DEVELOPMENT OF METHODS FOR MASS SPECTROMETRY ANALYSIS OF PROTEINS IN SUPERNATANTS HARVESTED FROM *IN VITRO* HUMAN LUNG CELLS (A549)

by

Edward Gee-Ming Lau
B.Sc., Simon Fraser University, 2006

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

In the
Department of
Chemistry

SIMON FRASER UNIVERSITY

Fall 2010

APPROVAL

Name: **Edward Gee-Ming Lau**

Degree: **Master of Science**

Title of Thesis: **Development of Methods for Mass Spectrometry Analysis of Proteins in Supernatants Harvested from *in vitro* Human Lung Cells (A549)**

Examining Committee:

Chair: **Dr. Robert Britton**
Assistant Professor

Dr. George Agnes
Senior Supervisor
Associate Dean of Graduate Studies
Professor – Department of Chemistry

Dr. Paul Li
Supervisor
Professor – Department of Chemistry

Dr. Charles Walsby
Supervisor
Associate Professor – Department of Chemistry

Dr. Luis Sojo
External Examiner
Adjunct Professor – Department of Chemistry

Date Defended/Approved: December 9, 2010

Declaration of Partial Copyright Licence

ABSTRACT

Proteins are primarily responsible for functionalizing cells and, with respect to an initiative in soft ionization mass spectrometry (MS), the relative abundances of selected proteins have been used as indicators of normal or stressed physiological states. Methods developed for this thesis describe the use of matrix assisted laser desorption ionization time of flight (MALDI-ToF) MS to monitor the ion signals of C-X-C motif chemokine 5 and ubiquitin protein. Identification of these ion signals were carried out by performing additional experiments using liquid chromatography interfaced to an electrospray ionization (LC-ESI) equipped MS capable of tandem MS. A total of 78 proteins were identified by LC-ESI tandem MS, but not all corresponded to MALDI-MS data. In addition to the use of commercially available instrumentation, a separate study was performed to investigate the potential for an AC trap to generate molecular ions from a single levitated droplet having undergone Coulomb explosion.

Keywords: Mass Spectrometry; Matrix Assisted Laser Desorption Ionization Time of Flight; Liquid Chromatography Electrospray Ionization; Tandem MS; Protein Identification; Single Droplet Levitation

DEDICATION

For my Mommy and Grandma

ACKNOWLEDGEMENTS

I am grateful beyond adequate acknowledgement to my supervisor George Agnes for his mentorship and guidance from my latter undergraduate semesters through my graduate studies. He provided me with a chance to advance in chemistry by showing me areas of research previously unknown to me. I remember him telling me just before I started my MSc., “You remind me of myself back when I was in school, just doing alright in school until you find an area you want to excel in.” I consider myself extremely fortunate to have been a student under his supervision, conducting research in an environment that allowed me to learn and investigate ideas with a tremendous amount of freedom.

I would like to thank my fellow lab members over the years who have helped me learn and provide insightful comments related to my research. Particularly, I thank Allen Haddrell (teaching me how to operate a MALDI-ToF and culture cells), Xinzhou (Siry) Hu (expanding my cell culturing abilities), Neil Draper (constant helpful insight), Mariana (helpful insight in to molecular biochemistry explained to a chemist), and of course the many others who helped make the Agnes Lab a comfortable second home.

I am thankful for the support of the British Columbia Proteomic Network. Through this organization I could access a licensed version of Mascot software, and make use of Thomas Clark’s (MS service specialist) expertise. Thanks to

Tom, he taught me how to assemble in-house an online nano-flow LC-ESI ion source, and prepared capillary reverse phase columns for me to use.

Lastly I have to thank my family, particularly my mom and grandmother for their strong belief in education, and their sacrifices for me to obtain one. I respect them more than anyone for their courage, conviction, and hard work for immigrating from their home in China, to Hong Kong, and ultimately Canada. I also thank my big sister and the rest of my family for a combination of their tolerance and enlightenment. Thank you for tolerating and entertaining what I can now imagine were endless annoying queries of “Why’s and How’s”, at the same time enlightening me with “Did you know’s and Want to see”.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

GLOSSARY

AC	alternating current
ACN	acetonitrile
ATC	Adrenocorticotropic
CAD	collision activation dissociation
d	diameter
DC	direct current
DTT	dithiothreitol
E_0	fragmentation threshold energy
ECD	electron capture dissociation
EDLT	electrodynamic levitation trap
EDTA	ethylenediamine-tetraacetic acid
EI	electron impact
ELISA	enzyme linked immunosorbent assay
ESI	electrospray ionization
e.g.	for example
f	frequency
FP	false positive

FWHM	full width at half max
GRAVY	grand average hydrophobicity
Hz	Hertz
IDA	independent data acquisition
IRMPD	infrared multi-photon dissociation
i.d.	inner diameter
i.e.	such as
ICAM	intercellular adhesion molecule
IQ	inter-quadrupole
kV	kilovolt
l	length
L	litre
LC	liquid chromatography
LIT	linear ion trap
LPS	lipopolysaccharide
m/z	mass-to-charge ratio
MALDI	matrix assisted laser desorption ionization
MCP	Journal of Molecular & Cellular Proteomics
MHz	megahertz
MOWSE	molecular weight search
MRM	multiple reaction monitoring

MS	mass spectrometry
MS/MS	tandem mass spectrometry (2 mass analyses)
MS^n	tandem mass spectrometry (n mass analyses)
MW	molecular weight
mL	millilitre
mm	millimeter
nm	nanometer
o.d.	outer diameter
PBS	phosphate buffered saline
PEEK	polyetheretherketone
PSD	post source decay
PTM	post translational modification
pI	isoelectric point
PM	particulate matter
q	quadrupole (ion guide)
Q	quadrupole (mass analyzing)
R	resolution
RF	radio frequency
RP	reverse phase
SA	sinapinic acid
SFM	serum-free medium

SRM	single reaction monitoring
SWISS-PROT	Swiss proteomic database
t	time
TFA	trifluoroacetic acid
TLF	time lag focusing
TNF-α	tumour necrosis factor-α
TOF	time of flight
TP	true positive
TPCK	L-(tosylamido-2-phenyl) ethyl chloromethyl ketone
USA	United States of America
V	volts
V_{AUX}	auxiliary alternating current signal
v/v	volume-to-volume ratio
ºC	degree Celsius
µg	microgram
µL	microlitre
µm	micrometre
Ω	angular frequency

1: A BRIEF PERSPECTIVE OF PROTEOMICS, HUMAN INHALATION OF AMBIENT PARTICULATE MATTER, AND ANALYTICAL DETERMINATIONS OF PROTEINS

1.1 Proteomics and its Applications

The synthesis of proteins by ribonucleic acids, coded by genetic information carried by deoxyribonucleic acids is well known (Table 1.1).[1] In general, proteins are composed of a chain of smaller molecular units called amino acids, which are covalently modified to varying degrees by a range of post translational modifications (PTM), plus additional co-factors such as metal ions or organic molecules. However, detailed understanding of protein structure and function from the cellular to organismal level is still not well known, and this has been the motivation behind the development of proteomics. Proteomics can be broadly defined as the study of proteins for: function, structure, abundance, factors that control synthesis, and localization. The word proteome was coined by Marc Wilkins to describe the complete set of proteins and their possible modifications produced by, as encompassed by an experiment that can span in scope from a population of cells, organs, or an organism.[2]

Table 1.1 Amino Acids Resulting from RNA and DNA Codons[1]

Amino Acid	Abbreviation	Letercode	RNA codons						DNA codons					
Glycine	gly	G	GGU	GGC	GGA	GGG			GGT	GGC	GGA	GGG		
Alanine	ala	A	GCU	GCC	GCA	GCG			GCT	GCC	GCA	GCG		
Valine	val	V	GAU	GUC	GUA	GUG			GTT	GTC	GTA	GTG		
Leucine	leu	L	CUU	CUC	CUA	CUG	UUA	UUG	CTT	CTC	CTA	CTG	TTA	TTG
Isoleucine	ile	I	AUU	AUC	AUA				ATT	ATC	ATA			
Prolyine	pro	P	CCU	CCC	CCA	CCG			CCT	CCC	CCA	CCG		
Phynylalanine	phe	F	UUU	UUC					TTT	TTC				
Tyrosine	tyr	Y	UAU	UAC					TAT	TAC				
Tryptophan	try	W	UGG						TGG					
Serine	ser	S	UCU	UCC	UCA	UCG	AGC	AGU	TCT	TCC	TCA	TCG	AGT	AGC
Threonine	thr	T	ACU	ACC	ACA	ACG			ACT	ACC	ACA	ACG		
Cysteine	cys	C	UGU	UGC					TGT	TGC				
Methionine	met	M	AUG						ATG					
Asparaginine	asn	D	AAU	AAC					GAT	GAC				
Glutamine	gin	Q	CAA	CAG					CAA	CAG				
Aspartic Acid	asp	N	GAU	GAC					AAT	AAC				
Glutamic Acid	glu	E	GAA	GAG					GAA	GAG				
Lysine	lys	K	AAA	AAG					AAA	AAG				
Arginine	arg	R	CGC	CGA	CGG	AGA	AGG		CGT	CGC	CGA	CGG	AGA	AGG
Histidine	his	H	CAU	CAC					CAT	CAC				

The development of technologies, techniques, and methods for proteomics has enabled the qualitative and quantitative detection of multiple analytes. This has advanced our capabilities to characterize biological systems, and as a result healthcare strategies for individuals are anticipated to evolve from being reactive to preventative.[3] Each individual is predisposed to a different combination of diseases, and current technologies have the potential to characterize biological samples from an individual to identify molecules (e.g. proteins, metabolites, DNA) that are markers of diseases. Described in this thesis is an application of methodology for proteomics to study cultures of human epithelial lung cells in response to mimics of atmospheric particulate matter, *in vitro*.

1.2 Particulate Matter and Human Health

1.2.1 Particulate Matter and Human Epidemiology

Of all the pollutants in the atmosphere, particulate matter (PM) with an aerodynamic diameter less than 10 μm (PM_{10}) shows the strongest association

with adverse effects on human health.[4] In 2000, a study published by the Health Effect Institute found 90 of the largest cities in the United States of America (USA) revealed an ~0.5% increase in death rate for each PM_{10} increase of 10 $\mu g/m^3$, calculated to be ~60, 000 excess deaths in the USA per year.[5] In 1997 J. Schwartz calculated a 2.75% increase in hospital admissions for each 23 $\mu g/m^3$ increase in suspended particles, with no associations with the common gas phase pollutants SO_2, O_3, or NO_2.[6] (For reference, remote ambient particle loadings in the troposphere are typically ~10 $\mu g/m^3$, and in Vancouver, BC, Canada, values are typically 18-25 $\mu g/m^3$ but can approach 50 $\mu g/m^3$ during temperature inversions.[7]) Findings from studies like these have led to substantial efforts and investments to regulate and reduce PM in the USA, resulting in improved air quality. For example, a 2009 study by Pope *et al.* attempted to address the question: does improvement in air quality result in measurable improvements in human health and longevity?[8] Their research concluded that a decrease of 10 $\mu g/m^3$ of PM_{10} between 1980 and 2000 was associated with an estimated mean life expectancy increase by 0.61+/-0.20 years for the same approximate period (P=0.004).

1.2.2 Particulate Matter

Non-anthropogenic sources of primary PM include sea spray, wind erosion, and natural disasters (e.g. forest fires and volcanic eruptions).[9] Anthropogenic sources of PM come from the incomplete combustion of fossil fuels, in particular diesel. Particles vary in size, but aerodynamic diameters less than 10 μm are of particular interest because Brownian motion due to wind

allows them to remain suspended in the troposphere for hours to days.[10] In the atmosphere, numerous additional chemical processes (oxidation, vapour condensation, adsorption, etc.) can occur to change the composition and complexity of PM.[11-13] Sub classifications based on size, according to the location in the human respiratory tract PM deposit, are upper tract; coarse (>10-2.5 μm), alveolar region; fine (2.5-0.1 μm), alveolar region with diffusion across the endothelium; ultra fine <0.1 μm.[14] Numerous studies have been performed to determine the composition of ambient particulate matter throughout the world, and some material collected has been registered as reference material.[15-18] In 1993, one such study of the outside air of Ottawa was collected and analyzed to characterize components of ambient particulate matter; this sample of PM is referred to as Environmental Health Centre 93 and is a representation of urban air pollution.

1.2.3 Inhalation of Particulate Matter into the Lung

In general, individuals susceptible to adverse health effects of PM increase their risks of myocardial infarction, stroke, arrhythmia, and related vascular and heart diseases within hours to days of inhalation.[19, 20] The possible biological mechanisms to these responses are numerous, and the ability to monitor many different protein mediators would be useful to elucidate the biochemical basis of the varied physiological outcomes. PM induced injury in the lung starts with inhaled ambient particles interacting (in contact, and/or undergoing endocytosis) with epithelial cells or macrophages causing inflammation.[21] These cells secrete mediator proteins (cytokines) into the

circulatory system which go on to stimulate bone marrow as part of a systemic inflammatory response.[22,23] Stimulated bone marrow release neutrophils and monocytes (leukocyte cells of the immune system involved in defending the body against foreign material) into the circulatory system. Cytokines also activate the vascular endothelium for the arrival of leukocytes by promoting the expression of protein for their adhesion. Activated endothelium also releases the protein Endothelin, which activates monocytes and modulates leukocyte-endothelial cell interactions. Other cytokines secreted can function as chemo-attractants (chemokines) for leukocytes to facilitate smooth muscle proliferation and migration back to the site of PM interaction. Studies suggest that inflammation in the lung increase the concentration of markers of endothelial dysfunction secreted in the circulatory system.[24]

Current understanding of ultrafine particles is that they are capable of inducing injury differently, by translocating from the lung into the circulatory system where they cause or promote endothelial dysfunction, possibly by directly damaging blood vessels.[14, 20]

1.3 Established Techniques for Protein Analysis

Numerous methods for protein analysis require proteins to be first isolated, followed by their detection. Enabling criteria for their isolation start with physical and chemical properties such as: charge, mass, hydrophobicity, and affinity to antibodies. Proteins can then be detected by techniques including staining, spectroscopy, or mass spectrometry for further studies. Techniques

involving antibodies or mass spectrometry (MS) are often favoured because they have been demonstrated to possess good selectivity and sensitivity for analysis.

Depending on the techniques used, there is 'hypothesis-driven' or 'discovery-oriented' proteomics. Hypothesis driven proteomics rely on prior knowledge about a system under study, such as proteins that regulate a biological pathway. Discovery-oriented proteomics are studies designed and performed with minimal knowledge of any protein complement specifics; unknown as well as known proteins may be identified by the experiment.[25, 26]

1.3.1 Properties of Proteins for Separation

1.3.1.1 Charge

Proteins placed in a medium with a pH gradient and an electric field, migrate in the electric field towards electrodes of opposite charge (isoelectric focusing).[27, 28] During migration through a pH gradient, proteins gain or lose protons depending on the pK_a of its functional groups; subsequently, net charge and mobility decreases. The isoeletric point (pI) of a molecule is defined as the pH at which it carries no net electrical charge. The pI of a protein is determined by the pH surrounding where it is immobilized, which is a direct consequence of its amino acid sequence.

1.3.1.2 Hydrophobicity

Some functional groups in the amino acids present in proteins contribute hydrophilic (e.g. arginine, lysine, asparagine) or hydrophobic (e.g. tyrosine, phenylalanine, tryptophan) character.[29, 30] In addition, proteins have secondary

structure motifs, described as alpha-helixes and beta-sheets, which alter the protein's hydrophobicity as presented at its water-protein's interface, by either positioning the functional groups of amino acids near the surface or burying them within its structure.[31] Protein tertiary structure is the overall assembly, and protein separation does depend on the extent to which the tertiary and secondary structures remain unchanged (or not) during separation.

In a chromatography column, proteins are separated when carried by a mobile phase through a stationary phase, and hydrophobic/hydrophilic regions allow proteins to become immobilized or retained.[32] The distribution coefficient (Equation 1.1) of a molecule describes its ratio of molar concentrations at a given chromatographic condition, where c_S is the concentration of the molecule in the stationary phase, and c_M is the concentration in the mobile phase.

Equation 1.1 $$D = c_S / c_M$$

Proteins are immobilized when their distributions are exclusively in the stationary phase. Conversely, proteins can move along the length of the column when $D \neq \infty$. Mixtures of proteins become separated by the column if each species has a distribution constant that is different from each other, resulting in a different migration time for each protein.

1.3.1.3 Size

Protein molecules have finite sizes with radii ranging from ~1 to 5 nm for proteins ~5 to 500 kDa large (based on the erroneous assumption that all

proteins can be approximated as spheres).[33] Mixtures of proteins in solution can be separated by their size in size-exclusion chromatography by forcing the solution through a solid support having uniformly sized pores for protein and solvent molecules to diffuse into.[32] Ideally, analyte molecules do not have strong inter-molecular interactions with the solid support. Proteins that become trapped in pores are retained in the column relative to proteins that are too large to fit into the pores. The average residence times for proteins in pores are determined by the relative size of pores versus the proteins. Proteins that are much smaller in size than the pores penetrate deeper and are retained for long periods. Proteins having sizes larger than the pores do not enter and are not retained. Between the two extremes, that of total penetration to total exclusion, proteins can be separated to a first approximation based on size. Protein shape has a lesser role in determining separation on a size exclusion column.

1.3.1.4 Affinity to Antibodies

Antibodies have high affinities for their corresponding antigens (e.g. proteins), and are often used for 'focused' proteomics due to their advantages of being selective and sensitive.[26, 34-37] There are numerous variations in the general theme of using antibodies to perform selective extraction, but the most common embodiment is to immobilize antibodies on a solid support. Methodologies often recommend the use of a universal blocker during the binding step to increase the overall selectivity of antibodies in the presence of a large range of different proteins because binding a target protein to an antibody is thermodynamically regulated. If universal blocker is not introduced, a mixture

of proteins may remain as they can undergo physical adsorption with the surface of the support or antibodies being presented. Upon binding of target protein, they are separated from a mixture as wash steps remove proteins that are not immobilized by antibodies.

1.3.2 Methodologies for Detecting Proteins

1.3.2.1 Staining

'Staining' a protein means that a compound is added to facilitate easier and more sensitive detection. Proteins are stained for detection (e.g. imaging) and/or quantitation purposes.[38] Ideally, the detection limit should be as low as possible while maintaining a wide linear relationship between protein quantity and staining intensity. Colourimetric methods such as Coomassie Brilliant Blue and silver staining are widely accepted for their non-destructive nature allowing additional studies to be performed after staining.[39-41] Fluorescence staining has the same non-destructive nature and has by comparison improved sensitivity, dynamic range, and reproducibility. There are a variety of fluorophores that that are commercially available ranging in sensitivity and wavelength.

1.3.2.2 Spectroscopy

The detection and quantitation (with extinction coefficients) of proteins and peptides is possible by measuring the absorption of light.[42-44] The absorption of 185-220 nm radiation by amide bonds in proteins and peptides is routinely used. The absorption of radiation at 275-280 nm can also be used to quantitate aromatic amino acids (e.g. tyrosine and tryptophan). Laser-induced fluorescence

when exciting at ~275 nm has also been found to be successful for the detection of aromatic residues emitting in the range of 300-350 nm. However, the challenge remains that unless absorption coefficients are known, variations in amino acid composition lead to different absorption values, which complicate quantitation, particularly for mixtures.

1.3.2.3 Reporter Molecules

Fluorescent molecules as reporters covalently labelled to antibodies or proteins can be used for detection and quantification.[26] In one approach termed "direct", a mixture of proteins immobilized onto a surface can be selectively analyzed with labelled antibodies. Related is the immobilization of antibodies for binding target proteins, an approach termed "antigen capture" uses pre-labelled target protein within a mixture for their direct detection. In a separate approach termed "sandwich", a target protein that is bound by an antibody is introduced to labelled antibodies (secondary antibodies) for detection. This method is advantageous over direct techniques because there are washing steps that remove non-specific protein, which decreases cross reactivity. This method is also advantageous over antigen capture techniques because multiple secondary antibodies adhere to the target protein, which amplify signal compared to only a single labelled protein.

Reporting molecules do not have to be bound to proteins of interest, one such example is enzyme linked immunosorbent assay (ELISA).[45] In general ELISA builds upon the sandwich technique involving antibodies. For economic reasons, target protein bound is introduced to secondary antibodies that are not

labelled, but adhere specifically. An additional antibody labelled (linked) with an enzyme is introduced which adheres specifically to the secondary antibody. A substrate is then added to the entire extended "sandwich" and the linked enzyme converts a substrate to a detectable form. The advantage of ELISA is the concentration of detectable substrate increases with time, which can provide up to ~3 orders of magnitude more sensitivity than fluorescent labelled antibody sandwich techniques.

1.3.2.4 Radioactive Labelling

Radio labelling is commonly accomplished by incorporating ^{3}H, ^{14}C, ^{35}S, ^{32}P, ^{33}P or ^{125}I either covalently with reporting molecules or metabolically during cell culturing.[38] Signal detection is accomplished by using photographic film, phosphor imagers, or charge transducers. Though sensitive detection is readily achieved, this technique can only be practiced after receiving appropriate safety training and certification.

1.3.3 Edman Degradation for Sequencing Amino Acids

Edman degradation in combination with automation can provide reliable sequence information for up to 60 consecutive residues.[46] The full potential was realized when Humpback myoglobin (145 amino acid residues) was sequenced in a time of four days, a sensational accomplishment in 1967.[47] The technique sequentially cleaves the N-terminal amino acid, followed by separation, and identification of the amino acid by electrophoresis. The disadvantages include any modification (e.g. glyclosylation or acetylation) to the amine ceases the serial

degradation reaction, and micro-molar quantities of material are required for analysis.

1.4 Characterization of Proteins by Mass spectrometry

1.4.1 Protein Identification by MS

Currently, there are two fundamental strategies for protein identification by MS.[48, 49] In one approach termed bottom-up, a mixture of proteins is subjected to separation and the resulting fractions, containing isolated proteins, are subsequently subjected to degradation by chemicals (e.g. cyanogen bromide) or proteases (e.g. trypsin or pepsin). The resulting peptide mixture is then usually separated by one or two-dimensional chromatography strategies prior to MS analysis. Peptide masses observed from a digested protein can be thought of as a fingerprint, and can be compared with theoretical digests for peptide masses predicted from known proteins and their sequences. A sub-technique of the bottom-up approach is termed shot-gun, where prior to MS analysis, a protein mixture is first digested to peptides. Peptide fingerprinting in this case is not feasible, as multiple peptides from different proteins are too complex for analysis. Shotgun strategies are almost exclusive to laboratories having one or, ideally, two-dimensional LC coupled via an electrospray ionization source to a MS capable of tandem MS (MS/MS). Experiments performed by tandem MS have the potential to fragment precursor ions, leading to sequence information for that peptide. If the sequence observed is unique to a protein, then an identification may be made. A major drawback of the bottom up strategy is the prevalence of incomplete peptide coverage, which in turn makes the analysis of proteins with

unknown sequences, modifications, or isoforms difficult. Reasons for this include differences in peptide ionization efficiency, suppression effects, and irreversible chromatography retention.

In another approach termed top-down, ions of intact or incompletely digested fragments of protein are isolated (by the mass analyzer of the instrument) in the gas phase, and are then fragmented within the context of a tandem MS experiment. The primary advantage of this strategy is the analysis of the entire protein, particularly for PTMs and without the sample handling steps of chemical or enzymatic digestion. Major factors that top-down approaches face are higher instrumentation costs and performance loads, increased complexity with increasing mass range, and software tools for data analysis (e.g. bioinformatics).[50] As a result, development of this approach has been slow.

1.4.2 Quantitative Proteomics

To study biological entities at the system level, in addition to identifying components, their responses to change is hypothesized to be indicative of biological states. However, quantification of all proteins in a biological system is still a challenge because proteins and their proteolytic peptides exhibit a wide range of physical and chemical properties (charge, hydrophobicity, size, etc) which lead to large differences in their MS response.[51, 52] Currently the most accurate approach is based on stable isotope dilution theory, where identical proteins or peptides derived from control versus test samples are labelled with different isotopes of the same element, and are thus chemically identical to each other.[53] These proteins and peptides behave nearly identically during sample

preparation, separation, and MS analysis. For example, after digestion of proteins for peptides, a sample from a negative control labelled with one isotope (e.g. ^{14}N) is mixed with a sample from a positive control labelled with a different isotope (e.g. ^{15}N); an MS analysis can differentiate between the two masses, and within-spectrum quantification is achieved by comparing respective intensities. Methods to introduce stable isotopes to peptides include metabolism, chemically, and via synthetic peptides.

Another approach to quantitation is by label-free methods, however they are considered the least accurate because all systematic and non-systematic variations between experiments are reflected in the data. Two strategies include the direct measuring and comparing of MS signal intensity, or counting and comparing the number of spectra containing an ion signal of interest. Aside from easier use and no additional costs required, advantages for label free methods compared to stable isotope labelling include no limit to experiments for comparison, and ion signals detected do not increase spectrum complexity due to potential overlapping ion signals.

1.4.3 Protein Structure

Studying proteins for their structure contribute to understanding their functions. Common approaches to studying protein structure involve crystallizing them for X-ray crystallography, MS is able to occupy a niche because crystals are not required. The use of MS for structure analysis is relatively new and methods are not nearly well developed. One promising utilization of MS is to exploit hydrogen exchange, in which hydrogen atoms along the amide backbone

of a protein exchange with deuterium in solvent (deuterated water).[54] If a portion of the protein structure is buried within itself and is water-inaccessible, those hydrogen atoms will not undergo exchange. Upon analysis of proteins (fragmented within the gas phase during a top down experiment) with and without hydrogen exchange, fragments that are heavier with deuterium are deduced to be from a position exposed (i.e. surface) to solvent, where as fragments that have no mass change were buried within its structure.

2: APPARATUSES AND METHODS FOR MASS SPECTROMETRY

2.1 Mass Spectrometry (MS)

Mass spectrometry is performed by a mass spectrometer and are both commonly denoted by MS. The development of technology for MS is rich in history regarding improvements that have led to new measurements.[49] The first MS instruments developed were used to study atoms and molecules already in the gas phase (i.e. gases or compounds with high vapour pressures). In incremental steps, efforts to study less volatile compounds eventually led to techniques that have enabled the study of low-volatility compounds in the gas phase. Along the way, technologies to improve the many facets of MS have led to more than just molecular weight measurements being performed. Some MS platforms now include techniques that can determine molecular structure, gas phase acidity/basicity, ionization thresholds, gas phase kinetics/reaction.[55-57]

Modern mass spectrometers contain five main components: an ionization source, a mass analyzer, a detector, a vacuum system, and a data manager (Figure 2.1). The mass analyzer and detector are held under vacuum at a pressure appropriate to establish a mean free path for ions that lead to optimal performance of the instrument. The optimal mean free path is dependent on the type of analyzer used, balances sensitivity versus resolution, as well as controls the extent of unwanted ion collisions with neutral gas molecules. Collisions

cause scattering or fragmentation, where the latter increases spectrum complexity and its interpretation.[58]

The detectors used rely on some form of a transducer for the detection of ions.[32] One type of transducer is an electron multiplier that functions by allowing ions of interest to strike an electron emissive surface in the presence of a strong electric field. Subsequently each electron ejected in turn strikes the surface again initiating a cascade of electrons (amplifying signal) that all arrive within a narrow period of time at an anode, which facilitates ready detection of single events using a pulse discriminator, or average signal levels at higher event rates. MS instruments measure the intensity of a signal for ions based on the instrumental setting, which only after calibration allows knowledge of the ion's mass-to-charge ratio (m/z). The mass of an ion depends on the composition of atomic isotopes, whereas the charge of an ion can depend on a variety of factors including: composition, ionization mechanism, and matrix effects.[59-61] Data is typically presented as a plot of ion signal intensity versus m/z.

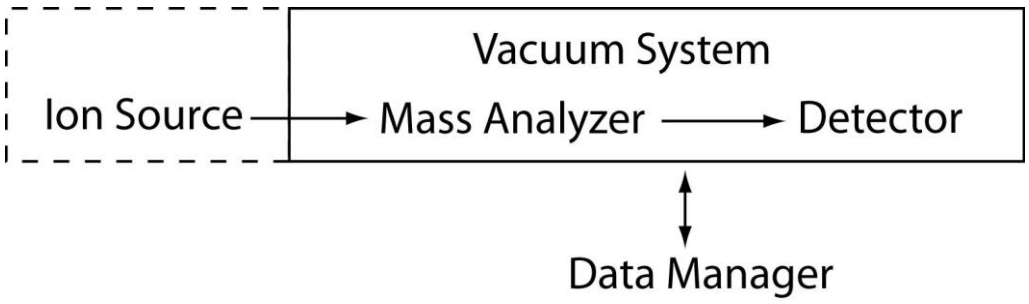

Figure 2.1 Block Diagram of the main components of a modern mass spectrometer.

2.1.1 Ionization Sources

Ionization sources are responsible for producing molecular ions in the gas phase for mass analysis. Early instruments analyzed exclusively gaseous molecules because they only needed to be ionized, typically by electron impact. In incremental fashion, additional molecules found to be volatile (by heating) that were stable in the gas phase were also studied by electron impact.[49] Mass spectrometry as a tool to study the molecules that underpin the foundation of biology began with limited success in that only relatively small and thermally stable molecules such as fatty acids, steroids, and small carbohydrates were amenable.[62] Their ionization by electron impact was performed but the process was much too 'hard' for these molecules as there was a lack of molecular ions and fragmentation patterns were very complex, which precluded spectral interpretation for most purposes. The drive to study other low volatility, thermally fragile, molecules of life drove the demand for developing new ionization sources that were 'soft', meaning that the ionization event (e.g. chemical ionization and plasma desorption) proceeded with minimal internal energy transmitted, thereby producing ions with minimal to no fragmentation.[63-65] Demand for ion sources that could ionize even larger bio-molecules (e.g. DNA and proteins) that are thermally labile, and cause minimal fragmentation led to more recent soft ion sources that include matrix assisted laser desorption, electrospray, and desorption electrospray.[66-69]

2.1.2 Mass Analyzers

There is a variety of mass spectrometers available to separate ions having different m/z. Spectrometers that separate ions “in space” include dispersing ions based on differences in momentum or kinetic energy (magnetic or electric sectors), velocity (time of flight), or periodic movement (quadrupoles).[57, 70-73] The techniques just mentioned are described as separating “in space” because ions need to travel between one point to another for separation prior to detection. Conversely, instruments that trap ions are described as separating “in time”. Trapping analyzers like quadrupole ion traps, ion cyclotron, and orbitraps, separate ions based on the frequency of their motion within a defined space, and the measured regular motion of ions in these devices (e.g. frequencies) are readily converted to masses.[74-76]

2.1.3 Data Management

Modern mass spectrometers have computers to control the various components of the instrument during the collection, storage, and display of data. Development of methodologies to improve analyses of sample with minimal prior knowledge have led to data dependant acquisitions. All MS sold commercially can be controlled by software to assess data in real time to autonomously determine which (if any) additional experiments are to be performed, with no additional supervision from the researcher.[77, 78] These software packages perform well for well characterized sample types.

2.1.4 Interpretation of Mass Spectra for Molecular Weight

A mass spectrum is the result of not only which analyte present but also the processes (e.g. ionization, clustering, fragmentation) to which they were subjected to. To illustrate some of these processes, myoglobin, cytochrome c, adrenocorticotropic (ATC) hormone peptide fragment were mixed together and an aliquot of the mixture was analyzed by MALDI-ToF-MS (Figure 2.2). The ionization of myoglobin by protonation (addition of H^+) was observed: a single myoglobin molecule with one H^+, a cluster consisting of a single myoglobin molecule with a single sinapinic acid fragment and one H^+ (I), a single myoglobin molecule with two H^+ (D), and a cluster consisting of a single myoglobin molecule with a single sinapinic acid fragment and two H^+ (E). Note when a myoglobin (MW 16,950.5) obtains one proton its mass is 16,951.5, charge is 1+, and a signal is expected for m/z 16,951.5. When myoglobin obtains two protons, its mass is 16,952.5, charge is 2+, and a signal is expected for m/z 8,476.3. Similar deductions regarding the identity of other ion signals in this spectrum can be performed, largely because an aliquot of solution containing standards was used as the sample in this instance.

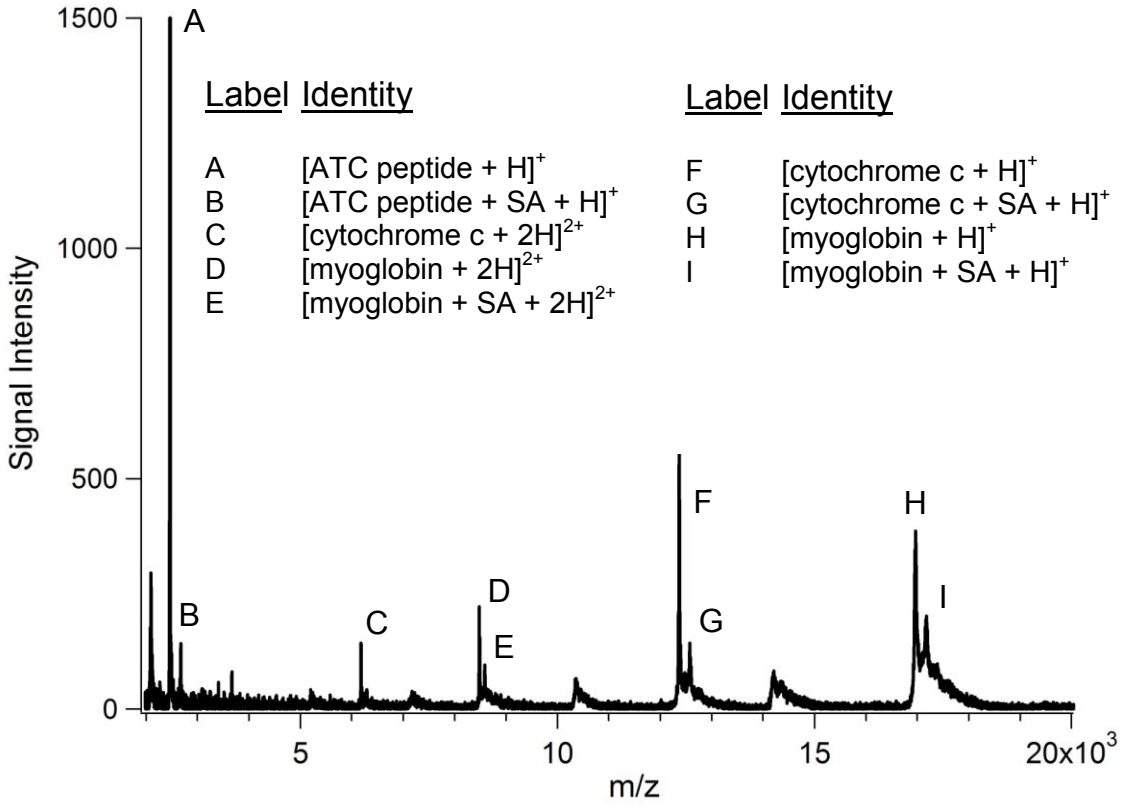

Figure 2.2 Mass spectra of myoglobin, cytochrome c, and ATC peptide fragment in the mass range from m/z 2 - 20 x 10^3 using matrix assisted laser desorption time of flight MS.

When the molecular weight of unknowns is sought, the m/z of the ion signals observed are used, however the subsequent interpretation is not always straightforward because the resolution (Equation 2.1) of this instrument's mass analyzer, a drift tube measuring time of flight, is only ~750 when used in the linear mode. A factor that complicates the interpretation of m/z is 'z', as the number of charges present for the ion is not always known. Two common methods for determining the number of charges on an ion relies on observing other ions that are related. One method of determining the charge is by

observing ion signals from isotopologues, molecules that have the same elemental composition but differ in isotopes. Isotopologues containing stable isotopes of carbon are commonly observed by MS, molecules comprised of ^{12}C with zero, one, two, and so on ^{13}C would each have a distinct m/z. By measuring the difference in m/z (Δm/z) between the ion signals of isotopologues, the reciprocal can determine 'z' (e.g. Δm/z = 0.50 Figure 4.6A). To observe isotopologues, the resolution of an MS instrument needs to be able to differentiate between such ion signals. Different MS platforms have their theoretical limits in resolution, as a result only ions of appropriate m/z will have isotopologues detected separately.

