ABSTRACT

Title of Dissertation: ANALYSIS OF INTACT PROTEINS IN COMPLEX MIXTURES

Avantika Dhabaria
Doctor of Philosophy
2013

Directed By: Catherine Fenselau, Professor,
Department of
Chemistry and Biochemistry

Our goal is to develop an effective work flow for analysis of intact proteins in a complex mixture using the LC-LTQ-Orbitrap XL. Intact protein analysis makes the entire sequence available for characterization, which allows for the identification of isoforms and post translational modifications. We focus on developing a method for top-down proteomics using a high-resolution, high mass accuracy analyzer coupled with bioinformatics tools. The complex mixtures are fractionated using 1-dimensional reversed-phase chromatography and basic reversed- phase, and open tubular electrophoresis. The analysis of intact proteins requires various fragmentation methods such as collisional induced dissociation, high energy collisional dissociation, and electron transfer dissociation. This overall method enables us to analyze intact proteins, providing a better understanding of protein expression levels and post transitional modification information. We have used standard proteins to optimize HPLC conditions and to compare three methods for

ion activation and dissociation. Furthermore, we have extended the method to analyze low mass proteins in MCF7 cytosol and in *E. coli* lysate as a model complex mixture. We have applied this strategy to identify and characterize proteins from extracellular vesicles (EVs) shed by murine myeloid-derived suppressor cells (MDSC). MDSCs suppress both innate and adaptive immune responses to tumor growth and prevent effective immunotherapy. Recently some of the intercellular immunomodulatory effects of MDSC have been shown to be propagated by EVs. Top-down analysis of intact proteins from these EVs was undertaken to identify low mass protein cargo, and to characterize post-translational modifications.

Analysis of Intact Protein in Complex Mixtures

By

Avantika Dhabaria

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2013

Advisory Committee:
Professor Catherine Fenselau, Chair
Professor Alan Kaufman, Dean's Representative
Professor Philip DeShong
Professor Neil Blough
Professor Nathan Edwards

Dedication

This work is a dedication to my grandfather, Mathurdas Dhabaria, who has always been my biggest supporter. He has inspired me to never let anything stand in the way of achieving my accomplishments

Acknowledgements

I would like to thank Dr. Catherine Fenselau for giving me the opportunity to work in her lab. Her guidance and training has shaped me in being the scientist that I am today. She has always been available to answer any questions and provide valuable advice. I would also like to thank Dr. Yan Wang, the director of the proteomics core facility, with her chromatography and instrumentation suggestions. I would like to acknowledge Dr. Nathan Edwards for his advice and help with any bioinformatics for the work in this thesis and Dr. Peter Gutierezz for his advice with cell culture and cell biology.

The members of the Fenselau lab, Meghan Burke, Rebecca Rose, Sara Moran, Maria Oei, Dr. Colin Wynne, Dr. Karen Lohnes, and who have provided a warm and supportive environment. I would especially like to thank Dr. Joe Cannon and Dr. Waeowalee Choksawankam for their support and friendship through the last four years.

Finally, I would like to thank my parents, Pramod and Amita Dhabaria, my uncle and aunt, Jatin and Meena Shah and my sister Priyanka Dhabaria for their advice, encouragement and love through graduate school. My friends, who have supported me and given me strength to move forward by being family away from home.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

1D	One dimensional
2D LC	Two-dimensional liquid chromatography
2DE	Two dimensional gel electrophoresis
AC	Alternating current
bRP-aRP-LC	Basic reversed-acidic reversed phase high performance liquid chromatography
ACN	Acetonitrile
CID	Collisionally induced dissociation
ESI	Electrospray ionization
ETD	Electron transfer dissociation
E-value	Expectation value
EVs	Extracellular vesicles
FTICR	Fourier transform ion cyclotron resonance
FTMS	Fourier transform mass spectrometer
GELFrEE	Gel-Eluted Liquid Fraction Entrapment Electrophoresis
HCD	High energy collision dissociation
HILIC	Hydrophilic interaction chromatography
HPLC	High pressure liquid chromatography
LC-MS/MS	Liquid chromatography tandem mass spectrometry
LTQ	Linear ion traps
m/z	Mass to charge ratio
MALDI	Matrix-assisted laser desorption ionization
MDSC	Myeloid-derived suppressor cells
MS	Mass spectrometry
MS/MS	Fragment ion spectrum
MS2	Fragment ion spectrum
Nano-ES	Nano electrospray
ppm	Parts per million
PTMs	Post-translational modifications
RP	Reverse phase
RPLC	Reverse phase liquid chromatography
SAX	Strong anion exchange
SCX	Strong cation-exchange
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
TOF	Time-of-flight

List of Appendices

Chapter 1: Introduction

Protein Mass Spectrometry

The word proteomics can be defined as the entire protein component expressed by a genome, or by a cell or tissue type, under a given condition. The main subfields of proteomics analysis are characterization, which provides a survey of proteins present in a cell tissue or biofluid; differential proteomics, which provides identification of differentially expressed proteins in different physiological states; and functional proteomics, which provides identification of a group of proteins involved in specific functions.[1] Proteomics focuses on the dynamic description of gene regulation. Thus it offers more information than a protein equivalent of DNA databases.[2] In the early history of proteomics, proteins were fractionated using two-dimensional polyacrylamide gel electrophoresis (2DE) followed by visualization using protein stains, such as Coomassie or silver stain, subsequently the protein spots were identified using mass spectrometry.[3] Mass spectrometry (MS) has become a ubiquitously useful tool for proteomic research. It provides comprehensive knowledge about systems biology, including proteomic profiling, protein quantity, and post-translational modifications (PTMs) of cellular and organellar proteomes.[1] Many different aspects of MS have led to its prominent position within the field of proteomics. The sensitivity of MS allows for routine identification of proteins in femtomole to high attomole range.[4] The ability to

identify proteins with confidence is aided by mass measurement accuracy available using current MS technology. The accuracy is typically less than 50 parts per million (ppm) and is often less than 5 ppm.[5] The ability of tandem MS to obtain partial sequence information enables confident bioinformatic identification of proteins and peptides in complex mixtures. In recent years, protein and peptide fractionation methods coupled with various mass spectrometry technologies have evolved as the dominant tools in the field for protein identification.[6] Mass spectrometry is a technique that measures the masses and relative abundances of atoms and molecules. In order to accomplish this, each mass spectrometer is composed of an ion source, an analyzer, and a detector. The ionizer generates gas phase ions from the sample.[1] The analyzer separates those ions by mass/charge and allows for fragmentation of the precursor ion to create other, smaller ions. The detector records the ions and provides the signal to be interpreted by the instrument software. The typical mass spectrum has two important pieces of information, the mass-to charge ratio (m/z) and the relative abundance or intensity. The most intense signal is generally set at 100% and the other signals have their height plotted in proportion to this "base" peak.[1] A common platform for analysis of complex protein mixtures consists of high pressure liquid chromatography (HPLC), electrospray ionization, and high resolution mass spectrometry. Complex mixtures are first fractionated by reverse phase HPLC, which is interfaced to an ion source of the mass spectrometer, where they are first ionized to acquire positive or negative charges. The ions travels through the mass analyzer and arrive at the

detector based on m/z ratio and the ions makes contact with the detector the signals are then generated and recorded.[7]

Proteomics Methods

The three widely used methods in proteomic analysis are bottom-up proteomics, top-down proteomics, and middle-out proteomics.

Bottom-Up Proteomics

The most common approach used in proteomic analysis is the bottom-up method. Proteins are proteolytically digested into peptides prior to mass spectrometric analysis, typically using trypsin, which is an enzyme that hydrolyzes peptide bonds C-terminal of arginine and lysine residues, except when followed by proline. The peptides are fractionated using LC methods and fragmented using collisionally induced dissociation as a fragmentation technique. The fragment ion spectra are compared with the predicted spectra that are in silico–generated fragmentation patterns of the peptides from proteins in a user-defined database. The drawback of utilizing bottom up proteomic methods is that the proteins maybe identified on the basis of a limited number of peptides, low abundant peptides can be lost during chromatography and due to the complexity of the sample, and thus some peptides may not be analyzed in data dependent experiments. In addition, a single amino acid substitutions or an unexpected post translational modification can prevent identification of the peptide. Furthermore, even when a protein is confidently identified in bottom-up experiments, information on this protein

sequence is only partially obtained, thus there is gain of information about proteoforms and PTMs of the proteins identified.[8]

Top-Down Proteomics

The top-down approach in proteomics ionizes the whole protein in the mass spectrometer, fragments the protein, then matches the fragments against the database of fragment masses generated *in-silico* from a data base of intact protein sequences. A major advantage of top down methods is that performing an MS/MS experiment on an intact protein ion, in principle, makes the entire sequence available for complete characterization and localize any post-translational modifications on the protein.[9] In addition, the method allows the identification of proteoforms, mutations and splice variants.[10] For example, the two proteins thymosin beta-4 and thymosin beta-10 have very similar amino acid sequences, as shown in Figure 1. If these two proteins are present in a complex mixture that is digested using trypsin and peptide **ETIEQEK** is identified in the bottom up analysis, then it would be very difficult to confidently say if the peptide was identified from thymosin beta-4, thymosin beta-10 or both. Thus, using the top-down methodology and recording the intact mass of the protein, it is possible to differentiate and identify two proteoforms present in the complex mixture.

(A) MADKPDMGEIASFDKAKLKKTETQEKNTLPTKETIEQEKRSEIS

(B) MSDKPDMAEIEKFDKSKLKKTETQEKNPLPSKETIEQEKQAGES

Figure 1: Protein sequences of (A) thymosin beta-10, (B) thymosin beta-4. The amino acid residues highlighted in blue are unique to each protein

High resolution mass spectrometers are needed to resolve monoisotopic peak of the highly charged protein molecules. The instruments that provide high resolving power for intact proteins and their fragment ions are Fourier transform based instruments such as the Fourier transform ion cyclotron resonance (FTICR), and the hybrid LTQ Orbitrap, and high end time-of-flight (TOF) analyzers. These instruments usually employ ESI as an ionization technique for analysis of intact proteins with multiply charged ions for the same protein.[11] The molecular weight of the intact protein is calculated on the basis of its multiply charged spectrum for ESI. During bioinformatics analysis the mass difference between the calculated and the measured mass of the protein is sufficient to indicate the presence of PTMs. Furthermore, fragment ion data generated by the top-down approach is unambiguously assigned to a particular protein whose precursor ion was selected for fragmentation, thus providing information about the complete sequence of the protein itself.[12] The top-down method can integrate and support information obtained from bottom-up analysis, in particular regarding complete protein sequence and PTM localization, as well as any combination possibly existing between modifications on distinct parts of the protein sequence.

Middle-Out Proteomics

The middle-out approach in proteomics is a compromise between bottom-up and top-down analysis. While heavy proteins have been successfully identified

using the top-down strategy, it is difficult to perform the analysis on a LC time scale, especially for low abundant proteins. Middle-out analysis uses alternative enzymes for the cleavage of proteins to generate longer peptides than conventionally obtained peptides in bottom-up proteomics; usually peptides larger than 20 amino acid residues (5-10 KDa). These large peptides can sometimes provide information on PTMs. The larger peptides are generated using reagents such a CNBr or enzymes such as Lys-C, Asp-N, Glu-C. Additionally, microwave assisted acid cleavage on the aspartic acid residue also generates peptides with an extended mass range. The advantage of middle-out strategy, as opposed to bottom-up proteomics, is that it provides higher confidence of identification and sequence coverage.[13]

Fractionation Methods for Top-Down Proteomics

Due to the complexity of biological samples, it is necessary to fractionate the complex mixtures prior to their measurement with mass spectrometry, so that the complexity of what is introduced to the mass spectrometer is compatible with the performance of the instrument. Among various separation modes available, electrophoresis and chromatography are most widely used at this time. For the top-down approach, 2-DE[3] is one of the most commonly used approaches for proteome analysis, because of its unparalleled resolving power. However, there are many downsides to using this approach. It is labor-intensive, and has low sensitivity, poor quantitative accuracy, and limited dynamic range. The limiting factor for 2-DE

methodology is its inability to investigate very hydrophobic/hydrophilic[14, 15] and low copy number proteins[16,17] in whole cell lysates.[18]

In recent years, HPLC has become a popular liquid phase based separation technique. HPLC coupled with mass spectrometry has become an indispensable tool for proteomics due to its compatibility with ESI. Chromatographic separation of a protein mixture is based on the interaction of proteins with the stationary phase, and equilibration between adsorption on the stationary phase and in the mobile phase. High resolution separation prior to mass spectrometer analysis minimizes ion suppression and under-sampling associated with the analysis of highly complex proteomes. In order to improve the separation capacity, detection sensitivity, and analysis throughput of micro- and nano-HPLC, recent advances have been made by increasing column length, reducing inner diameter, using sub-micrometer sized packing materials and monolithic columns.[19] Successful detection by the mass spectrometer is closely related to the flow rate of the HPLC. Lower flow rates in the nanoliter ranges result in smaller eluent droplets, more charges per analyte molecule, and higher ESI efficiency. Most protein separations are performed with columns packed with packing material made of silica and operated with an reversed phase HPLC.[20] The chemical and physical properties of a protein, such as hydrophobicity, length, net charge, and solubility influence the retention time and separation. From protein complexes to whole cell lysate, proteome analysis deals with highly complex mixtures, requiring more than one analytical dimension to achieve the high resolving power necessary for reliable analysis.[21] Usually, no separation method is capable of completely resolving complex

mixtures in a single analytical dimension. Consequently, multiple proteins, or peptides, enter the mass spectrometer at any given time, leading to fewer numbers of identifications. Multidimensional methods, having orthogonal separation modes, are expected to overcome the problem of insufficient resolution in the analysis of a complex mixture. The first step in a multidimensional separation is fractionation into simpler mixtures. These fractions are then chromatographically separated by reverse phase HPLC that allows maximum detection by the mass spectrometer. The first dimension fractionation methods include, strong cation-exchange (SCX), strong anion exchange (SAX), reverse phase (RP), size exclusion chromatography (SEC), or hydrophilic interaction chromatography (HILIC).[22] Along with HPLC, sample fractionation techniques, and bioinformatics, mass spectrometry provides unbiased analyses of components in complex mixtures within a short time frame.

In addition to HPLC and 2DE as a fractionation method for proteins, sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), is a popular method to fractionate proteins based on their molecular weight. The biggest limitation of SDS- PAGE as a fractionation technique is the recovery of protein from the gel. An alternative, to extracting proteins from a gel, is Gel-Eluted Liquid Fraction Entrapment Electrophoresis (GELFrEE). This is molecular weight-based separation involves continuous elution SDS-PAGE in a tube format, in which proteins are constantly eluted from the gel column and collected in the solution phase (i.e., free of the gel), providing broad mass range fractionation with good resolution, reproducibility, and recovery.[23] Appropriate combinations of available tools are promising for future proteomic research.

Electrospray Ionization

The mass spectrometer is made up of three major components: the ionization source, the mass analyzer, and the detector. The sample of interest is ionized and then desorbed into the gas phase within the ionization source.[24] The two most common methods to ionize biological molecules prior to the entrance into the analyzer region of the mass spectrometer are matrix-assisted laser desorption ionization (MALDI) and electrospray ionization (ESI). In 1989, John Fenn introduced a soft ionization technique, ESI, to ionize intact chemical species (proteins) by multiple charging.[25, 26] The ionization is soft because very little residual energy is retained by the analyte, and generally no fragmentation occurs upon ionization. In addition, weak noncovalent interactions can be preserved in the gas phase.[25, 27] Because of the multiple charging, the m/z values of the resulting ions become lower and fall in the mass ranges of all common mass analyzers. Thus ESI became very useful in the production of gas-phase ions from large biologically important macromolecules like proteins and nucleic acids, and their subsequent mass spectrometric analysis of structural characterization as well as their rapid identification on the basis of molecular mass, a very specific property of the analyte.[25, 28]

The molecular mass of macromolecules can be calculated using the following formula:

$$M=n(m'- H) \quad \text{(Equation1)}$$

In the relationship above, M is the neutral mass of the macromolecule, m' measured mass to charge ratio of the molecule, n is the number of charges and H is the mass of a proton.[1] The number of charges on the protein molecule will depend on the molecular weight of the protein and the number of accessible basic sites (e.g. arginine, histidine and lysine). Proteins exhibit different charge state distribution profiles in their ESI-mass spectra.[25]

The mechanism by which ESI works is not completely understood. ESI requires the sample of interest to be in solution so that it may flow into the ionization source region of the mass spectrometer.[24] The solution must be a conducting solution. The ionization occurs in three different processes: droplet formation, droplet shrinkage, and desorption of gaseous ions. The sample is ionized by applying high voltage through the stainless steel needle through which the sample flows.[24] At the onset of the electrospray process, the electrostatic force on the liquid leads to the partial separation of charges. In positive-ion mode, cations concentrate at the tip of the metal capillary and tend to migrate towards the counter electrode. The migration of the accumulated positive ions towards the counter electrode is counterbalanced by surface tension of the liquid, giving rise to a Taylor cone at the tip of the capillary.[1] As the sample exits the spray tip, the solution produces submicrometer-sized droplets containing both the conducting solute and analyte ions. The droplets are subsequently desolvated into gaseous ions and analysis of highly charged molecular ions can be analyzed by the mass spectrometer.[29]

Desorption is achieved by evaporation of the solvent by passing the solution through a heated capillary or a curtain of drying gas, typically nitrogen.[24]

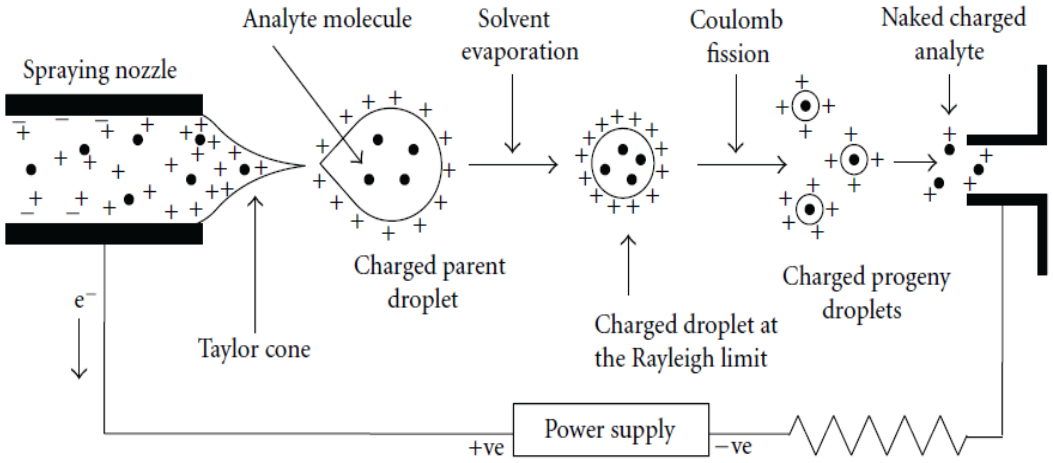

Figure 2: Schematic representation of electrospray ionization.[25]

Nano electrospray (nano-ES) is a miniaturized version of standard ESI and is desigened to operate at submicroliter flow rates.[30] There are several practical differences in the operation of the two modes. Nano-ES utlizes a smaller spraying tip aperture, the stable spray is obtained at lower voltages, and the droplets produced are about 200 nm in diameter. The rate of desorption of ions from small droplets and the mass molar sensitivity of ESI is inversely proportional to the flowrate. The nano-ES spray disperses the liquid purely by electrostatic means, and no assistance via the sheath flow gas is used. Thus, it is a very stable source, which can spray a variety of buffers in both positive and negative mode. The stability of the spray helps in measuring protein masses accurately.[30] Nano-ES provides several orders of additional sensitivity in comparison to ESI.

