THE DEVELOPMENT OF HIGH EFFICIENCY SEPARATION TECHNIQUES FOR THE

PROTEOMIC ANALYSIS OF THE CANINE PROSTATE GLAND AND EXPLORATION OF

THE CANINE AS AN ANIMAL MODEL OF HUMAN PROSTATE CANCER

by

DARRYL LEE JOHNSON

(Under the Direction of Ron Orlando)

ABSTRACT

Mass spectrometry (MS) is the most widely utilized analytical tool for the large-scale study of an organism's proteome. This has created lofty expectations for the field of proteomics; however the study of biological systems continues to be daunting task due to the extreme complexity and wide dynamic range of protein expression. Multidimensional separation techniques have been incorporated into MS-based proteomic workflows to overcome this challenge. Unfortunately, improvements in separating power come at the expense of MS analysis time, thus implementation of highly efficient separation strategies becomes necessary to achieve high throughput.

In this work, we describe the development of high efficiency reversed-phase liquid chromatography (LC) separation methods. The use of superficially porous column packing materials permitted fast LC separations, and optimization of data-dependent acquisition (DDA) parameters allowed for the collection of high quality MS/MS spectra when experiment time was reduced. The use of formic acid and ammonium formate (FA/AF) as a mobile phase modifier was found to be compatible with electrospray ionization, and provided a significant improvement

in peptide separations over formic acid alone. This combination of high efficiency LC separations and optimized DDA parameters lead to a significant reduction in experiment time and substantial increases in proteome coverage.

The efficiency of 1D gel electrophoresis and LC (GeLC) separations were evaluated to determine how to maximize protein identifications in a fixed instrument time format. This work demonstrates that the number of gel slices collected in GeLC analysis has very little impact on protein identifications. The most significant factor is GeLC protein identification efficiency is the percentage of instrument time dedicated to LC gradient elution conditions.

These newly developed, high efficiency GeLC separation methods were applied to the proteomic analysis of the canine prostate gland. Canines and humans are the only two large mammals that spontaneously generate prostate cancer, thereby suggesting a potential predictive model of the human disease. Our work identified several proteins with association to prostate cancer development and progression, suggesting the canine could be a relevant predictive model of androgen-insensitive, highly aggressive human prostate cancer.

INDEX WORDS: Liquid Chromatography, Mass Spectrometry, Tandem Mass Spectrometry, Proteomics, Gel Electrophoresis, Animal Model, Canine, Prostate Cancer

THE DEVELOPMENT OF HIGH EFFICIENCY SEPARATION TECHNIQUES FOR THE

PROTEOMIC ANALYSIS OF THE CANINE PROSTATE GLAND AND EXPLORATION OF

THE CANINE AS AN ANIMAL MODEL OF HUMAN PROSTATE CANCER

by

DARRYL LEE JOHNSON

BS, Piedmont College, 2005

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial

Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2013

THE DEVELOPMENT OF HIGH EFFICIENCY SEPARATION TECHNIQUES FOR THE

PROTEOMIC ANALYSIS OF THE CANINE PROSTATE GLAND AND EXPLORATION OF

THE CANINE AS AN ANIMAL MODEL OF HUMAN PROSTATE CANCER

by

DARRYL LEE JOHNSON

Major Professor:	Ron Orlando
Committee:	Jon Amster
	Jeff Urbauer
	Barry Boyes

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2013

DEDICATION

To my parents: I am forever grateful that you always pushed me to do my best and never allowed me to settle for anything less. You stressed the value of hard work, preached the importance of education and the entire time you said "one day you'll thank me for it"…well you were right, THANK YOU!!!

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

The most widely used analytical tool for the systematic study of a proteome is mass spectrometry (MS). MS-based proteomics has provided a significant upgrade over methods such as 2-dimensional (2D) gel electrophoresis for large-scale protein identification, and has created lofty expectations for biomarker discovery and improved therapeutics for human diseases.[1,2] These outcomes are challenging due to the large sample size (tens to hundreds of thousands) required for the identification and validation of biomarkers and other biologically relevant proteins.[1] To further complicate matters, biological systems are extremely complex and display a very wide dynamic range of protein expression. Therefore, various separation strategies are employed to reduce complexity and increase the dynamic range of protein identifications.[3,4] Unfortunately, increased separating power comes at the expense of analysis time, thus creating a need for high efficiency separation strategies. The aim of this work is to improve proteomic analysis throughput by increasing separation efficiency, and applying these methods to the proteomic analysis of the canine prostate gland, exploring the potential of the canine as an animal model of human prostate cancer.

Chapter 3 investigates the implementation of high speed LC separations to increase the throughput of LC-MS/MS analysis of tryptic digests. Superficially porous particles provide significant improvements in peptide separations, and unlike sub-2 μm diameter particles, operate at back pressures compatible with standard capillary high performance liquid chromatography (HPLC) instruments. At elevated flow rates, a minimal loss in separation efficiency is observed,

allowing for reduction of MS analysis time.[5-8] Initially, a decrease in protein coverage was observed when implementing these faster LC separations, due to data-dependent acquisition (DDA) settings that were not properly set to match the narrow peak widths resulting from newly implemented fast separation techniques. This chapter identifies several problems associated with developing fast LC-MS/MS methods, and uses a systematic approach to establish DDA settings to match the narrow chromatographic peak widths. Analysis of an authentic proteomic sample demonstrates an improvement in protein identifications while reducing experiment time by a factor of 5 with the application of fast LC separations and optimally adjusted DDA settings.

Chapter 4 examines LC mobile phase modifiers to improve separation efficiency, and increase peptide identifications during LC-MS/MS analysis. LC mobile phase modifiers are used to improve peak shape and increase sample load tolerance. Trifluoroacetic acid (TFA) produces peptide separations that are far superior to other additives, but is not compatible with electrospray ionization (ESI) due to signal suppression.[9-11] Therefore, formic acid is commonly used in LC-MS analysis. The drawback to formic acid is the substantial chromatographic band broadening and poor sample load tolerance particularly in the case of basic peptides.[12-14] An alternative mobile phase modifier is the combination of formic acid and ammonium formate which has been shown to improve peptide separations, but its compatibility with ESI has not been investigated.[15] This chapter compares the separation metrics of mobile phases modified with formic acid and formic acid/ammonium formate (FA/AF) and explores the utility of FA/AF for the mass spectrometry analysis of tryptic digests. When incorporated into an LC-MS/MS workflow, FA/AF provided an improvement in on-line RP-LC separations, leading to a significant increase in peptide identifications and improved protein sequence coverage.

Protein identification efficiency is analyzed in Chapter 5, looking at 1D gel electrophoresis and RP-LC (GeLC) separations. The high degree of complexity displayed by biological systems presents a challenge for MS based proteomic analysis, leading to the use of 2D separation techniques such as GeLC.[16] As separating power in either dimension increases, the amount of time required for MS analysis also increases. The aim of this work is to evaluate GeLC separations, particularly to determine how to maximize protein identifications when mass spectrometer analysis time is kept constant. This work evaluates separations with high numbers of 1st dimension fractions and fast LC-MS/MS analysis, few 1st dimension fractions and long LC-MS/MS analysis or with moderate separating power in both dimensions. While gel overlap is widespread and independent of the number of gel slices collected, we show a correlation of proteomic identifications and the percentage of total experiment time allocated for LC gradient elution. Also included in this chapter is a discussion about the measurement of GeLC peak capacity, and analysis of factors that influence proteomic identifications when instrument time is fixed.

Lastly, in Chapter 6 these newly developed high-throughput separation strategies are used for the proteomic analysis of the canine prostate gland. Prostate cancer research is currently hindered by the lack of a relevant animal model; therefore this work explores the potential of the canine as a model of the disease.[17] Canines and humans are the only two large mammals to spontaneously generate prostate cancer, thereby suggesting the potential of this animal model.[18-21] This study examines the protein expression of several histologically different canine prostates. Protein expression was significantly different between the healthy and carcinoma tissue. Numerous proteins were identified that have biological significance to prostate cancer and many proteins that were differentially expressed play an important role in androgen-insensitive, highly

metastatic tumor development and progression. The identification of many suggested therapeutic targets, current targets of anticancer drugs and potential biomarkers demonstrate the relevancy of the canine as an animal model of androgen-independent, highly aggressive human prostate cancer.

CHAPTER 2

LITERATURE REVIEW

From Genomics to Proteomics

The study of an organism's genome is essential for understanding its development and proliferation.[22,23] The Human Genome Project (HGP), initially launched in 1990, set out to identify the genes of the human genome, and generate a highly accurate reference sequence.[24,25] The results of this effort produced whole genome sequencing for humans, as well as a variety of other organisms. With the completion of genomic sequencing, a far more challenging analytical project was presented. Whereas genome sequencing cannot predict where proteins are localized within a cell, nor is it capable of identifying relative expression, specific activity, state of modification, or association with other proteins or biomolecules, the direct study of proteins becomes necessary to answers complex biological questions.[26] First introduced in 1995, the term "proteome" has been used to describe the protein complement of a genome.[27] The process to identify and characterize all proteins expressed within a cell or tissue became known as proteomics.[28]

Proteins are considered to be the most important biomolecule as they are involved in every biological process.[29] The systematic approach to identify all proteins expressed in a proteome is a far more daunting task than identifying all genes within an organism's genome. The genome is much smaller than the proteome, since the same gene can produce multiple protein products as a result of combinatorial splicing and different post translational modifications (PTM). In humans, the genome contains 20,000 to 25,000 genes, while the

proteome is estimated to contain 500,000 to 10.5 million different protein products.[30,31]   To further complicate matters, the proteome of an organism is very dynamic, as protein expression can change based on the state of the organism and environmental factors.[3]   Lastly, unlike genomics, proteomics has no equivalent to polymerase chain reaction (PCR) for amplification.[32]

Mass Spectrometry

Until recently, the powerful analytical tool used to determine molecular mass, better known as mass spectrometry (MS), was exclusively used to analyze small, volatile molecules. MS measures an unknown molecule's mass-to-charge (m/z) ratio, with the requirement that the ionized molecule must be in the gas phase.[33]   Historically, ions were generated via electron impact ionization (EI) or chemical ionization (CI). Due to excessive fragmentation by both EI and CI, analysis of peptides and proteins was not possible.[34]   The development of "soft ionization" techniques in the late 1980's, changed this however, allowing large biomolecules to be analyzed via MS.

Martix-Assisted Laser Desorption Ionization (MALDI)

Matrix-assisted laser desorption ionization (MALDI) is a soft ionization technique utilizing a UV laser to ionize biomolecules that have been co-crystalized in an organic matrix.[35] The matrix has the same absorbance wavelength as the laser, and is usually made of small, acidic compounds that contain an aromatic ring. The matrix protects the analytes from laser irradiation by absorbing most of the energy. During the ionization process, laser pulses induce rapid heating, and a partial charge transfer allows for ionization of the gas phase analyte.[36,37]   For biomolecule ionization, MALDI preferentially ionizes basic and aromatic residues, usually producing singly charged ions.[38,39]

Electrospray Ionization (ESI)

Electrospray ionization (ESI) is another soft ionization technique used for the MS analysis of biomolecules. In this technique, an analyte-containing solvent is passed through a capillary. A high voltage electric field is applied to the capillary at atmospheric pressure, producing charged solvent-analyte droplets. These drops then split due to high charge density, and migrate through a heated capillary. The solvent is then evaporated and a charge transfer produces an ionized analyte. The gas phase ion then travels to the high vacuum mass analyzer for detection.[40] ESI of peptides commonly produces doubly and triply charged ions.[41] This allows for analysis of large biomolecules at relatively low m/z ratios. In the case of proteins, several ionizable sites produce a highly charged analyte. ESI preferentially ionizes hydrophobic peptides, and the production of multiply charged ions allows for fragmentation at lower collisions energies.[42] This technique occurs at atmospheric pressure, allowing for online coupling of liquid chromatography (LC) and MS.

Mass Analyzers

Mass spectrometers consist of three basic components: ionization source, mass analyzer and ion detector.[33] There are several different types of mass analyzers available, each with its own advantages and limitations. For protein studies these various types of mass analyzers can measure the m/z ratio for intact proteins, protein complexes, fragment ions produced by gas phase activation of protein ions (top-down proteomics), peptides from enzymatic and chemical digestion (bottom-up proteomics), and fragmented ions produced by gas phase activation of m/z selected peptide ions (tandem MS or MS/MS).[43]

Quadrupole Mass Analyzer

The quadrupole mass analyzer consists of 4 circular rods set parallel to each other. This analyzer works as a mass filter, selectively isolating ions based on stability of their trajectories in the oscillating electric field applied to the rods. Each opposing rod is connected together and fixed direct current (DC) and alternating radio frequency (RF) potentials are applied between each pair of rods.[44] Only ions of a stable trajectory will be detected. The quadrupole mass analyzer can scan over of a range of m/z ratios by continuously varying the voltage potential, or can be used to select a specific m/z to detect.[45]

The most popular instrument utilizing a quadrupole mass analyzer is the triple quadrupole (QQQ). This instrument consists of 3 separate quadrupoles, with Q1 operating as a mass filter for selection of a precursor ion, Q2 serves as a collision cell to fragment ions via collision induced dissociation (CID) and Q3 is a mass filter for fragment ions. This instrument is known for very high selectivity and sensitivity, but is hindered by low resolution. Although a complete mass spectrum can be obtained by scanning, this significantly reduces sensitivity.[46]

Time of Flight (TOF) Mass Analyzer

The time of flight (TOF) mass analyzer measures m/z ratios from the time required for ions to travel through a high vacuum flight tube. Ions are accelerated through a fixed strength electric field, travel through a flight tube and are detected. Flight time is proportional to the square root of the m/z ratio.[47] This analyzer does not require scanning, so its advantages include speed, sensitivity and a wide mass range. The major drawback to TOF analyzers is the low mass resolution (500 units).[48]

Modern TOF analyzers have significantly improved resolution with the use of delayed extraction (DE) and a reflectron. Following ionization, ions with a higher kinetic energy will

travel faster, so after a short delay an acceleration voltage is applied. This acceleration voltage helps push lower kinetic energy ions, allowing for ions of different kinetic energies to reach the detector plate at the same time.[49] The use of a reflectron also improves mass resolution. A reflectron is an electrostatic field which reflects ions, allowing for charged species of different kinetic energies to reach the detector at the same time. Ions of higher kinetic energy will travel farther in the electrostatic field, while lower kinetic energy ions will travel a shorter distance before reaching the detector. With application of these techniques, a resolving power of 10,000 can be achieved.[50]

TOF analyzers can be stand-alone instruments or can be used in hybrid instruments including the TOF/TOF and the quadrupole-TOF (qTOF). The TOF/TOF is a very fast instrument, capable of very high throughput analysis.[51] The qTOF provides high mass accuracy, but unlike the TOF/TOF, this instrument is not capable of high throughput analysis. In a qTOF, the quadrupole is an ion guide for TOF analysis and in MS/MS, a second quadrupole provides a collision cell to produce fragment ions.[52] TOF analyzers are the most common analyzer for MALDI, and can also be used in combination with ESI.[53,54]

Quadrupole Ion Trap Mass Analyzer

The quadrupole ion trap (IT) is similar to the quadrupole, with the difference being the presence of ring electrode creating a 3 dimensional trap. The IT focuses ions into a small volume, where they become resonantly activated with DC and RF voltages. Helium is present in the trap to reduce the kinetic energy of ions that have been trapped. Scanning the electric field excites ions causing them to become unstable, thus they are ejected from the trap to the detector. Ion accumulation allows the IT to have high sensitivity.[55,56] The other advantage of the IT is since ion accumulation, fragmentation and detection occur in the same location, multi-stage

analysis is possible (MS[N]). This also allows for rapid shifting between MS and MS[N] detection.

The IT has limited resolution and a low ion tapping capacity due to space-charge effects.[57]

The linear ion trap (LTQ) utilizes a 2-dimensional quadrupole field as opposed to the 3-dimensional field used by the tradition IT, which has improved resolution and increased trapping capacity.[58-60] This, combined with fast ion accumulation and detection (cycle time), has allowed for high throughput LC-MS[N] analysis of complex mixtures. The LTQ can also be combined with a high resolution mass analyzer (TOF, FT, Orbitrap) for high accuracy precursor scans, and fast accumulation of MS/MS spectra using the LTQ.

Fourier Transform Ion Cyclotron Resonance (FTICR)

The FTICR analyzer is able to attain very high mass accuracy and resolving power by combining the use of a high magnetic field and vacuum along with cyclotron response to excite and detect ions.[61,62] The FTICR has a mass accuracy of 1-2 ppm and resolving power in excess of 10^5 can be achieved.[63,64] The FTICR is combined with an LTQ for MS/MS analysis, obtaining very high precursor mass accuracy. MS/MS spectra are collected in the LTQ allowing for fast acquisition. However, the FTICR is still limited by the slow acquisition speed required to obtain high mass accuracy and is still subjected to dynamic range limitations of IT analyzers.[65]

Orbitrap

The orbitrap analyzer is also an ion trap, differing in that ions are separated by their oscillations in an electrostatic field.[66,67] Ion packets oscillate around an electrode and produce an image current, which will be converted using FT into frequency spectra. Frequency spectra can then be converted to mass spectra. This allows for very high mass accuracy and resolution, similar to the FTICR, but without the added expense of a superconducting magnet. Similar to the FT-LTQ, the orbitrap is combined with an LTQ to achieve high mass accuracy precursor ion

acquisition and fast MS/MS analysis.[68,69] The limitation of the orbitrap is slow acquisition speed required to obtain high mass accuracy, similar to limitations of the FT-LTQ.[70]

Proteomic Strategies: Top Down vs. Bottom Up

MS analysis of an organism's proteome presents challenges due to the high degree of complexity and wide dynamic range displayed by biological systems. The range of protein expression in humans exceeds 10 orders of magnitude, far exceeding the dynamic range of any available MS.[31,71] These challenges can be addressed with the use of a wide variety of separation techniques. Separations based on chemical and physical properties of peptides and proteins can be combined with MS to increase the depth of proteome analysis. These techniques include fractionation (differential, extraction, centrifugation), chromatographic (reversed-phase, ion exchange, affinity, gel filtration), electrophoretic (1D and 2D) and MS (gas phase fractionation).[16,72-90] The two analysis strategies employed in MS-based proteomics is the top down approach which fragments proteins, and the bottom up approach utilizing an enzymatic or chemical digestion of proteins and peptides are fragmented during analysis.[91,92]

Top-Down Proteomics

The top-down proteomics approach ionizes intact proteins, and gas phase fragmentation/detection provides comprehensive sequence and modification identification.[91] This approach is advantageous for obtaining high protein sequence coverage, and is used for posttranslational modification (PTM) discovery.[93] Development of electron transfer dissociation (ETD) and electron capture dissociation (ECD) have improved analysis, but these techniques require long ion accumulation, activation and detection times.[94-96] Top-down proteomics is typically limited to analysis of proteins smaller than 50 kDa, and separation of proteins is a requirement which can be a challenging task.[91] High mass accuracy and resolution instruments,

such as the LTQ-Orbitrap and the FT-LTQ, are necessary to resolve the highly charged ions and their isotopes.[97-100] Another limitation is the protein dissociation mechanism, which is not well understood; therefore bioinformatic tools are not well developed for top-down analysis.[101-105] Due to the low throughput, this technique is not well suited complete proteome analysis.[97]

Bottom-Up Proteomics

The bottom-up proteomics approach is the most widely used, and is better suited for high throughput analysis. In this technique, proteins are enzymatically or chemically digested, and fragmentation of peptides is conducted during MS analysis. Bottom-up proteomics are divided into two classifications: peptide mass fingerprinting (PMF) and tandem mass spectrometry (MS/MS).[106]

PMF consists of 3 steps: protein separation, chemical or enzymatic digestion and fragmentation of peptides with subsequent detection in the mass spectrometer. In PMF, experimental peptide fragment m/z ratios are compared to in-silico produced theoretical peptide masses via database search algorithms. Statistical analysis is used to evaluate potential peptide and corresponding protein assignments.[107] While this is an effective technique, it requires extensive separation methods. The most common analysis strategy is 2D gel electrophoresis, followed by MALDI-TOF analysis of peptide fragments. 2D gels separate proteins in the 1st dimension based on their isoelectric point (pI) and in the 2nd dimension by size. The limitation of this separation strategy is proteins that are very large, very small or have an extremely high or low pI are not well resolved.[108] 2D gels are also not well suited for membrane protein separation, and difficulties arise when attempting to analyze low abundance species.[45]

The second approach, MS/MS, utilizes a chemical or enzymatic digestion, followed by separation of peptides and MS detection of peptide fragments. During MS detection, precursor

ions are selected for fragmentation via software algorithms, and fragmented via collision induced dissociate (CID). CID is conducted with an inert gas, usually helium or nitrogen, which produces a, b and c or x, y and z ions. Following CID, if the charge is located on the N-terminal side, a, b or c ions are produced. If the charge is present on the C-terminal side x, y or z ions are generated. MS/MS techniques are well suited for high throughput analysis, and are commonly used to analyze complex mixtures. The drawback to this technique is the very high complexity of digests. This often requires multiple separation steps to increase the dynamic range of identifications. Bottom-up proteomics can be further classified into a gel free (shotgun proteomics) or a gel electrophoresis approach (GeLC).[76,109,110]

Gel Free (Shotgun) Proteomics

Shotgun proteomics involves enzymatic or chemical digestion, followed by various peptide separation strategies before MS analysis.[85] Digestion and subsequent separations are carried out in-solution, making this approach well suited for high throughput.[22] Peptide separations are based on various chemical or physical properties.[111] Extensive research on multidimensional separations was conducted by Giddings *et al.*, with the separating power being defined as the product of peak capacity produced in each separation dimension. Orthogonal separation methods are required, that is the separations used must not separation based on the same physicochemical properties.[76,111-113]

The most common separation technique employed with ESI is reversed phase LC (RP-LC), and this has been combined with 1st dimension chromatographic separations including strong cation exchange (SCX), anion exchange (AE), size exclusion (SEC) and affinity (AFC). Separations can be conducted off-line meaning the 1st dimensional separation is not directly coupled to RPLC-MS or on-line with both dimensions of separation occurring during MS

analysis. On-line separations permit fast analysis with minimal sample losses. The most well-known on-line 2D technique, multidimensional protein identification technology (MudPIT) combines SCX and RP-LC, producing very high peak capacities allowing for an extensive increase in the dynamic range of identifications.[76,109,114,115] MudPIT has been used for a variety of applications including extensive analysis of an organism's proteome and subcellular components, quantification and characterization of protein complexes.[116-119]

Gel-based Analysis

Gel electrophoresis combined with LC-MS (GeLC) combines an offline protein separation in an SDS gel followed by in-gel digestion, and the 2nd dimension of separation is combined on-line with MS for peptide analysis. 1D gels separate based on size, and 2D gels separate by isoelectric point in the first dimension, and size in the second dimension.[120,121] Proteins can be visualized within the gel with the use of dyes such as Coomassie Blue or silver. Silver staining has significantly better sensitivity than Coomassie, but it not commonly used with MS due to very low peptide recovery and extensive clean-up process necessary to remove the stain. GeLC has many advantages, as in-gel digestion removes detergents, buffer salts and other non-protein containments which are not compatible with ESI. Therefore, this technique is good for proteins with poor solubility as detergents can be used and removed before MS analysis.[110] 1D gel electrophoresis can also provide long term storage of proteins, without detrimental sample degradation.[122,123]

While the advantages of GeLC are numerous, this technique also suffers a few drawbacks. Peptide yield is considered to be low, but work by Speicher *et al.* (2000), and Fang *et al.* (2010), reported recovery ranging from ~70% - 85%.[124,125] In-gel digestion is time consuming, and extensive sample handling can introduce varies containments. 1D gel

electrophoresis also displays poor reproducibility, and gel slices produce a high degree of overlap, meaning the same protein is identified in multiple fractions.

Reversed-Phase Liquid Chromatography (RP-LC)

RP-LC is the most widely utilized separation technique for ESI-MS.[33] In essence, this technique employs a polar mobile phase and a hydrophobic nonpolar stationary phase to separate peptides based on hydrophobicity. The more hydrophobic a peptide is, the more it will be retained. Silica-bonded phase is the most commonly used stationary phase material, with octadecyl carbon chain (C18)-bonded silica being the most popular for peptide separations.[126] RP separations of peptides employ organic mobile phases, with most proteomic applications utilizing acetonitrile or methanol. Flow rates can range from microliters (<10 µL/min) down to the nanoliter range (< 500 nL/min). The use of nanoliter flow rates can increase sensitivity, but the technique can be difficult due to the very small volumes of solvent flow. These flow rates can be achieved on nanoflow HPLC instruments or standard capillary HPLC systems, but considerations of back pressure from column length, particle size and column inner diameter usually determine which LC systems is used in analysis.[126,127]

RP separations of peptides at most frequently conducted at low pH, where the basic peptides lysine, arginine and histidine are protonated. Addition of a low concentration of acid (mobile phase modifier) provides the low mobile phase pH and aids in peptide solubilization. Mobile phase modifiers are also added to improve peak shape and increase sample load tolerance. While trifluoroacetic acid (TFA) is known for providing superior peptide separations, it is not compatible for ESI-MS due to ion pairing and surface tension effects.[9-11,128] Therefore, acetic acid and formic acid are used for LC-MS analysis, but display poor separation metrics

such as wide peak widths and poor sample load tolerance, especially true for basic analytes such as peptides.[12,13,15]

LC column efficiency is measured by theoretical plates, and is expressed with the height equivalent to a theoretical plate (HETP) which takes column length into consideration. Although columns do not actually contain "plates", the measure of efficiency is marked by either a high number of theoretical plates per meter (N > 30,000 plates/meter) or small plate heights (H < 5 μm) at the optimal mobile phase velocity.[129] The factors that influence column efficiency include column length, particle size, how well a column is packed, linear velocity, instrument efficiency (dwell volume) and retention factor. Column efficiency is commonly described using The Van Deemter equation.[130] HETP, also known as H, is expressed as: $H = A + B/u + C \cdot u$. The A term, Eddy diffusion, describes the different path lengths at which a mobile phase can travel through different flow channels within the column. Since the column is packed with spherical particles, analyte molecules can travel different paths, thus exciting the column at different times producing band dispersion. This is considered a constant as this does not change with respect to changes in linear velocity. The B term, longitudinal diffusion, describes molecular diffusion analytes experience within the column. At low flow rates, this has significant impact on efficiency, but since separations are rarely conducted below the optimal flow rate, this term is usually negligible. The C term, resistance to mass transfer, has the most significant impact on efficiency. During the separation process, the chromatographic system is in dynamic equilibrium. The speed at which an analyte can reach equilibrium between the mobile and stationary phases has significant impact on column efficiency and plate height.[131] Columns displaying large plate heights will display chromatographic band broadening, thus producing lower efficiency separations.

The Van Deemter equation is expressed with a Van Deemter plot, measuring column efficiency at multiple linear velocities. The optimal flow rate is determined when the lowest plate height is measured. Linear velocities below the optimal flow rate will require long analysis time and due to contribution from longitudinal diffusion, the separation efficiency will be low. Most practical applications utilize a flow rate that is two to three times above the optimal flow rate. This allows for faster analysis with a minimal loss in separation efficiency.[129] The size of the stationary phase particle and quality of column packing will determine how much efficiency is lost when operating above the optimal flow rate.

Column packing materials can be porous, meaning the analyte can penetrate completely through the material or superficially porous, composed of a solid core and a porous outer shell. The diameter of particles can range from 5 μm to 1.7 μm. For porous particles, the size influences the efficiency of the separation, with smaller particles producing higher efficiency separations. Sub-2 μm particles produce very high efficiency as the diffusion path length is shorter when compared to 3 μm diameter particles.[6,132,133] While sub-2 particles are known for their high efficiency, they generate high back pressure requiring ultra-high pressure LC instruments.[134] The attractive alternative is the superficially porous particle which has a solid core and porous outer layer. This column packing material provides separations equivalent to sub-2 μm particles, but operates at back pressures allowing for use of standard capillary HPLC instruments. Superficially porous particles have short diffusion path lengths (~0.5 μm), allowing for fast mass transfer. The solid core makes the particle diameter 2.7 μm, which produces a back pressure about half of that produced by sub-2 μm particles.[5,8,135]

The evaluation of separating power in gradient elution chromatography is common assed with peak capacity, measuring the number of theoretical chromatographic peaks that can be

resolved during a given gradient.[136] Peak capacity is determined by taking the difference in retention time from the first and last eluting peptides and dividing by the average peak width. In proteomic analysis, a correlation between peak capacity and proteomic identifications has been reported, highlighting the importance of efficient separation strategies.[137,138]

Peak capacity improvements can be achieved with the use of higher efficiency column packing materials, mobile phase modifiers, increased gradient length and elevated temperature. Chapter 3 examines the use of high efficiency 2.7 μm diameter superficially porous particles. Improvements in peak capacity are demonstrated, and due to the minimal loss in column efficiency at high flow rates, an improvement in throughput can be achieved with fast LC separations. ESI-MS compatible mobile phase modifiers are evaluated in Chapter 4, with the combination of formic acid and ammonium formate improving peak capacity, thus increasing proteomic identifications. While increasing gradient length will increase peak capacity, this will also increase MS analysis time. Therefore, Chapter 5 analyzes GeLC separations to determine how to improve separation efficiency, in turn producing more protein identifications without decreasing throughput.