Equation 2.1 **Resolution = (m/z)/(Δm/z)**

When the resolution of an instrument is insufficient to differentiate between isotopolouges (as in the situation for the data plotted in the spectrum Figure 2.2), another method which relies on observing other ion signals that are of the same molecule with different charge states is used (Equation 2.2).[79] If the difference in charge is known, 'b', the ion signal with greater numeric $(m/z)_a$ is assigned to have 'a' charge, and the lower numeric $(m/z)_{a+b}$ has 'a + b' charge. For example, peak H and D in Figure 2.2 are two ion signals observed from the same molecule with a difference in charge of 1 (b = 1), using Equation 2.2 the number of charges is solved for both ion signals.

Equation 2.2

$$\alpha\left[\left(\frac{m}{z}\right)_a + b\right] = \alpha + b\left(\frac{m}{z}\right)_{a+b}$$

2.2 Matrix Assisted Laser Desorption Ionization Time of Flight Mass Spectrometry (MALDI-ToF-MS)

MALDI is one type of soft ionization technique commonly used for MS. The ionization process relies on having matrix and analyte co-crystallize with an abundance ratio of ~10,000:1[61] The matrix is usually an organic acid, having aromaticity enabling absorption of the photons output by a laser pulse (due to lower costs, the N_2 laser with an output wavelength of 337 nm is most frequently used). Upon irradiating a co-crystallized matrix & analyte mixture, the matrix absorbs sufficient energy that results in a plume of neutral and charged matrix and analyte molecules being desorbed into the gas phase (Figure 2.3). This occurs within a low pressure region of an MS, that for a MALDI ion source, is also the ion acceleration region. Charged molecules leave this acceleration stage, ideally at the same time and all having the same kinetic energy, and enter into an electric field-free flight tube wherein ions having different m/z are separated by their respective different velocities, and therefore reach the ion detector at different times.

2.2.1 Ionization by MALDI

There are numerous ionization processes considered when describing how molecular ions may be formed by MALDI.[61, 67] The varied mechanisms for ion formation are often subcategorized as either “primary” or “secondary”.

Primary ionization refers to those processes that lead to gas phase ions from neutral molecules in the sample, after the absorption of photons from the pulsed laser. Secondary ionization processes lead to molecular ions not directly generated by primary processes.

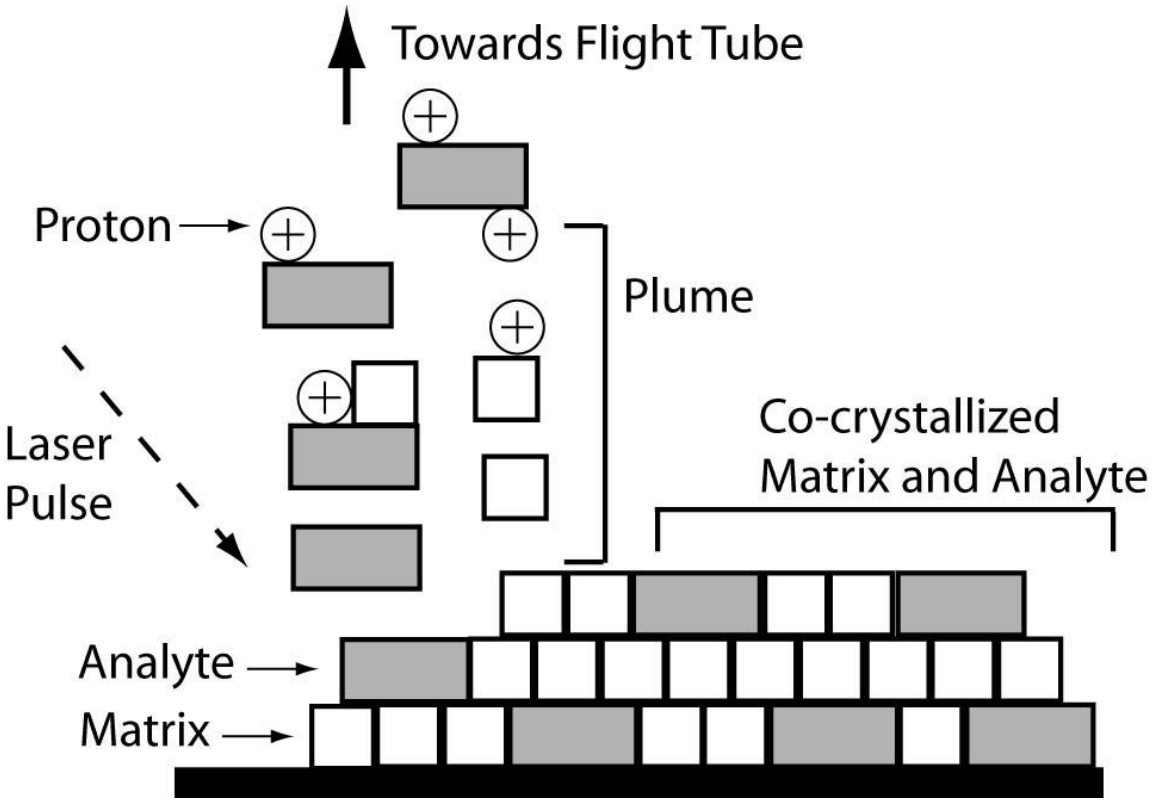

Figure 2.3 MALDI ionization process. Matrix-analyte co-crystallized sample is irradiated with a laser pulse that desorbs and excites mostly matrix, which in turn, forms a plume of neutral and charged matrix that expands into the gas phase while carrying with it analyte. Within the plume, primary and secondary ionization processes lead to gas phase ions.

2.2.1.1 Primary Ionization

One process is described as the desorption of preformed ions, in which ionized analyte are complexed with counter ions in the neutral matrix-analyte solid and are liberated upon laser irradiation. This process is thought of to have

more relevance to compounds intrinsically ionic in character, such as proteins and peptides due to their basic and acidic residues.

Another process is described as excited-state proton transfer in the MALDI plume. It presumes a single matrix molecule (M) is excited by a single photon, weakening a bond with hydrogen, and becoming much more acidic in the gas phase (Equation 2.3). A nearby analyte molecule (A) is then ionized by accepting a labile proton before the excited molecule relaxes (Equation 2.4).

Equation 2.3 $M + h\nu \rightarrow M^*$

Equation 2.4 $M^* + A \rightarrow (M-H)^\bullet + AH^+$

2.2.1.2 Secondary Ionization

Conditions are believed to exist in the MALDI plume that allows secondary ionization processes to occur. One such process is a gas-phase proton transfer that is thought to occur by a two step process. The first is a matrix-matrix reaction that involves a primary ionized matrix molecule colliding and transferring a labile proton to a ground state matrix molecule, protonating it (Equation 2.5). Second, a matrix-analyte reaction occurs that results in the transfer of the proton to an analyte molecule in a gas-phase acid-base reaction (Equation 2.6).

Equation 2.5 $M^* + M \rightarrow (M-H)^{\bullet} + MH^+$

Equation 2.6 $MH^+ + A \rightarrow AH^+ + M$

2.2.2 Time of Flight Mass Analyzers

Time of flight (ToF) analyzers make use of transit times through a fixed length drift tube that is free of electric fields to determine the m/z of ions. After the UV laser pulse event, an electric field is applied to accelerate ions into the field-free drift tube, and the time for ions to travel through the region and reach the detector is measured. In general, the time required for ions to travel through the drift tube increases with m/z (Figure 2.4A). Given a potential difference (V) from the electric field in the ion acceleration stage, drift tube length (L), and transit times (t) the m/z of ions can be calculated with the use of Equation 2.7 to Equation 2.9.

Equation 2.7 $\text{P.E.} = zeV = K.E = (1/2)mv^2$

Equation 2.8 $v = L/t$

Equation 2.9 $m/z = 2eV(t/L)^2$

Variation in kinetic energy and location within the plume during ionization is one factor that affects the resolution, and subsequently sensitivity, of ToF analyzers. To circumvent this, one method is to utilize a delay (<1 ms) prior to

the application of the electric field used to accelerate ions, is often termed delayed extraction or time lagged focusing.[80, 81] The purpose of the delay is to allow ion production from the variety of proposed paths. After the delay, when the electric field is applied, ions that are in different spatial locations experience a corresponding difference in force. This provides an energy correction, provided the detector is positioned at the focal point of that ion source, which improves the simultaneous detection of ions of the same m/z regardless of their initial energy. Another method of narrowing the variation of ion energies is the use of an ion optic device, often termed a reflectron, to change the path of ions as shown in Figure 2.4B.[82, 83] Ions with greater kinetic energy penetrate deeper into the electric field of the reflectron compared to ions with less kinetic energy. Note that ions leave the mirror with the same kinetic energy as they had when they entered, meaning that ions of the same m/z exit the reflectron with a correction in distance travelled, leading to improved simultaneous detection (again, provided the detector is positioned at the focal point for the reflectron). Though a reflectron is capable of improving the resolution of ion signals, a trade off is a loss of sensitivity due to imperfect ion geometries for reflection.

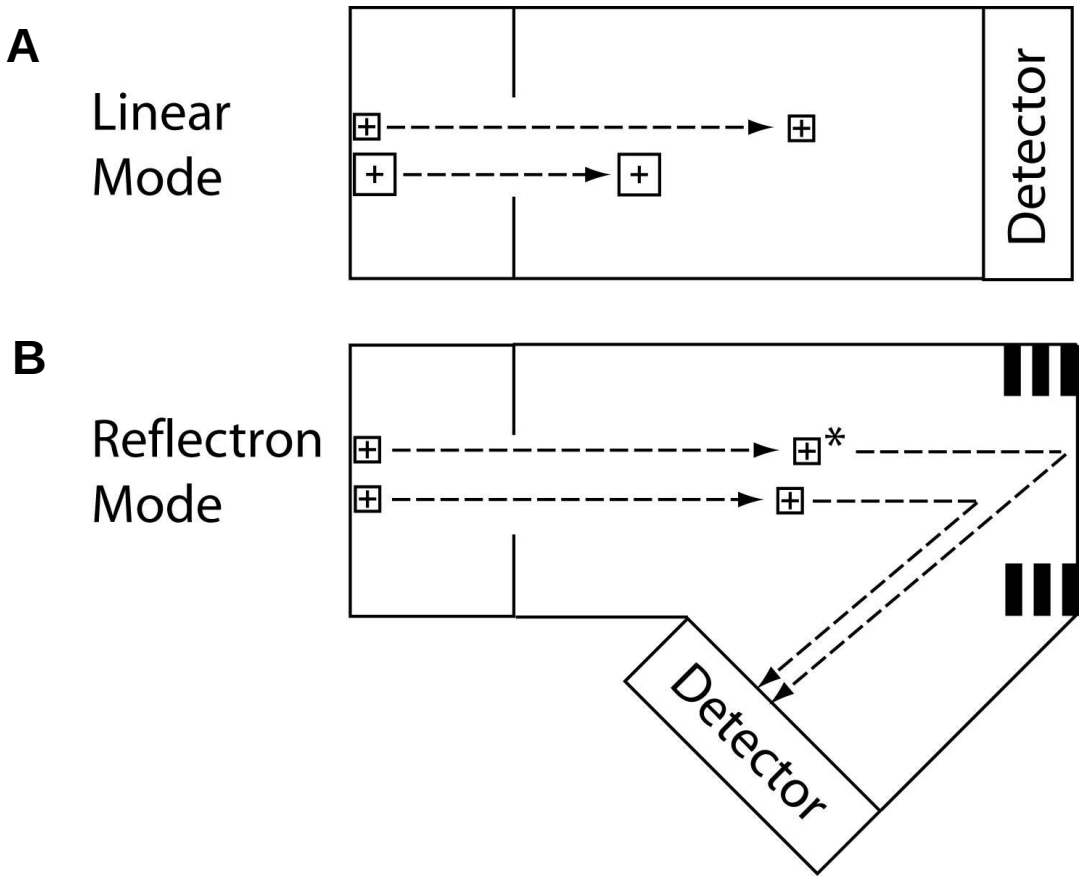

Figure 2.4 Representations of time of flight analyzers, a linear type and a reflectron type (addition of electrodes to help redirect and focus ions leading to improve resolution).

2.3 Reverse Phase Liquid Chromatography Coupled to Electrospray Ionization (RPLC-ESI)

2.3.1 Reverse Phase Liquid Chromatography (RPLC)

The early practices of liquid chromatography (LC) involved the use of bare silica or alumina as the stationary phase for the retention of polar compounds using organic solvents as the mobile phase.[32] Based on that history, materials for the stationary phase subsequently developed to retain hydrophobic compounds were termed “reverse phase” (RP) because the stationary phases in

these column are non-polar. How a column separates hydrophobic molecules was briefly described in Section 1.3.1.2. The RP column's stationary phase is typically straight chain hydrocarbons (typically 4 to 18 carbons in length) covalently bonded to porous (100 to 350 Angstrom in diameter) polymeric beads (2.5 - 5 um in diameter) packed inside a stainless-steel tube, all of which is then referred to as the separation column.[84] The mobile phase for RPLC is usually a miscible mixture of water and organic solvent, optionally buffered or pH modified. Separations performed using a fixed composition of the mobile phase are termed isocratic, where as separations involving changes in mobile phase composition (usually increasing in organic composition) are termed gradient elutions.

Two generic processes by which the stationary phase retains analyte molecules have been discussed in the literature. In one, the hydrocarbon bonded to the stationary phase behaves like a liquid coating, and analyte distribution behaves like a liquid-liquid interaction. In another, the hydrocarbons form a surface for physical adsorption to occur, and solvent molecules from the mobile phase compete with analyte molecules for the surface. Regardless of the mechanism for retention, analyte molecules with greater hydrophobicity have greater distribution coefficient; this typically translates into more time spent in the column and higher volumes of mobile phase to elute the analyte out of the column.

Proteins and peptides are large molecules that each have variable hydrophilic and hydrophobic amino acid sequences. These molecules are suitable for separation by RPLC provided the overall molecular composition are

not too hydrophobic, meaning that those molecules have a propensity to aggregate and/or precipitate in the solvents typically used for reverse phase chrmatography.[85, 86] To help address this, chemical modifications such as derivatization, proteolysis, and denaturation have been employed with some success to circumvent hydrophobic challenges. Proteins separated by RPLC have distribution coefficients with empirically observed characteristics that have been described simply as being “on-off”. In such instances, isocratic elution is poorly suited because distribution constants either greatly favour interactions with the stationary phase (requiring impractical amounts of time to elute) or they greatly favour interactions with the mobile phase (poorly separated). To address this “on-off” character, gradient elution is commonly used to increase the organic content of the mobile phase until the analyte molecule has $D \neq \infty$, at which point it can be eluted “off” the column. At the end of chromatography the mobile phase contains relatively high concentrations of organic solvent to remove hydrophobic molecules potentially remaining in the column. To prepare for the next analysis, the column is typically regenerated by flowing roughly 20 times the volume of the column with the ‘starting’ mobile phase of the next analysis. Column regeneration is needed for improved chromatographic reproducibility.

2.3.2 Electrospray Ionization

Electrospray Ionization (ESI) is a soft ionization technique that has allowed low volatility and thermally labile molecules to be studied by MS.[49, 87, 88] The ESI source operates at atmospheric pressure, thus necessitating interfacing with a MS equipped with an atmospheric pressure gas sampling inlet. Now, all

manufacturers of MS instruments have their own interface that allows gas at atmospheric pressure to be sampled into the MS (more in Section 2.5.2). [89, 90]

The ESI process is initiated by introducing a liquid sample to a capillary (~100 μm i.d.) that has a high DC potential bias (~2-4 kV) relative to a counter electrode, at flow rates 10-2000 μL min^{-1}. The electric field accelerates ions of like polarity in the volume of solution outside the charged capillary towards the counter electrode, forming a Taylor cone with the apex pointing towards the counter electrode. The formation of the cone is the result of two opposing forces; the Columbic acceleration of ions believed to be near the liquid-air interface of the solution toward the counter electrode, and the surface tension of the solution. At the apex of the cone the forces are believed balanced; thereafter ion acceleration causes jet flow in which the ions are accelerated to such an extent that the jet ruptures, leading to the formation of primary droplets having net charge. When ESI is coupled to higher liquid flow rates as encountered with some chromatography apparatuses, ionization is often assisted with pneumatic nebulisation which introduces a concurrent flow of inert gas to aid with solvent evaporation. As primary droplets evaporate, the charges within any one droplet get closer together increasing Columbic repulsive energy, ultimately leading to Columbic explosion that has been observed as uneven with respect to the segregation of net charge and mass. A typical Columbic explosion has been estimated to generate ~20 progeny droplets that together account for ~2% mass but ~15% charge of the initial droplet.[91] This sequence of solvent evaporation and uneven Coulomb explosion is believed to continue ~2 to 4 times, each time

generating progeny droplets evaporating to ~10 nm in dimension, theoretically containing one analyte molecule and a low number of ions and solvent molecules. The final processes for obtaining gas phase ions free of solvent remains unclear, but two processes for such are often discussed in the literature. It is noted that the dimensional and time scale of the final steps of ion generation makes its direct study exceedingly difficult, and hence the reason for the lack of direct experimental data (Figure 2.5).

Iribarne and Thompson proposed a theory they termed ion evaporation theory for the transfer of analyte ions, particularly low MW, in a droplet to the gas phase.[92] They reasoned that molecular ions in solution experience Coulombic repulsion between the excessive charges on the droplet and cause the molecular ion to exceed surface tension leading to its desorption to the gas phase. A different process proposed by Dole termed charge residue theory describes the complete removal of solvent in progeny droplets, through evaporation and repeated Coulomb explosion events; until a single ion remains, which was either formed in solution, acid-base, or adduct formation chemistry.[93] In work by Enke, *et al*., the chemistry in the droplet's having net charge was described for peptides based on equilibrium partitioning as a way to augment Iribarne's initial model.[94, 95] They reasoned peptides needed to contain hydrophilic and hydrophobic portions for successful ionization from a droplet with net charge; hydrophilic regions help form stable ions in solution, and hydrophobic regions tend to become desolvated at the surface where progeny droplets are ejected, leading to molecular ions.

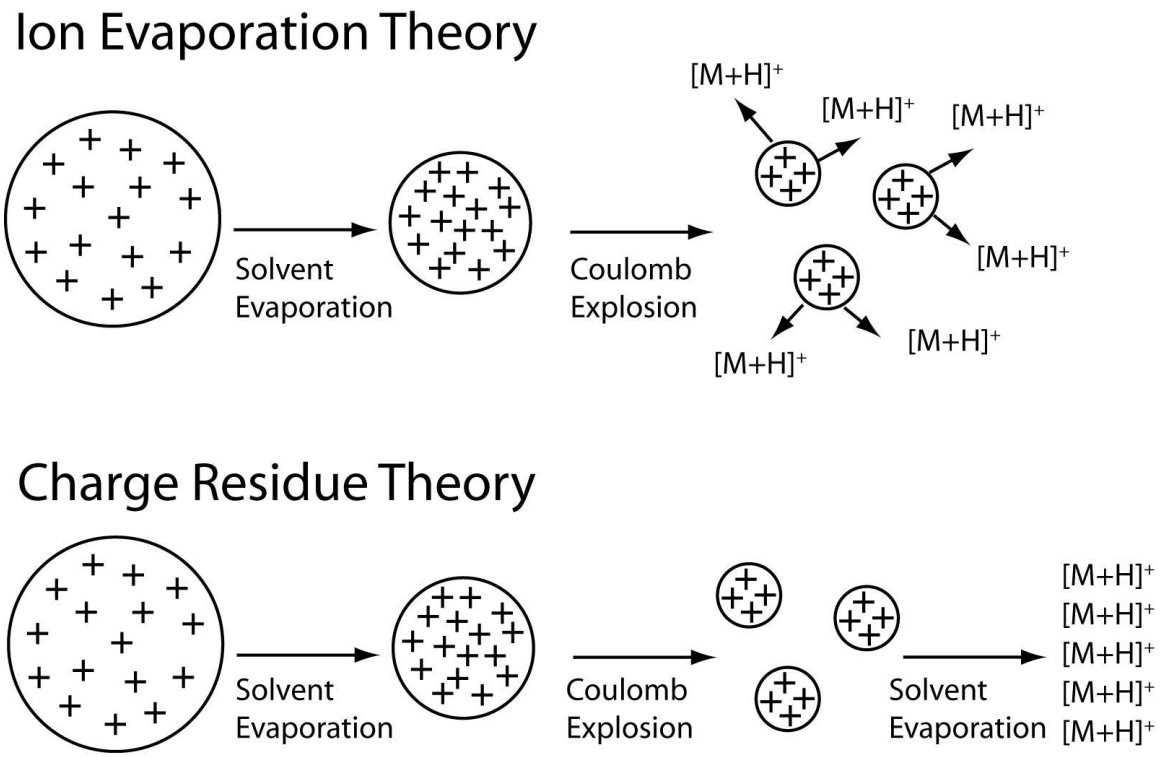

Figure 2.5 Charged primary droplets evaporating yielding smaller progeny droplets that will eventually lead to gas phase ions. In Ion Evaporation Theory, preformed ions desorb from the surface into gas phase (Top). In Charge Residue Theory, solvent is completely removed due to evaporation, leaving molecular ions. (Bottom).

Factors that are found to reduce the sensitivity of ESI include electrolytes and space charge effects of droplets. Electrolytes cause ions to form adducts (with Na^+ or K^+) that decrease the amount of molecules available to become protonated ions, and the generation of clusters from the precipitation of the electrolyte. Sources of electrolytes come from salts used in buffers or biological samples. Space charge effects (increased diffusion due to charge-charge repulsion) within the plume of droplets between the ESI tip and MS has been described as restricting the flux of ions sampled through the orifice of an

atmospheric pressure gas sampling interface.[96] Two strategies to overcome this effect lead to the development of nano-ESI and ion funnel (discussed more in chapter 5).

2.3.3 Nano Electro Spray Ionization

The term nano-ESI describes the nano litre flow rates (1 to 1000 nL/minute) used when reducing the dimensions of the capillary tip of an ESI source to under 10 μm in diameter.[97-99] Though the processes for ion formation are believed to be the same as ESI, miniaturizing the emitter leads to several performance improvements observed in experiments, which are believed to be due to smaller initial droplet formation (~150 nm radius), and lower sample consumption (Figure 2.6).

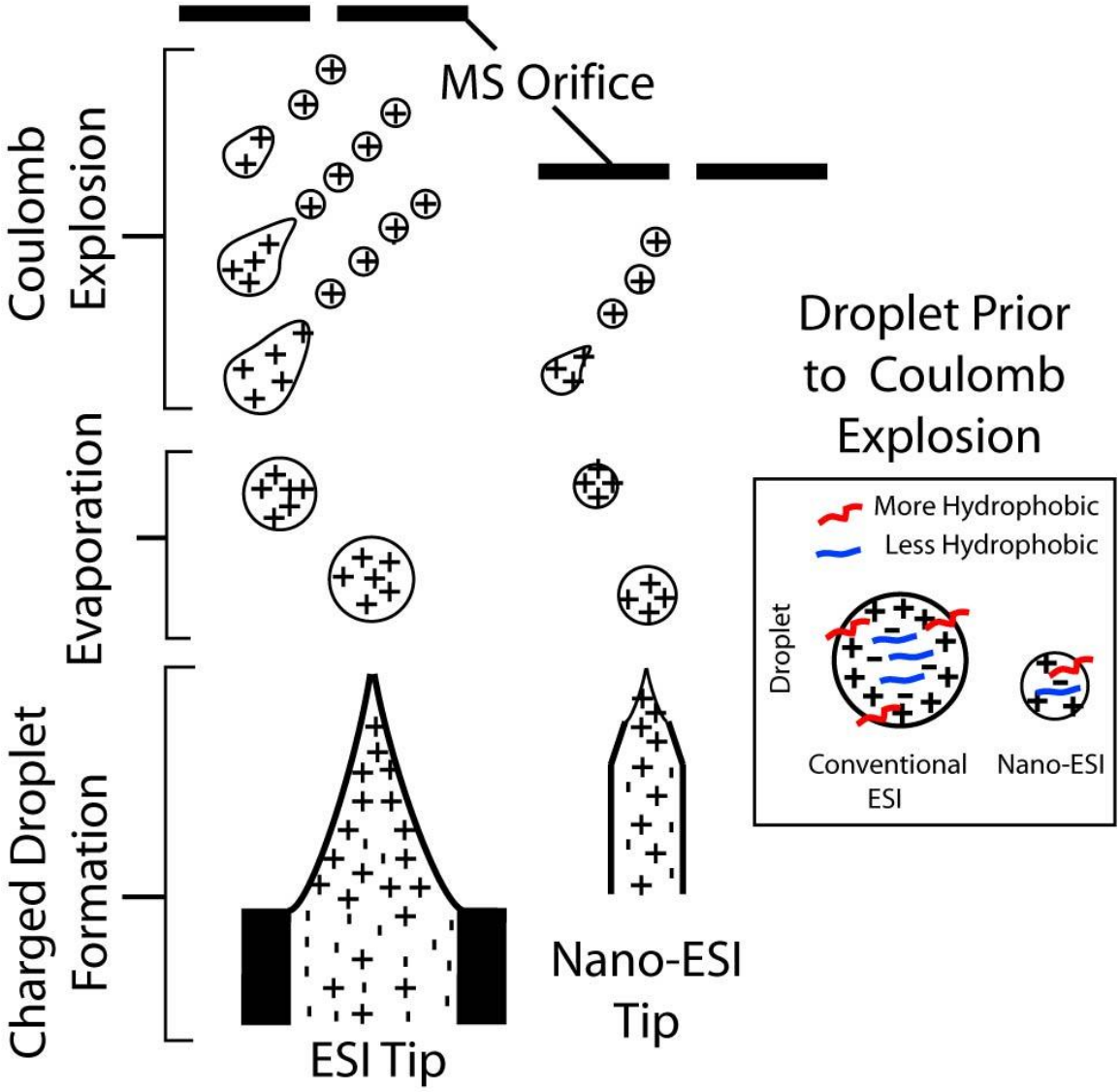

Figure 2.6 **Cartoon representation of ESI and nano-ESI sources, and snapshots of the processes that span from producing charged droplets that evaporate and undergo Coulomb explosions to the processes that lead to molecular ions in the gas phase. Inset, cartoon comparison of relatively smaller droplet sizes which is theorized to improve the ionization efficiency of less surface activated molecules.**

Droplets that are smaller have a greater surface area to volume ratio, which is purported to place the analyte molecules (e.g. peptides, proteins) in the bulk of the droplet closer to the net charges at the droplet-air interface where the ejection of progeny droplets leads to molecular ions. This is argued to improve the ionization efficiencies (ratio of analyte in the solution becoming gas phase ions) of analytes and reduce biasing of molecules classified as being non-surface active. Interference by electrolyte is reduced, because the electrolyte is less concentrated (due to evaporation) when the final steps of gas phase ion formation takes place, again, all attributed to a smaller initial droplet size. Even though space charge effects are still present, forming smaller and fewer droplets leads to physically smaller plumes. This contributes to improved ion signals because the entire plume can be sprayed directly at the orifice of the MS, with, what is believed to be a high fraction of the plume actually being sampled through to the MS .[100]

2.4 Tandem Mass Spectrometry

A generic description of tandem MS is a process in which a mixture of ions enter an instrument, become isolated by a mass analyzer in the first stage, collide with a bath of neutral gas molecules to induce fragmentation, and is followed by the analysis of the resulting charged products in a second stage (MS/MS or MS^2).[57] If additional fragmentation is performed, the resulting tandem MS experiment is termed MS^n where n is the number of mass analyses performed. The type and quality of data collected can vary greatly depending upon the type of reaction or analyzers used.

2.4.1 Dissociation of Molecular Ions for Structural Information

The most common tandem MS experiment performed is the fragmentation of a molecular ion for structural information.[101] The onset of fragmentation can be defined by a characteristic threshold energy, $E_{0,}$ and fragment ions are produced only if the ions internal energy is raised above $E_{0.}$[102] Many different techniques have been developed that vary in mechanism leading to different types of fragments formed, and therefore generate different information about the ion.

Roepstorff-Fohlmann-Biemann nomenclature is often use for the fragments of peptides and proteins. The resulting fragment ions after dissociation are labelled depending on which bond along the peptide backbone was broken and which fragment retains the charge (Figure 2.7).[103] For any given peptide or protein containing n numbers of amino-acid residues, fragments that contain the N-terminus ionized are labelled consecutively, a_m, b_m, and c_m where m is the number of residues in the fragment ion. For charged fragments that contain the C-terminus, they are labelled consecutively $z_{[n-m]}$, $y_{[n-m]}$, and $x_{[n-m]}$ from the end.

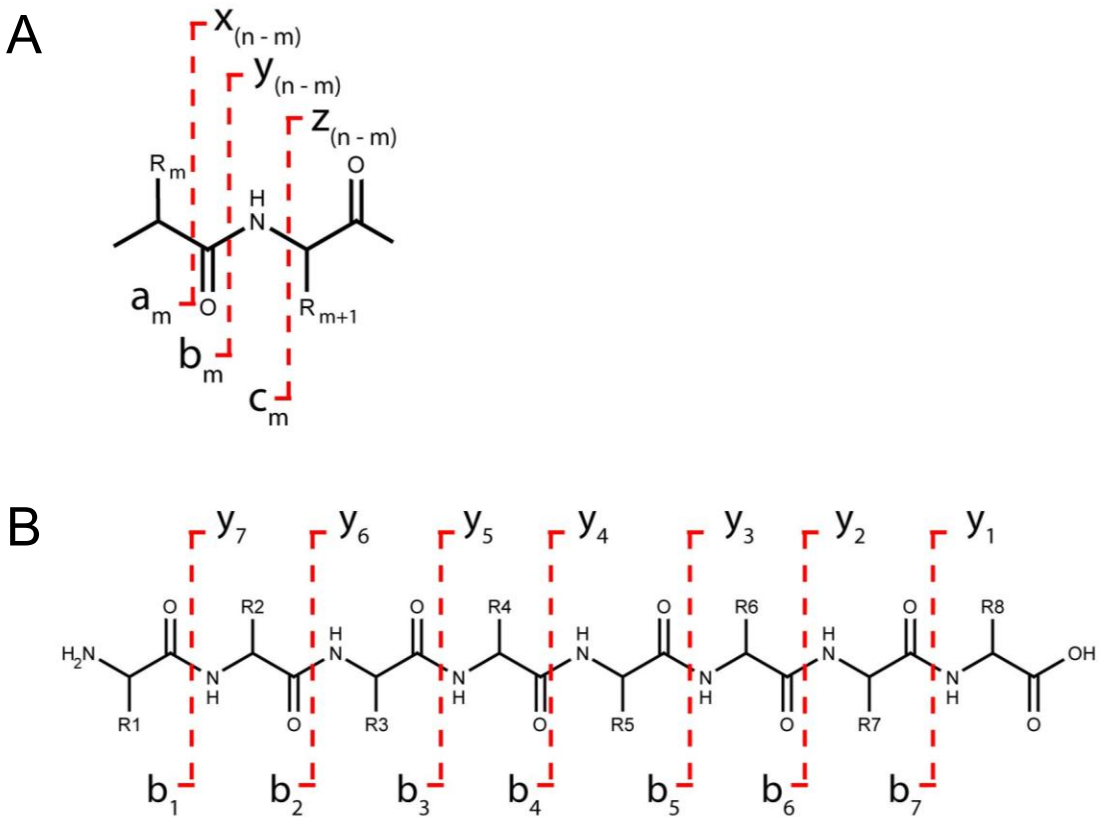

Figure 2.7 Roepstorff-Fohlmann-Biemann nomenclature for naming peptide fragments. (A) Ions labelled consecutively away from the amino terminus of the intact peptide are a, b, and c. When charges remain on fragments containing the carboxy terminus, they are labelled z, y, and x away from the end. m is the number of amino-acid side groups the ion contains, and n is the total from the peptide. (B) A sample peptide depicting broken CO-NH bonds is used by way of an example to illustrate the alpha-numeric nomenclature for b and y ions.

2.4.1.1 Metastable Unimolecular Dissociation by Electron Impact (EI)

Though rarely used for the analysis of biomolecules, fragmentation of molecules by EI is relatively well understood and serves as a point of comparison for other methods.[102] EI equipped mass spectrometers are capable of mass analysis and structure elucidation. First, a ~70 eV electron beam ionizes molecules generating a radical cation having excess energy above the ionization

threshold. An accelerating voltage is applied, and the resulting electric field directs ions towards the mass analyzer. As the ion is directed away from the source, its internal energy re-distributes from an excited electronic state to a ground state that is vibrationally and translationally excited. If enough energy accumulates in the appropriate modes, fragmentation of the ion may occur. This process occurs in a high vacuum environment (10^{-5} to 10^{-9} Torr) primarily to improve electric field focussing of the ions. It also minimizes over complicating the spectrum due to collisions between ions and neutral gas molecules inducing additional fragmentation.

2.4.1.2 Collision Activated Dissociation (CAD)

When an ion collide(s) with a neutral gas (e.g. Ar or N_2), a fraction of the kinetic energy is transferred into the ion's internal energy.[102, 104] From the studies of CAD for small molecules, it has been demonstrated that high-energy (fast) collisions excite an ion to a higher electronic state. If the energy accumulated is not randomized, it is often released into a vibrational mode for dissociation. CAD for structural information of ions is normally performed under conditions to obtain multiple collisions to deposit sufficient energy for several fragmentation pathways to occur with sufficient intensity. However, with an increase in number of collisions, mechanistic details are less understood. Difficulties also arise from larger molecules having less efficient energy transfers from collisions, and much greater degrees of freedom for energy redistribution. This imposes practical limitations on the molecular weight of ions that can be efficiently fragmented using CAD.

One rationalization for the observed dissociation of peptides is described by the “mobile proton model”, which assumes that protonation occurs at the most basic site and activation by collisions enables the proton to explore less basic sites along the peptide.[104-106] Wysocki and co-workers have shown that activation of an amide bond (due to collisions) weakens the adjoining bonds. This leads to ‘charge-directed’ cleaving which rationalizes the domination of fragmentation at NH-CO sites forming ‘b’ and ‘y’ ions.

2.4.1.3 Electron Capture Disassociation (ECD)

Electron capture reactions require instruments that can trap ions long enough to introduce thermal electrons.[50] Though the details of the mechanism are not well understood, this technique involves the capture of low-energy electrons by precursor ions that are positive and multiply charged, followed by charge state reduction and fragmentation. Fragment ions are produced because electron-ion recombination is an exothermic process that causes amine sites to fragment along the peptide backbone (generating primarily ‘c‘ ions), while typically preserving PTM. Fragmentation patterns from ECD are significantly different from CAD and the two techniques are complementary for sequence analysis.

2.4.1.4 Infrared multi-photon dissociation (IRMPD)

Ions can be excited and subsequently fragmented by the absorption of multiple IR photons.[102] IRMPD has been well utilized in the study of small molecules, and recently this method has witnessed increased applications to

proteins due to the growth in popularity of trapping instruments that have long ion trapping times. Typically, trapped ions are activated by a low-power (<100 W) continuous-wave CO_2 (10.6 μm) laser for an irradiance period (10^{-2} - 10^{-1} s) followed by the detection of resulting fragment ions. The mechanism for activation is assumed to be through stepwise absorption of IR photons, with rapid redistribution over all vibrational degrees of freedom. The outcome is an internal energy redistribution leading to similar fragmentation patterns as observed in CAD.

2.4.2 Tandem MS Scans

With the advent of different ion fragmentation techniques for MS, methods utilizing ion fragmentation for information vary in the amount of prior knowledge of a sample.[107-109]

2.4.2.1 Product Ion Scan

Product ion scans are typically performed for samples with very little prior knowledge, particularly with regard to the identity of ion signals. For this tandem MS experiment, all ions other than the analyte ion of interest from the ion source are removed by the first mass analyzer, and then the precursor ion is subjected to conditions leading to fragmentation. Fragments are then analyzed (requiring another stage of mass analysis) for their m/z values, and the resultant spectrum can be interpreted to deduce the original structure or identity of the analyte ion.

2.4.2.2 Precursor Ion Scan

Precursor ion scans are used to determine which unknown precursor ions will fragment to produce an ion of known m/z. Typically this scan is useful for studying small molecules such as the compounds in a metabolic pathway. For example, all acylcarnitines and carnitines of the same derivatization with either methyl or butyl ester (m/z 99 or 85, respectively) have product ions in common.[110] While the final mass analyzer is set to monitor the m/z of a fragment ion of interest, different precursor ions are fragmented to be monitored for producing interpretable ion signals at the detector. At the end of the experiment, the m/z of precursor ions are reviewed to learn which precursor ions led to the product ions of interest.