The LTQ-Orbitrap

A mass analyzer separates the ionic species based on their mass-to-charge ratios. The different analyzers have benefits and deficiencies in protein mass spectrometry. Four types of mass analyzers used widely in the field are linear ion traps (LTQ), time-of-flight (TOF), Orbitrap, and the Fourier-transform ion cyclotron resonance (FTICR). The two Fourier transform based analyzers used for top-down proteomics are the Orbitrap and the FTICR. Historically top-down protein analysis has been successful with FTICR as a mass analyzer. This requires a super conducting magnet, which is expensive to acquire and to maintain. The LTQ-hybrid Orbitrap, also a Fourier transform based instrument, is a good alternative to the FTICR. The two mass analyzers predominantly used for the work in this thesis are the linear ion trap and the Orbitrap. The working principles of these mass analyzers are discussed in the sections following.

Linear Ion Traps

Wolfgang Paul introduced the ion trap often referred as the Paul trap in 1969[31], and his contribution to mass spectrometry was recognized by the award of the 1989 Nobel Prize.[32] LTQ, also known as two-dimensional quadrupole ion traps are rapidly finding new applications in many areas of mass spectrometry. Instruments such as TOF, FT-ICR, and Orbitrap have been coupled with the LTQ. The advantage of the new generation hybrid instruments is their ability to combine

the MS^n feature of an LTQ, with the mass accuracy and resolution of the FT or TOF based mass analyzers.[33] Ion traps function based on oscillating electric and radio frequency (rf) potentials applied in three dimensions to maintain ion populations in regions of high stability.[34]

Helium is used as a dampening gas inside the ion trap, due to its ability to energetically cool the ions without fragmenting them. When the ions enter the mass spectrometer they collide with the helium gas. This helps to slow the ions so that they can be trapped in the mass analyzer by the rf field. The true power of the ion-trap analyzer is its ability to isolate and fragment peptide ions (MS/MS) from complex mixtures, such as those found in many proteomic analyses. To perform MS/MS analysis, specific ions are selected and the trapping voltages are adjusted to eject all other ions from the trap. The trapped ions undergo collision with helium gas, causing them to fragment. These fragments are then trapped and are scanned according to their m/z values.[34]

Orbitrap

The Orbitrap mass analyzer was invented by Alexender Makarov and is an alternative to traditional superconducting magnet based FTMS systems for high resolution analysis. It is an ion trap with an oblong shape consisting of an outer barrel like electrode and a central spindle like electrode along the axis.[35] The ions are injected orthogonally to the central electrode and are attracted to an increasing voltage on this central electrode. The outer electrodes oscillate polarity, causing the ions to orbit around the central electrode; this is shown in Figure 3.[36]

.

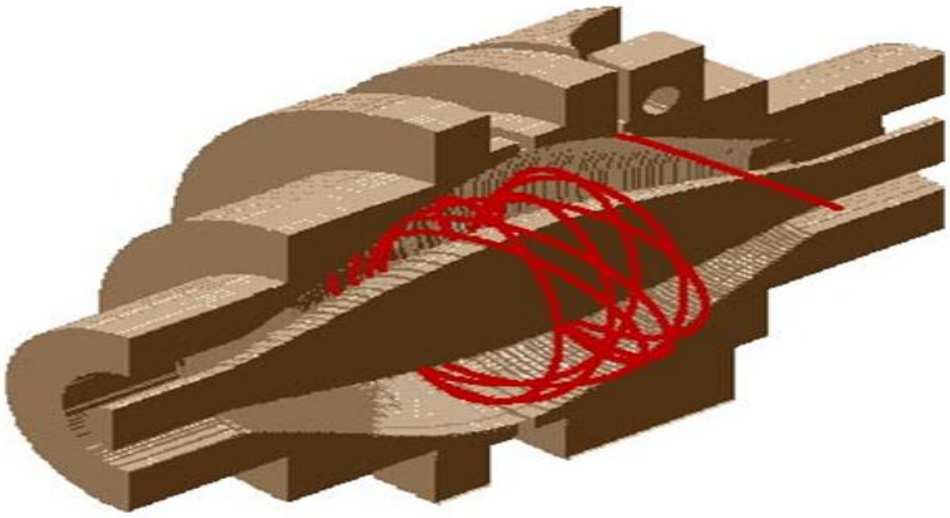

Figure 3: Image of the Orbitrap mass analyzer, with ions orbiting around the central electrode.[36]

All ions in the Orbitrap have exactly the same amplitude, although ion packets of different m/z have a given oscillation at frequencies, which can be calculated using:

$$\omega = \sqrt{(\frac{z}{m}).k}$$

Where ω the frequency of ion oscillation, m/z is is the mass to charge ratio and k is field curvature

The outer electrode is split in half, allowing the ion image current due to the axial motion to be collected. The current is amplified from each half of the outer electrode and undergoes analog-to-digital conversion before processing, and image current is recorded and converted to a frequency spectrum using a Fourier transform.[37]

The resolving power, in Orbitrap MS is given by:

$$\frac{m}{\Delta m} = \frac{1}{2\Delta\omega}\left(\frac{kq}{m}\right)^{1/2}$$

For a given acquisition time the resolving power of the Orbitrap mass analyzer diminishes as the square root of *m*/*z*. In addition, as the cross section of the background gas increases, the resolving power decreases with increasing mass. Collisions with the background gas can lead to fragmentation of ions, loss of ions, or ejection of ions from the trap. The Orbitrap mass analyzer produces spectra with mass accuracies in the range of 2–5 ppm. The main reason for a mass error in the Orbitrap for heavier proteins is the low signal to noise (S/N) of the precursor and product ions. Thus, as the S/N ratio increases, the accuracy of mass measurement improves.[38] The hybrid LTQ-Orbitrap has two mass analyzers. The Orbitrap achieves higher resolution and mass accuracy which is important for the analysis of intact proteins. Whereas, the ion trap allows the isolation and fragmentation of ions of interest.

Fragmentation of Intact Proteins

Fragmentation plays an important role in identification of proteoforms and post-translational modification of proteins in complex mixtures. It is important to optimize fragmentation conditions, as well as choose the suitable fragmentation method for a given analysis. The full MS scan detects the m/z values of precursor ions that are ionized into the mass spectrometer; while the MS2 scans are derived

from the dissociation of multiply charged protein ions and typically comprise ions with charges ranging from unity up to the charge of the precursor ion.[39] High mass accuracy of both the precursor and product ion fragmentation patterns is very useful for protein and peptide identification using bioinformatic analysis.[40]

Collisionally Induced Dissociation

A common technique used to activate peptides and proteins is collisionally induced dissociation (CID). Peptide or protein molecular ions in the gas phase are collide with an inert gas, e.g. helium and nitrogen, in the collision cell.[41] Collision with neutral gas atoms, leads to conversion of the kinetic energy of the ions into internal vibrational energy. As the vibrational energy exceeds a certain threshold, covalent peptide bonds break. The energy is randomly distributed between bonds and the weakest peptides bonds break.[42] Typically, the preferred sites of cleavage in gas phase peptide ions are the amide bonds of the peptide bond. In this type of fragmentation, the amide bond of the peptide backbone will fragment to produce a series of b and y-type ions as shown in Figure 4. Due to the nature of CID, the fragmentation of large peptides is likely to be incomplete, and information on labile chemical modification on the protein is normally undetectable.[43]

Electron Transfer Dissociation

Electron transfer dissociation (ETD) is a new method to fragment peptide and proteins that complements the data obtained from CID. It is suitable for fragmentation of larger molecular ions and is less destructive of modifications on peptides or proteins.[44] ETD fragments peptides by transferring an electron from a

radical anion to a protonated peptide bond. This transfer induces fragmentation of the peptide backbone, causing cleavage of the Cα-N bond generating series of c- and z- ions, and PTM linkages are not broken.[45] The fragmentation patterns are shown in Figure 4.

High Energy Collision Dissociation

High energy collision dissociation (HCD) can also be performed on the LTQ-Orbitrap. HCD employs higher energy dissociations than those used in ion trap CID, enabling a wider range of fragmentation pathways.[46] The fragment pattern for HCD is due to the higher energy and the shorter activation time when compared to CID. Both b-ions and y-ions are observed in CID, whereas higher energy levels with HCD lead to more y-ions than b-ions. The b-ions may fragment further to a-ions or smaller pieces.[42] Compared with traditional ion trap-based collision-induced dissociation, HCD fragmentation with the Orbitrap provides increased ion fragments and results in higher quality MS/MS spectra of proteins.

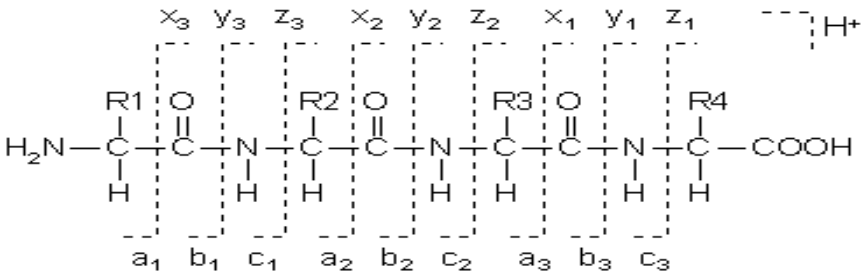

Figure 4: Fragmentation pattern of protein backbone showing series of a- and x- ions, b- and y- ions, and c- and z- ions (www.matrixscience.com).

Bioinformatics

One of the major bottlenecks of top-down proteomics is the limited number of bioinformatics tools for top-down data management and interpretation. Tandem mass spectra of proteins can be difficult to interpret because of the number of ions, the charge states of these different ions and the possible interferences. ProSightPC 2.0 (Thermofisher scientific, San Jose, CA) is the search program that is widely used in the high mass accuracy top-down approach to match protein fragment ions against the database of protein sequences.[47] First, the program uses the algorithm THRASH[47] or Xtract to deconvolute the multiply charged precursor fragment ions so that each protein, and each fragment, has only one mass. Using a given mass tolerance it analyzes the data from a chromatographic time scale. The program then seeks as match to a theoretical protein based on matching the fragment masses within the assigned tolerance. ProSightPC assigns an expect value (E–value) to every matched protein, which measures how likely the match is, compared to a match to a random protein. The expectation value (E-value) is calculated based on two parameters, the number of sequences (N) in the database and the Poisson based p-score (p).[9]

$$E = N * p(n)$$

A lower E-value gives higher confidence to the match of the observed and theoretical data. Additionally, ProSightPC has a tool that allows the user to check for a previously unknown modification after the search has been completed.[48] Due

to all the features described above, ProSightPC 2.0 is widely used for the analysis in this work.

Intact protein analysis workflow

A general intact protein analysis workflow is shown in Figure 5. A complex mixture of interest is usually fractionated using chromatographic methods. Following elution the proteins are ionized, typically using ESI ionization method. The MS1 scan or the precursor ion scan of the ionized proteins are recorded using high resolution mass analyzers. The precursor ion of interested is isolated and fragmented using CID, ETD or HCD fragmentation methods. The fragment ion scan, or the MS2, of the resulting fragments are recorded using a high resolution mass analyzer. The data is searched against a customized database and specialized search programs, such as ProSightPC 2.0.[9]

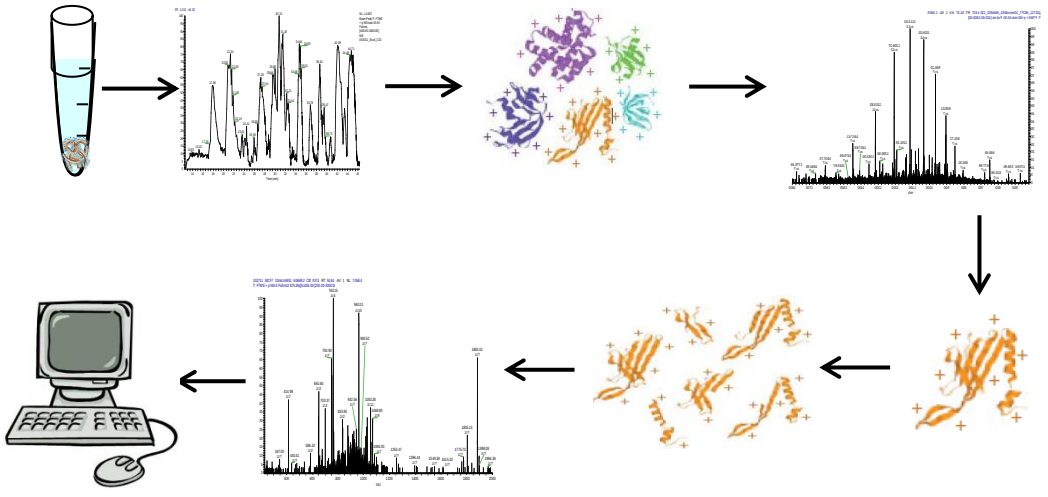

Figure 5: Workflow for analysis of intact proteins

Objective

The goal of this thesis is to develop an effective work flow for analysis of intact proteins using an LC-LTQ-Orbitrap, including front end chromatographic separation and the high resolution analysis of both precursor ions and product ions that is required to assign charge states and thus interpret the spectra. We have used standard proteins to optimize HPLC conditions and to compare three methods for ion activation and dissociation. The optimized method is extended to analyze low mass proteins in cancer cell MCF7 cytosol and *E. coli* lysate. Molecular weight cut-off filters, basic reversed- phase high performance liquid chromatograph and tubular electrophoresis are evaluated as fractionation methods for sample preparation in these model complex mixtures. Finally, this work flow is applied to identify and characterize proteins from EVs shed by murine MDSC cells. The top-down analysis of intact proteins from these EVs was utilized to identify low mass proteins and to characterize post-translational modifications.

Chapter 2: Evaluation of Chromatographic Methods for Intact proteins

Introduction

Despite the availability of high performance mass spectrometers, methods for intact protein separation, identification of intact proteins and their proteoforms are yet underdeveloped and remain a challenge for solution-based proteomics platforms. A variety of fractionation techniques including isoelectric focusing, capillary electric focusing, 2DE, GELFrEE, reverse phase liquid chromatography (RPLC), SCX, have been utilized in fractionating intact proteins prior to MS analysis. While gel based separations were widely adapted in top-down proteomics, its limitations for extraction of proteins from the gel makes in solution-based fractionation a suitable alternative. HPLC coupled with ESI-MS is an essential tool in proteomics, due to its compatibility with the electrospray ionization technique. To reduce the complexity of a sample it is important to fractionate protein mixtures prior to liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis. The evolution of both column material and packing designs contribute to strong efforts in increasing the improvement of fractionation of intact proteins, mainly in order to miniaturize the process and avoid sample loss. Progress has been made in column chromatography with both porous and non-porous packing materials.

The application of any single fractionation method provides insufficient peak capacity. The number of proteins identified with a conventional one

dimensional LC-MS/MS analysis is typically limited, and fractionating protein samples prior to analysis is important for increasing both the analytical dynamic range and proteome coverage. Two-dimensional liquid chromatography (2D LC) provides higher peak capacity and dynamic range for complex mixtures. In addition, it enhances the probability of identifying lower abundance proteins whose ions may be suppressed in complex samples. Although 2D LC significantly increases the peak capacity of the chromatographic conditions, identical proteins are often present in multiple fractions. The effectiveness of a 2D LC separation depends on the compatibility of the two separations, the separation efficiency and the separation orthogonality. Since strong cation exchanged (SCX) employs a different separation mechanism and provides good orthogonality to RPLC, it has been widely used as the first dimension for 2D LC-MS/MS. SCX as choice of fractionation methods in the first dimension has its limitations, including reduced sample recovery and sample losses due to sample desalting prior to second dimension analysis. Thus basic reversed-acidic reversed phase high performance liquid chromatography (bRP-aRP-LC)[49] has gained popularity as method that provides more effective separation than SCX. It generates cleaner fractions and that reduces both sample processing steps and sample loss. In this fractionation workflow, the first dimension is an offline fractionation with a solvent system at pH 10.0, and the second dimension is an online LC-MS/MS analysis with a solvent system at pH 2.0.

The objective of this experiment is to evaluate various reverse phase chromatographic columns using protein standards to improve intact protein

separation of complex mixtures. The multidimensional chromatographic technique bRP-aRP-LC was evaluated using *E. coli* lysate as a model complex mixture to improve the resolving power for separation of intact proteins, and to increase the number of protein and proteoform identifications.

Evaluation of Chromatographic Columns

Material and Methods

Sample preparation and HPLC method: Ten micromolar solution of lysozyme, cytochrome c, myoglobin and ribonuclease A (Sigma Aldrich, St. Louis, MO) was prepared in solvent A (97.4% water, 2.5% acetonitrile (ACN) and 0.1% formic acid). Reversed phase chromatography with 50 µL protein standard mix was performed using a Thermo Accela LC (San Jose, CA) or Shimadzu Prominence LC (Columbia, MD) pump with the chromatographic columns, at flow rates of 300 nL/min to 50 µL/min depending on the inner diameter of the column. A gradient elution was employed and the concentration of solvent B (97.5% acetonitrile, 2.4% water, 0.1% formic acid) was increased linearly from zero to 85% in 60 minutes. The chromatograms were acquired with a SPD-10A UV (Shimadzu, Columbia, MD) spectrometer at a wavelength of 214 nm.

Results and Discussion

The six acid reversed phase columns that were evaluated were Kinetex core shell column (Phenomenex, Toraance, CA), C3 capillary column (Agilent, Wilmington, De), Xbridge (Waters, Milford, MA), Grace C18 (Vydac, Deerfield, IL), PLRP-S capillary column (Agilent, Wilmington, De) and Proswift Monolithic

column (Dionex, Bannockburn, IL). The column chemistry, diameter and particle size for the six columns are indicated in Table 1. All columns were evaluated using standard protein mix. The Kinetex core shell column and the Agilent Zorbax columns were evaluated for the second dimension in the bRP-aRP LC. The Waters Xbridge column was evaluated for the first dimension for the bRP-aRP LC.

Column	**Chemistry**	**Particle Size (µm)**	**Pore Size (Å)**	**Dimension (mm)**	**Number of Theoretical Plates**	**Flow Rate**
Kinetex Core-shell	C18	2.1	100	150x2.1	27,300	50 µL/min
Agilent Zorbax	C3	5	300	150x0.1	17,700	300 nL/min
Waters Xbridge	C18	3.5	300	250x4.6	18,300	300 µL/min
Grace Vydac	C4	5	300	250x1.0	3,600	50 µL/min
Michrom PLRP-S	PS/DVB	5	1000	150x0.1	2,800	500 nL/min
Dionex ProSwift Monolithic	Phenyl	NA	NA	50X1.0	2,200	100 µL/min

Table 1: Chromatographic columns evaluated for intact proteins analysis.

The number of theoretical plates for each column was calculated based on the retention time and peak width for cytochrome c for each run. From the columns evaluated, the number of theoretical plates observed was highest in the Kinetex core shell, Agilent C3 and the Xbridge, respectively. Consequently, these three columns were evaluated in the analysis of LC-MS/MS analysis of *E. coli* lysate.

