_{LC-UV}} \right) \times \left(1 - \frac{s_{\left(\frac{A_{PIC}}{A_{*PIC}} \right)_{MS}}}{x_{\left(\frac{A_{PIC}}{A_{*PIC}} \right)_{MS}}} \right) \qquad (\mathbf{4.2})$$

In **Equation 4.2**, A_{PIC} and A_{*PIC} are the abundances of PIC-angiotensin I and *PIC-angiotensin I, respectively, determined from ESI-FT-ICR mass spectra (MS) or LC-UV chromatograms (LC-UV). The sample standard deviation is represented by the symbol s, and $\bar{x}$ signifies the average. Although accuracy and precision can be maximized without a corresponding improvement in total ion abundance, dynamic range suffers. In other words, the absolute ion abundance is important for quantitative measurements of species with vastly different concentrations. Therefore, we placed equal weight on total ion abundance, accuracy, and precision in the calculation of MQF. The analysis of more complex mixtures may warrant different weighting functions for each of the metrics in future studies.

In **Figure 4.2**, the ratio of the signal decay rate for the 1:1 PIC/*PIC-angiotensin I is plotted as a function of MQF value. Not surprisingly, instrumental parameters that result in a τ ratio nearer to 1 generally have MQF values approaching 1. This result corroborates previous observations that accurate quantification can be achieved if two species have similar signal decay rates [31]. Importantly, our results extend this observation to include a wider range of experimental conditions.

Figure 4.2 also demonstrates MQF improves with a decrease in acquisition time over all experimental parameters investigated. These results lend to the argument that shorter acquisition times produce more accurate ion abundance ratios (*i.e.,* peak height ratios) [25, 32, 33, 36]. Improvements in MQF were also observed for the 1:50 dataset when the acquisition time was decreased (data not shown). However, owing to our chosen τ-calculation method, the τ ratio could not be determined for many of the experimental parameters, thus is was difficult to observe trends for the 1:50 dataset.

In addition to the effect of acquisition time, we also explored the role of the type of excitation waveform in the distribution of τ ratios as a function of the MQF value. The τ ratios for the chirp excitation dataset are clustered around low MQF values with only a few

experimental conditions producing optimal MQF values. Furthermore, τ ratios from the chirp excitation dataset have a broader y-axis distribution. In contrast, SWIFT excitation generally results in τ ratios nearer to 1.

The data presented in **Figure 4.2** demonstrates that species with similar decay rates, resulting in a τ ratio near 1, provide the most accurate quantification measurements. Moreover, SWIFT excitation with shorter acquisition times are more likely to result in accurate abundance ratios.

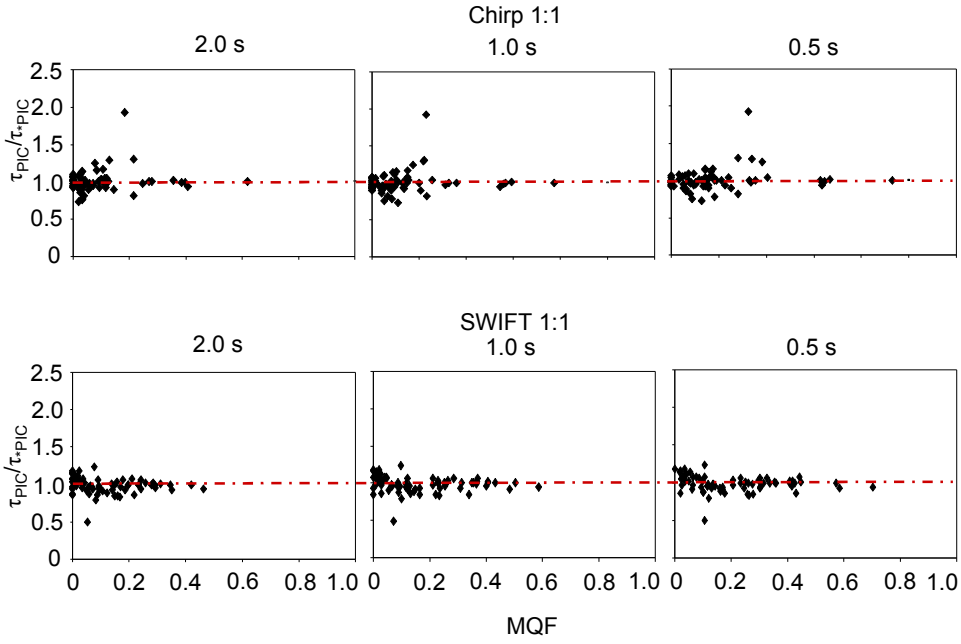


Figure 4.2 Signal decay rates of PIC/*PIC-angiotensin I. The ratio of the signal decay rates for the PIC-angiotensin I to *PIC-angiotensin I are plotted versus the multiplicative quality factor (MQF). The top and bottom three plots are from the chirp excitation dataset and SWIFT excitation dataset, respectively. The acquisition times of 2 s, 1s, and 0.5 s are shown at the top of each plot. The dotted line is placed at the ideal τ ratio of one.

4.3.3 Evaluation of Optimal Instrumental Parameters for Equal Ion Populations

Initially, the optimal instrumental parameters for quantification were explored for a system consisting of two equal population ion clouds. This case represents an "ideal" system, and therefore we expected the optimal instrumental parameters (MQF values approaching 1) to encompass a wide range of experimental parameters. **Figure 4.3** displays several heat maps showing the MQF for each experimental parameter investigated.

In the top and bottom rows are three heat maps, which correspond to the chirp and SWIFT excitation datasets and to each acquisition time, respectively. A decrease in acquisition time improves the MQF values and expands the range of optimal excitation and trapping voltages for both excitation waveforms.

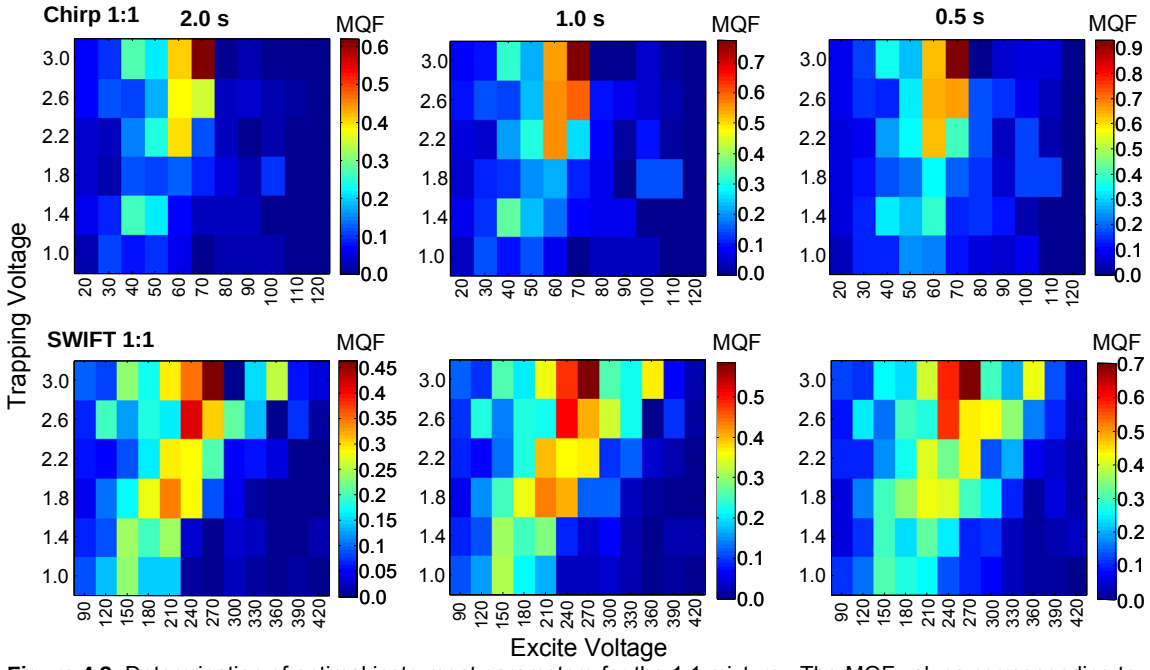


Figure 4.3 Determination of optimal instrument parameters for the 1:1 mixture. The MQF values corresponding to each excitation voltage and trapping voltage combination are displayed in heat maps for the 1:1 PIC-/*PIC-angiotensin I. The top and bottom rows of heat maps represent the chirp and SWIFT excitation datasets, respectively. Additionally, there are three heat maps for each excitation waveform that correspond to the three acquisition times. More red colors indicate optimal conditions.

Chirp excitation generated a higher maximum MQF value, but a closer examination of **Figure 4.3** reveals that only the 70 V excitation and 3.0 V trapping combination resulted in a higher MQF. The remaining combinations of excitation and trapping voltages produced similar, if not lower, MQF values as the SWIFT dataset. Since chirp excitation did not result in a higher MQF value for the 1:50 dataset (*vide infra*), the higher MQF values for the 1:1 dataset may be coincidental.

Generally, SWIFT excitation results in a wider range of optimal excitation and trapping voltages, which follow a diagonal trend in the MQF values. The MQF calculation incorporates total ion abundance (A_T), which is known to increase with increasing post-

excite radius (shown in **Equation 4.1**). Consequently, it is no surprise that larger excitation voltages (*i.e.,* post-excite radii) generated larger MQF values. Higher excitation voltages necessitate higher trapping voltages to axially confine the ions, thus the diagonal trend. Moreover, excitation and trapping voltages with a higher A_T also resulted in high accuracy and precision, and produced τ ratios near 1 (see **Figure 4.2**).

Many of the experimental conditions investigated produced an abundance ratio within 10% accuracy and precision, which is within acceptable limits for quantitative proteomics measurements. Chirp excitation yielded accuracy and precision values of the PIC-/*PIC-angiotensin abundance ratio within 10% for 20% of the 2 s acquisition dataset (*i.e.,* combination of excitation and trapping voltage), 26% of the 1 s acquisition dataset, and 38% of the 0.5 s acquisition dataset. The SWIFT dataset encompassed more experimental combinations with 33%, 47%, and 54% of the combinations generating accuracy and precision values within 10% for the 2 s, 1 s, and 0.5 s acquisition times, correspondingly.

In **Table 4.1** the average percent relative error (accuracy) and coefficient of variation (precision) for the PIC/*PIC-angiotensin I abundance ratio are listed for the optimal experimental parameters for both chirp and SWIFT excitation, which match to red or dark orange squares in the corresponding heat map shown in **Figure 4.3**. The empirically-determined optimal excitation and trapping voltages are defined as conditions that produce a MQF value within 0.1 of the maximum MQF for that dataset. These optimal values are

Table 4.1 Accuracy and precision of optimal experimental conditions (within 0.1 of the maximum MQF for that dataset) for 1:1 mixture (total concentration 1 μM)

$T_{acq'n}$ (s)			2 s		1 s		0.5 s	
Waveform	**Excite (V)**	**Trapping (V)**	**% RE[1]**	**% CV[2]**	**% RE[1]**	**% CV[2]**	**% RE[1]**	**% CV[2]**
Chirp	70	3.0	6.5	2.0	6.6	1.4	5.4	1.5
SWIFT	240	2.6	8.1	1.0	7.9	1.0	not optimal[3]	
	270	3.0	7.2	3.9	7.1	2.7	6.6	2.1

[1]Percent relative error of PIC/*PIC-angiotensin I abundance ratio determined by FT-ICR MS from LC-UV ratio
[2]Coefficient of variation of PIC/*PIC-angiotensin I ratio determined by FT-ICR MS
[3]MQF value does not fall within 0.1 of maximum MQF for that dataset

actually contained within a range of optimal voltages, whose width is determined by the instrument precision. Notably, all the values listed in **Table 4.1** are within 10% accuracy and precision.

4.3.4 Evaluation of Optimal Instrumental Parameters for Unequal Ion Populations

One of the most challenging aspects of quantification in proteomics is the wide concentration range of proteins present in the biological matrix [46]. Driven by this challenge, the next step in our evaluation of optimal instrumental parameters was to investigate the behavior of a 1:50 PIC/*PIC-angiotensin I mixture. The results from these experiments are represented by heat maps shown in **Figure 4.4**.

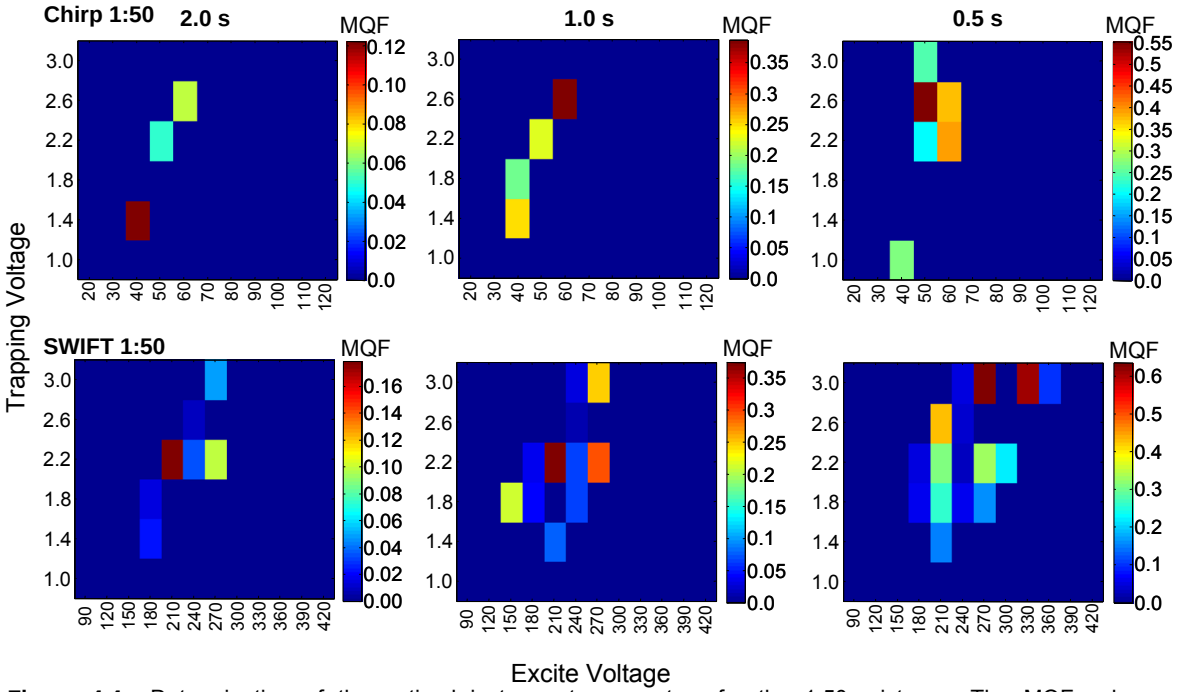


Figure 4.4 Determination of the optimal instrument parameters for the 1:50 mixture. The MQF values corresponding to each excitation voltage and trapping voltage combination are displayed in heat maps for the 1:50 PIC-/*PIC-angiotensin I. The top and bottom rows of heat maps represent the chirp and SWIFT excitation datasets, respectively. Additionally, there are three heat maps for each excitation waveform that correspond to the three acquisition times. More red colors indicate optimal conditions. Note the overall decrease in the scales of MQF values and the decreased range of optimal conditions from the 1:1 dataset.

The most apparent difference from the 1:1 data is the range of optimal excitation and trapping voltages is narrowed considerably. Moreover, the MQF values overall are smaller for the 1:50 dataset, which is a result of the reduced accuracy and precision of the abundance ratio. Although the optimal 1:50 MQF values also produced optimal MQF values for the 1:1 data, the best excitation and trapping voltage combination are not the same between the 1:1 and 1:50 datasets. For example, the excitation and trapping voltages that resulted in the best MQF for the SWIFT 2 s acquisition dataset were 270 V, 3.0 V for the 1:1 mixture and 210 V, 2.2 V for the 1:50 mixture. This difference in optimal experimental parameters for mixtures of species with equal and unequal concentrations should be considered during the experimental set up, especially for complex mixture analysis where species are present with a wide range of concentrations.