2.4.2.3 Neutral Loss Scan

A neutral loss scan is usually performed to determine which ions contain a neutral component of known mass that was connected by a labile bond. For example, neutral loss scans are found to be useful when studying which peptides have PTMs with adducts of known masses. Precursor ions by order of m/z are selected by the first mass analyzer, and are directed to subsequent fragmentation. Next, the final stage of mass analysis is set to monitor the m/z of the current precursor ion less the mass of the neutral portion under study - note that this type of experiment can only be readily performed on a triple quadruple MS. By the end of this experiment, the m/z of precursor ions are reviewed to learn which ions led to losses in mass of interest.

2.4.2.4 Selective Reaction Monitoring (SRM)

Selective reaction monitoring requires prior knowledge regarding a pair of precursor and fragment ions to monitor.[111, 112] The first mass analyzer selects a specific m/z ion for fragmentation while the second mass analyzer is set to monitor the signal of a specific fragment m/z. A related technique is multiple reaction monitoring (MRM) in which multiple known pairs of precursor and product ions are to be investigated, some of which are from the same precursor. SRM's and MRM's are typically employed for improved reproducibility, sensitivity, and quantitation performance. Sensitivity is usually improved because the duty cycle for any one analyte ion of interest is increased, due to eliminating the time to measure other ions that are non-information bearing. With a longer duty cycle, the experimental precision is improved because the instrument monitors the ion signal intensity of the pair of precursor and product ions more frequently, and in an LC-MS experiment, this minimizes the chance of missing the analyte's elution from the column. For instruments that have ion trapping capabilities, a linear ion trap hybrid system was demonstrated in operation with SRM as being capable of improving sensitivity by a factor of 3 versus that obtained when conducting full mass scans.[107] In addition, from the same study with an ion trap, the linear dynamic range was found to be 5 orders of magnitude in dynamic range, an improvement of 2 to 3 orders of magnitude over that achieved when performing full mass scans.

2.4.3 Hybrid Instruments for Tandem MS

Hybrid instruments use various combinations of mass analyzers to obtain desired performance characteristics.[57] Typically, resolution, sensitivity, and mass accuracy in each stage of the tandem MS analysis are of primary importance, but consideration of duty cycles and cost result in trade-offs. The development of many early commercial hybrid instruments was focused on fast tandem MS acquisition due to the time scale of gas-chromatographic separations. Early hybrid instruments (1980s) used analyzers that separated “in space”, as trapping instruments for separation “in time” were just being investigated in academic laboratories. Subsequently (e.g. ~1990s), combinations of analyzers separating “in space” and "in time" instruments were explored. More recently (e.g. ~2000), trap-trap type hybrid instruments have been commercially developed.[113, 114]

2.5 Hybrid Triple Quadrupole Linear Ion Trap Mass Spectrometer (Q-q-LIT-MS)

2.5.1 Linear Quadrupoles

Linear quadrupoles have been described and demonstrated to be functional as ion guides, mass analyzers, and ion traps.[115-117] The difference in how a linear quadrupole is operated is with regards to the electric fields applied to it. Ideally a linear quadrupole is constructed with four electrodes having hyperbolic (though most commercially available instruments are circular) cross sections, mounted equidistance in a radial array (Figure 2.8). In the instance of an ac field only, the quadrupole functions as an ion guide.

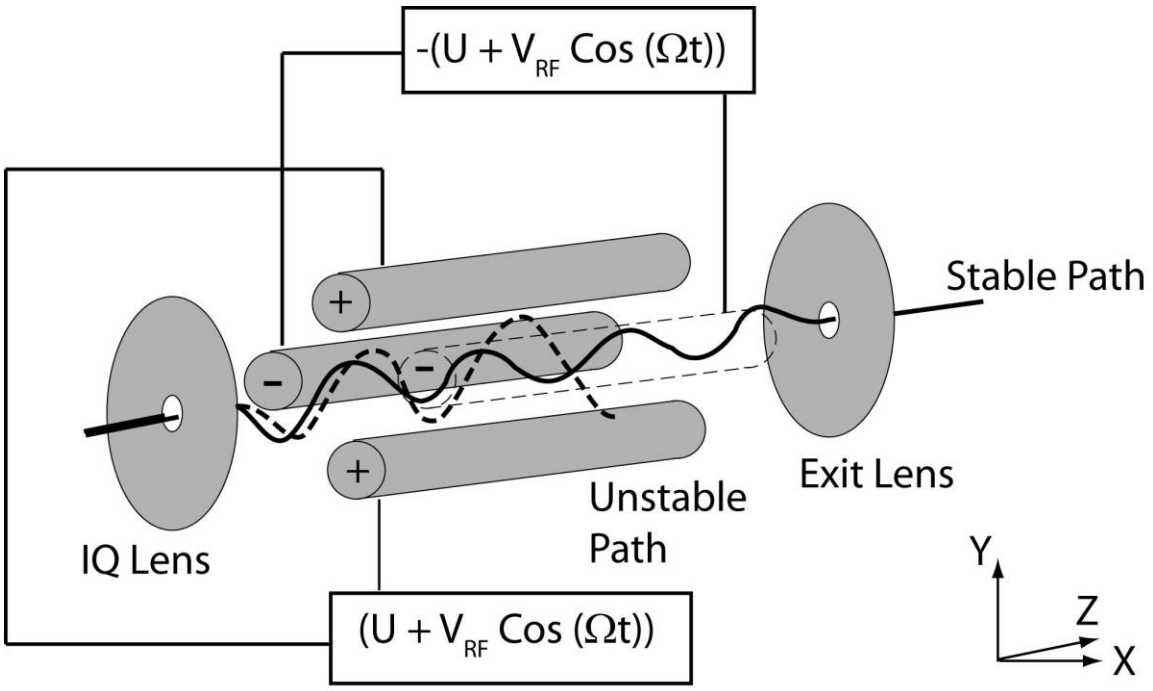

Figure 2.8 **Linear quadrupole for use as an ion guide, mass analyzer, or ion trap depending on the electric fields (and electrodes) applied. As an ion guide, an RF waveform (no DC potentials applied) maintains ion trajectories within the quadrupoles, preventing those ions from impacting, and neutralizing, on the rods. An RF-only ion guide transmits ions within a range of m/z. A linear quadrupole operated as a mass filter, has an RF waveform and DC potential applied to the quadrupole rods so as to reduce the size of the window of ions transmitted to a small range, ideally to a resolution of ~1,000 (R =M/Δm). A linear quadrupole operated as an ion trap, has lenses (inter-quadrupole and exit) at the ends of the quadruple rods charged to prevent ions from escaping. In the figure, an ion of the proper m/z has a trajectory, depicted by the solid line, which allows that ion to be transmitted through the quadrupole. Conversely, ions whose m/z is not selected has a trajectory that causes them to impact on the quadrupole rods (dotted line) that leads to their removal from the ion beam.**

A linear quadrupole when operated as an ion guide has applied to it a radio frequency (RF) waveform (typically ~1MHz sine wave). The waveform is applied to the quadrupole rods in pairs, with one pair being the rods mounted opposite to one another across the central axis (i.e. X-Z plane). The RF waveform is applied at 0^0 phase difference, with the second pair of rods (i.e. Y-Z plane) having the

same RF waveform applied but 180^0 out of phase relative to the first pair. The electric fields formed can be adjusted to permit ions of interest to traverse through the quadruple (i.e. stable trajectory) versus unwanted ions that are caused to change trajectory in a manner that causes them to impact into the electrodes.

$\Phi_{x,y,z,t}$ is the potential at a given time and position in Cartesian coordinates: x- and y-axis are perpendicular to electrodes, z-axis is down the centre of the quadrupole set, r_0 is the distance from the central axis to any of one the quadrupole rods, U is the DC potential applied, V_{RF} is the amplitude of the AC waveform (often fixed in radio frequency regime ~ 1MHz) applied, and Ω ($2\pi f$) is the angular frequency of the sinusoidal AC waveform (**Error! Reference source not found.**). As $\Phi_{x,y,z,t}$ does not contain a z-component, there is no force acting on an ion along the z-axis, which allow ions to traverse along it freely, unless additional charged electrodes are placed at the ends as barriers to facilitate trapping of ions along the z-azis.

Equation 2.10

$$\Phi(x,y,t) = \left(\frac{x^2 - y^2}{r_0^2}\right)(U - V_{RF}\cos\Omega t)$$

The motion of ions within quadrupoles have been well studied and can be described mathematically by the solutions of second-order linear differential equations originally described by E. Mathieu; from investigating the mathematics of vibrating stretched skins (e.g. drums), solutions were described in terms of

regions of stability and instability. These solutions and ideas of stability and instability can also be used to describe the trajectories of ions for analysis or trapping, and are commonly refer to as the Mathieu equation. From solutions of the Mathieu equation, terms can be collected into two parameters ‘a’ and ‘q’, referred to as stability parameters, to describe the stability of an ion within a quadrupole (Equation 2.11 and 2.12). Where *m* is the mass of the ion, e the elementary charge, and z is the number of net elementary charges on the ion.

Equation 2.11

$$a_x = -a_y = \frac{8zeU}{m\Omega^2 r_0^{\,2}}$$

Equation 2.12

$$q_x = -q_y = \frac{4zeV_{RF}}{m\Omega^2 r_0^{\,2}}$$

The stability regions of an ion can be viewed graphically by plotting stable and unstable conditions for ion trajectory in both a and in q space. By overlaying the a and q plots, areas of overlap are regions in which conditions for ion trajectory are stable in both the x and y direction (Figure 2.9A).

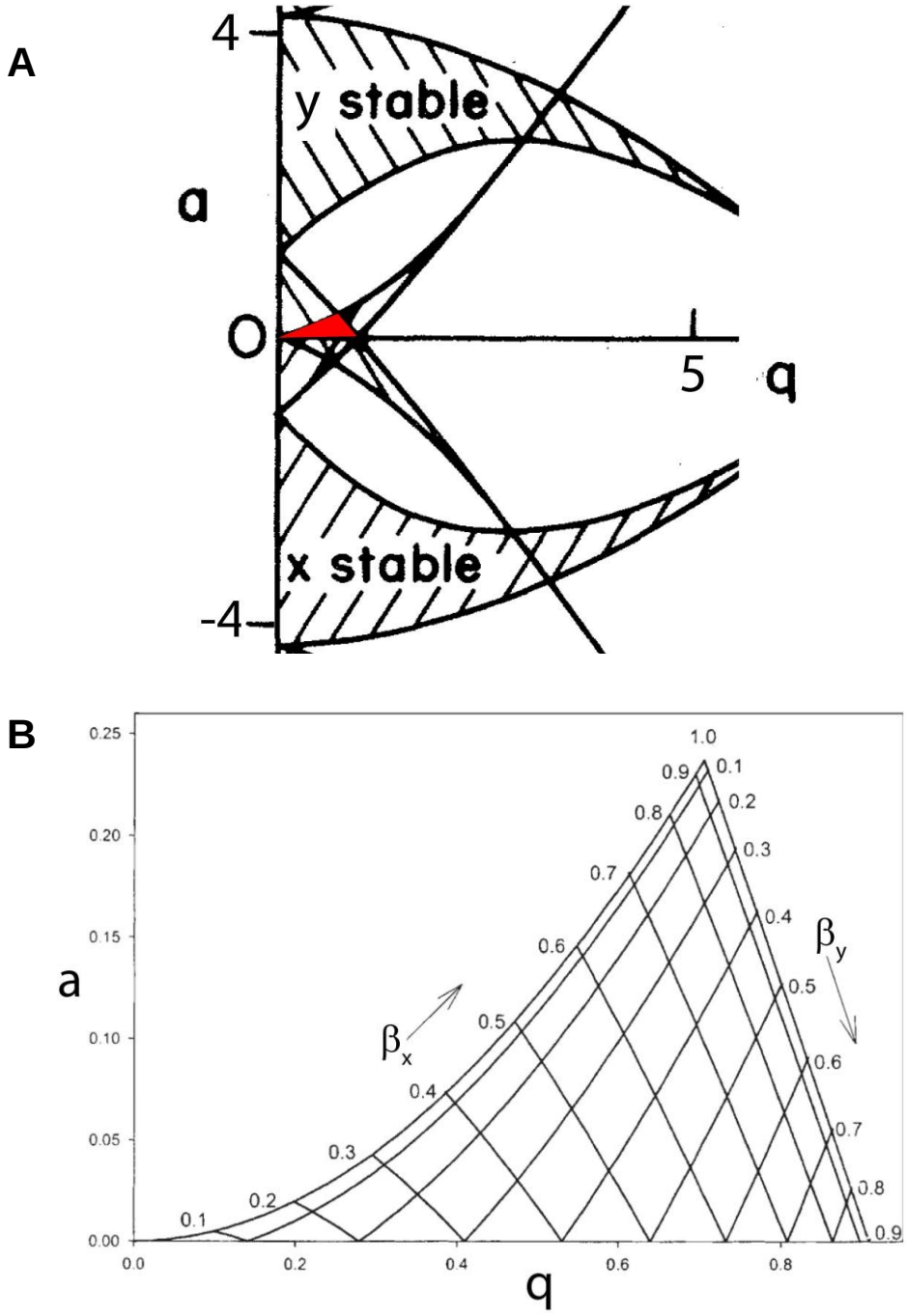

Figure 2.9 **Graphical representation of the stability regions of a quadrupole ion trap determined by Mathieu equation. (A) Mathieu stability diagram for both the x- and y- direction plotted in a and q space (i.e. a_x a_y and q_x q_y). Figure adapted with permission from an article by March, R. E.[118] (B) Stability diagram for the first overlap region, identified in panel (A) by the region in red. Figure reprinted with permission from an article by Douglas, D.J.[117]**

Quadrupoles operating as mass filters (sometimes denoted using a capital 'Q') have the background gas (e.g. N_2) pressure maintained such that the mean free path of an ion is greater than the length of the quadrupoles, meaning that an ion will not usually experience a collision with a neutral background gas molecule while being transmitted. Conversely, when quadrupoles are operated as an ion guide, denoted by small letter "q", improved ion transmission has been demonstrated when operating with the background gas pressure at elevated levels, such that on average, an ion being transmitted has multiple collisions with background gas molecules. The actual number of collisions an ion experiences depends on the background gas pressure and the collision cross section of the ion and target molecules. Some RF-only linear quadrupole ion guides are operated with a background pressure in the range from ~$1x10^{-3}$ to ~$5x10^{-2}$ Torr. In a 3-dimensional quadrupole ion trap, also normally operated with RF-only albeit with different electrodes (to be described), the normal operational background gas pressure for optimal performance is ~5 x 10^{-3} torr. In an ion trap, background gas is introduced to remove kinetic energy from the ions which acts to cause their median location within the ion trap to move toward the center of the device and also causes the mean displacement of ion motion to decrease. Hence the added gas is often referred to as a collisional cooling gas.

When a quadrupole is used as a mass filter, denoted Q_x after the discussion in the preceding paragraph, two methods of using the stability region exist. In one, the quadrupole is operated at the apex of the stability diagram, which allows only ions with values of q = 0.706 and a = 0.237 to have stable

trajectories and thus be transmitted (Figure 2.9B). By substituting the apex value back into Equation 2.11 and Equation 2.12, it is found if V_{RF} and U increased while maintaining a constant V_{RF}/U ratio of 5.96, ions increasing in m/z are selectively transmit through, and abundance is plotted for a spectrum.

Another method is when the linear quadrupole operates as a trap (LIT) and only V_{RF} is applied while U=0 (i.e. a = 0). Based on the stability diagram ions of varying q values are stable within the quadrupole. The addition of inter-quadrupole (IQ) electrodes (lenses) biased with DC potentials at the ends forms a potential well that trap ions, provided their translational energy was not too high. The plot of the first stability region also contains the isobaric lines for β, which describe the values of a and q needed to maintain a constant β (for either x or y). β values are useful at boundaries between stability and instability, one in particular is β_x = 1 (right edge of Figure 2.9B) as it intersects with the q-axis at 0.908 when a = 0. This point is termed the low-mass cut-off. The trapping conditions can be manipulated (i.e. increasing amplitude of V_{RF}) such that ions having increasing q values (i.e. m/z) move from stability to instability radially over the length of the quadrupole rod structure, subsequently allowing abundance versus m/z to be plotted. A disadvantage with radial ejection in these LIT devices is their significant regions for ion ejection from the device, leading to many ions simply not detected. To address this, axial ejection was developed by the addition of an auxiliary quadrupole field from an auxiliary AC signal (V_{Aux}) applied to the lens for which ions are to leave the linear trap through the exit lens. In this configuration, the improvement in ion extraction was reported to be ~50 %.

(For expert readers, the V_{aux} is analogous to the widely known parameter (and methodology) in the 3-dimensional quadrupole ion trap literature as the 'tickle' voltage, or mass-range extension supplemental voltage.[119])

2.5.2 Hybrid Q-q-LIT-MS for Mass Spectrometry

State-of-the-art instruments like a hybrid Q-q-LIT-MS belong to a platform family of triple quadrupole MS instruments because of historical reasons.[101, 120] Today, such instruments contain four or more sets of quadrupoles present as manufacturers have discarded DC-biased electrodes as ion guides and replaced them with high pressure RF-only linear quadrupoles. They improve ion transmission efficiency between successive linear quadrupoles by collisionally dampen ion motion in the X-Y plane while still transmitting them along the z-axis.[121]

Modern mass spectrometers that are designed to be front ended by ESI necessitate an atmospheric pressure gas sampling interface. For this instrument, the interface is an orifice plate-skimmer setup. On the atmospheric pressure side of this interface is an additional electrode (face plate) that has a hole drilled through it (~ 3 - 10 mm diameter) and that plate is biased with a potential that repels species with net charge of like polarity (e.g. droplets, ions, particles) to minimize their striking the surface and neutralizing (Figure 2.10).[122] In between the face plate and orifice plate, there is a usually counter current flow of a gas that is low in the partial pressure of the solvent being used in samples so that it flows out of the face plate away from the MS to aid in redirecting neutral ambient contaminants, droplet evaporation, and molecule-solvent ion de-

clustering. Ions that travel through the orifice experience jet expansion due to a drop in pressure, to ~ 1 - 5 Torr) depending on the design of the interface in the region between the orifice plate and the skimmer. In passing through the skimmer, ions experience another pressure drop to (~$1x10^{-3}$ - $5x10^{-2}$ Torr), again, dependent on the design of the ion optic path of the MS.

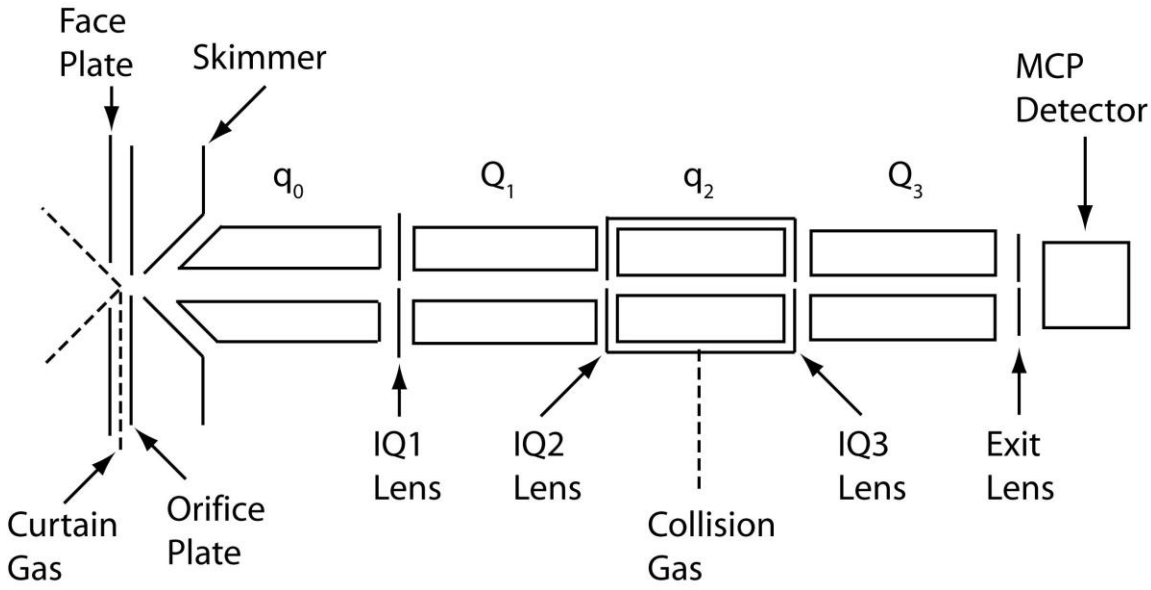

Figure 2.10 Components of a hybrid triple quadrupole linear ion trap MS. A face plate, orifice plate, and skimmer are common elements in most commercial atmospheric pressure gas sampling interfaces. The linear quadrupoles q_0, Q_1, q_2, and Q_3 are axially aligned. In the MS used in this work, Q_3 was also able to be controlled as a linear ion trap. In the space between the quadrupoles are lenses that function to improve ion acceleration, extraction, trapping, and in the case of q_2, to assist in establishing a partial pressure of target gas (e.g. collision) for collision activated dissociation (CAD).

Behind the skimmer are ion optics, and on most quadrupole-based MS instruments manufactured since ~2005, the ion optic in this region is an RF-only linear quadrupoles (q_0). The function of q_0, is to collisionally dampen ion motion

in the X-Y plane while still transmitting them along the z-axis. Throughout the instrument, all charged surfaces (plates, quadrupoles, and IQ lenses) are biased to controllably accelerate ions from the orifice to the detector in a cascading fashion (a good analogy is water flowing over a vertical drop where there are a series of drops).

Ions that enter the instrument travel through the quadrupoles in order of their subscripts. The quadrupole Q_1 is capable of mass analysis, followed by the quadrupole q_2 commonly referred to as a collision cell, which operates within a chamber pressurized with an inert gas (e.g. CAD gas, P = ~5 x 10^{-3} Torr). In the case of this hybrid mass spectrometer, the last quadrupole Q_3 can operate as a mass filter or as an LIT.

To collect a mass spectrum of ions sampled by the instrument, ions travel through q_o, $Q_{1,}$ and q_2 operating as ion guides until they arrive at Q_3. For conventional triple quadrupole MS instruments, ions continuously enter Q_3 and the V_{RF} and U potentials are continuously adjusted to sequentially allow ions to have stable trajectories for transmission to the detector to construct a spectrum. Note that Q_1 is operated independently of Q_3, but for improved ion transmission efficiency, the V_{RF} of q_0 is related to that of Q_1, and similarly, V_{RF} of q_2 is related to that of Q_3. Hybrid instruments that have a LIT for Q_3 can operated to analyze ions continuously, or as discrete populations that are trapped and accumulated within a set amount of time, and then have V_{RF} in conjunction with the V_{Aux} on the exit lens adjusted to sequentially scan out ions to generate a spectrum.

The rate at which ions are scanned for analysis is usually performed as fast as possible, to maintain high duty cycle, and because there is a trade-off, ion signals detected under such conditions are often lower in resolution (e.g. scan rate 1000 Da/s have R = ~1000) than can be achieved by the instrument. Such low resolution data preclude the observation of isotopolouges. To address this, ion signals of interest are often re-analyzed by filling the LIT with ions, followed by a narrow window mass scan out of m/z (30 m/z) centred around the ion of interest at a slower rate (e.g. scan rate of 250 Da/s which has the effect of increasing resolution to ~6000), and data acquired under such conditions are at sufficient resolution to enable observation of isotopologues.

When a tandem MS instrument is coupled with an LC to first separate unknown analytes (e.g. there is no apriori information regarding the m/z or elution times for what molecules available and at what abundance), the automatic selection of ions for tandem MS “on the fly” is sometimes referred to as information dependant acquisition (IDA). Prior to beginning an analysis, a set of experimental protocols for the MS can be defined, such as ion intensity thresholds and the total number of different ions (based on order of intensity) to be queried are entered. Options include whether a dynamic exclusion list is to be developed as ions are queried (e.g. abundant background ions), this determines whether subsequent ion queries with matching m/z are to be repeated, ignored, or to be ignored for a defined period of time (in case of two separate molecules with same m/z and eluting at similar times). Once IDA conditions are entered, analysis starts by collecting a mass spectrum to

determine which ion(s) will be queried by tandem MS. Just prior to tandem MS, a mass spectrum for the precursor ion is re-collected with higher resolution to determine the charge state, which in turn identifies the mass of the precursor ions to be subjected to CAD. After which, the precursor ions are then serially subjected to tandem MS. Once all ions selected by IDA are analyzed, the dynamic exclusion list is updated, and a new cycle of selecting precursor ions for tandem MS from a new mass spectrum begins. Ideally, the time required for the instrument to complete all MS scans directed by IDA should be within an analytes elution period from an LC.

2.6 Sample Preparation

2.6.1 Reduction of Disulfide Bonds by Dithiothreitol

Proteins that have cysteine residues contain thiol functional groups readily oxidize in the presence of oxygen and react with other cysteines to form disulfide bonds (Protein-S-S-Protein).[123] These disulfide bonds provide structural stability for proteins, assisting to maintain the proper conformation to function. Disulfide bonds are problematic for MS analysis, because they limit the extent of protein digestion because the disulfide bond provides increased structural stability which may preclude access to sites along the protein for cleaving, there by complicating interpretation of the protein fragments detected. Traditionally, to convert and maintain these groups in their reduced state, other thiols (RSH) such as cysteine, glutathione, mercaptoethanol, and etc., were added so that interchange takes place according to reactions 1 and 2.

Reaction 1 ***Protein-S-S-Protein + RSH ←→ Protein-SH + R-S-S-Protein***

Reaction 2 ***R-S-S-Protein + RSH ←→ R-S-S-R + Protein-SH***

The second thiol addition, Reaction 2, has RSH added in large excess to minimize the reverse reaction. DL-Dithiothreitol (DTT) is currently the reagent of choice because Reaction 2 proceeds by an intra-molecular path.

2.6.2 Carbamidomethylation: Covalent Modification of Cysteine Residue with Iodoacetamide

Methodologies to modify specific amino acids (e.g. arginine, lysine, tyrosine, etc) are possible; modifications are often performed to study their effects on activities of specific residues, or for mass shifts to be observed by MS.[124] The modifications of cysteines are of particular importance to prevent reformation of disulfide bonds after being broken by a reducing agent. One strategy to prevent the reformation is to form a covalent bond with the thiol groups to "cap" them. This strategy was practiced as early as 1935 by Goddard and Michaelis studying keratin.[125] Generally, thiols treated with base result in thiolate ions (RS^-), which can undergo S_N2 reactions with primary or secondary alkyl halides to give a sulphide (Figure 2.11).[126] Currently the reagent of choice to modify cysteines is iodoacetamide, and reactions are reported to produce ~70% yield after two minutes.[127] Each resulting cysteine residue modified with a carbamidomethyl group results in an increase of 58 (or 57 if cysteine is not

participating in a disulfide bond) Da. Mass shifts as detected by MS are then measured to deduce the number of cysteines that underwent modification.

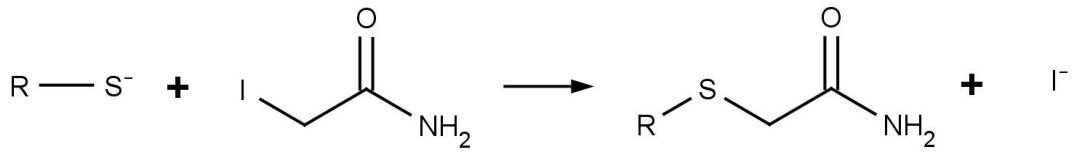

Figure 2.11 Reaction of thiolate with iodoacetamide to yield a product with a carbamidomethyl functional group.

2.6.3 Solid Phase Extraction Embedded Pipette Tips

Electrolytes which commonly accompany protein samples also cause interferences during ionization for both MALDI and ESI. Prior to MS analysis, one method for removing electrolyte from samples is by using pipette tips having immobilized solid phase extraction material on which analytes of interest are retained. For example, when sample is drawn through the tip and passes through reverse phase embedded material, molecules having hydrophobic properties can be retained. Drawing and aspirating aqueous solutions is then performed to wash away electrolytes. Hydrophobic molecules are then recovered from the stationary phase by drawing and aspirating an appropriate solvent that will elute the analyte molecules that had been retained on the tip.

2.6.4 Mechanism of Trypsin for Enzymatic Digestion

Trypsin is a protease that hydrolyses proteins containing either lysine or arginine (except when followed by proline) on the carboxy-side of the amino acid.[128] Trypsin belongs to the serine protease family as the mechanism involves

a catalytic triad of residues, for trypsin they are serine-195, aspartate-102, and histidine-57 (Figure 2.12). The serine first becomes deprotonated by transferring a proton to the histidine, which can stabilize the positive charge with the help of the aspartate. The deprotonated serine performs a neucleophilic attack on the carbonyl carbon of the target residue on the peptide backbone. The negatively charged intermediate is stabilized by an “oxyanion hole” formed by the amide hydrogen atoms of glycine-193 and serine-195. The carbonyl-amide bond is broken when the amide nitrogen extracts a proton from the charged histidine as the negative charge on the oxygen returns to the carbonyl. The intermediate containing the carboxy side of the amino acid is still stabilized in the “oxyanion hole”. Next, a nearby water molecule is deprotonated by the histidine, and the resulting OH^- attacks the carbonyl to form carboxylic acid. As the negative charge on the oxygen is again stabilized by the carbonyl, a proton on the histidine is returned to the serine freeing the peptide and catalytic centre.

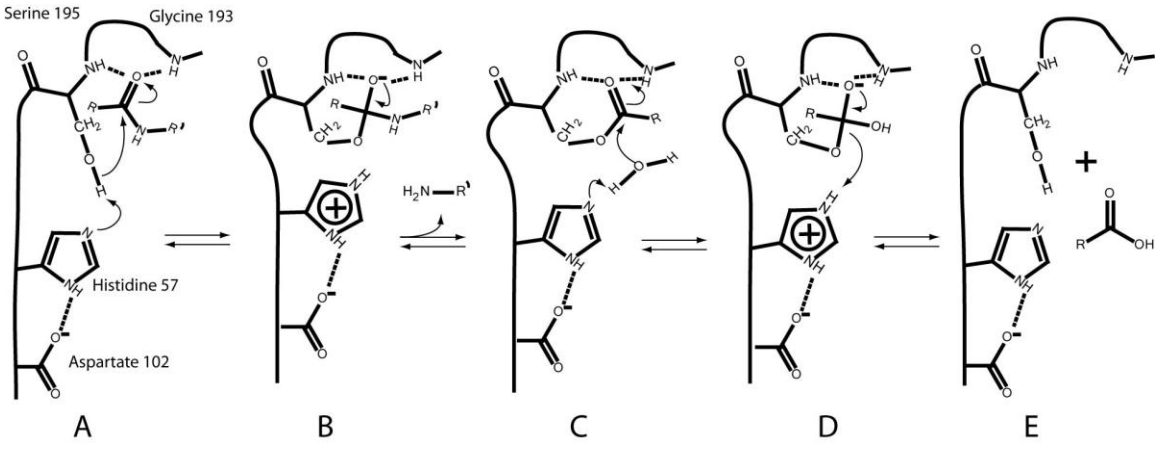

Figure 2.12 Generally accepted mechanism for trypsin proteolysis involving a catalytic triad of residues (serine-195, aspartate-102, and histidine-57).

2.7 Databases of Proteins and Search Engines for MS Data

2.7.1 Swiss-Proteomic (SWISS-PROT) Database

SWISS-PROT is a protein sequence and knowledge database that is valued for its annotations, usage of standardized nomenclature, and minimal redundancy.[129] Essential facts mandatory for each protein entry in this database include the complete amino acid sequence, name, originating taxonomy, and citation information. If available, additional annotation includes items such as: function, enzyme specific information (catalysis, cofactors, pathways, mechanism, etc.), relevant domains and sites, post translational modifications, MW determined by MS, sub-cellular location, tissue specific expression, development specific expression, secondary, tertiary and quaternary structures, splice isoform(s), polymorphism(s), similarity to other proteins, role in biological processes, postulated and theoretical association with diseases, etc. To maintain up-to-date knowledge regarding a protein, information is not only obtained from publications reporting new sequences, but from review articles and experts to maintain annotations periodically. For various SWISS-PROT items, controlled vocabulary for annotations is listed in documents (www.expasy.org/sprot/sp-docu.html). As of June 2010, the database contained 517,802 entries, 20,283 of which for humans.

2.7.2 TagIdent Search Tool

TagIdent is a protein identification tool that is available through the EXPASY World Wide Web Server (www.expasy.org/tools/tagident.html).[129] This tool is a search engine that lists proteins within a user defined search criterion for

pI, molecular weight, or sequence tags (up to six amino acids long). TagIdent searches the Swiss-Prot database for proteins, and can refine results for proteins associated with annotated keywords (i.e. secreted or Homo sapiens).

2.7.3 Mascot

Pappin *et al.* developed a search engine algorithm based on the proteolysis of proteins, named **mo**lecular **w**eight **se**arch (MOWSE), to analyze peptide ion signals (peptide fingerprint) by using only the number and mass of peptides assignable to protein entries within a database to score protein matches.[130] To improve upon this algorithm, Pappin *et al,* continued its development by including calculations for the statistical probability the best-observed match in a data set to an entry in a database is a chance event.[131] The Mascot algorithm is the latest evolution of this search engine software, also commonly described as probability-based MOWSE scoring. The program is capable of interpreting a variety of MS derived data such as, peptide fingerprints, peptide masses with partial sequence strings, or raw un-interpreted tandem MS data

The model for which the statistical analysis used is based on selected physical processes in addition to proteolysis to balance ideal representation of the system, with efficient programming code to complete analyses in a reasonable amount of time. These selected physical processes which users can input parameters for includes: reagent for proteolysis, missed cleavage, instrument platform, modifications, mass accuracy and database/taxonomy. Indicating the reagent (e.g. trypsin or chymotrypsin) for proteolysis and the

number of missed cleavages (incomplete digestion resulting in residues which are expected to be cleaved remaining in a peptide) determines which theoretical peptide sequences and masses Mascot generates. Proteolysis by trypsin allow the user to consider semi-trypsin, the difference being the theoretical peptides generated, one end is either arginine or lysine and the rest of the sequence can terminate anywhere between the next arginine(s) or lysine(s) depending on the number of missed cleavages considered.

Identifying the instrumental platform used for the fragmentation is important for tandem MS, as different instruments can produce fragments differently and this is be reflected in how the data are interpreted. For example, data collected by ESI-quadrupole with CAD type instruments produce fragment ions that are predominantly –b and-y ions and exclude the consideration of –c ions, whereas data from ECD trap types include –c ions and exclude –b ions for analysis (due to the difference in mechanism leading to fragmentation Section 2.4.1.2 and Section 2.4.1.3). Modifications occur due to post translational modifications or derivitization (e.g. addition of carbohydrates, oxidation, carbamidomethylation, etc.) change the molecular weight of peptides, and whether a specific location, functional group, or residue is modified consistently or variably can be considered. Mascot assumes that measurement errors should be treated as being uniform, as a result mass error are not treated as a variable in the probability calculation. This means that for an estimated mass accuracy error of +/- 0.25 Da, a match with an error of 0.2 Da is assumed to be “as good as” one with an error of 0.02 Da. Database and taxonomy selection should best

match the organism the sample originated from, though sometimes entries for one organism may be small and related organisms (which may contain more or different entries) maybe useful.

Resulting probabilities (P) calculated by Mascot are often expressed in scientific notation and are cumbersome to report, instead scores calculated by Equation 2.13 are reported instead.

Equation 2.13 $$Score = -10 * Log_{10}(P)$$

Knowing the size of the database being searched allows Mascot to calculate a significance ($P<0.05$) threshold for the results. For example, a given peptide mass may return 1.5×10^5 theoretical peptide sequences that satisfy a mass tolerance. If the tolerance for a match being a random event is 5%, then a Mascot score of 65 is calculated to be the significance threshold based on Equation 2.13. Individual ion scores for precursor ions depend on the quality of data and mass tolerance collected during tandem MS. Searches of data collected can include variable modifications or proteolysis semi-trypsin, but both increase the number of theoretical entries to consider. For example, searches considering semi-tryptic peptides roughly increase the number of theoretical peptides generated for comparison by ~10 times more, and this raises the probability of random matches and in turn requires a higher score for a match to be significant.[132]

When tandem MS spectra are poor with respect to low signal-to-noise ratios or gaps in fragmentation, ambiguity is introduced and it may not be

possible to reach a significant score for protein identification. When this happens, if the score of the best match is a clear outlier from the distribution of random scores, a second and lower, homology threshold is reported.