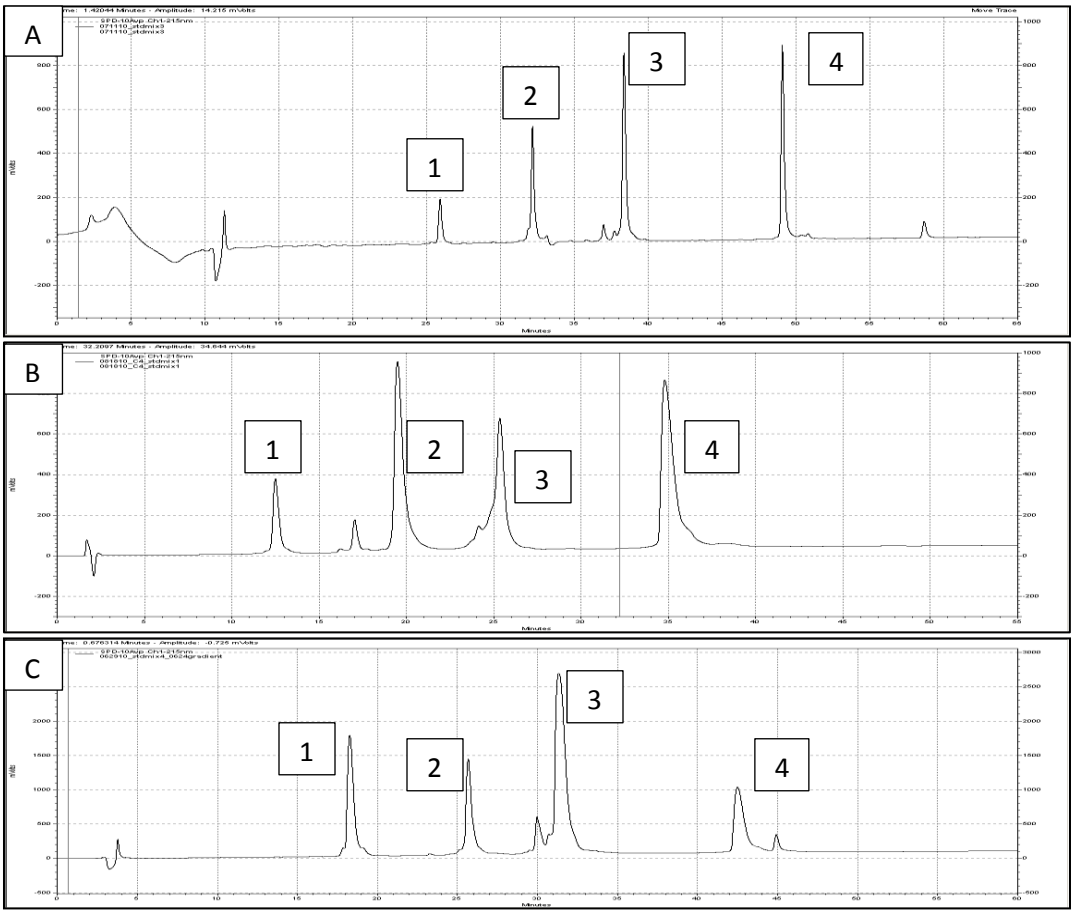

Figure 6: Chromatograms for reversed phase separation of four protein standard mix using, using (A) Kinetex core shell column, (B) Agilent C3 column and (C) Xbridge column Standards used were: 1. Lysozyme; 2. Ribonuclease A; 3. Cytochrome c; 4. Myoglobin.

The advantage of the Kinetex core shell and the Agilent C3 column were that the particles were not fully porous resulting in a reduced diffusion path and thus maximizing separation efficiency. The porous packing material of these columns also results in lower back pressure on the HPLC system. Thus the amount of intact protein that can be loaded on these columns is 2 to 3 fold times higher compared to columns packed with non-porous particles. The Agilent C3 column has a smaller inner diameter; therefore it is a nanoflow column and it requires less starting material for intact protein analysis. Due to the lower flow rate of 300 nL/min, the C3 column is compatible with the Nano ES source, therefore, it provides a higher sensitivity and dynamic range for the analysis of intact proteins. In addition, the Xbridge column provided a higher number of theoretical plates, since it is 250 mm in length. While longer columns increase the peak capacity for the analysis of complex mixtures, the back pressure on the system is also higher, thus the amount of sample that can be loaded onto the Xbridge column is much lower than Kinetex core shell column. The biggest advantage of the Xbridge column was that the particles were stable at extreme solvent pH and thus could be employed with various solvent systems. In the literature, PLRP-S and Monolithic columns have showed reduced chromatographic peak widths, fast mass transfer, low back pressure and high loadability of sample.[50-53] However in our hands, we found that the performance of these columns, based on the number of theoretical plates, separation and peak width, was lower for intact proteins for the lower molecular weight proteome. For the reasons stated above, the three

chromatographic columns utilized for the work in this thesis are the Kinetex core shell column, Aglient C3 column and the Waters Xbridge column.

Evaluation of bRP-aRP LC-MS/MS analysis of *E. coli* lysate

Material and Methods

E. coli Lysate: Forty milligrams (mg) of lyophilized K12 strain *E. coli* (Sigma Aldrich, St. Louis, MO) was suspended in 500 µl of 10% formic acid and vortexed. The sample was centrifuged at 14,000 rotation per minute (RPM) for 15 minutes. Proteins were precipitated with cold acetone for 30 minutes and centrifuged at 10,000 g for 20 minutes. The resultant pellet was re-suspended in 500 µL of Solvent A (97.5% water, 2.4% ACN, 0.1% formic acid) and filtered through a 0.22 µM filter. The protein concentration of the sample was determined using RC/DC assay (BioRad).

1D-LC analysis: Seventy-five micrograms of *E. coli* lysate was injected onto a Kinetex core shell column with a flow rate of 100 µL/min. Reversed phase chromatography was carried out on a Shimadzu Prominence LC system and Autosampler, with a linear gradient of increasing concentration of Solvent B (97.5% ACN, 2.4% water and 0.1% formic acid) from 10% to 85% over 110 minutes.

bRP-LC analysis: Two hundred micrograms (µg) of *E. coli* lysate was injected on the Xbridge column with a flow rate of 50 µL/min. The two solvents

used for the first dimension were: solvent A composed of 50 mM ammonium hydroxide in water and solvent B composed 50 mM ammonium hydroxide in water and ACN (2:8).[54] Reversed phase chromatography was carried out on a Thermo Accela LC system with a step gradient of increasing concentration of solvent B from 10% to 90% over 60 minutes. The samples were detected with a SPD-10A UV spectrometer at 214 nm. A total of five fractions were collected, lyophilized and re-suspended in solvent A (97.5% water, 2.4% ACN, 0.1% formic acid) for the second dimension. Each fraction was injected onto a Kinetex core shell column with a flow rate of 100 µL/min. Reversed phase chromatography was carried out on Shimadzu Prominence LC system and autosampler with a linear gradient of increasing concentration of solvent B (97.5% ACN, 2.4% water and 0.1% formic acid) from 10% to 85% over 110 minutes.

MS analysis for 1D and aRP LC samples: A LC was connected in line with an LTQ-Orbitrap-XL (ThermoFisher, San Jose, CA) and the precursor scans were recorded in the Orbitrap at a resolution of 60,000 at 400 m/z. The four most abundant signals for each precursor scan were subjected to CID fragmentation with activation energy at nominal 35. MS/MS spectra were recorded in the Orbitrap at a resolution of 30,000 at 400 m/z. Data dependent analysis was set to isolate precursor ions with unassigned charges and charges greater than +4. The isolation width for the precursor ions was set to 10 Da. The automatic gain control (AGC) targets were set to 1E6 for precursor scan and 1E5 for the four MS/MS scans.

Bioinformatics: Database searches were performed using ProSightPC 2.0[55] (ThermoFisher, San Jose, CA) against a custom UniProt *E. coli* database

consisting of proteins with molecular weight less than 30KDa. The THRASH[56] algorithm was used to decharge both the precursor and the fragment ions. Precursor mass tolerance was set to 250 Da and the fragment mass tolerance was set to 15 ppm. The ΔM mode on ProSightPC PC 2.0 was used to localize any mass shifts at the N- or C- terminus of the protein. Post-translational modifications and mass shifts were investigated manually using Sequence Gazer available in the software. The proteins identified were automatically assigned an E-value. Identifications with E-values lower than 10E-4 were considered as strong identifications.

Results and Discussion

Fractionation of intact proteins in complex mixtures plays an important role in increasing the number of proteoforms and proteins identified in a given analysis. Fractionation using multidimensional LC-MS/MS analysis enables the identification of proteins with lower abundances, proteins that may co-elute and proteins whose signal may be suppressed in a 1D- LC-MS/MS analysis. The step gradient employed and the UV chromatogram for the first dimension fractionation are shown in Figure 7.

It was important to employ the step gradient in the first dimension, so that we observe a valley in the chromatogram at a fixed time interval. Five fractions were collected, each at the valley in the chromatogram as shown in Figure 7. This ensures that a group of proteins with different hydrophobicity can be collected in each different fraction and it aids in minimizing overlap between fractions. In this report, a total of 14 unique proteins were confidently identified using 1D-LC-

MS/MS analysis, while 38 unique proteins were identified from the five fractions using bRP-aRP LC-MS/M.

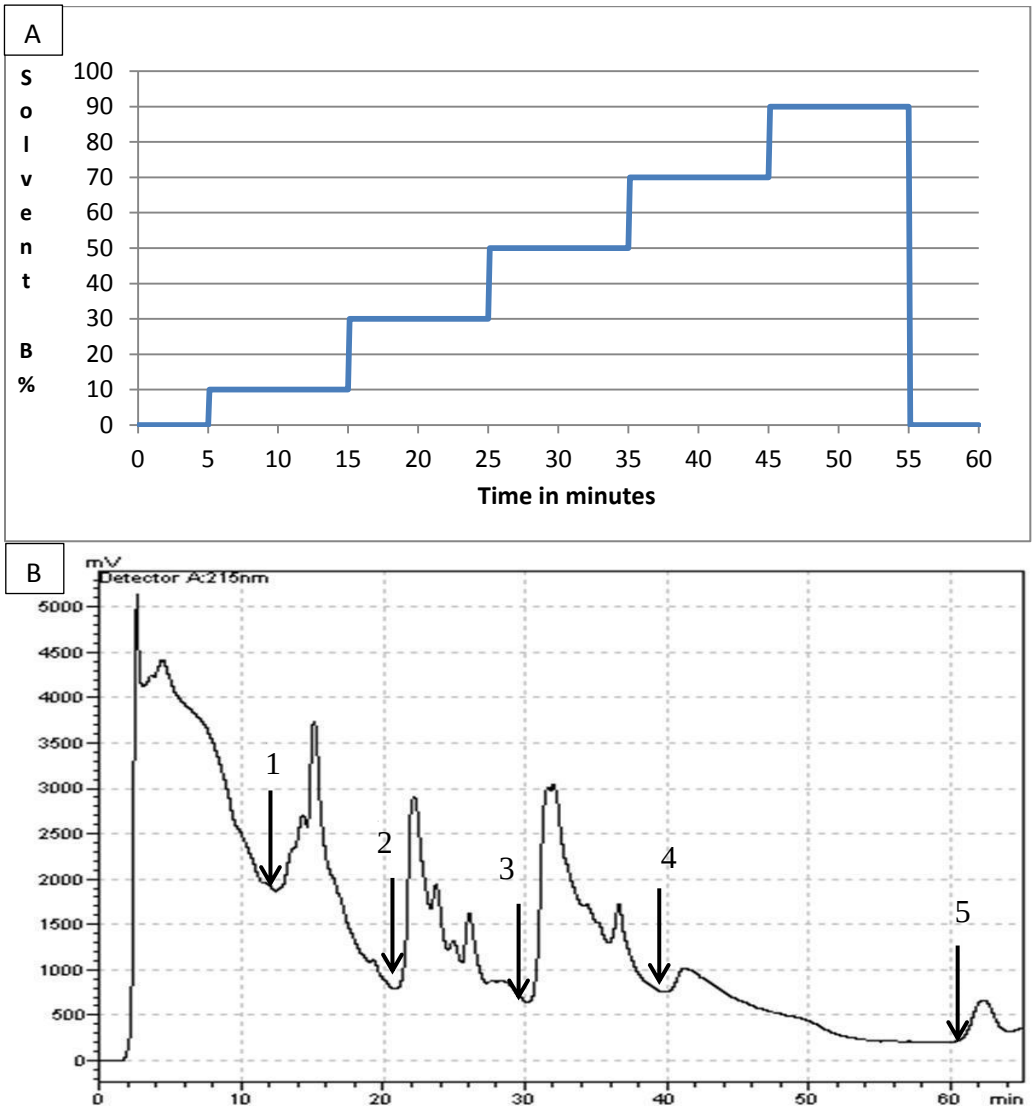

Figure 7: (A) Step gradient employed for the first dimension (B) Chromatogram of the first dimension separation of *E. coli* lysate at pH of 10;

Three of the proteins identified in 1D LC-MS/MS were not observed in the bRP-aRP-LC-MS/MS work flow. The mass ranges of the proteins identified in the bRP-aRP LC-MS/MS were from 4.9 KDa to 13.6 KDa and the combined number of proteins identified doubled compared to 1D LC-MS/MS analysis. A list of the proteins identified from both the analyses is provided in Appendix Tables 1 and 2. Figure 8 and 9 shows the mass ranges of the proteins identified and the number of unique proteins identified in each fraction. There were 22 proteins identified between mass ranges of 7000 Da to 10,000 Da whereas, 8 proteins were identified between 10,000 Da and 13,000 Da.

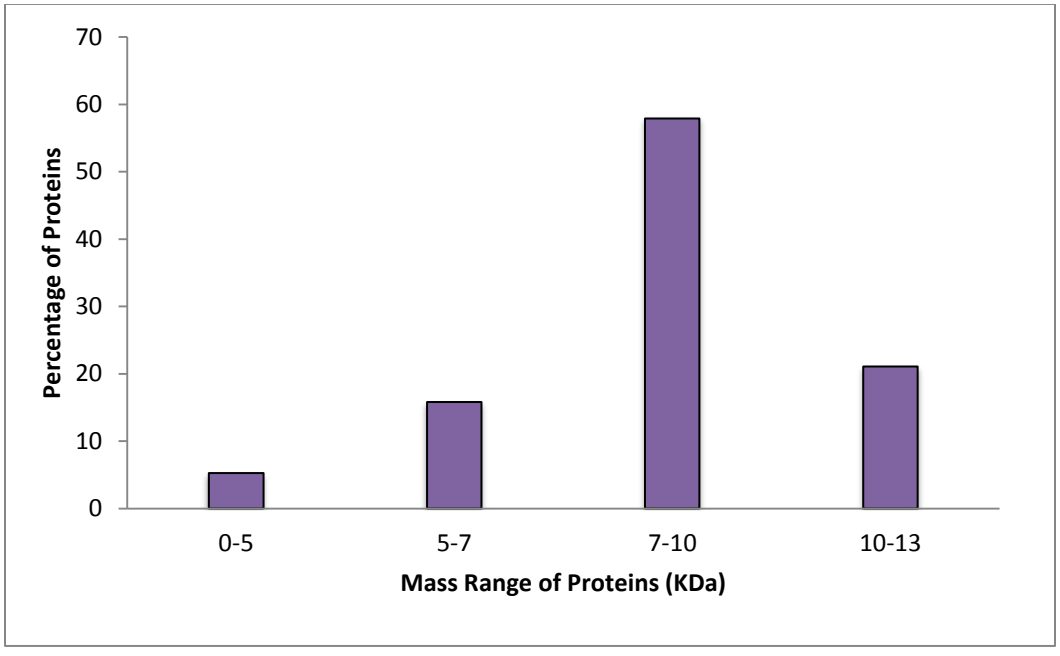

Figure 8: Mass distribution of all the proteins identified in all five fractions.

The number of unique proteins identified in fractions 5, 4 and 3 were 20, 2, and 6 proteins, respectively. While, there was some overlap between fractions, 80% of the proteins identified were unique to a given fraction. There were no proteins identified in fractions 1 and 2, and a majority of the proteins were identified in fraction 5.

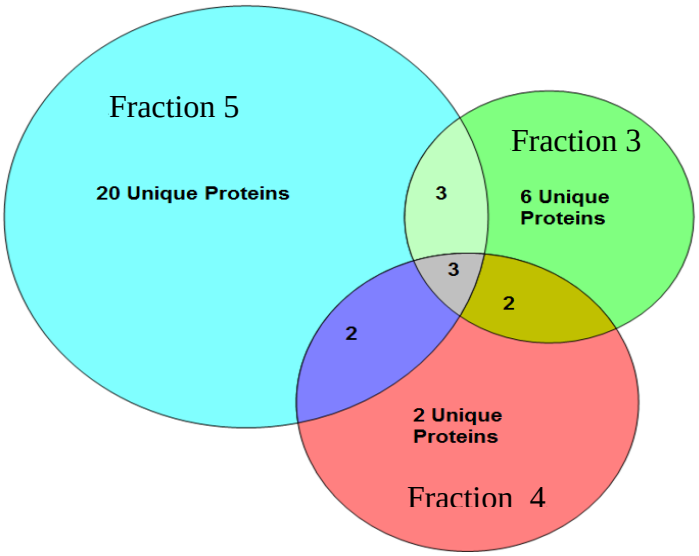

Figure 9: Distinct proteins identified in each fraction in bRP-aRP LC-MS/MS analysis.

The heaviest protein identified was in bRP fraction 5 and eluted at 61.42 minutes in the aRP gradient; a precursor ion of 1246.12 with a charge state of +11 was isolated and fragmented. The protein was identified with a reliable E value of 9.11E-06 and was identified as hypothetical protein ECB_03458 with an intact mass of 13688.3 Da which was 0.05 Da less than the theoretical mass. The precursor and product ion scans, the decharged product ions and the fragments matched are shown in Figure 10.