Interestingly, the maximum MQF values for the chirp and SWIFT 1:50 dataset are basically the same, so there is no distinct advantage for a particular excitation waveform with respect to the accuracy and precision of relative abundances for a system with unequal ion populations. However, SWIFT excitation produces optimal MQF values with a wider range of excitation and trapping voltages, which is advantageous for complex mixtures containing many species present at varying relative concentrations.

Table 4.2 lists the accuracy and precision for the optimal experimental parameters, defined as resulting in a MQF value within 0.1 of the maximum MQF for that dataset and correspond to red and dark orange squares in **Figure 4.4**. Unlike the 1:1 dataset for which there were many conditions resulting in acceptable accuracy and precision values, only two combinations of excitation and trapping voltages produced accuracy and precision values within 10%, which are both from the SWIFT dataset. Interestingly, only one of these combinations results in an optimal MQF value (0.61). The second combination of 300 V, 2.2 V for excitation and trapping voltages ($T_{acq'n}$ = 0.5s) resulted in a mass spectrum with a low total ion abundance, and therefore low MQF value (0.23). Our laboratory is currently

investigating methods to increase ion abundances, and therefore improve the limits of detection [47-49]. These experimental strategies would also improve the accuracy and precision of quantitative proteomics measurements.

Table 4.2. Accuracy and precision of optimal experimental conditions (within 0.1 of the maximum MQF value for that dataset) for 1:50 mixture (total concentration1 μM)

$T_{acq'n}$ (s)			2 s		1 s		0.5 s	
Waveform	Excite (V)	Trapping (V)	% RE[1]	% CV[2]	% RE[1]	% CV[2]	% RE[1]	% CV[2]
Chirp	40	1.4	55.8	37.8	not optimal[3]		not optimal[3]	
	50	2.6	not optimal[3]		not optimal[3]		11.6	18.1
	60	2.6	not optimal[3]		25.6	19.6	not optimal[3]	
SWIFT	210	2.2	16.5	25.3	19.2	11.6	not optimal[3]	
	270	2.2	not optimal[3]		14.8	18.6	not optimal[3]	
	270	3.0	not optimal[3]		not optimal[3]		11.4	19.7
	330	3.0	not optimal[3]		not optimal[3]		3.6	4.9

[1]Percent relative error of PIC/*PIC-angiotensin I abundance ratio determined by FT-ICR MS from LC-UV ratio

[2]Coefficient of variation of PIC/*PIC-angiotensin I ratio determined by FT-ICR MS

[3]MQF value does not fall within 0.1 of maximum MQF for that dataset

4.4 Conclusions and Future Experiments

We have shown that small changes in experimental conditions can significantly affect accuracy and precision of quantitative proteomics measurements by ESI-FT-ICR mass spectrometry, especially for systems with unequal ion populations. These results have important implications for the quantification of proteomes, and therefore experimental parameters should be carefully considered prior to data acquisition to ensure accurate abundance ratios.

The systematic exploration of several of the instrumental parameters that are known to affect quantification, explicitly acquisition time, excitation voltage (*i.e.,* post-excite radius), trapping voltage, excitation waveform, afford a thorough assessment of the optimal parameters for quantitative proteomics measurements. Moreover, the multiplicative quality factor (MQF) combines the total ion abundance, accuracy, and precision into one value, thus simplifying large datasets. Ideally, MQF values could be used to optimize instrumental parameters before commencing quantification experiments, and used for all subsequent experiments for a particular instrument.

Future experiments are planned to extend these investigations to include the role of complexity on the accuracy of quantitative proteomics measurements by ESI-FT-ICR mass spectrometry. We will also explore the effect of acquisition time, excitation voltage, and trapping voltages on the accuracy and precision of quantification using MQF values for a more complex system. However, a narrower range and a higher resolution of excitation and trapping voltages will be investigated in the next set of experiments. Moreover, we will focus our next investigations on SWIFT excitation, owing to the improved accuracy and precision over chirp excitation. Additionally, the signal decay rate of both the PIC- and *PIC-angiotensin I in the complex mixture will be examined. Ultimately, the data presented and acquired from planned experiments will be utilized to develop a correction factor for the signal decay rate, which when employed would improve the accuracy of quantitative proteomics measurements.

4.5 References

1. Gygi, SP, Rist, B, Gerber, SA, Turecek, F, Gelb, MH, and Aebersold, R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnology,* 1999. **17**: 994-999.

2. Ong, SE, Blagoev, B, Kratchmarova, I, Kristensen, DB, Steen, H, Pandey, A, et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & Cellular Proteomics,* 2002. **1**: 376-386.

3. Mirgorodskaya, OA, Kozmin, YP, Titov, MI, Korner, R, Sonksen, CP, and Roepstorff, P. Quantitation of peptides and proteins by matrix-assisted laser desorption/ionization mass spectrometry using O-18-labeled internal standards. *Rapid Communications in Mass Spectrometry,* 2000. **14**: 1226-1232.

4. Ross, PL, Huang, YN, Marchese, JN, Williamson, B, Parker, K, Hattan, S, et al. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Molecular & Cellular Proteomics,* 2004. **3**: 1154-1169.

5. Wiener, MC, Sachs, JR, Deyanova, EG, and Yates, NA. Differential mass spectrometry: A label-free LC-MS method for finding significant differences in complex peptide and protein mixtures. *Analytical Chemistry,* 2004. **76**: 6085-6096.

6. Johnson, KL, Mason, CJ, Muddiman, DC, and Eckel, JE. 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. *Analytical Chemistry,* 2004. **76**: 5097-5103.

7. Ong, SE, Foster, LJ, and Mann, M. Mass spectrometric-based approaches in quantitative proteomics. *Methods,* 2003. **29**: 124-130.

8. Ong, SE and Mann, M. Mass spectrometry-based proteomics turns quantitative. *Nature Chemical Biology,* 2005. **1**: 252-262.

9. Bronstrup, M. Absolute quantification strategies in proteomics based on mass spectrometry. *Expert Review of Proteomics,* 2004. **1**: 503-512.

10. Aebersold, R and Mann, M. Mass spectrometry-based proteomics. *Nature,* 2003. **422**: 198-207.

11. Hawkridge, AM, Heublein, DM, Bergen, HR, Cataliotti, A, Burnett, JC, and Muddiman, DC. Quantitative mass spectral evidence for the absence of circulating brain natriuretic peptide (BNP-32) in severe human heart failure. *Proceedings of the National Academy of Sciences of the United States of America,* 2005. **102**: 17442-17447.

12. Marshall, AG. Milestones in Fourier transform ion cyclotron resonance mass spectrometry technique development. *International Journal of Mass Spectrometry,* 2000. **200**: 331-356.

13. Grosshans, PB and Marshall, AG. General-theory of excitation in ion-cyclotron resonance mass-spectrometry. *Analytical Chemistry,* 1991. **63**: 2057-2061.

14. Marshall, AG. Fourier-transform ion-cyclotron resonance mass-spectrometry. *Accounts of Chemical Research,* 1985. **18**: 316-322.

15. Marshall, AG, Hendrickson, CL, and Jackson, GS. Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrometry Reviews,* 1998. **17**: 1-35.

16. Marshall, AG and Hendrickson, CL. Fourier transform ion cyclotron resonance detection: Principles and experimental configurations. *International Journal of Mass Spectrometry,* 2002. **215**: 59-75.

17. Frahm, JL, Howard, BE, Heber, S, and Muddiman, DC. Accessible proteomics space and its implications for peak capacity for zero-, one- and two-dimensional separations coupled with FT-ICR and TOF mass spectrometry. *Journal of Mass Spectrometry,* 2006. **41**: 281-288.

18. Shen, YF, Zhao, R, Belov, ME, Conrads, TP, Anderson, GA, Tang, KQ, et al. Packed capillary reversed-phase liquid chromatography with high-performance electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry for proteomics. *Analytical Chemistry,* 2001. **73**: 1766-1775.

19. Comisarow, MB. Theory of Fourier-transform ion-cyclotron resonance mass-spectroscopy .2. Signal modeling for ion-cyclotron resonance. *Journal of Chemical Physics,* 1978. **69**: 4097-4104.

20. Gorshkov, MV and Nikolaev, EN. Optimal cyclotron radius for high-resolution FT-ICR spectrometry. *International Journal of Mass Spectrometry and Ion Processes,* 1993. **125**: 1-8.

21. Uechi, GT and Dunbar, RC. Space-charge effects on relative peak heights in Fourier transform-ion cyclotron-resonance spectra. *Journal of the American Society for Mass Spectrometry,* 1992. **3**: 734-741.

22. Taylor, PK and Amster, IJ. Space charge effects on mass accuracy for multiply charged ions in ESI-FTICR. *International Journal of Mass Spectrometry,* 2003. **222**: 351-361.

23. Ledford, EB, Ghaderi, S, White, RL, Spencer, RB, Kulkarni, PS, Wilkins, CL, et al. Exact mass measurement by Fourier-transform mass-spectrometry. *Analytical Chemistry,* 1980. **52**: 463-468.

24. Hawkridge, AM, Nepomuceno, AI, Lovik, SL, Mason, CJ, and Muddiman, DC. Effect of post-excitation radius on ion abundance, mass measurement accuracy, and isotopic distributions in Fourier transform ion cyclotron resonance mass spectrometry. *Rapid Communications in Mass Spectrometry,* 2005. **19**: 915-918.

25. Gordon, EF and Muddiman, DC. Impact of ion cloud densities on the measurement of relative ion abundances in Fourier transform ion cyclotron resonance mass

spectrometry: Experimental observations of coulombically induced cyclotron radius perturbations and ion cloud dephasing rates. *Journal of Mass Spectrometry,* 2001. **36**: 195-203.

26. Jackson, GS, Canterbury, JD, Guan, SH, and Marshall, AG. Linearity and quadrupolarity of tetragonal and cylindrical penning traps of arbitrary length-to-width ratio. *Journal of the American Society for Mass Spectrometry,* 1997. **8**: 283-293.

27. Beu, SC and Laude, DA. Open trapped ion cell geometries for Fourier-transform ion-cyclotron resonance mass-spectrometry. *International Journal of Mass Spectrometry and Ion Processes,* 1992. **112**: 215-230.

28. Guan, SH and Marshall, AG. Ion traps for Fourier-transform ion-cyclotron resonance mass-spectrometry - principles and design of geometric and electric configurations. *International Journal of Mass Spectrometry and Ion Processes,* 1995. **146**: 261-296.

29. Amster, IJ. Fourier transform mass spectrometry. *Journal of Mass Spectrometry,* 1996. **31**: 1325-1337.

30. Hofstadler, SA, Bruce, JE, Rockwood, AL, Anderson, GA, Winger, BE, and Smith, RD. Isotopic beat patterns in Fourier-transform ion-cyclotron resonance mass-spectrometry - implications for high-resolution mass measurements of large biopolymers. *International Journal of Mass Spectrometry,* 1994. **132**: 109-127.

31. Gordon, EF and Muddiman, DC. Quantification of singly charged biomolecules by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry utilizing an internal standard. *Rapid Communications in Mass Spectrometry,* 1999. **13**: 164-171.

32. de Koning, LJ, Kort, CWF, Pinkse, FA, and Nibbering, NMM. Segmented Fourier-transform and its application to Fourier-transform ion-cyclotron resonance (FT-ICR) mass-spectrometry - ion abundances and mass measurements. *International Journal of Mass Spectrometry and Ion Processes,* 1989. **95**: 71-92.

33. Mitchell, DW and Smith, RD. Cyclotron motion of two coulombically interacting ion clouds with implications to Fourier-transform ion-cyclotron resonance mass-spectrometry. *Physical Review E,* 1995. **52**: 4366-4386.

34. Peurrung, AJ and Kouzes, RT. Long-term coherence of the cyclotron mode in a trapped ion cloud. *Physical Review E,* 1994. **49**: 4362-4368.

35. Kaiser, NK and Bruce, JE. Observation of increased ion cyclotron resonance signal duration through electric field perturbations. *Analytical Chemistry,* 2005. **77**: 5973-5981.

36. Bresson, JA, Anderson, GA, Bruce, JE, and Smith, RD. Improved isotopic abundance measurements for high resolution Fourier transform ion cyclotron resonance mass spectra via time-domain data extraction. *Journal of the American Society for Mass Spectrometry,* 1998. **9**: 799-804.

37. Farrar, TC, Elling, JW, and Krahling, MD. Application of linear prediction to Fourier-transform ion-cyclotron resonance signals for accurate relative ion abundance measurements. *Analytical Chemistry,* 1992. **64**: 2770-2774.

38. Goodner, KL, Milgram, KE, Williams, KR, Watson, CH, and Eyler, JR. Quantitation of ion abundances in Fourier transform ion cyclotron resonance mass spectrometry. *Journal of the American Society for Mass Spectrometry,* 1998. **9**: 1204-1212.

39. Marshall, AG, Comisarow, MB, and Parisod, G. Theory of Fourier-transform ion-cyclotron resonance mass spectroscopy-III .1. Relaxation and spectral-line shape in Fourier-transform ion resonance spectroscopy. *Journal of Chemical Physics,* 1979. **71**: 4434-4444.

40. Du, Y, Parks, BA, Sohn, S, Kwast, KE, and Kelleher, NL. Top-down approaches for measuring expression ratios of intact yeast proteins using Fourier transform mass spectrometry. *Analytical Chemistry,* 2006. **78**: 686-694.

41. Mason, DE and Liebler, DC. Quantitative analysis of modified proteins by LC-MS/MS of peptides labeled with phenyl isocyanate. *Journal of Proteome Research,* 2003. **2**: 265-272.

42. Guan, SH and Marshall, AG. Stored waveform inverse Fourier transform (SWIFT) ion excitation in trapped-ion mass spectrometry: Theory and applications. *International Journal of Mass Spectrometry and Ion Processes,* 1996. **158**: 5-37.

43. Anderson, GA, Bruce, JE, and Smith, RD. *ICR-2LS*. 1996: Richland.

44. Francl, TJ, Sherman, MG, Hunter, RL, Locke, MJ, Bowers, WD, and McIver, RT. Experimental-determination of the effects of space-charge on ion-cyclotron resonance frequencies. *International Journal of Mass Spectrometry and Ion Processes,* 1983. **54**: 189-199.

45. Jeffries, JB, Barlow, SE, and Dunn, GH. Theory of space-charge shift of ion-cyclotron resonance frequencies. *International Journal of Mass Spectrometry and Ion Processes,* 1983. **54**: 169-187.

46. Anderson, NL and Anderson, NG. The human plasma proteome - history, character, and diagnostic prospects. *Molecular & Cellular Proteomics,* 2002. **1**: 845-867.

47. Hawkridge, AM, Zhou, L, Lee, ML, and Muddiman, DC. Analytical performance of a Venturi device integrated into an electrospray ionization Fourier transform ion cyclotron resonance mass spectrometer for analysis of nucleic acids. *Analytical Chemistry,* 2004. **76**: 4118-4122.

48. Frahm, JL, Burke, MJ, and Muddiman, DC. Leveling response factors in the electrospray ionization process using a heated capillary interface. *Journal of the American Society for Mass Spectrometry,* 2005. **16**: 772-778.

49. Frahm, JL, Bori, ID, Comins, DL, Hawkridge, AM, and Muddiman, DC. Achieving augmented limits of detection for peptides with hydrophobic alkyl tags (ALiPHAT). *Analytical Chemistry,* 2007. **accepted, Accelerated Article**.

CHAPTER 5

Leveling Electrospray Ionization Response Factors of Biomolecules Using a Heated Capillary Interface

5.1 Introduction

Electrospray ionization (ESI) has become indispensable in the analysis of large biomolecules. Although ESI mechanisms may not be completely characterized, investigators have recognized numerous factors that influence ESI response. One of these factors is the analyte's hydrophobicity. The correlation between ESI response and hydrophobicity of the analyte has been the subject of several investigations [1-5].

The effect an analyte's hydrophobicity on ESI response can be quantitatively estimated using an Arrhenius-type rate equation described by Fenn [1]:

$$N_{iz} = 3Ae^{\left(-\Delta G_{iz}^{o} / RT\right)}\left(N_i / r\right)e^{\left(\Delta r^{*} zQ / 4\pi\varepsilon_0 RTr^2\right)} \quad (\mathbf{5.1})$$

Equation 5.1 estimates the ion flux from the droplet surface, N_{iz}, of an ion *i* with *z* charges. Ion flux is a function of a constant (*A*) that relates bulk concentration to the surface density, free energy of solvation (ΔG_{iz}°), gas constant (*R*), temperature (*T*), number of moles (N_i), droplet radius (*r*), distance an ion has to travel to be free of the droplet (Δr^{*}), number of excess charges (*Q*), and the gas permittivity constant (ε_0). The variables *A,* related to surface activity, and ΔG_{iz}°, the amount of work necessary to desorb the ion from the droplet, are characteristic to the species in solution [5]. The remaining variables will be constant for a given droplet.