In the analyses of large-scale tandem MS data, the calculation of a false-positive rate is complementary to probabilities, as it permits estimation of the likelihood that a match is correct.[133] Mascot evaluates the false positive rate in each database query by conducting a target-decoy entry search in parallel.[134] During such a target-decoy search, every time a peptide sequence from a target database is tested for a match during a search, a random sequence of the same length and composition is also tested, effectively a database of decoy peptides. As a result, the false positive rate for protein identification can be approximated (Equation 2.14), where FP is false positive and TP is true positive, the total matches in a target database are from true positives and false positives, and all matches from a decoy database are false positives.

Equation 2.14 $$\text{False Positive Rate} = FP_{\text{Decoy Database}} / (FP_{\text{Decoy Database}} + TP_{\text{Target Database}})$$

2.7.4 Guidelines to Tandem MS Derived Protein Identification

Increased MS derived bio-molecule identification reported in the literature is a result of the widespread availability of instruments, methods, and easy-to-use software for the collecting and interpreting of large amounts of data.[135] In 2004, editors at the journal Molecular & Cellular Proteomics (MCP) wanted to

address the likelihood that a significant but undefined number of proteins reported as “identified” are likely false positives in manuscripts published in their, as well as other journals. The editors pointed out the availability of different search engines, each using different algorithms, with different filtering criteria as a source for incorrect identification. A principle concern to address was that it was usually possible to match a MS/MS spectrum to a peptide in a database; how can the match be validated? As methods and software continue to be developed, the absence of universally accepted standards led MCP to create guidelines for protein identification based on MS data. Since 2004 there have been more reviews suggesting minimum guidelines when reporting MS derived data for the sequencing and identification of proteins.[136, 137]

Selected MCP Guidelines from 2004 for the Identification of Proteins by tandem MS or peptide fingerprinting include:

1. Sufficient information must be provided regarding which search engine was used, and how assignments were made using that software. Examples include whether mass spectral data had smoothing applied, signal to noise criteria, centroid applied, and whether charge states were calculated.
2. Information regarding sequence coverage observed for each protein should be provided, listing all identified peptides. At the minimum, the total number of peptides belonging to each protein must be explicitly stated in the text or table presented. To compute this number, different

forms (i.e. due to charge states or oxidation) of the same peptide are considered once, and only other unique peptides fragments are added.

3. Protein assignments based on a single-peptide must include the following in the table (see point #2): 1) the sequence used to make the protein assignment, together with the amino acids N- and C- terminal to the observed sequence, 2) the precursor mass and charge observed, not just m/z, 3) the score for the best peptide, in the case of multiple spectra assigned.
4. Proteins identified by peptide mass fingerprint data must contain a list of the number of masses matched to the identified protein, the number of masses not matched, and the sequence coverage observed.

3: COMPARISON OF DATA GENERATED BY CONVENTIONAL AND NANO FLOW LC-ESI FOR TANDEM MS OF PEPTIDES FROM THE AUTO-DIGESTION OF TRYPSIN

3.1 Introduction

Performing tandem MS on peptide ions to promote covalent bond breakage along the peptide's 'backbone' is widely used to deduce the amino acid sequence of the peptide, and when the peptide had a unique sequence from a single protein, to also identify the protein from which the peptide originated. [103, 138] Triple quadrupole mass spectrometers are well suited for CAD of peptides from protein proteolysis by trypsin because the resulting peptide fragments are typically under 3 kDa in mass, and these peptides are often able to accommodate multiple elementary charges. Thus the mass-to-charge of most peptides are within the m/z limit of <~4,000 for most commercial quadrupole MS, and the peptide backbone is also more susceptible to charge-directed cleavage.

ESI is popular for ionizing small volumes of liquid sample for MS, in particular nano-ESI, which is a variation of ESI imposed because of the reduced dimension of the liquid emitter.[97-99] Factors purported to be important regarding improvement in performance for nano-ESI include reduced sample consumption and reduced initial droplet sizes. Again, it is believed that the reduced dimension of this ion source that is directly related to the natural flow rate of this source (i.e. un-pumped), which is in the nano-litre per minute regime that therefore reduces

sample consumption rate leading to longer analysis times (per aliquot of sample solution when used in an infusion mode); this allows a greater number of replicate analyses for improved S/N, or alternatively, it allows the MS software to automatically build an exclusion list in real time that enable queries of low abundant ion signals.

All of the improvements that nano-ESI provides over ESI are typically described for nano-ESI conducted in its "pure" form. "Pure" nano-ESI starts with injecting a fixed volume of liquid sample into a conductive capillary needle (1-10 μm in inner diameter) having a sample flow rate of <100 nL/min towards the tip for the formation of droplets with net charge is due to just charging the capillary (no LC pumping). By coupling an LC column for online separation prior to ionization for ESI (nano-flow LC-ESI), flow rates higher (100-1000 nL) than those normally observed by "pure" nano-ESI are necessarily used because nano-flow LC technology is currently immature. Many investigators have commented that this coupling reduces the improvements of "pure" nano-ESI, as higher flow rates increase initial droplet and spray plume sizes, thus operating more like a conventional ESI.

Based on an equation by Faramawy *et al.* in their work whose objective was to provide improved characterization of ionization efficiencies for ESI at different flow rates, a predictor of analyte ion to molecule ratio was developed (Equation 3.1).[99] The denominator is calculated from the molar concentration and sample loop volume, and no assumptions are made about preformed ionized analyte in solution. The calculation reflects upon all potential losses including:

irreversible binding to LC columns, incomplete pre-ionization in solution, ejection of ions from droplets, throughput by the sampling orifice (orifice for conventional ESI is charged while grounded for nano-ESI), and ion throughput by the ion optics of the MS.

Equation 3.1 $\sum$***(Signal Ion Count) / Analyte Introduced = Analyte Ion to Molecule Ratio***

This chapter presents results that are discussed with respect to the differences between two different LC-ESI set-ups that operate at micro-flow or nano-flow. To streamline this comparison, the analytes studied are simply the peptides from the auto-digestion of trypsin (Table 3.1). Trypsin is a common protease used for the enzymatic digestion of proteins, and these peptides are often utilized as internal mass standards in LC-MS assays. Moreover, the hydrophobicity and pI of these peptides are used as background information with regard to the discussion of the results. Two methods of describing the hydrophobicity for a peptide include the aliphatic index value and grand average hydrophobicity (GRAVY) value. The aliphatic index value is the sum of alanine, valine, isoleucine, and leucine amino acid mole fractions multiplied by coefficients representing volume occupied relative to alanine.[139] A GRAVY value is the average value describing the hydrophobicity of every amino acid within a peptide. Peptides that are hydrophobic have higher aliphatic index values, or GRAVY values up to a maximum of 4.5. Hydrophilic peptides have smaller aliphatic indexes or GRAVY values as negative as –4.5. The pI of a peptide

varies with its composition of number and types of charged functional groups present. The pI indicates the pH at which a molecule (e.g. amino acid, peptide, or protein) contains no net charge, and is used to predict whether a peptide in a solution of known pH is charged (positive or negative) or neutral.

Table 3.1 Peptides from the Auto-Digestion of Bovine Trypsin and their Theoretical GRAVY, Aliphatic and pI Values

Fragment	Theoretical Peptide Sequence	Mass /Da	Aliphatic	GRAVY	pI
1-2	MK	277.38			
3-23	TFIFLALLGAAVAFPVDDDDK	2238.57	120.95	0.92	3.77
24-66	IVGGYTCGANTVPYQVSLNSGYHFC GGSLINSQWVVSAAHCYK	4553.12	76.98	0.23	7.94
67-72	SGIQVR	658.76	113.33	-0.08	9.47
73-92	LGEDNINVVEGNEQFISASK	2163.32	92.00	-0.43	4.00
93-112	SIVHPSYNSNTLNNDIMLIK	2273.59	112.00	-0.17	6.46
113-114	LK	259.35			
115-122	SAASLNSR	804.86	73.75	-0.38	9.47
123-148	VASISLPTSCASAGTQCLISGWGNTK	2552.9	82.69	0.40	8.03
149-159	SSGTSYPDVLK	1153.25	61.82	-0.53	5.55
160-162	CLK	362.49			
163-172	APILSDSSCK	1020.17	88.00	0.12	5.87
173-193	SAYPGQITSNMFCAGYLEGGK	2194.46	46.67	-0.16	5.72
194-209	DSCQGDSGGPVVCSGK	1495.6	36.25	-0.41	4.21
210-223	LQGIVSWGSGCAQK	1433.65	83.57	0.16	8.22
224-225	NK	260.29			
226-231	PGVYTK	663.77	48.33	-0.62	9.01
232-240	VCNYVSWIK	111.324	107.78	0.56	8.17
241-246	QTIASN	632.67	81.67	-0.37	5.52

3.2 Methods

3.2.1 Reagents and Solutions:

A 1.2 mL solution containing 2.1 x 10^{-6} M trypsin (MW 23.8 kDa) was prepared with 100 mM ammonium-bicarbonate in water. The trypsin (Product number T1426, Sigma Aldrich, St. Louis, MO) from bovine pancreas was treated

with L-(tosylamido-2-phenyl) ethyl chloromethyl ketone (TPCK) to inactivate any remaining activity of chymotrypsin present as an impurity during the purification process. This trypsin solution was allowed to auto-digest by incubating at 37 ^{0}C for 16 hours (overnight), and was stopped by adding 20 μL of 1.5 M formic acid. Individual aliquots of 20 μL for analysis were then prepared and stored in a -20^{0}C freezer.

Solvents A (25 mM formic acid in 95:5 v/v water methanol) and B (25 mM formic acid in methanol) was used as mobile phases for liquid chromatography. Formic acid was purchased from Anachemica (Montreal, Quebec) and water (Barnstead Nanopure Infinity system, Dubuque, IA, USA) used in this chapter was distilled, deionized (~18 MΩ when dispensed) and filtered through a 0.22 μm membrane. A polypropylene glycol solution, provided by the manufacturer of the MS (Product PPG-300, Applied Biosystems Sciex, Foster City, CA), was used for mass axis calibration.

3.2.2 Micro-Flow Liquid Chromatography

MassLynx V4.0 was software used to interface the user with the LC pump (model CapLC, Waters Corp., Milford, MA, USA). A 150 μm (i.d.) x 15 cm stainless steel capillary column (Waters Symmetry, Cambridge, MA) was packed with 5 μm polystyrene particles having 150 Angstrom pores with the surface modified with C18 alkyl chains. The column was kept at ambient room temperature. Samples were injected manually into a 5 μL sample loop. Chromatography was performed at 13 μL/min with initial conditions at 75 % A

and 25 % B for 60 minutes of isocratic elution. After which, the mobile phase was switched to 15% A and 85% B for 15 minutes as a wash step. Mobile phase conditions were then changed back to 75% A and 25% B for a minimum of 45 minutes prior to commencing with the next analysis. The LC and MS instruments communicate with the same computer, but their respective software do not operate concurrently. After sample injection, the software that was used to interface with the LC was terminated, and software for the MS was activated for manual data collection. Instrumental and computer communication challenges that occurred resulted in computer crashes and delays in data acquisition ranging from 5 to 15 minutes after sample injection.

3.2.3 Hybrid Q-q-LIT-MS Instrument Conditions for Micro-flow ESI

Analyst 1.4.2, updated May 2008, was the software used to control, collect, and interpret data (interface) from hybrid triple quadrupole linear ion trap MS (Q-Trap 2000, Applied Biosystems, Foster City, CA,USA).

To maintain mass accuracy of the instrument, external mass calibration were performed monthly by infusing a solution of polypropylene-glycol provided by the manufacturer (un-diluted) through a syringe at 7 μL/min. During calibration, the m/z of the ions and their intensities in Table 3.2 are observed to construct a mass calibration curve, and to validate the sensitivity.

Table 3.2. m/z of Polypropylene Glycol ions for mass calibration

m/z	Intensity to pass (cps)
59.1	3.0×10^7
175.1	2.7×10^7
616.5	5.0×10^6
906.7	1.5×10^7
1254.9	3.0×10^6
1545.1	5.0×10^5

Eluent from the LC, or sample in a syringe pump, is directed to an ESI ion source (Turbo Ion Spray, Applied Biosystems Sciex, Foster City, CA) and is analyzed with a hybrid triple quadrupole-LIT-MS. At the orifice of the MS is an electrode (faceplate) having a small hole (d = 3 mm) in it, and is charged with +250 V to help repel large charged droplets from entering the MS. Instrument conditions at the ion source interface include: +5500V potential at ESI tip, 20 curtain gas (units set in the software), 10 ion source gas for pneumatic flow (units set in the software), 225 ^{o}C pneumatic gas temperature, 45 V de-clustering potential, and High CAD (for collisional cooling prior to trapping). Mass scans for precursor selection were performed by filling (allowing ions to accumulate) the LIT for 250 ms (page 50), followed by m/z readout at a rate of 1,000 Da/s. Settings for IDA included: tandem MS for top 5 ion intensities above 2×10^5 cps, exclude ions after 3 occurrences for 300 s (dynamic exclusion list), and dynamic collision energy (Equation 3.2) that depend on the charge state and mass of precursor ion (Table 3.3). Mass scans for ion fragments were performed by filling the LIT for 350 ms and m/z readout at a rate of 1,000 Da/s.

Equation 3.2 **Collision Energy (V) = Slope x (m/z) + Intercept**

Table 3.3 Slope and intercept parameters used for Equation 3.2 when applying collision energy to precursor ions for fragmentation

Maximum Allowed Collision Energy: 100.000 V

Charge State	Slope	Intercept
Unknown	0.044	4
1+	0.063	9
2+	0.05	2
3+	0.045	4

3.2.4 Nano flow LC-ESI Set-up

The use of commercially available components for the assembly of an online nano flow LC-ESI ionsource has been presented by Gatlin *et al.*[140] Reducing unions and tees were selected from the Cheminert and Nanobore for purchase from Vici Valco, Houston, TX to assemble the apparatus depicted in Figure 3.1. All fused silica-capillaries were purchased from New Objective, Cambridge, MA.

Poly ether ether ketone (PEEK) tubing (1.6 mm o.d., 0.13 mm i.d.) connects the outlet of an injection port with a 5 μL sample loop (Figure 3.1). The PEEK tubing is connected by a reducing union to fused-silica capillary tubing (360 μm o.d., 20 μm i.d.). Connected to the fused-silica tubing is a fused-silica capillary column (360 μm o.d., 75 μm i.d., 18 cm length, product label “PicoFrit”) which was packed with 3 μm polystyrene particles having 100 A pores with the surface modified with C18 alkyl chains (Reprosil, Dr. Maisch, Ammerbuch, Germany). The capillary column is connected by a micro bore liquid handling tee to a charging electrode, and a fused-silica needle (360 μm o.d., 20 μm i.d., 10

mm length, “picotip”) for nano-ESI. The fused-silica needle is threaded through a tee (0.5 mm bore) with sleeves (1.6 mm o.d. 1 mm i.d.) allowing gas flow provided to exit as a sheath gas (A). All other connections between tubing and tees were fitted with appropriate sleeves to prevent leaks.

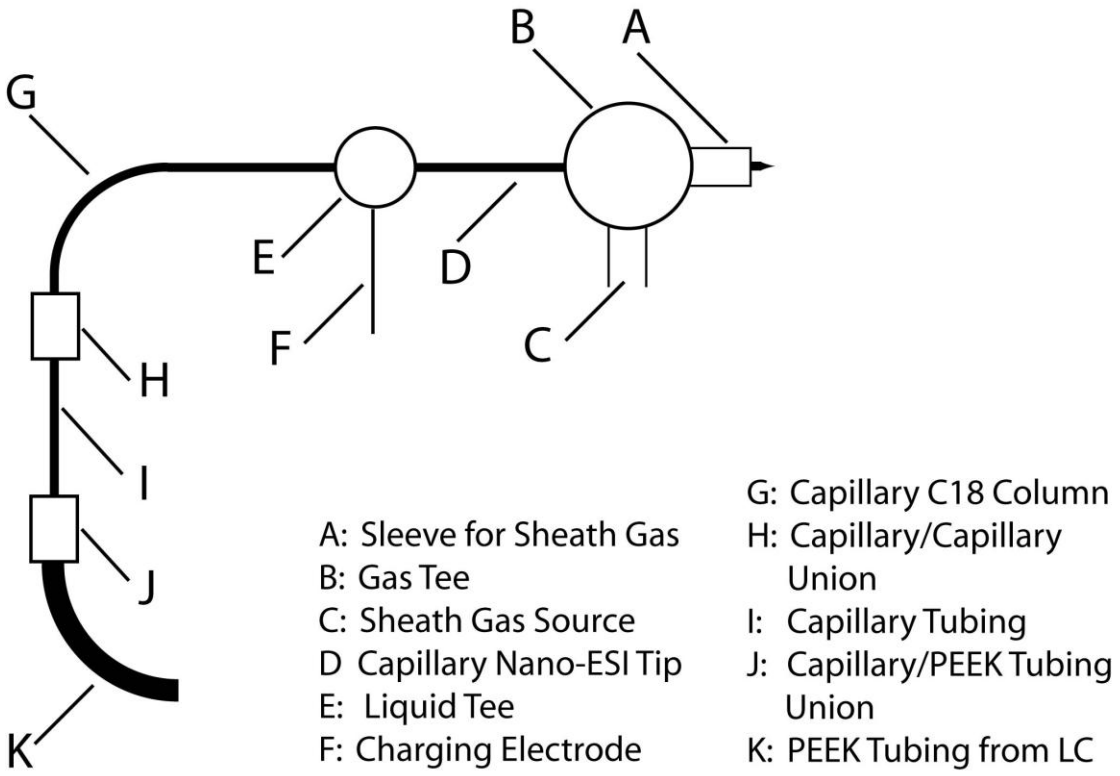

Figure 3.1 Cartoon schematic of an online nano flow LC-ESI ion source. A fused silica capillary that had its tip tapered to 10 μm i.d. (D) is threaded though a gas tee (B), and is connected to a liquid tee (E). Inside the cavity of the liquid tee, an electrode (F) is exposed to mobile phase to charge the solution. Beads coated with C18 was packed into a capillary column (G), and connected to the LC pump by additional capillary tubing (I) and unions (H and J).

The same LC pump system, 5 μL sample loop, and manual injection port was used. Chromatography was performed at 0.5 μL/min with initial conditions at 75 % A and 25 % B for 300 minutes of isocratic elution. After which, the mobile

phase was instantly switched to 15% A and 85% B for 120 minutes as a wash step. Mobile phase conditions are then changed back to 75% A and 25% B for a minimum of 12 hours (over night) before the next analysis.

3.2.5 Q-q-LIT-MS Instrument Conditions for Nano-ESI

The same mass spectrometer as Section 3.2.3 was used for data acquisition. Due to the reduced flow rate filling the same tubing within the LC system, no peptide signals were ever observed within 60 minutes of injection, and so data acquisition was also delayed by this time. The faceplate to the MS was changed to allow it to be electrically grounded when a nano- flow ESI source is in use, and to compensate, a higher curtain gas flow was used to repel large droplets from entering the MS. Instrument conditions at the ionsource interface include: 1,500V potential charging the ESI tip, 45 curtain gas, 3 ion source gas for sheath gas flow, ambient gas temperature, 45 V de-clustering potential, and High CAD. Settings for IDA were the same as those during micro-flow LC-ESI except the top 8 ion signals (instead of top 5) above 3 x 10^5 cps are selected for tandem MS.

3.2.6 Mascot Search

Prior to uploading tandem MS data to MASCOT, the following settings were selected in the Analyst 1.4.2 software: query molecules calculated to be between 600 and 3,500 Da, group spectra within +/- 0.05 m/z of precursor, a maximum of 5 cycles between grouping (1 cycle consists of a precursor scan and all of the software selected high resolution mass scans, and tandem MS scans),

remove peaks with <1% of highest, centroid spectra, and reject spectra if less than 8 peaks.

Data containing tandem MS spectra submitted to MASCOT (Matrix Science, London, UK) were initially searched with the following parameters: (i) SWISS-Prot. Database, (ii) other mammalian taxonomy, (iii) proteolysis by trypsin, (iv) max. one missed cleavage, (v) peptide mass tolerance of 750 ppm, (vi) and MS/MS mass tolerance of 0.75 Da. Apart from scores for identification, the mass difference between ion signals observed and theoretical masses are reported by Mascot. To determine the mass accuracy of the instrument, results for ion signals leading to confident identification from both (micro and nano flow) LC-ESI are reviewed for the single largest error reported. The tandem MS spectrum of the largest peptide by molecular weight was also reviewed for the largest error in fragment assignment. After which, Mascot searches for reporting were performed with the same search criteria, except, peptide mass tolerance of 500 ppm, MS/MS mass tolerance of 0.5 Da (entered as different units due to the programming of the software), and the addition of a separate search with proteolysis by semi-trypsin

3.2.7 Sample Calculation of Analyte Ion to Molecule Ratios Using Equation 3.1

An estimated analyte ion to molecule ratios (Equation 3.1) was determined for the peptide SIVHPSYNSNTLNNDIMLIK, formed from auto-digestion of trypsin, without any missed-cleavages. The signal ion count depends on the intensity and duration of all the observed charge states over the single most

abundant time period observed. The identity and time point for the ion signal to be integrated are obtained from the results of a Mascot analysis, and were verified manually. The signal from an ion was a range of m/z, and was determined by two (blue) lines from the peak intensity passing through 50% and intersects the x-axis of the precursor scan (Figure 3.2A). After which, an ion chromatogram (Figure 3.2B) for each peptide charge state is generated based on the range of ion m/z obtained in the previous sentence. To obtain the signal ion count, the duration for which to integrate the ion chromatograms is also determined by the region between two (blue) lines from the peak intensity passing through 50% and intersects the x-axis. The total signal ion count for the peptide is the sum of the area (4.5×10^9 counts for peptide with 3+ charge state and 5.0×10^{10} counts for peptide with 2+ charge state) for all charge states observed.

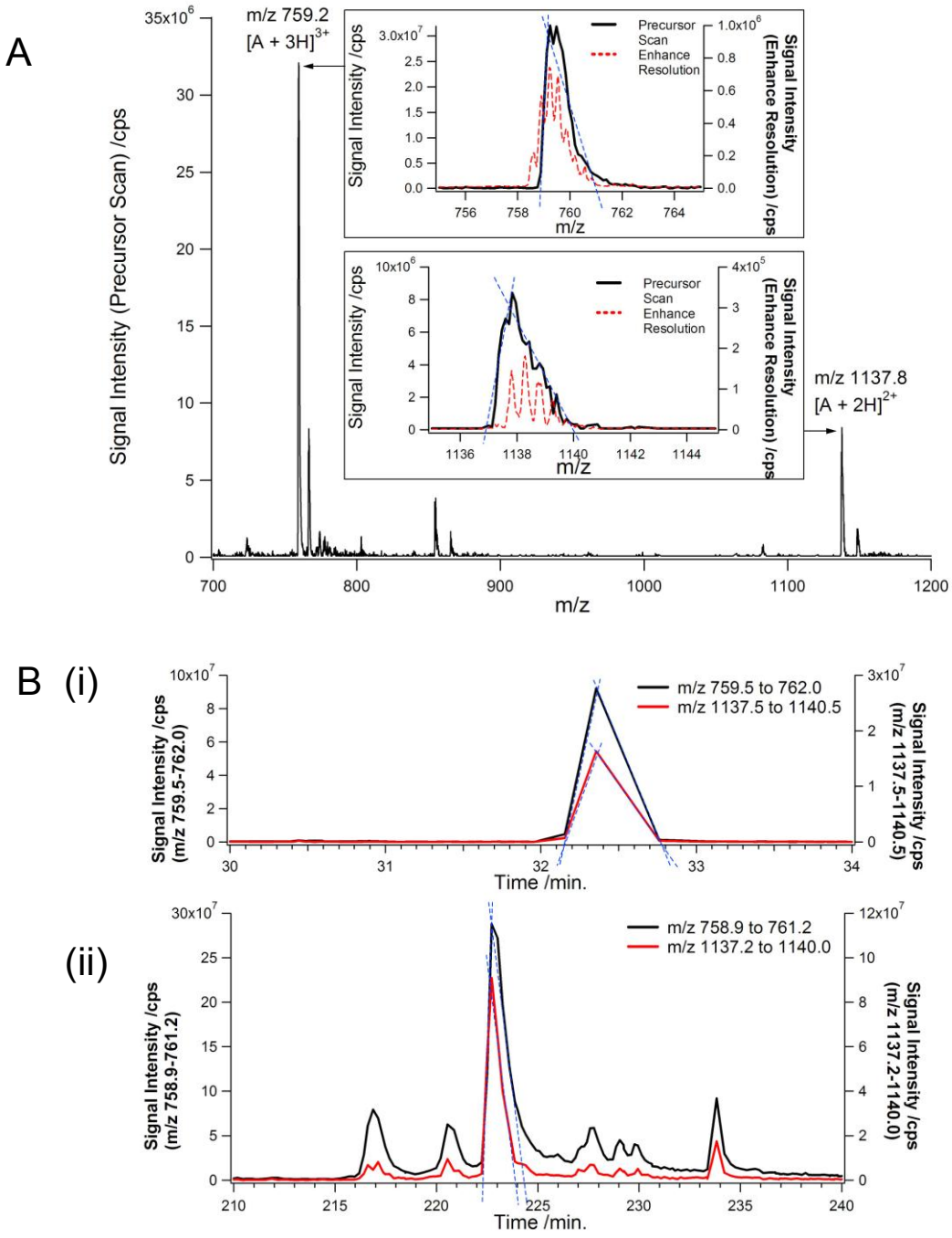

Figure 3.2 **MS derived data for the calculation of analyte ion to molecule ratio for the peptide SIVHPSYNSNTLNNDIMLIK. (A) Mass spectrum from nano-flow LC-ESI indicating signals for peptide ions with 2+ or 3+ charge state, insets are of m/z region recollected at higher resolution to determine distribution of isotopolouges. The m/z regions between the gray dotted lines were integrated to generate ion chromatograms (B) Ion chromatograms from the data generated by (i) micro-flow LC-ESI and (ii) nano-flow-LC ESI.**

3.3 Results and Discussion

Micro-flow LC-ESI generated chromatograms followed by mass spectral detection of peptides eluted yield variations in elution times by 3 - 4%, and signal intensity variation of 24 – 41% for the three most abundant peaks in the total ion chromatograms containing peptides (Figure 3.3A). In comparison, reproducibility of analyses by the nano-flow LC-ESI used was poor, as evident by every analysis resulting in spectra in which the trypsin peptides that could be identified had different elution times and signal intensities (Figure 3.3B). Problems with the nano-flow occurred in every analysis performed. One particularly persistent problem was a period of time where elution from the column would stop, re-start, and persist non-reproducibly. It was found this stoppage would rarely be corrected by itself. A resumption of flow was found to be aided by wiping the tip with lint-free tissue wetted with ethanol. After ~45 minutes of stoppage, if the flow did not resume the nano-ESI capillary was immediately replaced, as particulates were often found to be inside the bore at the tip. Not all injections led to detection of peptide signals; and in general, it was difficult to obtain replicate analyses where the same ion signals were present. For conditions just described, no more than 2 satisfactory replicate analyses of the same trypsin solution were collected after several attempts. Single analyses of peptides from the auto-digestion of trypsin are discussed for comparing data by micro-flow (Figure 3.3 Aiii) or nano-flow LC-ESI (Figure 3.3 Biv).

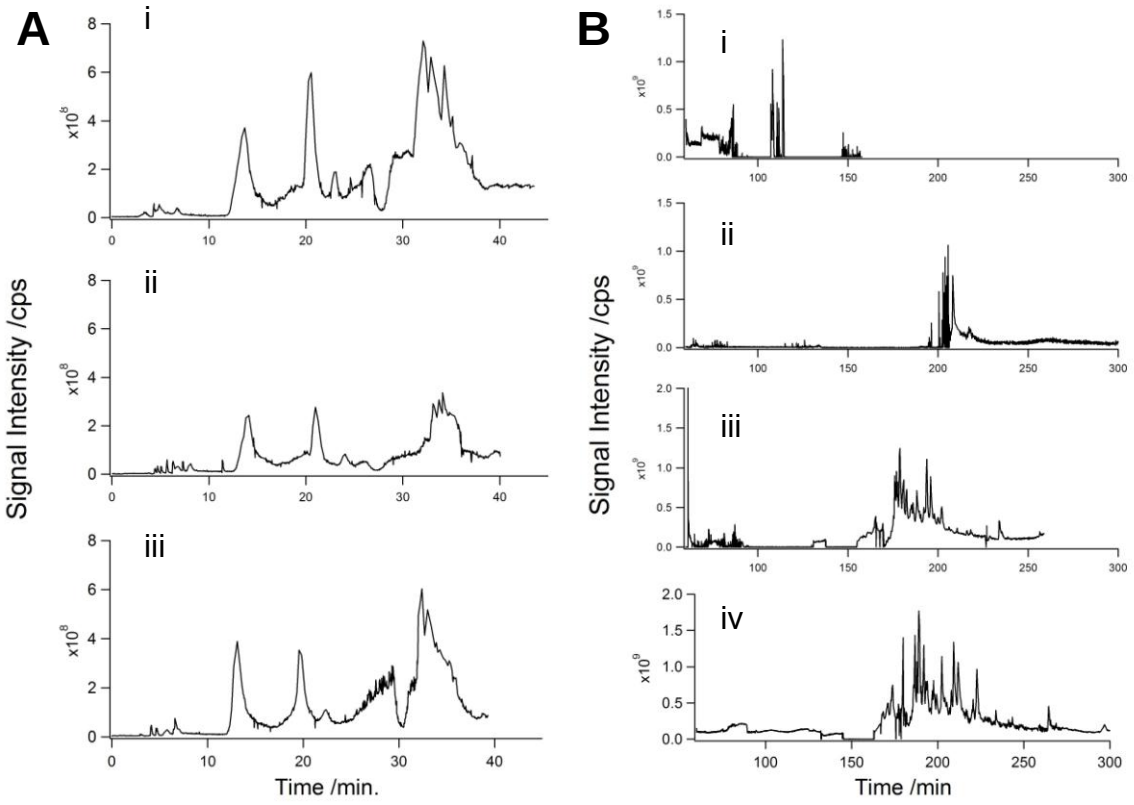

Figure 3.3 Total ion chromatograms from the analyses of auto-digested trypsin by either micro-flow (A) or nano-flow (B) LC-ESI.

The mass spectral data generated using either micro- or nano-flow LC-ESI did lead to identification of trypsin by Mascot with scores above the minimum to be significant (Mascot scores > 29 indicate at least 95% confidence in identification). Results tabulated in (Table 3.4) were from searches performed considering semi-trypsin proteolysis, as it lead to the greatest number of peptides identified to be the result of trypsin auto-digestion. .The data generated from LC-ESI at nano-flow resulted in identification with a greater Mascot score (1222, due to combination of all replicate and unique peptide ion signals having statistically

significant tandem MS spectra) compared to data from micro-flow LC-ESI (175), yet both were above the minimum score of 29 (uniquely calculated for each set of tandem MS data to be queried, Section 2.7.3) for identification having statistical significance. All peptides identified by LC-ESI performed at micro-flow (5 different peptides in total) were also identified when performed at nano-flow with an additional 15 peptides, and identifying more peptides helped improve sequence coverage from 35% (micro-flow) to 49% (nano-flow). Two factors that contributed to this improvement in scoring were, a greater number of tandem MS spectra from precursor ions indicating unique identity collected (different from each other or replicates), and individual peptide ions often scored higher.

Table 3.4 Peptides identified by tandem MS when ionized by a micro-flow or nano- flow LC-ESI. ND indicates the peptide was not identified.

	Micro-flow		Nano-flow	
	Total Mascot Score: 175 Secuence Coverage: 35%		Total Mascot Score: 1222 Secuence Coverage: 49%	
Sequence Identified	Mascot Score	Analyte ion/ Molecule ratio	Mascot Score	Analyte ion/ Molecule ratio
Tryptic Peptides				
LGEDNINVVEGNEQFISASK	63	3.2×10^{-2}	70	9.2×10^{-1}
SIVHPSYNSNTLNNDIMLIK	60	2.4×10^{-2}	56	8.5×10^{-1}
SAASLNSR	43	9.1×10^{-3}	43	1.2×10^{-1}
SSGTSYPDVLK	51	4.8×10^{-2}	71	6.4×10^{-1}
APILSDSSCK	ND	(Not Detected)	29	
Semi-Tryptic Peptides				
LGEDNINVVEGNEQF	44	1.1×10^{-2}	63	1.6×10^{-1}
VVEGNEQFISASK	ND		30	
SIVHPSYNSNTLNNDIM	ND		38	
SIVHPSYNSNTLNN	ND		43	
SIVHPSYNSNTLN	ND		36	
SIVHPSYNSNTL	ND		41	
SIVHPSY	ND		33	
SNTLNNDIMLIK	ND		40	
NSNTLNNDIMLIK	ND		44	
YNSNTLNNDIMLIK	ND		39	
SSGTSYPDVL	ND		35	
APILSDSSCKSAYPGQITSN	ND		29	
SAYPGQITSN	ND		31	
SAYPGQITSNMF	ND		33	
NKPGVYTK	ND		29	

Nano-flow LC-ESI produced ion intensities that were greater for all 5 peptides that were identified by both LC-ESI (because only these 5 peptides were observed by both LC-ESI analyses, they are the focus of subsequent comparisons). The intensity for the most abundant charge state of the peptides at their peak elution during nano-flow LC-ESI were found to be greater than elution during micro-flow LC-ESI by 4.2 to 6.9 times. In addition to absolute ion intensities, ion to molecule ratios for analytes were found to be greater as well,

ranging from 13 to 36 times greater. Table 3.4 contains peptides identified that directly contradict assumptions made when calculating analyte ion to molecule ratios, as some peptides (i.e. SIVHPSYNSNTLNNDIMLIK to SIVHPSY) have undergone degradation (possible reasons include hydrolysis, thermal degradation, and post source decay (PSD) in the collision cell of the MS), calculations were still made for comparing ratios from both LC-ESI set-ups because the same digest sample was analyzed.

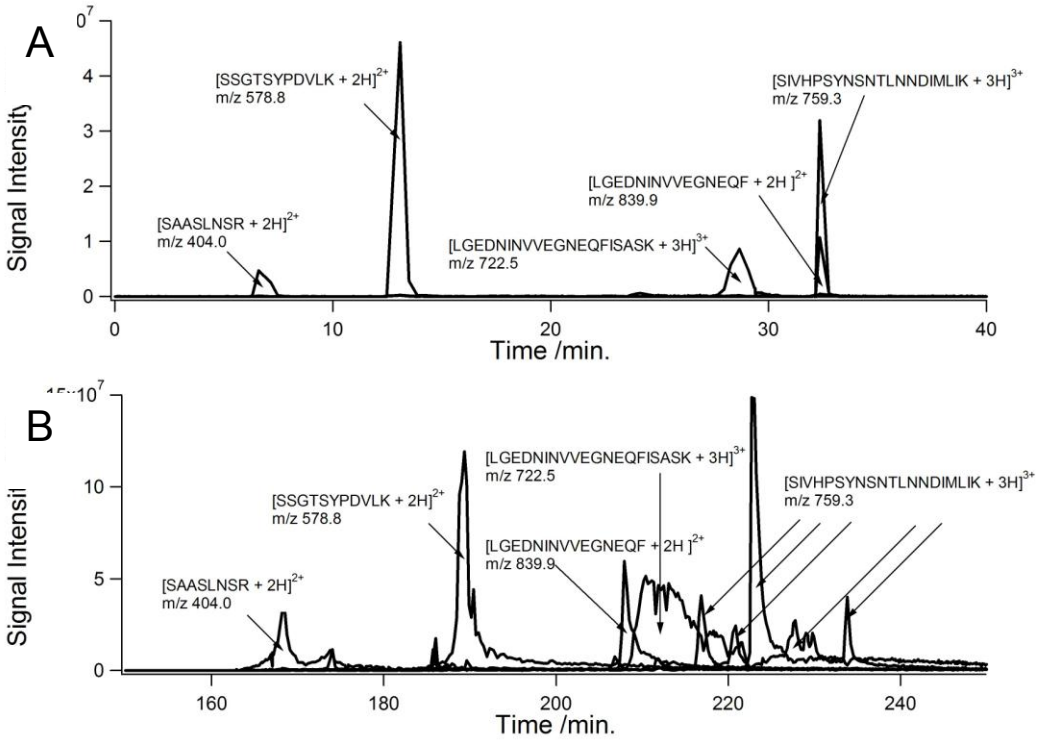

Figure 3.4 Ion chromatograms of most abundant charge state for peptides identified in both (A) micro-flow and (B) nano-flow LC-ESI.