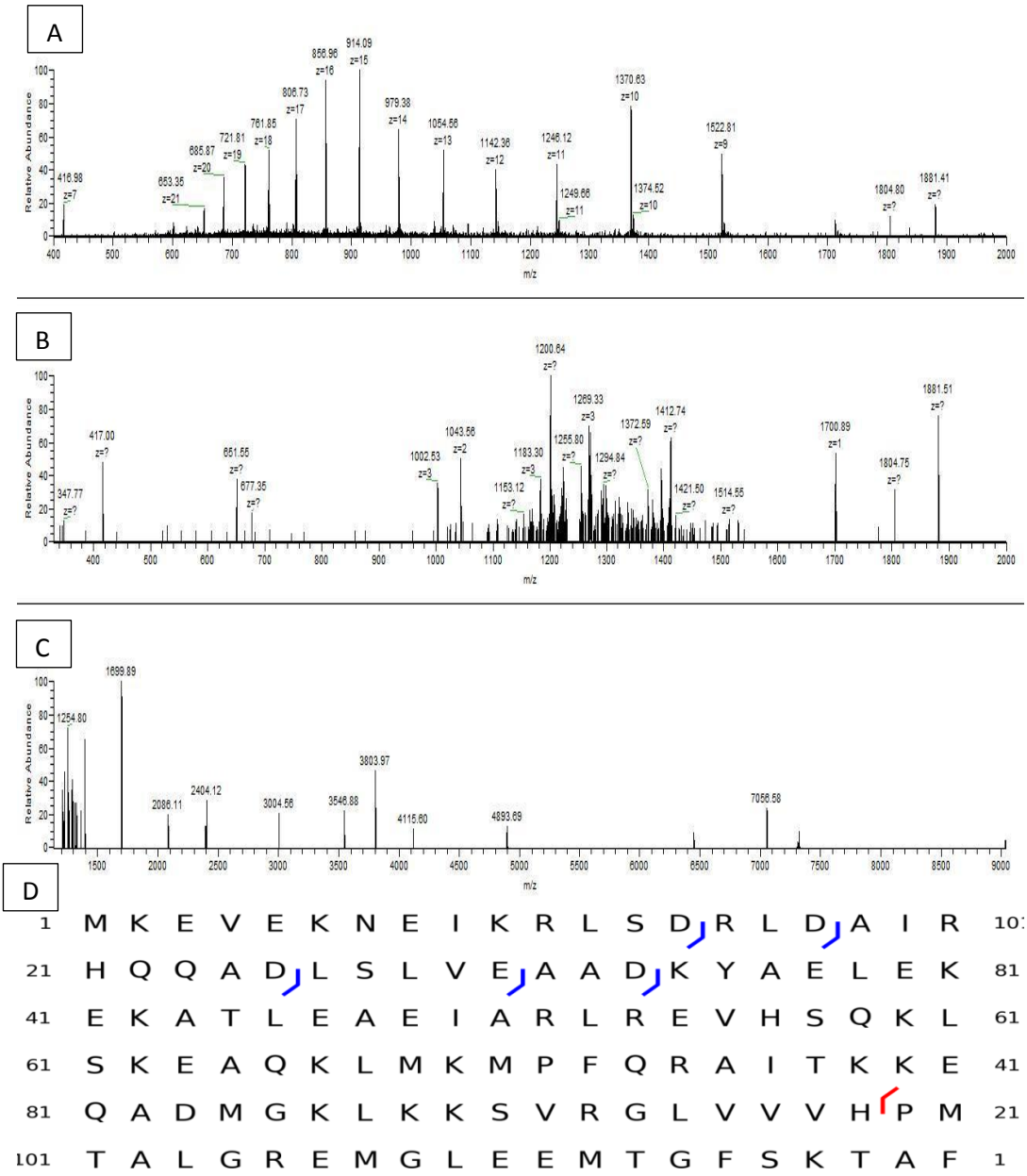

Figure 10: (A) Precursor ions scan at retention time 61.42 in fraction 5, (B) Product ions from m/z=1246.12 with charge state +11, (C) Decharged product ion scan; (D) protein sequence and observed 5 b-ions and 1 y-ion from hypothetical protein ECB_0345.

Post translational modifications were observed on 8 proteins and localized on 7 for 1D LC-MS/MS analysis. While, there were modifications observed on 9 proteins and localized on 8 for bRP-aRP LCMS/MS analysis of the *E. coli* lysate. Modifications on the proteins were manually investigated using the Sequence Gazer tool in ProSightPC 2.0 for proteins where a mass difference was observed between the theoretical and experimental intact masses. The number of matched fragments increases when a PTM is localized on a protein, which lowers the E-value of the protein, and increases the confidence of the identification of the protein. 50S Ribosomal protein is identified in fraction 5 of the bRP-aRP LC analysis. The precursor ion spectrum at 19.73 minutes and the product ion spectrum for ions of m/z 894.52 with a charge state of +7 are shown in Figure 11. The mass observed for the protein was 6250.57 Da which 13.99 Da is heavier than the theoretical mass of the protein. The localization of a post translational modification on 50S ribosomal protein L33 is shown in Figure 12. The mass difference between the observed and experimental intact mass was 13.98 Da. In addition, since only the 11 y-ions were matched, the PTM on the protein should theoretically be on the N-terminus. Thus, when a mass shift of 13.98 Da or methylation was applied to the alanine residue on the N-terminus, the number of fragments matching the protein increased from 11 fragments to 25 fragments. Also, the E-Value of the protein dropped from 2.06E-18 to 8.86E-49 making it a stronger identification.

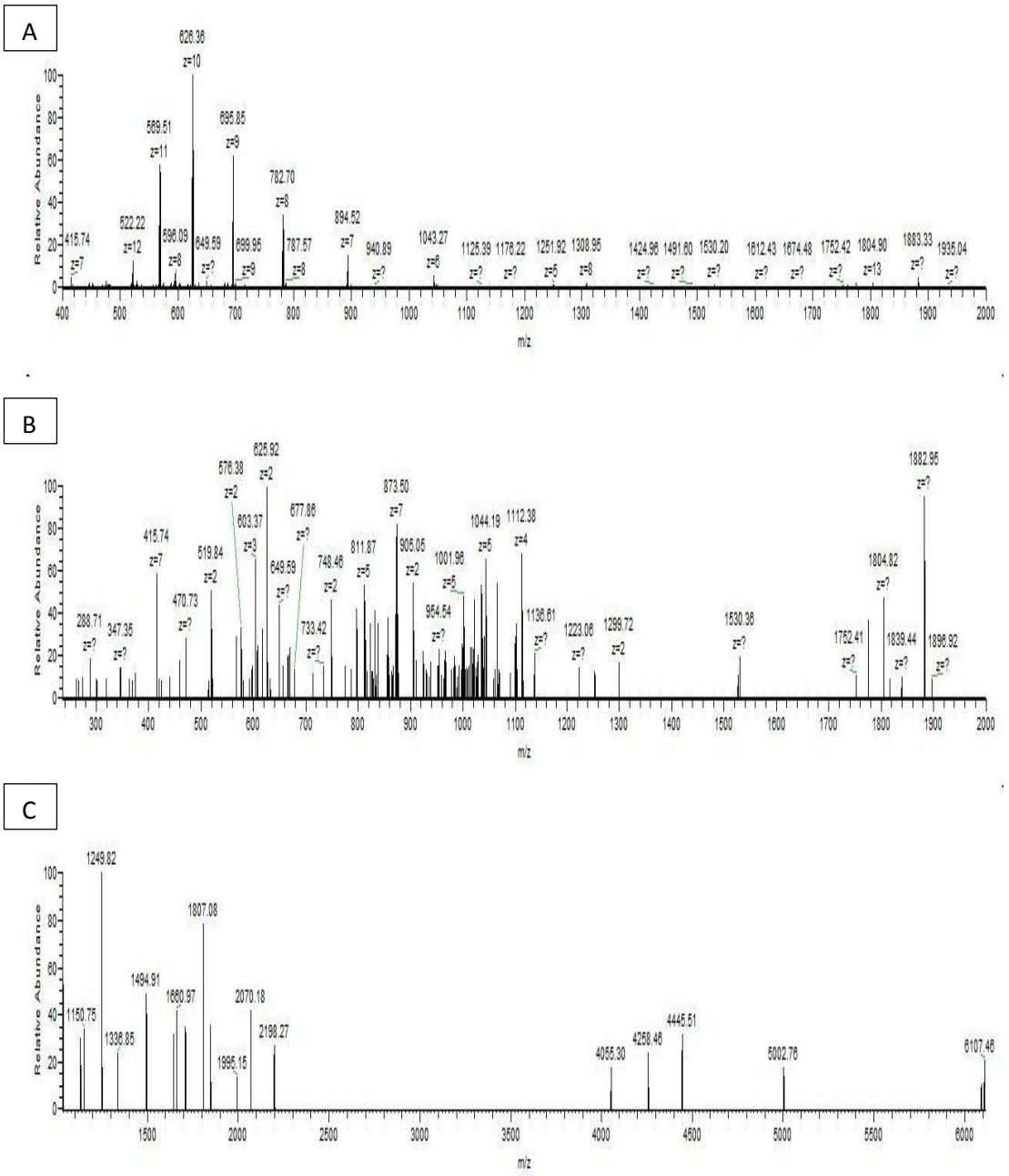

Figure 11: (A) Precursor ion spectrum at retention time 19.73 minutes from fraction 5 of bRP-aRP LC analysis of 50S ribosomal protein L33 (B) Product ions from m/z 894.53 with charge of +7 (C) Decharged product ion spectrum.

Figure 12: Protein sequence of 50S ribosomal protein L33 (A) based on 11 y-ions observed with E value of 2.06E-18 (B) based on 14 b-ions and 11 y-ions with E value of 8.86E-49 when methylation is assigned first alanine amino acid residue.

DNA-binding transcriptional regulator, alpha subunit was identified in fraction 5 from the bRP-aRP LC analysis. The precursor ion spectrum at 50.55 minutes and the product ion spectrum for ions of m/z 737.76 with a charge state of +13 and the fragmentation observed are shown in Figure 13. The protein was identified with a strong E-value of 4.27E-19 and the observed mass of the protein was 9259.08 Da which was 0.01 Da higher than the theoretical mass.

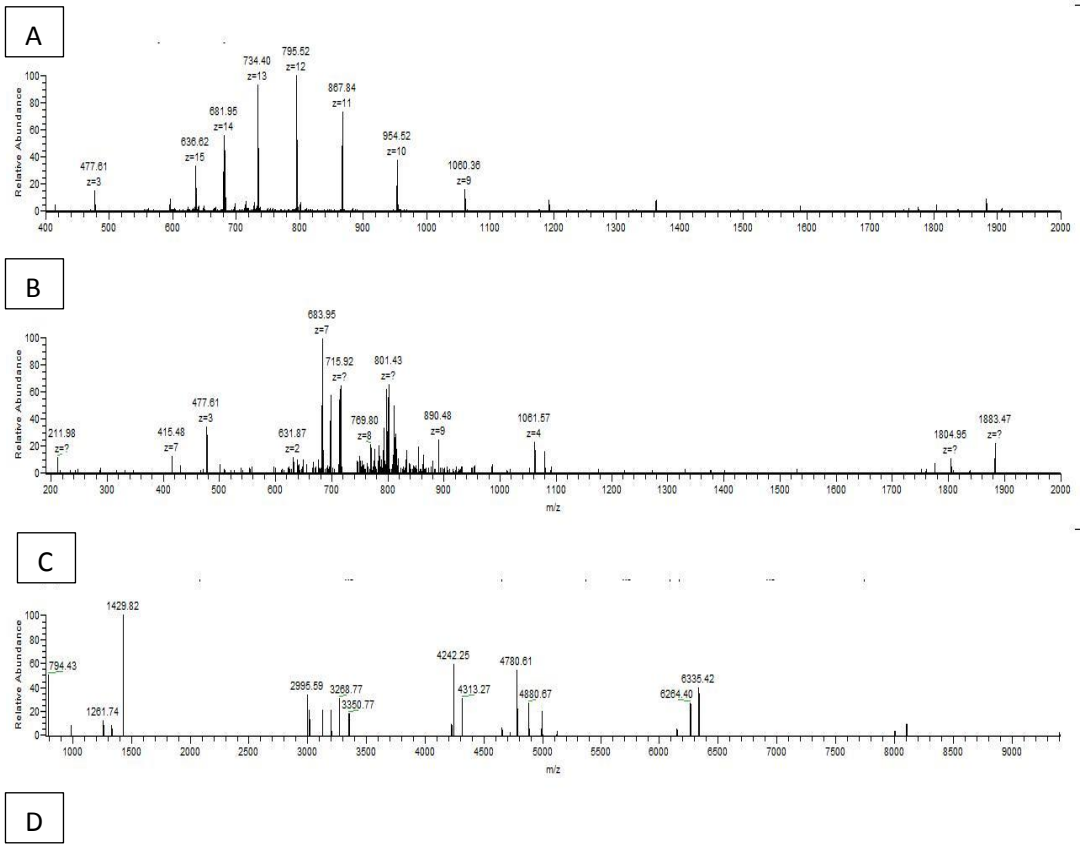

Figure 13: (A) Precursor ions scan at retention time 50.55 in fraction 5, (B) Product ions from m/z 737.76 with charge state +13, (C) Decharged product ion scan; (D) protein sequence and observed 8 b-ions and 11 y-ions from DNA-binding transcriptional regulator, alpha subunit.

Three experimental replicates of bRP-HPLC-MS/MS analysis were performed to evaluate the reproducibility of the method. The number of proteins identified in the three experiments was 38, 36, and 32, respectively. While most of the proteins identified between the analyses were the same, there were seven proteins identified that were not detected in all three experiments. The reason for the difference in identification of proteins is the complexity of the sample. During the LC-MS/MS analysis in the second dimension, at any given time there is more than one protein eluting from the column. Thus, the mass spectrometer could be isolating and fragmenting different precursor ions for a given fraction between experiments. A total of forty five unique proteins were identified from the three experimental replicates.

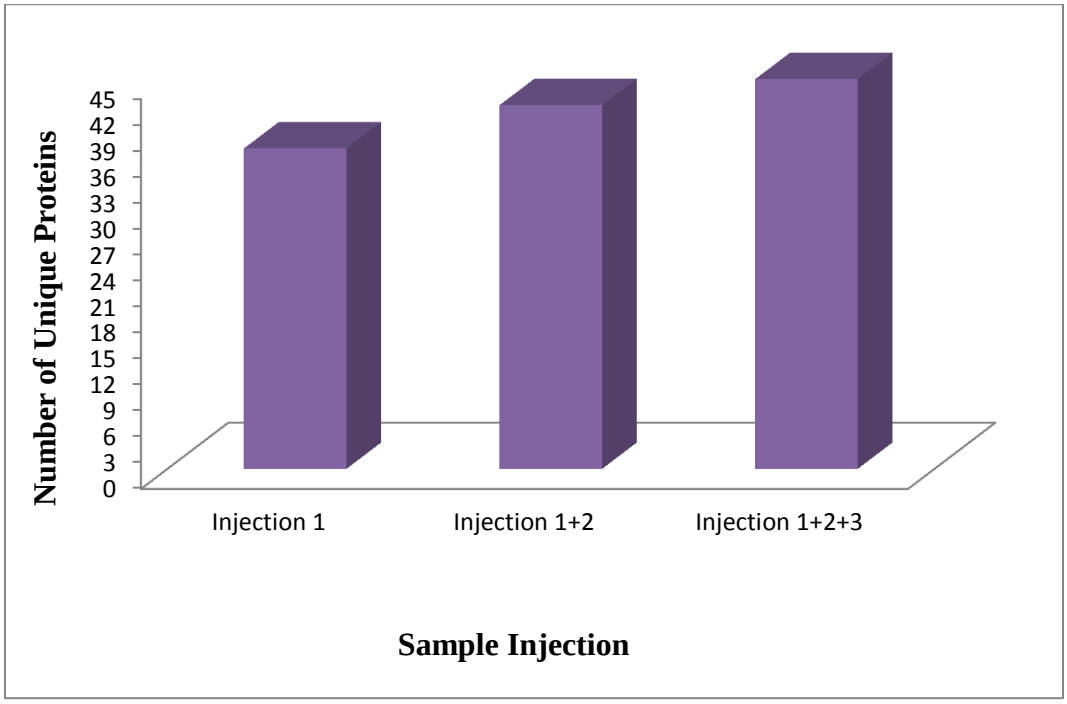

Figure 14: Number of unique proteins identified from three replicate analyses of bRP-aRP LC.

Conclusion

Optimized front end fractionation is essential to increase the number of identifications and proteoforms of intact protein from complex mixtures. From the columns evaluated, the Kinetex core shell column, Agilent C3 and the Waters Xbridge column were employed for the 1D-LC-MS/MS and bRP-aRP-LC MS/MS analysis. These columns provided a higher number of theoretical plates and were compatible with the HPLC system, as well as, solvents used in the analysis. In the bRP-aRP LC-MS/MS analysis, the step gradient in the first dimension was important so as to observe valleys at a fixed time interval in the chromatogram. This led in reducing the overlap of proteins identified between fractions. The number of identifications using bRP-aRP LC-MS/MS analysis increases more than two fold. Therefore, there is a need for offline fractionation of intact proteins in complex mixtures prior to LC-MS/MS analysis.

Chapter 3: Optimization of Fragmentation Conditions: CID, HCD and ETD

Introduction

The use of tandem mass spectrometry is important in proteomics for protein identification and characterization. In tandem mass spectrometry the charged ions within a specific mass to charge ratio (m/z) are isolated, subjected to fragmentation and the mass to charge ratio of the resulting fragment ions recorded. Optimized fragmentation of proteins is a key step in top-down approaches using mass spectrometry based proteomics. Employing a fragmentation method that produces a homologous series of fragment ions is important for confident protein identifications in complex mixtures. The most common fragmentation techniques for intact proteins are via collisions of an inert-gas or electron-based fragmentation. The three type of fragmentation available on LTQ-Orbitrap XL are low energy collisionally induced dissociation (CID), high energy collision dissociation (HCD) and electron transfer dissociation (ETD). The basic principle of all three fragmentation methods are discussed in Chapter 1.

The Orbitrap provides isotopic resolution of the analyte, which allows the charge state of ions to be determined. The charge state of a protein can be calculated using the following formula:

$$Z = {}^{1}\Big/\left({}^{\Delta m}\Big/_{Z_{\left({}^{13}C_{x+1}-{}^{13}C_{x}\right)}}\right)$$

In addition to fragmentation methods, the number of averaged scans and the resolution also play an important role in assigning the fragment ions of the proteins being analyzed. In the relationship above the charge state is the inverse of the mass to charge difference between two isotopic peaks, since the mass difference between two ^{13}C peaks is 1 Da

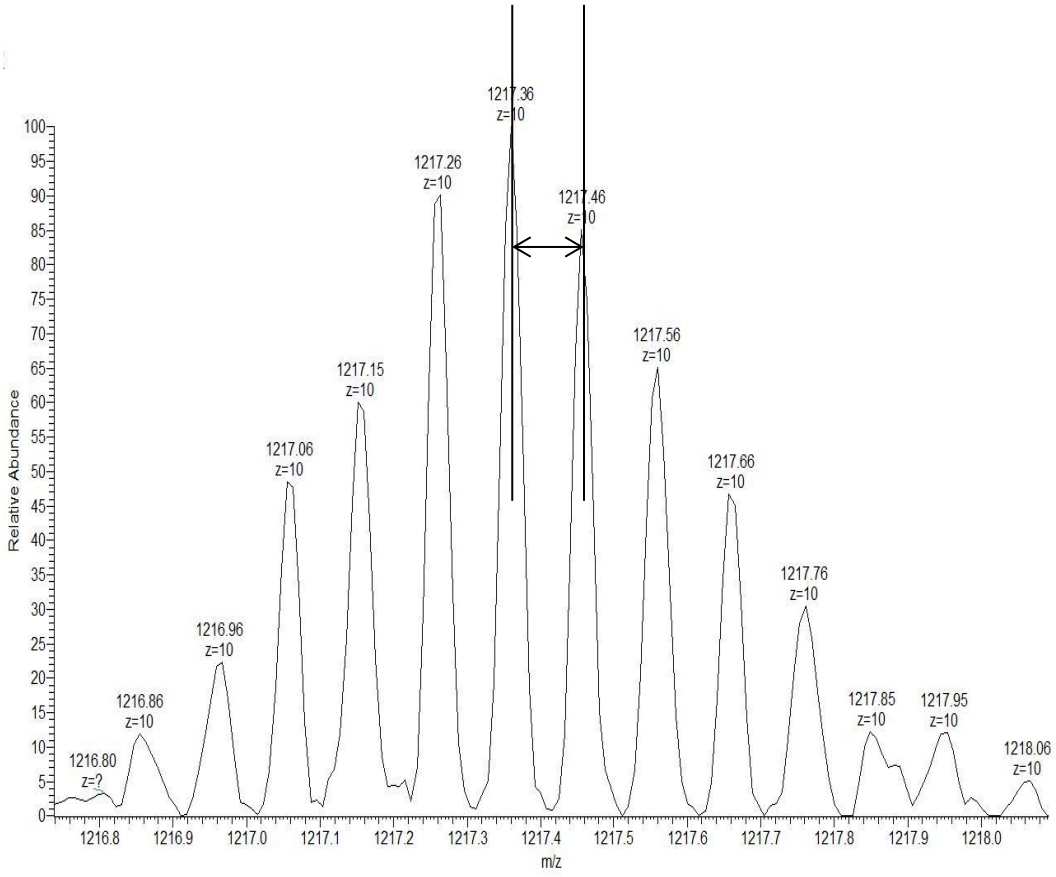

Figure 15: Expanded view of precursor ion spectrum from an *E. coli* LC-MS/MS analysis, showing the mass difference between isotopic peaks.