Null, *et al.* further developed **Equation 5.1** in order to predict the ESI responses of various oligonucleotides using their base composition; this predictor of ESI response is an analyte's ion flux score [5]. An ion flux score is an estimation of the ion flux (**Equation 5.1**) for an oligonucleotide calculated when the first offspring droplet reaches 75% of the Rayleigh

limit during the electrospray process normalized to the length of the oligonucleotide (number of bases in the sequence) [5]. Species with higher ion flux scores will be less abundant in the mass spectrum.

A correction to the values reported for the ion flux and the ion flux scores is presented here. The ion flux values should be 2.1×10^{-28}, 2.4×10^{-33}, 2.4×10^{-36}, and 1.2×10^{-28} mol for A, C, G, and T, respectively, when the nucleobase desorbs from the first offspring droplet at 75% of the Rayleigh limit. The negative log of the ion flux values is then used to determine the ion flux score, shown in **Equation 5.2**. **Equation 5.2** is the corrected version of equation 3 reported in reference [5], which has been amended to reflect the ion flux corrections.

$$ion\ flux\ score = \frac{\sum -\log(N_{iz})}{L} = \frac{(\#A \times 27.7) + (\#C \times 32.6) + (\#G \times 35.6) + (\#T \times 27.9)}{L} \quad (\mathbf{5.2})$$

Ion flux scores suggest a semi-quantitative estimation of ion abundances that can be used to compare oligonucleotides. However, owing to the many assumptions required to develop **Equation 5.2** and the dynamic nature of the electrospray process, any quantitative evaluation, albeit a crude approximation is difficult to make. Therefore, Null, *et al.* also utilized the number of guanines in an oligonucleotide to predict relative ESI responses [5]. Guanine's lower hydrophobicity and more negative free energy of solvation relative to the other nucleobases result in its lower ion abundance (higher ion flux score). Counting the number of guanines is a simple and quick way to predict the relative ESI responses of oligonucleotides.

As **Equation 5.**1 suggests, an increase in the hydrophobicity of an analyte would improve its ion abundance in the mass spectrum. Empirical observations from our laboratory have corroborated the relationship between electrospray ionization response and hydrophobicity[4-6]. Null, *et al.* demonstrated that the addition of an alkyl chain (hydrophobic tail) to the 5' terminus of PCR amplicons selectively enhances the signal of the modified

strand over the complementary unmodified strand from denatured amplicons [5]. More recently, this strategy has been applied to peptides [6] (see **Chapter 6**). The addition of an alkyl chain increases the hydrophobicity of the molecule, consequently promoting its escape from the droplet [5]. As a molecule becomes more hydrophobic, the surface activity (encompassed in constant *A* in **Equation 5.1**) increases leading to higher ion flux (*i.e.*, ion abundance).

Improvements in ion abundance have also been observed when the electrospray solution is heated [7]. Ikonomou and Kebarle describe two rationales for the increased ion abundance when heat is applied 1) better solvent evaporation, which thereby aids analyte fission from the droplet and 2) reduction of water surface tension as the temperature increases, which reduces the electric field required for the onset of electrospray and thus reduces the adverse effects of corona discharge [7].

Several investigators have reported the use of a heated electrospray ionization interface for a variety of applications [7-12]. Originally our laboratory designed a heated capillary interface in an attempt to denature duplex DNA [8]. Herein, we describe a completely redesigned source that is more stable allowing for broader application. During implementation of the new source, we observed a leveling effect on the ion abundance of two complementary strands of denatured DNA with different hydrophobicities as the temperature of the ESI solution was increased. Furthermore, we demonstrate that the heated capillary interface denatures duplex DNA, including amplicons containing the more stable 7-deaza purines. In 7-deaza analogs the N7 nitrogen is replaced with carbon, eliminating the preferred protonation site and thereby making the site resistant to fragmentation [13]. The reduced fragmentation facilitates gas phase sequencing by reducing the complexity of MS/MS spectra [14, 15]. Yet, the effect of the carbon substitution on the duplex stability is less clear. Grein, *et al.* report that 7-deaza purines stabilize the duplex, but this stability may depend on the sequence and the salt conditions [16]. Our laboratory has

observed that denaturation of oligonucleotides containing deaza analogs is more difficult compared to the unmodified oligonucleotides (unpublished observations).

5.2 Experimental

5.2.1 PCR Amplification

A 53-base pair region containing a single nucleotide polymorphism (SNP) (dbSNP, rs#717336) was amplified from synthetic template designed from a sequence on the long arm of chromosome 22 (SNP 22). Each PCR reaction contained 1X *AmpliTaq* Gold PCR buffer, 3 mM $MgCl_2$, 0.2 mM each deoxynucleotide triphosphate, 12.5 pmol of each primer (forward sequence, 5' GGG GCC ATC AGA GAA ATA TAC CT 3', reverse sequence, 5' CCA GGG TGG TAA TTT CCA GGT 3') (Midland Certified Reagent Co., Midland, TX), 1 fmol template (Midland Reagent Co., Midland, TX), and 1.25 U of *AmpliTaq* Gold (Applied Biosystems, Foster City, CA) in a total volume of 50 µL. Amplification was performed in a 96-well MJ Research PTC 200 DNA Engine System Peltier Thermal Cycler (Watertown, MA) using a hot-start PCR program. Cycling conditions were as follows: an initial ten-minute heat step at 95°C (required for enzyme activation), followed by 34 cycles at 95°C for 30 s (denaturation), 58°C for 45 s (annealing), and 72°C for 45 s (elongation). A final elongation step was accomplished at 72°C for 10 minutes. PCR reactions containing 7-deaza analogs were prepared in accordance with the above, except that 7-deaza-2'-deoxy-GTP (Roche Diagnostics Corp., Indianapolis, IN) and 7-deaza-2'-deoxy-ATP (TriLink BioTechnologies, San Diego, CA) were substituted for deoxyguanosine triphosphate and deoxyadenosine triphosphate, respectively in separate reactions with a final concentration of 0.4 mM per reaction. The resulting amplicons were purified before ESI-Fourier transform-ion cyclotron resonance (FT-ICR) mass spectrometry analysis according to standard practice consisting of ethanol precipitation [17, 18] and microdialysis [19, 20].

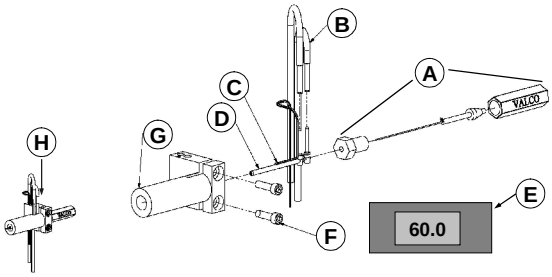
5.2.2 Mass Spectrometry

Mass spectra were acquired in negative-ion mode on a modified ESI-FT-ICR mass spectrometer (IonSpec FTMS Systems, Varian, Inc., Lake Forest, CA) with a 7 Tesla superconducting magnet (Cryomagnetics, Oak Ridge, TN). PCR products were electrosprayed from a 15 μm inner diameter PicoTip™ Emitter (New Objective, Woburn, MA) and remotely coupled to a potential of –2000 V. PCR products were electrosprayed from a buffer consisting of 60:20:20 acetonitrile/isopropanol/10 mM ammonium acetate in water with a final concentration of 20 mM each piperidine and imidazole and infused at a rate of 5 nL s^{-1}. Single-acquisition spectra were collected with 1024 k data points with a digitization rate of 1 MHz and a Hann window function applied, then zero-filled three times prior to fast-Fourier transform. Spectra were externally calibrated using polyethylene glycol with an average molecular weight of 1000 Da.

5.2.3 Heated Capillary Interface

A schematic of the heated capillary interface is shown in **Figure 5.1**. The new design allows for the heater and the electrospray transfer line (**Figure 5.1 A**) to remain separate entities, thus permitting easy replacement of the electrospray emitter as needed. Two electrical contact pins (**Figure 5.1 B**) are located at the upstream end of the ESI emitter spray. A thermocouple (**Figure 5.1 C**) is attached to two soldered concentric hypodermic needle tubes (**Figure 5.1 D**). The tight fit of the ESI emitter (**Figure 5.1 A**) in the inner tubing is too small to accommodate a thermocouple, so the thermocouple is attached only to the outer tubing (**Figure 5.1 C**). The outer tubing temperature is monitored by a K type thermocouple (**Figure 5.1 C**) (Omega 5SRTC-TT-K-36-36, Stamford, CT) bonded to the surface about one-third of the way from the emitter of the heater, where the temperature is near its maximum over the length of the heater. The set/measured temperature is displayed on a four digit LED readout (**Figure 5.1 E**). The thermocouple signal is fed to a closed loop

controller (Omega CNi3254) (**Figure 5.1 E**), which controls the current to the heater through a voltage to current converter circuit. The proportional, integral, derivative (PID) controller parameters are optimized to bring the temperature to the set point quickly and maintain the temperature within error limits (± 0.5˚C accuracy). PID parameters are proportional 42%, integral 9 resets/second, and derivative rate 2.3 seconds.



Part	Name
A	PicoTip™ Emitter
B	Heater Lead
C	Thermocouple
D	Capillary-nozzle heater assembly
E	Controller display
F	Hex screws
G	Heater insulator
H	Assembled heated ESI interface

Figure 5.1 Illustration of the heated capillary interface. (**A**) PicoTip™ Emitter (**B**) Heater lead (**C**) Thermocouple (**D**) Capillary-nozzle heater assembly (concentric hypodermic needles) (**E**) Controller display (**F**) Hex screws to adjust the fit of the insulator around the Valco fitting (**G**) Capillary-nozzle heater insulator (**H**) Assembled heated ESI interface.

The temperature of the inner tubing is linearly related to the outer temperature. Inner and outer tubing refers to the soldered concentric hypodermic needles in **Figure 5.1 D**. This relationship was empirically determined over the controller temperature range of 23.7 - 7.0°C; the relationship determined by linear least squares regression analysis is shown in **Equation 5.3**. All further discussion of temperatures of the heated capillary interface will refer to the inside temperature.

$$inner\ T = 1.4(outer\ T) - 10.4 \quad (\mathbf{5.3})$$

Current flows down the inner hypodermic tubing and back on the outer concentric tubing, heating both sections of tubing. Since the inner tubing has a smaller cross-sectional

area and therefore a higher resistance, it becomes hotter than the outer tubing. The outer tubing is still warm enough to limit heat losses from convection and radiation from the inner tubing.

5.3 Results and Discussion

Figure 5.2 shows an ESI-FT-ICR mass spectrum of the completely-denatured SNP 22 amplicon containing 7-deaza adenine, which is indicated by the presence of both the coding and noncoding strands and the absence of the double-stranded SNP 22 amplicon in the mass spectrum. In **Figure 5.2**, the monoadenylated noncoding strand is the sole species observed for the noncoding strand. Adenylation is an artifact of the *Taq* polymerase, which preferentially adenylates the 3' end during PCR amplification. The heated capillary interface also denatured the unmodified SNP 22 and SNP 22 containing 7-deaza guanine (data not shown). The ability of the newly fashioned heated capillary interface to denature amplicons containing 7-deaza analogs is a salient result, especially since the previous heated capillary interface [8] was unable to denature these species (unpublished observations).

We believe the heater element more efficiently and evenly heats the electrospray transfer line, primarily due to the smaller diameter of the heater. As the solution begins to heat, a gas layer begins to form on the inside perimeter of the capillary. Since this gas layer acts as an insulator, preventing even heating of the solution beneath the gas, a reduction in the diameter of the heated source thereby reduces the surface area to volume ratio providing more even heating. Furthermore, the new source was designed so that the heater element surrounds the electrospray emitter, thus allowing the solution to be heated just prior to electrospray droplet formation at the tip. In the previous design, the transfer line was heated further from the tip [8], allowing the liquid ample opportunity to cool before reaching the orifice of the mass spectrometer. The consequence of solution cooling may have

resulted in ineffective denaturation of oligonucleotides containing 7-deaza moieties with the previous source.

In **Figure 5.2 A**, a discrepancy in the ion abundance is apparent between the coding and monoadenylated noncoding strands obtained from the denatured SNP 22 amplicon (sequence shown in **Figure 5.2**). It is important to note that the solution is inherently equimolar in both strands since the species were obtained by denaturing the duplex amplicon. Therefore, the monoadenylated noncoding strand is underrepresented in the mass spectrum. The signal bias between the resulting complementary strands of this amplicon has been previously observed [5].

The observed signal bias between the coding and monoadenylated noncoding strands can be attributed to the difference in ion flux for the two species which is a result of the different base compositions of $A_{16}C_{15}G_{10}T_{12}$ and $A_{13}C_{10}G_{15}T_{16}$, respectively. The monoadenylated noncoding strand contains 15 guanines, whereas the coding strand only includes 10 guanines. A higher number of molecules of the coding strand are expected to escape the droplet (higher ion abundance) based on this measure of ion flux (*vide supra*).

When the heated capillary interface was utilized we observed an increase in the ion abundance of the monoadenylated noncoding strand relative to the ion abundance of the coding strand, shown in **Figure 5.2 B** and **C**. Additionally, an increase in temperature resulted in an improved signal-to-noise ratio, which is apparent from a comparison of **Figures 5.2 A-C**. Importantly, all spectra were collected under the same conditions, excluding temperature.

Figure 5.3 is a graph representing the ratio of ion abundances of the noncoding strand (NCS) to the coding strand (CS) as a function of temperature. The ratio of ion abundances approaches unity as the temperature increases, demonstrating the enhanced ion abundance of the noncoding strand (*i.e.*, less hydrophobic strand) with respect to the ion abundance of the coding strand. A statistically significant difference in the mean ion

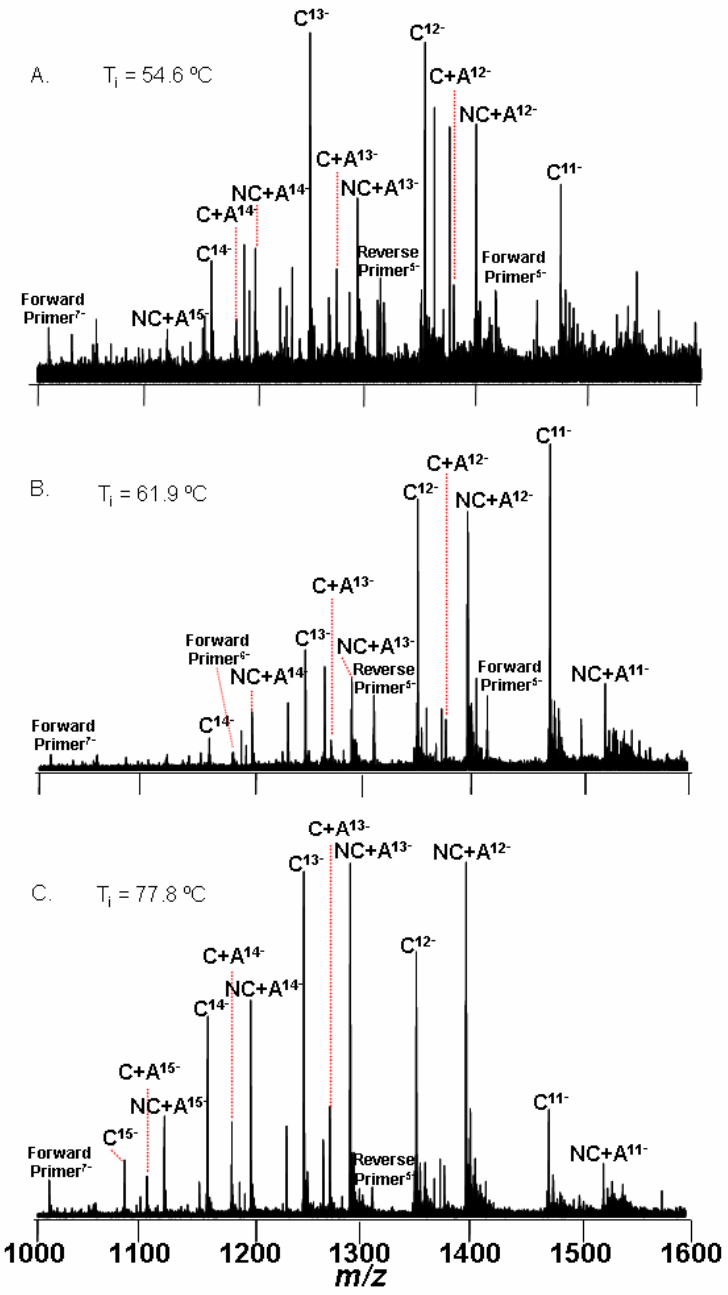


Figure 5.2 Single-acquisition ESI-FT-ICR mass spectra of the denatured SNP 22 amplicon containing 7-deaza adenine acquired at different temperatures; sequence is shown above the spectra. Underlined portions indicate the primer sequences and 7-deaza adenines are bolded. Primer peaks are labeled accordingly. Charge states are labeled according to the strand the peak originated from, either C (coding) or NC (noncoding). C+A and NC+A indicate the monoadenylated coding and noncoding strands, respectively. For each spectrum, the temperature inside the heater element (T_i) is listed. Mass spectrum of the amplicon acquired at (**A**) 54.6°C. (**B**) 61.9°C. (**C**) 77.8°C.

abundance was found between the 54.6°C and 77.8°C ($p = 0.0016$), 61.9°C and 77.8°C ($p = 0.0002$), 54.6°C and 80.7°C ($p = 0.0012$), and 61.9°C and 80.7°C ($p = 0.0002$) temperatures (two-sided two-sample t-test, Bonferroni-corrected alpha level = 0.008). Comparisons of temperatures to the 69.1°C and 74.8°C groups were not included as it is not statistically valid to include groups containing only one observation.