To help determine whether semi-tryptic peptides identified were fragments of tryptic peptides having undergone post source decay (PSD), the ion chromatograms of related peptides were examined together. For example, from

the analysis by nano-flow, the semi-tryptic peptide LGEDNINVVEGNEQF and its related intact tryptic peptide LGEDNINVVEGNEQFISASK have ion currents not related to each other as indicated by the peak intensities at different times (i.e. 29 versus 32 min.), thus not related to each other due to PSD (Figure 3.4A).

To accommodate a narrower capillary column and tip for nano-flow LC-ESI, the flow rate was reduced to 0.5 μL/min. However, the duration of an entire analysis increased. For example, the elution time for peak intensity of the peptide SIVHPSYNSNTLNNDIMLIK required 32 minutes by micro-flow LC and was 222 minutes with nano-flow LC (Figure 3.4). The slower flow rate also increased the elution periods as described by the full width half max (FWHM). Differences in FWHM ranging from 1 to 6 minutes were observed among the 5 common peptides used for comparison, the largest of which was for the peptide LGEDNINVVEGNEQFISASK.

The abundance of a precursor ion is a factor that determines how useful the subsequent tandem MS generated data will be with respect to submission to a database for possible identification. For example, the peptide SAYPGQITSNMF was selected for tandem MS in both LC-ESI analyses, yet the signal intensity of this peptide ion during a precursor scan was 15 times greater at nano-flow as compared to micro-flow (Figure 3.5). This discrepancy was magnified by the output of the Mascot query where the peptide was identified in data generated by nano-flow but not using the micro-flow data. When reviewing the tandem MS spectra by both LC-ESI by Mascot for this peptide ion, the spectra by nano-flow contained a greater number of peaks with greater intensity.

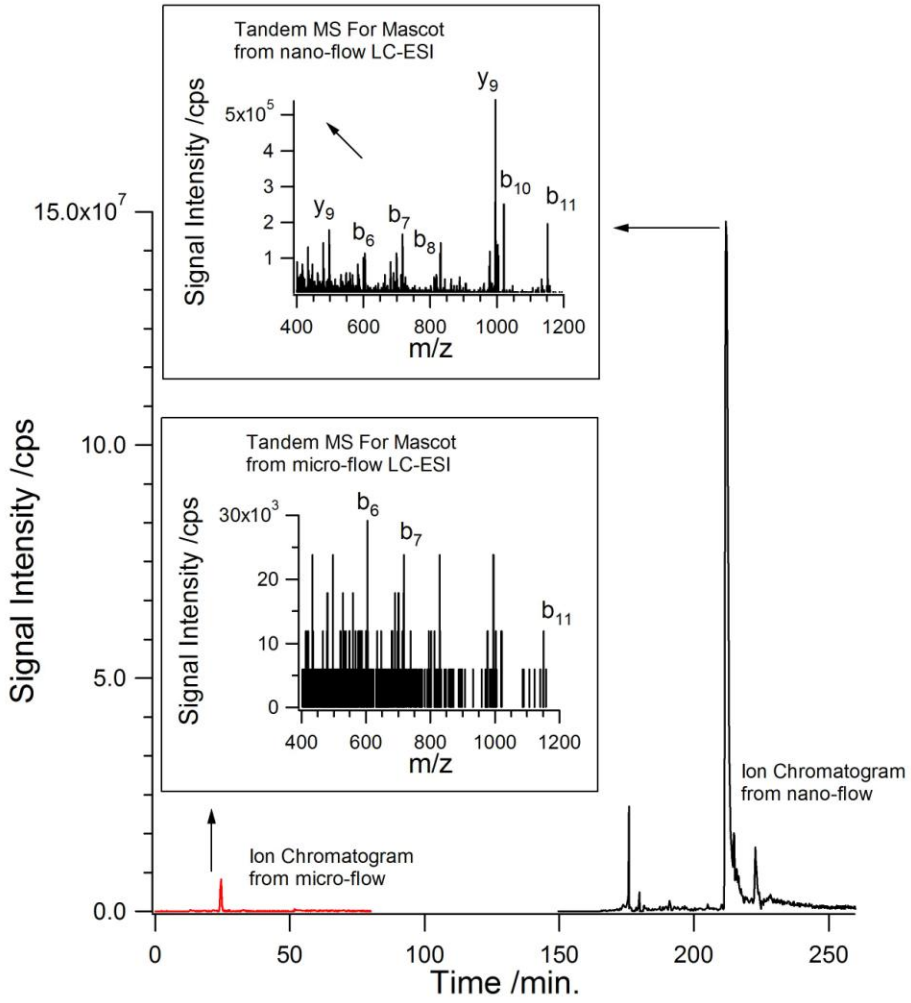

Figure 3.5 Ion Chromatograms and tandem MS spectra for m/z 658.4 by micro- and nano-flow LC-ESI from 2 different experiments plotted together, as indicated by the different colours for the chromatograms. Tandem MS of m/z 658.4 generated from nano-flow LC-ESI had a greater number of ion signals identified, leading to a higher statistical confidence the peptide identity as SAYPGQITSNMF.

The peptide SSGTSYPDVLK (m/z 577.8) illustrates severe tailing that occurs during chromatography by this nano-flow LC (Figure 3.6), possibly attributable to the mobile phase travelling from PEEK tubing (1.3 x 10^2 μm i.d.) to fused-silica tubing (20 μm i.d.). There was a large difference in inner radius of

these two tubes which may have lead to turbulent flow through the union. Tandem MS performed during the peak elution (t = 189 minutes) of this peptide and 31 minutes later (t = 220 minutes) both led to the confident identification of trypsin with the same peptide sequence, archived ion chromatogram data indicates a continuous elution during this period, hence tailing. The acquisition of a tandem MS spectra for m/z 577.8 at t = 220 minutes highlights the utility of a software generated dynamic exclusion list, as its abundance was not in the group of top 8 most abundant peaks at the time of precursor selection (right inset).

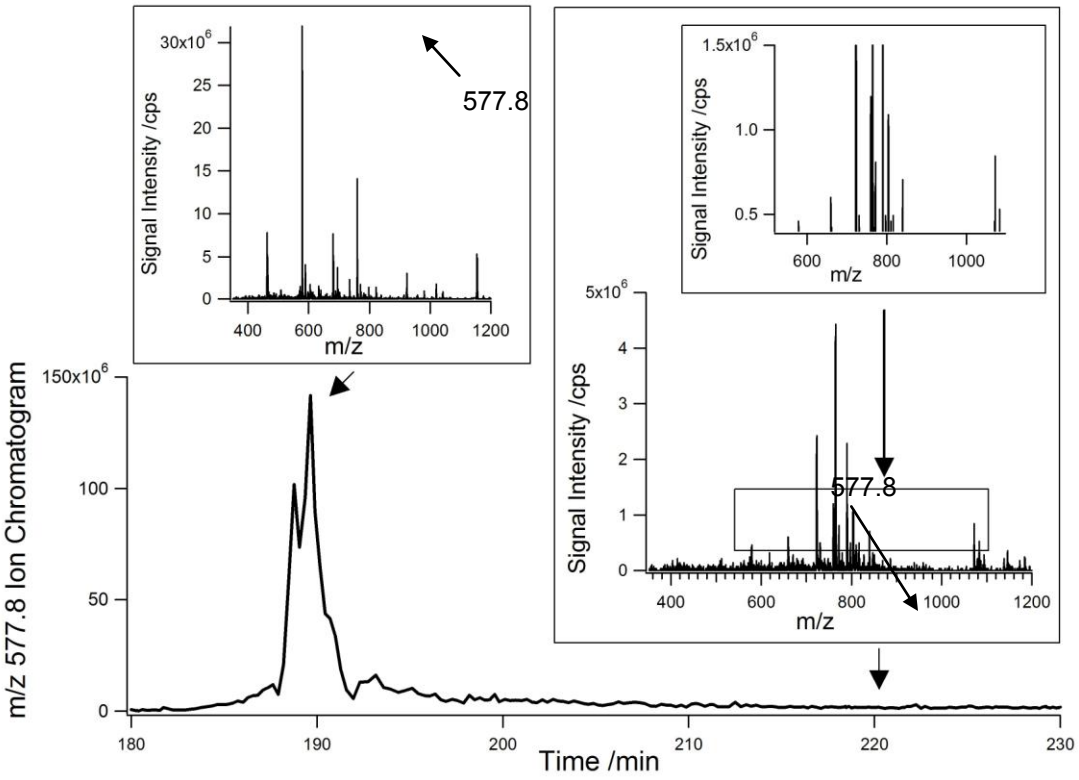

Figure 3.6 MS data for m/z 577.8 peptide ion for SSGTSYPDVLK. Mass spectrum for precursor selection of m/z 577.8 leading to confident identification of trypsin are shown for time points at 189 and 220 minutes in ion chromatogram. At 189 min., m/z 577.8 is at the peak of elution and is the most abundant ion detected. At t = 220, m/z 577.8 is not the top 8 abundant ions, but was selected due to the dynamic exclusion list.

The 5 common peptides identified by either LC-ESI were compared against the additional 15 peptides identified by using the nano-flow LC, for their average aliphatic index, GRAVY, and pI values. No significant differences in their averages were observed between the two sets. This result indicates that the detection of the additional 15 peptides by nano-flow LC is not due to a preference for analyte hydrophobicity or pI by reducing the size of droplets that leave the ESI tip. For comparison, in another study for peptide characteristics and successful tandem MS by either MALDI or ESI, the researchers found ionization by MALDI tend to favour basic peptides, while ESI tended to favoured aliphatic peptides.[141] As the number of peptide analyses in this study was much lower compared to the previous mentioned MALDI-ESI study (667 peptides) broad characteristic trends maybe revealed by studying a greater number of peptides for comparison.

Table 3.5 Peptides observed by both micro and nano-flow LC-ESI or nano-flow LC-ESI only. Included are the peptides Aliphatic, GRAVY and pI values on their average for either group

observe by both	Aliphatic	GRAVY	pI	
LGEDNINVVEGNEQFISASK	92.00	-0.425	4.00	
SIVHPSYNSNTLNNDIMLIK	112.00	-0.170	6.46	
SAASLNSR	73.75	-0.375	9.47	
SSGTSYPDVLK	61.82	-0.527	5.55	
LGEDNINVVEGNEQF	90.67	-0.620	3.50	
	86.05	-0.423	5.80	average
	19.16	0.170	2.37	stdev
Observed by nano				
APILSDSSCK	88.00	0.120	5.87	
NKPGVYTK	36.25	-1.387	9.70	
VVEGNEQFISASK	82.31	-0.185	4.53	
SIVHPSYNSNTLNNDIM	85.88	-0.459	5.06	
SIVHPSYNSNTLNN	76.43	-0.764	6.46	
SIVHPSYNSNTLN	49	-0.63	5.24	
SIVHPSYNSNTL	89.17	-0.308	6.46	
SIVHPSY	97.14	0.143	6.46	
SNTLNNDIMLIK	130	-0.075	5.55	
NSNTLNNDIMLIK	120	-0.338	5.84	
YNSNTLNNDIMLIK	111.43	-0.407	5.83	
SSGTSYPDVL	68	-0.19	3.8	
APILSDSSCKSAYPGQITSN	68.5	-0.255	5.87	
SAYPGQITSN	49	-0.63	5.24	
SAYPGQITSNMF	40.83	-0.133	5.24	
	79.46	-0.367	5.81	average
	27.68	0.252	0.78	stdev

3.4 Summary

Data generated from nano-flow (0.5 μL/min) and micro-flow (13 μL/min) LC-ESI-MS/MS were compared to learn any difference in performance due to differences in dimension and flow-rates, beyond simply using Mascot to confidently ($P = < 0.05$) identify the protein trypsin from solutions of it after allowing for partial auto-digestion. Analyses by micro-flow LC-ESI were

completed within ~35 minutes with elution times having reproducibility of 3 – 4 % RSD. The nano-flow LC-ESI operated under non-ideal conditions of the LC system (e.g. flow rate used, large difference in tubing diameter, and problematic instrument communication) resulting in analyses that were by comparison long in duration at ~240 minutes and with poor reproducibility. However, nano-flow LC-ESI generated data resulting in greater confidence in identification of trypsin auto-digested peptides (Mascot score of 1222 compared to 175) and greater sequence coverage of trypsin observed (49% compared to 35%, respectively).

Ion signals observed by nano-flow LC-ESI were greater in abundance and duration (having elution peaks with FWHM ranging from 1-6 minutes), in combination with a software generated exclusion list this helped to select low abundant ions to be queried, which helps identify as many unique peptides as possible. Using Equation 3.1, estimates of analyte ion to molecule ratios for peptides identified by both LC-ESI sources were higher for nano-flow LC-ESI, and this is believed to be an important factor that leads to improved peptide signal intensity, which in turn lead to increased sequence coverage and confidence in identification of trypsin. From a comparison of peptide ions observed by either micro-flow or nano-flow, there were no obvious trends for preferences based on hydrophobicity or pI values. Applications of this study are directly related to chapter 4, as the apparatus set-up and methodology for Mascot data base searching used in Chapter 4 was developed using solutions of partially auto-digested trypsin.

4: IDENTIFYING PROTEINS FROM HUMAN EPITHELIAL LUNG CELL (A549) DOSE RESPONSE ASSAYS BY MASS SPECTROMETRY

Several members of the Agnes group have conducted studies of the differential expressions of bio-molecules by human lung cells upon dosage with mimics of atmospheric particulate matter. Methodology adopted in this thesis to semi-quantitatively monitor bio-molecule expressions were primarily developed by Alice Kardjaputri and Xin-Zhou (Siry) Hu. My research objective was to develop LC tandem-MS based methodologies to be used in the identification of differentially expressed proteins in the supernatants of human epithelial lung cells (A549).

4.1 Introduction

A link between exposure to PM and adverse health effects is known.[4, 6, 8] However, the role of each compound found in PM and their effects on human health remain unclear. Compounds known to be present in PM in the atmosphere have been studied, and though thousands of different compounds have been identified, generalized categories that can be used to bin these compounds include metals, salts, carbonaceous, organic, and compounds of biological origin.[9, 15-18] Given this context, a long-term objective of MS based proteomics would be to learn about the identity and quantity of ideally all proteins present and expressed by lung tissues, as a function of the particulate stimulus

using an experimental platform that mimics *in vivo* conditions as closely as possible. MALDI-ToF or ESI-Q-q-LIT are MS platforms that each offer different operating characteristics, and these instruments allow for hypothesis driven and de novo proteomic studies because these techniques have been demonstrated to enable detection of compounds having a wide range of physical and chemical properties.

4.1.1 Selected Compounds for Dose Response Experiments

In this work, lung cells were not dosed with mimics of atmospheric PM, rather, selected compounds that have been identified in PM were introduced in aqueous state. The compounds used were TNF-α, India ink (mostly carbon black solubalized by detergent, exact composition is proprietary knowledge), lipopolysaccharide (LPS), $Ni(NO_3)_2$, ethylenediamine-tetraacetic acid (EDTA), and selected combinations thereof. The pro-inflammatory cytokine TNF-α was chosen for use as the positive control as it is a known stimulator of cellular response as evidenced by the activation of signaling pathways that lead to the synthesis of other cytokines.[142] India ink was used to mimic the carbon soot core of PM from the incomplete combustion of diesel fuels.[21, 143] LPS is an endotoxic molecule from the outer membrane wall of Gram-negative bacteria found throughout our environment, and $Ni(NO_3)_2$ salt was chosen to represent a soluble metal ion capable of redox chemistry. Metals are ubiquitous in ambient PM, including metal like nickel due to the wearing of engines in soot from the high temperature combustion of fossil fuels. The concentrations of LPS, $Ni(NO_3)_2$ and carbon in the starting solutions used for dose response experiment are related to

those found in literature.[18, 143] EDTA is typically not found in the atmosphere, but was added in some samples to chelate Ni^{2+}. Its inclusion in the study was based on A. Kardjaputri's observations of particles containing carbon and Ni^{2+}. It was observed by including EDTA, the expression of a pro-inflammation molecule (ICAM-1) increased compared to samples dosed with similar particles excluding EDTA.[144]

4.2 Methodologies

The water used in this chapter was distilled, deionized, filtered through a 0.22 μm membrane, and had a resistance >18 MΩ when dispensed (Barnstead Nanopure Infinity system, Dubuque, IA, USA).

4.2.1 Obtaining Samples of Supernatants from Cells Cultured and Dosed

4.2.1.1 Solutions for Cell Culturing

Nutrient Mixture F-12 Ham (N3220, Sigma Aldrich, St. Louis, MO) in combination with fetal bovine serum was used to make growth medium for cell culturing. The nutrient mixture was first prepared 0.9 L at a time according to instructions provided by the supplier. It was then sterilized by filtering in a biological safety cabinet (Forma Scientific Inc. # 14567-386, Marietta, OH, USA) through a sterile cellulose acetate membrane filter (mean pore size of 0.2 μm) into a sterilized container. Following nutrient mixture, 100 mL of fetal bovine serum was filtered and added to make the final volume of the growth medium 1L.

Serum free medium (SFM) is sterilized growth medium without fetal bovine serum. In the experiments performed, SFM was used with its contents

(nutrient mixture) dissolved in 90% of the recommended volume of water. The remaining 10% of liquid was delivered when cell cultures were dosed with an aliquot of a SFM solution. Solutions of phosphate buffered saline (PBS) were used to wash cells to prevent injury due to osmosis. This solution was prepared according to the instructions of the supplier (Oxoid Ltd., Basingstoke, Hampshire, England) and then either sterilized by autoclaving at 120 ^{o}C for 20 minutes or filtered through a cellulose acetate membrane (mean pore size of 0.2 μm) into sterilized bottles prior to use.

4.2.1.2 Cell-line Passages

Vials of alveolar type II epithelial cells from the A549 cell line (passage #89, American Type Culture Collection, Manassas, VA, USA) in ~1mL of MEM containing 10% dimethyl sulfoxide were retrieved from a liquid nitrogen freezer. Standard techniques to culture cells were performed in biological safety cabinets. Polystyrene tissue culturing plates were used for supports (100 x 20 mm, BD Falcon #353003, BD Biosciences, Franklin Lakes, NJ). Culture plates were stored in an incubator with an environment of 5 % carbon dioxide, 100 % humidity, at 37.5 ^{o}C. Cell cultures were grown to 80-95% confluence prior to further passaging or dose response experiments.

4.2.1.3 Introducing Serum-Free Medium for Experiments

Fetal bovine serum (FBS) contains proteins, growth factors, hormones, and minerals for growth. To minimize interference by these compounds during MS analyses, all experiments were conducted in serum-free medium (SFM), as

these cultures no longer need these nutrients for growth. Once cell cultures were 75-95% confluent, prior to dosing, growth medium (which contains FBS) was removed by aspiration. The cultures were then washed by delivering and removing 3 mL aliquots of PBS five times. After washing, 2.5 mL of SFM was added to the culture, and were ready to be dosed for experiments.

4.2.1.4 Solutions Prepared for Dosing Cell Cultures

Reagent compounds TNF-α, LPS, $Ni(NO_3)_2$, EDTA were purchased from Sigma Aldrich (St. Louis, MO.) Carbon black was assumed to be the principle material in India ink, yet it is recognized that it is a proprietary mixtures that contains other compounds, including surfactants, glycol, and sodium chloride (Speedball, product #3338, Statesville, NC). To dose cell cultures (and dilute the SFM to the suggested final concentration), 275 μL of a dosing solution, containing for example one or more of the compounds mentioned in the first sentence of this paragraph, was added to 2.5 mL of SFM. The final concentration of these compounds of interest in the aliquots delivered to the cell cultures was 4.9×10^2 ng/mL TNF-α, 1.0 μg/mL LPS, 1.0×10^2 nM $Ni(NO_3)_2$, 3.3×10^{-3} % v/v ink (6.5 μg/mL of non-volatile material by gravimetric analysis), and 3.0×10^2 nM EDTA. Note that any one solution delivered to the cultures may have contained one, or a combination of these compounds. The cultures were then stored in an incubator (37 ^oC, 100% humidity, 5% CO_2) for 23-25 hours (24 hour experiment), though in select experiments the incubation period was 72

hours. After incubation, the SFM conditioned by cells (supernatant) was collected and stored at -80 ^{o}C until sample analysis at a later time.

4.2.1.5 Viability of Cell Cultures After Dosing

To investigate the viability of cells after the experiment, 2 mL of PBS was delivered to the culture plates in addition to 100 μL of Trypan blue (Sigma Aldrich, St. Louis, MO, USA). A Trypan blue assay is an exclusion technique to investigate the integrity of cell membranes. A Trypan blue dye molecule is a negatively charged ion that cannot permeate through intact lipid bi-layer cell membranes. Compromised (no longer viable) cell membranes will allow Trypan blue to permeate and increase in concentration by interacting with positively charged residues in proteins, dying the cell blue. With the use of an optical microscope, supernatants from cultures were observed to have <5 % of cells dyed blue (Figure 4.1).

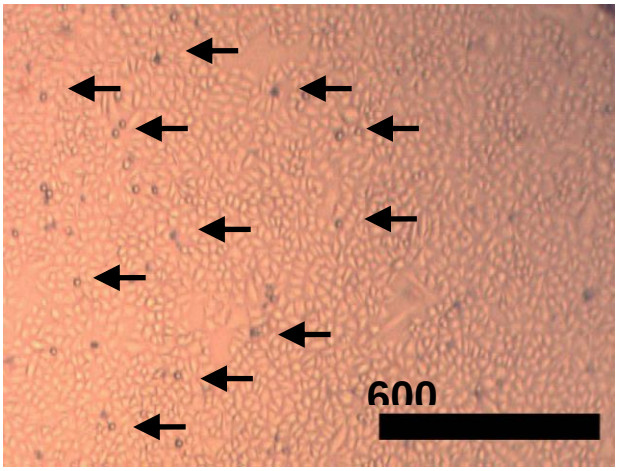

Figure 4.1 Image of cell culture dosed with 100 nM $Ni(NO3)_2$ and 6.5 μg/mL India ink. Cells that are blue (arrows indicate examples) have compromised cell walls allowing trypan blue dye to enter.

4.2.2 Sample Preparation of Supernatants from Cell Cultures

4.2.2.1 Denaturing, Reduction and Carbamidomethylation of Proteins

The amount of supernatant necessary to perform an analysis was dictated by the technique used (i.e. MALDI-ToF or LC-ESI-MS). The following protocol describes the amount of reagents used per 100 μL of supernatant for analysis. Polypropylene vial(s) containing supernatant were bathed in a hot water bath at 95 ^{0}C for 2 minutes to denature proteins. After cooling to room temperature, 2 μL of 1.0 M ammonium bicarbonate was added to raise the pH, so that the reactions of DTT and iodoacetamide that were added next were carried out at favourable pH conditions. To the sample, 1 μL of 1.0 M DTT was added and incubated at 37^{o}C for 20 minutes. After which, 2 μL of 1.0 M iodoacetamide was added to the sample to alkylate cysteines present (Sections 2.6.2), and then left at room temperature in the dark for 15 minutes. To stop the reaction alkylation reaction, 1 μL of 10 % trifluoroacetic acid (TFA) is added. The aliquot volume of supernatants analyzed using MALDI-ToF-MS was 400 μL, offline LC-MALDI-ToF-MS used 150 μL, and LC-ESI-MS used 800 μL.

4.2.2.2 Solid Phase Extraction for desalting and concentrating Sample for MS analysis

Supernatants contain electrolytes that contribute to decreased ionization performances due to adducts or cluster formations. Prior to MS analysis, samples were desalted and concentrated by solid phase extraction using ZipTips. ZipTips are commercially available pipette tips that have a bed of C18 coated beads (e.g. the stationary phase of a liquid-liquid extraction cartridge)

trapped in a small volume at the end of disposable micro-pipette tips (C18 ZipTips, Millipore, Billerica, MA,). To minimize sample preparation error due to abundant proteins having higher affinity for the stationary phase, the entire denatured, reduced and modified sample was divided into 33 μL portions that were then processed individually. Drawing and aspirating out sample is a method used to increase the effective surface area in a short period of time, allowing hydrophobic molecules to become retained on the C18 coated beads. Drawing and aspirating out aqueous solutions of 0.1% TFA removed salt ions, while maintaining a similar pH environment within the tip. The retained molecules were then eluted by drawing and aspirating back 10 μL from a 60 μL aliquot of 60:40 acetonitrile (ACN) 0.1% TFA solution. Using the same ZipTip, hydrophobic molecules drawn and retained from the other fractions of a supernatant sample were also subjected to the same steps to optimize the removal of electrolytes, and all elutions of retained proteins in the ZipTips were combined into a single 60 μL aliquot. After the last aliquot of the supernatant sample had been eluted from the ZipTip, a final 10 μL of fresh 60:40 ACN containing 0.1% TFA was applied to the ZipTip (e.g. drawn and aspirated), followed by 10 μL of fresh ACN. These last 2 washings of the stationary phase on the ZipTip were added to the previously collected 60 μL of eluent. The 80 μL of eluent was then concentrated by vacuum desiccation until <2 μL remained.

4.2.2.3 Tryptic Digestion of Supernatant for ESI-MS

For micro and nano-flow LC-ESI-MS analyses, 10 μL of 100 mM ammonium bicarbonate was added to sample (~2 μL) after vacuum desiccation. To the resulting solution, 1 μL of 0.50 mg/mL L-(tosylamido-2-phenyl) ethyl chloromethyl ketone (TPCK) treated (to inactivate any chymotrypsin impurities present) bovine trypsin (Product number T1426, Sigma Aldrich, St. Louis, MO) in 25 mM acetic acid was added. The sample containing trypsin was then left in an incubator at 37 ^{o}C for 16-20 hours (overnight). Just prior to MS analysis, the samples were then acidified with 0.3 μL of 1.3 M formic acid to stop Trypsin activity and protonate basic functional groups.

4.2.3 Instrumentation

4.2.3.1 Software and Instruments

Analyst software version 1.4.2 updated to May 2008 was the software used to control, collect, and interpret data from hybrid Q-q-LIT-MS (model Q-Trap 2000, Applied Biosystems, Foster City, CA,USA). MassLynx software V4.0 was used to interface with both the LC pump and MALDI-ToF-MS (models: CapLC and M@LDI LR, Waters Corp., Milford, MA, USA). Igor Pro 4.05 (WaveMetrics, Inc., Lake Oswego, OR, USA) is the plotting software package that used to prepare all figures of MS derived data.

4.2.3.2 MALDI-ToF-MS

A two-layer method developed by Dai *et al.* was adapted to spot sample with matrix for MALDI-MS analysis.[145, 146] In the two-layer method, two different

solutions of sinapinic acid concentrations are prepared and applied, referred to as the thin or thick layer solution. The thin layer solution is 6 mg/mL of sinapinic acid dissolved in 60:40 methanol and ACN. The thick layer solution is 10 mg/mL of sinapinic acid dissolved in 50:50 methanol and 0.1% TFA solution. A sample well on a stainless steel analysis plate was first spotted with 2 μL of the thin layer solution that, when dry, a layer of sinapinic acid crystals remain. Next, a 2 μL aliquot of sample solution and 4 μL of the thick layer solution was mixed, and the entire mixture was spotted onto the well containing the thin layer of sinapinic acid.

Data for the m/z region between 2 - 20 x 10^3 was collected with the instrument set to linear mode (settings: 1200 V pulse voltage, 15000 V source voltage, 1850 V MCP detector, 1000 m/z matrix suppression (m/z cut-off by instrument entering flight tube), 700 ns time lag focusing delay, 1 ns sampling time, 337 nm laser wavelength, 10 Hz laser firing rate, and 10 laser shots were fired per spectrum). The laser shot flux was attenuated by a neutral density filter controlled by the instrument by selecting 'medium' under course laser control, and an aperture setting in the 75 - 100 % open was selected under fine laser control. After data collection, an averaged spectrum for the sample was generated by combining 15 spectra (each from 10 laser shots), box car smoothed over +/- 3 data channels, and a baseline subtraction (with a 15^{th} order polynomial leaving 10% of data points collected below the curve).

Calibration of the instrument was performed with external mass standards at the beginning of any given data acquisition. A solution containing 3.6 x 10^{-5} M

myoglobin (MW 16,950.5 Da), 3.0 x 10^{-5} M cytochrome c (MW 12,361.0 Da), and 4.0 x 10^{-6} M adrenocorticotropic (ATC) hormone fragment 18-39 (MW 2465.7 Da) was prepared and spotted onto an analysis plate (Figure 2.2). The detection of ion signals of calibration compounds that contain a single or double charge allow the m/z region between 2465 and 16965 to be calibrated. A polynomial order of 2 and residuals of +/- 10 was allowed for the construction of a calibration curve when operating in linear mode, as per the MALDI-ToF-MS instruction manual.

4.2.3.3 Offline Micro-flow LC-MALDI-ToF-MS

150 μL of supernatant from a positive control incubated for 72 hours was desalted, concentrated (Sections 4.2.2), diluted with 10 μL of 0.1 % TFA, and that solution was injected into an LC system for separation at mico-flow (described in chapter 3). Solvent A consisted of 0.1% v/v TFA in water, and solvent B was 0.1% v/v TFA in ACN. Mobile phase conditions starting at 0% B changed linearly using conditions in

Table 4.1, at a flow rate of 15 μL/min. After injection of sample, a 15 minute delay was used before manually collecting fractions by touching the droplet forming at the tip of the tubing exiting the column into a micro centrifuge tube. Starting from the 15th minute mark, 30 second fractions (~7.5 μL) were collected, totalling 96 fractions.

Table 4.1 Table of conditions for offline-LC separation of supernatants

% B	time /min	Flow rate: 15 μL/min
0	5	Solvent A: 0.1% TFA in Water
27	15	Solvent B: 0.1% TFA in Acetonitrile
34	25	
40	30	
45	33	
95	43	

A 2 μL portion of each fraction was spotted (two-layer method) onto an analysis plate for MALDI-ToF-MS. Spectra of the fractions were collected using automation functions of the instruments software. Instrument settings (linear mode, pulse voltage, source voltage, time lag focusing delay, etc.) for spectra acquisition of m/z 2 - 20 x 10^3 were the same as described previously. The averaged spectrum generated was from 12 individual spectra (each an average of 20 laser shots and spectra) collected from 12 random locations chosen by the instrument within each well. The mass accuracy (+/- 5 m/z) for this experiment was determined by calculating the largest difference observed between an ion signal for a standard and its theoretical m/z.

4.2.3.4 Micro Flow LC-ESI-MS

Solution A (25 mM formic acid in 95:5 water and methanol) and B (25 mM formic acid in methanol), were prepared using formic acid purchased from Anachemica (Montreal, Quebec, Canada). The LC system, sample loop, and column described in section 3.2.2 was used. The mobile phase flow rate throughout the chromatography was set at 13 μL/min. Gradient elutions were performed with the mobile phase composition starting at 10% B, and that linearly increased to 75% B over 65 minutes. The mobile phase was kept at 75% B for

an additional 20 minutes, followed by a linear change back to 10% over 20 minutes, and the column was then equilibrated for a minimum of 90 minutes prior to the next injection. Conditions for ESI and instrument were the same as those described in section 3.2.3.

4.2.3.5 Nano-Flow LC-ESI-MS

The solvents and apparatus described in section 3.2.4 were used. The flow rate throughout the chromatography was 0.3 μL/min. Gradient elutions were performed with the mobile phase composition starting at 10% B, and increased using a linear ramp to 75% over a period of 300 minutes. The mobile phase was kept at 75% for an additional 120 minutes, followed by a linear change back to 10% over 120 minutes. The column was then equilibrated for a minimum of 12 hours. After injection of sample, a 180 minute delay was used prior to initiating data acquisition on the MS, which was operated using the same conditions as described in section 3.2.5.

4.2.4 Database Searching

4.2.4.1 TagIdent Searches of MALDI-ToF-MS Derived Data

The Swiss-Prot database was searched by TagIdent accessed through http://expasy.org/tools/tagident.html (accessed March 2008). The only search parameter used was organism classification (Homo sapiens), molecular weight (kDa), and range (%) of molecular weight. Molecular weights queried were based on average ion m/z observed less 1 for the mass of a proton for charging,

and the range was calculated from three times the standard deviation of their average m/z observed.

4.2.4.2 SWISS-PROT Database Search by MASCOT of Peptide Fingerprint and Tandem MS data

A list of peptide masses (from the digestion of fractions which have mass spectrums containing ion peaks at m/z 8356 or 8570) was compiled manually from observing ion signals differentially observed compared to auto-digested trypsin. The list of masses were entered manually along with search parameters: proteolysis by trypsin with up to 1 missed cleavage, peptide tolerance of 200 ppm, and the m/z of the most abundant signal from pooled fractions was entered for the protein mass.

The post-acquisition treatment of the tandem-MS data, including integration of spectra and elimination of ion signal below a threshold was performed as described previously in chapter 3. Each data file generated from analyses with micro-flow LC-ESI was first searched in the SWISS-PROT database to identify trypsin present, the same data was then resubmitted for a search with a maximum mass tolerance determined from the peptides identified for trypsin, as described in Section 3.2.6. Due to instrumental drift, the mass accuracy of higher m/z (> ~1000) was decreasing, and was not the same in all analyses. For some analyses, by accommodating the poorer mass accuracy at the higher m/z (> 750 ppm or 0.7 Da for MS/MS as determined by trypsin peptides) tandem MS data for lower m/z ions were becoming insignificant, though their precursor data (m/z and charge state) and fragmentation patterns

were similar to data in other analyses leading to identification. As a result, Nano-flow LC-ESI-MS experimental data were searched with the same mass tolerance parameters. Parameters used include: other mammalian taxonomy (based on mammals available for searching), precursor mass tolerance of 750 ppm, MS/MS mass tolerance of 0.7 Da, allow up to 1 missed cleavage, and ESI-Trap as the instrument were selected. Each data file was first searched to identify bovine trypsin present, this helped to indicate whether molecules from the digestion was being analyzed. Afterwards the exact same data (including peptides for trypsin) was resubmitted to search for proteins with the same parameters except fixed modification carbamidomethylation of cysteines, and human taxonomy.

4.3 Results and Discussion

4.3.1 MALDI-ToF-MS for Analysis of Supernatants

MALDI-ToF-MS was employed to acquire spectra of the compounds in the A549 cell culture supernatants. Supernatants from positive (dosed with TNF-α) and negative (no dosing) controls were first analyzed and compared with respect to the m/z of ions observed and their relative signal intensity (Figure 4.2). In prior work from previous group members (A. Kardjaputri and X. Hu) using MALDI-ToF-MS to characterize ion signals observed in A549 cell culture supernatants, an ion signal in all samples was consistently observed at m/z 8570 +/- 3. This ion signal was interpreted as being from a homeostatic (maintained to be constant or stable) protein.[144, 147] Subsequently, the ion signal at m/z 8570 was used in processing of the data as a reference to try to improve the quantitation of other ion signals observed. X. Hu further characterized this by introducing an internal

standard in duplicate analyses and found ratios of ion signals for m/z 8570 and the internal standard comparable. I continued this method of processing other ion signals using the ion signal at m/z 8570 as a reference, using Equation 4.1.

Equation 4.1 **Ratio of analyte to reference = $(S_{analyte} - S_b)/ (S_{m/z\ 8570} - S_b)$**

For Equation 4.1, S_b is obtained by averaging 100 data points from a region at m/z ~ 19,900 were there was no discernable ion signals. From the analysis of three separate positive and negative control experiments, ion signals that were observed in at least two of three replicate analyses, having signal intensity greater than 5 times S_b, and have FWHM greater than 5 m/z were tabulated. Averages of analyte to reference ratios were subjected to a Student's t-test, and ion signals found to be differentially expressed with confidence >90% were noted (Table 4.2). In generating the table of ion signal ratios, it was assumed that all in signals were because of proteins, either intact or portions thereof, that had been secreted by A549 cells. Upon a review of analyte to reference ratios, one ion of interest was m/z 8354 +/- 4 as it was found to be one of the most abundant ion signal in the positive control, and was calculated to be one of the most significantly differentially expressed.