Data Searching

Bioinformatics plays an essential role in MS-based proteomic analysis. The three main bioinformatic processes used in MS proteomic analysis are conversion of instrument data to a database readable file, database searching and statistical validation. MS/MS analysis can produce thousands to hundreds of thousands of fragmentation spectra, which makes manual interpretation impractical. Therefore, database search programs have been developed to quickly analyze and assign experimental MS/MS spectra to their corresponding peptides.[139-141]

The three most widely used programs, SEQUEST, Mascot and X!Tandem, essentially perform the same process, comparing experimental MS/MS spectra to in-silico derived theoretical MS/MS spectra. A score is then assigned to determine how similar the experimental and theoretical spectra are to each other. In Mascot, this is probability based (Mowse) and SEQUEST utilizes threshold scoring (Xcorr value).[142-144] This is a very effective method for determining protein identifications for very large-scale proteomic analysis. However the drawback to these programs is they cannot identify peptides or proteins that are not included in the database used to create theoretical spectra. Thus, they cannot identify unknown proteins and novel modifications.

Identification of novel sequences and unknown modifications can be done with the use of *De novo* sequencing. Programs such as PEAKS and DenovoX determine potential peptide sequences based on m/z differences in fragmentation spectra. The difference in the m/z ratio between adjacent peaks corresponds to cleavage of a specific amino acid residue.[145,146]

The third step, validation, is necessary since many MS/MS spectra can be assigned to an incorrect peptide, thus producing a false positive identification. Different methods have been applied to filter correct and incorrect peptide identifications based on Xcorr values and Mowse scores. They include p-value, false positive rate, family wise error rate, false negative rate and false discovery rate.[147] Validation represents a critical to step to minimize reporting of false positive protein identifications.[148-151]

Animal Models

MS-based proteomics are commonly applied for the analysis of human disease such as cancer. Comparative analysis between proteins expressed in healthy and disease state cells or tissue can provide insight into disease development and progression. Research aimed at

expanding the understanding of human diseases, as well as identifying potential diagnostic markers and therapeutic targets can often times be difficult and time consuming due to regulatory and ethical issues that arise from studies involving human subjects.[152]

Animal models have become a useful alternative as neoplasms occurring in animals often display many similarities with corresponding human cancers.[153,154] These models have played a valuable role in improving our understanding of various neoplasms including breast, bladder, gastric and pancreatic.[155-159] The mouse has traditionally been the most widely used animal model, but may not be the best candidate as most human tumors arise spontaneously, while in mice most tumors must be induced.[152,160] Canines are becoming a more extensively used model, as approximately 400 inherited diseases similar to those found in humans are characterized in dogs.[161,162] In addition, the canine model offers many other additional advantages including genetic diversity and are exposed to similar environments as humans, age five to eight-fold faster than humans, receive specialized healthcare and spontaneously generate many human diseases; all which make canines a more relevant animal model of human diseases over mice.[152]

Canine Prostate Cancer

The only two large mammals in which spontaneous prostate cancer occurrence is common are humans and canines.[19] Canines have previously been reported as an animal model of prostate cancer, but the relevance to human prostate cancer is still not well defined.[18,20] Many similarities between the two species exist suggesting a potentially useful animal model. The prostate gland in humans and canines share many anatomical and functional similarities. In both species, the prostate is an ovoid-shaped retroperitoneal gland that surrounds the neck of the urinary bladder and proximal urethra, with association to clinical disease appearing to be only in humans, canines and felines.[163] Canines and humans also share similarities associated with

prostate carcinoma development such as increased incidence with age, experience similar clinical symptoms, a lack of hormone responsiveness by the cancer cells in advanced stages and both experience frequent occurrence of osteoblastic bone metastases.[163,164] Therefore, Chapter 6 will examine protein expression of various states of canine prostate to explore the utility of the canine as an animal model of human prostate cancer.

CHAPTER 3

OPTIMIZATION OF DATA DEPENDENT ACQUISITION PARAMETERS FOR COUPLING

HIGH SPEED SEPARATIONS WITH LC-MS/MS FOR PROTEIN IDENTIFICATIONS[1]

_____

Abstract:

Recent developments in chromatography, such as ultrahigh-pressure liquid chromatography (uHPLC) and superficially porous particles, offer significantly improved peptide separation. The narrow peak widths, often only several seconds, can permit a 15 minute LC run to have a similar peak capacity as a 60 minute run using traditional HPLC approaches. In theory these larger peak capacities should provide higher protein coverage and/or more protein identifications when incorporated into a proteomic workflow. We initially observed a decrease in protein coverage when implementing these faster chromatographic approaches, due to data-dependent acquisition (DDA) settings that were not properly set to match the narrow peak widths resulting from newly implemented fast separation techniques. Oversampling of high intensity peptides lead to low protein sequence coverage, and MS/MS spectra from lower intensity peptides were of poor quality because automated MS/MS events were occurring late on chromatographic peaks. These observations led us to optimize data-dependent acquisition (DDA) settings to utilize these fast separations. Optimized DDA settings were applied to the analysis of *T. brucei* peptides, yielding peptide identifications at a rate almost 5 times faster than previously used methodologies. The described approach significantly improves protein identification workflows that use typical available instrumentation.

Introduction:

Mass spectrometry is at the center of many proteomic workflows often times used with online liquid chromatography separation techniques.[33,165] An issue with this approach is the large amount of instrument time required to run traditional liquid chromatography tandem mass spectrometry (LC-MS/MS) experiments. Recent developments in chromatography, such as sub-2 micron particles and superficially porous particles, can offer significantly improved peptide separation.[132,133] High efficiency columns allow increased mobile phase velocities that shorten experimental analysis time, and can provide narrow peak widths, typically a few seconds or less.[166] Combined, these two factors can, permit a 15 minute LC run to have similar peak capacity as a 60 minute run using traditional HPLC approaches.[135,167,168] A drawback to sub-2 micron particles is the high back pressure created by these particles, which exceeds the pressure limit of most available standard HPLC instruments.[134] Superficially porous particles offer similar efficiencies in terms of number of theoretical plates and improved plate height, but operate at about half the back pressures of sub-2 micron particles, which allow for use on a standard capillary LC system.[5,6,8,169] These column packings exhibit rapid solute mass transfer with minimal loss of chromatographic resolution when operated with high flow rates to reduce experimental analysis time.[170] Based on these characteristics and current instrumentation, superficially porous columns were integrated into our lab's proteomic workflow.

Fast separation methodologies were implemented with an expected outcome of increased peptide identifications in a shorter amount of time. However, we initially observed decreased protein sequence coverage when implementing these separation methods. Analysis of peak capacities showed the fast gradient conditions improved separation metrics, such as reduced peak widths which should improve peptide identifications, as has been previously displayed in other

proteomic applications.[137] Detailed inspection of raw LC-MS data revealed that most MS/MS spectra were obtained on peptides of high abundance. This lead to the acquisition of few MS/MS events being performed on peptide ions with low signal intensity, which in turn lead to poor protein sequence coverage. It became apparent that improved chromatographic separations would not improve proteome coverage without enhanced MS/MS acquisition strategies.

The selection of peptide ions for MS/MS analysis is controlled by data-dependent acquisition (DDA) settings, which include repeat count, repeat duration, minimum MS signal and dynamic exclusion. Previous work has shown that proteome coverage can be increased when DDA settings are optimized. Andrews, *et al*. (2011), found that the ionization settings and the number of MS/MS events per cycle were the most important settings associated with improving proteome coverage on an LTQ-Orbitrap XL. They also stated that parameters such as dynamic exclusion and minimum signal counts had minimal influence on proteome coverage. [171] Zang, *et al*. (2009), stated increased dynamic exclusion can improve proteome coverage, however this hinders spectral count quantitation.[172] While these reports provided good information for the basis of our experiments, they did not specifically apply to matching DDA settings with fast separation techniques and materials.

In this current work, we present a process of optimizing DDA settings on an LTQ to match fast LC separations with improved peak capacities, and demonstrate that this combination can provide a large number of peptide identifications in significantly less time.

Materials and Methods:

Data was acquired using an Agilent 1100 Capillary LC system (Palo Alto, CA), with on-line MS detection using a Thermo-Fisher LTQ ion trap (San Jose, CA) fitted with a Michrom (Michrom Bioresources, Auburn, CA) captive spray interface. Gradient delay volume was

reduced by removing the mixing column in the Agilent LC system. These changes allowed for faster separations and re-equilibration of the chromatograph. Measured gradient delay volume determined by analytical column bypass was observed to be 56 µL. Analytical columns were either a 0.2 x 50 mm Halo Peptide ES-C18 capillary column packed with 2.7 µm diameter superficially porous particles (Advanced Materials Technology, Inc., Wilmington, DE), or a 0.2 x 50 mm Magic C18AQ Column packed with 3 µm diameter porous particles (Michrom Bioresources, Auburn, CA). Proteomic sample analysis utilized the LTQ divert valve fitted with either an EXP Stem Trap 2.6 µL cartridge packed with 2.7 µm diameter superficially porous particles (Optimize Technologies, Oregon City, OR) or with a 0.5 x 2 mm CapTrap packed with 3 µm diameter porous particles (Michrom Bioresources, Auburn, CA). Experiments using trap columns were matched for stationary phase materials of the analytical and guard columns. Slow gradient conditions operated at a flow rate of 4 µL/min increasing mobile phase B concentration from 5% to 50% B over 30 minutes with a total experiment time of 90 minutes. Fast gradient conditions operated at a flow rate of 9 µL/min increasing mobile phase B concentration from 5% to 50% B over 12.5 minutes with a total experiment time of 21 minutes. Mobile phases used formic acid and acetonitrile from Sigma-Aldrich (St. Louis, MO).

Tryptic peptides from bovine serum albumin (Michrom Bioresources, Auburn, CA) were used to evaluate mass spectrometry and separation metrics during the evaluation and optimization phase of this work. Peak capacity data was analyzed using Xcalibur software. Peak widths at 50% height were measured by manual inspection with the use of extracted ion chromatograms. Calculations were made from duplicate experiments of 1 pmol injections, measuring both peak widths and retention times. Experiments were conducted without the collection of MS/MS spectra, in order to obtain more MS1 data points to better define each

chromatographic peak. Peak capacities were calculated using two methods, differing in the expression of gradient time (T_g). Equation 1, referred to as actual peak capacity, was calculated using the retention times of the first and last eluting peptides from Table 1. Equation 2, is referred to as theoretical peak capacity, using the actual length of gradient generation (T_g), which is independent of actual peptide retention times. For fast gradient conditions, this T_g value was 750 seconds, while for slow gradient conditions 1800 seconds was used for T_g. Due to the use of multiple stationary phases and gradients conditions in this analysis, peak capacity and proteomic experiments were conducted for each type of stationary phase material under both fast and slow gradient conditions.

Preferred operation conditions were tested for application to proteomic sample analysis using a whole cell lysate from procyclic form *trypanosome brucei*. Soluble proteins from *T. brucei* underwent electrophoresis through a NuPAGE 12% Bis-Tris Gel (Invitrogen, Carlsbad, CA). This step was aimed to remove undesirable non-protein containments, with negligible protein separation, hence the entire lane was treated as a single sample. Each gel lane was cut into 1 x 1 mm squares for digestion. Water was then added to the gel pieces and discarded to waste. Gel pieces were then washed with a mixture of 50% acetonitrile and 50% water, with solution removed as waste after 15 minutes. 100mM ammonium bicarbonate was then added, and after 15 minutes, an equal volume of acetonitrile was added to make a 1:1 (v/v) solution. After incubation at room temperature for 15 minutes, the ammonium bicarbonate/acetonitrile solution was removed to waste. Acetonitrile was again added to the gel slices and incubated at room temperature for 15 minutes. This solution was then removed as waste. A solution of 10 mM dithiothreitol (DTT) in 100 mM ammonium bicarbonate was then added and incubated in a 70°C water bath for 1 hour. The reducing solution was then removed and a 55 mM

iodoacetamide (IDA) in 100 mM ammonium bicarbonate alkylating solution was added. Samples were incubated at room temperature for 1 hour in the dark. The alkylating solution was then removed and 100 mM ammonium bicarbonate was added. After 5 minutes an equal volume of acetonitrile was added to make a 1:1 (v/v) solution. After 15 minutes of incubation, this solution was removed to waste. A solution of 0.1% sequencing grade trypsin (Promega, San Luis Obispo, CA) was made in 50 mM ammonium bicarbonate and added to the gel pieces at 1:50 (w/w) enzyme to protein and incubated overnight at 37°C. The following day peptides were extracted by pooling and saving the solution from gel pieces. A solution of 50% acetonitrile and 0.1% formic acid was then added to the gel pieces. After 15 minutes this solution was extracted, and added to the solution previously pulled from the digested gel slices. *T. brucei* sample analysis was conducted using 2 µg injections preformed in duplicate for each of the 4 gradient/stationary phase conditions.

Raw tandem mass spectra were converted to mzXML files, then into Mascot generic files (MGF) via the Trans-Proteomic Pipeline (Seattle Proteome Center, Seattle, WA). MGF files were searched using Mascot (Matrix Scientific, Boston, MA) against separate target and decoy databases obtained from the National Center for Biotechnology Information (NCBI). The target database contained all *Trypanosome brucei* protein sequences and the decoy database contained the reversed sequences from the target database. Mascot settings were as follows: tryptic enzymatic cleavages allowing for up to 2 missed cleavages, peptide tolerance of 1000 parts-per-million, fragment ion tolerance of 0.6 Da, fixed modification due to carboxyamidomethylation of cysteine (+57 Da), and variable modifications of oxidation of methionine (+16 Da) and deamidation of asparagine or glutamine (+0.98 Da). Mascot files were loaded into ProteoIQ

(NuSep, Bogart, GA), where a 5% false discovery rate was applied for confirmation of protein or peptide identifications.

Results:

Separation Condition Analysis:

The initial experiments were performed to evaluate the impact of fast gradient separations when applied to LC-MS/MS based proteomic experiments. Figure 1a shows a base peak chromatogram for slow gradient conditions along with the timescale of events occurring during the experiment. The narrow window of peptide elution displayed under these conditions highlights a poor efficiency in utilization of instrument time. This inefficiency can in part be attributed to the fixed LC system dwell volume, which when considered with a low flow rate, accounts for almost 15 minutes of experiment time. A high efficiency gradient was developed reducing experiment time from 90 minutes to 21 minutes using a superficially porous particle stationary phase allowing efficient separations at elevated flow rates. This reduction in experiment time was achieved by doubling the flow rate, optimizing gradient conditions, and employing a highly efficient reverse phase column packed with superficially porous particles. The increased flow rates affected the time for all aspects of the experiment, in particular this halved the times required for loading, washing and re-equilibrating of the chromatograph. This gradient included a 10x column volume high organic mobile phase wash and a 10x column volume reequilibration of the chromatograph. Figure 1b displays a base peak chromatogram along with a timescale of experimental events demonstrating the improvements in instrument time utilization. Although a further reduction in experiment time would be beneficial, the instrumentally-fixed gradient delay volume cannot be altered.

The LC resolving power was evaluated by calculating the peak capacity for both gradient conditions using both stationary phase materials. To conduct this assessment, bovine serum albumin tryptic digest was employed due to its high availability and relative simplicity, compared to a proteomic sample. Measured peak capacity allowed for an evaluation of separation conditions dependent upon sample retention times, while theoretical peak capacity allowed for evaluation independent of peptide retention times. Theoretical peak capacity calculations demonstrated the maximum peak capacity obtainable for each gradient/column condition.

Peak capacities showed significant improvement with the use of superficially porous packing material under both gradient conditions examined. This was achieved by substantial reduction in peak widths as superficially porous particles produced peaks widths about half that of porous particles, as shown in Table 2. The use of superficially porous particles operated at elevated flow rates maintains high peak capacity, as shown by comparison of the results for fast and slow conditions. Figure 2 displays an extracted ion chromatogram for peptide QTALVELLK for all four experimental conditions, showing the reduction in peak width and improvement in peak shape using the superficially porous particle stationary phase materials during operation with increased flow rates.

Variation of the chromatographic peak widths were evaluated to assess the reliability of the separation system, as well as to inform the range of peak widths that should be considered for MS and MS/MS parameter optimization. The porous particle column produced standard deviations in the mean peak width about 5 times greater than the superficially porous particle packed columns under identical gradient conditions. The small variance displayed by the

superficially porous particle column confirms conditions effective for the broad range of tryptic digest peptides present in a complex proteomic sample.

Mass Spectral Condition Analysis:

The overarching purpose of the present work was to evaluate the fitness of fast gradient separations for high throughput proteomic analysis. While peak capacity is a useful measure of a separation, the utility of fast gradient separations was strongly dependent upon appropriate mass spectrometer settings. To acquire appropriate MS settings, mass spectral data generated from slow gradient conditions with porous particles was compared to data collected using fast gradient conditions with superficially porous particles. Mass spectrometer settings were evaluated by comparing Mascot scores and protein sequence coverage using 1 pmol injections of BSA tryptic digests. Typical "lab favorite" mass spectrometer instrument settings, utilized with slow gradient conditions and porous particle columns produced a protein score of 10,109 with 82% sequence coverage. Application of fast gradient conditions using superficially porous particles combined with the same mass spectrometer settings produced a protein identification score of 2,664 with 61% sequence coverage. The poor Mascot score and protein sequence coverage from conditions displaying superior separation metrics demonstrated MS/MS acquisitions events were not properly matched to the narrow chromatographic peak widths.

The significant difference in Mascot scores and sequence coverage can be attributed to multiple factors associated with MS/MS acquisition. First, reducing gradient time decreased the number of MS/MS spectra collected during an experiment. Experimentally measured duty cycles produced an average cycle time of approximately 1.5 seconds. The duty cycle includes 1 MS1 acquisition of m/z 300 to 2000, with the top 5 most intense precursor ions subjected to MS/MS analysis. From this data, slow gradient conditions would produce about 6000 MS/MS

spectra during the 30 minute gradient, while fast gradient conditions would produce approximately 2500 MS/MS spectra during the 12.5 minute gradient. Based on these calculations, the number of MS/MS spectra collected for the fast gradient conditions would be about 2.5 times lower than obtained with slow gradient conditions, indicating more efficient MS/MS spectra collection would be necessary to produce similar data.

The significant differences in Mascot scores are also attributed to the high number of redundant MS/MS spectra collected under slow gradient conditions. Poor separation performance conditions resulting in wide peaks and significant peak tailing, which allowed peptide precursor ions to go on and off dynamic exclusion lists multiple times, provided multiple opportunities to collect redundant MS/MS spectra. Slow gradient conditions matched 741 MS/MS spectra to 97 peptides. This produced an average of 7.7 MS/MS spectra per peptide identification. Fast gradient conditions matched 152 MS/MS spectra to 70 peptides, averaging 2.2 spectra per identification. Figure 3a shows the chromatographic peak for peptide LFTFHADICTLPDTEK using slow gradient conditions with an "X" indicating the location where MS/MS acquisition occurred. The majority of MS/MS spectra for this peptide were collected on the tail of the peak, as seen in Figure 3b, which displays the occurrence of MS/MS acquisition and dynamic exclusion on the timescale of this chromatographic peak. The collection of 9 MS/MS spectra on the peak tail provided redundant spectra that increased the protein score. Alternately, fast gradient conditions showed significantly less peak tailing of this peptide, as shown in Figure 4. The result is fewer MS/MS events performed on this peptide, leading to reduction in the protein score. The wide chromatographic peaks associated with the slow gradient conditions lead to the collection of redundant spectra, accounting for more matching MS/MS spectra and higher protein scores.

The poor sequence coverage obtained using fast gradient conditions can be attributed to the reduced experiment time, which decreased the total number of MS/MS events during an LC-MS/MS experiment. The reduced experiment time will require more efficient precursor ion selection for MS/MS acquisition so that low abundance ions are less likely to be missed. For example, the slow gradient conditions produced an average peak width at half height of 25.4 seconds, during which time 17 duty cycles acquiring 85 MS/MS spectra occurred. Fast gradient conditions were not afforded the same luxury. In this case, the average peak widths were 5.1 seconds allowing approximately 3 duty cycles to obtain 15 MS/MS spectra. Fast gradient conditions have smaller windows of opportunity to collect MS/MS spectra on individual peptides during their chromatographic peak elution, which necessitates collection of MS/MS spectra on a wider variety of precursor ions. This can be seen in the fast LC-MS/MS run, 152 MS/MS spectra were matched to 70 peptides, 22 of these peptides were identified with 3 or more spectra. Spectra matching these 22 peptides accounted for 126 of the 152 (83%) matched MS/MS spectra demonstrating the oversampling of high abundance peptides, leading to poor sequence coverage.

Peptides identified with the slow gradient conditions, but not identified using the fast gradient conditions, were examined via extracted ion chromatograms. These missed peptides typically had the following characteristics: low signal intensity, co-elution with high signal intensity peptides, and very narrow peak widths. These issues are demonstrated by an extracted ion chromatogram for peptide NYQEAK, Figure 5a, which elutes in a peak that is 2.7 seconds wide. For this peptide to be identified, this ion would need to be selected in one of the ~10 MS/MS spectra collected during this time. The MS1 spectrum (Figure 5b) acquired during this peak, displays an ion at m/z 752.3, which corresponds to the $M+H^+$ of this peptide, and at least six ions with higher abundance. These other ions would have been subjected to MS/MS analysis

prior to the ion at m/z 752.3, because of their higher abundance. Furthermore, the repeat count setting allowed up to 3 MS/MS spectra to be obtained on these higher abundance ions, consequently the ion at m/z 752.3 was not subjected to MS/MS analysis during the 2 cycles it eluted from the LC. For other low abundance peptides, the presence of co-eluting peptides caused the MS/MS event to occur past the apex of the chromatographic peak, as shown in the extracted ion chromatogram for peptide YICDNQDTISSK and the corresponding MS spectrum for this peak (Figures 6a&b, respectively). This peptide has a wider chromatographic peak than the peptide NYQEAK, discussed above, permitting more MS/MS acquisitions, which in turn allowed the higher abundance peptides to be placed on the dynamic exclusion list. Although this peptide was subjected to MS/MS analysis, the spectrum was of low quality due to its acquisition late on the chromatographic peak. Although signal intensity is not directly factored into a peptide score via database searching, poor signal intensity will produce less distinguishable fragment ions leading to lower scores and decrease the chance of correct identification.

These issues made it apparent that DDA settings needed to be optimized for the sharper chromatographic peaks observed using fast gradient conditions. Specifically, minimum MS signal, repeat count, repeat duration and dynamic exclusion were examined to improve MS/MS acquisition and quality. Minimum MS signal controls the minimum signal intensity required for a precursor ion to be selected for MS/MS acquisition. After a mass to charge ratio is subjected to MS/MS acquisition, repeat count is the number of times that specific mass to charge ratio is eligible for additional MS/MS acquisition. Repeat duration is the length of time the mass to charge ratio is eligible for MS/MS acquisition before it is no longer subjected to further fragmentation spectra. Dynamic exclusion is the amount of time a mass to charge ratio is put on

an exclusion list preventing it from MS/MS fragmentation to allow other mass to charge ratios to be subjected to MS/MS fragmentation.

These experiments were performed using a DDA setting to select the "top 5" most abundant ions for MS/MS analysis. Changes to this setting were not explored in this work. Increasing this DDA setting would allow for the selection of lower abundance ions, but at the expense of increased cycle time. The increase in cycle time was not conducive for fast gradient conditions producing very narrow peak widths. For example, the cycle time increases to over 3 seconds with the selection of the top 9 most abundant ions, which would not allow for multiple cycles to occur over the average 5 second wide chromatographic peaks obtained using these conditions. The reduction of this DDA setting was also not explored due to the reduction in the number of MS/MS spectra that would be collected during each experiment.

Systematic optimization of DDA settings to match peak widths, with decreased numbers of duty cycles acquired will lead to a desired condition of operation. Table 3 presents the DDA setting procedure employed for fast gradient separation conditions and the impact this systematic change had on the number of matched MS/MS spectra, Mascot score and protein sequence coverage. Reduction of dynamic exclusion time to 10 seconds was selected to match the average peak width at base, while maintaining the dynamic exclusion time to peak width ratio used with the slow gradient. This change improved sequence coverage and increased the number of matched MS/MS spectra, however, this also lead to an increased number of redundant MS/MS spectra collected. The next variable investigated was the minimum MS signal, which was decreased from 1500 to 1000. This change further increased the number of matched MS/MS spectra, peptide identifications and improved sequence coverage. Further reductions in the value were not considered useful, with the expectation that MS/MS spectra quality would significantly

suffer. The number of redundant MS/MS spectra was decreased by lowering the repeat count parameter. A value of 1 for this setting was found to produce similar sequence coverage to slow gradient conditions, although the collection of redundant spectra was still prevalent. The repeat duration parameter was also examined, but changes to this value appeared to have insignificant effect on evaluation metrics. A value of 10 seconds for repeat duration was selected to correspond with observed average peak width at base.

The ultimate DDA setting to be optimized was dynamic exclusion. Preferred conditions for repeat count, repeat duration and minimum MS signal were applied for dynamic exclusion data collection to determine a best fit for fast gradient conditions. While longer dynamic exclusion times produced less redundant spectra, fewer peptide identifications were observed, shown in Table 4. Most tryptic peptides carrying a +2 charge state will have a mass-to-charge ratio in the range of 400-700. Long dynamic exclusion times, when combined with the limited resolution of the LTQ mass spectrometer will significantly hamper peptide identifications in a proteomic application, with many peptides displaying a similar mass to charge for a complex tryptic digest. The preferred dynamic exclusion time was 30 seconds. This was selected as approximately 3 times the average peak width at base, thereby providing sufficient time for exclusion of peaks displaying some degree of tailing, while not being too long and thereby cover a significant portion of the short gradient. Although additional dynamic exclusion times could be examined in more detail, the objective was to improve MS/MS spectra collection, not to optimize DDA settings specifically to BSA. This strategy improved MS/MS acquisition for application to fast gradient conditions, permitting reduction of experiment time. Based on preferred MS parameter settings, fast gradient conditions produced similar peptide identifications

and protein sequence coverage as did the previous "typical" conditions, when employed for standard tryptic digest peptides.

Proteomic Sample Analysis:

The preferred DDA settings and separation conditions were evaluated with an authentic proteomic sample mixture, using peptides from a tryptic digest of a whole cell lysate of procyclic *T. brucei* applying all four experimental conditions described above. The superficially porous particles displayed increased peptide and protein identifications, as shown in Table 5. Using experimentally optimized DDA settings, fast gradient conditions with superficially porous particles produced the highest number of peptide and protein identifications. This condition also produced the fewest number of redundant spectra, as shown by the lowest spectra per peptide identification average. An analysis of peptide identifications per minute also confirmed the most efficient use of instrument time was with the use of fast gradient conditions with superficially porous particles, with an improvement of nearly 6 fold relative to previously used separation conditions and instrument settings.

Discussion:

The increase in instrument utilization efficiency using fast separation conditions required optimization of MS instrument DDA settings to acquire good quality MS/MS spectra. Relevant DDA settings included minimum MS signal counts, dynamic exclusion, repeat count and repeat duration. An appropriate scheme to optimize DDA parameters to match the narrow chromatographic peaks resulting from fast separation conditions was uncovered. The change from columns packed with porous particles to columns of superficially porous particles allowed for operation at elevated flow rates while doubling the peak capacity. Analysis of T. *brucei* tryptic digests demonstrated the utility of optimized DDA settings and fast separation techniques.

In the evaluation of optimized conditions, the number of protein identifications doubled, the number of peptide identifications increased by over 50%, and the experiment time was reduced by a factor of 5.

Acknowledgments:

Table 3.1

BSA Tryptic Peptides Used for Chromatographic Analysis and Peak Capacity Calculations

Sequence	Peptide Mass (Da)	Retention Time: Fast Gradient	Retention Time: Slow Gradient
CASIQK	705.8237	6.62	22.37
LVTDLTK	788.9290	7.80	25.87
QTALVELLK	1014.2164	8.92	27.29
EACFAVEGPK	1107.2371	6.59	21.58
CCTESLVNR	1139.2607	8.16	25.45
SLHTLFGDELCK	1419.6016	8.77	27.12
LGEYGFQNALIVR	1479.6783	9.39	28.70
MPCTEDYLSLILNR	1724.9953	11.12	33.25
LFTFHADICTLPDTEK	1908.1352	9.52	29.10
DAIPENLPPLTADFAEDKDVCK	2459.6801	9.63	29.75
EYEATLEECCAKDDPHACYSTVFDK	3039.2431	8.60	27.47
TVMENFVAFVDKCCAADDKEACFAVEGPK	3310.7085	11.20	32.95

[a]Fast Gradient Retention Time provides the average retention time under the fast gradient conditions with the use of the superficially porous particle column.
[b]Slow Gradient Retention Time displays the average retention time from slow gradient conditions using the porous particle column.