The increased ion abundance with increasing temperature is a consequence of the reduced free energy of solvation. The free energy of solvation can be defined as the activation-energy barrier required for the ion to escape the droplet supplied in the form of thermal energy [1]. At higher temperatures a larger number of ions will reach the required energy of activation. Therefore, a larger number of ions will escape the droplet resulting in increased ion abundance.

Moreover, we observed an increase in the average charge state at higher temperatures. Temperatures in the range of 54.6 - 69°C resulted in an average charge state of 12^- for the coding and noncoding strands, while temperatures in the range of 74.9 - 83.6°C produced an average charge state of 13^- for the coding and noncoding strands. Generally, ions with higher charge states indicate that the analyte desorbed later in the electrospray ionization process when there was a higher surface charge density [1, 5]. Alternatively, ions with higher charge states can be formed via a fast evaporation rate, which also increases the surface charge density [1]. Higher surface-charge densities of the droplet increase the probability that the ion will be in contact with more charges, which the ion will take upon desorption thus leading to a higher charge state [1]. In our experiments the heated capillary increases the evaporation rate by heating the electrospray solution.

One potential application of the heated capillary interface could be in the determination of allelic frequencies occurring in as little as 1% or less of the population. Previously, it has been shown that allelic frequencies occurring with a frequency of 1% in pooled DNA can be determined by increasing the electrospray ionization response using

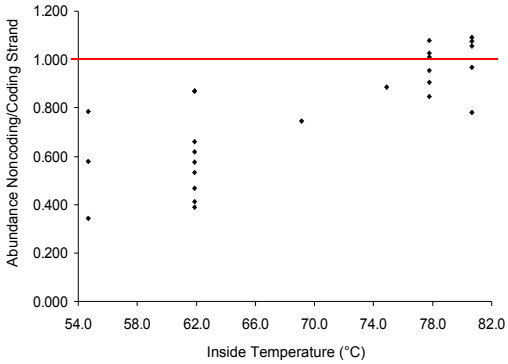


Figure 5.3 Plot of the ratio of ion abundances of the noncoding strand and coding strand versus the inside temperature of the heated ESI source. The line indicates equal ESI response for the noncoding strand and coding strand. Notice that as the temperature increases the relative ratio of ion abundances approaches one, indicating equal electrospray ionization response of both species.

primers with hydrophobic tails during PCR amplification [5]. Utilizing higher temperatures with the heated capillary interface to increase ion abundance, in addition to the use of primers with hydrophobic tails, the potential to detect species below the 1% level exists, which would be important for mutation sequence variations in the population that exist with a frequency of less than 1% of the population. Furthermore, the heated capillary interface could be applied in proteomics measurements to increase the electrospray ionization response of low-abundant peptides. The idea of increasing the ion abundance for proteomics applications is discussed further in **Chapter 6**.

5.4 Conclusions and Future Experiments

The heated capillary interface provides a simple means to obtain single-stranded species of DNA. Importantly, amplicons containing the more stable 7-deaza purines, compared to the parent purines, can be fully denatured. Information from both strands is retained (molecular weight) and can be used to delineate base composition (sequence) using a published algorithm [21]. Furthermore, the heated capillary interface can be used to improve the electrospray ionization response of species that are less hydrophobic. Usefulness of the heated capillary interface lies in its ability to aid the desolvation process thereby increasing ion abundance and improving signal-to-noise. The utility of the heated capillary interface is not limited to DNA; use of the source may be extended to other species, such as RNA, peptides, and proteins. Some applications including on-line LC-MS may also benefit from a heated capillary interface. We plan to investigate the utility of the heated capillary interface in proteomics applications.

5.5 References

1. Fenn, JB. Ion formation from charged droplets -Roles of geometry, energy, and time. *Journal of the American Society for Mass Spectrometry,* 1993. **4**: 524-535.

2. Cech, NB, Krone, JR, and Enke, CG. Predicting electrospray response from chromatographic retention time. *Analytical Chemistry,* 2001. **73**: 208-213.

3. Zhan, DL and Fenn, JB. Gas phase hydration of electrospray ions from small peptides. *International Journal of Mass Spectrometry,* 2002. **219**: 1-10.

4. Gordon, EF, Mansoori, BA, Carroll, CF, and Muddiman, DC. 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. *Journal of Mass Spectrometry,* 1999. **34**: 1055-1062.

5. Null, AP, Nepomuceno, AI, and Muddiman, DC. Implications of hydrophobicity and free energy of solvation for characterization of nucleic acids by electrospray ionization mass spectrometry. *Analytical Chemistry,* 2003. **75**: 1331-1339.

6. Frahm, JL, Bori, ID, Comins, DL, Hawkridge, AM, and Muddiman, DC. Achieving augmented limits of detection for peptides with hydrophobic alkyl tags (ALiPHAT). *Analytical Chemistry,* 2007. **accepted, Accelerated Article**.

7. Ikonomou, MG and Kebarle, P. A heated electrospray source for mass-spectrometry of analytes from aqueous-solutions. *Journal of the American Society for Mass Spectrometry,* 1994. **5**: 791-799.

8. Mangrum, JB, Flora, JW, and Muddiman, DC. Solution composition and thermal denaturation for the production of single-stranded PCR amplicons: Piperidine-induced destabilization of the DNA duplex? *Journal of the American Society for Mass Spectrometry,* 2002. **13**: 232-240.

9. Mirza, UA, Cohen, SL, and Chait, BT. Heat-induced conformational-changes in proteins studied by electrospray ionization mass-spectrometry. *Analytical Chemistry,* 1993. **65**: 1-6.

10. Daneshfar, R, Kitova, EN, and Klassen, JS. Determination of protein-ligand association thermochemistry using variable-temperature nanoelectrospray mass spectrometry. *Journal of the American Chemical Society,* 2004. **126**: 4786-4787.

11. Benesch, JLP, Sobott, F, and Robinson, CV. Thermal dissociation of multimeric protein complexes by using nanoelectrospray mass spectrometry. *Analytical Chemistry,* 2003. **75**: 2208-2214.

12. Rychlovsky, P, Cernoch, P, and Sklenickova, M. Application of a heated electrospray interface for on-line connection of the AAS detector with HPLC for detection of organotin and organolead compounds. *Analytical and Bioanalytical Chemistry,* 2002. **374**: 955-62.

13. Kirpekar, F, Nordhoff, E, Kristiansen, K, Roepstorff, P, Hahner, S, and Hillenkamp, F. 7-deaza purine-bases offer a higher ion stability in the analysis of DNA by matrix-assisted laser-desorption ionization mass-spectrometry. *Rapid Communications in Mass Spectrometry,* 1995. **9**: 525-531.

14. Hannis, JC and Muddiman, DC. Tailoring the gas-phase dissociation and determining the relative energy of activation for dissociation of 7-deaza purine modified oligonucleotides containing a repeating motif. *International Journal of Mass Spectrometry,* 2002. **219**: 139-150.

15. Frahm, JL, Mason, CJ, and Muddiman, DC. Utility of accurate monoisotopic mass measurements to confidently identify lambda exonuclease generated single-stranded amplicons containing 7-deaza analogs by electrospray ionization FT-ICR mass spectrometry. *International Journal of Mass Spectrometry,* 2004. **234**: 79-87.

16. Grein, T, Lampe, S, Mersmann, K, Rosemeyer, H, Thomas, H, and Seela, F. 3-deazapurines and 7-deazapurines - Duplex stability of oligonucleotides containing modified adenine or guanine bases. *Bioorganic & Medicinal Chemistry Letters,* 1994. **4**: 971-976.

17. Stults, JT and Marsters, JC. Improved electrospray ionization of synthetic oligodeoxynucleotides. *Rapid Communications in Mass Spectrometry,* 1991. **5**: 359-363.

18. Limbach, PA, Crain, PF, and McCloskey, JA. Molecular-mass measurement of intact ribonucleic-acids via electrospray-ionization quadrupole mass-spectrometry. *Journal of the American Society for Mass Spectrometry,* 1995. **6**: 27-39.

19. Liu, CL, Muddiman, DC, Tang, KQ, and Smith, RD. Improving the microdialysis procedure for electrospray ionization mass spectrometry of biological samples. *Journal of Mass Spectrometry,* 1997. **32**: 425-431.

20. Hannis, JC and Muddiman, DC. Characterization of a microdialysis approach to prepare polymerase chain reaction products for electrospray ionization mass spectrometry using on-line ultraviolet absorbance measurements and inductively coupled plasma-atomic emission spectroscopy. *Rapid Communications in Mass Spectrometry,* 1999. **13**: 323-330.

21. Muddiman, DC, Anderson, GA, Hofstadler, SA, and Smith, RD. Length and base composition of PCR-amplified nucleic acids using mass measurements from electrospray ionization mass spectrometry. *Analytical Chemistry,* 1997. **69**: 1543-1549.

CHAPTER 6

Improving the Electrospray Ionization Response of Cysteine-Containing Peptides with Hydrophobic Alkyl Tags

6.1 Introduction

Chemical tags have long been employed in the field of mass spectrometry for a multitude of purposes, thereby dramatically expanding the effectiveness of the measurements. Early in the history of chemical tagging for mass spectrometry investigators showed trimethylsilyl (TMS) derivatization improved the volatility of alcohols, thus allowing faster analysis of alcohols in hydrocarbon solutions without prior separation [1]. Later, the utility of TMS derivatization was applied to GC-MS analysis [2], which is now ubiquitous. Chemical derivatization has also been utilized to allow the ionization of typically nonionizable compounds by electrospray ionization [3-5]. Additionally, sulfonic acid tags added to the N-terminus of tryptic peptides has been shown to result in a single post-source decay (a tandem mass spectrometry alternative in MALDI TOF [6]) fragment ion series to facilitate identification [7].

The recent growing interest in exploiting the quantitative potential of mass spectrometry has also shifted the function of unlabeled and labeled chemical tags in this direction. Cysteine residues in peptides have been modified via alkylation [8] including the addition of a chlorine-containing tag [9] to improve identification, while ICAT™ [10], solid-phase ICAT [11], cleavable ICAT [12], and IDBEST™ (Target Discovery, Palo Alto, CA) [13] have allowed for the relative quantification of proteins in two different states. Additionally, reagents that target primary amine groups (*i.e.*, N-termini and lysine side chains) have been utilized for relative quantification [14, 15], including the iTRAQ reagent [16]. C-terminal tagging has also been explored; guanidination of C-terminal lysine for relative quantification is one C-terminal tagging strategy [17]. Moreover, tagging of phosphorylation sites has also been

developed for the relative quantification of phosphoproteins [18]. The numerous chemical tagging approaches for quantitative proteomics are presented in several recent reviews [19-21]. Clearly, the basic concept of combining chemical derivatization and mass spectrometry has a long and productive history.

Since the drafting of the human genome project and subsequent advances in available databases, mass spectrometry has emerged as a leading technology in proteome characterization [22-24]. In order to facilitate proteomic analysis, cystines are normally reduced and alkylated to eliminate the protein's tertiary structure [8, 25]. The alkylation step is a logical choice to exploit for the chemical modification a peptide or protein, since the reaction is simple and no further chemical labeling reactions would be required. Several chemical tagging strategies utilize alkylation chemistry, reviewed in references [20, 26]. The majority of these reports employ alkylation chemistry to introduce a stable-isotope label for quantitative measurements. In these studies we used alkylation chemistry to add a hydrophobic tag to cysteine-containing peptides in order to increase the electrospray ionization (ESI) response, thereby lowering the limit of detection. The notable features of this reaction are: 1) quantitative labeling of peptides under standard alkylation conditions, 2) increase in retention times by reversed phase-LC, 3) increase in ion abundance during ESI-MS.

Several investigators have previously demonstrated a correlation between ESI response (*i.e.,* ion abundance) and surface affinity, for review see reference [27]. Moreover, it has been shown that more hydrophobic (nonpolar character) analytes have larger ESI responses [28], which is attributable to greater surface affinity [27, 29-31]. Despite the name implication, hydrophobic molecules are driven to the surface in an attempt to reduce the surface free energy by reducing the number of hydrophobe-water van der Waals interactions and maximizing the number of water-water interactions (hydrogen bonding and

van der Waals), not because of repulsion forces between water and the hydrophobe. Increased ESI responses correspond to lower limits-of-detection.

Additionally, molecules that reside near the surface are in contact with a greater number of charges than molecules in the interior of the droplet [28]. However, surface molecules desorb from the droplet earlier in the ESI process, and therefore have a lower average charge state. Cech and Enke demonstrated via mathematical modeling that the amount of charged analyte is dependent on the molecule's surface affinity [32].

Early studies in our group have explored the role of hydrophobicity on relative quantification of heart cytochrome c from bovine and equine sources, which have 97 % sequence similarity [31]. Interestingly, the three amino acid difference results in GRAVY scores of -0.866 and -0.902 for bovine and equine cytochrome c, respectively. These few unique amino acid residues between the two species produced significantly different slopes in the calibration curves, which has implications for both absolute and relative quantification [31]. Recently, Kelleher's group demonstrated that hydrophobicity of a peptide can also influence tandem mass spectrometry quantification [33].

Hydrophobic modifications have been chemically added to a variety of biological substrates to improve the ESI response of a particular analyte. Hydrophobic tagging of nucleobases [34], nucleotides [34, 35], oligonucleotides [30], and PCR amplicons [30] have been reported to increase the ESI response for DNA analysis. Notably, the report by Null, *et al.* is the first to describe a hydrophobic tagging strategy for large (>500 Da) biomolecules [30]. Similarly, modifications have been added to amino acids for metabolite analysis [36-38].