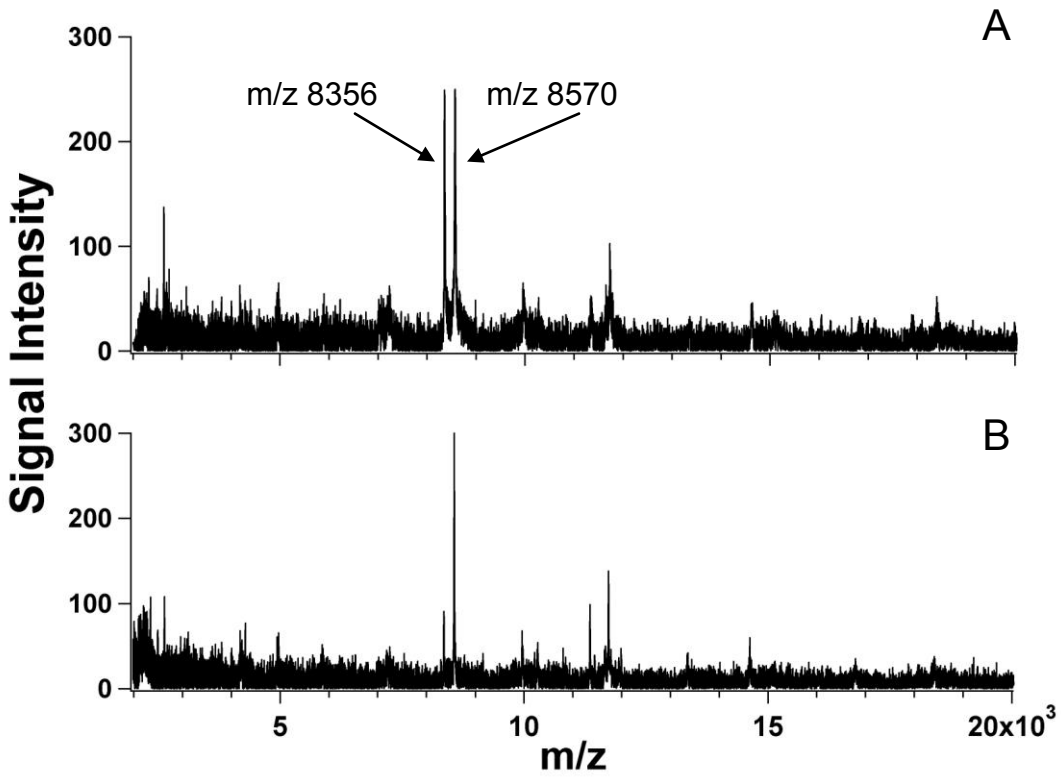

Figure 4.2 Representative MALDI-ToF-MS spectra of supernatant from A) a positive control (dosed with TNF-α) B) a negative (no dosing) control for m/z 2 - 20 x 10^3.

Table 4.2 Ion signals from m/z 2 - 20x10^3 observed in supernatants from positive and negative controls. Average m/z observed, control experiment with greater expression, and confidence (>90%) for differential expression are tabulated. For ions with a confidence value for differential expression determined, it was observed in both controls.

#	Control Experiment Observed in	Confidence	Average m/z Observed	Standard Deviation	Supported by TagIdent & Mascot Search	Supported by Mass Shifts due to Modification of Cysteines
1	Pos.	95	2497	7		
2	Both		2628	3		
3	Both		2699	7		
4	Both		2733	4		
5	Neg.		2777	5		
6	Pos.	92	3094	3		
7	Pos.	91	3805	3		
8	Both		4177	4		
9	Both		4283	5		
10	Both		4939	3		
11	Both		4981	4		
12	Pos.	90	5868	3		
13	Pos.		6632	3		
14	Pos.		7014	3		
15	Pos.	98	7244	3		
16	Pos.		8192	3		
17	Pos.		8284	3		
18	Pos.	98	8354	4	CXCL-5	Yes
19	Pos.	97	8394	3		
20	Both		8570	3	Ubiquitin	Yes
21	Both		8667	2		
22	Neg.		9830	3		
23	Both		9957	3		
24	Pos.		10214	3		
25	Pos.	96	10273	9		
26	Both		10847	6		
27	Pos.		11071	3		
28	Pos.		11307	3		
29	Both		11346	6		
30	Pos.		11515	6		
31	Neg.		11600	3	Thioredoxin	
32	Pos.		11652	3		
33	Both		11727	4	Beta-2-microglobulin	
34	Both		11941	3		
35	Both		14630	3		
36	Both		14961	8		
37	Pos.		15088	3		
38	Pos.		15139	3		
39	Pos.		15841	3		
40	Pos.		16798	3		
41	Pos.		17885	3	Peptidyl-prolyl cis-trans isomerase A	
42	Pos.		18064	5		
43	Pos.		18406	3		
44	Pos.		18619	8		

4.3.1.1 TagIdent and Carbamidomethylation Chemistry for Tentative Identification of Ion Signals by MALDI-ToF-MS

To help identify the ion signals observed by MALDI-ToF-MS, TagIdent was used to search the m/z of ion signals observed for matches to molecular weights of intact proteins or portions thereof annotated in the SWISS-PROT database. Search results by TagIdent returned multiple possible protein identities due to the low mass resolution of the data. For example, the search results for m/z 8356 and 8570 are tabulated (Table 4.3). Subsequently, the list of possible proteins in Table 4.3 was queried for the number of cysteine residues known to be present per protein using the Swiss-Prot database via the website http://expasy.org/ (accessed March 2010).

Table 4.3 TagIdent Search results for the query of m/z 8353 (+/- 12) and 8569 (+/- 9)

Observed Average	Theoretical mass /Da	Protein	Cysteines in Sequence
8353	8353	C-X-C motif chemokine 5	4
	8349	Platelet basic protein-like 2	4
	8344	Rab3 GTPase-activating protein catalytic subunit	0
	8343	C-C motif chemokine 14	4
	8360	Dual specificity protein phosphatase 7	1
	8356	Transcription factor NF-E4	1
	8342	Prenylcysteine oxidase-like	2
8569	8575	Amyloid protein A	0
	8565	Ubiquitin	0
	8560	NEDD8	0
	8547	Keratin-associated protein 21-2	18
	8563	Proactivator polypeptide-like 1	6
	8571	Brain and acute leukemia cytoplasmic protein	2
	8569	Leucine zipper putative tumor suppressor 1	3

Search results by TagIdent were not subjected to statistical analyses for confidence. Carbamidomethylation chemistry (Section 2.6.2) to modify cysteine residues accessible to solvent can be used to augment the data-set used to identify a protein. The number of cysteine residues in the proteins tabulated in Table 4.3 can be used to predict mass shifts observable by MALDI-ToF-MS.

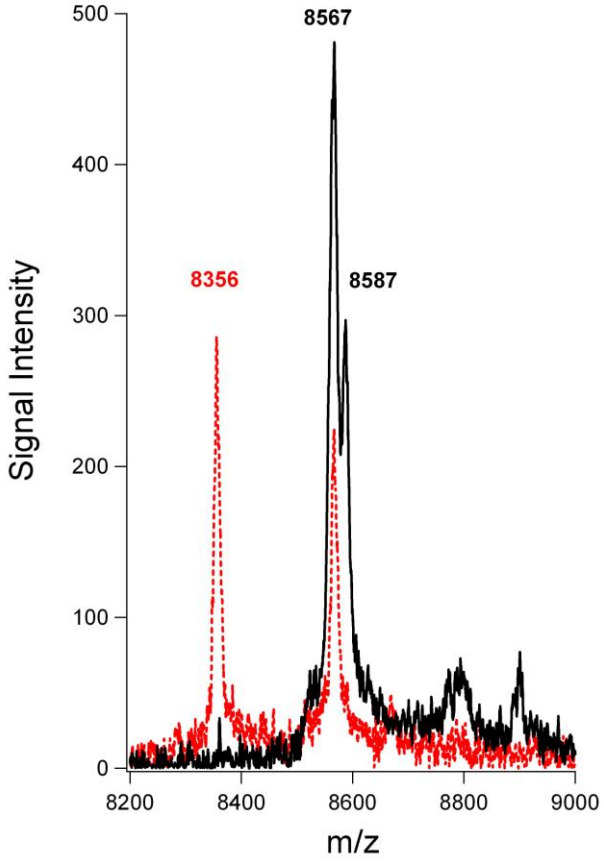

Figure 4.3 MALDI-ToF-MS spectra of m/z region for 8200 – 9200 for two supernatants from the same positive control experiment. (Red) Ion signals from supernatant with no addition of reagents for protein modification added. (Black) Ion signals from supernatant with iodoacetamide added for carbamidomethyl chemistry.

In the case of the ion signal at m/z = 8356, a mass shift of 231 was observed and based on a mass increase of 58 Da per modification by

carbamidomethyl chemistry, the mass shift was equivalent to that expected for 4 cysteine residues (Figure 4.3). This information was used to eliminate some tentative assignments listed in Table 4. 3 that were known to have 3 cysteines or less, leaving three or more (due to proteins not yet identified) possible identities for the ion peak including C-X-C motif chemokine 5, platelet basic protein-like 2, or C-C motif chemokine 14. For the ion signal at m/z 8567, no mass shift was observed, indicating no cysteine residues modified. This result cannot be used to eliminate any of tentative assignments, as it is possible to have cysteine residues present and not modified.

Factors that contribute to unsuccessful identification of ion signals by MALDI-ToF-MS include limitations of the instrument and the biology of the sample. When using TagIdent to query an observed ion signal, those ion signals were assumed to be from proteins that were intact. However, because other studies of A549 cells have identified proteases, such as matrix metalloprotease-9 and matrix metalloprotease-1 that are both secreted, the observed ion signals could also have been portions of proteins which these proteases generate that are not annotated into Swiss-Prot.[148, 149]

Instrumental limitations include limits to mass accuracy and resolution. In linear mode, the ToF MS attains a resolution of $M/\Delta m$ = 600-750 when the method of external mass calibration was applied. If C-X-C motif chemokine 5 and platelet basic protein-like 2 were both present, their ion signals would not be differentiated well, as their theoretical mass difference of 4 Da would not have

been resolved at 50% intensity, and to achieve that, a resolution of 2000 would be required.

4.3.1.2 Offline RPLC and MALDI-ToF-MS to Fractionate and Analyze Supernatants

MALDI ion sources operate under vacuum and require samples to have low volatility; analyte eluted from an LC needs to be in the solid phase prior to ionization by this source. As a result, offline RPLC-MALDI-ToF-MS was attempted, by first fractionating the supernatant to reduce its complexity then followed by MALDI-ToF-MS analyses. Fractionating the supernatant resulted in a coarse separation of molecules and new ion signals were observed compared to an un-separated supernatant observed (Figure 4.4 Spectrum A). Spectra B, C, and E are examples of coarse separation of molecules by observing m/z 4081, 8357 and 8567 in separate fractions eluting from 26.3 - 26.7, 30.3-30.7, and 37.0-37.3 minutes respectively. New ion signals observed in the fractionated samples include m/z 14,015 and 15,841 (spectrum D). In addition, S/b of many ion signals observed in the analyses of fractions from the separated supernatant were improved compared to an un-separated analysis , for example m/z 11,727 has an S/b of 36.8 (spectrum D) versus 9.0 (spectrum A). Data processing for total abundance ratios of analyte versus a reference compound's ion signal (Equation 4.1) in spectra of fractions was not performed due to these ions signals being observed in several fractions. The common understanding for observed increases in ion S/b after fractionating is the reduction of ion suppression due to competition effects in the process for ionization where there is a mixture of molecules.[61]

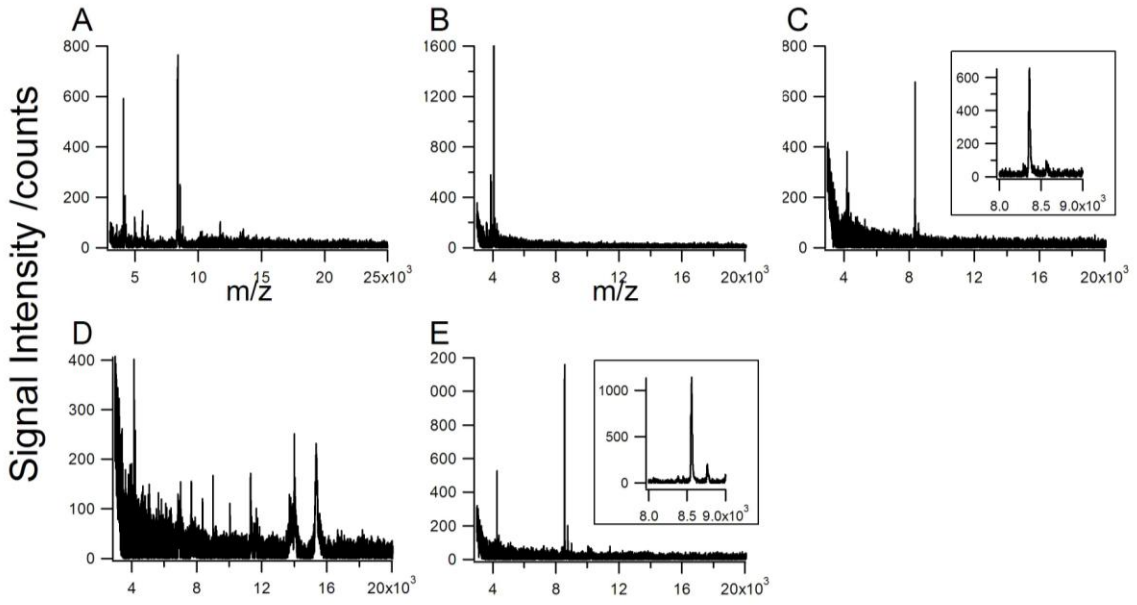

Figure 4.4 MALDI-ToF-MS spectra (m/z 3 - 20 x10^3) of a positive control incubated for 72 hours. (A) Spectrum is the result of combining two spectra from an un-separated sample (m/z 3 - 13 x 10^3 and 10-20 x 10^3). Spectrum B to E are from fractions that all demonstrated increases in S/S_b with some indicating coarse separation. (B) Fraction 35, eluted between 26.3-26.7 min indicate coarse separation of m/z 4,283 (C) Fraction 47, eluted between 30.3-30.7 min. indicate coarse separation of m/z 8,570 (D) Fraction 56, eluted 33.3-33.7 min. indicate increases in S/S_b for m/z 14,015 and 15,841 (E) Fraction 67, eluted 37.0-37.3 min. indicate coarse separation of m/z 8,557.

4.3.1.3 Digestion of Fractionated Supernatant for Peptide Fingerprinting by MALDI-ToF-MS

With the observation that fractions contained molecules giving rise to ion signals at m/z 8,354 and 8,570 were relatively isolated from the mixture, an attempt was made to digest fractions and try to identify proteins present by peptide fingerprinting. Spectra collected by MALDI-ToF-MS were reviewed for fractions containing similar ion signals to pool for trypsin digestion. Fractions 46 to 48 were pooled to investigate m/z 8354 ($[M+H]^+$) and 4178 ($[M+2H]^{2+}$) (Figure

4.5). Fractions 69 and 70 were pooled to investigate m/z 8570 ($[M+H]^+$) and 4284 ($[M+2H]^{2+}$).

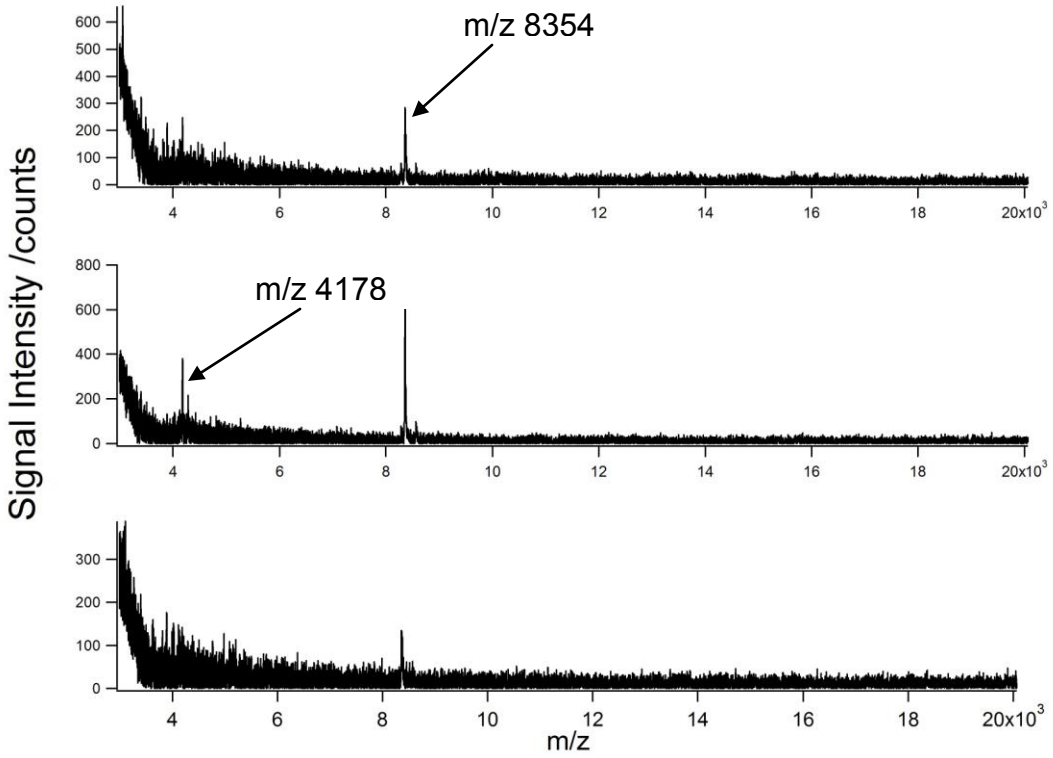

Figure 4.5 MALDI-ToF-MS spectra of fractions 46 to 48 observed to have m/z 8354 were pooled together for trypsin digestion. The ion signal at m/z 4178 is doubly protonated.

The resulting trypsin digest of fractions 46 - 48 or 69 – 70 pooled together was analyzed by a MALDI-ToF-MS in reflectron mode for m/z 1000-3000. Compared to the peptide signals from the auto-digestion of trypsin, ion signals differentially observed from the digests of fractions 46 - 48 or 69 - 70 pooled were noted for analysis by Mascot. The resulting searches for either peptide fingerprint list did not return any proteins with confident identification or related

homology. No other peptide fingerprinting was attempted based on the inability to obtain identifications for m/z 8354 and 8570, as they were the two most abundant ions isolated in the fractions. The supernatant analyzed was not subjected to protein denaturing or disulfide bond reduction; this meant that proteins could have remained folded and that may have prevented trypsin from accessing protein cleavage sites, or peptides to be linked by disulfide bonds. Both preventing trypsin from cleaving, and intact disulfide bonds contribute to provide ion signals that are not predicted by Mascot for comparing and scoring. To improve this analysis in future attempts, methodology described in sections 4.2.2.1 to denature, break disulfide bonds, and modify cysteine residues should be included.

4.3.2 Analysis of Trypsin Digested, Cysteine Carbamidomethylated, Supernatant by Micro-flow LC-ESI-MS

After attempting the methodology described in the previous paragraphs, no identification of an ion signals by a MALDI-ToF-MS could be achieved, and no additional method development for it was attempted. Supernatants from A549 cells after proteolysis by trypsin was analyzed by micro flow LC-ESI for its relatively short analysis time, and ability to acquire tandem-MS data for protein identification (Chapter 3). From the data collected, an ion with m/z = 716.4 was doubly charged as evidenced by the ~0.5 m/z spacing between isotopolouges (Figure 4.6). That precursor ion was selected for tandem MS. Submission of the LC-ESI tandem MS data collected from negative (n=2) and positive controls (n=6) to the Mascot search engine identified the precursor ion as

EICLDPEAPFLK from which was assigned to originate from the protein C-X-C motif chemokine 5. C-X-C motif chemokine 5 was the only protein identified and was observed in positive control supernatants in 2 of 6 analyses.

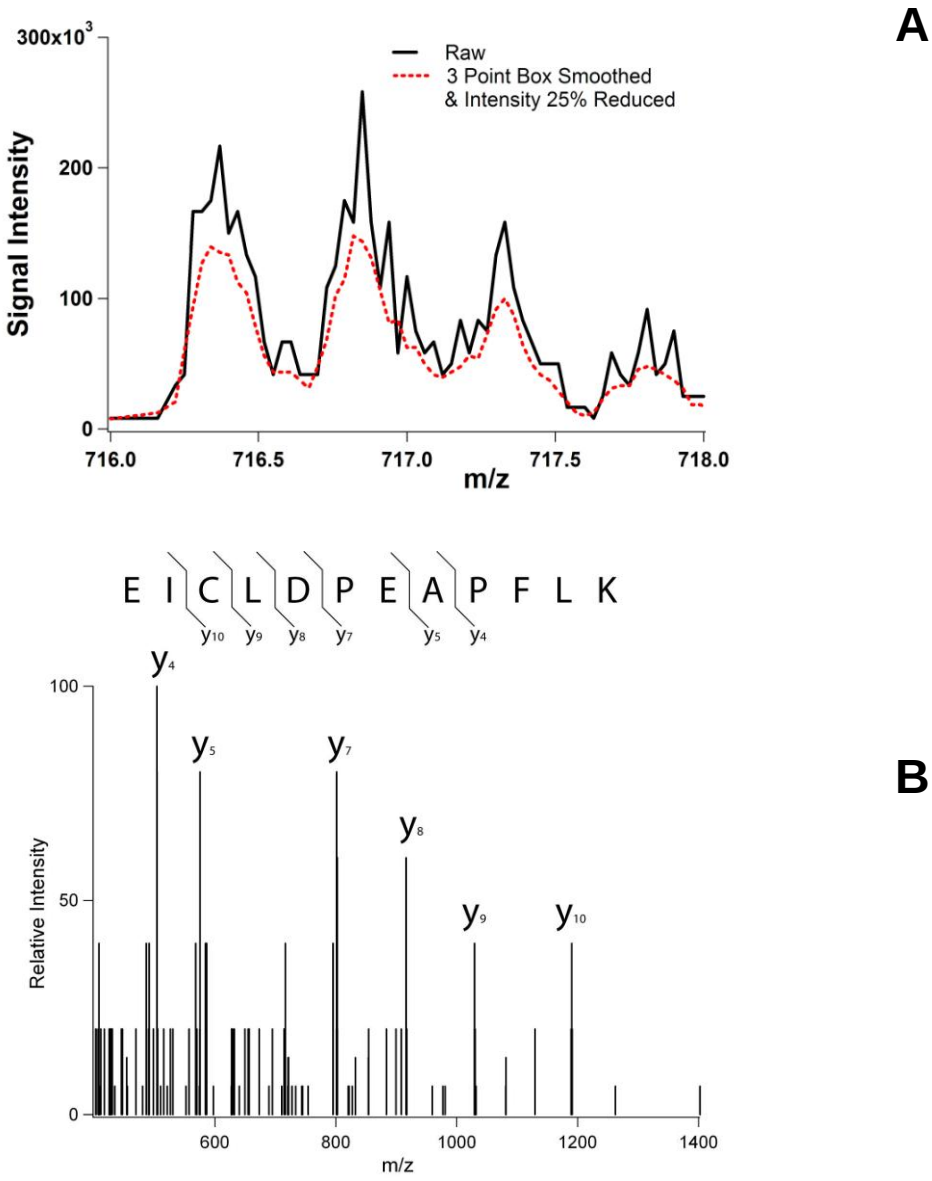

Figure 4.6 **(A) Mass spectrum of a peptide with monoisotopic peak at m/z 716.4 determined to be doubly charged because the separation between isotopolouges was ~0.5 Da (Section 2.1). (B) Tandem MS spectrum of the precursor ion m/z 716.4. The output from the Mascot database, regarding the sequence of the peptide from which it originated were added to the figure.**

The detection of C-X-C motif chemokine 5 (sequence 37-14) less the signalling peptide (sequence 1-36) in supernatants from A549 cells has been reported in literature. In a study of C-X-C chemokine 5 for potency in chemotaxis with varying sequence lengths (37-114, 41-114, 45-114, and 46-114), Nufer *et al.* demonstrated that the fragment 41-114 exhibited a 10 fold increase in potency, and used ESI-MS to confirm the MW of the different fragments.[150] Keshamouni *et al.* in their investigation of select chemokines for tumour growth in A549 cells measured, by ELISA, differential up-regulation of C-X-C chemokine 5 upon stimulation by TNF-α as compared to negative controls.[151] To my knowledge, at the time of writing this thesis, there was no published report indicating the use of tandem MS for the direct identification of C-X-C chemokine 5 by MS. The identification of C-X-C chemokine 5 is of particular interest because it is a protein that participates in an inflammatory response, as learned using ELISA in previous studies by other investigators.

Additional analyses of supernatant were not performed by micro-flow LC-ESI. Rather, nano-flow LC-ESI was used exclusively thereafter. This decision was based on several reasons including, only one protein was identified even though many ion signals were observed by MALDI-ToF-MS, an identification was made in only 2 of 6 analyses, and nano flow LC-ESI is purported in the literature as offering improved analyte ion signal abundance as compared to micro flow LC-ESI (Chapter 3).

4.3.3 Analysis of Trypsin Digested, Cysteine Carbamidomethylated, Supernatant by Nano-flow LC-ESI- MS

Replication of the sample preparation for the micro-flow LC-ESI analyses was repeated for analysis by nano-flow LC-ESI for supernatants from positive and negative control experiments, and resulted in 35 proteins being identified from the tandem MS data (Table 4.4). Proteins identified were further queried against published literature regarding their involvement in known biological processes. Many of the proteins are involved in multiple processes, and the most common function encountered was their role in “house-keeping” for homeostasis, which includes mRNA directed synthesis of proteins and their chaperones (for modification, folding, or transport), metabolism, cell growth, and maintenance of tissue structure. Processes like apoptosis and oxidation/reduction are part of homeostasis, but are affected by external stimuli, and as such, the proteins and their post-translational modifications that are known to be involved in an immune or stress induced-response are of considerable interest.

Supernatants from experiments dosed with compounds identified in PM, were analysed to learn of their effect when introduced in aqueous form as background information prior to further studies of these compounds' effect when introduced on particulate matter. These analyses resulted in an additional 43 proteins, for a total of 78, having been identified by tandem MS. Proteins identified by tandem MS and reported in Table 4.4 are consistent with MS analyses for the complete (intra- and inter-cellular proteins) protein analysis of the A549 cell line.[152, 153]

Table 4.4 Proteins identified by tandem MS in media (i.e. supernatant) conditioned by human epithelial lung cells (A549)

Biologicial Processes

a: Apoptosis
b: Cell signaling
c: Immune/stress response
d: House keeping: mRNA processing, synthesis, protein folding, transport, modification, metabolism, cell growth
e: Oxidation/reduction
f: Structure
g: Not well known

Legend

* Accociated with a biological process or dose type

Dose Type

U: Negative Control
V: 0.10 μM $Ni(NO_3)_2$
W: 0.86 μM $Ni(NO_3)_2$
X: 1 μg/mL LPS
Y: 1 μg/mL LPS + 6.5 μg/mL India Ink
Z: Positive Control

Bio. Processes							Protein	Peptides obseved	Sequence	Mass /da	Charge	Score	Dose type					
a	b	c	d	e	f	g							U	V	W	Z	Y	Z
*			*				10 kDa heat shock protein, mitochondrial	2	K.VLQATVVAVGSGSK.G	1314.75	2	77		*		*	*	*
*	*						14-3-3 protein epsilon, zeta, delata, gamma, eta	1	K.DSTLIMQLLR.D	1188.65	2	87	*			*	*	
			*				40S ribosomal protein S28	1	K.FAAATGATPIAGR.F	1202.64	2	37		*		*		
			*				40S ribosomal protein S30	1	R.FVNVVPTFGK.K	1106.61	2	36				*		
			*				60S acidic ribosomal protein	1	K.NIEDVIAQGIGK.L	1255.68	2	47	*			*	*	
			*				Actin, cytoplasmic 1 or 2	2	K.EITALAPSTMK.I	1160.61	2	79		*		*	*	*
		*	*	*			Alcohol Dehydrogenase [NADP+] / Aldose reductase	1	K.MPLIGLGTWK.S	1114.62	2	42					*	
			*	*			Aldo-keto reductase family 1 member B10	1	K.MPIVGLGTWK.S	1100.61	2	60		*		*	*	
			*	*			Aldo-keto reductase family 1 member C1 or C4	1	K.LNDGHFMPVLGFGTYAPAEVPK.S	2359.17	3	43		*		*		
							Aldose reductase	2	K.MPILGLGTWK.S	1114.62	2	48		*				
			*				Alpha-enolase / Gamma-enolase	2	R.AAVPSGASTGIYEALELR.D	1803.94	2	96	*	*	*	*	*	
			*				Annexin A2	1	K.TDLEKDIISDTSGDFR.K	1810.86	3	36					*	
							Beta-2-Microglobulin Precursor	1	K.VEHSDLSFSK.D	1147.55	2	35		*				
	*						Calmodulin	2	K.EAFSLFDKDGDGTITTK.E	1843.88	3	72	*	*		*	*	
						*	Calpastatin	1	K.AAAPAPVSEAVCR.T	1297.64	3	34					*	
*		*	*				Clusterin precursor	2	R.IDSLLENDR.Q	1073.54	2	198	*	*	*	*	*	*
						*	Coactosin-like protein	1	K.EVVQNFAK.E	933.49	2	37					*	
*						*	Cofilin-1	3	K.LGGSAVISLEGKPL.-	1339.77	2	135	*	*	*	*	*	*

Biologicial Processes

a: Apoptosis
b: Cell signaling
c: Immune/stress response
d: House keeping: mRNA processing, synthesis, protein folding, transport, modification, metabolism, cell growth
e: Oxidation/reduction
f: Structure
g: Not well known

Legend

* Accociated with a biological process or dose type

Dose Type

U: Negative Control
V: 0.10 μM $Ni(NO_3)_2$
W: 0.86 μM $Ni(NO_3)_2$
X: 1 μg/mL LPS
Y: 1 μg/mL LPS + 6.5 μg/mL India Ink
Z: Positive Control

Bio. Processes												Dose type					
a	b	c	d	e	f	g	Protein	Peptides obseved	Sequence	Mass /da	Charge	Score	U	V	W	Z	Y	Z
		*					Complement component 1 Q subcomponent-binding protein, mitochondrial precursor	1	K.AFVDFLSDEIKEER.K	1696.83	3	67	*				*	
	*	*					C-X-C motif chemokine 5 precursor	2	K.EICLDPEAPFLK.K	1430.71	2	101				*	*	*
	*	*					Cystatin-C	1	R.ALDFAVGEYNK.A	1225.6	2	62	*					
			*				Elongation factor 1-alpha 1	1	K.IGGIGTVPVGR.V	1024.6	2	44		*		*	*	
							FK506-binding protein 1A	1	M.GVQVETISPGDGR.T	1313.66	2	64					*	
							Fructose-bisphosphate aldolase A	1	R.ALANSLACQGK.Y	1131.57	2	60					*	
			*	*			Glyceraldehyde-3-phosphate dehydrogenase	2	K.LVINGNPITIFQER.D	1612.89	2	99		*		*	*	*
			*	*			Glyoxylate reductase/hydroxy pyruvate reductase	1	R.TRNTMSLLAANNLLAGLR.G	1928.06	3	38					*	
*		*	*				Heat Shock Protein beta-1	1	R.VSLDVNHFAPDELTVK.T	1982.92	23	33				*		
			*				Heterogeneous nuclear ribonucleoprotein K	1	R.NTDEMVELR.I	1105.51	2	37					*	
			*				Heterogeneous nuclear ribonucleoproteins A2/B1	1	K.IDTIEIITDR.Q	1187.64	2	66	*		*	*	*	*
			*				Heterogeneous nuclear ribonucleoproteins A1	1	K.IEVIEIMTDR.G	1217.63	2	52				*		
			*				Heterogeneous nuclear ribonucleoproteins C1/C2	1	R.VPPPPPIAR.A	942.57	2	42					*	*
			*				Histone H2A type 1, 2, or 3	2	R.AGLQFPVGR.V	943.52	2	70		*			*	

Biologicial Processes							Dose Type	
a: Apoptosis				f: Structure			U: Negative Control	
b: Cell signaling				g: Not well known			V: 0.10 μM $Ni(NO_3)_2$	
c: Immune/stress response							W: 0.86 μM $Ni(NO_3)_2$	
d: House keeping: mRNA processing, synthesis, protein folding, transport, modification, metabolism, cell growth				Legend			X: 1 μg/mL LPS	
				* Accociated with a biological process or dose type			Y: 1 μg/mL LPS + 6.5 μg/mL India Ink	
e: Oxidation/reduction							Z: Positive Control	

Bio. Processes													Dose type					
a	b	c	d	e	f	g	Protein	Peptides obseved	Sequence	Mass /da	Charge	Score	U	V	W	Z	Y	Z
			*				Histone H2B type 1-A/C/E/F/G/I	1	R.LLLPGELAK.H	952.6	2	50	*	*		*	*	*
			*				Histone H4	2	R.ISGLIYEETR.G	1179.61	2	48		*		*	*	
			*				Histone H3.1t	2	R.EIAQDFK.T	849.42	2	55						*
			*				Insulin-like growth factor-binding protein 4 precursor	3	R.EDARPVPQGSCQSELHR.A	1964.91	3	102	*	*	*	*	*	*
					*		Keratin, type I cytoskeletal 10	2	R.SQYEQLAEQNR.K	1364.62	2	57		*			*	
							Keratin, type I cytoskeletal 14	1	K.VTMQNLNDR.L	1089.52	2	41		*				
*			*		*		Keratin, type I cytoskeletal 18	2	R.TVQSLEIDLDSMR.N	1505.62	2	81				*	*	
					*		Keratin, type I cytoskeletal 9	2	R.SGGGGGGGLGSGGSIR.S	1231.59	2	57					*	
					*		Keratin, type II cytoskeletal 1	2	K.LNDLEDALQQAK.E	1356.69	2	43					*	*
					*		Keratin, type II cytoskeletal 8	3	R.SLDMDSIIAEVK.A	1418.74	2	220		*		*	*	
					*		Lamin-A/C	1	K.EGDLIAAQAR.L	1042.54	2	37					*	*
					*		LIM and SH3 domain protein 1	2	K.GFSVVADTPELQR.I	1417.72	2	55				*		*
			*	*			L-lactate dehydrogenase A chain	1	R.VIGSGCNLDSAR.F	1247.59	2	50		*			*	*
*	*	*					Macrophage Migration Inhibitory Factor	1	M.PMFIVNTNVPR.A	1286.68	2	53	*	*				
			*				Metalloproteinase inhibitor 1 precursor	1	K.GFQALGDAADIR.F	1232.61	2	34					*	
			*			*	Nucleobindin-1 precursor	3	R.YLQEVIDVLETDGHFR.E	1932.96	3	127	*	*		*	*	
*	*	*	*				Nucleophosmin	1	K.MSVQPTVSLGGFEITPPVVLR.L		3	88				*	*	
	*	*					Osteopontin precursor	1	K.AIPVAQDLNAPSDWDSR.G	1853.89	2	43			*			
				*			Oxidoreductase NAD-binding domain-containing protein 1 precursor	1	M.ACAAVMIPGLLR.C	1270.69	2	31				*		
		*					Parathymosin	1	K.SVEAAAELSAK.D	1074.56	2	46					*	

Biologicial Processes

a: Apoptosis
b: Cell signaling
c: Immune/stress response
d: House keeping: mRNA processing, synthesis, protein folding, transport, modification, metabolism, cell growth
e: Oxidation/reduction
f: Structure
g: Not well known

Legend
* Accociated with a biological process or dose type

Dose Type

U: Negative Control
V: 0.10 μM $Ni(NO_3)_2$
W: 0.86 μM $Ni(NO_3)_2$
X: 1 μg/mL LPS
Y: 1 μg/mL LPS + 6.5 μg/mL India Ink
Z: Positive Control

Bio. Processes							Protein	Peptides obseved	Sequence	Mass /da	Charge	Score	Dose type					
a	b	c	d	e	f	g							U	V	W	Z	Y	Z
			*				Peptidyl-prolyl cis-trans isomerase A	4	K.EGMNIVEAMER.F	1277.57	2	112	*	*	*	*	*	*
			*				Peptidyl-prolyl cis-trans isomerase B	1	K.TVDNFVALATGEK.G	1363.7	2	38				*		
			*	*	*		Peroxiredoxin-1	5	K.ATAVMPDGQFK.D	1163.56	2	58	*	*	*	*	*	*
			*				Phosphatidylethanolamine-binding protein 1	1	K.VLTPTQVK.N	884.53	2	36					*	
			*				Phosphoglycerate kinase 1	1	K.ACANPAAGSVILLENLR.F	1767.93	3	42				*		
*		*	*				Plasminogen activator inhibitor 1 precursor	3	R.QFQADFTSLSDQEPLHVAQALQK.V	2600.29	3	64	*	*	*		*	
			*				Proactivator polypeptide precursor [Contains: Saposin-A	6	R.LGPGMADICK.N	1060.5	2	61	*		*	*	*	*
					*		Profilin-1	4	K.STGGAPTFNVTVTK.T	1378.71	2	118	*	*	*	*	*	*
	*						ProSAAS precursor	1	R.AADHDVGSELPPEGVLGALLR.V	2115.1	3	56					*	
		*	*				Protein S100-A6	2	R.LMEDLDR.N	890.42	2	45	*	*			*	*
*			*				Protein SET	1	R.VEVTEFEDIK.S	1207.6	2	39				*		
			*			*	Putative hexokinase HKDC1	1	R.AIPDGSENGEFLSLDLGGSK.F	2004.96	3	36	*			*	*	
			*				Putative Heterogeneous Nuclear Ribonucleoprotein A1-Like Protein 3	1	K.IEVIEIMTDR.G	1217.63	2	71	*					
*			*				Pyruvate kinase isozymes M1/M2	1	R.GDLGIEIPAEK.V	1140.6	2	58		*				
			*			*	Regulator of G-protein signaling 13	1	K.TMQSNNSF.-	927.38	2	35			*			
			*	*			Retinal dehydrogenase 1	1	K.VAFTGSTEVGK.L	1094.56	2	50						*
*		*	*	*			Superoxide dismutase [Cu-Zn]	1	K.GDGPVQGIINFEQK.E	1500.76	2	48				*		
			*				Synaptophysin-like protein 1	1	M.APNIYLVRQR.I	1228.7	2	35		*				

Biologicial Processes			
a: Apoptosis	f: Structure	Dose Type	
b: Cell signaling	g: Not well known	U: Negative Control	
c: Immune/stress response		V: 0.10 μM $Ni(NO_3)_2$	
d: House keeping: mRNA processing, synthesis, protein folding, transport, modification, metabolism, cell growth	Legend	W: 0.86 μM $Ni(NO_3)_2$	
	* Accociated with a biological process or dose type	X: 1 μg/mL LPS	
		Y: 1 μg/mL LPS + 6.5 μg/mL India Ink	
e: Oxidation/reduction		Z: Positive Control	

Bio. Processes								Peptides					Dose type					
a	b	c	d	e	f	g	Protein	obseved	Sequence	Mass /da	Charge	Score	U	V	W	Z	Y	Z
	*		*	*			Thioredoxin	4	K.TAFQEALDAAGDK.L	1335.63	2	195	*	*	*	*	*	*
					*		Thymosin beta-4	1	K.NPLPSKETIEQEK.Q	1511.78	2	35					*	
					*	*	Transgelin-2	2	K.NVIGLQMGTNR.G	1201.62	2	43				*	*	
			*				Transketolae	1	R.IIALDGDTK.N	994.52	2	45		*				
			*				Triosephosphate isomerase	2	K.SNVSDAVAQSTR.I	1233.59	2	84				*	*	
					*		Tropomyosin alpha/beta	1	K.LVILEGELER.A	1169.67	2	47					*	
	*		*				Tuberin	1	K.DMEGLVDTSVAK.I	1263.6	2	32		*				
			*				Tubulin alpha-1A Chain	1	R.AVFVDLEPTVIDEVR.T	1700.9	3	33	*					
*	*		*				Ubiquitin	2	K.TITLEVEPSDTIENVK.A	1786.92	2	116	*	*	*	*	*	*
			*				Ubiquitin carboxyl-terminal hydrolase isozyme L1	2	K.LGFEDGSVLK.Q	1063.56	2	39	*	*	*			
			*				Vacuolar ATP synthase subunit E 2	1	R.NDLISDLLSEAKLR.L	1585.87	3	35			*			
			*		*		Vimentin	5	K.ILLAELEQLK.G	1168.71	2	57	*	*	*	*	*	*

Individual scores (Equation 2.13) reported in Table 4.4 are based on the probability the identification of the reported protein is a random event, and the scores are the sums of all (unique or replicate analyses) tandem MS results. Note, minimum scores needed for identification and false positive (Equation 2.14) rates for an analysis varies depending on the mass accuracy of the data and number of tandem MS queries. Minimum scores of 28-31 and false positive rates of 3-8 % were reported in the analyses of supernatants.