As the mass of the protein increases, the number of charge states usually increases and the mass difference between the isotopic peaks decreases, which is why top-down analysis of intact proteins is challenging on a HPLC time-scale.

The objective of the experiments presented in this chapter was to optimize the parameters employed to obtain product ion spectra; specifically, the fragmentation conditions, number of averaged scans and resolution for different mass range of the proteins desired to be analyzed. Intact ubiquitin (8.5 KDa), cytochrome c (12.7 KDa), myoglobin (16.8 KDa) were employed to evaluate the reaction time for ETD and activation energies for CID and HCD fragmentation methods. Fragmentation conditions were also evaluated for intact proteins from the cytosol of MCF-7 cancer cells on a LC-MS/MS timescale. The effect of averaged scans for precursor ions was evaluated by infusing myoglobin and by the LC-MS/MS analysis of cytosolic proteins from MCF-7 cancer cells.

.

Evaluation of Fragmentation Conditions for Intact Proteins

Materials and Methods

Sample preparation and MS analysis: One micromolar solution each of ubiquitin, cytochrome c, myoglobin (Sigma Aldrich, St. Louis, MO) was prepared in 50% ACN, 49.9% water and 0.1% formic acid. Each protein solution was infused individually into the LTQ-Orbtrap XL using a syringe that was set at a flow rate of 1 μL/min. The precursor ion scan was acquired at a resolution of 60,000 at 400 m/z and the fragment ion scans were automatically acquired at a resolution of 15,000 at 400 m/z for the top three most abundant ions. Data dependent analysis was set to isolate precursor ions with unassigned charges and charges greater than +4. The isolation width for the precursor ions was set to 10 Da. The AGC targets

were set to 1E6 for precursor scans and 1E5 for the MS/MS scans. Dynamic exclusion was employed with a repeat count of 1, repeat duration of 30 seconds and exclusion duration of 30 seconds. The normalized collision energy was evaluated for CID and HCD for all three proteins. The energy was varied from 15 to 30 at increment of 5. ETD fragmentation was evaluated for the three proteins by varying the reaction time from 5 milliseconds (ms) to 25 ms at increments of 5 ms. The instrument was set to collect the spectrum for each repetitively for 2 minutes

Bioinformatics: Database search was performed using ProSightPC 2.0[55] against a custom UniProt database consisting of the sequences from the three proteins. The THRASH[56] algorithm was used to deconvolute both the precursor and the fragment ions. Precursor mass tolerance was set to 250 Da and the fragment mass tolerance was set to 15 ppm. The proteins identified were automatically assigned an E-value, identifications with E-value lower than 10E-4 were considered as a strong identification

Results and Discussion

Ubiquitin, cytochrome c and myoglobin were chosen for the evaluation of the fragmentation conditions since all three weigh less than 30 KDa. The number of fragments from the three proteins for the different conditions from the three fragmentation methods was identified using ProSightPC.

The results presented in Figures 16, 17 and 18 indicate the number of fragments identified using ProSightPC from ubiquitin, cytochrome c and myoglobin, respectively, for the three fragmentation techniques. The number of

fragments identified for all three proteins using CID and HCD methods decrease at higher activation energy. The molecular weight did not play an important role in CID fragmentation. In the case of HCD at higher activation energy internal fragmentation was observed, which the software could not match. It was observed that as the molecular weight of the protein increased the number of fragments matched using HCD decreased. In the case of the myoglobin due to extensive internal fragmentation, no fragments were identified at even low activation energy using HCD.

For ETD, results show that the charge state and the molecular weight of the protein play an important role in producing fragment ions. For ubiquitin, the reaction time of 20 ms yielded the highest number of matched fragments. Whereas, a reaction time of 5 ms yielded the higher number of matched fragments for myoglobin. In both cases it was observed that the number of fragments identified decreases with increasing reaction time for the analysis. The fragmentation using ETD is protein dependent. Thus there were no confident fragments identified from the ETD analysis of cytochrome c. Extensive fragmentation was observed from lower charge states precursor ions for CID and HCD compared to ETD for all three proteins. In literature, it is observed that ETD favors higher charge states.[57-59]

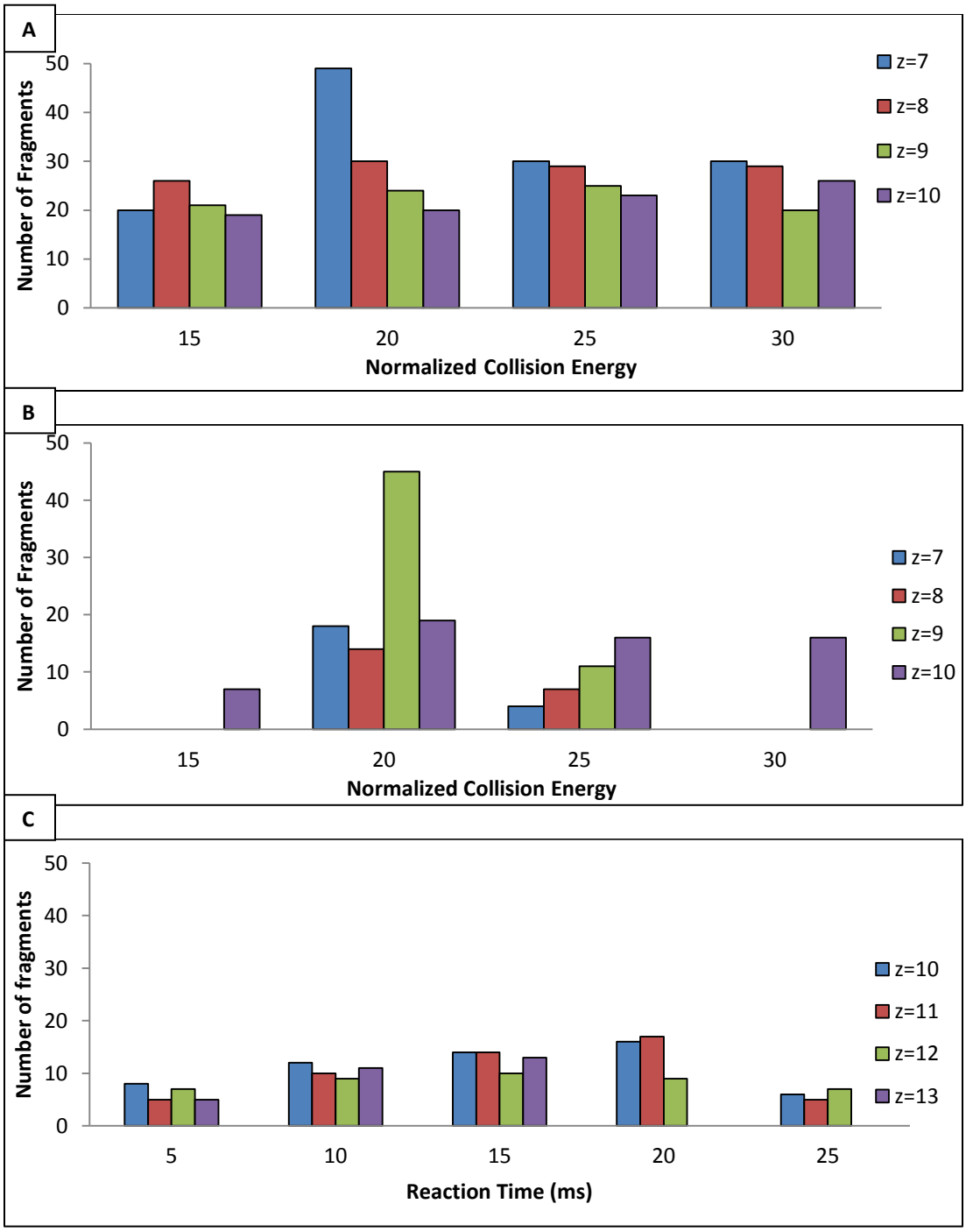

Figure 16: (A) Number of Fragments from CID for ubiquitin with varying normalized collision energy, (B) Number of fragments from HCD for ubiquitin with varying normalized collision energy, (C) Number of fragments from ETD for ubiquitin with varying reaction time.

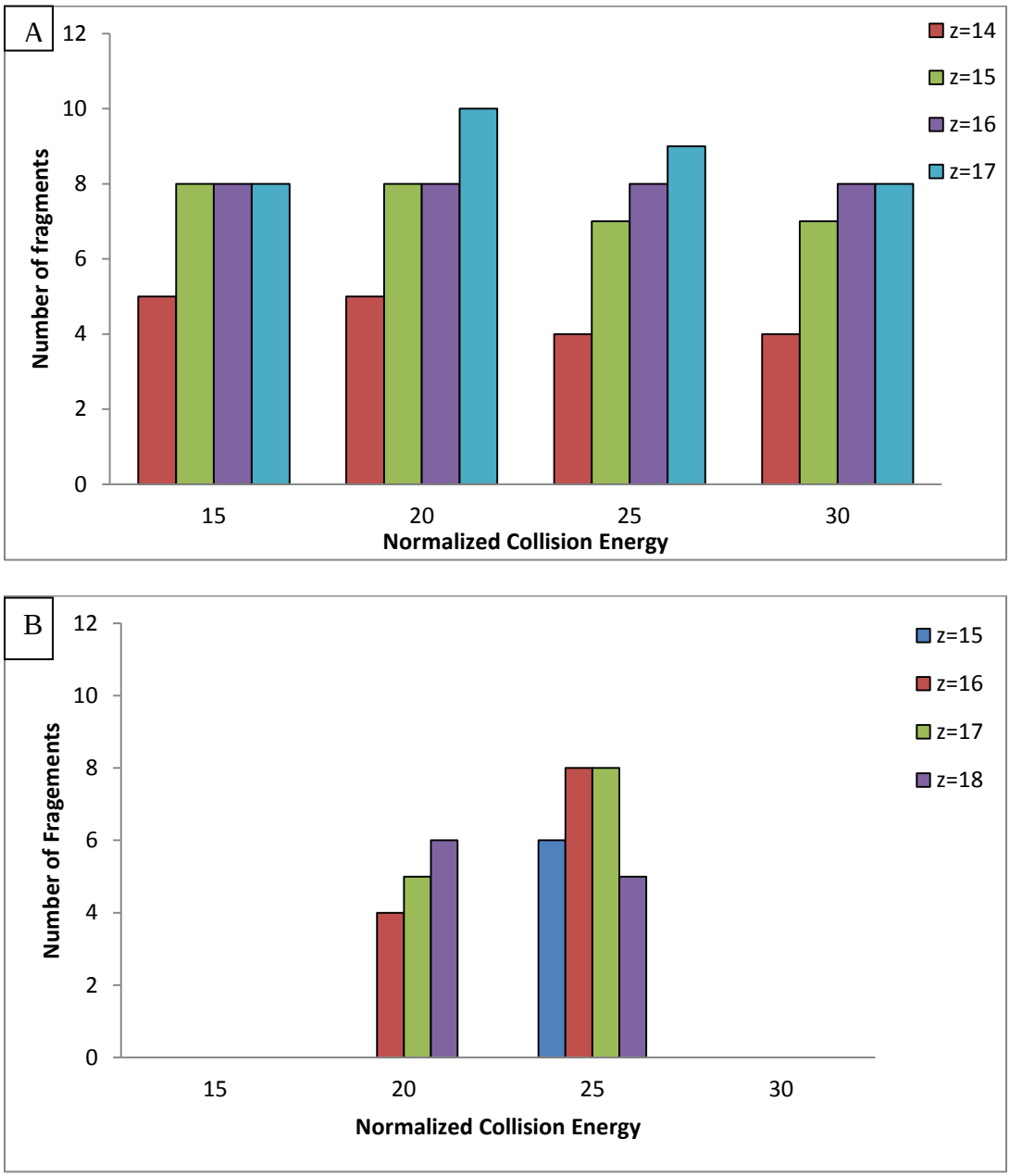

Figure 17: (A) Number of Fragments from CID for cytochrome c with varying normalized collision energy (B) Number of fragments from HCD for cytochrome c with varying normalized collision energy.

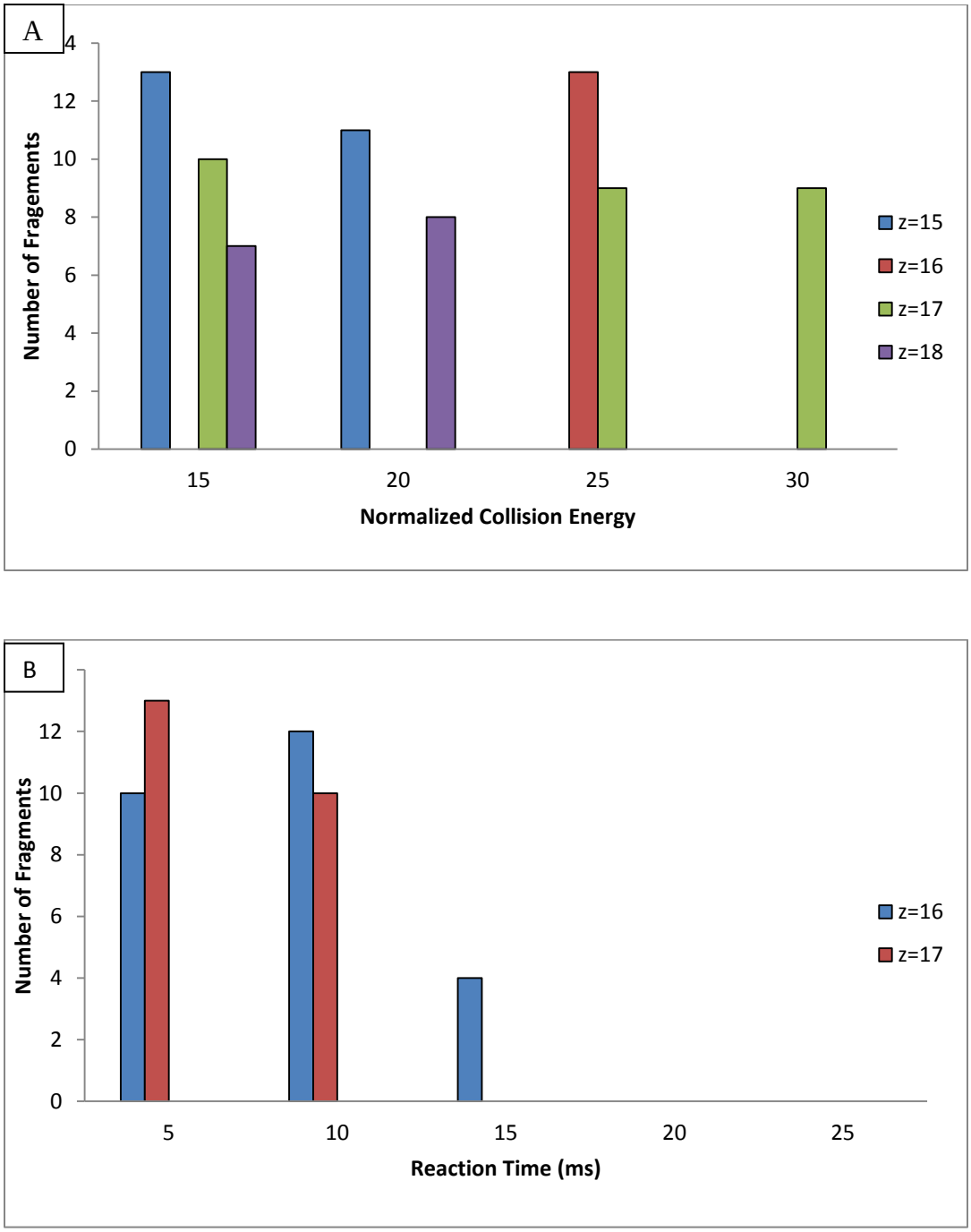

Figure 18: (A) Number of Fragments from CID for myoglobin with varying normalized collision energy, (B) Number of fragments from ETD for myoglobin with varying reaction time.

Effect of Averaged Scans for Precursor Ions

Materials and Methods

Sample preparation and MS analysis of myoglobin: One micromolar solution of myoglobin was prepared in 50% ACN, 49.9% water and 0.1% formic acid. The solution was directly infused into the LTQ-Orbtrap XL at using an ESI source and a syringe that was set at a flow rate of 1 µL/min. The precursor ion scans were acquired at resolutions of 15K, 30K 60K and 100K at 400 m/z and the number of scans averaged for each resolution setting was 1, 3, 5 and 10. The instrument was set to collect the spectra for each for 0.5 minutes.

Sample preparation of cytosolic proteins from MCF-7 cells: MCF-7 breast cancer cells were grown to confluence in Improved Minimal Essential Medium (ATCC, Manassas, VA) supplemented with 1% penicillin and 10% fetal bovine serum (Sigma Aldrich, St. Louis, MO) with 5% carbon dioxide and at 37 °C. The cells were detached by adding 5 mL of trypsin and incubating the flask at 37 °C for 5 minutes. The cells were washed two times with 5 volumes of phosphate buffered saline (PBS) and once with 10 volumes of 10 mM NaCl. The cells were collected by centrifugation at 500g for 10 minutes. Cytosolic proteins from the cells were extracted using digitonin buffer (10 mM PIPES, 0.015% digitonin, 300 mM sucrose, 100 mM NaCl, 3 mM MgCl2, 5 mM EDTA, and 1 mM protease inhibitor (PMSF), pH 6.8) at 4 °C on a shaker for 15 minutes. The cytosolic proteins were collected by centrifugation at 10,000 g for 10 minutes.[60] The protein concentration of the sample was determined using RC/DC assay. The proteins were precipitated

using $CHCl_3$/MeOH/H_2O (4:1:3)[61] and the pellet was re-suspended in 90% water, 5% ACN and 5% formic acid.