In the last year, various reports of hydrophobic tagging for peptides have been published [39-41]. Foettinger, *et al.* derivatized peptides with malondialdehyde, which reacts with arginine under strong acidic conditions (12M HCl), and observed a 2.5-10-fold increase in ESI response [39]. However, side reactions with tryptophan potentially hindering identification and the strong acidic reaction conditions requiring extensive clean-up prior to

mass spectrometry and conceivably leading to peptide degradation limit the scope of this labeling reagent [39]. Furthermore, the tag would not introduce an adequate mass shift when stable-isotope labeled for quantitative proteomics measurements. Ullmer, *et al.* reported ~10-fold increase in ESI response when the free amine groups of peptides and glycopeptides were reacted with 6-aminoquinolyl-*N*-hydroxyscuuinimideyl carbamate; however, incomplete labeling and occasional loss of the tag during collision induced dissociation (CID) was observed [40]. Mirzaei and Regnier reacted peptides with [3-(2,5)-dioxopyrrolindin-1-yloxycarbonyl)-propyl]dimethyloctylammonium to afford 1- to 500-fold improvements (average 10-fold improvement) in ESI response [41]. However, this study used a tag that contained both a hydrophobic moiety and a permanent charge. Thus, it is difficult to ascertain the factor (*i.e.,* permanent charge and/or hydrophobicity) leading to the improved ESI response. Furthermore, the observed increases in ESI response were quite possibly the result of improved analyte retention on the reversed-phase LC column; therefore, the improved ESI response would be consequence of increased concentration of the hydrophobic-tagged peptides. This is supported by the observation that generally a greater increase in ESI response was detected for smaller peptides, especially for peptides less than 500 Da [39-41]. In an attempt to fundamentally understand the role of hydrophobicity on the ESI response of peptides, we synthesized and characterized 2-iodo-*N*-octylacetamide, the first in a series of new cysteine-specific peptide/protein tags.

6.2 Experimental

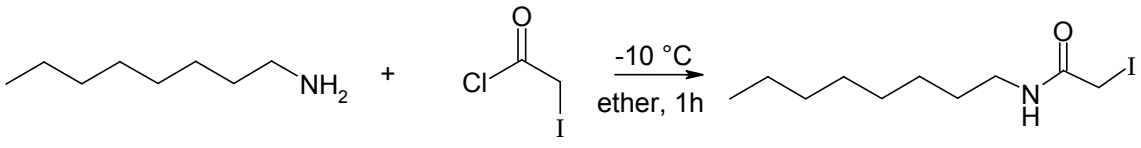
6.2.1 Reagents

Iodoacetyl chloride was purchased from Alfa Aesar (Ward Hill, MA). Anhydrous diethylether was obtained via the distillation of ether. Saturated sodium biocarbonate was made by adding sodium bicarbonate (Fisher Scientific, Inc, Pittsburgh, PA) to water until a saturated solution was obtained. Similarly, brine was made by adding sodium chloride

(Fisher Scientific, Inc, Pittsburgh, PA) to water until saturation. Celite was purchased from Acros Organics (Morris Plains, NJ). Sodium thiosulfate, anhydrous magnesium sulfate, b-type natriuretic peptide-32, formic acid, iodoacetamide, 1-octylamine were obtained from Sigma-Aldrich (St. Louis, MO). Laminin nonapeptide was purchased from EMD Biosciences (San Diego, CA). Tris•HCl buffer (1 M, pH 8.0), ethyl acetate, and ethanol were acquired from Fisher Scientific, Inc. (Pittsburgh, PA). Tris(2-carboxyethyl)phosphine was purchased from Pierce Biotechnology, Inc. (Rockford, IL). HPLC-grade Burdick and Jackson brand water, acetonitrile, isopropanol were purchased thru VWR International (West Chester, PA). All reagents were used as received unless otherwise noted.

6.2.2 2-Iodo-*N*-octylacetamide Synthesis

The reagent 2-iodo-*N*-octylacetamide was synthesized in the laboratory of Professor Daniel L. Comins (Department of Chemistry, North Carolina State University). A solution of iodoacetyl chloride (0.50 mL, 5.41 mmol, 1.0 eq) in anhydrous ether (15 mL) at -10 °C was treated with 1-octylamine (1.81 mL, 10.84 mmol, 2.0 eq); see **Scheme 6.1**. The mixture was stirred for one hour at -10 °C and filtered through a pad of Celite. The pad of Celite was then washed with cold ether. The combined extracts were washed with 10% sodium thiosulfate, water, saturated sodium bicarbonate and brine. Next, the washed extracts were dried over anhydrous magnesium sulfate and concentrated. Purification by radial PLC (silica gel, ethyl acetate/hexanes, 10:90, 25:75) yielded the desired product as a yellow solid.

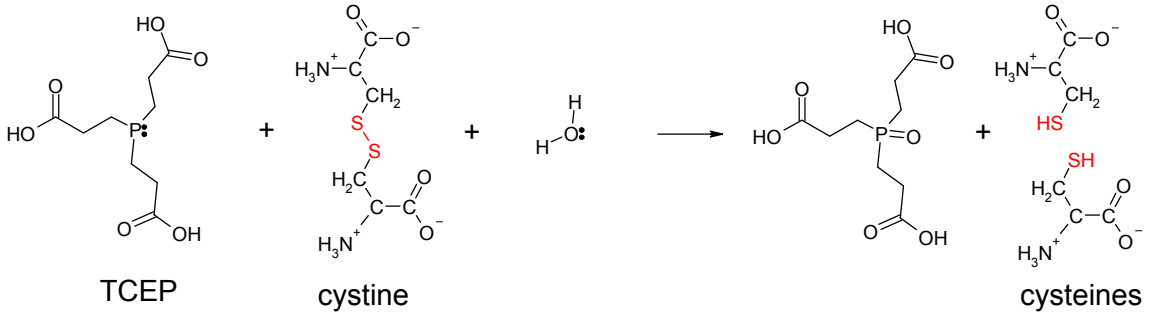


1-octylamine iodoacetyl chloride 2-iodo-*N*-octylacetamide

Scheme 6.1 2-Iodo-*N*-octylacetamide synthesis

6.2.3 Reduction of Cystines with Tris(2-carboxyethyl)phosphine

0.03 M tris(2-carboxyethyl)phosphine (TCEP) was added to a 0.5 mg/mL standard peptide solution to achieve a final concentration of 9 mM; TCEP and standard peptides were dissolved in 0.1M Tris buffer (pH 8.0). Tubes were then incubated for 20 minutes at 37 °C. TCEP reduces disulfide bonds according the reaction shown in **Scheme 6.2**. The large-scale synthesis of TCEP [42] and the manifold advantages of TCEP over other sulfhydral reductants, including the decreased oxidation rate over time (increased storage stability) [43, 44], activity over a wide pH range [42, 44], and rapid reduction [42, 44], make TCEP an ideal reagent for biochemical applications.



Scheme 6.2 Reduction of cystine using tris(2-carboxyethyl)phosphine

6.2.4 Alkylation with Iodoacetamide

Succeeding reduction with TCEP, 20 mM iodoacetamide (in 0.1 M Tris, pH 8.0) was added to the peptide solution (final iodoacetamide concentration 5 mM). The reaction proceeded for 1 hour at 37 °C in the dark.

6.2.5 Alkylation with 2-Iodo-*N*-octylacetamide

A 20 mM stock solution of 2-iodo-*N*-octylacetamide was prepared by first dissolving ~6 mg in 200 µL of ethanol. Then, we added 800 µL 0.1 M Tris to the solution. After adding the Tris buffer, some 2-iodo-*N*-octylacetamide did precipitate out of solution; therefore, we had to compensate by adding more 2-iodo-*N*-octylacetamide than iodoacetamide during the

peptide akylation step. In a separate reaction tube from that of the iodoacetamide, 20 mM 2-iodo-*N*-octylacetamide (in 0.1 M Tris, pH 8.0) was added to the reduced peptide solution to achieve a final concentration of 10 mM. The reaction proceeded for 1 hour at 37 °C in the dark.

6.2.6 Off-line Liquid Chromatography-Tandem Mass Spectrometry

An 8 µL aliquot of the reduced, alkylated peptide solution was separated on a 150 mm long by 2 mm i.d. column packed with Magic C18AQ, 5 µm, 200 Å silica (Michrom BioResources, Auburn, CA). Mobile phase A was water/acetonitrile/formic acid (98/2/0.1% by volume) and mobile phase B was acetonitrile/water/formic acid (90/10/0.1% by volume). A linear gradient was performed with 5% B initially and after four minutes increasing to 100% B at 60 minutes. Conditions were held at 100% B for five minutes and then initial conditions resumed. The column was re-equilibrated for 15 minutes prior to the next injection. The flow was maintained at 0.2 mL/min by the pumps (Shimadzu LC-20AD, Shimadzu, Columbia, MD). During the optimization of alkylation reactions, one minute fractions were collected and subsequently analyzed with an ESI-LTQ XL mass spectrometer (Thermo Fisher Scientific, Inc., Waltham, MA). Once optimal alkylation conditions were achieved, fractions containing the alkylated peptide were collected according to the UV absorbance at 205 nm. The integrated area under the peak was determined using the EZStart software (Shimadzu, Columbia, MD) and used to calculate the approximate concentration with the Scopes method [45]. Each solution was then diluted to approximately 1 µM with a 50:50 acetonitrile/water solution containing 0.1% formic acid. MS/MS spectra of each species were acquired with an LTQ-FT Ultra (Thermo Fisher Scientific, Inc., Waltham, MA) utilizing a previously reported nano-electrospray source [46]. The AGC limit was set to 1×10^6 and mass resolving power (at *m/z* 400) was set to $100{,}000_{FWHM}$. External calibration was implemented according to the manufacturer's protocol.

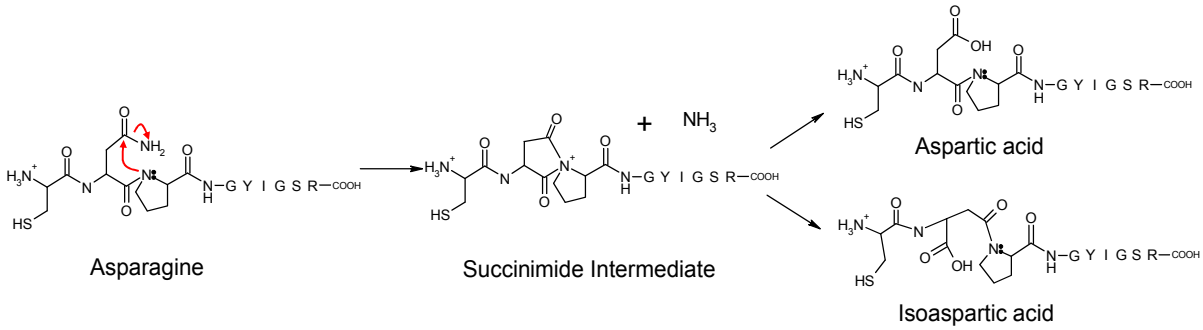
6.2.7 Liquid Chromatography-Mass Spectrometry of Laminin Nonapeptide

Prior to LC-MS, laminin nonapeptide alkylated with iodoacetamide and 2-iodo-N-octylacetamide were combined in a 1:1 volumetric ratio. Reversed-phase separation was achieved on a 150 mm long by 0.5 mm inner diameter Jupiter Proteo column packed with 4 μm, 90 Å silica (Phenomenex, Torrance, CA). Mobile phase A was water/isopropanol/acetonitrile/formic acid (98/1/1/0.2% by volume) and mobile phase B was acetonitrile/water/isopropanol/formic acid (80/10/10/0.2% by volume). Pumps (Shimadzu LC-20AD, Shimadzu, Columbia, MD) were set to a 20 μL/min flow rate and were initially maintained at 2% B for two minutes, followed by a linear gradient to 95% at 60 minutes. After 10 minutes at 95% B, the initial conditions were reinstated followed by a 10 minute reequilibration period. The autosampler (Shimadzu SIL-20AC, Shimadzu, Columbia, MD) was set to inject 10 μL of each sample. The needle was rinsed with water/isopropanol/acetonitrile/formic acid (50/25/25/0.2% by volume) between injections. The dwell time of this set-up is ~14.5 minutes computed from the volume of connecting tubing, autosampler dwell volume reported by the manufacturer (Shimadzu, Columbia, MD), and the column volume determined using length and diameter.

Mass spectra of eluted species were acquired in MS mode with an LTQ-FT Ultra (Thermo Fisher Scientific, Inc., Waltham, MA) utilizing the IonMax Source. The AGC limit was set to 1×10^6 and mass resolving power (at *m/z* 400) was set to $100{,}000_{FWHM}$. External calibration was implemented according to the manufacturer's protocol.

In all experiments we observed the deamidation of asparagine to aspartic acid in the laminin sequence. Although the large, bulky side chain of proline, the N+1 residue (the nucleophile for deamidation), does not favor deamidation, the post-translational modification has been observed in several proteins containing Asn-Pro dipeptides [47, 48]. Capasso developed a method based on the published deamidation rates of various proteins, to calculate the deamidation rate constant of any protein based on the amino acids preceding

and succeeding aspargine at 37°C, pH 7.4, in 0.0146 M $H_2PO_4^-$/0.062 M HPO_4^- buffer [49]. Using this method the first order rate constant for the deamidation of Cys-Asn-Pro (in the laminin sequence), is -7.22 s^{-1} (with rate constant of -5.07 s^{-1} for the fastest reaction and -7.96 s^{-1} for the slowest reaction) [49]. Obviously, the calculated rate constant depends on a number of factors (*e.g.,* buffer conditions, pH, temperature), however, it does provide an evaluation of the tripeptide sequence relative to other sequences. The reaction mechanism is shown in **Scheme 6.3**. Note, aspartic acid and isoaspartic acid are isobaric, and therefore, indistinguishable by the mass spectrometer.



Scheme 6.3 Asparagine deamidation to aspartic acid and isoaspartic acid

6.2.8 Liquid Chromatography-Mass Spectrometry of B-type Natriuretic Peptide-32

Prior to LC-MS, b-type natriuretic peptide-32 (BNP-32) alkylated with iodoacetamide and 2-iodo-N-octylacetamide were combined in a 1:1 volumetric ratio. Reversed-phase separation was achieved on a 150 mm × 0.5 mm Magic C18AQ column packed with 5 µm, 200 Å silica (Michrom BioResources, Auburn, CA). Mobile phase A was water/acetonitrile/formic acid (98/2/0.1% by volume) and mobile phase B was acetonitrile/water /formic acid (90/10/0.1% by volume). Pumps (Shimadzu LC-20AD, Shimadzu, Columbia, MD) were set to a 20 µL/min flow rate and were initially maintained at 2% B for two minutes, followed by a linear gradient to 95% at 60 minutes. After 10 minutes at 95% B, the initial conditions were reinstated followed by a 10 minute reequilibration

period. The autosampler (Shimadzu SIL-20AC, Shimadzu, Columbia, MD) was set to inject 10 µL of each sample. The needle was rinsed with water/isopropanol/acetonitrile/formic acid (50/25/25/0.2% by volume) between injections. The calculated dwell time of this set-up was the same as in the laminin nonapeptide experiments (~14.5 minutes). Mass spectrometry parameters were the same as for the laminin nonapeptide experiments.

6.2.9 Data Analysis

6.2.9.1 Ion Abundance Method to Determine the Relative Increase in Electrospray Ionization Response

ICIS peak detection algorithm, included in the Xcalibur™ software package (Thermo Fisher Scientific, Inc., Waltham, MA), was applied to the extracted ion chromatogram (XIC) of the entire isotopic distribution for each alkylated peptide. The mass spectrum corresponding to the apex in the XIC was used to determine the ion abundance. Ion abundances of the monoisotopic peak for each species were calculated from the ionization time and abundance [50]. The ionization time refers to the time required to accumulate a set number of ions in the ICR cell (mass analyzer), which is set by the automatic gain control feature. We empirically determined that the Xcalibur™ software normalizes the ion abundance to the ionization time, therefore to obtain an absolute ion abundance we must multiply the ion abundance found in Xcalibur by the ionization time [50]. Ionization times and the abundance of each *m/z* peak were obtained from the scan header and spectrum list for each spectrum, respectively. The ion abundance ratio was calculated as shown in **Equation 6.1**:

$$Ion\ Abundance\ Ratio = \frac{\left(\sum_{z=1}^{i}\left(abundance_{\mathrm{MIM}} * \mathrm{ionization\ time}\right)\right)_{octyl}}{\left(\sum_{z=1}^{i}\left(abundance_{\mathrm{MIM}} * \mathrm{ionization\ time}\right)\right)_{IAM}} \quad (6.1)$$

Ion abundances of the monoisotopic peak (MIM) were summed for each charge state in the distribution (from z= 1 to the i^{th} charge state).