Analyses performed by the nano-flow LC-ESI were conducted under non-ideal chromatography conditions as illustrated by irreproducible elution times and long tailing observed in chapter 3. Reproducibility in protein identification within replicate analyses by information dependant analysis (IDA, page 53) was poor and is commonly commented in the literature.[141, 154, 155] The use of IDA strategies is primarily for qualitative analysis and its ineffectiveness for quantitative analysis (e.g. monitoring protein expressions) has been commented on in literature.[107, 111] As a result, Table 4.4 is best interpreted simply as a list of proteins identified in the supernatants of A549 cells from different experiments.

4.3.3.1 Comparison of m/z 8356 and m/z 8567 Ion Signal Intensity Ratios in Supernatants from A549 Cell Cultures Dosed with Selected Compounds

An application of identifying the proteins present in the supernatants is improved interpretation of ion signals (i.e. protein expression) related to biological processes. MALDI-MS spectra were collected for supernatants from cultures of controls and experiments that involved dosages of solutions containing LPS, $Ni(NO_3)_2$, India ink, and combinations thereof. After processing the ion signal intensities with Equation 4.1, the ion signal intensity ratio m/z 8356 to m/z 8567 (C-X-C motif chemokine 5 to ubiquitin as learned by observing mass shifts due to carbamidomethyl chemistry and tandem mass spectrometry) was plotted (Figure 4.7).

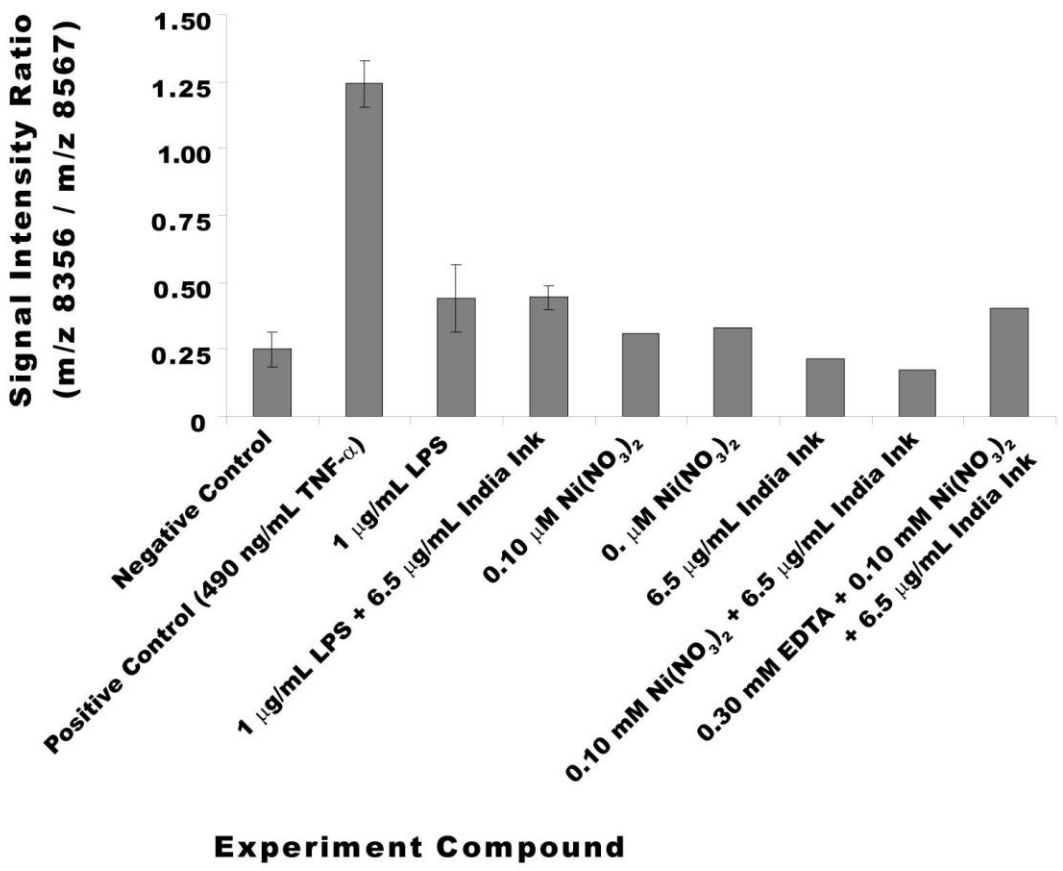

Figure 4.7 Ratio of ion signal intensities observed by the MALDI-ToF-MS at m/z 8356 to m/z 8567 (C-X-C chemokine 5 to ubiquitin) signals as a function of selected 275 μL doses to A549 cells bathed in 2.5 mL of SFM. The final concentration of compounds were as follows: positive control 490 ng/mL TNF-α, LPS 1 μg/mL, India ink 6.5 μg/mL, and $Ni(NO_3)_2$ 0.10 or 0.86 μM, and EDTA 0.30 μM.

Only the ratios of have C-X-C motif chemokine 5 to ubiquitin between negative (n = 3) and positive (n = 3) controls were found to differentially expressed with a confidence > 95 % from a Students-T test. Based on that observation, we hypothesize the ratio of C-X-C motif chemokine 5 to ubiquitin can be used as an indicator of cellular stimulation. Experiments dosed with LPS (n =3 each with or without ink) were also performed, though their average expression of C-X-C motif

chemokine 5 to ubiquitin was observed to be elevated over a negative control, confidence from a multiple hypothesis Student's-t test was less than 90%. Results from this preliminary study of selected compounds indicate C-X-C motif chemokine 5/ubiquitin can be semi-quantitatively measured to study their effects on human epithelial lung cell cultures. Based on the ratio of C-X-C motif chemokine 5 to ubiquitin, of the selected compounds dosed, ink, $Ni(NO_3)_2$, and LPS (with or without India ink) was found to have increasing ratios. Results for $Ni(NO_3)_2$ with ink were unexpected as the C-X-C chemokine 5/ubiquitin ratio was smaller than $Ni(NO_3)_2$ alone given that $Ni(NO_3)_2$ was at equal concentrations. However $Ni(NO_3)_2$ with ink and EDTA did result in a higher C-X-C chemokine 5/ubiquitin ratio and is consistent with A. Kardjaputri's work of observing an increase in a different pro-inflammation molecule, ICAM-1. As this study of selected compounds for dose responses is preliminary, it is incomplete and additional replicates are needed for experiments with India ink, $Ni(NO_3)_2$, and EDTA to determine whether there ratios are reproducible. In addition to replicates, different concentrations should be studied to learn what concentrations are required to illicit a differential response compared to a negative control.

4.4 Summary

MS methods described in this chapter were developed with obtaining identification of proteins in A549 cell culture supernatants as the focus. The use of MALDI-ToF-MS and related methods was not able to produce identifications. However, it did provide complementary information regarding the molecular

weights of possible intact proteins and their mass shifts following selective chemistry for cysteines (breaking disulfide bonds and carbamidomethylation). The use of an LC for fractionation prior to MALDI-ToF-MS did demonstrate limited success in reducing the complexity of different analytes present in the supernatant, as evidenced by improved ion signal intensities when analyzing their fractions compared to un-separated samples.

The use of LC-ESI for tandem MS led to protein identifications. In particular nano-flow LC-ESI generated 78 identifications of proteins compared to 1 by micro-flow LC-ESI. Due to the poor reproducibility of the nano-flow LC the list of proteins compiled for Table 4.4 cannot be used quantitatively as interpretation of that data set should be restricted to the identification of proteins within the A549 cell proteome, with no indication of quantitative information.

In retrospect, after reviewing all the MS derived information regarding supernatants analyzed, MALDI-ToF-MS alone can be employed to characterize supernatants, to rapidly assess the presence of an ion signal for a previously identified protein. For example, from this study, following the data set that includes the tandem MS spectra, theoretical molecular weights, and carbamidomethyl chemistry, C-X-C motif chemokine 5 and ubiquitin were identified to be responsible for the ion signals at m/z 8356 and 8570 respectively. In future work, MALDI-ToF-MS can be used to rapidly assess supernatants of A549 lung cell cultures for their relative abundance with the identities of C-X-C motif chemokine 5 and ubiquitin known.

Furthermore, based on the data from the preliminary dose response experiment conducted, MALDI-ToF-MS can be used to semi-quantitatively rank the compounds studied based on the ion signal of C-X-C motif chemokine 5 to ubiquitin. For example, based on comparing negative and positive controls, elevated ratios of C-X-C motif chemokine 5 to ubiquitin indicated a response to cellular stimulation. Of the selected compounds in PM dosed onto A549 cells, India ink, $Ni(NO_3)_2$, and LPS were found to cause, increased A549 cell response in such a way that the ion signal intensity ratio of C-X-C chemokine 5 to ubiquitin increased in the order written.

5: SINGLE CHARGED DROPLET AS A SOURCE OF IONS FOR MS

5.1 Introduction

The history of MS can readily be told from the perspective of advances in accessibility to molecules having different properties which tracks the development of new ion sources. For example, the introduction of MALDI and ESI as soft ionization sources illustrate a significant foray into high molecular weight biological molecules.[48, 51, 156, 157]With ESI being broadly introduced to the MS community in the early 1990s, the application base of ESI is now considered mature. Though ambiguity persists regarding aspects of the mechanism for ion production, characterization of space charge effects in the plume of droplets between the ESI tip and MS that were found to restrict ion signal helped motivate developments to improve ESI, and nano-ESI is one outcome.[96]

Other developments that have lead to improved performance for ESI include the addition of supplemental apparatuses. For example, described as an ion-funnel, additional electrodes positioned between the ESI and MS orifice, or offset orifices to exploit differences in mobility of charged entities, have lead to modest improvements in directing charged droplets or ions depending on the device and application toward the sampling orifice of an atmospheric pressure gas sampling interface.[158] One other strategy being developed in several laboratories world-wide to negate the limitations imposed by space charge

effects is a single droplet at a time approach. The strategy of manipulating single droplets is neither new nor unique, but it is relatively new with respect to the production of ions of interest for MS. For example, various apparatuses have been reported that utilize electric fields (static and/or dynamic), acoustic waves, and inkjet dispensers to isolate or produce droplets with net charge.[159-161]

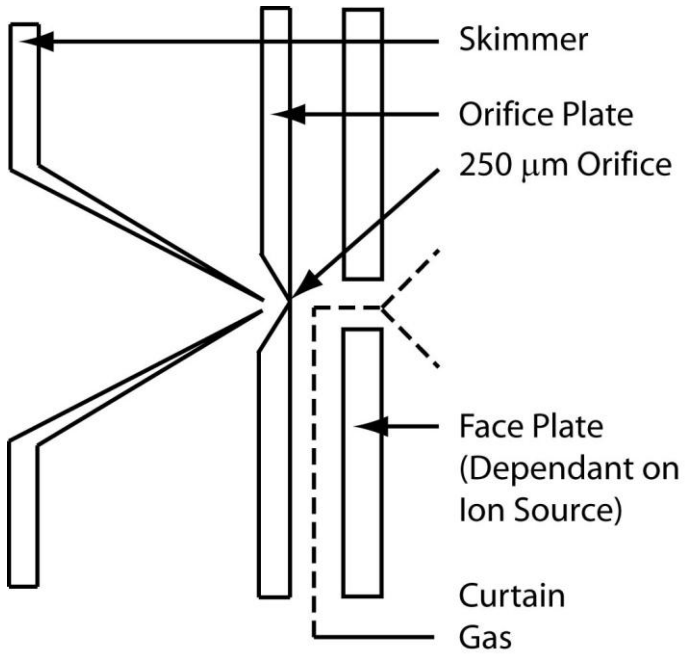

Figure 5.1 Front end of a hybrid Q-q-LIT-MS used, employing a orifice plate-skimmer with curtain gas flow.

The proposed ion source developed for this chapter is based on trapping and levitating droplets with net charge within an electro dynamic levitation trap (EDLT). An EDLT is based on a 3-dimensional quadrupole ion trap, which functions similarly to a LIT for which the theory was reviewed in section 2.5.1 (Figure 5.2). A significant difference, is that the trapping is done at atmospheric pressure, resulting in is additional dampening due to collisions with air and

because of that, the m/z of the entity trapped is also different. Another difference is an outcome of the collisional dampening of charge object motion in the EDLT, which often leads to non-optimal electrode geometry being used (e.g. the electrode structure of the EDLT are poor mimics of an 3-dimensional QIT) yet the performance of the device is not negatively affected. As such, the EDLT used in the Agnes laboratory is often termed an AC trap to avoid possible misrepresentation of how we use the levitation device.

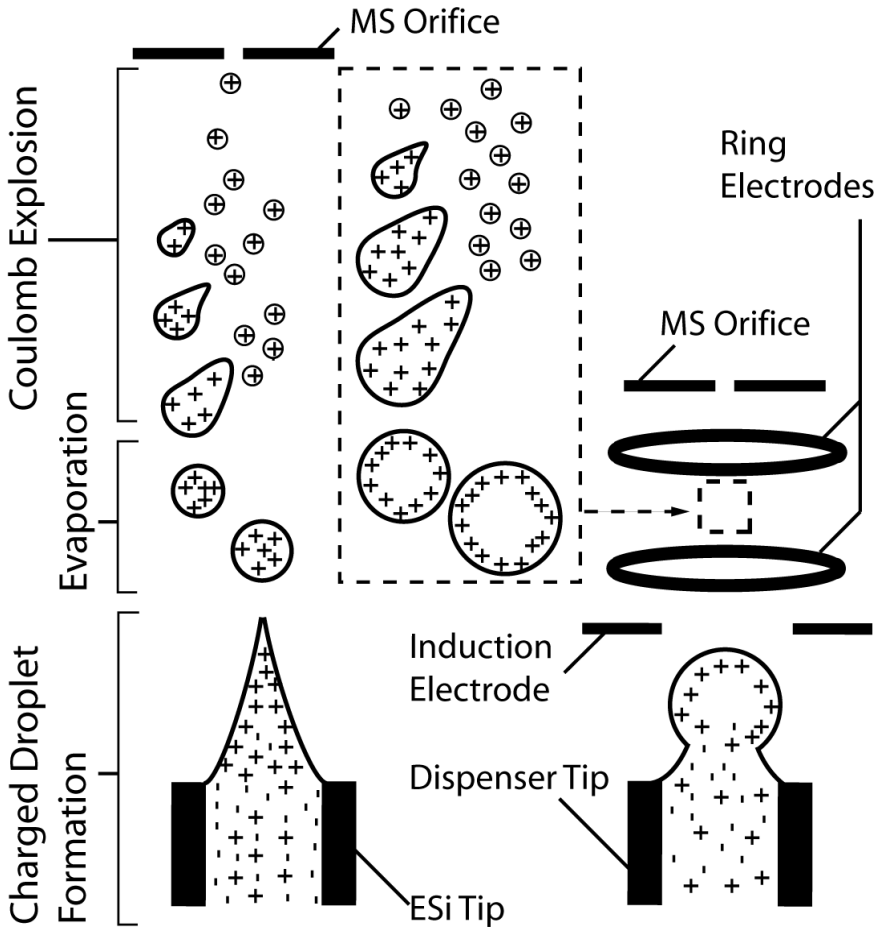

Figure 5.2 Cartoon comparison of conventional ESI and AC trap for droplets with net charges undergoing evaporation leading to coulomb explosion events and ion formation.

The general strategy was to trap and levitate mono-disperse droplets (primary) from starting solutions for analysis, and levitating the droplet for a period of time to allow most of its solvent to evaporate. Upon reaching the Coulomb limit, the primary droplet undergoes an electrostatic-driven explosion that causes the primary droplet to release smaller droplets, termed progeny. Ideally progeny droplets would be directed towards the orifice of the MS such that molecular ion formation occurs immediately in front of the orifice to the MS.

Motivation for investigating an AC trap and associated methodology as a source of ions was to investigate whether detection limits could be improved, and to potentially enable studies on the ionization mechanism that are otherwise essentially precluded using an ESI. The reasoning is when a single levitated droplet loses solvent due to evaporation, the concentration of analyte was estimated to increase by ~2 orders of magnitude (it is not known to what extent solute concentration occurs in the droplets generated in a conventional ESI that contribute to ion production). Even though studies for nano-ESI indicate ionization efficiencies decrease with increasing droplet size, it was hypothesized the combination of decreased space charge effects and an increase in total analyte abundance could lead to improved sensitivity, provided that methodology was available to preferentially pre-concentrate analyte compounds in starting solution while not concentrating matrix compounds and electrolyte. An example of studies that could be made feasible with a single droplet ion source would be a droplet desolvation time-course study. Demonstrating that would help address

whether there is a chronological order to the ionization of molecules hypothesized to be surface activated versus solvated within droplets.

5.2 Methodology

Glycerol and tetrabutylammonium bromide were purchased from Sigma Aldrich (St. Louis, MO) and were used to prepare solutions. Solutions of 2% v/v glycerol were dispensed to develop methodology for reproducibly directing droplets out of an EDLT. Solutions of a 1 μM tetrabutylammonium bromide + 10 μM NaCl were dispensed for detection by MS.

5.2.1 Dispensing, Trapping, and Levitating, Droplets with Net Charge

A droplet-on-demand generator (model 201, Uni-photon Systems, Brooklyn, NY) was fitted with a commercially available droplet dispenser (MJ-AB-01-60, MicroFab Technologies Inc., Plano, TX). Solutions (~10 μL) for dispensing were loaded into the reservoir of the droplet dispenser, using an auto-pipette. An electrode biased with a negative DC potential (-75 to -150V) positioned ~2 mm away from the nozzle tip of the dispenser was used to induce a net positive charge onto each droplet (Figure 5.3). Controlled by a Variac, the output of a 10X line-voltage amplifier (60 Hz) was applied to two ring electrodes (0^0 difference in phase). The amplitude of the AC potential applied to the ring electrodes ranged from 1400 - 1680 $V_{0\text{-}P}$. The electrodes of the AC trap were enclosed in a (5 x 5 x 5 cm) aluminum chamber. To illuminate the droplets by scattering light, the output of a 5W laser (HeNe, uniphase, 5W minimum power)

was directed into the chamber enclosing the ring electrode and through the volume within the AC trap that the droplet was levitated.

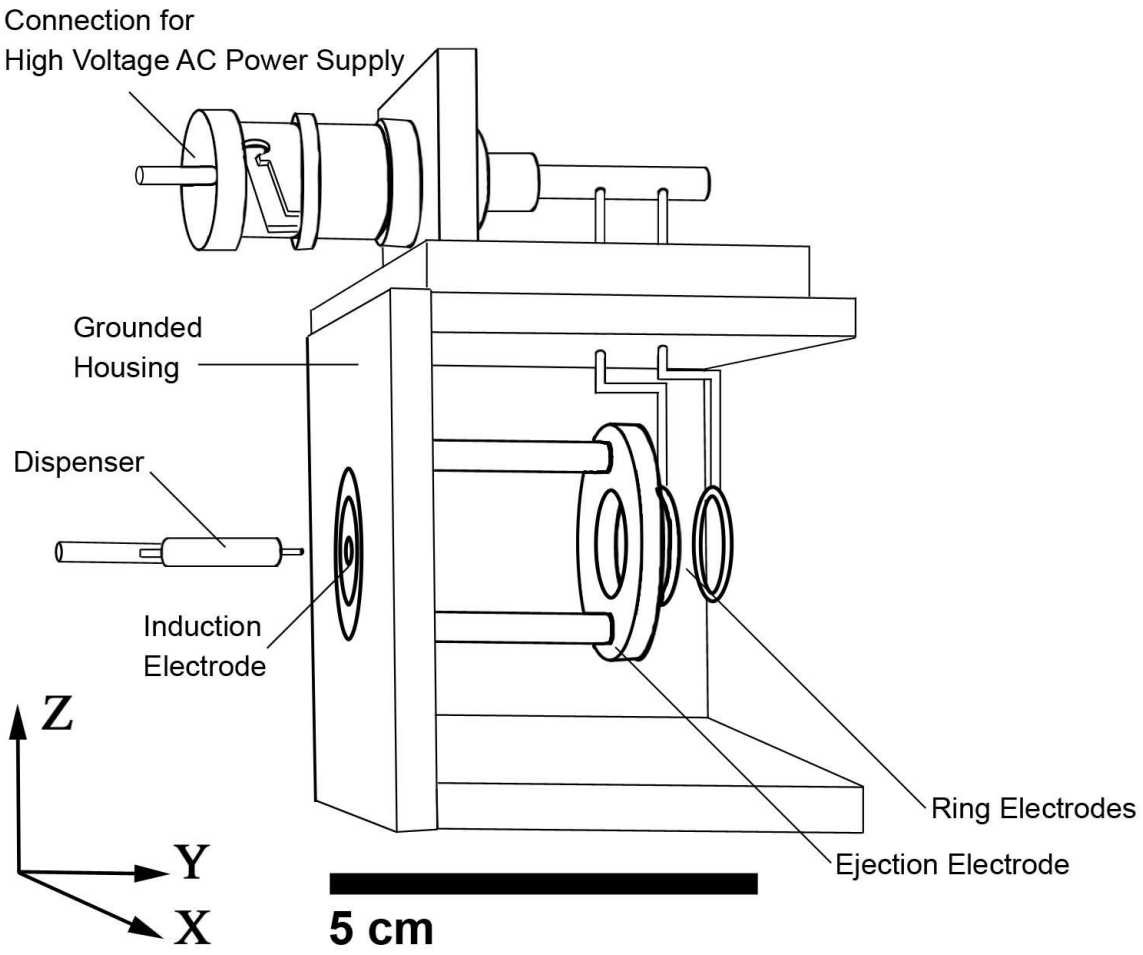

Figure 5.3 Schematic diagram of the AC trap for droplet trapping and delivery in the horizontal direction to couple with bench-top MS equipped with an atmospheric pressure gas sampling interface.

Throughout the entire process, an electrode that was located between the induction electrode and the ring electrodes was charged with a positive DC potential (350 to 1200V) to provide an electric field that ideally directed trapped

droplets, that had been levitated but whose levitation had become unstable due to a loss of mass (e.g. solvent evaporation), towards the orifice of the MS. The entire assembly was mounted onto the translation stage of a Nanospray ion source (Applied Biosystems, Foster City, CA) provided by Thomas Clark. The electrical conditions just described were for dispensing, trapping, and levitating droplets, and those conditions were pre-determined by using the AC trap as a stand-alone device to capture and levitated one droplet at a time.

When the droplet dispenser is inactive, solution accumulates at the nozzle of the dispenser. As such, the “first” droplet that is dispensed is usually dissimilar in velocity (thus mass/volume) compared to subsequent droplets dispensed, which are more mono-dispersed. During an experiment, the path between the AC trap and the MS is blocked when the first droplet and subsequent droplets (~10) were being dispensed. Once more mono-disperse droplets are serially trapped and levitated, the path was opened for droplets to traverse. Note that the pattern of location of droplet deposition (i.e. ejection from the AC trap) onto a coverslip fixed at a single position in one experiment was highly variable when the AC trap was mounted for horizontal droplet injection and ejection from the device. Factors that contributed to this variability are the AC trap's imperfections, gravity, and variation in the initial droplet mass and net charge.

5.2.2 MS Conditions

A hybrid Q-q-LIT-MS (Q-Trap 2000, Applied Biosystems, Foster City, CA) was used for this study. A face plate designed for nano-ESI was used, and this plate was electrically grounded to the instrument. Instrument conditions at the

front end of the MS were: 5 (units set in the software) curtain gas, 45 V de-clustering potential, and medium CAD. The LIT was configured for a trap fill-time of 150 ms followed by a mass scan read-out at a rate of 1000 Da/s from m/z 50 to 1700 for a duration of 1.67s.

5.3 Results and Discussion

The strategy for the development of an methodology was to direct droplets (primary or progeny) with net charge towards the sampling orifice of the MS. Upon arriving near the front of the sampling orifice, Coulomb explosion would lead to the production of molecular ions at a location where the sampling through the orifice into the MS would be favourable. To achieve this, after leaving the AC trap, a primary droplet would have to travel through a face plate with a 3 mm diameter aperture, and then on to an orifice plate having an 250 μm diameter aperture, and the orifice plate was located 2 mm behind the face plate. In total, there was a separation gap of 2 cm from the centre of the AC trap to the orifice plate.

Two orientations of the AC trap were studied for reproducibility in delivering droplets to the orifice. The first orientation studied maintained the orientation of the ion path in the bench-top MS horizontal, or parallel to the desk-top (horizontal ejection). This orientation demonstrated droplets ejected non-reproducibly due to the restoring force of the AC trap causing vertical oscillations. Vertical oscillations and horizontal ejection resulted in vectors that varied the path of droplets by >3.5 mm, as evidenced by viewing their residues after impacting on a glass slide positioned ~2 cm away from the AC trap. Note, this

data was acquired using the AC-trap offline to mimic the relative position of the orifice versus the AC trap (Figure 5.4).

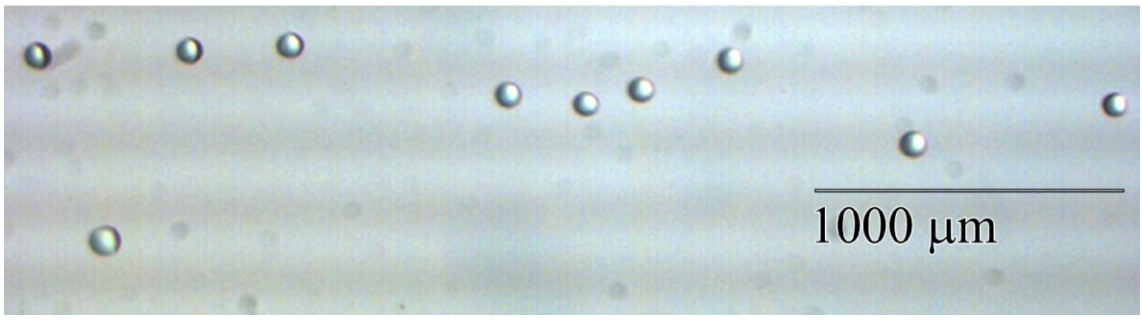

Figure 5.4 Image from a glass slide position ~2 cm away from the centre of the ring electrodes onto which droplets had been levitated illustrates their deposition location.

To overcome the variation of pathways caused by the vertical oscillation and horizontal ejection, the assembly was rotated 90^0 to eject droplets vertically upward. This was found to improve the reproducibility of where droplets would impact on a glass slide. Data from a single test of this configuration indicated that in serially dispensing a population of 40 droplets, 58% (23 droplets) deposited onto a coverslip (positioned with a separation gap to the AC trap to mimicked that realized when interface with the MS, and the droplets deposited) in a circular region having a diameter of 250 μm diameter circle (Figure 5.5A). To improve mimicry of the conditions at the front end of an MS, the face plate with a 3 mm diameter aperture used by the MS had also been placed flush against the housing of the AC trap. In serially dispensing a population of 20 droplets, repeated three times, individual droplets (12, 7, and 11 droplets) were found to

deposit within a 250 μm diameter circle 50 +/- 13 % of the time (Figure 5.5 B to D).

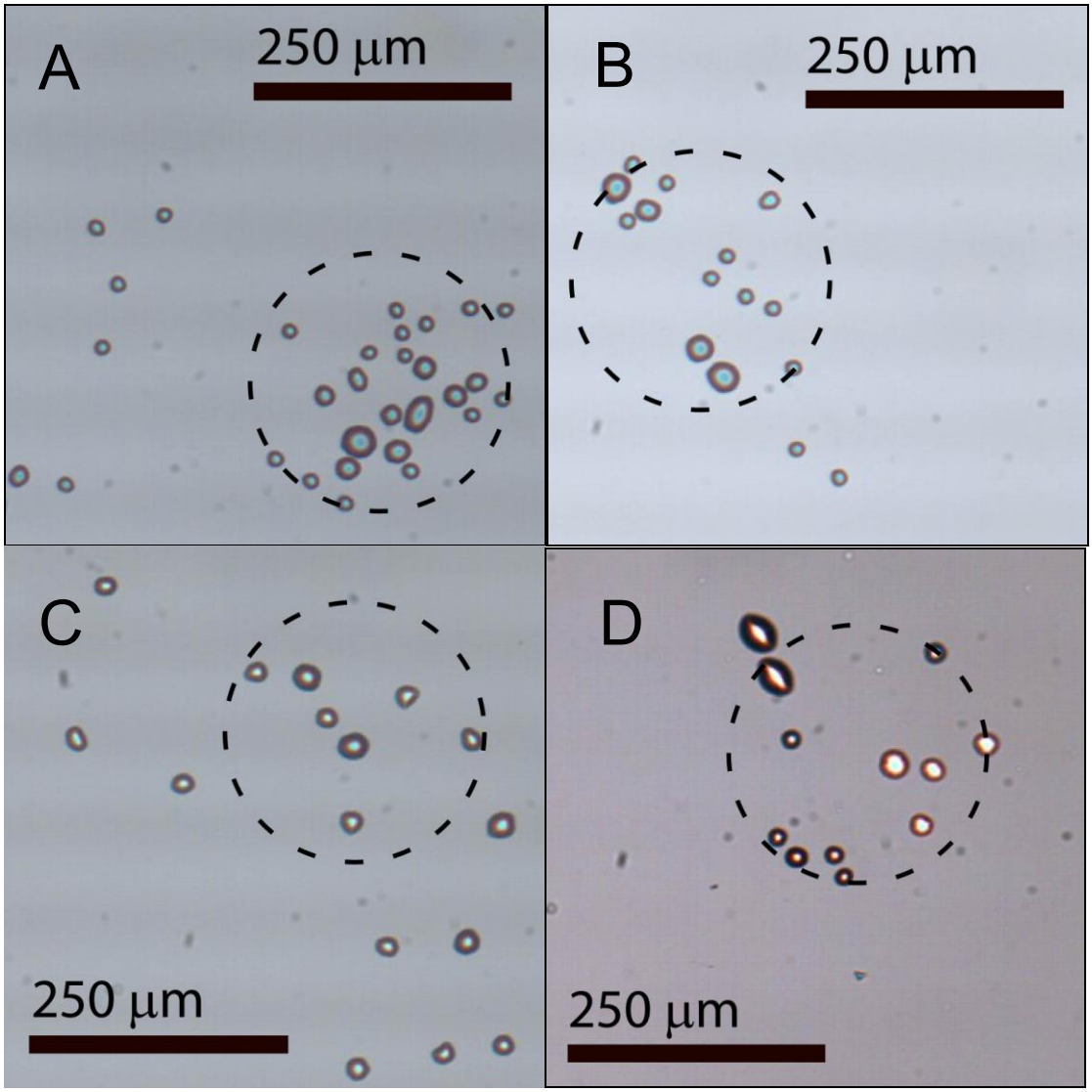

Figure 5.5 **Droplet deposition location after being briefly levitated in an AC trap, and then ejected vertically up so that they impacted onto a glass coverslip positioned above the trap to study how droplets will travel. All images have a dotted circle drawn in to represent a 250 μm diameter orifice of an MS, and the starting solution used contained 2% v/v glycerol. (A) Image of 23 individual droplets, even though 40 droplets were dispensed. Images B-D had an additional face plate with a 3 mm hole positioned to mimic conditions in front of an MS (B) 12 of 20 droplets dispensed (C) 7 of 20 droplets dispensed (D) 11 of 20 droplets dispensed.**

After developing a method for the vertical upward ejection of droplets from the AC trap with their impact on a remote target, the apparatus and method was transferred for testing with a MS. To accommodate droplets travelling in the vertical direction, an aluminium scaffold was constructed to support the rotation of the MS instrument 90^0 from its normal orientation of being a bench-top instrument, such that the orifice of the MS could face vertically up or down. Without confidence in the size of droplets being directed towards the MS, fear of large particulates entering and damaging the instrument precluded any investigation of dispensing droplets downwards. A solution of 1 μM $N((CH_2)_3(CH_3))_4Br$ + 10 μM NaCl was prepared to be detected by the MS as quaternary ions are known to be easily observed by MS instruments. Attempts to direct droplets towards the MS in the vertically upward direction for molecular ions were ultimately unsuccessful as no ion signals in a mass spectrum could be observed to represent the tetrabutylammonium quaternary ions being introduced. In Figure 5.6, representative ion chromatograms were acquired over many replicate cycles, with each cycle lasting up to 30 s for dispensing, trapping, and levitating single droplets of the quaternary salt solution.