MS analysis LC analysis for cytosolic proteins: Shimadzu Prominence nano LC system and autosampler was used for the LC analysis. Fifty micrograms of cytosolic proteins from MCF-7 cells was loaded onto a Zorbax (5 μm, 5 mm x 0.3 mm) trap column (Agilent, Wilmington, De) and washed with Solvent A (97.5% Water, 2.5% ACN and 0.1% formic acid) for 10 minutes at 10 μL/min. The proteins were then fractionated using Agilent C3 (5 μm, 150 mm x 0.1 mm) column at a flow rate of 300 nL/min and a linear gradient of increasing concentration of Solvent B (97.5% ACN, 2.4% water and 0.1% formic acid) from 10% to 85% over 90 minutes. The LC was connected in line with a LTQ-Orbitrap-XL and the precursor scans were recorded in the Orbitrap at a resolution of 60,000 at 400 m/z with either 1 or 5 averaged scans. The most abundant signal for each precursor scan were subjected to CID fragmentation with activation energy of 35, MS/MS spectra were recorded in the Orbitrap at a resolution of 60,000 at 400 m/z with 5 averaged scans. Data dependent analysis was set to isolate precursor ions with unassigned charges and charges greater than +4. The isolation width for the precursor ions was set to 10 Da. The AGC targets were set to 1E6 for precursor scan and 1E5 for the four MS/MS scans. Dynamic exclusion was employed with a repeat count of 1, repeat duration of 15 seconds and exclusion duration of 30 seconds.

Bioinformatics: Database searching was performed using ProSightPC 2.0[55] against a custom UniProt human database consisting of the sequences of proteins with molecular weight less than 30 KDa. The THRASH[56] algorithm was

used to deconvolute both the precursor and the fragment ions. Precursor mass tolerance was set to 2000 Da and the fragment mass tolerance was set to 15 ppm. The ΔM mode on ProSightPC 2.0 was used to localize any mass shifts at the N- or C- terminus of the protein. Post-translational modifications and mass shifts were investigated manually using Sequence Gazer available in the software. The proteins identified were automatically assigned an E-value, identifications with E-value lower than 10E-4 was considered as a strong identification.

Results and Discussion

An intact protein ionized using ESI has multiple charges. These charges are calculated using the mass difference between the isotopic peaks as mentioned earlier in this chapter. As the mass of the protein increases, the difference between the isotopic peaks decreases. This is one of the main reasons why the analysis of intact proteins require high resolution FTMS or TOF based instrument. It is observed that the resolution of the Orbitrap diminishes with an increase of mass of the protein, even if m/z is unchanged. This is due to the collision with background noise that leads to the fragmentation of ions and the formation of fragment ions.[36] Due to the reason mentioned above it important the precursor and fragments ions for the intact proteins are acquired at high resolution.

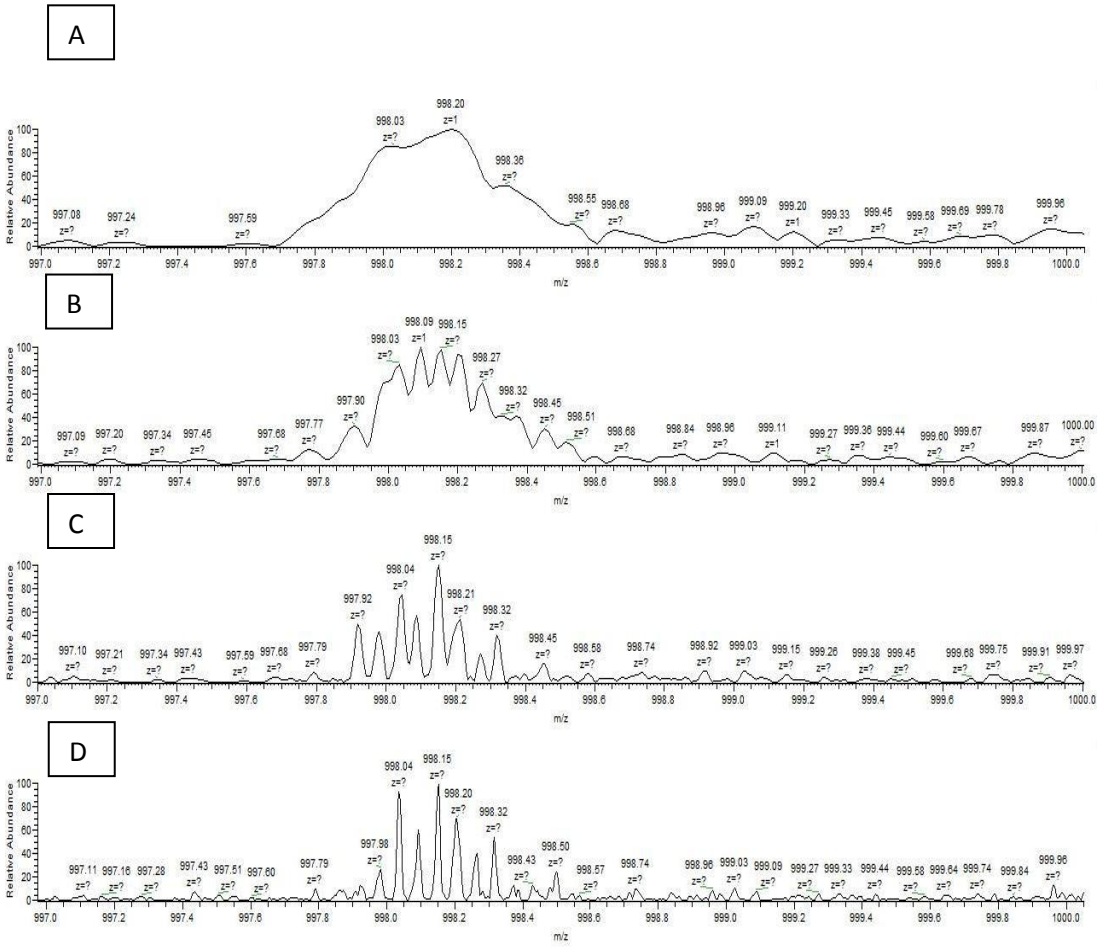

Figure 19: Zoomed in precursor ion scan of myoglobin at charge state of +17, showing the isotopic peaks at; (A) 15,000 resolution at m/z 400, (B) 30,000 resolution at m/z 400, (C) 60,000 resolution at m/z 400, (D) 100,000 resolution at m/z 400.

The precursor ion scan of the myoglobin was acquired by infusing the protein at various resolutions available on the LTQ-Orbitrap XL. In Figure 19, the ion at 998.15 with charge state +17 is evaluated for the for resolution settings. The isotopic peaks of the ions are well separated at higher resolution. Although, even at a resolution of the 100,000 at m/z 400 the charge state of this ion cannot be detected by a single scan. Thus, there is a need to average scans especially for higher molecular weight proteins.

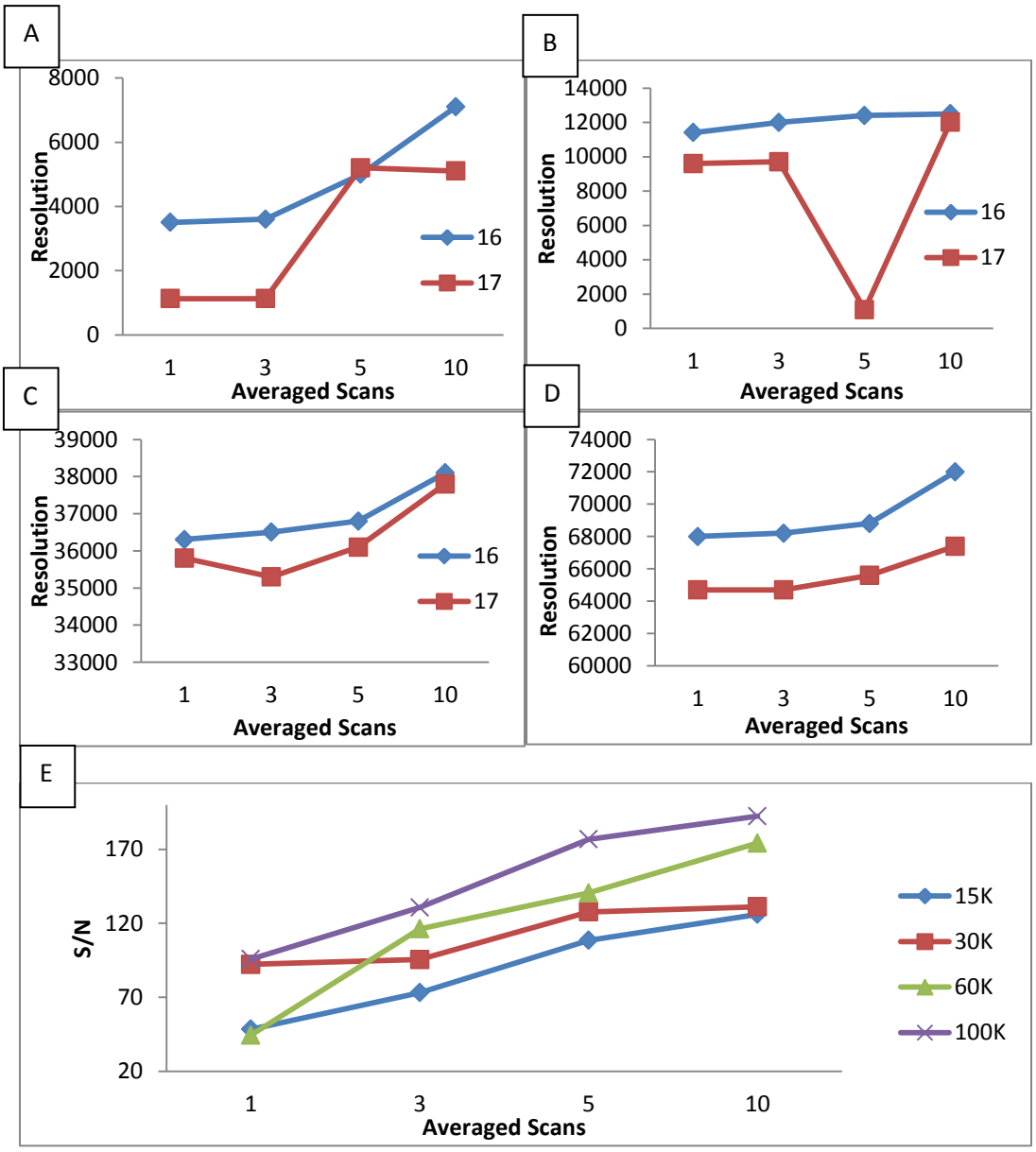

Figure 20: Effect of averaged scan on ions from myoglobin with charge states +16 and +17 for prescursor ion scans. (A) resolution of 15K at400 m/z, (B) resolution of 30K at 400 m/z (C) resolution of 60K at 400 m/z, (D) resolution of 100K at 400 m/z, (E) S/N of precursor ion with averaged scans for the four resolution settings.

The effect of resolution by averaging scans for the precurosr ion is shown in Figure 20. It is observed that the resolution is increased by averaging scans for the precursor ion. In the case of resolution of 100,000, the resolution of the ion with charge state of +16 increases from 68,000 for one averaged scan to 72,000 for ten averaged scans. The reason being, by averaging scans the instrument acquires the transient for a longer time due to which there is an increase in the resolution of the ion. In addition, the signal to noise for the +16 ion at the same resolution increases two fold from 95 using one averaged scan to 192 using ten averaged scans.

Even though it is experimentally evident that the averaged scan improves both the resolution and S/N to noise for ions from heavier proteins, it also increases the time taken by the instrument to acquire the scan. Typically, a single scan at resolution 15,000 takes 0.25 seconds to acquire; whereas, its takes 0.5 second, 1 second and 2 seconds to acquire a single scan at resolution of 30,000, 60,000 and 100,000, respectively. Ten averaged scans at 100,000 resolution takes approximately 20 seconds to acquire by the Orbitrap. While this time scale would be feasible for an experiment where the protein is being directly infused, it would be difficult to average so many scans at a high resolution on an LC-MS/MS time scale. It is important to utilize high resolution and averaged scans for both the precursor ions and fragment ions, but it is also important to balance the duty cycle for LC-MS analysis. In the case of complex mixtures, typically there is more than one protein eluting at a given time. Therefore, if the duty cycle on the mass spectrometer is very long, the low abundance ions may not be sampled by the mass

spectrometer. As a compromise, a resolution of 60,000 and five averaged scans for both precursor and fragment ions was employed in most experiments.

While the effects of averaged scans were analyzed by infusing proteins into the mass spectrometer, it was important to evaluate averaging the precursor ions scan for a complex mixture on a LC-MS/MS time scale. Cytosolic proteins from MCF-7 cancer cells were employed for this analysis. Two fifty-microgram samples were injected onto a C3 column in line with the LTQ-Orbitrap XL. The two experiments used the same LC gradient and MS fragmentation conditions. The first sample was analyzed using one precursor ion scan at resolution of 60,000 and the second with five averaged precursor scans at the same resolution.

The total number of unique proteins identified with one averaged precursor ion scan is 14 proteins, while the number of unique proteins identified with five average precursor ion scan is 13. The list of proteins identified from the two analyses is listed in the Appendix Table 3 and 4. There were 11 proteins overlapping between the two samples. The mass range of proteins identified using one averaged precursor ion scan is 4.0 KDa-12 KDa, whereas the mass range of the proteins identified with five averaged scans is from 5.0 KDa to 22.7 KDa. The percentage of proteins identified in the mass range of 9.0 KDa and 22.7 KDa is higher in the experiment where there are five averaged scans for the precursor ion. The heaviest protein identified is the heat shock beta protein-1 with a mass of 22.7 KDa

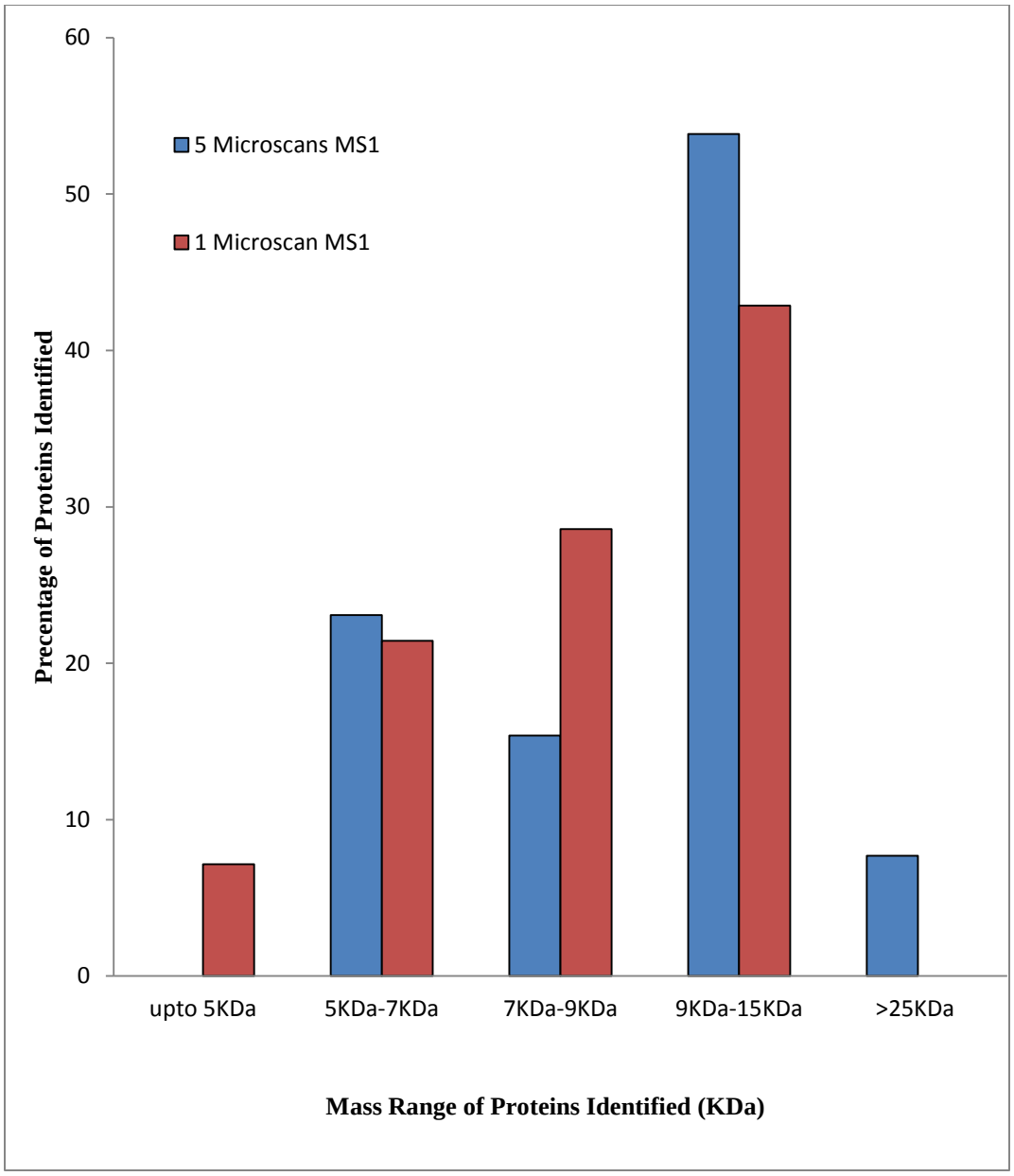

Figure 21: Percentage of proteins of different masses identified from MCF-7 cytosol by LC-MS/MS with one averaged and five averaged precursor ion scans.

The heavist protein in the analysis is identified a heat shock beta protein-1. The precursor and fragment ion spectra are shown in Figure 22 and 23. The precursor ion of 912.88 with charge state of +25 was selected for fragmentation The protein was identfied with a 7 y-ions, a mass of 22,678.61 Da and with an E-value of 7.35E-05 . The theoretical mass of the protein is 22,768.5 Da. Thus, the mass difference between the theortical and observed mass was -89.89 Da. All the fragments identifed were only y-ions, indicating that the PTM was on the first 150 residues of the protein. Thus, using Sequence Gazer in ProSightPC if the initial methionine is removed (-131.19 Da) and the mass of an acetylation (+42.01 Da) is applied to theronine residue, the number of fragment ions matched go from 7 to 11 ions with additional 4 b-ions matched. In addition, there is a small mass difference between the observed and theortical mass and the E-value is 4.3E-18 making it a more confident identifcation.