6.2.9.2 Peak Area Method to Determine the Relative Increase in Electrospray Ionization Response

Extracted ion chromatograms (XIC) of the entire isotopic distribution for each charge state of each alkylated peptide were obtained. After applying the ICIS peak detection algorithm, included in the Xcalibur™ software package (Thermo Fisher Scientific, Inc., Waltham, MA), the integrated area under the peak for each XIC was displayed. The peak area ratio was calculated according to **Equation 6.2**:

$$Peak\ Area\ Ratio = \frac{\left(\sum_{z=1}^{i} peak\ area\right)_{octyl}}{\left(\sum_{z=1}^{i} peak\ area\right)_{IAM}} \qquad (6.2)$$

It is important to note that the peak area and the ion abundance used to calculate the relative improvement in ESI response using **Equations 6.1** and **6.2** are both measured by the mass analyzer. However, the peak area method (**Equation 6.2**) reveals the ESI response over the entire elution window, whereas the ion abundance method provides a "snapshot" of the ESI response at the eluted peak apex (**Equation 6.1**).

6.3 Results and Discussion

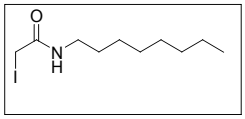
Peptides and proteins are generally reduced and alkylated in order to eliminate tertiary structure, thereby increasing accessibility of interior residues to proteolytic enzymes during the digestion step and available charges during the ionization process. In this study, the ubiquity of the alkylation step in peptide/protein analyses was exploited as a means to add a chemical moiety to cysteines that would enhance the electrospray ionization response

of modified peptides. In an initial evaluation of this strategy we synthesized 2-iodo-*N*-octylacetemide, which we subsequently reacted with two cysteine-containing peptide substrates; laminin nonapeptide prevents metastasis *in vitro* [51] and, therefore, is important for the study of tumor cell invasion and metastasis [52] and BNP-32, which is a widely-used (>10 million assays annually) cardiac marker for congestive heart failure [53].

The synthesis of 2-iodo-*N*-octylacetamide is a one step reaction with sufficient yield (86%, 1.4 g) to alkylate approximately 40 mg protein. NMR, FTIR, and high resolution mass spectrometry (**Figure 6.1**) analyses confirm the purity and generation of the desired product. Collectively, the availability of reagents, ease of synthesis and purification, and yield of 2-iodo-*N*-octylacetamide contribute to the desirability of the reagent.

Following synthesis, we reacted iodoacetamide (control) and 2-iodo-*N*-octylacetamide with two peptide substrates, laminin nonapeptide (MW 966.4 Da) and BNP-32 (MW 3463.7 Da). Alkylation conditions were optimized to reduce dialkylation and trialkylation products observed with higher iodoacetamide concentrations or with longer incubation times; observation of these products is consistent with other reports [54]. One modification per cysteine is achieved with the reaction conditions shown in **Figure 6.2**. Cysteine groups reacted with iodoacetamide result in the addition of a carboxyamidomethyl (CAM) modification; likewise, cysteines reacted with 2-iodo-*N*-octylacetamide have octylcarboxyamidomethyl (OCAM) modifications. The alkylation reactions appear to be complete based on the absence of unmodified substrate in fractions collected from off-line LC-UV followed by mass spectrometry analysis.

Compound: 2-Iodo-*N*-octylacetamide



Formula: $C_{10}H_{20}INO$
Molecular Weight: 297.1764
Yield(s): 86 %
Appearance: Yellow solid
Stability: stable
mp: 62-64 °C

HRMS(M+1)	Calculated: 298.0668 Da	Found: 298.0681 Da

IR (thin film, neat, NaCl): cm^{-1}.
722, 1158, 1294, 1316, 1565, 1638, 2849, 2919, 2953, 3088, 3265

NMR ($CDCl_3$):

(^1H, 400 MHz): δ 0.86-0.90 (t, 3H, *J* = 6.4 Hz), 1.28-1.31 (m, 10H), 1.49-1.54 (m, 2H), 3.24-3.29 (m, 2H), 3.70 (s, 2H), 6.10 (s, 1H).

(^{13}C, 100 MHz): δ 0.047, 14.29, 22.83, 27.00, 29.37, 29.41, 29.44, 31.97, 40.74, 166.73.

Figure 6.1 Characterization of 2-iodo-*N*-octylacetamide

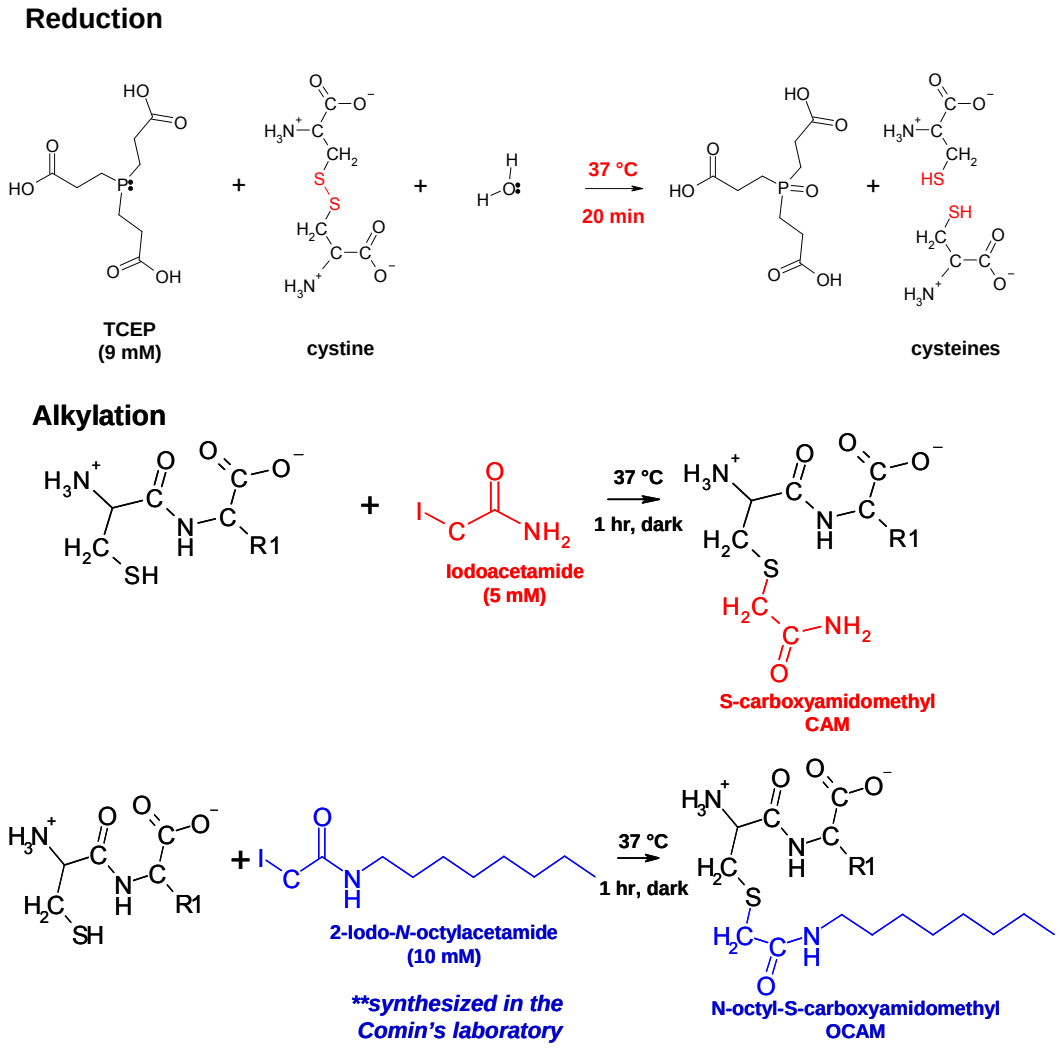


Figure 6.2 Reduction and alkylation with iodoacetamide and 2-iodo-*N*-octylacetamide

6.3.1 Relative Increase in Electrospray Ionization Response

Tandem mass spectrometry experiments were performed separately (*vide infra*) to maximize the MS duty cycle, thereby ensuring adequate profiling of the eluted peak. Relative improvements in the ESI response of the OCAM-modified peptides compared to CAM-modified peptides were assessed using two measures, 1) ratio of absolute ion abundances from mass spectra (**Equation 6.1**) and 2) ratio of peak area from XICs (**Equation 6.2**). The first measure corresponds to the increased ESI response with respect only to the mass analyzer. The second metric, the peak area ratio, describes both the

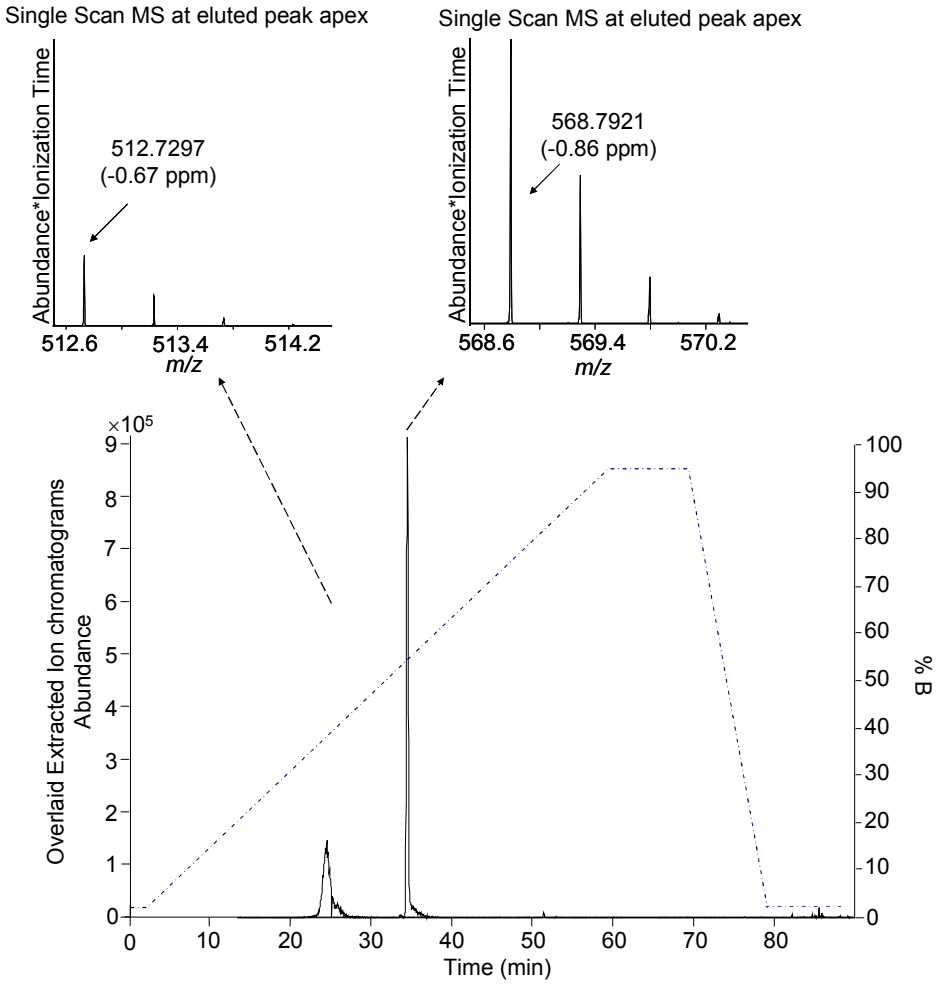
chromatographic and electrospray ionization consequence of adding the hydrophobic tag. Electrospray ionization is a concentration-dependent phenomenon, therefore increases in electrospray ionization response will be moderated by chromatographic peak broadening.

The addition of the OCAM-modification to the cysteine of laminin improved the chromatography and ESI response of laminin. **Figure 6.3** shows the extracted ion chromatograms for the 2^+ charge state of CAM-laminin and the OCAM-laminin, which are overlaid to demonstrate the difference in retention time resulting from increased alkyl chain length (hydrophobicity). Increased retention times of tagged peptides, compared to untagged peptides, have also been observed with the solid-phase ICAT strategy [11]. Accounting for the dwell time, CAM- and OCAM-laminin eluted in approximately 14% B and 30% B, correspondingly.

Noticeably, the peak width of the OCAM-laminin is much narrower. Increasing the hydrophobicity of the small peptide laminin with the OCAM-modification favors increased capture at the head of the column allowing sample concentration (focusing) prior to elution. The CAM-laminin interactions with the stationary phase at the head of the column are weak; therefore, the band diffuses and results in a broad peak [55]. Several authors have concluded that the larger ESI response of hydrophobic-tagged small peptides is a direct result of the additional hydrophobicity, which causes more analyte to reside on the surface [39-41]. However, our data suggests that the larger increase in ESI response may be a combination of both better chromatographic focusing (higher concentration) of more hydrophobic species as evidenced by the smaller peak width (**Figure 6.3**) (since ESI is a concentration-dependent process) and the higher percentage of analyte at the droplet surface.

In addition to the improved chromatographic performance of the OCAM-laminin, we also observed an increased ESI response of the modified peptide. Mass spectra of the 2^+ charge state for CAM- and OCAM-laminin from the eluted peak apex are inset in **Figure 6.3**. The calculated improvements in the ESI response are shown in the table (bottom of **Figure**

6.3). A 3.9-fold improvement is calculated using **Equation 6.1**, which is based solely on the differences in absolute ion abundance of each monoisotopic peak in the charge state distribution. Utilizing the peak areas from the XICs of the entire isotopic distribution of each charge state to calculate the improvement, the improvement in ESI response is calculated to be 2.8-fold (**Equation 6.2**). The variation in these calculations demonstrates the necessity



Equation 1	Equation 2
3.9	2.8

Figure 6.3 Relative increase in ESI response for laminin. Overlaid extracted ion chromatograms (XIC) for the 2+ charge state of CAM-laminin (*m/z* 512.72-514.74) and OCAM-laminin (*m/z* 568.79-570.80). Left y-axis indicates the absolute abundance of the XIC. The gradient plot is superimposed on the XICs, with the right y-axis indicating the percent B composition. Note the gradient plot does not account for dwell time, which is calculated as 14.5 minutes for this set-up. Additionally, the mass spectra of the 2+ charge state corresponding to the most abundant point in the XIC of the eluted peptide are inset. Mass spectra are plotted on the same y-scale. Values indicate the *m/z* values of the monoisotopic peaks and corresponding mass accuracies are shown in parenthesis. The relative improvement in the ESI response is shown in the table.

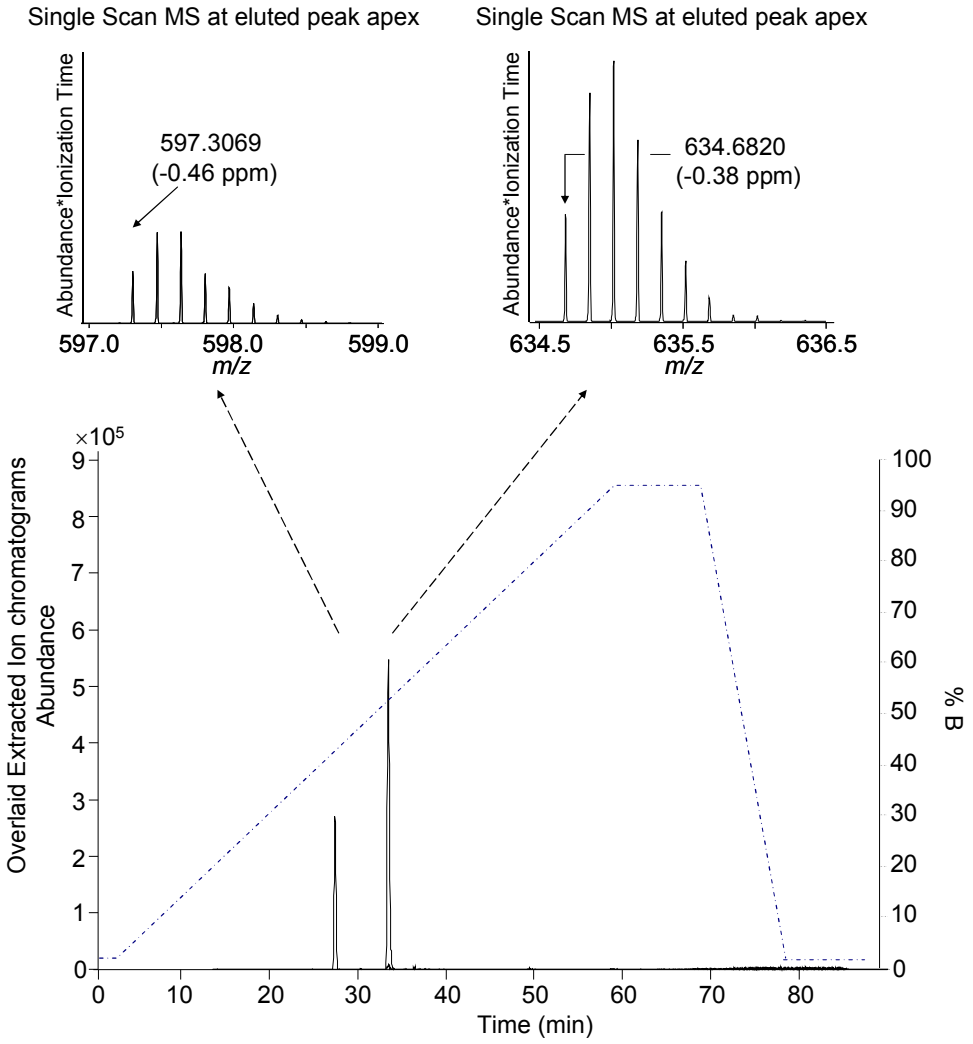
for careful consideration of reported improvements in ESI response.