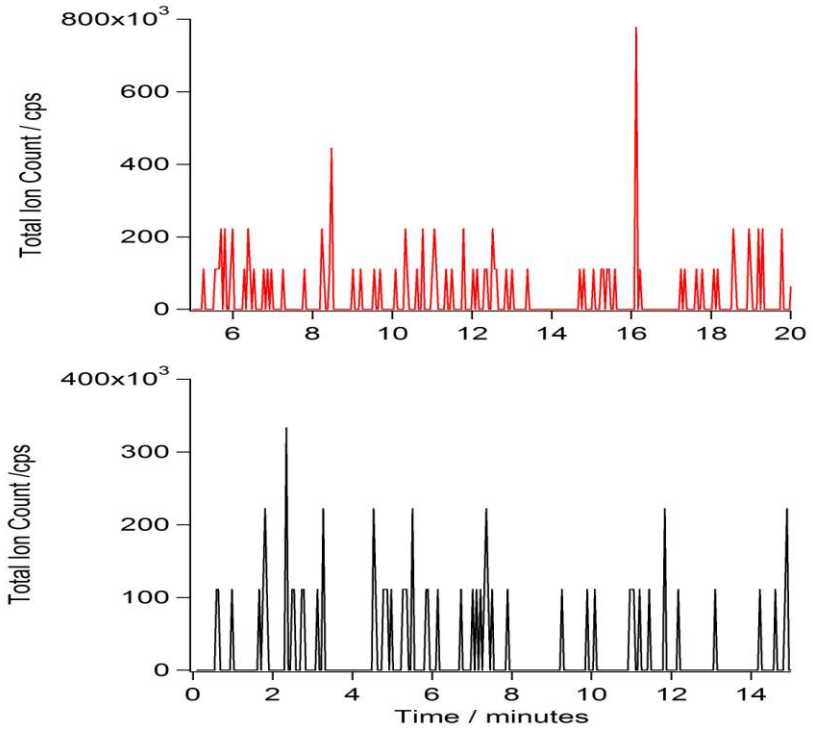

Figure 5.6 Total ion chromatograms from MS performing 150 ms of trapping followed by 1.65 s of mass scan out for analysis. (Bottom) Background (Top) Selected ion chromatogram collected curing dispensing, trapping, and levitating droplets of 1 μM $N((CH_2)_3(CH_3))_4Br$ + 10 μM NaCl.

The timing for ion trapping and mass analysis of ions during the experiment likely introduced opportunities for the MS to miss sampling ions. Typical coulomb explosion events from a single droplet during ESI are estimated to last <1 ms, and occur 2 to 3 times ~70 ms apart.[87] If coulomb explosions occurred during the 1.65 s period of time that ion readout for mass analysis was taking place, ions entering the MS will not be accumulated. The duty cycle for trapping ions could be improved significantly by re-configuring the MS software in consideration of the droplet dispensing rate and synchronizing these events.

Curtain gas turbulence, a charged orifice (200 V DC) from an operating MS, and the obvious lack of instrument synchronization, were all contributing factors that thwarted the goal of being able to detect analyte ions. The turbulent gas flow from the curtain gas is intended to aid in solvent evaporation, or directing large droplets away from the orifice when the MS is used as designed, with either ESI or nano-ESI sources. Also, with an ESI, the orifice plate is charged by the instrument to prevent droplets with net charge or ions from striking and neutralizing on its surface. Attempts to eject droplets with greater force by increasing the DC potential (up to 1200 V DC) on the ejection electrode did not lead to detection of analyte ions. At potentials greater than 1200 V DC, droplets with net charge were observed to be repelled away from the electrode used for ejection, preventing trapping within the AC trap.

To overcome the curtain gas and charged orifice plate issues, a unique configuration of the electrodes specifically customized for an electric potential gradient when the MS is interfaced with an AC trap is needed. Similar in concept to an ion-funnel or ion guide, time dependent waveforms applied to electrodes positioned between the ring electrodes and MS might improve electrostatic forces acting on the droplets to better guide them to the orifice of a MS.[162]

5.4 Conclusion

A mimic of an orifice to an MS was used during the initial development phase of this work to learn to reproducibly direct droplets towards a target When this apparatus was operated to eject droplet vertically out of the AC trap, 50 +/- 13% of the droplets ejected landed on a remotely positioned target within a circle

of diameter 250 μm (i.e. analogous to an orifice to a MS). When the same apparatus was interfaced to an operating MS, there were no ion signals differentiable from background detected. Of the many possible factors that likely contributed to the end result of no analyte ions being detected by the MS, certainly the poor synchronization of the AC trap to the MS was a factor and turbulent flow of the curtain gas used negatively affected the transport of droplets to the orifice plate.

6: CONCLUSIONS AND FUTURE DIRECTIONS

The use of MS to study samples collected from cultures of lung cells enabled the identification and semi-quantitation of proteins. Different MS technologies were exploited for their performance characteristics in this work. MALDI-ToF-MS is a technique that was shown to be capable of being used to characterize samples with respect to monitoring the relative intensity of ion signals from intact, previously identified proteins quickly and semi-quantitatively. Nano-flow LC-ESI was demonstrated to have improved performance over micro-flow LC-ESI for tandem MS with respect to identification of proteins. An attempt to develop an AC trap as a source of ions for MS failed, yet off-line data of droplet delivery to a remote target showed promise.

6.1 Future Work and Directions

Numerous studies hypothesize that changes in protein expressions are indicative of biological states. A large majority of the work undertaken during the course of this thesis was related to the identification of the ion signals at m/z 8357 and 8560 as originally observed using MALDI-ToF-MS in characterizing A549 supernatants. Though 76 other proteins were identified to be present in the A549 supernatants by nano-flow LC ESI-tandem-MS, many more ion signals observed by MALDI-ToF-MS remain unidentified. Experiments to improve the

number of ion signals identified will expand the utility of MALDI-ToF-MS to monitor ion signals rapidly with semi-quantitation.

An LC system designed to deliver reproducible flow in the nano-litre/min. regime is needed. With improved control of nano-flow rates in a separation column, the majority of the issues encountered in this work regarding analyte elution time irreproducibility, and nano-ESI signal intensities will be alleviated. To introduce quantitation into the methods developed, external calibration curves could then be constructed by monitoring ion signals of standards using other tandem MS scans, such as MRM techniques for known parent-product ion pairs to monitor, which can improve sensitivity and linear dynamic range. Another method to improve the quantitation of proteins in samples is the use of standards with or without labelling techniques.

References

1. Vallabhajosula, S. in *Molecular Imaging: Radiopharmaceuticals for PET and SPECT* (ed Vallabhajosula, S.) 109 (Springer Berlin Heidelberg, Heidelberg, 2009).

2. Wilkins, M., Sanchez, J., Golley, A. & Appel, R. HumpherySmith, I., Hochstrasser, DF and Williams, KL, Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol. Genet. Eng. Rev.* **13**, 19–50 (1996).

3. Sampath, R. *et al*. Rapid identification of emerging infectious agents using PCR and electrospray ionization mass spectrometry. *Ann. N. Y. Acad. Sci.* **1102**, 109-120 (2007).

4. Dockery, D. W. *et al*. An association between air pollution and mortality in six US cities. *N. Engl. J. Med.* **329**, 1753 (1993).

5. Samet , J. M. *et al*. The National Morbidity, Mortality, and Air Pollution Study. Part II: Morbidity and mortality from air pollution in the United States. *Res. Rep. Health Eff. Inst.* **94**, 5-70; discussion 71-9 (2000).

6. Schwartz, J. Air pollution and hospital admissions for cardiovascular disease in Tucson. *Epidemiology* **8**, 371-377 (1997).

7. Villeneuve, P. J. *et al*. A time-series study of air pollution, socioeconomic status, and mortality in Vancouver, Canada. *J. Exposure Sci. Environ. Epidemiol.* **13**, 427-435 (2003).

8. Pope III, C. A., Ezzati, M. & Dockery, D. W. Fine-particulate air pollution and life expectancy in the United States. *N. Engl. J. Med.* **360**, 376 (2009).

9. Poschl, U. Atmospheric aerosols: Composition, transformation, climate and health effects. *Angew. Chem. Int. Ed.* **44**, 7520-7541 (2005).

10. Noll, K. E. & Fang, K. Y. P. Development of a dry deposition model for atmospheric coarse particles. *Atmos. Environ.* **23**, 585-594 (1989).

11. Finlayson-Pitts, B. J. & Pitts Jr, J. N. Tropospheric air pollution: ozone, airborne toxics, polycyclic aromatic hydrocarbons, and particles. *Science* **276**, 1045 (1997).

12. Jacobson, M. C., Hansson, H. C., Noone, K. J. & Charlson, R. J. Organic atmospheric aerosols: Review and state of the science. *Rev. Geophys.* **38**, 267–294 (2000).

13. Ravishankara, A. R. Heterogeneous and multiphase chemistry in the troposphere. *Science* **276**, 1058 (1997).

14. Kampa, M. & Castanas, E. Human health effects of air pollution. *Environ. Pollut.* **151**, 362-367 (2008).

15. Fuzzi, S. *et al*. Critical assessment of the current state of scientific knowledge, terminology, and research needs concerning the role of organic aerosols in the atmosphere, climate, and global change. *Atmos.Chem.Phys* **6**, 859-860 (2006).

16. Harrison, R. M. & Yin, J. Particulate matter in the atmosphere: which particle properties are important for its effects on health? *Sci. Total Environ.* **249**, 85-101 (2000).

17. Vincent, R. *et al*. Regulation of promoter-CAT stress genes in HepG2 cells by suspensions of particles from ambient air. *Toxicol. Sci.* **39**, 18-32 (1997).

18. Adamson, I. Y. R., Prieditis, H. & Vincent, R. Pulmonary toxicity of an atmospheric particulate sample is due to the soluble fraction. *Toxicol. Appl. Pharmacol.* **157**, 43-50 (1999).

19. Brook, R. D. *et al*. Particulate Matter Air Pollution and Cardiovascular Disease: An Update to the Scientific Statement From the American Heart Association. *Circulation* **121**, 2331 (2010).

20. Brook, R. D. *et al*. Air pollution and cardiovascular disease: a statement for healthcare professionals from the Expert Panel on Population and Prevention Science of the American Heart Association. *Circulation* **109**, 2655 (2004).

21. Goto, Y. *et al*. Particulate Matter Air Pollution Stimulates Monocyte Release from the Bone Marrow. *Am. J. Respir. Crit. Care Med.* **170**, 891-897 (2004).

22. Khan, M. M. in *Immunopharmacology* (ed Khan, M. M.) 253 (SpringerScience+BusinessMedia, LLC, New York, 2008).

23. Buckland, K. F. & Hogaboam, C. M. in *Immunology of Fungal Infections* (eds Brown, G. D. & Netea, M. G.) 201-234 (Springer, Dordrecht, Netherlands, 2007).

24. van Eeden, S. F., Yeung, A., Quinlam, K. & Hogg, J. C. Systemic response to ambient particulate matter: relevance to chronic obstructive pulmonary disease. *Proc. Am. Thorac. Soc.* **2**, 61 (2005).

25. Bienvenut, W. V. *et al*. in *Acceleration and Improvement of Protein Identification by Mass Spectrometry* (ed Bienvenut, W. V.) 225-281 (Springer, 2005).

26. MacBeath, G. Protein microarrays and proteomics. *Nat. Genet.* **32**, 526-532 (2002).

27. Svensson, H. Isoelectric fractionation, analysis, and characterisation of ampholytes in natural pH gra-dients. I. The differential equation of solute concentrations at a steady state and its solution for simple cases. *Acta Chem. Scand.* **15** (1961).

28. Bierczynska-Krzysik, A. & Lubec, G. in *Proteomics : introduction to methods and applications* (eds Kraj, A. & Silberring, J.) 43 (John Wiley & Sons, Inc, Hoboken, New Jersey, 2008).

29. Nozaki, Y. & Tanford, C. The solubility of amino acids and related compounds in aqueous urea solutions. *J. Biol. Chem.* **238**, 4074 (1963).

30. Nozaki, Y. & Tanford, C. The solubility of amino acids and related compounds in aqueous ethylene glycol solutions. *J. Biol. Chem.* **240**, 3568 (1965).

31. Buxgaum, E. Introduction to Protein Structure and Function, 13-31 (2007).

32. Skoog, D. A., Holler, F. J. & Nieman, T. A. in *Principles of Instrumental Analysis* 849-674 (Sauders College Publishing, Orlando, Florida, 1998).

33. Erickson, H. P. Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy. *Biol. Proc. Online* **11**, 32-51 (2009).

34. Wofsy, L. & Burr, B. The use of affinity chromatography for the specific purification of antibodies and antigens. *J. Immunol.* **103**, 380 (1969).

35. Eisen, H. N. & Siskind, G. W. Variations in Affinities of Antibodies during the Immune Response. *J. Biol. Chem.* **3**, 996-1008 (1964).

36. Towbin, H., Staehelin, T. & Gordon, J. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc. Natl. Acad. Sci. U. S. A.* **76**, 4350 (1979).

37. Adair, W. S., Jurivich, D. & Goodenough, U. W. Localization of Cellular Antigens in Sodium Dodecyl Sulfate-Polyacrylamide Gels. *J. Cell Biol.* **79**, 281-285 (1978).

38. Patton, W. F. Detection technologies in proteome analysis. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences* **771**, 3-31 (2002).

39. Taylor, S. H. & Thorp, J. M. Properties and biological behaviour of Coomassie blue. *Br. Heart J.* **21**, 492 (1959).

40. Morrissey, J. H. Silver stain for proteins in polyacrylamide gels: a modified procedure with enhanced uniform sensitivity. *Anal. Biochem.* **117**, 307-310 (1981).

41. Switzer, R. C. A highly sensitive silver stain for detecting proteins and peptides in polyacrylamide gels. *Anal. Biochem.* **98**, 231-237 (1979).

42. Beaven, G. H. & Holiday, E. R. Ultraviolet absorption spectra of proteins and amino acids. *Adv. Protein Chem.* **7**, 319-386 (1952).

43. Issaq, H. J. The role of separation science in proteomics research. *Electrophoresis* **22**, 3629 (2001).

44. Aitken, A. & Learmonth, M. in *The Protein Protocol Handbook* (ed Walker, J. M.) 3-6 (Humana Press Inc., Totowa, NJ, 1996).

45. Butler, J. E. Enzyme-linked immunosorbent assay. *J. Immunoassay Immunochem.* **21**, 165-209 (2000).

46. Edman, P. & Begg, G. A protein sequenator. *Eur. J. Biochem.* **1**, 80-91 (1967).

47. Fietzek, P. & Kühn, K. in *Automation in Analytical Chemistry* (ed Adelmann, B. C.) 1-28 (Springer Berlin, Heidelberg, 1972).

48. Yates, J. R., Ruse, C. I. & Nakorchevsky, A. Proteomics by mass spectrometry: approaches, advances, and applications. *Annu. Rev. Biomed. Eng.* **11**, 49-79 (2009).

49. El-Aneed, A., Cohen, A. & Banoub, J. Mass spectrometry, review of the basics: electrospray, MALDI, and commonly used mass analyzers. *Appl. Spectrosc. Rev.* **44**, 210-230 (2009).

50. Chen, C. H. W. Review of a current role of mass spectrometry for proteome research. *Anal. Chim. Acta* **624**, 16-36 (2008).

51. Bantscheff, M., Schirle, M., Sweetman, G., Rick, J. & Kuster, B. Quantitative mass spectrometry in proteomics: a critical review. *Analytical and bioanalytical chemistry* **389**, 1017-1031 (2007).

52. Ong, S. E. & Mann, M. Mass spectrometry–based proteomics turns quantitative. *Nat. Chem. Biol.* **1**, 252-262 (2005).

53. Browne, T. R. Stable isotopes in pharmacology studies: present and future. *J. Clin. Pharmacol.* **26**, 485 (1986).

54. Wales, T. E. & Engen, J. R. Hydrogen exchange mass spectrometry for the analysis of protein dynamics. *Mass Spectrom. Rev.* **25**, 158-170 (2006).

55. Gronert, S. Mass spectrometric studies of organic ion/molecule reactions. *Chem. Rev.* **101**, 329-360 (2001).

56. Bouchoux, G., Salpin, J. Y. & Leblanc, D. A relationship between the kinetics and thermochemistry of proton transfer reactions in the gas phase. *Int. J. Mass Spectrom. Ion Processes* **153**, 37-48 (1996).

57. Glish, G. L. & Burinsky, D. J. Hybrid mass spectrometers for tandem mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **19**, 161-172 (2008).

58. Haddon, W. F. & McLafferty, F. W. Metastable ion characteristics. VII. Collision-induced metastables. *J. Am. Chem. Soc.* **90**, 4745-4746 (1968).

59. Kelly, M. A., Vestling, M. M., Fenselau, C. C. & Smith, P. B. Electrospray analysis of proteins: A comparison of positive-ion and negative-ion mass spectra at high and low pH. *Org. Mass Spectrom.* **27**, 1143-1147 (1992).

60. Wang, G. & Cole, R. B. Effect of solution ionic strength on analyte charge state distributions in positive and negative ion electrospray mass spectrometry. *Anal. Chem.* **66**, 3702-3708 (1994).

61. Zenobi, R. & Knochenmuss, R. Ion formation in MALDI mass spectrometry. *Mass Spectrom. Rev.* **17**, 337-366 (1998).

62. Chace, D. H. Mass spectrometry in the clinical laboratory. *Chem. Rev.* **101**, 445-477 (2001).

63. Munson, M. S. B. & Field, F. H. Chemical ionization mass spectrometry. I. General introduction. *J. Am. Chem. Soc.* **88**, 2621-2630 (1966).

64. Baldwin, M. A. & McLafferty, F. W. Liquid chromatography mass spectrometry interface–I: The direct introduction of liquid solutions into a chemical ionization mass spectrometer. *Org. Mass Spectrom.* **7**, 1111-1112 (1973).

65. Macfarlane, R. D. & Torgerson, D. F. Californium-252 plasma desorption mass spectroscopy. *Science* **191**, 920 (1976).

66. Karas, M., Bachmann, D. & Hillenkamp, F. Influence of the wavelength in high-irradiance ultraviolet laser desorption mass spectrometry of organic molecules. *Anal. Chem.* **57**, 2935-2939 (1985).

67. Karas, M. & Kruger, R. Ion formation in MALDI: the cluster ionization mechanism. *Chem. Rev.* **103**, 427-440 (2003).

68. Dole, M., Cox, H. L. & Gieniec, J. Electrospray mass spectroscopy. *Adv. Chem.* **125**, 73–84 (1973).

69. Takats, Z., Wiseman, J. M., Gologan, B. & Cooks, R. G. Mass spectrometry sampling under ambient conditions with desorption electrospray ionization. *Science* **306**, 471 (2004).

70. Berry, C. E. Effects of initial energies on mass spectra. *Phys. Rev.* **78**, 597-605 (1950).

71. Goudsmit, S. A. A time-of-flight mass spectrometer. *Phys. Rev.* **74**, 622-623 (1948).

72. Gordon, J. P., Zeiger, H. J. & Townes, C. H. The maser—new type of microwave amplifier, frequency standard, and spectrometer. *Phys. Rev.* **99**, 1264-1274 (1955).

73. Glish, G. L. & Vachet, R. W. The basics of mass spectrometry in the twenty-first century. *Nat. Rev. Drug Discovery* **2**, 140-150 (2003).

74. Paul, W. & Steinwedel, H. A new mass spectrometer without a magnetic field. *Z. Naturforsch.* **8**, 448-450 (1953).

75. Baldeschwieler, J. D. & Woodgate, S. S. Ion cyclotron resonance spectroscopy. *Acc. Chem. Res.* **4**, 114-120 (1971).

76. Hu, Q. *et al*. The Orbitrap: a new mass spectrometer. *J. Mass. Spectrom.* **40**, 430-443 (2005).

77. Li, L. *et al*. High-throughput peptide identification from protein digests using data-dependent multiplexed tandem FTICR mass spectrometry coupled with capillary liquid chromatography. *Anal. Chem.* **73**, 3312-3322 (2001).

78. Belov, M. E. *et al*. Dynamic range expansion applied to mass spectrometry based on data-dependent selective ion ejection in capillary liquid chromatography Fourier transform ion cyclotron resonance for enhanced proteome characterization. *Anal. Chem.* **73**, 5052-5060 (2001).

79. Covey, T. R., Bonner, R. F., Shushan, B. I., Henion, J. & Boyd, R. K. The determination of protein, oligonucleotide and peptide molecular weights by ion-spray mass spectrometry. *Rapid Commun. Mass Spectrom.* **2**, 249-256 (1988).

80. Wiley, W. C. & McLaren, I. H. Time-of-Flight Mass Spectrometer with Improved Resolution. *Rev. Sci. Instrum.* **26**, 1150 (1955).

81. Whittal, R. M. & Li, L. High-resolution matrix-assisted laser desorption/ionization in a linear time-of-flight mass spectrometer. *Anal. Chem.* **67**, 1950-1954 (1995).

82. Mamyrin, B. A., Karataev, V. I., Shmikk, D. V. & Zagulin, V. A. The mass-reflectron, a new nonmagnetic time-of-flight mass spectrometer with high resolution. *Sov.Phys.JETP* **37**, 45–48 (1973).

83. Lubman, D. M., Bell, W. E. & Kronick, M. N. Linear mass reflectron with a laser photoionization source for time-of-flight mass spectrometry. *Anal. Chem.* **55**, 1437-1440 (1983).

84. Kirkland, J. J. Microparticles with bonded hydrocarbon phases for high-performance reverse-phase liquid chromatography. *Chromatographia* **8**, 661-668 (1975).

85. Schluter, H. Reversed-phase chromatography. *J. Chrom. Lib.* **61**, 147-234 (2000).

86. Mahoney, W. C. & Hermodson, M. A. Separation of large denatured peptides by reverse phase high performance liquid chromatography. Trifluoroacetic acid as a peptide solvent. *J. Biol. Chem.* **255**, 11199 (1980).

87. Kebarle, P. & Tang, L. From ions in solution to ions in the gas phase-the mechanism of electrospray mass spectrometry. *Anal. Chem.* **65**, 972-986 (1993).

88. Cech, N. B. & Enke, C. G. Practical implications of some recent studies in electrospray ionization fundamentals. *Mass Spectrom. Rev.* **20**, 362-387 (2001).

89. Chowdhury, S. K., Katta, V., Beavis, R. C. & Chait, B. T. Origin and removal of adducts (molecular mass= 98 u) attached to peptide and protein ions in electrospray ionization mass spectra. *J. Am. Soc. Mass Spectrom.* **1**, 382-388 (1990).

90. Bruins, A. P., Covey, T. R. & Henion, J. D. Ion spray interface for combined liquid chromatography/atmospheric pressure ionization mass spectrometry. *Anal. Chem.* **59**, 2642-2646 (1987).

91. Gomez, A. & Tang, K. Charge and fission of droplets in electrostatic sprays. *Phys. Fluids* **6**, 404 (1994).

92. Iribarne, J. V. & Thomson, B. A. On the evaporation of small ions from charged droplets. *J. Chem. Phys.* **64**, 2287 (1976).

93. Dole, M. *et al*. Molecular beams of macroions. *J. Chem. Phys.* **49**, 2240 (1968).

94. Cech, N. B. & Enke, C. G. Relating electrospray ionization response to nonpolar character of small peptides. *Anal. Chem.* **72**, 2717-2723 (2000).

95. Cech, N. B., Krone, J. R. & Enke, C. G. Predicting electrospray response from chromatographic retention time. *Anal. Chem.* **73**, 208-213 (2001).

96. Busman, M., Sunner, J. & Vogel, C. R. Space-Charge-Dominated Mass Spectrometry Ion Sources: Modeling and Sensitivity. *J. Am. Soc. Mass Spectrom.* **2**, 1-10 (1991).

97. Karas, M., Bahr, U. & Dülcks, T. Nano-electrospray ionization mass spectrometry: addressing analytical problems beyond routine. *Fresenius J. Anal. Chem.* **366**, 669-676 (2000).

98. Schmidt, A., Karas, M. & Dülcks, T. Effect of different solution flow rates on analyte ion signals in nano-ESI MS, or: when does ESI turn into nano-ESI? *J. Am. Soc. Mass Spectrom.* **14**, 492-500 (2003).

99. El-Faramawy, A., Siu, K. W. M. & Thomson, B. A. Efficiency of nano-electrospray ionization. *J. Am. Soc. Mass Spectrom.* **16**, 1702-1707 (2005).

100. Wilm, M. S. & Mann, M. Electrospray and Taylor-Cone theory, Dole's beam of macromolecules at last? *Int. J. Mass Spectrom. Ion Processes* **136**, 167-180 (1994).

101. de Hoffmann, E. Tandem mass spectrometry: a primer. *J. Mass Spectrom.* **31**, 129-137 (1996).

102. Sleno, L. & Volmer, D. A. Ion activation methods for tandem mass spectrometry. *J. Mass Spectrom.* **39**, 1091-1112 (2004).

103. Steen, H. & Mann, M. The ABC's (and XYZ's) of peptide sequencing. *Nat. Rev. Mol. Cell Biol.* **5**, 699-711 (2004).

104. Summerfield, S. G. & Gaskell, S. J. Fragmentation efficiencies of peptide ions following low energy collisional activation. *Int. J. Mass Spectrom. Ion Processes* **165**, 509-521 (1997).

105. Shukla, A. K. & Futrell, J. H. Tandem mass spectrometry: dissociation of ions by collisional activation. *J. Mass Spectrom.* **35**, 1069-1090 (2000).

106. Dongre, A. R., Jones, J. L., Somogyi, A. & Wysocki, V. H. Influence of peptide composition, gas-phase basicity, and chemical modification on fragmentation efficiency: Evidence for the mobile proton model. *J. Am. Chem. Soc.* **118**, 8365-8374 (1996).

107. Bourgogne, E. & Leuthold, L. A. Triple quadrupole linear ion trap mass spectrometer for the analysis of small molecules and macromolecules. *J. Mass Spectrom.* **39**, 845-855 (2004).

108. Le Blanc, J. C. Y. *et al*. Unique scanning capabilities of a new hybrid linear ion trap mass spectrometer (Q TRAP) used for high sensitivity proteomics applications. *Proteomics* **3**, 859-869 (2003).

109. Hager, J. W. & Le Blanc, J. C. Y. Product ion scanning using a Qq-Qlinear ion trap (Q TRAPTM) mass spectrometer. *Rapid Commun. Mass Spectrom.* **17**, 1056-1064 (2003).

110. Millington, D. S., Kodo, N., Terada, N., Roe, D. & Chace, D. H. The analysis of diagnostic markers of genetic disorders in human blood and urine using tandem mass spectrometry with liquid secondary ion mass spectrometry. *Int. J. Mass Spectrom. Ion Processes* **111**, 211-228 (1991).

111. Wolf-Yadlin, A., Hautaniemi, S., Lauffenburger, D. A. & White, F. M. Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. *Proc. Natl. Acad. Sci.* **104**, 5860 (2007).

112. Pan, S. *et al*. Mass spectrometry based targeted protein quantification: methods and applications. *J. Proteome Res.* **8**, 787 (2009).

113. Syka, J. E. P. *et al*. Novel linear quadrupole ion trap/FT mass spectrometer: performance characterization and use in the comparative analysis of histone H3 post-translational modifications. *J. Proteome Res.* **3**, 621-626 (2004).

114. Makarov, A. *et al*. Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal. Chem.* **78**, 2113-2120 (2006).

115. Miller, P. E. & Denton, M. B. The quadrupole mass filter: basic operating concepts. *J.Chem.Education* **63**, 617-622 (1986).

116. Paul, W. Electromagnetic traps for charged and neutral particles. *Rev. Mod. Phys.* **62**, 531-540 (1990).

117. Douglas, D. J., Frank, A. J. & Mao, D. Linear ion traps in mass spectrometry. *Mass Spectrom. Rev.* **24**, 1-29 (2004).

118. March, R. E. An introduction to quadrupole ion trap mass spectrometry. *J. Mass. Spectrom.* **32**, 351-369 (1997).

119. Kaiser Jr, R. E., Cooks, R. G., Moss, J. & Hemberger, P. H. Mass range extension in a quadrupole ion-trap mass spectrometer. *Rapid Commun. Mass Spectrom.* **3**, 50-53 (1989).

120. Hager, J. W. A new linear ion trap mass spectrometer. *Rapid Commun. Mass Spectrom.* **16**, 512-526 (2002).

121. Sparkman, O. D. Mass Spectrometry PittCon® 2005. *J. Am. Soc. Mass Spectrom.* **16**, 967-976 (2005).

122. Agnes, G. R. & Horlick, G. Effect of operating parameters on analyte signals in elemental electrospray mass spectrometry. *Appl. Spectrosc.* **49**, 324-334 (1995).

123. Cleland, W. W. Dithiothreitol, a New Protective Reagent for SH Groups*. *Biochemistry (N. Y.)* **3**, 480-482 (1964).

124. Krell, T., Chackrewarthy, S., Pitt, A. R., Elwell, A. & Coggins, J. R. Chemical modification monitored by electrospray mass spectrometry: A rapid and simple method for identifying and studying functional residues in enzymes. *J. Pept. Res.* **51**, 201-209 (1998).

125. Goddard, D. R. & Michaelis, L. Derivatives of keratin. *J. Biol. Chem.* **112**, 361 (1935).

126. McMurry, J. Organic Chemistry. **1**, 1176-656 (2004).

127. Galvani, M., Hamdan, M., Herbert, B. & Righetti, P. G. Alkylation kinetics of proteins in preparation for two-dimensional maps: a matrix assisted laser desorption/ionization-mass spectrometry investigation. *Electrophoresis* **22**, 2058-2065 (2001).

128. Hedstrom, L. Serine protease mechanism and specificity. *Chem. Rev.* **102**, 4501-4524 (2002).

129. Boeckmann, B. *et al*. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* **31**, 365 (2003).

130. Pappin, D. J. C., Hojrup, P. & Bleasby, A. J. Rapid identification of proteins by peptide-mass fingerprinting. *Curr. Biol.* **3**, 327-332 (1993).

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

132. Alves, P. *et al*. Fast and accurate identification of semi-tryptic peptides in shotgun proteomics. *Bioinformatics* **24**, 102 (2008).

133. Elias, J. E. & Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **4**, 207-214 (2007).

134. http://www.matrixscience.com/help/decoy_help.html.

135. Mischak, H. *et al*. Clinical proteomics: a need to define the field and to begin to set adequate standards. *PROTEOMICS-Clinical Applications* **1**, 148-156 (2007).

136. Nesvizhskii, A. I., Vitek, O. & Aebersold, R. Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nat. Methods* **4**, 787-797 (2007).

137. Wilkins, M. R. *et al*. Guidelines for the next 10 years of proteomics. *Proteomics* **6**, 4-8 (2006).

138. Hernandez, P., Binz, P. A. & Wilkins, M. in *Proteome Research: Concepts, Technology and Application* (eds Wilkins, M. R., Appel, R. D., Williams, K. L. & Hochstrasser, D. F.) 41-67 (Springer-Verlag Berlin, Heidelberg, 2007).

139. Ikai, A. Thermostability and aliphatic index of globular proteins. *J. Biochem.* **88**, 1895-1898 (1980).

140. Gatlin, C. L., Kleemann, G. R., Hays, L. G., Link, A. J. & Yates, J. R. Protein identification at the low femtomole level from silver-stained gels using a new fritless electrospray interface for liquid chromatography–microspray and nanospray mass spectrometry. *Anal. Biochem.* **263**, 93-101 (1998).

141. Stapels, M. D. & Barofsky, D. F. Complementary use of MALDI and ESI for the HPLC-MS/MS analysis of DNA-binding proteins. *Anal. Chem.* **76**, 5423-5430 (2004).

142. Kempe, S., Kestler, H., Lasar, A. & Wirth, T. NF-$# x003ba; B controls the global pro-inflammatory response in endothelial cells: evidence for the regulation of a pro-atherogenic program. *Nucleic Acids Res.* **33**, 5308 (2005).

143. van EEDEN, S. F. *et al*. Cytokines involved in the systemic inflammatory response induced by exposure to particulate matter air pollutants (PM10). *Am. J. Respir. Crit. Care Med.* **164**, 826 (2001).

144. Kardjaputri, A. J. Dose Response Relationship Between The Chemical Composition Of 2.4±0.1 μM Particles And Human Lung Cell (A549) Cytokine Expression. (2009).

145. Dai, Y., Whittal, R. M. & Li, L. Two-layer sample preparation: a method for MALDI-MS analysis of complex peptide and protein mixtures. *Anal. Chem.* **71**, 1087-1091 (1999).

146. Dai, Y., Whittal, R. M. & Li, L. Confocal fluorescence microscopic imaging for investigating the analyte distribution in MALDI matrices. *Anal. Chem.* **68**, 2494-2500 (1996).

147. Hu, X. Improvement of Methodology to Study ICAM-1 and CXCL-5 expressions of Ambient Particles, *in vitro*. (2010).

148. Vliagoftis, H. *et al*. Proteinase-activated receptor-2-mediated matrix metalloproteinase-9 release from airway epithelial cells. *J. Allergy Clin. Immunol.* **106**, 537-545 (2000).

149. Ishida, T., Tsukada, H., Hasegawa, T., Yoshizawa, H. & Gejyo, F. Matrix metalloproteinase-1 activation via plasmin generated on alveolar epithelial cell surfaces. *Lung* **184**, 15-19 (2006).

150. Nufer, O., Corbett, M. & Walz, A. Amino-Terminal Processing of Chemokine ENA-78 Regulates Biological Activity. *Biochem.* **38**, 636-642 (1999).

151. Keshamouni, V. G. *et al*. PPAR-γ activation inhibits angiogenesis by blocking ELR CXC chemokine production in non-small cell lung cancer. *Neoplasia (New York, NY)* **7**, 294 (2005).

152. Forbus, J. *et al*. Functional analysis of the nuclear proteome of human A549 alveolar epithelial cells by HPLC-high resolution 2-D gel electrophoresis. *Proteomics* **6**, 2656-2672 (2006).

153. Huang, L. J. *et al*. Proteomics-based identification of secreted protein dihydrodiol dehydrogenase as a novel serum markers of non-small cell lung cancer. *Lung Cancer* **54**, 87-94 (2006).

154. E. C., Y., Marelli, M. & Lee, H. Approaching complete peroxisome characterization by gas-phase fractionation. *Electrophoresis* **23**, 3 (2002).

155. Resing, K. A. *et al*. Improving reproducibility and sensitivity in identifying human proteins by shotgun proteomics. *Anal. Chem.* **76**, 3556-3568 (2004).

156. Gingras, A. C., Gstaiger, M., Raught, B. & Aebersold, R. Analysis of protein complexes using mass spectrometry. *Nat. Rev. Mol. Cell Biol.* **8**, 645-654 (2007).

157. Cornett, D. S., Reyzer, M. L., Chaurand, P. & Caprioli, R. M. MALDI imaging mass spectrometry: molecular snapshots of biochemical systems. *Nat. Methods* **4**, 828-833 (2007).

158. Kelly, R. T., Tolmachev, A. V., Page, J. S., Tang, K. & Smith, R. D. The ion funnel: Theory, implementations, and applications. *Mass Spectrom. Rev.* **29**, 294-312 (2009).

159. Davis, E. J. A history of single aerosol particle levitation. *Aerosol Science and Technology* **26**, 212-254 (1997).

160. Westphall, M. S., Jorabchi, K. & Smith, L. M. Mass spectrometry of acoustically levitated droplets. *Anal. Chem.* **80**, 5847 (2008).

161. Berggren, W. T., Westphall, M. S. & Smith, L. M. Single-pulse nanoelectrospray ionization. *Anal. Chem.* **74**, 3443-3448 (2002).

162. Giles, K. *et al*. Applications of a travelling wave-based radio-frequency-only stacked ring ion guide. *Rapid Communications in Mass Spectrometry* **18**, 2401-2414 (2004).