Table 3.2

Peak Widths and Peak Capacities Obtained from LC-MS Analysis of Tryptic BSA Peptides Using Each Chromatographic Condition

Phase/ Particle Size	Flow Rate (µL/min)	Gradient Length (min)	Average Peak Width: FWHM (s)	Peak Width σ	Measured Peak Capacity	Theoretical Peak Capacity
Porous 3 µm	4	30	25.35	22.28	16.40	41.77
Porous 3 µm	9	12.5	8.68	4.12	18.71	50.85
S.P.P 2.7 µm	4	30	11.35	5.89	32.62	93.29
S.P.P 2.7 µm	9	12.5	5.08	0.84	32.16	86.93

[a] S.P.P.: superficially porous particles

Table 3.3

Mascot Search Results from LC-MS/MS Analysis of Tryptic Digested BSA Using Different Data-Dependent Acquisition (DDA) Parameters

Column/ Gradient Length	Condition Change	Mascot Score	Sequence Coverage	Matched Spectra	Peptides Identified	Spectra/ Peptide ID
Porous 50mm/90 min	RC = 3 RD = 10 s DX = 60 s Min MS = 1500	10109	82%	741	97	7.7
S.P.P. 50mm/20 min	RC = 3 RD = 10 s DX = 60 s Min MS = 1500	2664	61%	152	60	2.5
S.P.P. 50mm/20 min	RC = 3 RD = 10 s ***DX = 10 s** Min MS = 1500	7322	72%	406	68	6.0
S.P.P. 50mm/20 min	RC = 3 RD = 10 s DX = 60 s ***Min MS = 1000**	7372	78%	469	75	6.3
S.P.P. 50mm/20 min	***RC = 2** RD = 10 s DX = 60 s Min MS = 1500	6083	70%	381	72	5.3
S.P.P. 50mm/20 min	***RC = 1** RD = 10 s DX = 60 s Min MS = 1500	4651	82%	365	85	4.3
S.P.P. 50mm/20 min	RC = 3 ***RD = 5 s** DX = 60 s Min MS = 1500	7065	71%	325	63	5.2
S.P.P. 50mm/20 min	RC = 3 ***RD = 15 s** DX = 60 s Min MS = 1500	6657	73%	347	67	5.2

[a]Experimental conditions used for analysis and metrics from Mascot database searching to assess the impact of each setting. (***** indicates conditional change)
[b]RC = repeat count, RD = repeat duration, DX = dynamic exclusion, Min MS = minimum MS signal
[c]All experiments included: Microscans = 1, Max inject time = 50 ms

Table 3.4

Mascot Search Results from LC-MS/MS Analysis of Tryptic Digested BSA Using Different Dynamic Exclusion Times

Application of Preferred Settings	Condition Change	Mascot Score	Sequence Coverage	Matched Spectra	Peptides Identified	Spectra/Peptide Identification
S.P.P. 50mm/20 min	DX = 15 s	2749	83%	248	83	3.0
S.P.P. 50mm/20 min	DX = 30 s	2845	83%	221	89	2.5
S.P.P. 50mm/20 min	DX = 60 s	1699	82%	152	74	2.1
S.P.P. 50mm/20 min	DX = 120 s	1385	72%	114	64	1.8

[a]Data was obtained using the optimum DDA settings determined in Table 3. Specifically, minimum MS signal 1000, repeat count 1 and repeat duration 10 seconds were applied.

Table 3.5

Proteomic Results from *T. brucei* Analysis under Various Chromatographic Conditions with the Application of Optimized DDA Settings

Column	Flow Rate	Exp. Time	Protein Identifications	Matched Spectra	Peptide Identifications	Spectra/Peptide	Peptide/min
Porous 50mm	4 uL/min	90 min	22	370	121	3.1	2.1
Porous 50mm	9 uL/min	21 min	15	376	86	4.4	9.4
S.P.P. 50mm	4 uL/min	90 min	43	439	139	3.2	2.4
S.P.P. 50mm	9 uL/min	21 min	45	477	185	2.6	11.9

[a]Results from the analysis of soluble proteins from *T. brucei* using all 4 experimental conditions examined. Each condition displays data from duplicate analysis with protein and peptide identifications validated using a 5% false discovery rate.
[b]Peptide/minute reflects the total number of peptides identified for each condition divided by total experiment time for duplicate sample analysis.

Equation 1

Measured Peak Capacity

$$n_{pc} = (t_f - t_i)/W_{4[\sigma]}$$

Equation 2

Theoretical Peak Capacity

$$n_{pc} = T_g/W_{4[\sigma]}$$
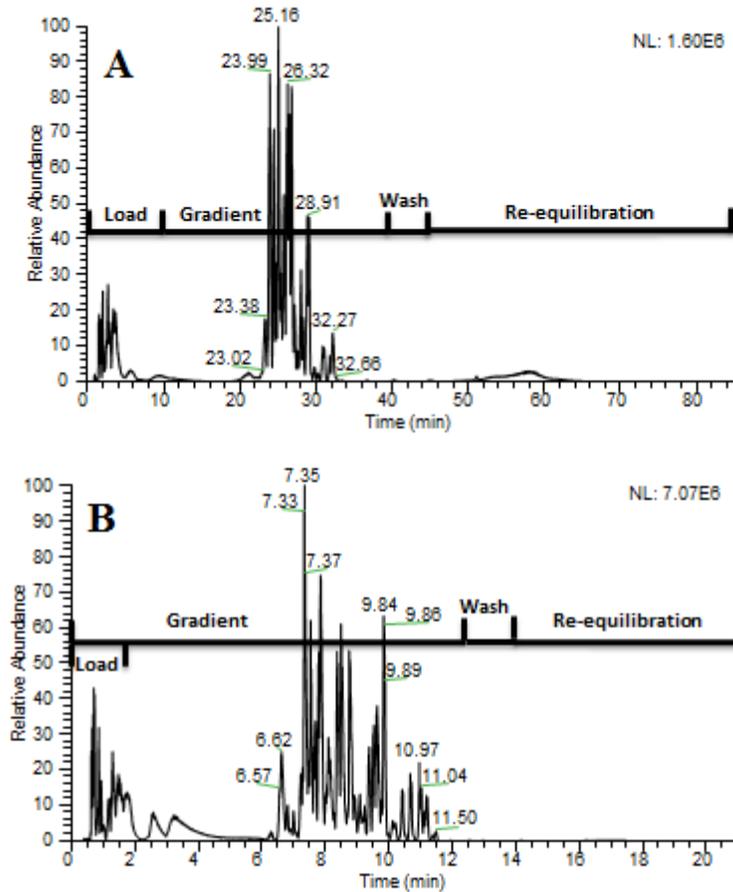
Figure 3.1 Base Peak Chromatograms of BSA Tryptic Peptides. Base peak chromatograms with a superimposed timescale of experimental events from the LC-MS analysis of tryptic digested BSA. (A) Chromatogram obtained using the 90 minute experimental protocol, which included a 0.2 x 50 mm porous particle column operated at 4 μL/min. (B) Chromatogram obtained using the 21 minute experimental protocol, which included a 0.2 x 50 mm superficially porous particle operated at 9 μL/min.
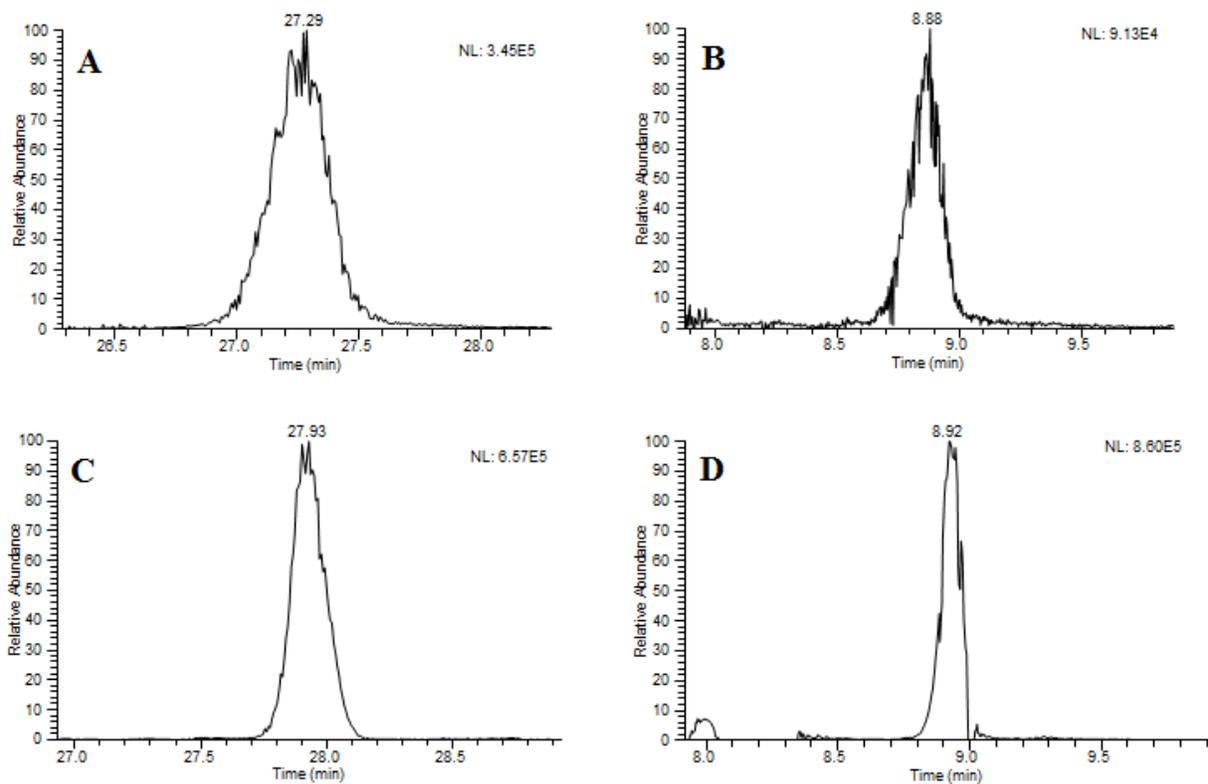
Figure 3.2 Extracted Ion Chromatograms of BSA Tryptic Peptide QTALVELLK.
Extracted ion chromatograms for LC-MS analysis of BSA tryptic peptide QTALVELLK using each of the four chromatographic conditions analyzed. (A) Slow gradient conditions using a 0.2 x 50 mm porous particle column operated at 4 µL/min. (B) Fast gradient conditions using a 0.2 x 50 mm porous particle column operated at 9 µL/min. (C) Slow gradient conditions using a 0.2 x 50 mm superficially porous particle column operated at 4 µL/min. Fast gradient conditions using a 0.2 x 50 mm superficially porous column operated at 9 µL/min.
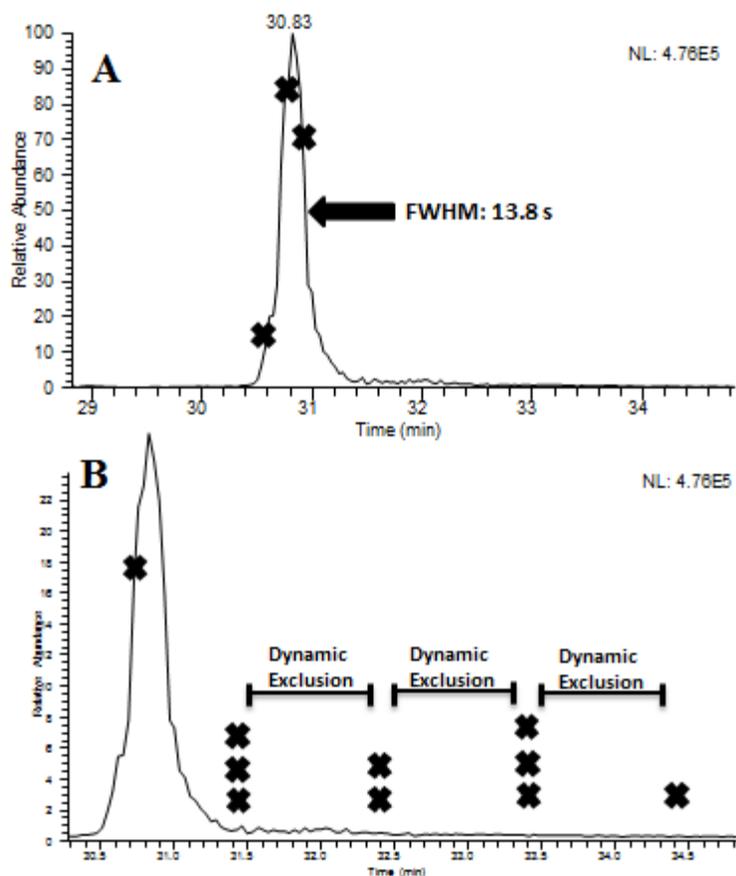
Figure 3.3 Magnified Extracted Ion Chromatogram of BSA Tryptic Peptide LFTFHADICTLPDTEK. (A) An extracted ion chromatogram and (B) a magnification of this chromatogram obtained from the LC-MS/MS analysis of the BSA tryptic peptide LFTFHADICTLPDTEK. Data was collected using slow gradient conditions, which included a 0.2 x 50 mm porous particle column operated at 4 μL/min. An "**X**" on the chromatogram indicates each location an MS/MS spectrum was collected for this peptide. A superimposed timescale for the application of dynamic exclusion is also shown.
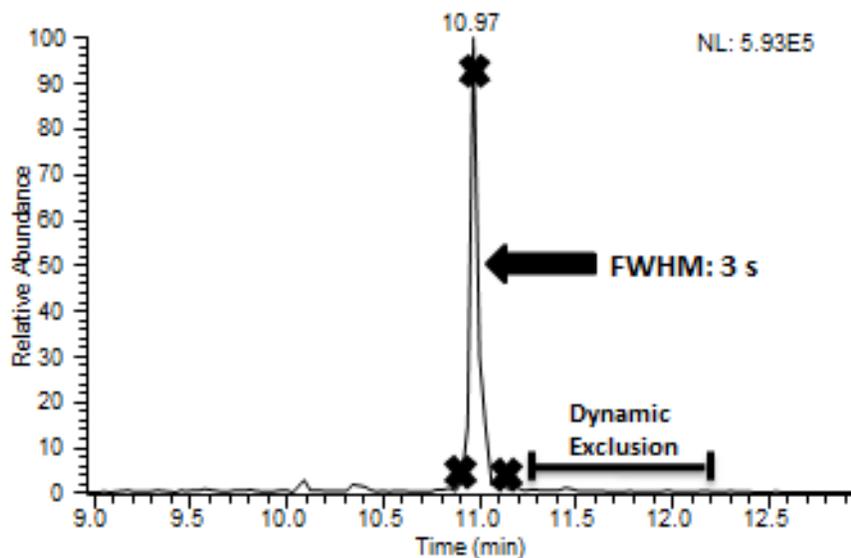
Figure 3.4 Extracted Ion Chromatogram of BSA Tryptic Peptide LFTFHADICTLPDTEK.

An extracted ion chromatogram for the BSA peptide LFTFHADICTLPDTEK using the fast

gradient conditions, which included a 0.2 x 50 mm superficially porous particle column operated

at 9 µL/min. An "**X**" on the chromatogram indicates each location an MS/MS spectrum was

collected for this peptide. A superimposed timescale for the application of dynamic exclusion is

also shown.
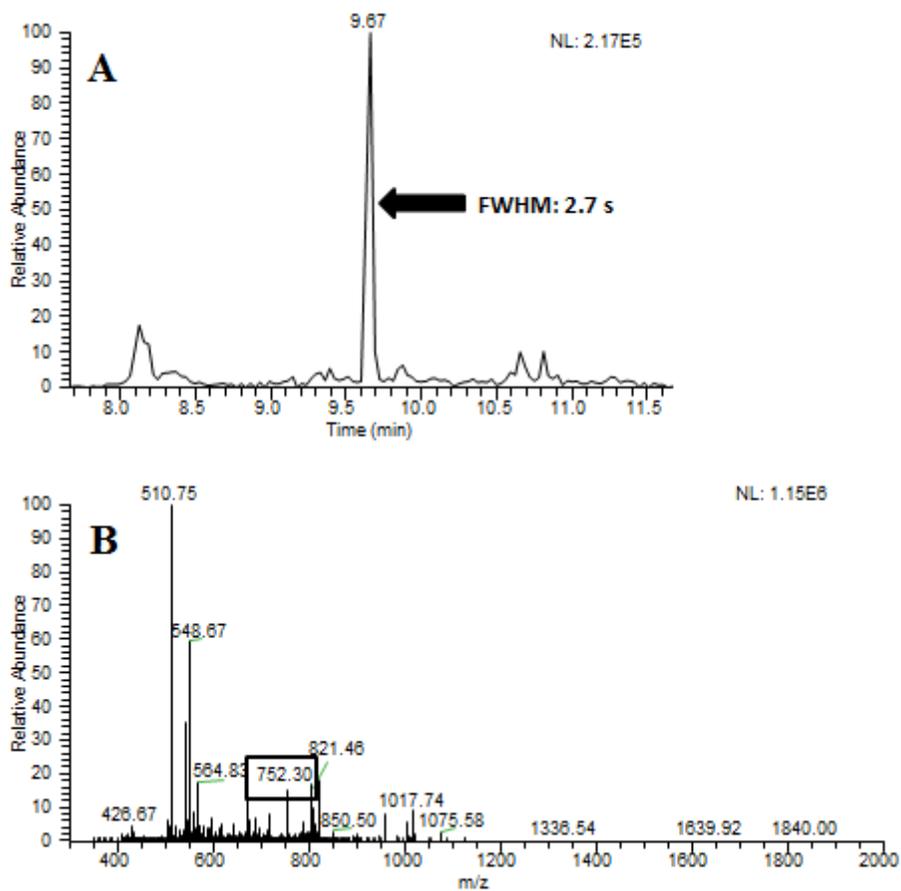
Figure 3.5 Extracted Ion Chromatogram of BSA Tryptic Peptide NYQEAK. (A) The extracted ion chromatogram for m/z 752.30, which corresponds to the BSA peptide NYQEAK, obtained with the fast gradient conditions. The extracted ion chromatogram contains a narrow chromatographic peak for this peptide. This data also shows that no MS/MS spectra were collected on this peptide while it was eluting. (B) A mass spectrum collected during the elution of peptide NYQEAK, showing that other ions, several of which had higher abundance, were present along with the ion of interest at m/z 752.30.
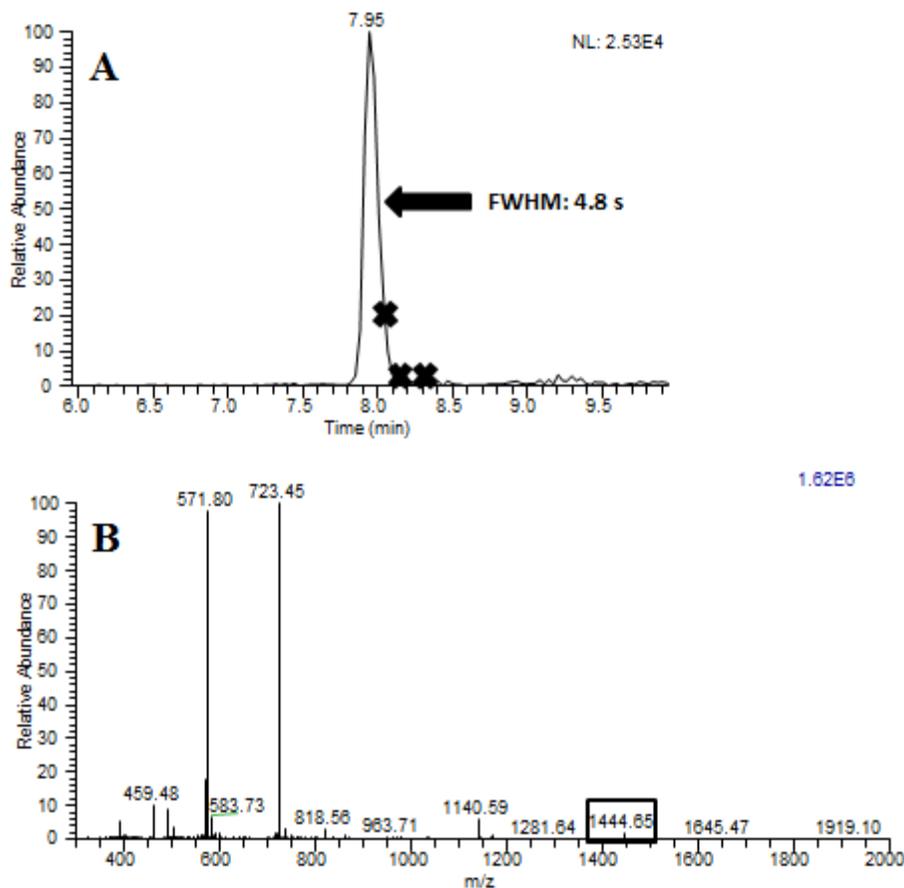
Figure 3.6 Extracted Ion Chromatogram of BSA Tryptic Peptide YICDNQDTISSK. (A) The extracted ion chromatogram for BSA peptide YICDNQDTISSK, obtained with the fast gradient conditions. An "**X**" on the chromatogram indicates each location an MS/MS spectrum was collected for this peptide. (B) A mass spectrum collected during the elution of peptide YICDNQDTISSK, showing that other ions, several of which had higher abundance, were present along with the ion of interest at m/z 1444.65.

CHAPTER 4

THE USE OF AMMONIUM FORMATE AS A MOBILE PHASE MODIFIER FOR LC-

MS/MS ANALYSIS OF TRYPTIC DIGESTS[2]

Abstract:

A major challenge facing current mass spectrometry based proteomics research is the large concentration range displayed in biological systems, which far exceeds the dynamic range of commonly available mass spectrometers. One approach to overcome this limitation is to improve online reversed-phase liquid chromatography (RP-LC) separation methodologies. LC mobile phase modifiers are used to improve peak shape and increase sample load tolerance. Trifluoroacetic acid (TFA) is a commonly used mobile phase modifier because it produces peptide separations that are far superior to other additives. However TFA leads to signal suppression when incorporated with electrospray ionization (ESI), and thus other modifiers, such as formic acid, are used for liquid chromatography mass spectrometry (LC-MS) applications. Formic acid exhibits significantly less signal suppression, however is not as effective a modifier as TFA. An alternative mobile phase modifier is the combination of formic acid and ammonium formate, which has been shown to improve peptide separations. The ESI-MS compatibility of this modifier has not been investigated, particularly for proteomic applications. This work compares the separation metrics of mobile phases modified with formic acid and formic acid/ammonium formate (FA/AF) and explores the utility of FA/AF for the LC-MS analysis of tryptic digests. Standard tryptic digest peptides were used for comparative analysis of peak capacity and sample load tolerance. The compatibility of FA/AF in proteomic applications was examined with the analysis of soluble proteins from canine prostate carcinoma tissue. Overall, the use of FA/AF improved on-line RP-LC separations, and led to significant increases in peptide identifications with improved protein sequence coverage.

Introduction:

A major challenge facing mass spectrometry based proteomics is the large concentration range of proteins expressed by biological systems, which far exceeds the dynamic range of any available mass spectrometer.[173] For example, the concentration range of human serum proteins exceeds 10 orders of magnitude.[174] While there are many techniques available to help improve upon this limitation, there is currently no definitive solution to resolve this challenge.[165,175,176] One way to address the dynamic range limitation is to improve separation methodologies. The most common measure of separation capability in gradient elution chromatography is peak capacity, a metric that has shown correlation to peptide identifications in proteomic applications. It has been demonstrated that improvements in peak capacity lead to increased peptide identifications and enhanced protein sequence coverage.[137,138] The increase in peptide identifications is due to the reduction of co-eluting components, which decreases the number of peptides that compete for ionization. Improved peak capacity can be achieved with several different approaches.

One manner to improve peak capacity involves the use of high efficiency reversed-phase column packing materials. Recent developments in HPLC instrumentation and column packing materials are permitting highly efficient separations with sub-2 μm diameter particles.[6,132,133] These sub-2 μm particles provide high peak capacities, but generate back pressures requiring ultra-high pressure LC instruments.[134] An attractive alternative is superficially porous particles, which show similar separation characteristics as sub-2 μm particles, including small theoretical plate heights and high peak capacities, but operate at back pressures (<350 Bar) compatible for standard capillary LC instruments.[5,8,135]

Another available technique to improve peak capacity is with mobile phase modifiers. Peptide separations using trifluoroacetic acid (TFA) as a modifier are far superior to those obtained when formic acid is used as a modifier. However, TFA is not compatible with electrospray ionization (ESI) mass spectrometry due to signal suppression from ion pairing and spray instability from surface tensions effects.[9-11,128] For this reason, formic acid is probably the most common mobile phase modifier used with liquid chromatography mass spectrometry (LC-MS) based proteomics. Unfortunately, this is a compromise since TFA is the superior modifier, and chromatograms obtained with formic acid exhibit significant band broadening and peak tailing.[6] An explanation of this poor performance is the very low ionic strength exhibited by formic acid solutions (0.2 M formic acid, ionic strength = 1.9 mM) compared to solutions of TFA (0.0079 M TFA, ionic strength = 7.8 mM).[177] Consequently a strategy to improve peptide separations would be the use of a mobile phase modifier with a higher ionic strength. McCalley (2004), reported the addition of 7 mM ammonium formate to formic acid mobile phases nearly quadruples (7.4 mM) ionic strength, while only slightly increasing the pH (Formic acid = 2.7; Formic Acid with ammonium formate = 3.3).[177] Other investigators have shown that mobile phases modified with formic acid and ammonium formate (FA/AF) provide similar separation characteristics to those obtained using mobile phases modified with TFA.[6,178,179]

The addition of ammonium formate to formic acid mobile phases can also improve sample load tolerance. Formic acid has consistently shown high sensitivity to column overloading, especially true for basic analytes.[12,13] Basic analytes, including peptides, exhibit poor peak shape and significant peak tailing due to repulsion effects between ions of the same charge.[177,180,181] Under reversed-phase conditions including low pH, it has been shown that overloading of basic peptides occurs at significantly lower sample loads when compared to

neutral molecules.[182-184] FA/AF has been shown to provide much higher sample load tolerance, similar to TFA, even for basic peptides.[6]

FA/AF's resemblance to TFA in terms of separation characteristics makes it a potential candidate for use as a mobile phase modifier in LC-MS, however no systematic investigation has been conducted evaluating FA/AF's ability to permit the efficient ionization of peptide analytes. Therefore, this work aims to explore the LC-MS/MS compatibility of FA/AF as a mobile phase modifier and assess its impact on chromatographic metrics, sample load tolerance, protein identification and protein sequence coverage.

Materials and Methods:

Data was acquired using an Agilent 1100 Capillary LC system (Palo Alto, CA), with a 0.2 x 50 mm or a 0.2 x 150 mm Halo Peptide ES-C18 capillary column packed with 2.7 μm diameter superficially porous particles (Advanced Materials Technology, Inc., Wilmington, DE). On-line MS detection used a Thermo-Fisher LTQ ion trap (San Jose, CA) with a Michrom (Michrom Bioresources, Auburn, CA) captive spray interface. Proteomic sample analysis utilized the LTQ divert valve fitted with an EXP Stem Trap 2.6 μL cartridge packed with Halo Peptide ES-C18 2.7 μm diameter superficially porous particles (Optimize Technologies, Oregon City, OR). Mobile phases used formic acid, ammonium formate and acetonitrile from Sigma-Aldrich (St. Louis, MO).

Mobile phase modifiers were evaluated using two sets of gradient and column conditions. The short column/fast gradient conditions were conducted using a 0.2 x 50 mm Halo Peptide ES-C18 column operated at a flow rate of 9 μL/min. Gradient conditions increased the concentration of mobile phase B from 5% to 60% over 12.5 minutes. Mobile phase modifiers were also evaluated under long column/slow gradient conditions using a 0.2 x 150 mm Halo

Peptide ES-C18 column operated at 4 µL/min, increasing mobile phase B concentration from 5% to 60% over 90 minutes. Formic Acid mobile phase modifiers (FA): Mobile phase A consisted of 99.9% water and 0.1% formic acid. Mobile phase B contained 99.9% acetonitrile and 0.1% formic acid. Formic Acid and Ammonium Formate mobile phase modifiers (FA/AF): Mobile phase A consisted of 99.9% water, 0.1% formic acid, adjusted to 10 mM ammonium formate (by addition of solid). Mobile phase B consisted of 80% acetonitrile, 0.1% formic acid and 10 mM ammonium formate. Under FA/AF conditions gradient slope was adjusted to account for the change in acetonitrile concentration by increasing mobile phase B concentration from 6.25% to 75% B over 12.5 or 90 minutes. Mass spectrometer settings performed MS/MS analysis on the top 5 most abundant ions for the 21 minute experiment, and top 8 most abundant ions for the 140 min experiment.