When the hydrophobic tag was added to a larger peptide, BNP-32 (monoisotopic mass 3463.7 Da), we did not observe an improved chromatographic response of the OCAM-BNP-32 compared to the CAM-BNP-32. In fact, the peak width is slightly larger for the OCAM-BNP-32, which we attribute to band broadening observed with longer retention times [55]. The alkylated BNP-32 eluted in approximately 20% B for CAM-BNP-32 and 29% B for OCAM-BNP-32. Initially, CAM- and OCAM-BNP-32 LC-MS experiments were performed with the same conditions as the laminin experiments; however, significant noise, peak tailing, and overall poor peak shape warranted further optimization. Eliminating isopropanol from the mobile phases and utilizing a column containing silica with larger pores sizes improved the shape of eluted peaks; the results reported were obtained with these optimized conditions.

An increase in the ESI response of OCAM-BNP-32 in comparison to CAM-BNP-32 was observed. Mass spectra of the 6^+ charge state of CAM- and OCAM-BNP-32 derived from the maximum point of the eluted peak in the XIC are inset in **Figure 6.4**. The calculated improvement in the ESI response, determined from **Equations 6.1** and **6.2**, is 2.1 (both methods resulted in the same fold-improvement). These two calculations resulted in similar improvements because CAM- and OCAM-laminin have similar elution volumes, therefore the sum of peak areas is directly proportional to the ion abundance at the peak apex. In the case of laminin the peak area of CAM-laminin was larger than the OCAM-laminin peak area, so the improved ESI response calculated using **Equation 6.2** was less than the ESI response at the peak apex (determined using **Equation 6.1**).

The two peptides we investigated have similar GRAVY scores (-0.611 and -0.506 for laminin and BNP-32, respectively). Since hydrophobicity is a dominant determinant of ESI response, the addition of the OCAM-modification would have a similar effect on the ESI response of both peptides. In fact, there is less than a one-fold difference between the two

peptides considering the improvement in ESI response calculated with **Equation 6.2**. Although, admittedly, we are hesitant to draw too many conclusions from the limited dataset. To this end, we are currently examining the role of the OCAM- hydrophobic modification, as well as other hydrophobic modifications, on a vast array of substrates varying in mass, hydrophobicity, and number of cysteines (which would result in various numbers of added hydrophobic modifications).



Equation 1	Equation 2
2.1	2.1

Figure 6.4 Relative increase in ESI response for BNP-32. Notation is the same as in **Figure 6.3**, except that the overlaid XIC are for the 6+ charge state of CAM-BNP-32 (*m/z* 597.30-598.65) and OCAM-BNP-32 (*m/z* 634.67-636.02). Calculated dwell time of this set up is 14.5 minutes. Mass spectra of the 6+ charge state are inset. XICs and mass spectra are plotted on the same y-scale as in Figure 2.

Importantly, the data presented suggest that improvements in ESI responses with the addition of a hydrophobic tag are not limited to small peptides. Specifically for BNP-32, the ability to enhance the ESI response is especially promising considering the clinical relevance of the molecule [56]. A two-fold improvement in the ESI response would be expected to correlate to a two-fold reduction in the limit-of-detection. In our own work, this could potentiate the discovery of novel BNP forms related to congestive heart failure [53, 56].

6.3.2 Tandem Mass Spectrometry of Tagged Peptides

Tandem mass spectrometry is an important step for identification of species present in a sample; therefore, the second part of our evaluation of the OCAM hydrophobic modification was to determine the behavior of the hydrophobic tag during collision-induced dissociation. After the reduction and alkylation steps, we purified the alkylated peptide from the reaction mixture using off-line LC-UV for tandem mass spectrometry experiments. **Figure 6.5 A** and **B** are the MS/MS spectra of CAM- and OCAM-laminin, respectively, which have the same y-scale. The y_7 product ion is the dominant fragmentation pathway of both the CAM- and OCAM-laminin. Cleavage N-terminal to proline is known to occur preferentially [50, 57]. Notably, the OCAM hydrophobic tag does not fragment or undergo neutral loss, which would complicate interpretation of MS/MS spectra. The average mass accuracy of the product ions for CAM- and OCAM-laminin is -0.83 ± 0.46 ppm and -0.14 ± 0.72 ppm (confidence interval of the mean at the 95% confidence level), respectively.

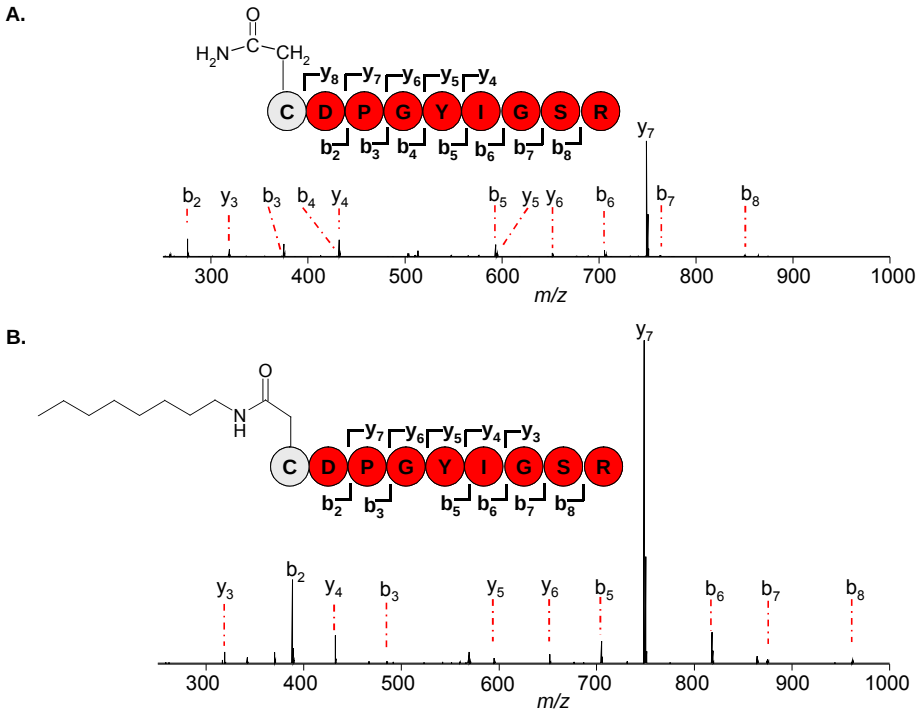


Figure 6.5 MS/MS spectra resulting from the isolation and collision-induced dissociation of the 2+ charge state of CAM-laminin (**A**) and OCAM-laminin (**B**). The average mass accuracy of the product ions is -0.83 ± 0.46 pmm and -0.14 ± 0.72 ppm (confidence interval of the mean at the 95% confidence level) for CAM-laminin and OCAM-laminin product ions, respectively. The b- and y-ions observed are labeled in the mass spectra, as well as indicated on the laminin sequence. Asterisks indicate loss of water. Both mass spectra in (**A**) and (**B**) are scaled the same.

The tandem mass spectra for the CAM- and OCAM-BNP-32 are shown in **Figure 6.6 A** and **B**. In addition to the increased ion abundance, more sequence ions were observed during the dissociation of OCAM-BNP-32 compared to CAM-BNP-32. The average mass measurement accuracy of the product ions is -0.18 ± 0.79 ppm and -0.52 ± 0.46 ppm (confidence interval of the mean at the 95% confidence level). The improved ESI response of the OCAM-modified peptides equates to better sequence coverage, which would facilitate peptide sequencing especially in cases when unexpected isoforms are present.

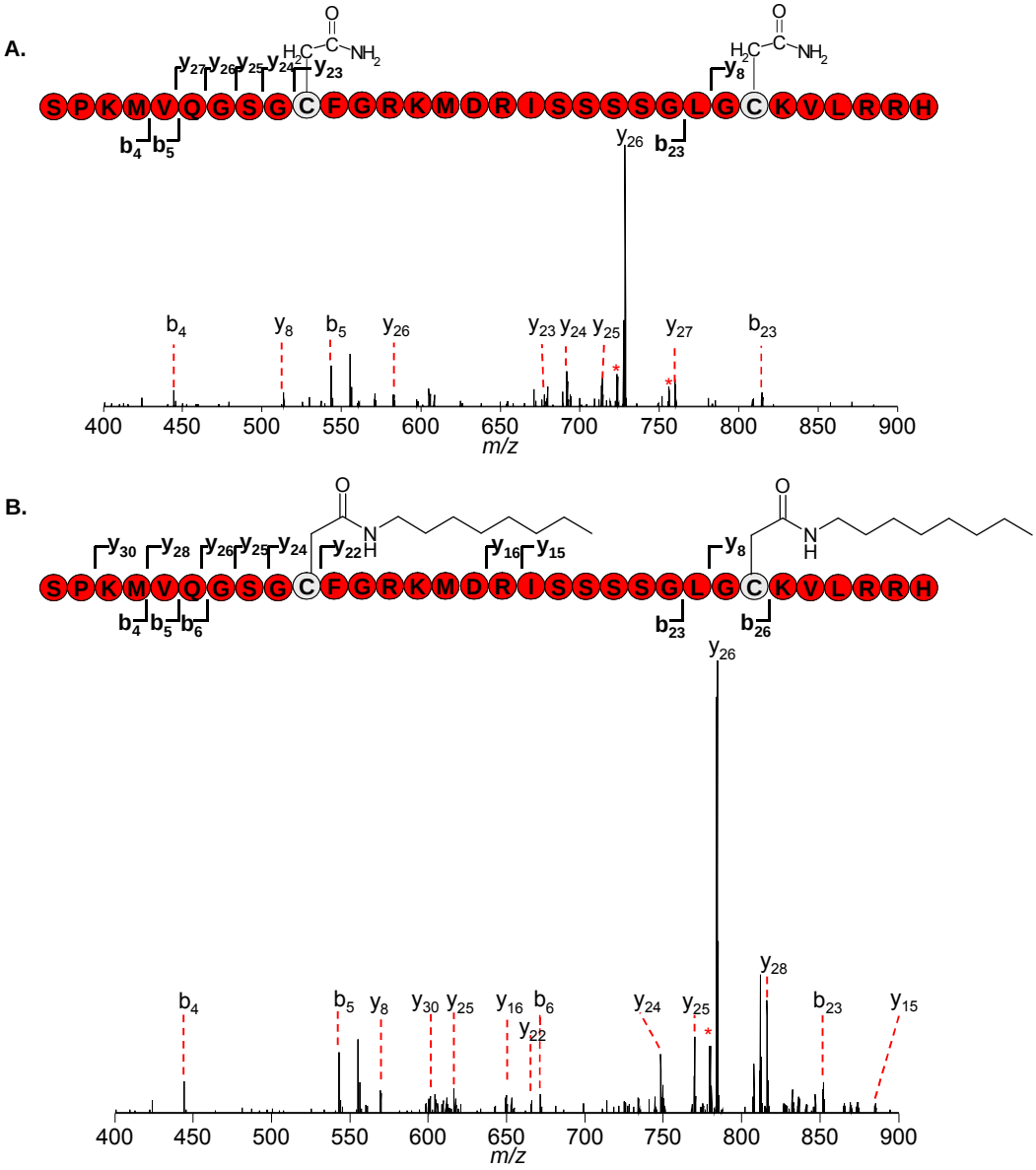


Figure 6.6 MS/MS spectra resulting from the isolation and collision-induced dissociation of the 6+ charge state of CAM-BNP-32 (**A**) and OCAM-BNP-32 (**B**). The average mass accuracy of the product ions is -0.18 ± 0.79 ppm and -0.52 ± 0.46 ppm (confidence interval of the mean at the 95% confidence level) for CAM-BNP-32 and OCAM-BNP-32 product ions, respectively. The b- and y-ions observed are labeled in the mass spectra, as well as indicated on the BNP-32 sequence. Asterisks indicate loss of water. Both mass spectra in (**A**) and (**B**) are scaled the same.

6.4 Conclusions and Future Experiments

A hydrophobic tag, namely an octylcarboxyamidomethyl modification, can be added to cysteine residues via simple alkylation chemistry. We observed a three-fold and two-fold improvement in the ESI responses of two peptides, laminin nonapeptide (MW 966.4 Da) and

b-type natriuretic peptide-32 (MW 3463.7 Da), respectively, in these proof-of-concept experiments. Importantly, the octylcarboxyamidomethyl tag does not undergo collision-induced dissociation or neutral loss during tandem mass spectrometry, thus facilitating peptide sequencing. Potential applications of this hydrophobic tagging strategy are innumerable, as the strategy could be applied to improve the limit of detection of any cysteine-containing peptide. Perhaps, the area of greatest benefit would be targeted (*i.e.,* affinity isolated) proteomics studies because the limit of detection for a known substrate can be decreased, consequently expanding the realm of accessible clinically-relevant markers to lower abundance peptides (*e.g.,* BNP-32, troponin). We are currently working to broaden our study of hydrophobic tags to include a repertoire of peptide substrates with varying molecular mass, hydrophobicities (GRAVY scores), and number of cysteines (number of tags). Furthermore, we are characterizing the behavior of various hydrophobic tags with respect to the improvement in ESI response through development of a diverse library of hydrophobic tags, all based on alkylation chemistry. Concurrently, the development of the hydrophobic tags for solid-phase alkylation, analogous to solid-phase ICAT [11], is being explored.

6.5 References

1. Sharkey, AG, Friedel, RA, and Langer, SH. Mass spectra of trimethylsilyl derivatives. *Analytical Chemistry,* 1957. **29**: 770-776.

2. Sweeley, CC, Bentley, R, Makita, M, and Wells, WW. Gas-liquid chromatography of trimethylsilyl derivatives of sugars and related substances. *Journal of the American Chemical Society,* 1963. **85**: 2497-2507.

3. Quirke, JME, Adams, CL, and Van Berkel, GJ. Chemical derivatization for electrospray-ionization mass-spectrometry .1. Alkyl-halides, alcohols, phenols, thiols, and amines. *Analytical Chemistry,* 1994. **66**: 1302-1315.

4. Van Berkel, GJ and Asano, KG. Chemical derivatization for electrospray-ionization mass-spectrometry. 2. Aromatic and highly conjugated molecules. *Analytical Chemistry,* 1994. **66**: 2096-2102.

5. Van Berkel, GJ, Quirke, JME, Tigani, RA, Dilley, AS, and Covey, TR. Derivatization for electrospray ionization mass spectrometry. 3. Electrochemically ionizable derivatives. *Analytical Chemistry,* 1998. **70**: 1544-1554.

6. Kaufmann, R, Spengler, B, and Lutzenkirchen, F. Mass-spectrometric sequencing of linear peptides by product-ion analysis in a reflectron time-of-flight mass-spectrometer using matrix-assisted laser-desorption ionization. *Rapid Communications in Mass Spectrometry,* 1993. **7**: 902-910.