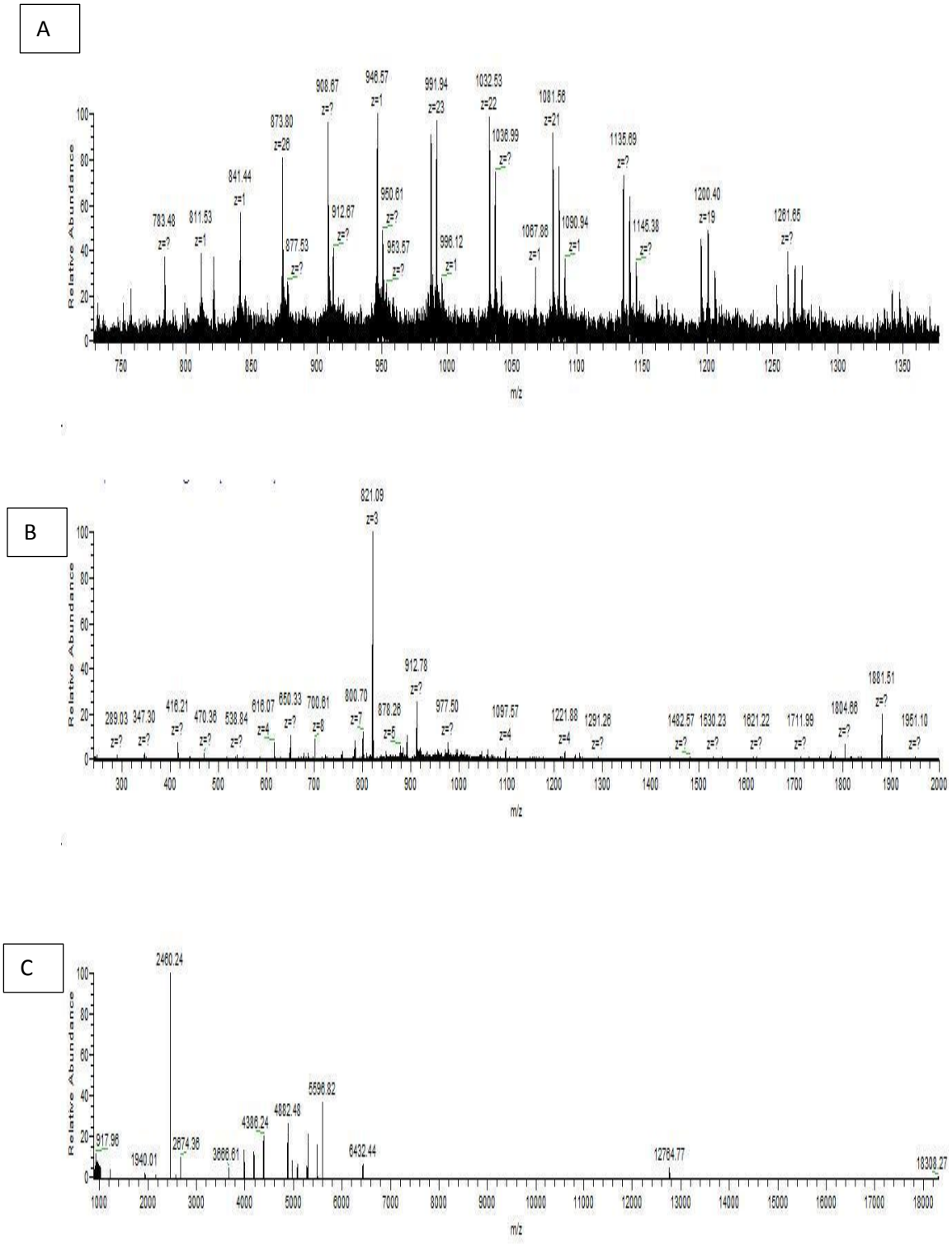

Figure 22: (A) Precursor Ion spectrum at retention time 39.08 minutes from analysis with five averaged precursor ion scans, (B) Product ions from m/z 912.88 with charge of +25, (C) Decharged product ion spectrum.

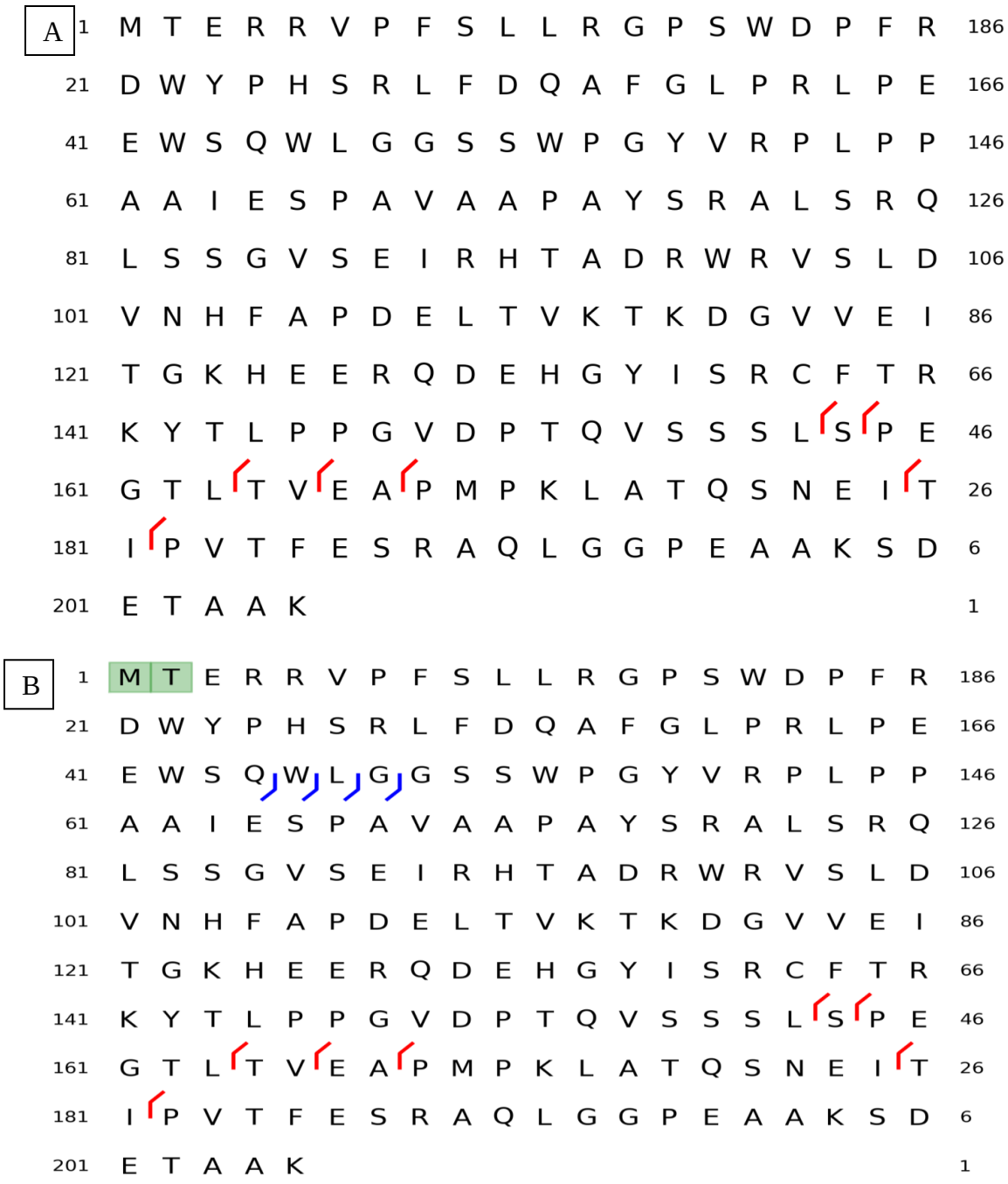

Figure 23: (A) Protein sequence and fragments observed from heat shock beta-1 protein identified by ProSightPC with an E-value of 7.35E-05 and mass difference of -89.89 Da compared to the theoretical mass, (B) PTMs localized and highlighted in green using Sequence Gazer in ProSightPC with a recalculated E-value of 4.3E-18.

In the analysis with five averaged scans for the precursor ion scan, co-elution of multiple proteins was observed at a retention time of 25.50 minutes. The precursor ion spectrum obtained at this retention time is shown in Figure 24A. The ion with 1332.61 and charge state of +9, was identified as prothymosin alpha. The observed mass for the protein was 11978.95, which is 217.05 Da less than the theortical mass of 12,196 Da. There were 18 y-ions that were matched using ProSightPC with a E-value of 9.43E-25. The mass difference corresponds to the truncation of the first two amino acid resides and matches the mass of methionine (131.19) and serine (87.08). If this change is applied to the protein using Sequence Gazer in ProSightPC, the E-value of the protein drops to 1.13E-29 and an additional 6 b-ions are matched making this a very confident match. Similarly, in the following product ion scan the ion with 881.02 and charge state of +13 was selected for fragmentation, was identified as parathymosin with an observed mass of 11,434.23 Da and E-value of 4.07E-11. The theoretical mass of the protein is 11,523.2 Da, which is 88.97 Da heavier than the theoretical mass. Using Sequence Gazer in ProSightPC, the PTM was localized in the first 40 residues by adding an acetylation to the serine (+42.01 Da) residue and removal of the initial methionine (-131.19 Da). The E-value of the protein dropped to 7.18E-18 and 7 additional b-ions were identified in addition to previously matched 3 y-ions. Identification of co-eluting proteins is feasible when averaging scans for the precursor ion spectrum at a resolution of 60,000 at 400 m/z despite the increase in time for the duty cycle.

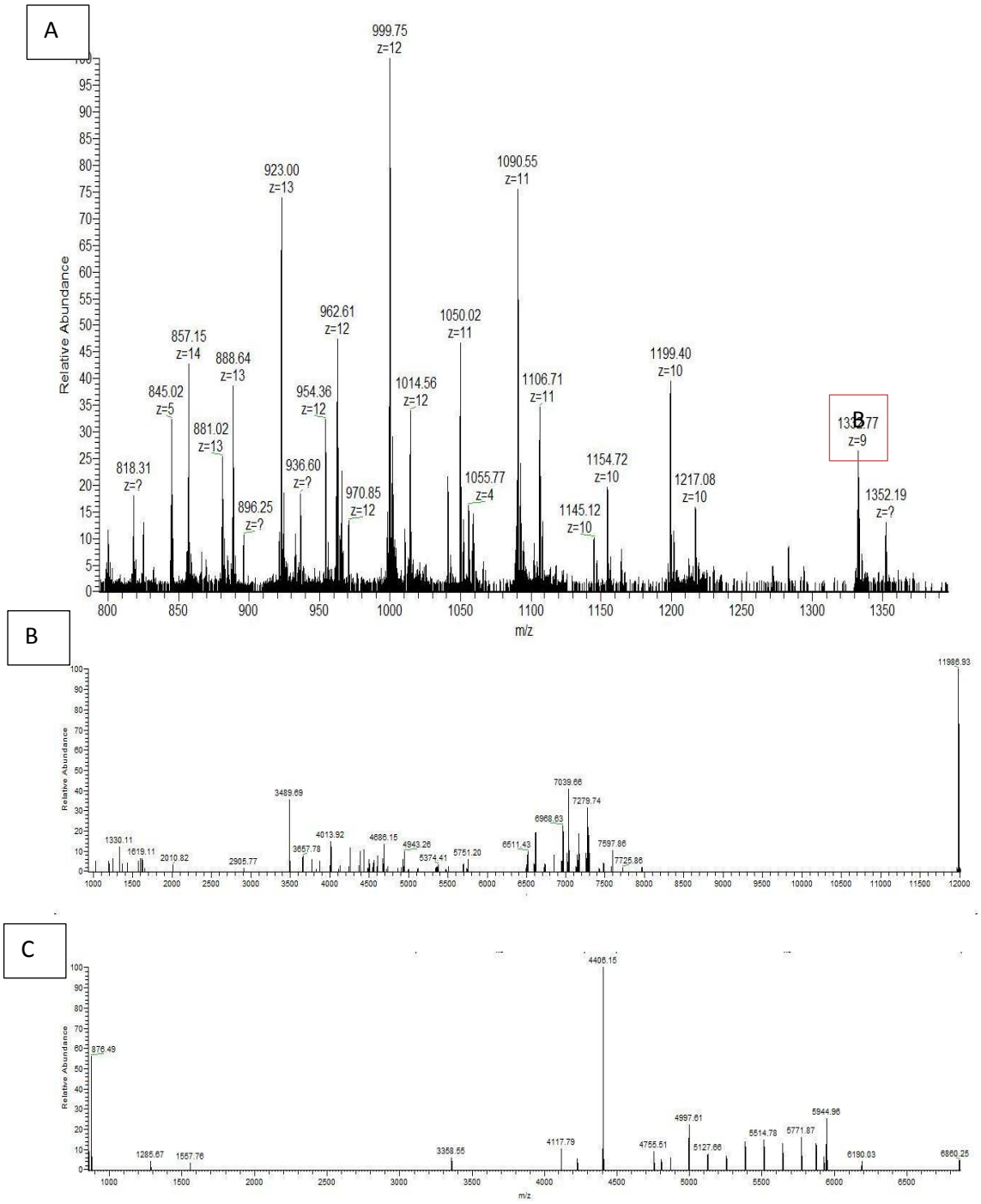

Figure 24: (A) Precursor Ion spectrum at retention time 25.50 minutes from analysis with five averaged precursor ion scans, (B) Decharged product ions from m/z 1332.61 with charge of +9, (C) Decharged product spectrum from m/z 881.02 with charge of +13.

Figure 25: (A) Protein sequence and fragments observed for prothymosin alpha with a theoretical mass of 12,196 Da. PTMs localized and highlighted in green using Sequence Gazer in ProSightPC, (B) Protein sequence and fragments observed for parathymosin with a theoretical mass of 11,434.23 Da, PTMs localized and highlighted in green using Sequence Gazer in ProSightPC.

Comparsion of CID and ETD on an LC-MS/MS Time Scale for Complex Mixtures

Material and Methods

Sample preparation of cytosolic proteins from MCF-7 cells: cytosolic proteins from MCF-7 cancer cells were prepared as described in the earlier section of this chapter

LC analysis and MS analysis for cytosolic proteins: A Shimadzu Prominence nano LC system and Autosampler was used for the LC analysis. Fifty micrograms cytosolic proteins from MCF-7 cells was loaded onto a zorbax (5 μm, 5 mm x 0.3 mm) trap column and washed with Solvent A (97.5% Water, 2.5% ACN and 0.1% formic acid) for 20 minutes at 10 μL/min. The proteins were then fractionated using an Agilent C3 column (5 μm, 150 mm x 0.1 mm) at a flow rate of 300 nL/min and a linear gradient of increasing concentration of Solvent B (97.5% ACN, 2.4% water and 0.1% formic acid) from 10% to 85% over 90 minutes. The LC was connected in line with an LTQ-Orbitrap-XL and the precursor scans were recorded in the Orbitrap at a resolution of 60,000 at 400 m/z with 5 averaged scans. The most abundant signal for each precursor ion was subjected to CID fragmentation with activation energy at 25 or ETD with reaction time of 5 ms, 10 ms or 20 ms. MS/MS spectra were recorded in the Orbitrap at a resolution of 60,000 at 400 m/z 5 averaged scans. Data dependent analysis was set to isolate precursor ions with unassigned charges and charges greater than +4. The isolation width for the precursor ions was set to 10 Da. The AGC targets were set to

1E6 for precursor scan and 1E5 for the four MS/MS scans. Dynamic exclusion was employed with a repeat count of 1, repeat duration of 15 seconds and exclusion duration of 30 seconds.

Bioinformatics: Database searching was performed using ProSightPC 2.0.[55] The precursor ion and fragment ions were deconvoluted using the THRASH[56] algorithm. The data was searched against a custom UniProt sub 30 KDa human database (consisting of the sequences of proteins with molecular weights less than 30KDa). A precursor mass tolerance was set to 2000 Da and the fragment mass tolerance was set at 15 ppm. Post-translational modifications were localized on the identifications using ΔM mode and Sequence Gazer on ProSightPC 2.0 was used to localize any mass shifts at the N- or C- terminus of the protein. Post-translational modifications and mass shifts were investigated manually using Sequence Gazer available in the software. The proteins identified were automatically assigned an E-value, identifications with E-value lower than 10E-4 were considered as a strong identification.

Results and Discussion

CID and ETD fragmentation conditions were evaluated based on the number of unique proteins identified, the mass range of the proteins identified and the number of fragments matched using a custom database and ProSightPC. The number of unique proteins identified using ETD was 2 proteins, 6 proteins, and 8 proteins with a reaction time condition of 5 ms, 10 ms, and 20 ms, respectively. The total number of unique proteins identified using CID fragmentation was 14

proteins. The lists of proteins identified from the four analyses are provided in the Appendix Tables 5-8.

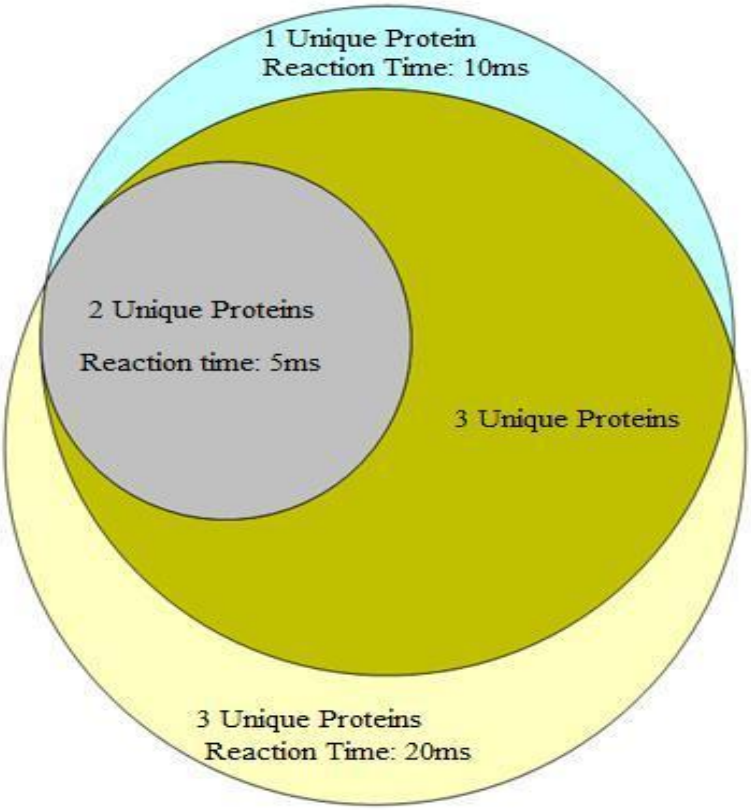

Figure 26: Unique protein identification and overlapping proteins from three ETD fragmentation conditions.

The two proteins identified using ETD fragmentation with a reaction time of 5 ms were identified in the analysis with a reaction time of 10 ms as well as 20 ms. There were 3 proteins overlapping between the analyses with reaction time of 10 ms and 20 ms. The 3 proteins identified in the analysis of the cytoslic proteins with reaction time of 20 ms were also sampled by the mass spectrometer in the other two analyses; but due to lower reaction times, fewer fragments were formed.

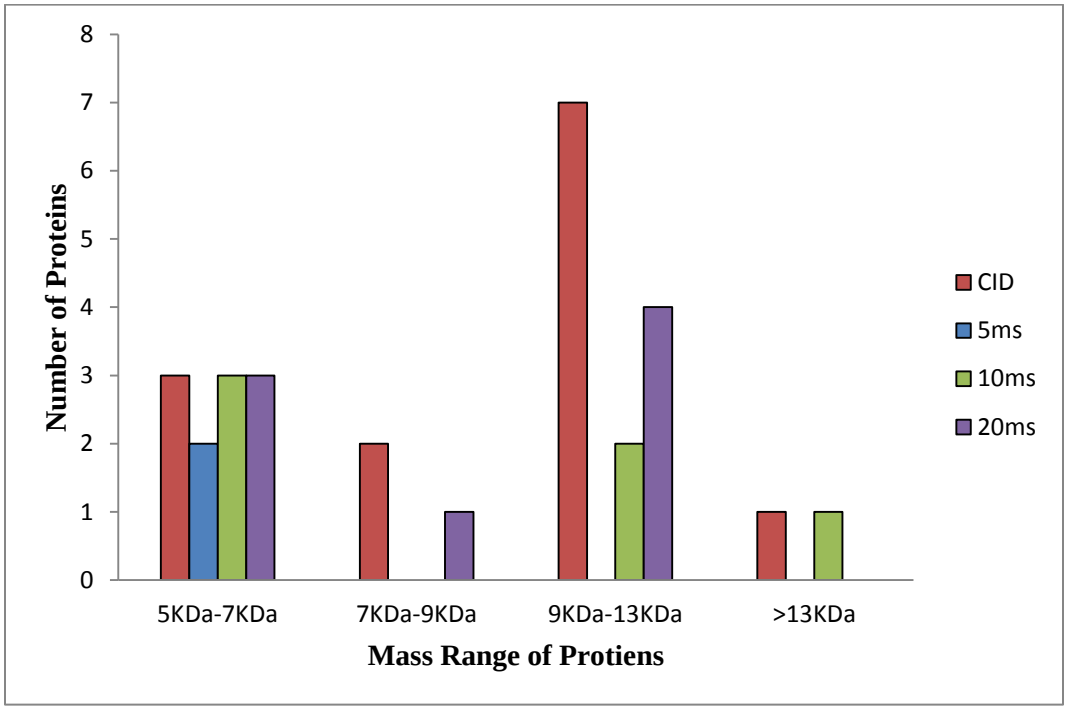

Figure 27: Molecular weights of proteins identified from MCF-7 cytosol on an LC-MS/MS time scale with CID and ETD fragmentation methods.