Tryptic peptides from bovine serum albumin (Michrom Bioresources, Auburn, CA) were used to evaluate mass spectrometry and separation metrics. All analysis of separation metrics with BSA tryptic peptides were conducted from duplicate experiments of 1 pmol or 8 pmol sample injections. Peak capacity and peak asymmetry measurements were made without the collection of MS/MS spectra, to increase the number of data points collected across each chromatographic peak. Calculation of peak capacity was based on the selection of 14 BSA tryptic peptides listed in Table 1. Theses peptides were selected to provide a wide range of retention for accurate evaluation of both mobile phase modifiers. Calculations of peak capacities were conducted using 2 methods, each with a different measurement of time. Equation 1, measured peak capacity, was calculated using the retention times of the first and last eluting peptides from Table 1. Equation 2, theoretical peak capacity, used the actual length of the gradient (T_g) for retention time, independent of actual peptide retention times. The value for T_g

was either 750 or 5400 seconds. Theoretical peak capacity was calculated to provide a measure of peak capacity independent of sample retention times to demonstrate the maximum peak capacity obtainable for each mobile phase/gradient condition. Peak capacity and peak asymmetry measurements were made using Xcalibur software. Peak asymmetry was calculated using peak widths at 10% peak height, and peak capacity calculations were conducted by measurement of peak widths at 50% peak height via manual inspection with the use of extracted ion chromatograms. Signal intensity measurements were made via extracted ion chromatograms through Xcalibur software.

Proteomic analysis was conducted using proteins extracted from canine prostatic carcinoma tissue. Soluble proteins from canine prostate tissue underwent electrophoresis through a NuPAGE 12% Bis-Tris Gel (Invitrogen, Carlsbad, CA), with the entire lane being removed for composite analysis. This step was aimed to remove undesirable non-protein containments, with negligible protein separation, hence the entire lane was treated as a single sample. Each gel lane was cut into 1 x 1 mm squares for digestion. Water was then added to the gel pieces and discarded to waste. Gel pieces were then washed with a mixture of 50% acetonitrile and 50% water, with solution removed as waste after 15 minutes. 100mM ammonium bicarbonate was then added, and after 15 minutes, an equal volume of acetonitrile was added to make a 1:1 (v/v) solution. After incubation at room temperature for 15 minutes, the ammonium bicarbonate/acetonitrile solution was removed to waste. Acetonitrile was again added to the gel slices and incubated at room temperature for 15 minutes. This solution was then removed as waste. A solution of 10 mM dithiothreitol (DTT) in 100 mM ammonium bicarbonate was then added and incubated in a 65°C water bath for 1 hour. The reducing solution was then removed and a 55 mM iodoacetamide (IDA) in 100 mM ammonium

bicarbonate alkylating solution was added. Samples were incubated at room temperature for 1 hour in the dark. The alkylating solution was then removed and 100 mM ammonium bicarbonate was added. After 5 minutes an equal volume of acetonitrile was added to make a 1:1 (v/v) solution. After 15 minutes of incubation, this solution was removed to waste. A solution of 0.1% sequencing grade trypsin (Promega, San Luis Obispo, CA) was made in 100 mM ammonium bicarbonate and added to the gel pieces at 1:50 (w/w) enzyme to protein; the resulting mixture was digested overnight at 37°C. The following day peptides were extracted by collecting the solution from gel pieces, then a solution of 50% acetonitrile and 0.1% formic acid was then added to the gel pieces. After 15 minutes this solution was extracted, for pooling with the solution previously pulled from the digested gel slices. Canine proteomic analysis was conducted using 1 μg of protein with, triplicate analysis for each mobile phase modifier condition.

Raw tandem mass spectra were converted to mzXML files, then into mascot generic files (MGF) via the Trans-Proteomic Pipeline (Seattle Proteome Center, Seattle, WA). MGF files were searched using Mascot (Matrix Scientific Inc, Boston, MA) against target and decoy National Center for Biotechnology Information (NCBI) databases for canine proteins. Mascot settings were as follows: tryptic enzymatic cleavages allowing for up to 2 missed cleavages, peptide tolerance of 1000 parts-per-million, fragment ion tolerance of 0.6 Da, fixed modification due to carboxyamidomethylation of cysteine (+57 Da), and variable modifications of oxidation of methionine (+16 Da) and deamidation of asparagine or glutamine (+0.98 Da). To examine peptide and protein identifications using a 5% false discovery rate, Mascot .dat target and decoy search files were loaded into ProteoIQ (NuSep, Bogart, GA). Protein identifications were

confirmed with the application of a 5% protein false discovery rate and inspection of peptide MS/MS data.

Results:

Standard Tryptic Peptide Analysis

Peak capacity, the most commonly used assessment of separation power in gradient elution chromatography, was calculated to examine the impact of ammonium formate addition on separation efficiency. In all conditions evaluated, ammonium formate improved peak capacity. For low sample load conditions (1 pmol of BSA), peak widths at half height were reduced by 15% with ammonium formate, compared to formic acid alone, producing average peak widths of 3.6 and 14 seconds for 50 mm and 150 mm columns respectively (Table 2). Ammonium formate also reduced variability of peak widths, evaluated by the improvements in the standard deviations of the peak widths. The decreased peak widths lead to a 30% increase in measured peak capacity for both column lengths analyzed. This reduction in peak width also increased theoretical peak capacity. The 50 mm column obtained a theoretical peak capacity over 120 during a 12.5 minute gradient, and the 150 mm column produced a theoretical peak capacity of almost 230 during a 90 minute gradient. A demonstration of the improved peak shape and reduced peak widths observed with the use of ammonium formate is provided in Figures 1 and 2. Figure 1 displays chromatographic peaks for an early eluting peptide, CASIQK, with and without the presence of ammonium formate in the mobile phase, while Figure 2 compares a late eluting peptide, GLVLIAFSQYLQQCPFDEHVK. As shown in both figures, chromatograms obtained with formic acid and ammonium formate (FA/AF) modified mobile phases contain narrower, more symmetrical peaks.

Formic acid is very sensitive to sample overload, particularly for basic analytes, including peptides.[12,13] The addition of ammonium formate has previously been shown to improve sample load tolerance compared to formic acid alone.[6] To examine sample load tolerance with application to LC-MS, peak capacity experiments were repeated with sample load increased by a factor of 8. Column overload was expected for both mobile phase conditions at this high sample mass, but FA/AF mobile phases showed significantly higher tolerance to sample overloading as displayed by less peak broadening and improved peak capacity compared to the mobile phases containing only FA, shown in Table 3. In FA mobile phases, the average peak width nearly doubled with the increased sample load, while peak widths only increased by about 25% with FA/AF. Figure 3 displays the chromatographic peaks for peptide CASIQK as the amount of sample is increased, and displays the significant improvements in peak widths and peak shape observed with the FA/AF mobile phases. FA/AF again displayed superior peak width variance with standard deviations significantly smaller than those from the FA only condition. The increased sample load decreased peak capacity for both mobile phase modifiers, but FA/AF produced significantly higher peak capacities for each column/gradient condition examined. The high sample mass injections decreased peak capacity by approximately 40% with FA, while FA/AF conditions only experienced an approximately 16% reduction in peak capacity. Direct comparison of peak capacities using 8 pmol BSA data showed FA/AF produced a peak capacity (measured and theoretical) that was nearly double that of FA conditions, regardless of column length. The importance of this observation is significant to the application of proteomics, as typical proteomic samples will display a wide concentration range with high abundance peptides routinely overloading typical analytical columns.

While peak capacity provides a good assessment of separation power, peak widths are measured at half height, which does not provide an assessment of peak shape or peak tailing. To evaluate the impact of ammonium formate on peak shape/tailing, peak asymmetry factors were calculated. Improved peak shape will improve the separation by reducing the number of peptides that co-elute, which in turn will help overcome the dynamic range limitations with mass spectrometers that employ data dependent acquisition strategies (for example see Johnson *et al.*, 2013).[185] Peak asymmetry factors obtained in the analysis of BSA tryptic peptides are presented in Table 1, with measurements conducted at 10% peak height. FA/AF produced better peak shape and decreased peak tailing as seen in Table 4. At a sample load of 1 pmol, FA/AF produced less peak tailing as demonstrated from peak asymmetry factors being closer to 1 when compared to FA conditions. Data also showed improvements in peak asymmetry standard deviation, indicating less variance of peak shape with FA/AF. When high sample mass (8 pmol) was examined, FA/AF significantly improved peak shape, with 150 mm column data showing peak asymmetry factors that were almost 3 times smaller than those produced with FA alone.

The impact of ammonium formate on ionization efficiency and evaluation of signal suppression was also examined during BSA tryptic peptide samples. Ammonium formate increases ionic strength of mobile phases, which should increase ion pairing and thus may lead to ESI signal suppression. Although some degree of signal suppression was expected, decreased peak widths should increase signal intensity due to increased peptide concentration as a result of narrower chromatographic peaks. Initially, 20 mM ammonium formate was explored for LC-MS which showed improved separation metrics, but mass spectral signal intensity was reduced by approximately one order of magnitude for all examined peptides (data not shown). A compromise of 10 mM was chosen, balancing the benefits of improved chromatography without

significant sacrifice of mass spectral signal. Extracted ion chromatograms were examined to compare signal intensity between the two mobile phase modifiers, analyzing BSA peptides identified in Table 1. Minor differences in signal intensities were observed, but these were less than half an order of magnitude for all peptides examined. For some peptides ammonium formate increased MS signal intensities, while in others the presence of ammonium reduced MS signal intensities.   Measured signal intensities of 4 peptides, each containing different combinations of the basic amino acid residues Lysine, Arginine, and Histidine are presented in Table 5 (50 mm column) and Table 6 (150 mm column). The potential ion pairing would occur at these basic amino acid residue sites due to protonation in low mobile phase pH. Since differences in mass spectra signal intensities seemed to even out between the mobile phase conditions, it was concluded that signal suppression would not be detrimental to mass spectrometry analysis.

Proteomic Sample Analysis

The utility of ammonium formate for proteomic analysis was examined using soluble proteins from canine prostate carcinoma tissue samples. Proteins underwent gel electrophoresis to remove non-protein containments, but no electrophoretic gel fractionation occurred, which provided a very complex, with which to compare mobile phase modifiers. The addition of ammonium formate lead to increased numbers of matched MS/MS spectra and peptide identifications, which resulted in an increase in protein identifications, as shown in Table 7. An approximate 35% increase in protein identifications and more than a 50% increase in the number of matched MS/MS spectra were observed with the 50 mm column using FA/AF mobile phases when compared to FA data.   The number of peptide identifications also increased by approximately 30% during triplicate sample analysis. Data collected with a 150 mm column

displayed similar trends seen with the shorter column. With the longer column, the addition of ammonium formate increased the number of protein identifications by almost 70%, and the number of matched MS/MS spectra nearly doubled. The number of peptide identifications also increased by approximately 50% with FA/AF mobile phases. This analysis demonstrates the expected correlation between peptide identifications and peak capacity, since improved peak capacity provided by the use of ammonium formate leads to increased proteomic identifications for both column lengths analyzed. The number of matched MS/MS spectra per peptide identification displayed similar numbers between each mobile phase modifier condition, also confirming that any signal suppression from ammonium formate use was not detrimental to the proteomic sample analysis.

An evaluation of proteomic identifications was conducted from 150 mm column data, specifically comparing identifications and matched MS/MS spectra for the 61 proteins identified in both mobile phase modifier conditions. Analysis of these common identifications showed protein sequence coverage improved with FA/AF, as the average number of peptide identifications per protein increased, as shown in Table 8. The average number of spectral counts per protein identification also increased using FA/AF. Single spectrum protein identifications were analyzed to determine if chromatographic improvements enhanced protein coverage. Data from FA conditions included 3 single spectrum protein identifications, while with FA/AF conditions these same proteins were identified with either multiple peptides and/or multiple MS/MS spectra. FA/AF conditions produced no single spectrum identifications when comparing proteins identified in both mobile phase conditions. The improvement in chromatography not only increased proteomic identifications, but also improved these

identifications as sequence coverage and peptide identifications increased, improving the confidence of these identifications.

The amino acid composition of identified peptides was examined, specifically focusing on peptides with multiple basic amino acid residues since formic acid has traditionally been reported as a poor mobile phase modifier for basic analytes. FA/AF conditions identified significantly more peptides with multiple basic amino acid sites, indicating the improvements achieved in the separation of these peptides resulted in increased proteomic identifications. After application of a 5% peptide false discovery rate, 150 mm column data produced 150 peptides commonly identified using both mobile phase conditions. Of those common identifications, 48% contained 2 or more basic amino acid residues. In FA conditions, 209 peptides were uniquely identified with 49% containing 2 or more basic amino acid residues. FA/AF identified 388 condition unique peptides, with 72% of those containing multiple basic amino acid residues, shown in Figure 4. While this increase in peptide identifications can be attributed to multiple factors, it appears the improved sample load tolerance from increased ionic strength with the ammonium formate allowed for better chromatographic behavior and separation of peptides with multiple basic amino acid residues.

The increase in ionic strength displayed by FA/AF mobile phases has the potential to increase ion pairing, which in turn could reduce peptide charge state. To examine this, peptide charge states were analyzed using the charge state provided from Mascot database identifications. Figure 5 displays the charge state of all identified peptides, 359 in FA and 538 in FA/AF. As expected, most peptides were identified in a +2 charge state. Only small differences were observed, with FA displaying a slightly higher percentage of peptides identified in a +3 charge state than FA/AF. For a direct comparison, the 150 peptides commonly identified in both

mobile phase conditions were examined, with almost 90% of these peptides identified in the same charge state, Figure 6. Of the 16 peptides identified in a different charge state between the two mobile phase modifiers, neither mobile phase condition consistently produced identification in a higher charge state. Further analysis of these 150 common peptides showed 72 peptides containing 2 or more basic amino acid residues. Of these 72 peptides, 83% were identified from the same charge state, as shown in Figure 7. Among the 12 peptides with multiple basic residues, FA produced a higher charge state in 8 peptides, while FA/AF produced the higher identified charge state in 4 peptides. Analysis of peptide charge states showed very little difference between the two mobile phase conditions, indicating ammonium formate did not significantly reduce the charge state in the peptides identified, nor was this mobile phase modifier found to be detrimental to proteomic analysis.

Conclusions:

The use of ammonium formate along with formic acid as a mobile phase modifier was found to be compatible with LC-MS/MS analysis of peptides. Peak widths were reduced by approximately 15% leading to an increase of approximately 30% in peak capacity when compared to FA modified mobile phases. Significant improvements in peak shape were displayed by FA/AF when peak asymmetry factors were compared. Although some degree of signal suppression occurs with FA/AF, mass spectral signal intensities were comparable, demonstrating signal suppression was not detrimental for peptide analysis.

Analysis of soluble proteins from canine prostate tissue displayed significant increases in peptide identifications and protein sequence coverage with ammonium formate addition. The increase in peptide identifications can be directly attributed to improvement in separation metrics displayed with BSA peptide analysis. FA/AF showed better retention and separation of basic

nature peptides, with improved identifications of peptides composed of multiple basic residues. Analysis of identified charge state showed no difference between FA/AF and FA modified mobile phases. An analysis of matched spectra and peptide analysis also confirmed the degree of signal suppression observed in standard peptide analysis was not detrimental to peptide identifications. The high number of peptide identifications as well as improve matched spectra per peptide identification displayed with FA/AF confirmed similar ionization efficiency between both mobile phase modifiers.

Acknowledgments:

Table 4.1

BSA Tryptic Peptides Used for Chromatographic Analysis and Peak Capacity Calculations

Sequence	Peptide Mass (Da)
CASIQK	705.8237
LVTDLTK	788.9290
QTALVELLK	1014.2164
EACFAVEGPK	1107.2371
KQTALVELLK	1142.3887
SLHTLFGDELCK	1419.6016
YICDNQDTISSK	1443.5353
LGEYGFQNALIVR	1479.6783
MPCTEDYLSLILNR	1724.9953
LFTFHADICTLPDTEK	1908.1352
DAIPENLPPLTADFAEDKDVCK	2459.6801
GLVLIAFSQYLQQCPFDEHVK	2492.8454
QEPERNECFLSHKDDSPDLPK	2541.7044
TVMENFVAFVDKCCAADDKEACFAVEGPK	3310.7085

Table 4.2

Peak Widths and Peak Capacities Obtained from LC-MS Analysis of 1 pmol Injections of Tryptic BSA Peptides Using Various Chromatographic Conditions with Each Mobile Phase Modifier

Column Length (mm)	Flow Rate (µL/min)	Gradient Length (min)	Mobile Phase Modifier	AVG Peak Width (s)	Standard Deviation	Measured Peak Capacity	Theoretical Peak Capacity
50	9	12.5	0.1% FA	4.14	0.94	35.21	106.56
50	9	12.5	0.1% FA, 10mM AF	3.56	0.69	46.60	123.93
150	4	90	0.1% FA	16.11	2.98	76.30	197.17
150	4	90	0.1% FA, 10mM AF	13.99	2.35	99.55	227.05

Table 4.3

Peak Widths and Peak Capacities Obtained from LC-MS Analysis of 8 pmol Injections of Tryptic BSA Peptides Using Various Chromatographic Conditions with Each Mobile Phase Modifier

Column Length (mm)	Flow Rate (µL/min)	Gradient Length (min)	Mobile Phase Modifier	AVG Peak Width (s)	Standard Deviation	Peak Width Change[a]	Measured Peak Capacity: 8pmol	Actual Peak Capacity Change[b]	Theoretical Peak Capacity: 8pmol	Theoretical Peak Capacity Change[b]
50	9	12.5	0.1% FA	7.78	4.57	+87.9%	21.19	-39.8%	56.71	-46.8%
50	9	12.5	0.1% FA, 10mM AF	4.43	1.21	+24.4%	38.72	-16.9%	99.59	-19.6%
150	4	90	0.1% FA	29.1	16.11	+80.6%	44.18	-42.1%	109.16	-44.6%
150	4	90	0.1% FA, 10mM AF	16.82	2.64	+20.2%	83.54	-16.0%	188.85	-16.8%

[a] Calculated by taking the difference in average peak width from 1 pmol and 8 pmol BSA tryptic peptide injections, and expressing as a percentage of the average peak width observed using 1 pmol BSA injections.

b Calculated by taking the difference in peak capacity from 1 pmol and 8 pmol BSA tryptic peptide injections, and expressing as a percentage of the peak capacity measured using 1 pmol BSA injections.

Table 4.4

Peak Asymmetry Obtained from LC-MS Analysis of 1 and 8 pmol Injections of Tryptic BSA Peptides Using Various Chromatographic Conditions with Each Mobile Phase Modifier

Column Length (mm)	Flow Rate (μL/min)	Gradient Length (min)	Mobile Phase Modifier	Peak Asymmetry: 1 pmol BSA	Standard Deviation	Peak Asymmetry: 8 pmol BSA	Standard Deviation
50	9	12.5	0.1% FA	1.49	0.88	2.14	1.48
50	9	12.5	0.1% FA, 10mM AF	1.22	0.43	1.63	0.82
150	4	90	0.1% FA	2.24	1.25	6.08	4.48
150	4	90	0.1% FA, 10mM AF	1.18	0.40	2.18	1.67

Table 4.5

The Extracted Ion Chromatogram Signal Intensity from LC-MS Analysis of BSA Tryptic Peptides Using a 0.2 x 50mm Column

Sequence	FA: 1pmol BSA	FA/AF: 1 pmol BSA	FA: 8 pmol BSA	FA/AF: 8 pmol BSA
QTALVELL**K**	6.03e4	2.67e5	3.61e5	3.09e5
SL**H**TLFGDELC**K**	4.78e5	6.70e5	6.94e5	3.26e5
MPCTEDYLSLILN**R**	9.98e4	3.10e4	6.55e5	1.80e5
DAIPENLPPLTADFAED**K**DVC**K**	5.21e5	2.21e5	2.03e6	9.95e5

Table 4.6

The Extracted Ion Chromatogram Signal Intensity from LC-MS Analysis of BSA Tryptic Peptides Using a 0.2 x 150mm Column

Sequence	FA: 1pmol BSA	FA/AF: 1 pmol BSA	FA: 8 pmol BSA	FA/AF: 8 pmol BSA
QTALVELL**K**	2.64e4	4.20e4	2.71e5	3.68e5
SL**H**TLFGDELC**K**	5.54e4	4.52e4	3.29e5	3.47e5
MPCTEDYLSLILN**R**	5.04e4	1.46e4	5.09e5	5.33e5
DAIPENLPPLTADFAED**K**DVC**K**	1.14e5	9.29e4	9.78e5	7.22e5

Table 4.7

Proteomic Results from Canine Prostate Carcinoma Analysis Under Various Chromatographic Conditions for Each Mobile Phase Modifier

Column Length (mm)	Flow Rate (μL/min)	Experiment Time (min)	Mobile Phase Modifier	Protein IDs[a]	Matched MS/MS Spectra	Peptide IDs[a]	Spectra/Peptide ID[b]
50	9	21	0.1% FA	44	455	196	2.32
50	9	21	0.1% FA, 10mM AF	60	697	255	2.73
150	4	140	0.1% FA	70	1142	359	3.18
150	4	140	0.1% FA, 10mM AF	118	2028	538	3.77

[a] Results for each mobile phase modifier generated from duplicate sample analysis with protein and peptide identifications validated using a 5% false discovery rate.

[b] Reflects the total number of database matched MS/MS spectra divided by the total number of peptide identifications for each condition from duplicate sample analysis.

Table 4.8

The Analysis of the 61 Proteins Commonly Identified Using Both Mobile Phase Modifier Conditions from LC-MS/MS Analysis Canine Prostate Carcinoma using a 0.2 X 150 mm Column

Mobile Phase Modifier	Average Peptide IDs/Protein[a]	Average Spectral Count/Protein ID[b]	Single Spectrum Protein IDs[c]
0.1% FA	6.60	20.71	3
0.1% FA, 10mM AF	9.64	28.56	0

[a]The number of peptides identified from the 61 common identification proteins divided by the number of common protein identifications.
[b]The total number of database matched MS/MS spectra from the 61 common identification proteins divided by total of common protein identifications.
[c]Protein identifications from only a single MS/MS spectra after application of a 5% false discovery rate.

Equation 1

Measured Peak Capacity

$n_{pc} = (t_f - t_i)/W_{4[\sigma]}$

Equation 2

Theoretical Peak Capacity

$n_{pc} = T_g/W_{4[\sigma]}$
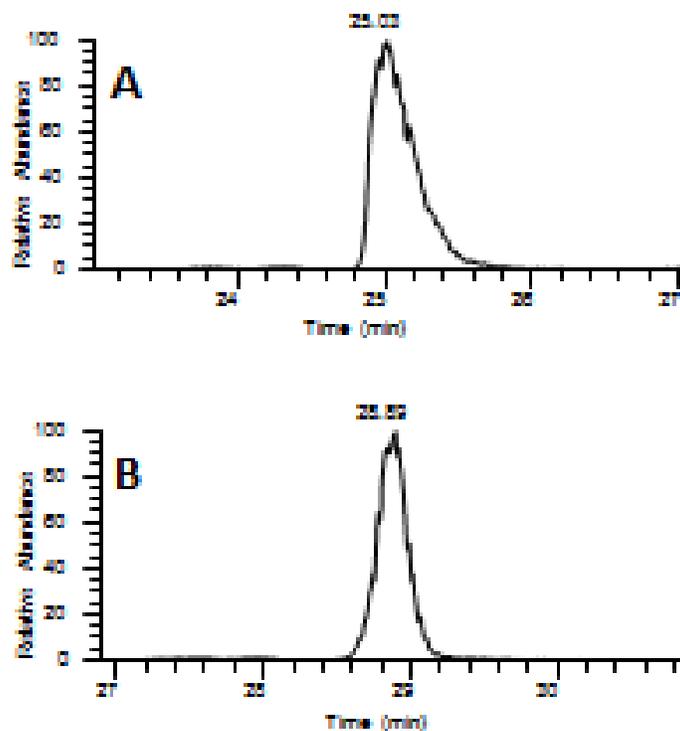
Figure 4.1 Extracted Ion Chromatograms of BSA Tryptic Peptide CASIQK. The Extracted

Ion Chromatograms of the peptide CASIQK, which eluted early in the 90 minute gradient, from

LC-MS analysis of 1 pmol of tryptic digested BSA. Data was collected using a 0.2 x 150 mm

column. Extracted Ion Chromatogram (A) was obtained using only formic acid as the mobile

phase modifier, while Extracted Ion Chromatogram (B) was obtained using formic acid and 10

mM ammonium formate as the mobile phase modifier.
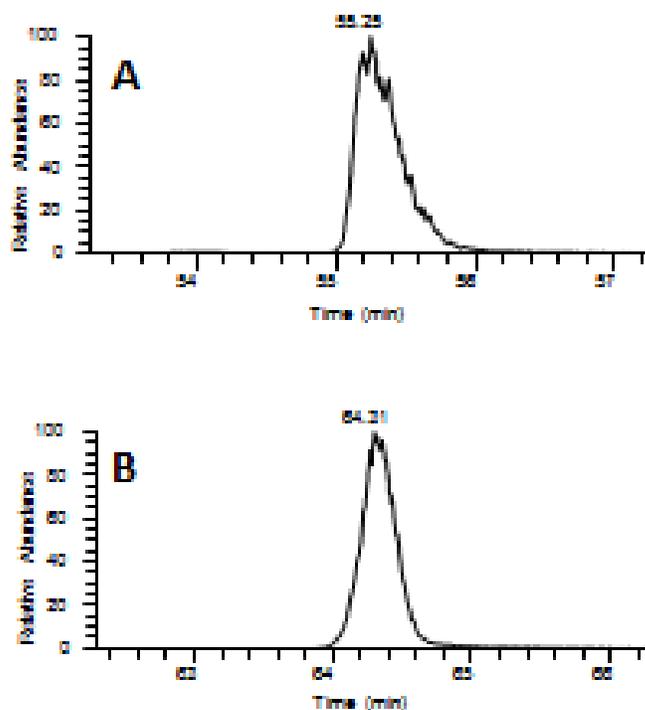
Figure 4.2 Extracted Ion Chromatograms of BSA Tryptic Peptide GLVLIAFSQYLQQCPFDEHVK. The Extracted Ion Chromatograms of peptide GLVLIAFSQYLQQCPFDEHVK, which eluted late in the 90 minute gradient, from LC-MS analysis of 1 pmol of tryptic digested BSA. Data was collected using a 0.2 x 150 mm column. (A) Extracted Ion Chromatogram obtained using only formic acid as the mobile phase modifier. (B) Extracted Ion Chromatogram obtained using formic acid and 10 mM ammonium formate as the mobile phase modifier.
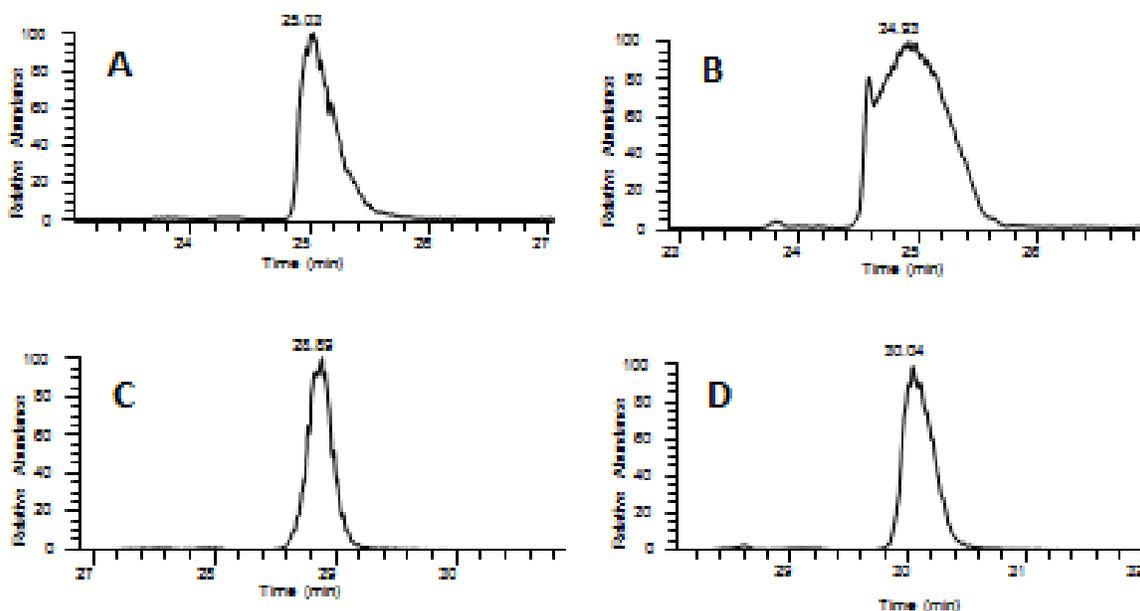
Figure 4.3 Extracted Ion Chromatograms of BSA Tryptic Peptide CASIQK. A comparison of Extracted Ion Chromatograms of the peptide CASIQK as sample load is increased by a factor of 8. Data was collected from LC-MS analysis of tryptic digested BSA using a 0.2 x 150 mm column. (A) Extracted Ion Chromatogram from 1 pmol of BSA with only formic acid as the mobile phase modifier. (B) Extracted Ion Chromatogram from 8 pmol of BSA with only formic acid as the mobile phase modifier. (C) Extracted Ion Chromatogram from 1 pmol of BSA with formic acid and 10 mM ammonium formate as the mobile phase modifier. (D) Extracted Ion Chromatogram from 8 pmol of BSA with formic acid and 10 mM ammonium formate as the mobile phase modifier.
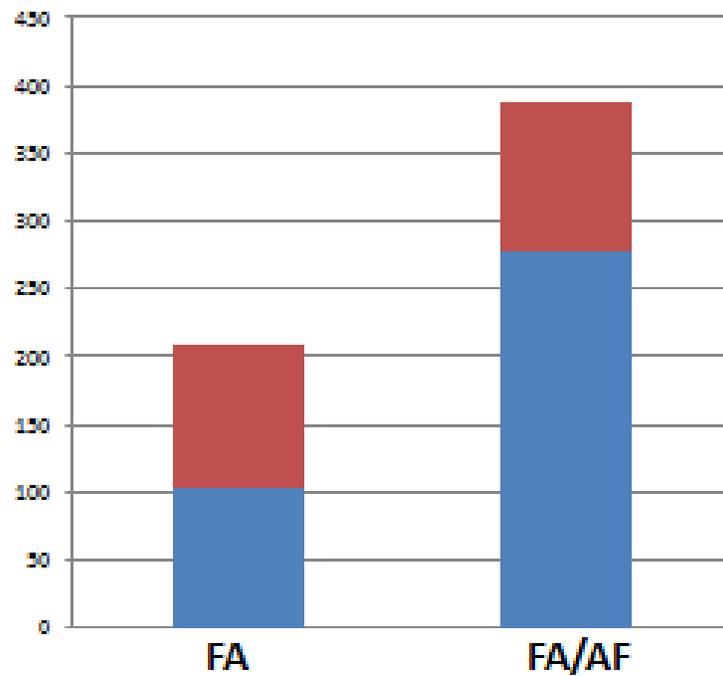
Figure 4.4 Analysis of the Number of Basic Amino Acid Residues. For all peptides not identified in both mobile phase modifier conditions, the amino acid composition was examined. Peptide identifications unique to formic acid only mobile phases are displayed on the left, and peptide identifications unique to FA/AF mobile phases are shown on the right. Peptides containing 2 or more basic amino acid residues are represented in blue, and peptides containing only one basic amino acid are displayed in red. Data was collected during triplicate LC-MS/MS analysis of canine prostate carcinoma using a 0.2 x 150 mm column.