7. Keough, T, Youngquist, RS, and Lacey, MP. A method for high-sensitivity peptide sequencing using postsource decay matrix-assisted laser desorption ionization mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America,* 1999. **96**: 7131-7136.

8. Sechi, S and Chait, BT. Modification of cysteine residues by alkylation. A tool in peptide mapping and protein identification. *Analytical Chemistry,* 1998. **70**: 5150-5158.

9. Goodlett, DR, Bruce, JE, Anderson, GA, Rist, B, Pasa-Tolic, L, Fiehn, O, et al. Protein identification with a single accurate mass of a cysteine-containing peptide and constrained database searching. *Analytical Chemistry,* 2000. **72**: 1112-1118.

10. Gygi, SP, Rist, B, Gerber, SA, Turecek, F, Gelb, MH, and Aebersold, R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnology,* 1999. **17**: 994-999.

11. Zhou, HL, Ranish, JA, Watts, JD, and Aebersold, R. Quantitative proteome analysis by solid-phase isotope tagging and mass spectrometry. *Nature Biotechnology,* 2002. **20**: 512-515.

12. Li, J, Steen, H, and Gygi, SP. Protein profiling with cleavable isotope-coded affinity tag (cICAT) reagents: The yeast salinity stress response. *Molecular & Cellular Proteomics,* 2003. **2**: 1198-1204.

13. Hall, MP and Schneider, LV. Isotope-differentiated binding energy shift tags (IDBEST(tm)) for improved targeted biomarker discovery and validation. *Expert Review of Proteomics,* 2004. **1**: 421-431.

14. Geng, MH, Ji, JY, and Regnier, FE. Signature-peptide approach to detecting proteins in complex mixtures. *Journal of Chromatography A,* 2000. **870**: 295-313.

15. Chakraborty, A and Regnier, FE. Global internal standard technology for comparative proteomics. *Journal of Chromatography A,* 2002. **949**: 173-184.

16. Ross, PL, Huang, YN, Marchese, JN, Williamson, B, Parker, K, Hattan, S, et al. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Molecular & Cellular Proteomics,* 2004. **3**: 1154-1169.

17. Cagney, G and Emili, A. De novo peptide sequencing and quantitative profiling of complex protein mixtures using mass-coded abundance tagging. *Nature Biotechnology,* 2002. **20**: 163-170.

18. Goshe, MB, Conrads, TP, Panisko, EA, Angell, NH, Veenstra, TD, and Smith, RD. Phosphoprotein isotope-coded affinity tag approach for isolating and quantitating phosphopeptides in proteome-wide analyses. *Analytical Chemistry,* 2001. **73**: 2578-2586.

19. Julka, S and Regnier, F. Quantification in proteomics through stable isotope coding: A review. *Journal of Proteome Research,* 2004. **3**: 350-363.

20. Leitner, A and Lindner, W. Current chemical tagging strategies for proteome analysis by mass spectrometry. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences,* 2004. **813**: 1-26.

21. Ong, SE and Mann, M. Mass spectrometry-based proteomics turns quantitative. *Nature Chemical Biology,* 2005. **1**: 252-262.

22. Kenyon, GL, DeMarini, DM, Fuchs, E, Galas, DJ, Kirsch, JF, Leyh, TS, et al. Defining the mandate of proteomics in the post-genomics era: Workshop report. *Molecular & Cellular Proteomics,* 2002. **1**: 763-780.

23. Aebersold, R and Mann, M. Mass spectrometry-based proteomics. *Nature,* 2003. **422**: 198-207.

24. Mann, M, Hendrickson, RC, and Pandey, A. Analysis of proteins and proteomes by mass spectrometry. *Annual Review of Biochemistry,* 2001. **70**: 437-473.

25. Herbert, B, Galvani, M, Hamdan, M, Olivieri, E, MacCarthy, J, Pedersen, S, et al. Reduction and alkylation of proteins in preparation of two-dimensional map analysis: Why, when, and how? *Electrophoresis,* 2001. **22**: 2046-2057.

26. Hamdan, M and Righetti, PG. Modern strategies for protein quantification in proteome analysis: Advantages and limitations. *Mass Spectrometry Reviews,* 2002. **21**: 287-302.

27. Cech, NB and Enke, CG. Practical implications of some recent studies in electrospray ionization fundamentals. *Mass Spectrometry Reviews,* 2001. **20**: 362-387.

28. Fenn, JB. Ion formation from charged droplets - roles of geometry, energy, and time. *Journal of the American Society for Mass Spectrometry,* 1993. **4**: 524-535.

29. Cech, NB and Enke, CG. Relating electrospray ionization response to nonpolar character of small peptides. *Analytical Chemistry,* 2000. **72**: 2717-2723.

30. Null, AP, Nepomuceno, AI, and Muddiman, DC. Implications of hydrophobicity and free energy of solvation for characterization of nucleic acids by electrospray ionization mass spectrometry. *Analytical Chemistry,* 2003. **75**: 1331-1339.

31. Gordon, EF, Mansoori, BA, Carroll, CF, and Muddiman, DC. 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. *Journal of Mass Spectrometry,* 1999. **34**: 1055-1062.

32. Cech, NB and Enke, CG. Effect of affinity for droplet surfaces on the fraction of analyte molecules charged during electrospray droplet fission. *Analytical Chemistry,* 2001. **73**: 4632-4639.

33. Pesavento, JJ, Mizzen, CA, and Kelleher, NL. Quantitative analysis of modified proteins and their positional isomers by tandem mass spectrometry: Human histone H4. *Analytical Chemistry,* 2006. **78**: 4271-4280.

34. Nordstrom, A, Tarkowski, P, Tarkowska, D, Dolezal, K, Astot, C, Sandberg, G, et al. Derivatization for LC electrospray ionization-MS: A tool for improving reversed-phase separation and ESI responses of bases, ribosides, and intact nucleotides. *Analytical Chemistry,* 2004. **76**: 2869-2877.

35. Flarakos, J, Xiong, W, Glick, J, and Vouros, P. A deoxynucleotide derivatization methodology for improving LC-ESI-MS detection. *Analytical Chemistry,* 2005. **77**: 2373-2380.

36. Shortreed, MR, Lamos, SM, Frey, BL, Phillips, MF, Patel, M, Belshaw, PJ, et al. Ionizable isotopic labeling reagent for relative quantification of amine metabolites by mass spectrometry. *Analytical Chemistry,* 2006. **78**: 6398-6403.

37. Yang, WC, Mirzaei, H, Liu, XP, and Regnier, FE. Enhancement of amino acid detection and quantification by electrospray ionization mass spectrometry. *Analytical Chemistry,* 2006. **78**: 4702-4708.

38. Chen, Y, Zhang, JC, Chen, J, Cao, XY, Wang, J, and Zhao, YF. Sensitivity improvement of amino acids by n-terminal diisopropyloxyphosphorylation in electrospray ionization mass spectrometry. *Rapid Communications in Mass Spectrometry,* 2004. **18**: 469-473.

39. Foettinger, A, Leitner, A, and Lindner, W. Derivatisation of arginine residues with malondialdehyde for the analysis of peptides and protein digests by LC-ESI-MSMS. *Journal of Mass Spectrometry,* 2006. **41**: 623-632.

40. Ullmer, R, Plematl, A, and Rizzi, A. Derivatization by 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate for enhancing the ionization yield of small peptides and glycopeptides in matrix-assisted laser desorption/ionization and electrospray ionization mass spectrometry. *Rapid Communications in Mass Spectrometry,* 2006. **20**: 1469-1479.

41. Mirzaei, H and Regnier, F. Enhancing electrospray ionization efficiency of peptides by derivatization. *Analytical Chemistry,* 2006. **78**: 4175-4183.

42. Burns, JA, Butler, JC, Moran, J, and Whitesides, GM. Selective reduction of disulfides by tris(2-carboxyethyl)phosphine. *Journal of Organic Chemistry,* 1991. **56**: 2648-2650.

43. Getz, EB, Xiao, M, Chakrabarty, T, Cooke, R, and Selvin, PR. A comparison between the sulfhydryl reductants tris(2-carboxyethyl)phosphine and dithiothreitol for use in protein biochemistry. *Analytical Biochemistry,* 1999. **273**: 73-80.

44. Han, JC and Han, GY. A procedure for quantitative-determination of tris(2-carboxyethyl)phosphine, an odorless reducing agent more stable and effective than dithiothreitol. *Analytical Biochemistry,* 1994. **220**: 5-10.

45. Scopes, RK. Measurement of protein by spectrophotometry at 205-nm. *Analytical Biochemistry,* 1974. **59**: 277-282.

46. Nepomuceno, AI, Muddiman, DC, Bergen, HR, Craighead, JR, Burke, MJ, Caskey, PE, et al. Dual electrospray ionization source for confident generation of accurate mass tags using liquid chromatography Fourier transform ion cyclotron resonance mass spectrometry. *Analytical Chemistry,* 2003. **75**: 3411-3418.

47. Robinson, NE and Robinson, AB. Prediction of protein deamidation rates from primary and three-dimensional structure. *Proceedings of the National Academy of Sciences of the United States of America,* 2001. **98**: 4367-4372.

48. Tarelli, E and Corran, PH. Ammonia cleaves polypeptides at asparagine proline bonds. *Journal of Peptide Research,* 2003. **62**: 245-251.

49. Capasso, S. Estimation of the deamidation rate of asparagine side chains. *Journal of Peptide Research,* 2000. **55**: 224-229.

50. Williams, DK and Muddiman, DC. 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. *Analytical Chemistry,* submitted.

51. Iwamoto, Y, Robey, FA, Graf, J, Sasaki, M, Kleinman, HK, Yamada, Y, et al. YIGSR, a synthetic laminin pentapeptide, inhibits experimental metastasis formation. *Science,* 1987. **238**: 1132-1134.

52. Stetler-Stevenson, WG, Aznavoorian, S, and Liotta, LA. Tumor-cell interactions with the extracellular-matrix during invasion and metastasis*. Annual Review of Cell Biology,* 1993. **9**: 541-573.

53. Anderson, L. Candidate-based proteomics in the search for biomarkers of cardiovascular disease*. Journal of Physiology-London,* 2005. **563**: 23-60.

54. Yang, ZH and Attygalle, AB. LC/MS characterization of undesired products formed during iodoacetamide derivatization of sulfhydryl groups of peptides*. Journal of Mass Spectrometry,* 2007. **42**: 233-243.

55. Snyder, LR, Kirkland, JJ, and Glajch, JL. *Practical HPLC method development*. 2nd ed. 1997, New York: John Wiley & Sons. xxvi, 765 p.

56. Hawkridge, AM, Heublein, DM, Bergen, HR, Cataliotti, A, Burnett, JC, and Muddiman, DC. Quantitative mass spectral evidence for the absence of circulating brain natriuretic peptide (BNP-32) in severe human heart failure*. Proceedings of the National Academy of Sciences of the United States of America,* 2005. **102**: 17442-17447.

57. Schwartz, BL and Bursey, MM. Some proline substituent effects in the tandem mass-spectrum of protonated pentaalanine*. Biological Mass Spectrometry,* 1992. **21**: 92-96.

APPENDIX

APPENDIX A

GLOSSARY: Definition of terms and abbreviations in the context of this proposal

0 D	zero-dimension separation or direct infusion
1 D	one-dimension separation; single-dimension separation
2 D	two-dimension separation; separation in one dimension is followed by separation in an orthogonal dimension. **Example SCX-RP-HPLC:** Separation in the first dimension is based on a protein's isoelectric point (pH at which the protein has no net charge) and separation in the second dimension is based on *hydrophobicity*.
absolute quantification	ion abundances of proteins/peptides are compared to the ion abundance of a stable-isotope labeled internal standard added at a known concentration; concentration can be expressed as copy number or mg/ml, for example
abundance-matched normalization	abundances of an endogenous peptide in a LC-MS run is matched to the internal standard in the same run with the closest abundance. The abundance of the chosen internal standard in that run is then divided by the abundance of the same internal standard in the first run to determine the *normalization* factor. The abundance of the endogenous peptide is multiplied by the

	normalization factor. This is repeated for all endogenous peptides in every LC-MS run.
biomarker	a protein or peptide indicative of disease
bottom-up proteomics	protein samples are digested with a proteolytic enzyme and protein identification is determined from the peptide masses and peptide ion masses from tandem MS experiments
calibration law	adjusts for frequency shifts in FT-ICR
CAM	S-carboxyamidomethyl modification on cysteine residues resulting from alkylation with iodoacetamide
capacity factor	describes the migration rate of an analyte through the column; calculated as the difference between the retention time of the analyte and unretained species divided by the retention time of unretained species
charge capacity	maximum number of ions that can be contained in an ICR cell before *peak coalescence*, large frequency shifts, and rapid loss of cloud coherence are observed
chirp excitation	systematic sinusoidal frequency sweep
cyclotron frequency	the number of time a charged-particle moving in a magnetic field passes the detector in a given time; directly correlated to *m/z*
electrospray response	ion abundance; ability of the ion to escape the droplet
ESI	electrospray ionization
exact mass	*monoisotopic mass*
forbidden zone	*m/z* region that is not occupied by peptides or proteins

free energy of solvation	energy required to transfer the ion from the solution to the gas phase
FT-ICR	Fourier transform-ion cyclotron resonance
global normalization	Average abundance of the internal standards in subsequent run is divided by the average abundances of the same internal standards in the first run to determine the *normalization* factor. Then, the abundance all endogenous peptides are multiplied by the *normalization* factor. This is repeated for every LC-MS run.
global proteomics	investigation of all proteins in a give system or resulting from a particular genome
heated electrospray interface	programmable modification used to heated the capillary transfer line
hydrophobic tagging	addition of an alkyl modification to increase ion abundance
hydrophobicity	revealed by a protein or peptide GRAVY (grand average of hydropathy) score, which determined from each amino acid value ranging from -4.6 to 4.6. More positive GRAVY scores indicate more hydrophobic proteins (such as membrane proteins).
iodopeptides	a set of peptide internal standards with 3,5-diiodotyrosine residues incorporated to shift the *mass excess* of the internal standards into *forbidden zone*s
ion suppression	changes in droplet formation or evaporation, and therefore the resulting ions during electrospray

	ionization caused by the presence of certain co-eluting compounds
label-free proteomics	*relative quantification* of proteins and peptides based on absolute ion abundance; no stable-isotope label is used
Lorentz force	force a charged particle experiences in an electromagnetic field resulting from the electric field force and magnetic field force
mass defect	difference between the *exact mass* and the mass of the constituent nucleons
mass excess	difference between the *exact mass* and the *nominal mass*
monoisotopic mass	sum of the masses of the most-abundant isotope occurring in nature for each element in the chemical formula
multi-dimensional separation	two or more orthogonal separation steps performed in tandem
nominal mass	*monoisotopic mass* rounded to the nearest integer value; sum of mass numbers of all elements in the molecule
normalization	elimination of statistical error in the abundance for repeated measurements
normalization factor	calculated by dividing the ion abundance of the internal standard in subsequent runs by the ion abundance of the internal standard in subsequent run; accounts for run to run variability in ion abundance of the internal standard

OCAM	*N*-octyl-S-carboxyamidomethyl modification on cysteine residues resulting from alkylation with 2-iodo-*N*-octyl-acetamide
peak coalescence	peaks for ions of similar *m/z* values combine
post-excite radius	orbit radius of the ion cloud after excitation
relative quantification	ion abundances of the analyte are compared to ion abundances of a stable-isotope labeled internal standard; concentration is often expressed as fold-change
retention time reproducibility	measure of the variability in retention time of an analyte
RP-HPLC	reverse-phase high performance liquid chromatography
Scopes method	method to determine the concentration of peptides and proteins; the concentration (mg/m) is absorbance measured at 205 nm, which is due to the peptide bonds, divided by 31 multiplied by the path length.
signal-damping constant (τ)	time it takes the signal to decrease to $1/e^1$ the initial amplitude
space-charge effects	Coulombic interactions between ions in the same ion cloud
SWIFT excitation	excitation waveform created by defining a mass domain profile, converting to the frequency domain, then inverse-Fourier transforming to obtain the time-domain profile
targeted proteomics	investigation of a single or small number of defined protein(s) or peptide(s)
TOF	time-of-flight

top-down proteomics	protein identifications are made from accurate mass measurements of the intact protein and the fragment ions resulting from gas-phase dissociation