Figure 4.5 Peptide Charge State. The charge state of all peptides identified during triplicate analysis of canine prostate carcinoma was investigated. Results display similar charge state identifications between the two mobile phase modifiers. Peptides identified in the +1 charge state are shown in blue, +2 in red and +3 in green. Data was collected using a 0.2 x 150 mm column.
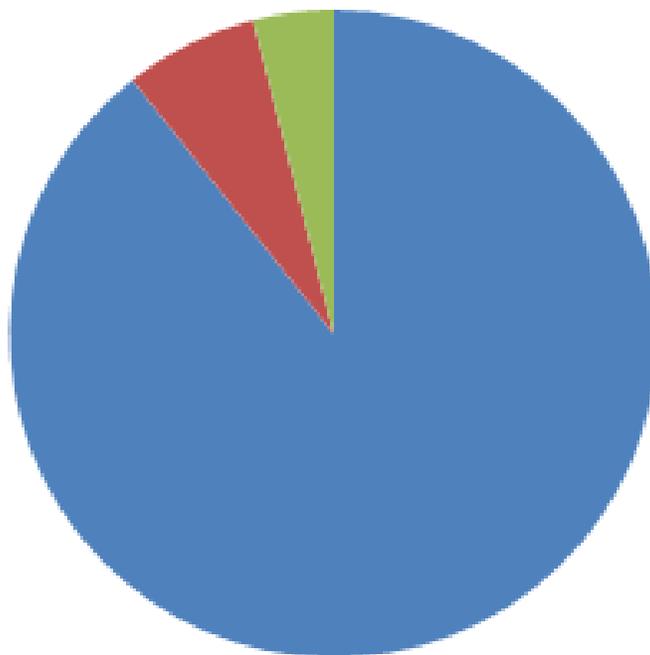
Figure 4.6 Analysis of Charge State for Peptides Identified by Both Mobile Phase Modifiers. An analysis of peptide charge state was conducted for the 150 peptides identified in both mobile phase modifier conditions from triplicate analysis of canine prostate carcinoma. Peptides identified in the same charge state are shown in blue. Peptides identified in a higher charge state in formic acid only conditions are shown in red, and in a higher charge state in formic acid/10 mM ammonium formate are shown in green. Data was collected using a 0.2 x 150 mm column.
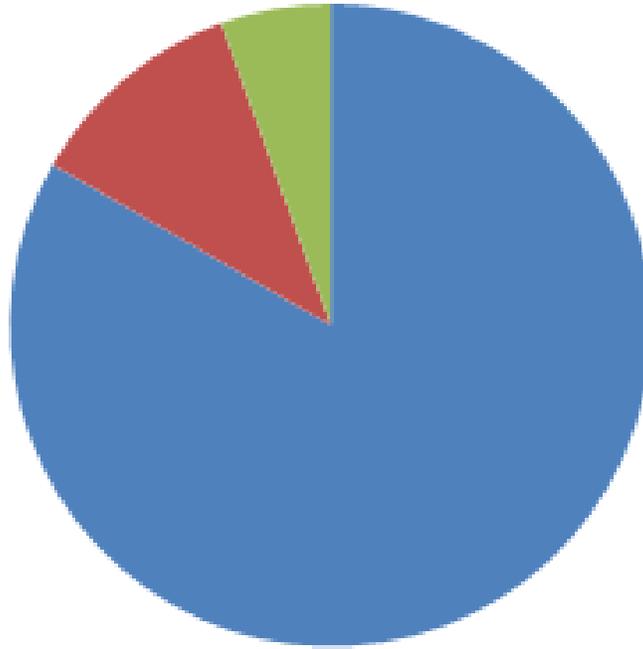
Figure 4.7 Charge State Analysis of Peptides Identified by Both Mobile Phase Modifiers Containing Multiple Basic Amino Acid Residues. Comparison of charge state was conducted for the 72 peptides that consisted of two or more basic amino acid residues and were identified in both mobile phase modifier conditions. Peptides identified in a higher charge state in formic acid only conditions are shown in red and, and in a higher charge state in FA/AF are displayed in green. Peptides identified in the same charge state are shown in blue. Data collected during triplicate analysis of canine prostate carcinoma using a 0.2 x 150 mm column.

CHAPTER 5

GELC SEPARATIONS: WHICH DIMENSION PLAYS THE MOST IMPORTANT ROLE IN

PROTEIN IDENTIFICATION EFFICIENCY?[3]

_____

[3]Johnson, D., Boyes, B., Orlando, R. To be submitted to *Journal of Biomolecular Techniques: JBT*.

Abstract:

The high degree of complexity displayed by biological systems presents a challenge for mass spectrometry based proteomic analysis, leading to the use of 2-dimensional (2D) separation techniques such as gel electrophoresis and reversed phase liquid chromatography (GeLC). Improvements in separation efficiency can increase the dynamic range of protein identifications. Unfortunately, as separation efficiency in either dimension increases, the amount of time required for mass spectrometry analysis also increases. Since instrument time is a practical limitation, the aim of this work is to evaluate GeLC separations, to determine how to maximize protein identifications when mass spectrometer analysis time is kept constant. GeLC separations were evaluated with high numbers of 1st dimension fractions and fast LC-MS/MS analysis, few 1st dimension fractions and long LC-MS/MS analysis or with moderate separation efficiency in both dimensions. We initially thought that the highest number of protein identifications would correlate with the GeLC condition displaying the highest 2D peak capacity; however this was not the case. The highest numbers of protein identifications were produced with moderate separation efficiency in each dimension, when the chromatographic reequilibration times constituted the lowest percentage of total experiment time. Our results show gel overlap is widespread and independent of the number of slices collected. Furthermore, for proteomic applications no optimal number of gel slices to collect was identified. These results also show that more efficient use of instrument time can help offset the reduction in peak capacity when LC gradient time was reduced.

Introduction:

The field of proteomics has witnessed significant advances in recent years due to rapid improvements in mass spectrometry (MS). Technological developments have improved instrument sensitivity, mass accuracy and resolution, but bioanalytical analysis still presents a formidable challenge due to the high degree of complexity and wide dynamic range displayed by biological systems.[3,4] MS is currently unable to overcome these challenges due to a relatively narrow dynamic range, therefore a variety of separation techniques have been utilized to improve upon this limitation.[16,33] One-dimensional (1D) separations improve proteome coverage, but lack the resolving power to match the complexity of a proteolytic digest which can contain tens to hundreds of thousands of different peptides.[33] To increase resolving power and improve proteome coverage, multidimensional separation techniques have been incorporated into MS-based proteomic workflows.

A wide variety of electrophoretic, chromatographic and mass spectrometry techniques have been combined to achieve 2, 3 and even 4D separations.[16,72-76,87,89,90] Unfortunately, as the number of separation dimensions increases, the required amount of MS analysis time also increases. For example, if the time required for liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis is 2 hours, adding a second dimension of separation with 10 fractions would increase MS analysis time to 20 hours. Likewise, the addition of a third dimension collecting 10 fractions would increase MS analysis time to 200 hours for a single replicate. While 3-and-4D separations are possible, this becomes impractical when analyzing numerous samples with multiple replicates.

A viable compromise between MS analysis time and separating power can be achieved with 2D separations. Reversed phase LC (RP-LC) is the most widely utilized separation

technique for electrospray ionization (ESI), and has been combined with various other techniques for 2D analysis.[33,72,73,75,76,87] Numerous reports have compared these 2D separation techniques, and results have varied concerning the selection of the "best" method.[86,125,186-188] However multiple investigators have reported 1D gel electrophoresis combined with RP-LC-MS/MS (GeLC-MS/MS) produced the highest numbers of proteomic identifications.[86,125]

While previous evaluations of GeLC separations have been based on comparative analysis with other 2D methods, questions can be raised about the separation efficiency contribution from each dimension. As separation efficiency in either dimension increases, the amount of time required for MS analysis also increases, similar to the pitfalls observed with 3- and-4D separations. The common question presented to researchers then becomes how many gel slices should be collected, and what LC conditions should be applied for GeLC-MS/MS analysis.

The optimal number of gel slices to collect has not been a subject of previous evaluation, and in most cases it appears that a number was chosen based on preference from the individual performing the analysis. In 2D comparative works, the number of gel slices collected has varied from 10 to 38.[86,125,186,188] Other GeLC-MS/MS proteomic works have utilized high numbers of gel slices (10-20) for the purpose of increasing the depth of analysis.[110,189] Is this increase in depth due to enhanced 1st dimension separation efficiency, the additional amount of MS analysis time now required to analyze more gel slices or a combination of both?

The evaluation of RP-LC separations has been much more thoroughly examined, with many reports demonstrating a correlation of peak capacity and proteomic identifications.[137,138,190,191] The combination of lengthy columns (i.e. 50 cm) and extensive gradients (i.e. 8 hrs.) can achieve the highest peak capacities, thus producing the most proteomic

identifications, but how much of an impact does LC peak capacity have on proteomic identifications when SDS-PAGE is applied upstream of LC-MS/MS analysis?[138,190,191]

As MS analysis time is a practical limitation, achieving the highest proteomic identification efficiency becomes an important consideration. The aim of this work is to evaluate GeLC separations to determining how to maximize protein identifications when MS instrument time is held constant. Does the highest protein identification efficiency occur with high numbers of 1st dimension fractions and short LC-MS/MS analysis, few 1st dimension fractions and long LC-MS/MS analysis or with moderate separation efficiency in both dimensions?

Materials and Methods:

In-gel Digestion

Soluble protein extracts from canine prostate tissue were loaded into a gel and underwent gel electrophoresis through a NuPAGE 12% Bis-Tris Gel (Invitrogen, Carlsbad, CA). Each gel lane contained 8 µg of protein measured by BCA assay. Gels were stained with Coomassie Brilliant Blue R-250 Staining Solution (Bio-Rad, Hercules, CA). After a 3 hour staining, a fixer solution of 40% methanol and 10% acetic acid was then added to the gel for 1 hour. Gels were then removed from the fixer solution, washed with water and cut into the desired number of slices. Gel lanes were cut into either 3, 5, 9 or 18 slices, with each slice cut into 1 x 1 mm squares for digestion. Water was then added to the gel pieces and discarded to waste. Gel pieces were then washed with a mixture of 50% acetonitrile and 50% water, with solution removed as waste after 15 minutes. 100mM ammonium bicarbonate was then added, and after 15 minutes, an equal volume of acetonitrile was added to make a 1:1 (v/v) solution. After incubation at room temperature for 15 minutes, the ammonium bicarbonate/acetonitrile solution was removed to waste. Acetonitrile was again added to the gel slices and incubated at room temperature for 15

minutes. This solution was then removed as waste. A solution of 10 mM dithiothreitol (DTT) in

100 mM ammonium bicarbonate was then added and incubated in a 65°C water bath for 1 hour.

The reducing solution was then removed and a 55 mM iodoacetamide (IDA) in 100 mM

ammonium bicarbonate alkylating solution was added. Samples were incubated at room

temperature for 1 hour in the dark. The alkylating solution was then removed and 100 mM

ammonium bicarbonate was added. After 5 minutes an equal volume of acetonitrile was added

to make a 1:1 (v/v) solution. After 15 minutes of incubation, this solution was removed to waste.

A solution of 0.1% sequencing grade trypsin (Promega, San Luis Obispo, CA) was made in 100

mM ammonium bicarbonate and added to the gel pieces at 1:50 (w/w) enzyme to protein; the

resulting mixture was digested overnight at 37°C. The following day peptides were extracted by

collecting the solution from gel pieces, then a solution of 50% acetonitrile and 0.1% formic acid

was then added to the gel pieces. After 15 minutes this solution was extracted, and pooled with

the solution previously pulled from the digested gel slices. Peptide solutions were then

evaporated to dryness and stored at -20°C until mass spectrometry analysis.

LC-MS/MS Analysis

Data was acquired using an Agilent 1100 Capillary LC system (Palo Alto, CA) with

either a 0.2 x 50 mm or 0.2 x 150 mm Halo Peptide ES-C18 capillary column packed with 2.7

µm diameter superficially porous particles (Advanced Materials Technology, Inc., Wilmington,

DE). On-line MS detection used the Thermo-Fisher LTQ ion trap (San Jose, CA) with a

Michrom (Michrom Bioresources, Auburn, CA) captive spray interface. Sample analysis

utilized the LTQ divert valve fitted with an EXP Stem Trap 2.6 µL cartridge packed with Halo

Peptide ES-C18 2.7 µm diameter superficially porous particles (Optimize Technologies, Oregon

City, OR). Mobile phase A consisted of 99.9% water, 0.1% formic acid, adjusted to 10 mM

ammonium formate. Mobile phase B contained 80% acetonitrile, 0.1% formic acid and 10 mM ammonium formate. Mobile phases used formic acid, ammonium formate and acetonitrile from Sigma-Aldrich (St. Louis, MO).

Gradient conditions increased the concentration of mobile phase B from 6.3% to 75% B over 12.5, 36, 74 or 90 minutes (complete LC experimental conditions provided in Table 1). Evaluation of GeLC separations were conducting using the following conditions; (1) 18 gel fractions with a 21 minute LC-MS/MS analysis for each fraction (18 x 21 min), (2) 9 gel fractions with a 45 minute LC-MS/MS analysis for each fraction (9 x 45 min), (3) 5 gel fractions with an 83 minute LC-MS/MS analysis (5 x 83 min) and (4) 3 gel fractions with a 140 minute LC-MS/MS analysis for each fraction (3 x 140 min). For each GeLC separation, a total of 3 gel lanes were analyzed in duplicate, requiring approximately 43 hours of MS analysis time.

The number of MS/MS spectra acquired per cycle was set to either: 5, 7 or 8. This was determined based on observed cycle times and chromatographic peak widths measured during peak capacity analysis.[185] The number of MS/MS spectra acquired per cycle was set to allow for completion of 3 cycles during chromatographic peak elution, Table 1.

Raw tandem mass spectra were converted to mzXML files, then into mascot generic files (MGF) via the Trans-Proteomic Pipeline (Seattle Proteome Center, Seattle, WA). MGF files were searched using Mascot (Matrix Scientific Inc., Boston, MA) against separate target and decoy databases obtained from the National Center for Biotechnology Information (NCBI). The target database contained all canine protein sequences and the decoy database contained the reversed sequences from the target database. Mascot settings were as follows: tryptic enzymatic cleavages allowing for up to 2 missed cleavages, peptide tolerance of 1000 parts-per-million, fragment ion tolerance of 0.6 Da, fixed modification due to carboxyamidomethylation of

cysteine (+57 Da), and variable modifications of oxidation of methionine (+16 Da) and deamidation of asparagine or glutamine (+0.98 Da). Mascot files were loaded into ProteoIQ (PREMIER Biosoft, Palo Alto, CA), where a 5% protein false discovery rate and a 0.9 peptide probability was applied for confirmation of proteomic identifications.

Peak Capacity

Tryptic peptides from bovine serum albumin (Michrom Bioresources, Auburn, CA) were used to evaluate peak capacity, measured from duplicate experiments of 1 pmol sample injections. Peak capacity measurements were made without the collection of MS/MS spectra, to increase the number of data points collected across each chromatographic peak. Calculation of peak capacity was based on the selection of 14 BSA tryptic peptides listed in Table 2. These peptides were selected to provide a wide range of retention for accurate evaluation of peak capacity. Calculations of peak capacities were conducted using 2 methods, each with a different measurement of time. Equation 1, measured peak capacity, was calculated using the retention times of the first and last eluting peptides from Table 2. Equation 2, theoretical peak capacity, used the actual length of the gradient (T_g) for retention time, independent of actual peptide retention times. Theoretical peak capacity was calculated to provide a measure of peak capacity independent of sample retention times to demonstrate the maximum peak capacity obtainable for each LC gradient condition. Peak capacity measurements were made using Xcalibur software, measuring peak widths at 50% peak height via manual inspection of extracted ion chromatograms.

Results:

Comparative analysis of GeLC separations were conducted with the calculation of peak capacity. The Giddings' method for determining peak capacity of 2-dimensional separations was

applied, taking the product of peak capacity from each separation dimension (N_{pc} x N_{pc}).[111] The number of gel slices collected was used for peak capacity in the 1st dimension, and the LC peak capacity was measured using 14 BSA tryptic peptides listed in Table 2. The GeLC separation producing the highest 1st dimension peak capacity was the 18 gel slices/21 minute LC-MS/MS analysis (18 x 21 min), which also produced the highest 2D peak capacity shown in Table 3. Although the 3 gel slices/140 minute LC-MS/MS analysis (3 x 140 min) produced the highest LC peak capacity, it provided the lowest 2D peak capacity due to the low number of 1st dimension fractions.

The next metric evaluated were the numbers of proteomic identifications produced by each GeLC separation. For this analysis, soluble proteins extracted from canine prostate tissue were used, and protein identifications were validated using a 5% protein false discovery rate and a 0.9 peptide probability.   Initially, it was thought that the highest numbers of protein identifications would correlate to the GeLC condition displaying the highest 2D peak capacity, but this was not the case. The combination of moderate separation efficiency in the 1st and 2nd dimension produced the highest numbers of proteomic identifications, and the 5 gel slices/83 minute LC-MS/MS analysis (5 x 83 min) provided the highest protein identification efficiency, shown in Table 4. The most surprising results came from the 18 x 21 min separation, which produced the fewest number of proteomic identifications, 4-6x less than any of the other GeLC separations evaluated. Another surprising observation was the GeLC separation displaying the highest LC peak capacity (3 x 140 min) produced significantly fewer protein identifications than separations having much lower LC peak capacities.   These results suggest the previously reported correlation of peak capacity and proteomic identifications does not apply for GeLC

separations, and other factors appear to provide more significant contributions than initially expected.

1D Gel Separation Analysis

These results raised many questions since they contradicted a previously reported correlation and our initial hypothesis. The first question presented is how to accurately measure GeLC peak capacity. The calculation of peak capacity for LC separations is well defined, but when combined with gel fractionation, this can be far more ambiguous. Peak capacities of 2D separations are commonly measured using the Giddings' method (N_{pc} x N_{pc}), but no reports have addressed peak capacity of 1D gel electrophoresis when applied to GeLC.

The measurement of peak width can easily be determined in chromatographic analysis, but this becomes far more difficult to determine in gel electrophoresis. Is a peak defined as an individual protein band or the number of gel slices collected? The measurement of protein bands can be difficult, and gel stains commonly used such as Coomassie Blue exhibit very low sensitivity. Most to all low abundance proteins are not observed with gel staining, so measuring protein bands does not provide an accurate assessment of separating power. The approached used in this work defined a peak as a gel slice since the actual separation occurs from cutting the gel. While this appeared to be the most logical approach, it still may not be the most accurate due to high sample overlap commonly observed in gel electrophoresis.

The next questions raised included how much overlap is occurring with gel electrophoresis and can this be accurately accounted for in a peak capacity measurement? It was initially hypothesized that higher numbers of gel slices created more sample overlap, leading to the same protein being identified in multiple gel slices. This was based in part on the high number of matched MS/MS spectra per peptide identification observed with the 18 x 21 min

analysis (Table 4). To measure the degree of overlap, an adjusted number of gel slices was calculated for each GeLC separation. Every protein identification was analyzed to determine how many different gel slices it was identified in within the same gel lane. Based on this, the adjusted gel slice number was determined, and an average was taken for each GeLC separation. Figure 1 displays an image of a stained gel, with red lines indicating the approximate location the gel was sliced at. If the same protein was identified in both slice 1A and 2A, the number of fractions collected for the separation was reduced from 3 to 2. Likewise, if a protein was only identified in slice 3A, the number of slices remained the same. Table 5 presents the specific adjusted gel slices numbers used for the 3 x 140 min analysis. The same logic was applied for the 5, 9 and 18 gel slice separations. Once all protein identifications were analyzed, adjusted gel slices number were calculated for each GeLC separation. This average was applied to the 2D peak capacities presented in Table 6. For each GeLC separation, ~40% of proteins found were identified in multiple gel slices, suggesting no correlation between overlap and the number of gel slices collected.

Due to the similarities in measured overlap and significant differences in protein identifications, high abundance proteins were specifically examined to determine how much impact they have on sample overlap. To assess this, adjusted gel slice numbers were determined for the top 10% and 25% most abundant protein identifications (based on spectral counts) for each GeLC separation. Surprisingly, analysis of the top 10% most abundant protein identifications showed similar overlap, with the highest numbers of gel slices producing the smallest degree of overlap. Results were also similar analyzing the top 25% of proteins, indicating high abundance proteins do not show an increasing degree of overlap as the number of gel slices is increased.

These results demonstrate a lack of correlation between 2D peak capacity and proteomic identifications for GeLC separations. Although the measurement of peak capacity in a 1D gel can be the subject of much debate, our results suggest N_{pc} x N_{pc} it is not applicable to GeLC separations. Peak capacity is a measurement of well-resolved peaks. Our results show ~40% of protein found were identified in multiple gel slices in the same lane, and this was independent of the number of slices collected. This high degree of overlap suggests no well-resolved bands are present, so gel fractionation separating power cannot be measured with peak capacity. Furthermore, GeLC may not be a true 2D separation since a chemical modification occurs between the 1st and 2nd dimension of separation with enzymatic digestion. These results also suggest there is no "optimal" number of gel slices to collect. The increase in depth of analysis previously reported is the result of increased MS analysis rather than increased separation efficiency from collecting additional gel slices.[110]

LC-MS/MS Analysis

LC-MS/MS conditions were also examined to determine factors that contributed to protein identification efficiency. While each GeLC separation required ~43 hours for MS analysis, the times required for sample loading onto a trap column and reequilibration were not the same due to differences in column lengths and flow rates. This in turn determined the gradient length for each LC separation. The percentage of total experiment time allocated to LC gradient elution showed a direct correlation with proteomic identifications. The 5 x 83 min analysis utilized 86% of instrument time on gradient elution (Figure 2C), in turn producing the most protein identifications. Likewise, the 18 x 21 min analysis spent the least amount of instrument time on gradient elution (52%), and identified the fewest number of proteins (Figure 2A).

While reequilibration is necessary, no relevant proteomic data should be collected during this time since high confidence peptide identifications should only be produced during gradient elution. To confirm this, data from the 3 x 140 min analysis was run through an in-house program to remove all spectra acquired during sample loading and reequilibration. Mascot database searching produced identical results with and without the presence of sample loading and reequilibration spectra, proving no proteomic identifications are produced during these times.

Each LC gradient included appropriate trap column sample loading time and a 10x column volume reequilibration (50 mm column volume: 0.8 µL; 150 mm column volume: 2.5 µL). The experimentally measured 56 µL chromatograph dwell volume was also taken into account when determining LC gradient conditions. Every LC experiment was operated at the highest flow rate possible staying below the chromatograph back pressure limit of 400 Bar. The 50 mm column was operated at 9 µL/min, so taking LC dwell volume into account, only 6.5 minutes were required for reequilibration, whereas the 150 mm column could only be operated at 4 µL/min so 50 minutes were required for reequilibration. The high flow rates provided an advantage in efficiency, most notably for the 5 x 83 min analysis which spent less than 10% of total instrument time on sample loading and reequilibration. Likewise, the 2 GeLC separations producing the fewest proteomic identifications required the most instrument reequilibration time (~30%).

These results also show that more efficient use of MS experiment time can help offset the reduction in peak capacity when LC gradient time was reduced. Both moderate separation efficiency conditions (5 x 83 min and 9 x 45 min) produced LC peak capacities of 65, which was

far less than that of the 3 x 140 min separation (N_{pc} = 100). However, they both displayed better instrument time efficiency, in turn identifying more proteins.

Analysis of instrument time utilization also highlighted another consideration when determining how to maximize protein identification efficiency. The amount of MS analysis time for each GeLC separation is presented in Table 1, which includes the run to run delay necessary for LC injection. Unfortunately each LC injection is subjected to a ~3 minute instrument-fixed delay. This puts GeLC separations with higher numbers of 1st dimension fractions at a disadvantage due to higher numbers of LC-MS/MS runs. To analyze all replicates from the 18 x 21 min analysis, a total of 108 LC injections were necessary, which corresponded to 12.5% of total instrument time (Figure 1A). This was time that no MS data was collected. Although this was not significant enough to explain the huge disparity in proteomic identifications, it is a source of inefficiency that provided some contribution to these results.

Conclusions:

The GeLC separation producing the most protein identifications in this fixed instrument time experiment was the 5 x 83 min analysis, which utilized moderate separation efficiency in both dimensions. Surprisingly, a correlation of 2D peak capacity and proteomic identifications was not observed. The 18 x 21 min analysis produced the highest 2D peak capacity, but identified the fewest number of proteins. Our data shows the amount of overlap is significant, and independent of the number of slices collected. Although the calculation of peak capacity in a gel is quite ambiguous, it appears this is not an appropriate measurement of separating power in a gel since no well-resolved bands are produced. Increased depth of analysis from collecting higher numbers of gel slices is the result of additional MS analysis rather than increased 1st

dimension separation efficiency. We can also conclude that no ideal number of gel slices to collect was identified.

A correlation of the percentage of total instrument time allocated to gradient elution and proteomic identifications was observed. Since no high confidence peptide identifications are produced during sample loading and LC reequilibration times, maximum protein identification efficiency occurs when these times can be minimized. The high flow rates used with the 50 mm column provided an advantage over the 150 mm column, as sample load and reequilibration time was significantly reduced. This in turn provided a more effective use of instrument time, producing the highest number of protein identifications. Results also show that increased instrument efficiency can help offset the reduction in LC peak capacity resulting from shorter LC gradients.

Acknowledgments:

Table 5.1

GeLC-MS/MS Experimental Conditions

Abbreviation Used	No. Gel Slices	LC Column Length (mm)	Flow Rate (µL/min)	Gradient Length (min)	Experiment Time (min)[a,b]	Total MS Analysis Time (hrs)[c]	No. MS/MS Events/Cycle	Approx. Cycle Time (s)[d]
18 x 21 min	18	50	9	12.5	24	43.2	5	1.5
9 x 45 min	9	50	9	36	48	43.2	7	2.7
5 x 83 min	5	50	9	74	86	43	8	3
3 x 140 min	3	150	4	90	143	42.9	8	3

[a] Experiment time required for the analysis of 1 gel slice. This includes an additional 3 minutes for LC injection delay.

[b] Each LC-MS/MS experiment included the following considerations: 56 µL LC system dwell volume, 10x column volume re-equilibration (Column volumes: 50 mm = 0.8 µL; 150 mm = 2.5 µL).

[c] Total mass spectrometer time required to analyze all gel slices and replicates (6 replicates).

[d] Cycle time includes acquisition of 1 MS scan and user defined number of MS/MS scans.

Table 5.2

BSA Tryptic Peptides Used for Chromatographic Analysis and Peak Capacity Calculations

Sequence	Peptide Mass (Da)
CASIQK	705.8237
LVTDLTK	788.9290
QTALVELLK	1014.2164
EACFAVEGPK	1107.2371
KQTALVELLK	1142.3887
SLHTLFGDELCK	1419.6016
YICDNQDTISSK	1443.5353
LGEYGFQNALIVR	1479.6783
MPCTEDYLSLILNR	1724.9953
LFTFHADICTLPDTEK	1908.1352
DAIPENLPPLTADFAEDKDVCK	2459.6801
GLVLIAFSQYLQQCPFDEHVK	2492.8454
QEPERNECFLSHKDDSPDLPK	2541.7044
TVMENFVAFVDKCCAADDKEACFAVEGPK	3310.7085

Table 5.3

2nd Dimension Peak Capacities Obtained from the Number of Gel Slices and LC-MS Analysis of 1 pmol Injections of Tryptic BSA Peptides Using Various Chromatographic Conditions

No. Gel Slices	Column Length (mm)	Flow Rate (µL/min)	Gradient Length (min)	AVG. Peak Width (s)	Peak Width SD	LC N_{pc}	LC Theoretical N_{pc}	2D N_{pc}[a]	2D Theoretical N_{pc}[a]
18	50	9	12.5	3.56	0.69	46.6	123.9	838.8	2230.7
9	50	9	36	7.30	1.33	64.6	174.1	581.1	1566.5
5	50	9	74	15.99	5.65	64.9	163.4	324.8	816.9
3	150	4	90	13.99	2.35	99.6	227.1	298.7	681.2

[a] 2D peak capacity is the product of LC peak capacity and the number of gel slices collected.

Table 5.4

Proteomic Identifications Results from Canine Prostate Analysis Using Various GeLC Separations

No. Gel Slices	Column Length (mm)	Gradient Length (min)	Protein IDs[a]	AVG Protein ID/ Replicate (SD)[b]	Peptide IDs[a]	AVG Peptide ID/ Replicate (SD)[b]	Matched MS/MS Spectra[b]	MS/MS Spectra/ Peptide ID	2D Measured N_{pc} (Theoretical N_{pc})
18	50	12.5	46	27 (5.4)	185	78.8 (20.4)	1105	6.0	838.8 (2230.7)
9	50	36	267	188.7 (5.8)	1228	549 (45.8)	5126	4.2	581.1 (1566.5)
5	50	74	285	198 (6.7)	1327	606.7 (33.7)	5389	4.1	324.8 (816.9)
3	150	90	199	129.3 (6.5)	765	320.7 (28.9)	2357	3.1	298.7 (681.2)

[a] Proteomic identifications validated using a 5% protein false discovery rate and 0.9 peptide probability
[b] Average protein and peptide identifications calculated from 6 replicates

Table 5.5

Calculation of Adjusted Gel Slice Number for the 3 x 140 min GeLC Separation

Actual Number of Gel Slices Collected	No. of Slices a Protein is Identified In[a]	Adjusted Gel Slice Number
3	1	3
3	2	2
3	3	1

a Reflects the number of individual gel slices a specific protein was identified in within the same gel lane.

Table 5.6

The Actual Number of Gel Slices Collected and Calculated Adjusted Number of Gel Slices Measuring Sample Overlap Observed During Gel Fractionation

No. Gel Slices	Percentage of Protein IDs in Multiple Slices[a]	ADJ No. Gel Slices[a,d]	Measured 2D Npc	ADJ No. Gel Slices: Top 10%[b,d]	ADJ No. Gel Slices: Top 25%[c,d]
18	39.4%	16.8 (93.3%)	783	12.1 (67.4%)	15.0 (83.5%)
9	47.0%	7.8 (86.8%)	504	5.3 (58.4%)	6.5 (71.9%)
5	42.6%	4.3 (85.8%)	279	2.8 (55.8%)	3.4 (67.2%)
3	36.9%	2.6 (85%)	254	1.6 (53%)	2.0 (66.3%)

[a] For every GeLC separation every protein was analyzed by the gel lane, so if the same protein was identified in two separate gel lanes, this was treated as two different identifications for the purpose of calculating adjusted gel slice numbers.
[b] The adjusted number of gel slices was only calculated for the top 10% of protein identifications (based on spectral counts).
[c] The adjusted number of gel slices was only calculated for the top 25% of protein identifications (based on spectral counts).
[d] The percentage reflects the comparison of the calculated adjusted number of gel slices compared to the actual number of gel slices collected.

Equation 1

Measured Peak Capacity

$$n_{pc} = (t_f - t_i)/W_{4[\sigma]}$$

Equation 2

Theoretical Peak Capacity

$$n_{pc} = T_g/W_{4[\sigma]}$$
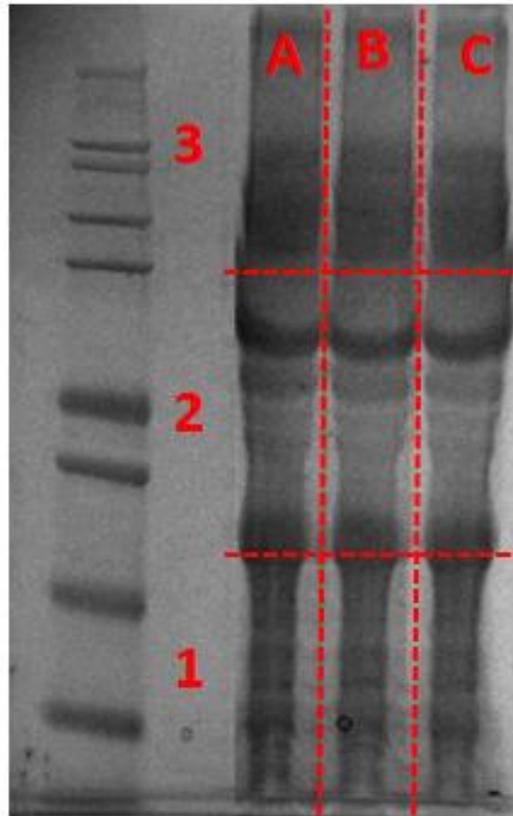
Figure 5.1 SDS PAGE Gel of Canine Prostate Tissue. 1D SDS gel of soluble canine prostate proteins stained with Coomassie Blue. Each lane contained ~8 μg of protein. Red dashed lines indicate the approximate location were the gel was sliced for the 3 x 140 min GeLC analysis. A, B and C are shown to denote gel lanes, and 1, 2 and 3 denote the location where each slice was collected on the gel.
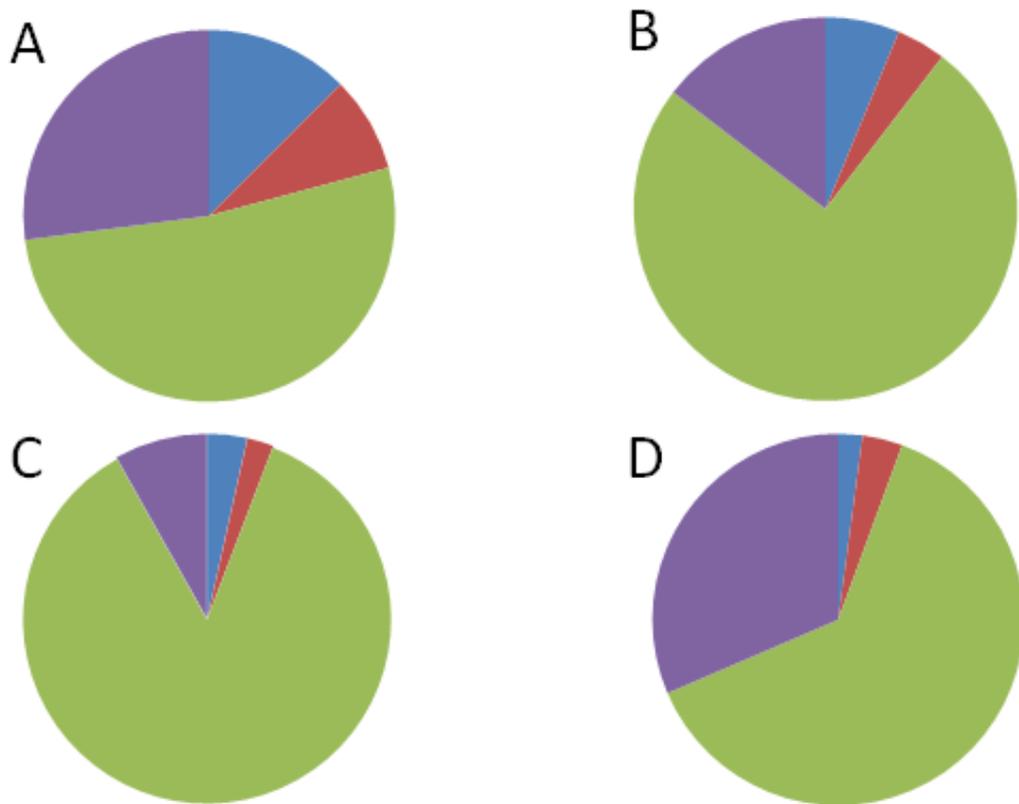
Figure 5.2 Mass Spectrometer Instrument Time Utilization. An analysis of LC-MS instrument time utilization for each GeLC separation was conducted. Each pie chart reflects total LC-MS/MS analysis time required to analyze all gel slices and replicates. LC injection time: Blue; time required for sample loading onto a trap column: Red; gradient elution time: Green; time required for LC reequilibration: Purple. (A) 18 x 21 min, (B) 9 x 45 min, (C) 5 x 83 min, (D) 3 x 140 min.

CHAPTER 6

PROTEOMIC ANALYSIS OF THE CANINE PROSTATE GLAND AND EXPLORATION OF

THE CANINE AS AN ANIMAL MODEL OF HUMAN PROSTATE CANCER?[4]

---

[3]Johnson, D., Goedeke, A., Machado, U., Orlando, R. To be submitted to *Journal of Proteome Research*.

Abstract:

Prostate cancer is the sixth leading cause of carcinoma death among men worldwide. Although it is the second most frequently diagnosed cancer, the healthcare industry still lacks a reliable diagnostic tool to distinguish between indolent and aggressive forms of the disease. Progress towards identifying new diagnostic and therapeutic strategies presents a variety of challenges, therefore animals models are typically used for the large-scale study of human diseases. Unfortunately, prostate cancer research has been hindered by the lack of a relevant animal model. The potential use of the canine shows promise, as humans and canines are the only two large mammals that spontaneously generate prostate cancer. Currently, no large-scale proteomic studies have been conducted evaluating protein expression of the canine prostate gland. The aim of this work is to analyze protein expression among several histologically different canine prostates, and explore the potential for a predictive model of human prostate cancer. Comparative analysis between canine prostate cancer and healthy prostate tissues identified numerous proteins with reported biological significance to human prostate cancer development and progression. These results suggest the canine can provide a relevant predictive model of androgen-insensitive, highly-metastatic human prostate cancer.

Introduction:

Prostate cancer is the sixth leading cause of carcinoma death and the second most frequently diagnosed cancer among men worldwide.[192] While increased media attention has heightened prostate cancer awareness, the healthcare industry still lacks a reliable diagnostic tool to distinguish between indolent and aggressive forms of the disease.[193] The two most commonly used noninvasive screening methods, the digital rectal exam (DRE) and the prostate specific antigen test (PSA), have not been shown to decrease mortality, and are prone to over-diagnosis of clinically insignificant tumors leading to over-treatment.[194,195] Currently, the U.S. Preventative Services Task Force recommends against PSA-based screening for prostate cancer, concluding the benefit does not outweigh the expected harm.[196]

The process of developing improved diagnostic strategies is time consuming and can be difficult due to the regulatory and ethical issues that arise from studies involving human subjects.[152] Therefore, animal models are frequently used for the robust study of human diseases. Unfortunately, prostate cancer research has been hindered by the lack of a relevant animal model.[17] Mice have traditionally been the most widely used model, however they may not be the best candidate for studying prostate cancer. Mouse models are limited by narrow genetic diversity and most tumors must be induced, whereas in humans, tumors arise spontaneously.[152,160] Due to these shortcomings, the canine has been suggested as a potential animal model.

Approximately 400 inherited diseases similar to those found in humans have been characterized in dogs.[161,162] The canine also offers many other advantages including a high degree of genetic diversity, exposure to similar environments, aging five to eight-fold faster than humans, receiving specialized healthcare and the spontaneous generation of many human

diseases.[152]   Furthermore, canines and humans are the only two large mammals that spontaneously generate prostate cancer, and in both species the prostate gland shares many anatomical and functional similarities. [18-20] The prostate is an ovoid-shaped retroperitoneal gland that surrounds the neck of the urinary bladder and proximal urethra.[163] Canines and humans also share many characteristics associated with prostate cancer development such as increased incidence with age, similar clinical symptoms and frequent occurrence of osteoblastic bone metastases.[163,164]   In addition, most canine prostate neoplasms are aggressive and androgen-independent due to a lack of androgen receptors, thereby suggesting a model specifically for metastatic prostate cancer.[163,197]

Previous work exploring the relevance of a canine animal model was conducted using a mass spectrometry (MS)-based proteomics approach. LeRoy et al. utilized two-dimensional differential in-gel electrophoresis (2-D DIGE) to compare protein expression between healthy and carcinoma prostate tissue. Proteins exhibiting more than a 2.5-fold difference in expression were subjected to MS analysis. This study identified several proteins that were differentially expressed and are of biological significance to prostate cancer.[21] While this work demonstrated the potential of the canine animal model, it did not provide a global proteome analysis to further study the disease.

The aim of this work is to produce the first large-scale analysis of protein expression in the canine prostate gland and further explore the potential of the canine as a predictive model of human prostate cancer. A 1D gel electrophoresis reversed-phase liquid chromatography tandem mass spectrometry (GeLC-MS/MS) method was used to examine protein expression profiles among several histologically different canine prostates. The following prostate samples were analyzed: prostate carcinoma, atrophied prostate, benign prostatic hyperplasia, suppurative

prostatitis and healthy prostate tissue. Comparative analysis of protein expression was conducted, and proteins that were differentially expressed in carcinoma samples were examined to determine their biological significance in prostate cancer.

Methods:

Collection of Samples

The prostate tissue samples were collected from cadavers at the time of necropsy. The breed, age, and reproductive status were recorded. From each animal, 1/2 of the sample was kept at -20C until needed for tissue preparation; the remaining tissues were preserved in neutral-buffered, 10% formalin solution. Fixed tissues were trimmed, routinely processed, embedded in paraffin, and cut at approximately 5 microns. Slides were stained with hematoxylin and eosin (HE) and with Masson's trichrome. A board certified veterinary pathologist evaluated the slides microscopically and provided histological diagnoses of the prostate gland. Histological slides are provided in Figure 1.

Tissue Preparation

Approximately 1 gram of tissue was placed in a 45 mL falcon tube and a lysis buffer (44 mL PBS pH 7.4, 5 mL NP-40, 0.5 mL 10% SDS, 5 µL 1 M dithiothreitol (DTT)) was added for a final sample/buffer volume of 7 mL. Ice cold tissue samples were then subjected to homogenization using a Polytron PT 10-35 (Kinematica, Inc, Bohemia, NY) for 30 seconds at a setting of 3, repeated 3 times with 1 minute of rest on ice between each homogenization. Tissue samples were then centrifuged at 1,000 x g for 5 minutes at 4°C. Samples were then transferred to Oakridge tubes for centrifugation at 35,000 x g for 30 minutes at 4°C. The soluble supernatant was then removed and concentration was measured using a BCA Protein Assay kit (Pierce, Rockford, IL).

Soluble protein extracts from each prostate sample were loaded into a gel and underwent gel electrophoresis through a NuPAGE 12% Bis-Tris Gel (Invitrogen, Carlsbad, CA). Each gel lane contained approximately 8 µg of protein measured by BCA assay. Each gel lane was cut into 14 slices, with each slice cut into 1 x 1 mm squares for digestion. Water was then added to the gel pieces and discarded to waste. Gel pieces were then washed with a mixture of 50% acetonitrile and 50% water, with solution removed as waste after 15 minutes. 100mM ammonium bicarbonate was then added, and after 15 minutes, an equal volume of acetonitrile was added to make a 1:1 (v/v) solution. After incubation at room temperature for 15 minutes, the ammonium bicarbonate/acetonitrile solution was removed to waste. Acetonitrile was again added to the gel slices and incubated at room temperature for 15 minutes. This solution was then removed as waste. A solution of 10 mM dithiothreitol (DTT) in 100 mM ammonium bicarbonate was then added and incubated in a 70°C water bath for 1 hour. The reducing solution was then removed and a 55 mM iodoacetamide (IDA) in 100 mM ammonium bicarbonate alkylating solution was added. Samples were incubated at room temperature for 1 hour in the dark. The alkylating solution was then removed and 100 mM ammonium bicarbonate was added. After 5 minutes an equal volume of acetonitrile was added to make a 1:1 (v/v) solution. After 15 minutes of incubation, this solution was removed to waste. A solution of 0.1% sequencing grade trypsin (Promega, San Luis Obispo, CA) was made in 100 mM ammonium bicarbonate and added to the gel pieces at 1:50 (w/w) enzyme to protein; the resulting mixture was digested overnight at 37°C. The following day peptides were extracted by collecting the solution from gel pieces, then a solution of 50% acetonitrile and 0.1% formic acid was then added to the gel pieces. After 15 minutes this solution was extracted, for pooling with

the solution previously pulled from the digested gel slices.   Peptide solutions were then evaporated to dryness and stored at -20°C for mass spectrometry analysis.

Mass Spectrometry Analysis

Data was acquired using an Agilent 1100 Capillary LC system (Palo Alto, CA) with a 0.2 x 150 mm Halo Peptide ES-C18 capillary column packed with 2.7 μm diameter superficially porous particles (Advanced Materials Technology, Inc., Wilmington, DE). On-line MS detection used the Thermo-Fisher LTQ ion trap (San Jose, CA) with a Michrom (Michrom Bioresources, Auburn, CA) captive spray interface. Sample analysis utilized the LTQ divert valve fitted with an EXP Stem Trap 2.6 μL cartridge packed with Halo Peptide ES-C18 2.7 μm diameter superficially porous particles (Optimize Technologies, Oregon City, OR). Gradient conditions increased the concentration of mobile phase B from 6.25% to 75% B over 90 minutes. Mobile phase A consisted of 99.9% water, 0.1% formic acid and 10 mM ammonium formate. Mobile phase B contained 80% acetonitrile, 0.1% formic acid and 10 mM ammonium formate. Mobile phases used formic acid, ammonium formate and acetonitrile from Sigma-Aldrich (St. Louis, MO).

Raw tandem mass spectra were converted to mzXML files, then into mascot generic files (MGF) via the Trans-Proteomic Pipeline (Seattle Proteome Center, Seattle, WA). MGF files were searched using Mascot (Matrix Scientific Inc, Boston, MA) against separate target and decoy databases obtained from the National Center for Biotechnology Information (NCBI). The target database contained all canine protein sequences and the decoy database contained the reversed sequences from the target database. Mascot settings were as follows: tryptic enzymatic cleavages allowing for up to 2 missed cleavages, peptide tolerance of 1000 parts-per-million, fragment ion tolerance of 0.6 Da, fixed modification due to carboxyamidomethylation of

cysteine (+57 Da), and variable modifications of oxidation of methionine (+16 Da) and deamidation of asparagine or glutamine (+0.98 Da). Mascot files were loaded into ProteoIQ (NuSep, Bogart, GA), where a 5% false discovery rate and a 0.9 peptide probability were applied for confirmation of protein identifications. Quantitative analysis was performed measuring spectral counts within ProteoIQ using the total spectra count normalization method.[198]

Results:

Proteomic analysis of carcinoma, hyperplasia, prostatitis, atrophy and healthy prostate tissue were compared to determine the effects of disease on protein expression. Overall, protein expression was similar between the healthy, hyperplastic and prostatitis samples. In all three conditions, less than 10% of protein identifications were unique to a specific sample (Table 1). The two outliers in this analysis were the atrophy and cancer samples. The number of proteins identified in the atrophy prostate was significantly less than all other conditions examined; however this was expected since this animal was neutered. A lack of hormonal signaling results in prostate atrophy; thus this sample was much smaller, and fewer proteins were identified. The carcinoma prostate produced biologically significant differences when compared to the healthy prostate. Our results were then compared with previous proteomic studies of prostate cancer in both humans and canines.

Proteins Up-regulated in Cancer

A total of 285 proteins were identified in both the healthy and cancer samples. Of these proteins, 52 were found to be upregulated by more than 3-fold in the cancer sample. Further analysis determined 20 of these proteins have been previously shown to have biological significance in human prostate cancer. Table 2 displays these 20 proteins, which have all been identified by 2 or more peptides. Alpha-enolase, transketolase variant, voltage-dependent anion

channel 2, and heat shock protein 90-alpha have been identified in multiple prostate cancer studies as proteins of biological importance. Alpha-enolase is a multifunctional protein involved in glycolysis, growth control, hypoxia tolerance and allergic response.[17,21,199-201] It has previously been identified in men with metastatic prostate cancer and overexpressed in androgen-insensitive cell lines. LeRoy et al. identified alpha-enolase as an overexpressed protein in canine prostate cancer samples, noting it may indicate enhanced glycolytic processes to meet the metabolic needs of rapidly growing and dividing cancer cells.[21]

Transketolase variant (TV) is an enzyme involved in the glycolysis pathway, and has been purposed to be a protein utilized by malignant cells.[202] During the pentose phosphate pathway, the nonoxidative part of the conversion of glucose to ribose is controlled by transketolase enzymes. The inhibition of these enzyme reactions suppresses tumor growth and metastasis, therefore suggested to be a potential target for cancer therapies.[203] Our results identified TV as upregulated more than 4-fold in carcinoma tissue. Other studies have also identified transketolase, finding it to be overexpressed in androgen-insensitive and highly metastatic cell lines.[17,204]

Another protein found to be upregulated in cancer was the voltage-dependent anion channel (VDAC) 2 protein. This is a channel forming protein in the outer mitochondrial membrane allowing anion transfer. The VDAC protein family plays a key role in transporting ATP, ADP, pyruvate and other metabolites, thus exerting global control over mitochondrial metabolism.[205] It has also been shown that global suppression of VDAC proteins leads to apoptosis, suggesting overexpression of this protein could be related to increased tumor cell metabolism.[206] In other proteomic studies, VDAC-2 has been identified as a tumor-derived

protein and has been shown to be overexpressed in androgen-insensitive prostate cancer cell lines.[17,200]

Heat shock protein (Hsp) 90-alpha was also identified, and significantly upregulated in the prostate cancer sample. Other proteomic analyses have identified Hsp90-alpha in both androgen-sensitive and androgen-insensitive prostate cancer cell lines.[17,207] Hsp90 is a chaperone molecule with a well-established role in tumor biogenesis, assisting in folding and stabilizing a number of proteins required for tumor growth.[208] Inhibition of Hsp90 has also been investigated as an anticancer drug.[209]

Another interesting observation presented in Table 2 is the overexpression of annexin 1 and annexin 2. Numerous reports have identified the annexin proteins as differentially expressed in neoplasms, however both annexin 1 and 2 have been found to be downregulated in prostate cancer.[210-212] Annexin 1 is involved in signal transductions, cell differentiation, proliferation, tumor invasion, and has been suggested as a possible tumor suppressor.[213,214] It has also been reported that increased annexin 1 expression reduced tumourigenicity by enhancing activation of pro-apoptotic signaling pathways.[215] Likewise, annexin 2 has also been reported as downregulated in prostate cancer.[214,216] Annexin 2 has also been described as a tumor suppressor, and reduced expression may contribute to prostate cancer development and progression.[213] While our results identified these proteins as overexpressed, this is most likely the result of a limited sample size used in this analysis.

Proteins Unique to Cancer

This study also identified 69 proteins that were unique to the cancer sample. Among these identifications, 10 proteins have been previously associated with human prostate cancer development and progression, shown in Table 3. Calreticulin is a calcium-binding chaperone

protein identified uniquely in the cancer sample in this work. Other studies have identified calreticulin in prostate cancer cell lines, and have shown it to be upregulated in highly metastatic and androgen-insensitive cell lines.[17,204,217] Calreticulin plays a key role in tumor proliferation by providing a pro-phagocytosis signal, which is then inhibited by CD47 to prevent cancer cell phagocytosis by the immune system. On the surface of healthy cells this protein is minimally expressed, however it is highly expressed on the surface of several human cancer cells.[218]

Elongation factor 1 alpha (eEF1A1) was identified in the prostate cancer tissue but not in the healthy sample, in agreement with other studies that have found it to be uniquely expressed in human prostate cancer tumors.[219] When protein expression was compared between the cancer and hyperplastic samples, eEF1A1 was significantly upregulated (Cancer Norm.SC: 35; Hyperplastic Norm.SC: 1). Other reports have also noted increased expression of eEF1A1 in non-progressing prostate cancer tissue when compared to hyperplasia samples.[220,221] eEF1A1 has been shown to interact with phosphor-Akt to regulate proliferation, survival and motility in breast cancer.[222] A downregulation of eEF1A1 by RNA interference in DU-145 cells was shown to reduced tumor cell proliferation and inhibited cell migration and invasion.[223]

Integrins are proteins that mediate the attachment of cells to surrounding cells, extracellular matrix and play an important role in cell signaling. Signals for cell growth, division, survival, differentiation and apoptosis are all received through integrins.[224] They have been found to play in integral role in the spread of cancer to bone, and have been found in most malignant cancers.[225] Currently, alpha V integrins are promising therapeutic targets in prostate cancer.[226]

Proteins Downregulated in Cancer

A total of 43 proteins were downregulated by more than 3-fold in our cancer sample. Further analysis determined 5 of these proteins have biological association with prostate cancer, shown in Table 4.[227] Hsp70 is a chaperone important in protein folding and prevention of aggregation. It may also play a role apoptosis. Although multiple reports have identified Hsp70 in prostate cancer analysis, conflicting data exists about its expression. One study reported increased expression of Hsp70 as the disease progresses, while another work found it to be downregulated in prostate cancer.[217,220] Desmin and Hsp beta-1 were also found to be downregulated in our cancer sample. Both proteins are phosphoproteins, and a previous study identified them in all prostate cancer samples analyzed; however, this was after the enrichment of phosphopeptides.[228] Hsp beta-1 has also been identified in both androgen-sensitive and insensitive cell lines, however quantitative comparison with healthy tissue was not reported.[17,201]

Proteins Unique to Healthy Tissue

Analysis of healthy tissue identified 27 proteins that were unique, and 3 have biological significance to prostate cancer. Although 10 kDa heat shock protein, mitochondrial and isoform of sodium/potassium-transporting ATPase were not identified in the carcinoma tissue in this study, they have previously been reported to be expressed in both androgen-sensitive and androgen-insensitive prostate cancer cell lines.[17] The fact that our study did not identify these proteins in the carcinoma tissue is most likely the result of a limited number of samples analyzed. Interestingly, myosin 6 was only identified in the healthy sample. Much debate remains about the expression of myosin 6 as Jansen et al. reported it to be a "tumor-derived" protein, whereas Lapek et al. found this protein to be expressed in healthy tissue, but was absent in tumor and metastatic cell lines.[200,220]

Comparative analysis to previous canine prostate cancer proteomics

A previous study by LeRoy et al. compared the protein expression between healthy and carcinoma prostate tissue in canines.[21] In total, 9 proteins were identified as being overexpressed (> 2.5-fold) in cancer samples, and 6 of these proteins have been found to be biologically significant in prostate cancer or in the progression of other types of cancer. Our analysis produced similar findings, identifying 7 of these proteins, and 6 showed greater than 2.5-fold upregulation in our analysis, shown in Table 6. While vimentin was identified in our study, it was not found to be differentially expressed between healthy and carcinoma prostate samples.

Conclusions:

This work produced the first large-scale study of protein expression in the canine prostate gland. Protein expression was similar in the healthy, prostatitis and hyperplastic prostate samples. Significant differences in expression were observed when comparing healthy and carcinoma tissue. Many differentially expressed proteins identified in this study have been associated with human prostate cancer, particularly late stage prostate cancer, or disease progression in other types of cancer. Furthermore, previous studies have identified several of these proteins as suggested therapeutic targets, current targets of anticancer drugs, or potential biomarkers. One major problems that has restricted prostate cancer research is the relatively limited number of metastatic human prostate cancer cell lines.[17] Canine prostate cancer is both androgen-insensitive and generally highly metastatic, thus suggesting a potential predictive model of late-stage human prostate cancer. Although the sample size in this work was limited, our results appear very promising as many proteins identified in the canine carcinoma sample have been identified in androgen-insensitive and highly metastatic human prostate cancer cell lines. Future studies will be needed to confirm, validate, and expand on these results.

This work also presents a basis for future research in prostatic disease and progression. Very little information is available on protein expression differences in benign hyperplasia versus malignant growth or other prostatic diseases. Analyzing the changes in protein expression among different diseases, as well as protein expression changes throughout the progression of the disease could produce new in-sites that would be useful in understanding pathophysiology, diagnosis, and treatment options.

Acknowledgments:

Table 6.1

Canine Prostate Samples Analyzed and Proteomic Identifications Produced During GeLC-MS/MS Analysis.

Prostate Sample	Breed	Reproductive Status	Age	Protein IDs [a,b]	Peptide IDs [a,b]	Sample Unique Protein IDs
Atrophy	English Bulldog	Neutered	3	304	1124	5
Cancer	Unknown	Unknown	Unknown	426	2318	69
Hyperplastic	Dachshund	Intact	9	354	1704	29
Healthy	Labrador Retriever	Intact	5	466	2231	27
Prostatitis	Unknown	Unknown	Unknown	432	2009	23

[a] Totals produced during triplicate analysis.

[b] All protein and peptide identifications validated with a 5% protein false discovery rate and 0.9 peptide probability.

Table 6.2

Proteins Upregulated by >3 fold in Canine Prostate Cancer Sample That Have Biological Significance to Prostate Cancer.

Protein[a]	Molecular Weight (kDa)	Cancer Norm. SC	Healthy Norm. SC	Biological Function	Reference No.
Alpha-enolase	48.827	32.33	7	Multifunctional: Glycolysis, growth control, hypoxia tolerance, allergic response	17, 200, 201
Annexin A1	38.584	102.67	2.33	Calcium binding protein involved in exocytosis, cell signal transduction associated with inflammation, cell differentiation and proliferation	17, 201,210, 220
Annexin A2	38.612	31.67	9.33	Involved in cell motility, linkage of membrane associated proteins, organization of exocytosis of intracellular proteins to extracellular domain	17, 210, 212, 213
Cathepsin D	44.274	37	4.33	Acid protease involved in intracellular protein breakdown	217
Endoplasmin precursor	92.438	30.67	5	Chaperone molecule important in transporting secreted proteins	17
Ezrin	69.362	18	0.67	Involved in cytoskeletal connections to plasma membrane and the formation of microvilli	217
Filamin B	281.02	36.67	10.33	Actin-binding involved in cell motility	17
Fructose-bisphosphate aldolase A	39.46	45.33	2.33	Involved in glycolysis	200, 207
GRP 78	72.2204	63.67	8.33	Calcium binding chaperone, assists in protein folding	17
Heat Shock Protein 90-alpha	77.168	32.33	7.33	Chaperone molecule involved in cell cycle control and signal transduction	207
Hemopexin	51.287	12	2.33	Plasma protein with the highest known binding affinity to heme, acts as an antioxidant to prevent heme-mediated oxidative stress	221
Lactate dehydrogenase A	36.544	27.33	2.67	Involved in anaerobic glycolysis	200
Myosin-9	226.199	78.67	0.67	Actin-binding involved in cell motility	17, 228
Protein disulfide-isomerase A6	48.25	35	1	Chaperone molecule involved in platelet aggregation	17
Serum albumin	68.5419	1673	442.33	Maintains oncotic pressure, carrier protein of various molecules throughout the body	17
Thrombospondin	129.512	4.33	1.33	Mediates cell interactions, involved in antiangiogenesis	200, 207
Transketolase variant	79.917	32.33	7.33	Enzyme in glycolysis pathway	17, 204
Tubulin beta chain	41.697	11.33	2	Major constituent of microtubules	17
Vitamin D binding Protein	52.922	42.67	7.67	Cell surface protein that binds vitamin D and transports to target tissue	221
Voltage-dependent anion channel 2	31.541	7.33	0.33	Channel forming protein in the outer mitochondrial membrane, allowing anion transport	200

[a] All protein identifications were validated with a 5% protein false discovery rate and a 0.9 peptide probability. All proteins were identified by 2 or more peptides.

Table 6.3

Protein Identifications Unique to the Cancer Sample That Have Biological Significance to Prostate Cancer.

Protein[a]	Molecular Weight (kDa)	Cancer Norm. SC	Healthy Norm. SC	Biological Function	Reference No.
Annexin A3	36.179	1.33	0	Inhibits phospholipase A2, anti-coagulant, and potentially an angiogenic mediator	210, 211
Calreticulin	48.133	8.33	0	A calcium-binding chaperone that interacts with the endoplasmic reticulum to promote oligomeric assembly and other functions	17, 204, 217
Calnexin	67.543	1.33	0	Calcium-binding protein that plays a major role in retaining misfolded proteins in the endoplasmic reticulum	17, 228
Cathepsin Z	41.515	2	0	Proteolysis, ubiquitously expressed in tumors, may be involved in tumorigenesis	200
Ceruloplasmin	122.466	6	0	Copper-binding protein important in transporting iron across cell membrane	221
Elongation factor 1 alpha	50.091	35	0	Important role in translation and nuclear exportation of proteins	220, 221
Fibronectin	270.867	2	0	Major role in cell adhesion, growth, migration, and differentiation	221
integrin alpha V	108.831	3	0	Role in signal transduction, also plays a role in angiogenesis	224
Junction plakoglobin	81.689	1	0	Role in cell adhesion and cell signaling	200, 207
Maltase-glucoamylase, intestinal	209.129	16	0	May play a role in starch metabolism	200

[a] All protein identifications were validated with a 5% protein false discovery rate and a 0.9 peptide probability. All proteins were identified by 2 or more peptides.

Table 6.4

Proteins Downregulated by >3 fold in Canine Prostate Cancer That Have Biological Significance to Prostate Cancer.

Protein[a]	Molecular Weight (kDa)	Cancer Norm. SC	Healthy Norm. SC	Biological Function	Reference No.
Collagen alpha 3(VI) chain precursor	342.435	9.33	83.67	Protein involved in the formation of basement membranes	227
Desmin	53.27	3.33	10.33	Forms a fibrous networks in striated muscle	228
Filamin A	280.432	10.67	194	Structural protein of cytoskeleton, plays a role in cell signaling	17
Heat Shock Protein 70	69.874	2.33	16.67	Chaperone protein important in folding of proteins and preventing aggregation; may play a role in apoptosis	212, 217
Heat Shock Protein beta-1	22.907	4.67	14.33	Chaperone protein involved in cell development, differentiation, and signal transduction	17, 201, 228

[a] All protein identifications were validated with a 5% protein false discovery rate and a 0.9 peptide probability. All proteins were identified by 2 or more peptides.

Table 6.5

Protein Identifications Unique to the Healthy Tissue Sample That Have Biological Significance to Prostate Cancer.

Protein[a]	Molecular Weight (kDa)	Cancer Norm. SC	Intact Norm. SC	Biological Function	Reference No.
10 kDa Heat Shock Protein, Mitochondrial	12.673	0	6	Important in mitochondrial biogenesis, also plays a role in inhibiting the immune response.	17
Isoform of sodium/potassium-transporting ATPase alpha-1 chain	112.577	0	7.33	Catalyzes the hydrolysis of ATP coupled with the exchange of sodium and phosphate ions across the plasma membrane	17
Myosin, light polypeptide 6	16.932	0	16.33	Non-calcium binding regulatory light chain of myosin	200

[a] All protein identifications were validated with a 5% protein false discovery rate and a 0.9 peptide probability. All proteins were identified by 2 or more peptides.

Table 6.6

An Analysis of Protein Expression Examining Proteins Previously Identified in Canine Prostate Cancer with Biological Significance to Disease Progression[a]

Protein	Protein MW (kDA)	Cancer Peptide IDs	Cancer Normalized Spectral Ct	Intact Peptide IDs	Intact Normalized Spectral Ct	Expression Difference
Serum Albumin	68.5419	182	1673	129	442.33	3.8
Vimentin isoform 12	53.547	31	53.33	29	47.33	1.1
Haptoglobin isoform 2	38.3403	44	91.33	33	32.67	2.8
Serotransferrin isoform 2 Serotransferrin isoform 6	78.0375 77.9863	44	88.67	15	23	3.9
GRP 78	72.2204	28	63.67	9	8.33	7.6
Alpha-enolase isoform 1	48.8274	16	32.33	11	7	4.6
Endoplasmin Precursor	92.4384	17	30.67	7	5	6.1

[a] All protein identifications were previously identified by LeRoy et al. displaying greater than 2.5-fold expression differences between healthy and carcinoma prostate tissue. All proteins were previously reported for their association with prostate cancer or progression of other types of cancer.[21]
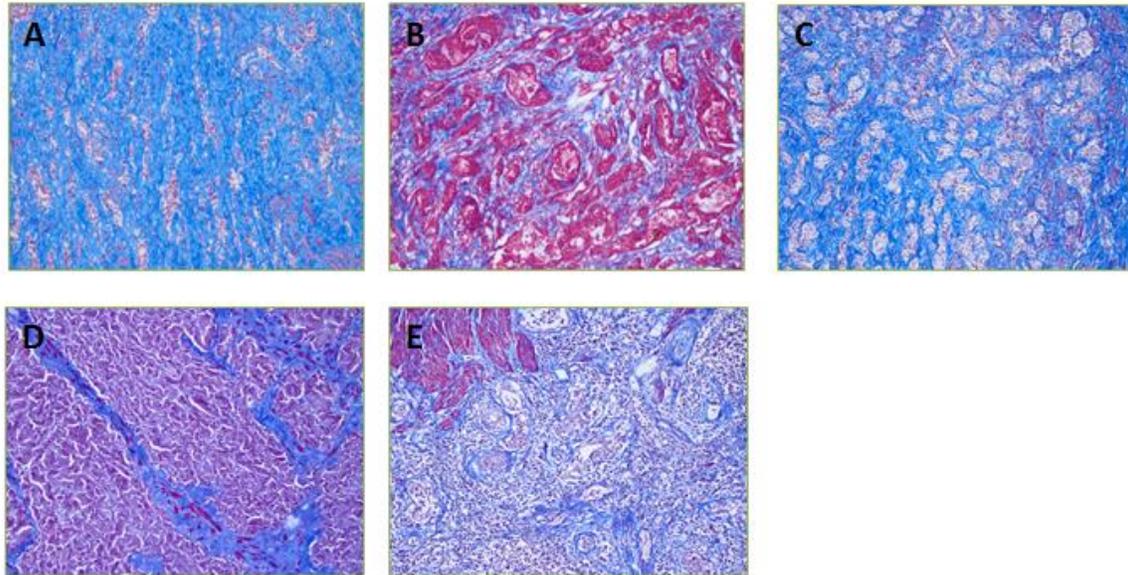
Figure 6.1 Trichrome Stain of Canine Prostate Gland Samples. (A) Atrophied canine prostate gland, denoted by the angulated glands and lack of cytoplasm. (B) Canine Prostate Cancer sample, denoted by the proliferation of glandular cells, lacking any organization or lumens. (C) Healthy canine prostate gland, denoted by nicely defined tubuloalveolar glands. (D) Canine Benign Hyperplasia Prostate gland, denoted by the presence of organized tubuloalveolar glands with proliferative cell growth. (E) Canine Prostatitis, denoted by the abundance of inflammatory cells present.

CHAPTER 7

CONCLUSIONS

The overall purpose of this work was to develop high efficiency separation techniques to improve the throughput of MS-based proteomic analysis. These separation techniques were then applied to the analysis of canine prostate tissue, identifying several proteins that suggest the canine is a relevant animal model of androgen-insensitive, highly aggressive human prostate cancer.

In Chapter 3, superficially porous particles were used to improve LC separations and reduce overall LC-MS experiment time. While these high efficiency particles doubled peak capacity, protein sequence coverage was diminished due to data-dependent acquisition (DDA) settings that were not properly set to match narrow chromatographic peak widths. Minimum MS signal intensity, repeat count, repeat duration and dynamic exclusion were systematically optimized to match chromatographic widths, allowing for reduction of experiment time without sacrifice of MS/MS spectra quality. Analysis of an authentic proteomic sample demonstrated the effectiveness of fast LC separations and optimized DDA settings, as the number of proteins identified doubled and sequence coverage was improved, with experiment time being reduced by a factor of 5.

Chapter 4 demonstrates the utility of formic acid and ammonium formate (FA/AF) as a mobile phase modifier for LC-MS analysis of tryptic digests. This work found FA/AF to be compatible with electrospray ionization, determining the loss in mass spectrometer signal intensity was minimal. FA/AF mobile phases reduced chromatographic peak widths and improved peak shape, leading to increased peak capacity. We also demonstrated sample load

tolerance can be significantly improved with FA/AF mobile phases. These improvements in LC separations lead to increased protein identifications and improved sequence coverage. Analysis of protein identifications determined FA/AF mobile phases provide improved retention and separation of basic peptides, and do not impact the charge state of tryptic peptides.

The protein identification efficiency of GeLC separations was analyzed in Chapter 5. Moderate 1st and 2nd dimension separation efficiency provided the most protein identifications in this fixed instrument time format. The biggest factor influencing protein identification efficiency was the percentage of total experiment time dedicated to LC gradient elution conditions. Our results show gel overlap is widespread and independent of the number of gel slices, thus no optimal number of slices to collect was determined. The increased depth of proteome analysis reported with higher numbers of gel slices is due to additional MS analysis time rather than increased separation efficiency from collecting higher numbers of gel slices. We also conclude peak capacity is not an appropriate measurement of GeLC separating power. This work also demonstrates that increased efficiency of MS instrument time can help offset the reduction in LC peak capacity resulting from reduced LC gradient lengths.

In Chapter 6 the canine prostate was analyzed, and the potential use of an animal model for human prostate cancer was explored. Protein expression in canine prostate carcinoma tissue was significantly different than in a healthy prostate gland. Our results suggest the canine can be a relevant animal model of androgen-insensitive, highly aggressive human prostate cancer. This proteomic analysis identified numerous proteins that have biological significance to prostate cancer development and progression. Many of the proteins identified in this work have been suggested to be therapeutic targets, current targets of anticancer drugs or potential biomarkers.

REFERENCES

1.   Harding A. LIFE SCIENCE TECHNOLOGIES: Panning the Proteome for Biomarker Gold. *Science.* 2012;337(6098):1120-1122.

2.   Tyers M, Mann M. From genomics to proteomics. *Nature.* 2003;422(6928):193-197.

3.   Aebersold R, Goodlett DR. Mass spectrometry in proteomics. *proteomics.* 2001;3:5.

4.   Cox J, Mann M. Is proteomics the new genomics? *Cell.* 2007;130(3):395-398.

5.   Schuster S, Wagner B, Boyes B, Kirkland J. Wider pore superficially porous particles for peptide separations by HPLC. *Journal of chromatographic science.* 2010;48(7):566-571.

6.   Schuster SA, Boyes BE, Wagner BM, Kirkland JJ. Fast high performance liquid chromatography separations for proteomic applications using Fused-Core® silica particles. *Journal of Chromatography A.* 2011.

7.   Hsieh Y, Duncan CJ, Brisson J-M. Fused-core silica column high-performance liquid chromatography/tandem mass spectrometric determination of rimonabant in mouse plasma. *Analytical chemistry.* 2007;79(15):5668-5673.

8.   Kirkland JJ, Langlois TJ, DeStefano JJ. Fused Core Particles for HPLC. *American Laboratory.* 2007:19.

9.   Apffel A, Fischer S, Goldberg G, Goodley PC, Kuhlmann FE. Enhanced sensitivity for peptide mapping with electrospray liquid chromatography-mass spectrometry in the presence of signal suppression due to trifluoroacetic acid-containing mobile phases. *Journal of Chromatography A.* 1995;712(1):177-190.

10.   Gustavsson SÅ, Samskog J, Markides KE, Långström B. Studies of signal suppression in liquid chromatography–electrospray ionization mass spectrometry using volatile ion-pairing reagents. *Journal of Chromatography A.* 2001;937(1):41-47.

11.   Eshraghi J, Chowdhury SK. Factors affecting electrospray ionization of effluents containing trifluoroacetic acid for high-performance liquid chromatography/mass spectrometry. *Analytical chemistry.* 1993;65(23):3528-3533.

12.   McCalley DV. Some practical comparisons of the efficiency and overloading behaviour of sub-2μm porous and sub-3μm shell particles in reversed-phase liquid chromatography. *Journal of Chromatography A.* 2011;1218(20):2887-2897.

13. McCalley DV. Rationalization of retention and overloading behavior of basic compounds in reversed-phase HPLC using low ionic strength buffers suitable for mass spectrometric detection. *Analytical chemistry.* 2003;75(14):3404-3410.

14. Schuster SA, Boyes BE, Wagner BM, Kirkland JJ. Fast high performance liquid chromatography separations for proteomic applications using Fused-Core< sup>®</sup> silica particles. *Journal of Chromatography A.* 2012;1228:232-241.

15. McCalley DV. Effect of buffer on peak shape of peptides in reversed-phase high performance liquid chromatography. *Journal of Chromatography A.* 2004;1038(1):77-84.

16. Hu L, Ye M, Jiang X, Feng S, Zou H. Advances in hyphenated analytical techniques for shotgun proteome and peptidome analysis—a review. *Analytica chimica acta.* 2007;598(2):193-204.

17. Glen A, Evans CA, Gan CS, et al. Eight-plex iTRAQ analysis of variant metastatic human prostate cancer cells identifies candidate biomarkers of progression: An exploratory study. *The Prostate.* 2010;70(12):1313-1332.

18. Leav I, Ling GV. Adenocarcinoma of the canine prostate. *Cancer.* 2006;22(6):1329-1345.

19. Priester WA. The occurrence of tumors in domestic animals. *United States. National Cancer Institute. Monograph.* 1980.

20. Cornell KK, Bostwick DG, Cooley DM, et al. Clinical and pathologic aspects of spontaneous canine prostate carcinoma: a retrospective analysis of 76 cases. *The Prostate.* 2000;45(2):173-183.

21. LeRoy B, Painter A, Sheppard H, Popiolek L, Samuel-Foo M, Andacht T. Protein expression profiling of normal and neoplastic canine prostate and bladder tissue. *Veterinary and Comparative Oncology.* 2007;5(2):119-130.

22. Yates JR. Mass spectrometry: from genomics to proteomics. *Trends in Genetics.* 2000;16(1):5-8.

23. Vidal M. A Biological Atlas of Functional Maps Review. *Cell.* 2001;104:333-339.

24. Schmutz J, Wheeler J, Grimwood J, et al. Quality assessment of the human genome sequence. *Nature.* 2004;429(6990):365-368.

25. Collins FS, Mansoura MK. The human genome project. *Cancer.* 2001;91(S1):221-225.

26. Patterson SD, Aebersold RH. Proteomics: the first decade and beyond. *Nature genetics.* 2003;33:311-323.

27. Wasinger VC, Cordwell SJ, Cerpa-Poljak A, et al. Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium. *Electrophoresis.* 1995;16(1):1090-1094.

28.   James P. Protein identification in the post-genome era: the rapid rise of proteomics. *Quarterly reviews of biophysics.* 1997;30(4):279-331.

29.   Aebersold R. Constellations in a cellular universe. *Nature.* 2003;422(6928):115-116.

30.   Collins F, Lander E, Rogers J, Waterston R, Conso I. Finishing the euchromatic sequence of the human genome. *Nature.* 2004;431(7011):931-945.

31.   Anderson NL, Anderson NG. The human plasma proteome history, character, and diagnostic prospects. *Molecular & Cellular Proteomics.* 2002;1(11):845-867.

32.   Blackstock WP, Weir MP. Proteomics: quantitative and physical mapping of cellular proteins. *Trends in biotechnology.* 1999;17(3):121.

33.   Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature.* 2003;422(6928):198-207.

34.   Domon B, Aebersold R. Mass spectrometry and protein analysis. *Science Signaling.* 2006;312(5771):212.

35.   Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical chemistry.* 1988;60(20):2299-2301.

36.   Hillenkamp F, Karas M, Beavis RC, Chait BT. Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. *Analytical Chemistry.* 1991;63(24):1193A-1203A.

37.   Beavis RC, Chait BT. Rapid, sensitive analysis of protein mixtures by mass spectrometry. *Proceedings of the National Academy of Sciences.* 1990;87(17):6873-6877.

38.   Krause E, Wenschuh H, Jungblut PR. The dominance of arginine-containing peptides in MALDI-derived tryptic mass fingerprints of proteins. *Analytical chemistry.* 1999;71(19):4160-4165.

39.   Stapels MD, Barofsky DF. Complementary use of MALDI and ESI for the HPLC-MS/MS analysis of DNA-binding proteins. *Analytical chemistry.* 2004;76(18):5423-5430.

40.   Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray ionization for mass spectrometry of large biomolecules. *Science.* 1989;246(4926):64-71.

41.   Huang L, Riggin RM. Analysis of nonderivatized neutral and sialylated oligosaccharides by electrospray mass spectrometry. *Analytical chemistry.* 2000;72(15):3539-3546.

42.   Cech NB, Enke CG. Relating electrospray ionization response to nonpolar character of small peptides. *Analytical chemistry.* 2000;72(13):2717-2723.

43.   Wysocki VH, Resing KA, Zhang Q, Cheng G. Mass spectrometry of peptides and proteins. *Methods.* 2005;35(3):211-222.

44.   Finnigan RE. Quadrupole mass spectrometers. *Analytical Chemistry.* 1994;66(19):969A-975A.

45.   Gygi SP, Aebersold R. Mass spectrometry and proteomics. *Current opinion in chemical biology.* 2000;4(5):489-494.

46.   Yost RA, Boyd RK. [7] Tandem mass spectrometry: Quadrupole and hybrid instruments. *Methods in enzymology.* 1990;193:154-200.

47.   Cotter RJ. Time-of-flight mass spectrometry for the structural analysis of biological molecules. *Analytical chemistry.* 1992;64(21):1027A-1039A.

48.   Cotter RJ. Time-of-flight mass spectrometry: An increasing role in the life sciences. *Biological Mass Spectrometry.* 1989;18(8):513-532.

49.   Brown RS, Lennon JJ. Mass resolution improvement by incorporation of pulsed ion extraction in a matrix-assisted laser desorption/ionization linear time-of-flight mass spectrometer. *Analytical chemistry.* 1995;67(13):1998-2003.

50.   Fancher CA, Woods AS, Cotter RJ. Improving the sensitivity of the end-cap reflectron time-of-flight mass spectrometer. *Journal of mass spectrometry.* 2000;35(2):157-162.

51.   Medzihradszky KF, Campbell JM, Baldwin MA, et al. The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer. *Analytical Chemistry.* 2000;72(3):552-558.

52.   Ens W, Standing KG. Hybrid quadrupole/time-of-flight mass spectrometers for analysis of biomolecules. *Methods in enzymology.* 2005;402:49.

53.   Shevchenko A, Loboda A, Shevchenko A, Ens W, Standing KG. MALDI quadrupole time-of-flight mass spectrometry: a powerful tool for proteomic research. *Analytical chemistry.* 2000;72(9):2132-2141.

54.   Krutchinsky AN, Zhang W, Chait BT. Rapidly switchable matrix-assisted laser desorption/ionization and electrospray quadrupole-time-of-flight mass spectrometry for protein identification. *Journal of the American Society for Mass Spectrometry.* 2000;11(6):493-504.

55.   Cha B, Blades M, Douglas D. An interface with a linear quadrupole ion guide for an electrospray-ion trap mass spectrometer system. *Analytical chemistry.* 2000;72(22):5647-5654.

56.   Stafford Jr G. Ion trap mass spectrometry: a personal perspective. *Journal of the American Society for Mass spectrometry.* 2002;13(6):589-596.

57. Jonscher KR, Yates III JR. The quadrupole ion trap mass spectrometer—a small solution to a big challenge. *Analytical biochemistry.* 1997;244(1):1-15.

58. Schwartz JC, Senko MW, Syka JE. A two-dimensional quadrupole ion trap mass spectrometer. *Journal of the American Society for Mass Spectrometry.* 2002;13(6):659-669.

59. Mayya V, Rezaul K, Cong Y-S, Han D. Systematic comparison of a two-dimensional ion trap and a three-dimensional ion trap mass spectrometer in proteomics. *Molecular & Cellular Proteomics.* 2005;4(2):214-223.

60. Hager JW, Le Blanc J. High-performance liquid chromatography–tandem mass spectrometry with a new quadrupole/linear ion trap instrument. *Journal of Chromatography A.* 2003;1020(1):3-9.

61. Marshall AG, Hendrickson CL, Jackson GS. Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass spectrometry reviews.* 1998;17(1):1-35.

62. Comisarow MB, Marshall AG. The Early Development of Fourier Transform Ion Cyclotron Resonance (FT-ICR) Spectroscopy. *Journal of mass spectrometry.* 1996;31(6):581-585.

63. Emmett MR, White FM, Hendrickson CL, Shi SD-H, Marshall AG. Application of micro-electrospray liquid chromatography techniques to FT-ICR MS to enable high-sensitivity biological analysis. *Journal of the American Society for Mass Spectrometry.* 1998;9(4):333-340.

64. Hernandez H, Niehauser S, Boltz SA, Gawandi V, Phillips RS, Amster IJ. Mass defect labeling of cysteine for improving peptide assignment in shotgun proteomic analyses. *Analytical chemistry.* 2006;78(10):3417-3423.

65. Bogdanov B, Smith RD. Proteomics by FTICR mass spectrometry: top down and bottom up. *Mass Spectrometry Reviews.* 2005;24(2):168-200.

66. Hardman M, Makarov AA. Interfacing the orbitrap mass analyzer to an electrospray ion source. *Analytical chemistry.* 2003;75(7):1699-1705.

67. Makarov A. Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Analytical chemistry.* 2000;72(6):1156-1162.

68. Makarov A, Denisov E, Kholomeev A, et al. Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Analytical chemistry.* 2006;78(7):2113-2120.

69. Olsen JV, de Godoy LM, Li G, et al. Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Molecular & Cellular Proteomics.* 2005;4(12):2010-2021.

70. Yates JR, Cociorva D, Liao L, Zabrouskov V. Performance of a linear ion trap-Orbitrap hybrid for peptide analysis. *Analytical chemistry.* 2006;78(2):493-500.

71. Fournier ML, Gilmore JM, Martin-Brown SA, Washburn MP. Multidimensional separations-based shotgun proteomics. *Chemical Reviews-Columbus.* 2007;107(8):3654-3686.

72. McDonald WH, Ohi R, Miyamoto DT, Mitchison TJ, Yates III JR. Comparison of three directly coupled HPLC MS/MS strategies for identification of proteins from complex mixtures: single-dimension LC-MS/MS, 2-phase MudPIT, and 3-phase MudPIT. *International Journal of Mass Spectrometry.* 2002;219(1):245-251.

73. Delmotte N, Lasaosa M, Tholey A, Heinzle E, Huber CG. Two-dimensional reversed-phase× ion-pair reversed-phase HPLC: an alternative approach to high-resolution peptide separation for shotgun proteome analysis. *Journal of proteome research.* 2007;6(11):4363-4373.

74. Tang H-Y, Beer LA, Speicher DW. In-depth analysis of a plasma or serum proteome using a 4D protein profiling method. *Serum/Plasma Proteomics*: Springer; 2011:47-67.

75. Wang H, Chang-Wong T, Tang H-Y, Speicher DW. Comparison of extensive protein fractionation and repetitive LC-MS/MS analyses on depth of analysis for complex proteomes. *Journal of proteome research.* 2010;9(2):1032-1040.

76. Wolters DA, Washburn MP, Yates JR. An automated multidimensional protein identification technology for shotgun proteomics. *Analytical chemistry.* 2001;73(23):5683-5690.

77. Malmström J, Lee H, Nesvizhskii AI, et al. Optimized peptide separation and identification for mass spectrometry based proteomics via free-flow electrophoresis. *Journal of proteome research.* 2006;5(9):2241-2249.

78. Peng J, Gygi SP. Proteomics: the move to mixtures. *Journal of Mass Spectrometry.* 2001;36(10):1083-1091.

79. Mohan D, Paša-Tolic L, Masselon CD, et al. Integration of electrokinetic-based multidimensional separation/concentration platform with electrospray ionization-Fourier transform ion cyclotron resonance-mass spectrometry for proteome analysis of Shewanella oneidensis. *Analytical chemistry.* 2003;75(17):4432-4440.

80. Simpson DC, Ahn S, Pasa-Tolic L, et al. Using size exclusion chromatography-RPLC and RPLC-CIEF as two-dimensional separation strategies for protein profiling. *Electrophoresis.* 2006;27(13):2722-2733.

81. Yin H, Killeen K, Brennen R, Sobek D, Werlich M, van de Goor T. Microfluidic chip for peptide analysis with an integrated HPLC column, sample enrichment column, and nanoelectrospray tip. *Analytical chemistry.* 2005;77(2):527-533.

82. Shen Y, Smith RD, Unger KK, Kumar D, Lubda D. Ultrahigh-throughput proteomics using fast RPLC separations with ESI-MS/MS. *Analytical chemistry.* 2005;77(20):6692-6701.

83. Heller M, Michel PE, Morier P, et al. Two-stage Off-Gel™ isoelectric focusing: Protein followed by peptide fractionation and application to proteome analysis of human plasma. *Electrophoresis.* 2005;26(6):1174-1188.

84. Xie H, Bandhakavi S, Griffin TJ. Evaluating Preparative Isoelectric Focusing of Complex Peptide Mixtures for Tandem Mass Spectrometry-Based Proteomics: A Case Study in Profiling Chromatin-Enriched Subcellular Fractions in Saccharomyces c erevisiae. *Analytical chemistry.* 2005;77(10):3198-3207.

85. Link AJ, Eng J, Schieltz DM, et al. Direct analysis of protein complexes using mass spectrometry. *Nature biotechnology.* 1999;17(7):676-682.

86. Breci L, Hattrup E, Keeler M, Letarte J, Johnson R, Haynes PA. Comprehensive proteomics in yeast using chromatographic fractionation, gas phase fractionation, protein gel electrophoresis, and isoelectric focusing. *Proteomics.* 2005;5(8):2018-2028.

87. Blonder J, Rodriguez-Galan MC, Lucas DA, et al. Proteomic investigation of natural killer cell microsomes using gas-phase fractionation by mass spectrometry. *Biochimica et biophysica acta.* 2004;1698(1):87.

88. Premstaller A, Oberacher H, Walcher W, et al. High-performance liquid chromatography-electrospray ionization mass spectrometry using monolithic capillary columns for proteomic studies. *Analytical chemistry.* 2001;73(11):2390-2396.

89. Friedman DB, Hill S, Keller JW, et al. Proteome analysis of human colon cancer by two-dimensional difference gel electrophoresis and mass spectrometry. *Proteomics.* 2004;4(3):793-811.

90. Wang H, Hanash S. Multi-dimensional liquid phase based separations in proteomics. *Journal of Chromatography B.* 2003;787(1):11-18.

91. Chait BT. Mass spectrometry: Bottom-up or top-down? *Science(Washington).* 2006;314(5796):65-66.

92. Resing KA, Ahn NG. Proteomics strategies for protein identification. *FEBS letters.* 2005;579(4):885-889.

93. Forbes AJ, Patrie SM, Taylor GK, Kim Y-B, Jiang L, Kelleher NL. Targeted analysis and discovery of posttranslational modifications in proteins from methanogenic archaea by top-down MS. *Proceedings of the National Academy of Sciences of the United States of America.* 2004;101(9):2678-2683.

94.   Zubarev RA, Horn DM, Fridriksson EK, et al. Electron capture dissociation for structural characterization of multiply charged protein cations. *Analytical chemistry.* 2000;72(3):563-573.

95.   Coon JJ, Ueberheide B, Syka JE, et al. Protein identification using sequential ion/ion reactions and tandem mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America.* 2005;102(27):9463-9468.

96.   Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America.* 2004;101(26):9528-9533.

97.   Reid GE, McLuckey SA. 'Top down'protein characterization via tandem mass spectrometry. *Journal of mass spectrometry.* 2002;37(7):663-675.

98.   Kelleher NL. Peer reviewed: Top-down proteomics. *Analytical chemistry.* 2004;76(11):196 A-203 A.

99.   Parks BA, Jiang L, Thomas PM, et al. Top-down proteomics on a chromatographic time scale using linear ion trap Fourier transform hybrid mass spectrometers. *Analytical chemistry.* 2007;79(21):7984-7991.

100.   Li W, Hendrickson CL, Emmett MR, Marshall AG. Identification of intact proteins in mixtures by alternated capillary liquid chromatography electrospray ionization and LC ESI infrared multiphoton dissociation Fourier transform ion cyclotron resonance mass spectrometry. *Analytical chemistry.* 1999;71(19):4397-4402.

101.   Taylor GK, Kim Y-B, Forbes AJ, Meng F, McCarthy R, Kelleher NL. Web and database software for identification of intact proteins using "top down" mass spectrometry. *Analytical chemistry.* 2003;75(16):4081-4086.

102.   Zamdborg L, LeDuc RD, Glowacz KJ, et al. ProSight PTM 2.0: improved protein identification and characterization for top down mass spectrometry. *Nucleic acids research.* 2007;35(suppl 2):W701-W706.

103.   LeDuc RD, Taylor GK, Kim Y-B, et al. ProSight PTM: an integrated environment for protein identification and characterization by top-down mass spectrometry. *Nucleic acids research.* 2004;32(suppl 2):W340-W345.

104.   LeDuc RD. *Bioinformatics of High Throughput Proteomics Using Tandem Mass Spectrometry of Intact Proteins.* ProQuest; 2007.

105.   Horn DM, Zubarev RA, McLafferty FW. Automated reduction and interpretation of high resolution electrospray mass spectra of large molecules. *Journal of the American Society for Mass Spectrometry.* 2000;11(4):320-332.

106.   Han X, Aslanian A, Yates III JR. Mass spectrometry for proteomics. *Current opinion in chemical biology.* 2008;12(5):483.

107. Henzel WJ, Watanabe C, Stults JT. Protein identification: the origins of peptide mass fingerprinting. *Journal of the American Society for Mass Spectrometry.* 2003;14(9):931-942.

108. Henzel WJ, Billeci TM, Stults JT, Wong SC, Grimley C, Watanabe C. Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proceedings of the National Academy of Sciences.* 1993;90(11):5011-5015.

109. Washburn MP, Wolters D, Yates JR. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature biotechnology.* 2001;19(3):242-247.

110. Shevchenko A, Henrik Tomas JH, sbreve, Olsen JV, Mann M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nature protocols.* 2007;1(6):2856-2860.

111. Giddings JC. Two-dimensional separations: concept and promise. *Analytical chemistry.* 1984;56(12):1258A-1270A.

112. Davis JM, Giddings JC. Origin and characterization of departures from the statistical model of component-peak overlap in chromatography. *Journal of Chromatography A.* 1984;289:277-298.

113. Davis JM, Giddings JC. Statistical method for estimation of number of components from single complex chromatograms: application to experimental chromatograms. *Analytical chemistry.* 1985;57(12):2178-2182.

114. Schirmer EC, Yates 3rd J, Gerace L. MudPIT: A powerful proteomics tool for discovery. *Discovery medicine.* 2003;3(18):38.

115. Florens L, Washburn MP. Proteomic analysis by multidimensional protein identification technology. *New and Emerging Proteomic Techniques*: Springer; 2006:159-175.

116. Florens L, Washburn MP, Raine JD, et al. A proteomic view of the Plasmodium falciparum life cycle. *Nature.* 2002;419(6906):520-526.

117. Tomomori-Sato C, Sato S, Parmely TJ, et al. A mammalian mediator subunit that shares properties with Saccharomyces cerevisiae mediator subunit Cse2. *Journal of Biological Chemistry.* 2004;279(7):5846-5851.

118. Zybailov B, Mosley AL, Sardiu ME, Coleman MK, Florens L, Washburn MP. Statistical Analysis of Membrane Proteome Expression Changes in Saccharomyces c erevisiae. *Journal of proteome research.* 2006;5(9):2339-2347.

119. Zybailov B, Coleman MK, Florens L, Washburn MP. Correlation of relative abundance ratios derived from peptide ion chromatograms and spectrum counting for quantitative

proteomic analysis using stable isotope labeling. *Analytical chemistry.* 2005;77(19):6218-6224.

120. Yang Y, Thannhauser TW, Li L, Zhang S. Development of an integrated approach for evaluation of 2-D gel image analysis: Impact of multiple proteins in single spots on comparative proteomics in conventional 2-D gel/MALDI workflow. *Electrophoresis.* 2007;28(12):2080-2094.

121. Rezaul K, Wu L, Mayya V, Hwang S-I, Han D. A systematic characterization of mitochondrial proteome from human T leukemia cells. *Molecular & Cellular Proteomics.* 2005;4(2):169-181.

122. Zhu W, Venable J, Giometti CS, et al. Large-scale μLC-MS/MS for silver-and Coomassie blue-stained polyacrylamide gels. *Electrophoresis.* 2005;26(23):4495-4507.

123. Shevchenko A, Loboda A, Ens W, Schraven B, Standing KG, Shevchenko A. Archived polyacrylamide gels as a resource for proteome characterization by mass spectrometry. *Electrophoresis.* 2001;22(6):1194-1203.

124. Speicher K, Kolbas O, Harper S, Speicher D. Systematic analysis of peptide recoveries from in-gel digestions for protein identifications in proteome studies. *Journal of biomolecular techniques: JBT.* 2000;11(2):74.

125. Fang Y, Robinson DP, Foster LJ. Quantitative analysis of proteome coverage and recovery rates for upstream fractionation methods in proteomics. *Journal of proteome research.* 2010;9(4):1902-1912.

126. Mann M, Hendrickson RC, Pandey A. Analysis of proteins and proteomes by mass spectrometry. *Annual review of biochemistry.* 2001;70(1):437-473.

127. Romijn EP, Krijgsveld J, Heck AJ. Recent liquid chromatographic–(tandem) mass spectrometric applications in proteomics. *Journal of Chromatography A.* 2003;1000(1):589-608.

128. Mirza UA, Chait BT. Effects of anions on the positive ion electrospray ionization mass spectra of peptides and proteins. *Analytical chemistry.* 1994;66(18):2898-2904.

129. McNair HM, Miller JM. Basic Concepts and Terms. *Basic Gas Chromatography, Second Edition.* 2009:29-52.

130. Van Deemter J, Zuiderweg F, Klinkenberg A. Longitudinal diffusion and resistance to mass transfer as causes of nonideality in chromatography. *Chemical Engineering Science.* 1956;5(6):271-289.

131. Knox JH, Scott HP. < i> B</i> and< i> C</i> terms in the Van Deemter equation for liquid chromatography. *Journal of Chromatography A.* 1983;282:297-313.

132.	Salisbury JJ. Fused-core particles: a practical alternative to sub-2 micron particles. *Journal of chromatographic science.* 2008;46(10):883-886.

133.	Fekete S, Fekete J, Ganzler K. Characterization of new types of stationary phases for fast liquid chromatographic applications. *Journal of pharmaceutical and biomedical analysis.* 2009;50(5):703-709.

134.	Nguyen DTT, Guillarme D, Rudaz S, Veuthey JL. Fast analysis in liquid chromatography using small particle size and high pressure. *Journal of separation science.* 2006;29(12):1836-1848.

135.	Marchetti N, Guiochon G. High peak capacity separations of peptides in reversed-phase gradient elution liquid chromatography on columns packed with porous shell particles. *Journal of Chromatography A.* 2007;1176(1-2):206-216.

136.	Neue UD. Theory of peak capacity in gradient elution. *Journal of Chromatography A.* 2005;1079(1):153-161.

137.	Fairchild JN, Walworth MJ, Horváth K, Guiochon G. Correlation between peak capacity and protein sequence coverage in proteomics analysis by liquid chromatography-mass spectrometry/mass spectrometry. *Journal of Chromatography A.* 2010;1217(29):4779-4783.

138.	Wang X, Stoll DR, Schellinger AP, Carr PW. Peak capacity optimization of peptide separations in reversed-phase gradient elution chromatography: fixed column format. *Analytical chemistry.* 2006;78(10):3406-3416.

139.	Chalkley RJ, Hansen KC, Baldwin MA. Bioinformatic methods to exploit mass spectrometric data for proteomic applications. *Methods in enzymology.* 2005;402:289.

140.	Patterson SD. Data analysis—the Achilles heel of proteomics. *Nature biotechnology.* 2003;21(3):221-222.

141.	Tabb DL, MacCoss MJ, Wu CC, Anderson SD, Yates JR. Similarity among tandem mass spectra from proteomic experiments: detection, significance, and utility. *Analytical chemistry.* 2003;75(10):2470-2477.

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

143.	Cottrell J, London U. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis.* 1999;20(18):3551-3567.

144.	Craig R, Beavis RC. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics.* 2004;20(9):1466-1467.

145. Ma B, Zhang K, Hendrie C, et al. PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid communications in mass spectrometry.* 2003;17(20):2337-2342.

146. Johnson RS, Taylor JA. Searching sequence databases via de novo peptide sequencing by tandem mass spectrometry. *Molecular biotechnology.* 2002;22(3):301-315.

147. Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Analytical chemistry.* 2002;74(20):5383-5392.

148. Nesvizhskii AI, Vitek O, Aebersold R. Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nature methods.* 2007;4(10):787-797.

149. Johnson RS, Davis MT, Taylor JA, Patterson SD. Informatics for protein identification by mass spectrometry. *Methods.* 2005;35(3):223-236.

150. Elias JE, Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nature methods.* 2007;4(3):207-214.

151. Qian W-J, Liu T, Monroe ME, et al. Probability-based evaluation of peptide and protein identifications from tandem mass spectrometry and SEQUEST analysis: the human proteome. *Journal of proteome research.* 2005;4(1):53-62.

152. Rowell JL, McCarthy DO, Alvarez CE. Dog models of naturally occurring cancer. *Trends in molecular medicine.* 2011;17(7):380-388.

153. Tamburini BA, Trapp S, Phang TL, Schappa JT, Hunter LE, Modiano JF. Gene expression profiles of sporadic canine hemangiosarcoma are uniquely associated with breed. *PLoS One.* 2009;4(5):e5549.

154. Paoloni M, Khanna C. Translation of new cancer treatments from pet dogs to humans. *Nature Reviews Cancer.* 2008;8(2):147-156.

155. Clarke R. Animal models of breast cancer: their diversity and role in biomedical research. *Breast cancer research and treatment.* 1996;39(1):1-6.

156. Chen J, Kähne T, Röcken C, et al. Proteome analysis of gastric cancer metastasis by two-dimensional gel electrophoresis and matrix assisted laser desorption/ionization-mass spectrometry for identification of metastasis-related proteins. *Journal of proteome research.* 2004;3(5):1009-1016.

157. Knapp DW, Glickman NW, DeNicola DB, Bonney PL, Lin TL, Glickman LT. Naturally-occurring canine transitional cell carcinoma of the urinary bladder A relevant model of human invasive bladder cancer. Paper presented at: Urologic Oncology: Seminars and Original Investigations2000.

158.   Misdorp W, Weijer K. Animal model of human disease: breast cancer. *The American Journal of Pathology.* 1980;98(2):573.

159.   Hingorani SR, Petricoin III EF, Maitra A, et al. Preinvasive and invasive ductal pancreatic cancer and its early detection in the mouse. *Cancer cell.* 2003;4(6):437-450.

160.   Gondo Y, Fukumura R, Murata T, Makino S. Next-generation gene targeting in the mouse for functional genomics. *BMB Rep.* 2009;42(6):315-323.

161.   Sargan D. IDID: inherited diseases in dogs: web-based information for canine inherited disease genetics. *Mammalian genome.* 2004;15(6):503-506.

162.   Parker HG, Shearin AL, Ostrander EA. Man's best friend becomes biology's best in show: genome analyses in the domestic dog. *Annual review of genetics.* 2010;44:309.

163.   LeRoy BE, Northrup N. Prostate cancer in dogs: comparative and clinical aspects. *The Veterinary Journal.* 2009;180(2):149-162.

164.   Rosol TJ, Tannehill-Gregg SH, LeRoy BE, Mandl S, Contag CH. Animal models of bone metastasis. *Cancer.* 2003;97(S3):748-757.

165.   Chen CHW. Review of a current role of mass spectrometry for proteome research. *Analytica Chimica Acta.* 2008;624(1):16-36.

166.   DeStefano J, Langlois T, Kirkland J. Characteristics of Superficially-Porous Silica Particles for Fast HPLC: Some Performance Comparisons with Sub-2-m Particles. *Journal of chromatographic science.* 2008;46(3):254-260.

167.   Ali I, Gaitonde VD, Grahn A. Halo Columns: New Generation Technology for High Speed Liquid Chromatography. *Journal of chromatographic science.* 2010;48(5):386-394.

168.   Kiss I, Bacskay I, Kilár F, Felinger A. Comparison of the mass transfer in totally porous and superficially porous stationary phases in liquid chromatography. *Analytical and Bioanalytical Chemistry.* 2010;397(3):1307-1314.

169.   Hsieh Y, Duncan CJG, Brisson JM. Fused-core silica column high-performance liquid chromatography/tandem mass spectrometric determination of rimonabant in mouse plasma. *Analytical chemistry.* 2007;79(15):5668-5673.

170.   Gritti F, Leonardis I, Abia J, Guiochon G. Physical properties and structure of fine core-shell particles used as packing materials for chromatography:: Relationships between particle characteristics and column performance. *Journal of Chromatography A.* 2010;1217(24):3819-3843.

171.   Andrews GL, Dean RA, Hawkridge AM, Muddiman DC. Improving Proteome Coverage on a LTQ-Orbitrap Using Design of Experiments. *Journal of the American Society for Mass Spectrometry.* 2011:1-11.

172. Zhang Y, Wen Z, Washburn M, Florens L. Effect of dynamic exclusion duration on spectral count based quantitative proteomics. *Anal. Chem.* 2009;81(15):6317-6326.

173. Domon B, Aebersold R. Mass Spectrometry and Protein Analysis. *Science.* April 14, 2006 2006;312(5771):212-217.

174. Anderson NL, Anderson NG. The human plasma proteome. *Molecular & Cellular Proteomics.* 2002;1(11):845-867.

175. Righetti PG, Castagna A, Herbert B, Reymond F, Rossier JS. Prefractionation techniques in proteome analysis. *Proteomics.* 2003;3(8):1397-1407.

176. Issaq HJ, Conrads TP, Janini GM, Veenstra TD. Methods for fractionation, separation and profiling of proteins and peptides. *Electrophoresis.* 2002;23(17):3048-3061.

177. McCalley DV. Effect of buffer on peak shape of peptides in reversed-phase high performance liquid chromatography. *Journal of Chromatography A.* 2004;1038(1-2):77-84.

178. McCalley DV. Comparison of peak shapes obtained with volatile (mass spectrometry-compatible) buffers and conventional buffers in reversed-phase high-performance liquid chromatography of bases on particulate and monolithic columns. *Journal of Chromatography A.* 2003;987(1):17-28.

179. Huber CG, Premstaller A, Kleindienst G. Evaluation of volatile eluents and electrolytes for high-performance liquid chromatography–electrospray ionization mass spectrometry and capillary electrophoresis–electrospray ionization mass spectrometry of proteins: II. Capillary electrophoresis. *Journal of Chromatography A.* 1999;849(1):175-189.

180. Buckenmaier SMC, McCalley DV, Euerby MR. Overloading study of bases using polymeric RP-HPLC columns as an aid to rationalization of overloading on silica-ODS phases. *Analytical chemistry.* 2002;74(18):4672-4681.

181. McCalley DV. The challenges of the analysis of basic compounds by high performance liquid chromatography: Some possible approaches for improved separations. *Journal of Chromatography A.* 2010;1217(6):858-880.

182. Cox G, Snyder L. Preparative high-performance liquid chromatography under isocratic conditions:: III. The consequences of two adjacent bands having unequal column capacities. *Journal of Chromatography A.* 1989;483:95-110.

183. Eble J, Grob R, Antle P, Snyder L. Simplified description of high-performance liquid chromatographic separation under overload conditions, based on the Craig distribution model:: II. Effect of isotherm type, and experimental verification of computer simulations for a single band. *Journal of Chromatography A.* 1987;384:45-79.

184. Snyder L, Cox G, Antle P. A simplified description of HPLC separation under overload conditions. A synthesis and extension of two recent approaches. *Chromatographia.* 1987;24(1):82-96.

185. Johnson D, Boyes B, Fields T, Kopkin R, Orlando R. Optimization of Data-Dependent Acquisition Parameters for Coupling High-Speed Separations with LC-MS/MS for Protein Identifications "Data-Dependent Acquisition Optimization". *Journal of Biomolecular Techniques: JBT.* 2013.

186. Barnea E, Sorkin R, Ziv T, Beer I, Admon A. Evaluation of prefractionation methods as a preparatory step for multidimensional based chromatography of serum proteins. *Proteomics.* 2005;5(13):3367-3375.

187. de Godoy LM, Olsen JV, Cox J, et al. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature.* 2008;455(7217):1251-1254.

188. Hubner NC, Ren S, Mann M. Peptide separation with immobilized pI strips is an attractive alternative to in-gel protein digestion for proteome analysis. *Proteomics.* 2008;8(23-24):4862-4872.

189. Seyfried NT, Xu P, Duong DM, Cheng D, Hanfelt J, Peng J. Systematic approach for validating the ubiquitinated proteome. *Analytical chemistry.* 2008;80(11):4161-4169.

190. Hsieh EJ, Bereman MS, Durand S, Valaskovic GA, MacCoss MJ. Effects of Column and Gradient Lengths on Peak Capacity and Peptide Identification in Nanoflow LC-MS/MS of Complex Proteomic Samples. *Journal of the American Society for Mass Spectrometry.* 2013;24(1):148-153.

191. Köcher T, Swart R, Mechtler K. Ultra-high-pressure RPLC hyphenated to an LTQ-Orbitrap Velos reveals a linear relation between peak capacity and number of identified peptides. *Analytical chemistry.* 2011;83(7):2699-2704.

192. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA: a cancer journal for clinicians.* 2011;61(2):69-90.

193. van Gils MP, Stenman U, Schalken J, et al. Innovations in serum and urine markers in prostate cancer: current European research in the P-Mark project. *European urology.* 2005;48(6):1031-1041.

194. Djulbegovic M, Beyth RJ, Neuberger MM, et al. Screening for prostate cancer: systematic review and meta-analysis of randomised controlled trials. *BMJ.* 2010;341(10):c4543.

195. Bastian PJ, Carter BH, Bjartell A, et al. Insignificant prostate cancer and active surveillance: from definition to clinical implications. *European urology.* 2009;55(6):1321-1332.

196. Moyer VA. Screening for prostate cancer: US Preventive Services Task Force recommendation statement. *Annals of internal medicine.* 2012.

197. Bullock A, Andriole G. Screening for prostate cancer: prostate-specific antigen, digital rectal examination, and free density, and age-specific derivatives. W: Kantoff PW, Carroll PR, D'Amico AV (red.). Prostate cancer principles & practice. *KantoffP, CarrollP, D'AmicoA, eds. Prostate Cancer Principles and Practice. Philadelphia: Lippincott Williams and Wilkins.* 2002:195-211.

198. Gokce E, Shuford CM, Franck WL, Dean RA, Muddiman DC. Evaluation of normalization methods on GeLC-MS/MS label-free spectral counting data to correct for variation during proteomic workflows. *Journal of the American Society for Mass Spectrometry.* 2011;22(12):2199-2208.

199. Rehman I, Azzouzi A, Catto J, et al. Proteomic analysis of voided urine after prostatic massage from patients with prostate cancer: a pilot study. *Urology.* 2004;64(6):1238-1243.

200. Jansen FH, Krijgsveld J, van Rijswijk A, et al. Exosomal secretion of cytoplasmic prostate cancer xenograft-derived proteins. *Molecular & Cellular Proteomics.* 2009;8(6):1192-1205.

201. RONQUIST KG, Ronquist G, Larsson A, Carlsson L. Proteomic analysis of prostate cancer metastasis-derived prostasomes. *Anticancer research.* 2010;30(2):285-290.

202. ARENDS J. Metabolism in cancer patients. *Anticancer research.* 2010;30(5):1863-1868.

203. Langbein S, Zerilli M, Zur Hausen A, et al. Expression of transketolase TKTL1 predicts colon and urothelial cancer patient survival: Warburg effect reinterpreted. *British journal of cancer.* 2006;94(4):578-585.

204. Glen A, Gan CS, Hamdy FC, et al. iTRAQ-facilitated proteomic analysis of human prostate cancer cells identifies proteins associated with progression. *Journal of proteome research.* 2008;7(3):897-907.

205. Colombini M. VDAC: the channel at the interface between mitochondria and the cytosol. *Molecular and cellular biochemistry.* 2004;256(1-2):107-115.

206. Lemasters JJ, Holmuhamedov E. Voltage-dependent anion channel (VDAC) as mitochondrial governator—thinking outside the box. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease.* 2006;1762(2):181-190.

207. Burgess EF, Ham AJL, Tabb DL, et al. Prostate cancer serum biomarker discovery through proteomic analysis of alpha-2 macroglobulin protein complexes. *PROTEOMICS-Clinical Applications.* 2008;2(9):1223-1233.

208. Stebbins CE, Russo AA, Schneider C, Rosen N, Hartl FU, Pavletich NP. Crystal structure of an Hsp90-geldanamycin complex: targeting of a protein chaperone by an antitumor agent. *Cell.* 1997;89(2):239-250.

209. Whitesell L, Lindquist SL. HSP90 and the chaperoning of cancer. *Nature Reviews Cancer.* 2005;5(10):761-772.

210. Mussunoor S, Murray G. The role of annexins in tumour development and progression. *The Journal of pathology.* 2008;216(2):131-140.

211. Wozny W, Schroer K, Schwall GP, et al. Differential radioactive quantification of protein abundance ratios between benign and malignant prostate tissues: cancer association of annexin A3. *Proteomics.* 2007;7(2):313-322.

212. Shishkin S, Kovalyov L, Kovalyova M, et al. " Prostate Cancer Proteomics" Database. *Acta naturae.* 2010;2(4):95.

213. Liu J-W, Shen J-J, Tanzillo-Swarts A, et al. Annexin II expression is reduced or lost in prostate cancer cells and its re-expression inhibits prostate cancer cell migration. *Oncogene.* 2003;22(10):1475-1485.

214. Paweletz CP, Ornstein DK, Roth MJ, et al. Loss of annexin 1 correlates with early onset of tumorigenesis in esophageal and prostate carcinoma. *Cancer research.* 2000;60(22):6293-6297.

215. Hsiang CH, Tunoda T, Whang YE, Tyson DR, Ornstein DK. The impact of altered annexin I protein levels on apoptosis and signal transduction pathways in prostate cancer cells. *The Prostate.* 2006;66(13):1413-1424.

216. Chetcuti A, Margan SH, Russell P, et al. Loss of annexin II heavy and light chains in prostate cancer and its precursors. *Cancer research.* 2001;61(17):6331-6334.

217. Nagano K, Masters JR, Akpan A, et al. Differential protein synthesis and expression levels in normal and neoplastic human prostate cells and their regulation by type I and II interferons. *Oncogene.* 2003;23(9):1693-1703.

218. Chao MP, Jaiswal S, Weissman-Tsukamoto R, et al. Calreticulin is the dominant pro-phagocytic signal on multiple human cancers and is counterbalanced by CD47. *Science translational medicine.* 2010;2(63):63ra94-63ra94.

219. Shen R, Su Z-Z, Olsson CA, Fisher PB. Identification of the human prostatic carcinoma oncogene PTI-1 by rapid expression cloning and differential RNA display. *Proceedings of the National Academy of Sciences.* 1995;92(15):6778-6782.

220. Lapek JD, McGrath JL, Ricke WA, Friedman AE. LC/LC–MS/MS of an innovative prostate human epithelial cancer (PHEC)< i> in vitro</i> model system. *Journal of Chromatography B.* 2012.

221.   Rehman I, Evans CA, Glen A, et al. iTRAQ Identification of Candidate Serum Biomarkers Associated with Metastatic Progression of Human Prostate Cancer. *PloS one.* 2012;7(2):e30885.

222.   Pecorari L, Marin O, Silvestri C, et al. Elongation Factor 1 alpha interacts with phospho-Akt in breast cancer cells and regulates their proliferation, survival and motility. *Mol Cancer.* 2009;8(8):58.

223.   Zhu G, Yan W, He H-c, et al. Inhibition of proliferation, invasion, and migration of prostate cancer Cells by downregulating elongation factor-1α expression. *Molecular Medicine.* 2009;15(11-12):363.

224.   Pontes-Júnior J, Reis ST, de Oliveira LCN, et al. Association between integrin expression and prognosis in localized prostate cancer. *The Prostate.* 2010;70(11):1189-1195.

225.   McCabe N, De S, Vasanji A, Brainard J, Byzova T. Prostate cancer specific integrin αvβ3 modulates bone metastatic growth and tissue remodeling. *Oncogene.* 2007;26(42):6238-6243.

226.   Goel HL, Li J, Kogan S, Languino LR. Integrins in prostate cancer progression. *Endocrine-related cancer.* 2008;15(3):657-664.

227.   Johansson B, Pourian MR, Chuan YC, et al. Proteomic comparison of prostate cancer cell lines LNCaP-FGC and LNCaP-r reveals heatshock protein 60 as a marker for prostate malignancy. *The Prostate.* 2006;66(12):1235-1244.

228.   Chen L, Fang B, Giorgianni F, Gingrich JR, Beranova-Giorgianni S. Investigation of phosphoprotein signatures of archived prostate cancer tissue specimens via proteomic analysis. *Electrophoresis.* 2011;32(15):1984-1991.