The mass range of proteins identified using CID is 5.0 KDa-14.7 KDa. Analysis with ETD allowed unique protein identification from 5.0 KDa to 6.6 KDa and 5.0 KDa to 14.7 KDa for analysis with ETD reaction time of 5 and 10 ms, respectively. Whereas, the study using ETD reaction time 20 ms allowed for unique protein identification from 5.0 KDa to 12.2 KDa. The 8 proteins identified from ETD with a reaction time of 20 ms were also identified using CID.

The protein thymosin beta-4 was confidently identified in all four analyses. The protein sequence and the fragment ions observed for the four analyses are shown in Figure 28. The observed mass of the protein in all four cases was 4072.04 Da and the theortical mass of the protein is 5049.52 Da thus, the mass difference

between the theoretical and observed mass was -89.89 Da. Since all the fragments identifed were y-ions for CID fragmentation or z-ions for ETD fragmentation, the identification of y-ions indicated that the modification was on the N-terminus of the protein. Thus, using Sequence Gazer in ProSightPC if the initial methionine (-131.19 Da) is removed and an acetylation (+42.01 Da) is added to serine residue, the number of fragment ions matched increased in all four cases and a lower E-value was observed increasing the confidence of the identification. Using CID as a fragmentation method, 23 b-ions and 13 y-ions were matched with thymosin beta-4 with a E-value of 6.71E-58. The number of fragments for ETD fragmention with a reaction time of 5 ms is 12 c-ions and 5-zion with a E-value of 9.55E-15. Whereas, 15 c-ions and 7 z-ions with a E-value of 1.39E-30 were matched utlizing ETD fragmentation and reaction time of 10 ms and 13 c-ions and 11 z-ion with an E-value of 1.17E-30 with a reaction time of 20 ms.

Although all four analyses identified the protein confidently, the number of fragments matched using CID was significantly higher. In addition, many proteins identified with CID fragmentation method were not confidently identified with ETD fragmentation method. The successful fragmentation using ETD method was observed to be very protein dependent, some proteins fragmented more successfully compared to others. The utlization of ETD fragmentation technique did not show any advantage over CID for analysis of intact proteins in a complex mixture.

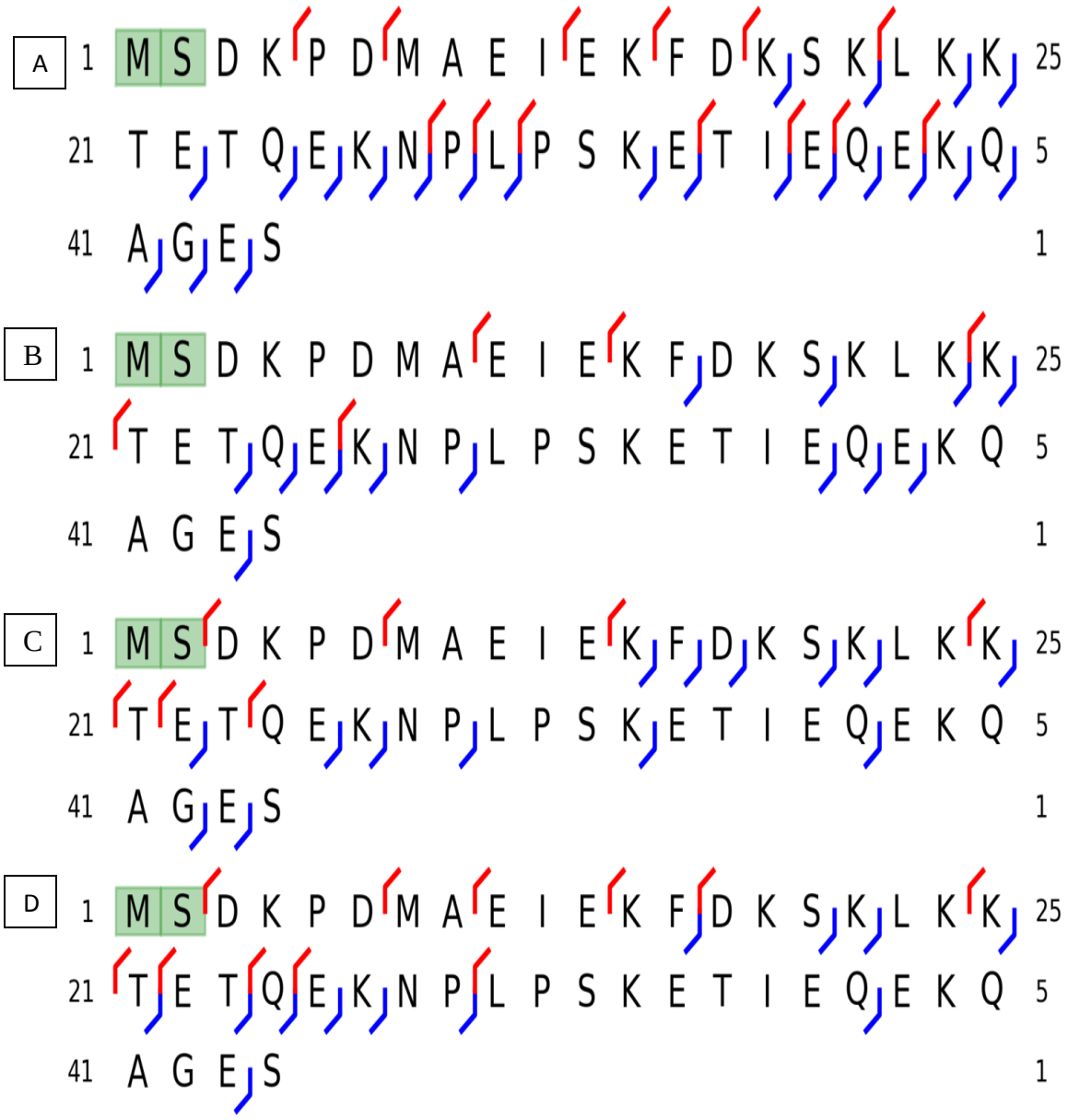

Figure 28: Protein sequence and fragments observed for thymosin beta-4. modifications are highlighted in green were localized using Sequence Gazer in ProSightPC, (A) CID fragmentation, (B) ETD fragmentation with reaction time 5 ms, (C); ETD fragmentation with reaction time of 10 ms, (D) ETD fragmentation with reaction time of 20 ms.

Conclusion

Optimized fragmentation conditions and averaged precursor scans play an important role in confidently identifying intact proteins from complex mixtures on an LC-MS/MS time scale. In the infusion experiment for intact proteins, more fragments were matched using lower activation energy with CID and HCD for intact proteins infused directly in the mass spectrometer. As the mass of the protein increased, internal fragments were observed using the HCD fragmentation technique due to which fewer fragment ions were matched. For ETD fragmentation, the number of fragments matched increased as the mass of the protein increased. In addition, ETD fragmentation favored higher charge states, while, CID and HCD fragmentation techniques favored lower charge states. For higher molecular weight proteins, the resolution and S/N for the precursor ion increases when scans are averaged. The mass range of the proteins increased when five averaged scans were acquired for the precursor ion at a resolution of 60,000. In the comparison of CID and ETD, a higher number of proteins were identified using CID. Fragmentation achieved using ETD was protein dependent.

Chapter 4: Comparative Study of Fractionation Methods for Top-down Analysis of Complex Protein Mixtures

Introduction

The advancement of mass spectrometric instrumentation and fragmentation methods has played an important role in the routine LC-MS/MS based analysis of mixtures of intact proteins. Long duty cycle and co-elution of proteins require front end fractionation methods are required to reduce the complexity of the sample. The reduction of complexity of the sample aids in increasing the number of protein identifications and the mass range accessible in a given analysis. Proteins are usually fractionated using gel electrophoresis or chromatographic methods. Traditionally, intact proteins are separated using electrophoretic methods such as 2D gel electrophoresis or SDS-PAGE. The limitations of these methods are laborious nature, biased protein identification, and poor extraction of proteins from the gels.[62, 63] Due to these reasons, separations in which the sample is retained in solution are preferred fractionation methods for top-down proteomics. The three in-solution fractionation methods popularly used for intact proteins analysis are the Agilent OFFGEL system, molecular weight cut- filters (MWCOF) and the Protein Discovery GELFrEE system. The fractionation using the OFFGEL system is based on the isoelectric point of the protein; whereas, fractionation is based on the molecular weight of the proteins on the GELFrEE system as well as with MWCOF.

Liquid chromatographic methods are an alternative to electrophoretic fractionation of complex mixtures where fractionation is based on physiochemical properties of the proteins. Usually, RP-LC is utilized for analysis as the solvents used in this method are compatible with ESI source; thus, RP chromatographic columns can be connected in line with the mass spectrometer. Employing orthogonal chromatographic techniques, bRP-aRP-LC increases the peak capacity and dynamic range of the analysis. The number of proteins identified in bRP-aRP-LC analysis increased almost 2 fold as demonstrated in chapter 2.

The objective of the experiment presented in this chapter is to compare the effectiveness of GELFrEE electrophoresis, orthogonal bRP-aRP-LC, and molecular weight cut-off (MWCO) filters. The evaluation is based on the number of proteins identified and their molecular masses. *E. coli* lysate is used as the model complex mixture.

Material and Methods

***E. coli* Lysate**: Forty milligrams (mg) *E. coli* (K-12 strain) was lysed by vortexing the lyophilized cells in 500 µl of 10% formic acid. The protein lysate is centrifuged for 10 minutes at 14,000 rotations per minute (RPM). Proteins were precipitated with cold acetone through 45 minutes at -20 ºC and precipitated proteins were collected by centrifugation at 10,000 g for 20 minutes. The resultant pellet was re-suspended in 500 µL of Solvent A (97.5% water, 2.4% acetonitrile, 0.1% formic acid). The RC/DC assay was used to determine the protein concentration of the sample.

Fractionation of *E. coli* Lysate

1) High Pressure bRP-LC analysis: *E. coli* lysate (300µg) was first fractionated on an Xbridge column using solvents with high pH and a flow rate of 50 µL/min. The two solvents used for the first dimension were: 50 mM ammonium hydroxide in water for solvent A and 50 mM ammonium hydroxide in water and ACN (2:8) for solvent B.[54] Reversed phase chromatography was carried out on a Thermo Accela LC system with a step gradient of increasing concentration of solvent B from 10% to 90% over 60 minutes. The samples were detected with a SPD-10A UV spectrometer at 214 nm. A total of five fractions were collected, lyophilized and re-suspended in solvent A (97.5% water, 2.4% ACN, 0.1% formic acid) for the second dimension. The second dimension was the low pH chromatography interfaced to the LTQ-Orbitrap mass spectrometer. This last separation was the same for all three of the fractionations evaluated here.

2) GELFrEE fractionation: Thirty microliters of acetate buffer provided by the manufacturer was added to 300 µg of *E. coli* lysate, followed by reduction in 53 mM DTT at 50 ºC for 10 minutes. The lysate was loaded onto a 12% Tris-acetate polyacrylamide cartridge (Expedeon, San Diego, CA) and fractionated using manufacturer's instruction by applying a voltage that increased from 50V to 85V over two hours. A total of 12 fractions were collected. Each fraction was precipitated using $CHCl_3$/MeOH/H_2O (4:1:3)[61] and re-suspended in solvent A (97.5% water, 2.4% ACN, 0.1% formic acid) for LC-MS/MS analysis. The fractionation was visualized by SDS-PAGE. Ten microliter aliquots from each fraction were electrophoresed on an 8-16% Tris-HCl polyacrylamide gel (BioRad,

Hercules, CA). The gel was run at 200 V using a BioRad apparatus and stained using silver stain.

3) MWCOF fractionation: Molecular weight cut-off filters (3 KDa and 30KDa) (Millipore, Billerica, MA) were equilibrated using solvent A (97.5% water, 2.4% ACN, 0.1% formic acid). *E. coli* lysate was centrifuged using a 30KDa membrane filter at 14,000 g for 12 minutes. The filtrate (lysate < 30KDa) was then centrifuged using a 3 KDa filter at 14,000 g for 30 minutes. The retentate was aspirated 30x times with 300 µL of solvent A to increase recovery of the sample. The volume was reduced to approximately 50 µL by lyophilizing the sample.

LC-MS/MS analysis: Seventy-five micrograms of *E. coli* lysate for the control sample or each of the fractions from the three methods was loaded onto a Zorbax (5 µm, 5 mm x 0.3 mm) trap column and desalted with solvent A (97.5% Water, 2.5% ACN and 0.1% formic acid) for 10 minutes at 15 µL/min. The proteins were then fractionated using an Agilent C3 column (5 µm, 150 mm x 0.1 mm) at a flow rate of 300 nL/min and a linear gradient of increasing concentration of solvent B (97.5% ACN, 2.4% water and 0.1% formic acid) from 0% to 85% over 180 minutes. The LC was connected in line with a LTQ-Orbitrap-XL and the precursor scans were recorded in the Orbitrap at a resolution of 60,000 at 400 m/z. The most abundant precursor scans were subjected to CID fragmentation in the LTQ ion trap with activation energy at 25. MS/MS spectra were recorded in the Orbitrap at a resolution of 60,000 at 400 m/z. Data dependent analysis was set to isolate precursor ions with either unassigned charges or charges greater than +4. Dynamic exclusion was employed with a repeat count of 2, repeat duration of 240

seconds and exclusion duration of 300 seconds. The isolation width for the precursor ions was set to 10 Da. The automatic gain control (AGC) targets were set to 1E6 for precursor scans and 5E5 for MS/MS scans.

Bioinformatics: Database searches were performed using ProSightPC 2.0[55] against a custom UniProt *E. coli* database consisting of proteins with molecular weight less than 30KDa. The THRASH[56] algorithm was used to deconvolute both the precursor and fragment ions. Precursor mass tolerance was set to 2500 Da and fragment ions mass tolerance was set to 15 ppm. The ΔM mode on ProSightPC 2.0 was used to evaluate mass shifts. Post-translational modifications and mass shifts were investigated manually using Sequence Gazer available in the software. The proteins identified were automatically assigned an E-value. Identifications with E-value lower than 10E-4 were considered as a strong identification.

Results and Discussion

Fractionation methods for analysis of intact proteins from *E. coli* lysate were evaluated based on the number of proteins identified and mass range of the proteins identified using each of the three methods. The GELFrEE and bRP-aRP-LC methods allow a higher amount of starting material because of the number of fractions collected in these analyses. It is easier to get the proteins in-solution using GELFrEE fractionation due to the detergent present in the buffers. Fractionation using GELFrEE is analogous to SDS-PAGE; however the sample is maintained in solution. To visualize the fractionation, a SDS-PAGE of aliquots from the 12 fractions was run on a polyacrylamide gel visualized with silver stain. The molecular weights of the proteins from the GELFrEE analysis increases in every fraction. The image of the gel is shown in Figure 29. It shows the narrow mass range of proteins in each fraction.

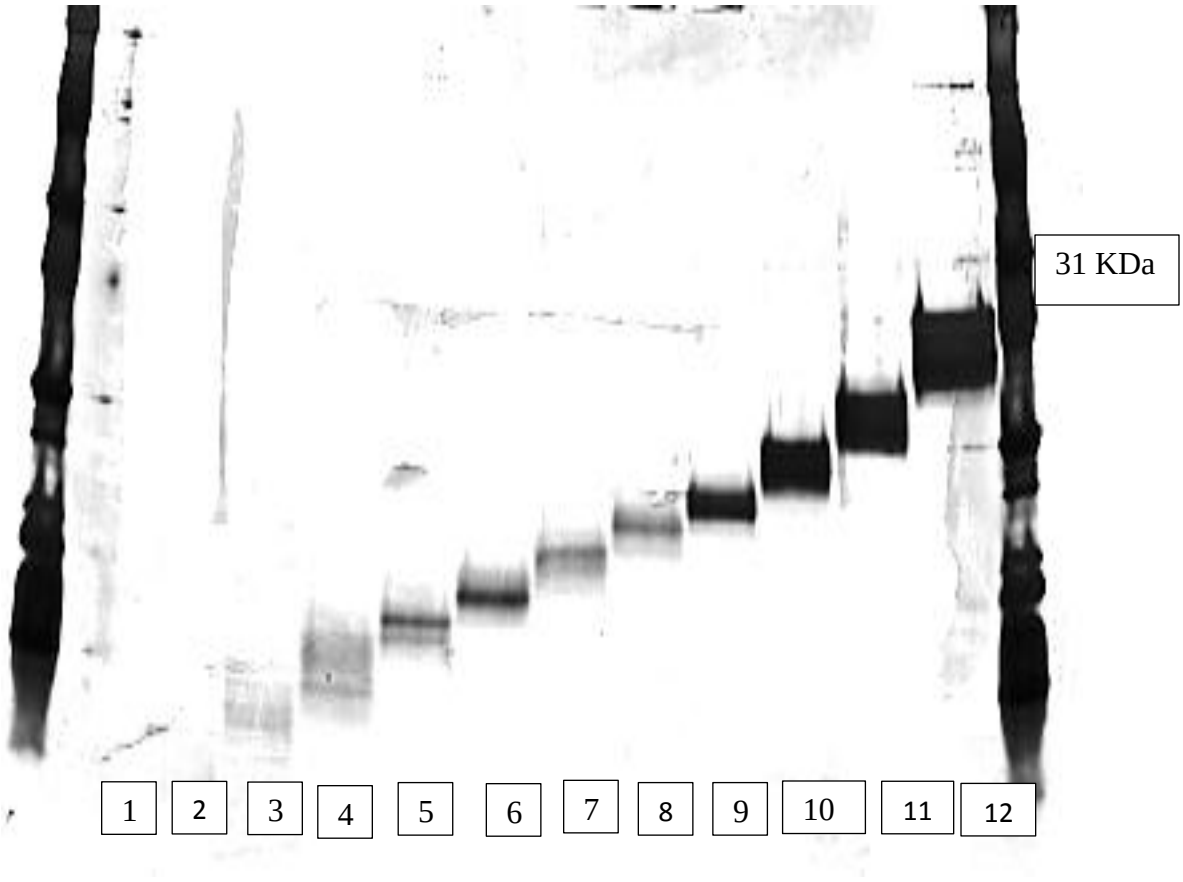

Figure 29: Silver stained 1D gel of the 12 fractions collected from GELFrEE analysis.

Twenty-one unique proteins were identified from analysis of a control sample injected directly into the LC-MS/MS without prior fractionation. The number of unique proteins identified when MWCOF, GELFrEE, or bRP-aRP-LC was implemented as a preliminary fractionation method is 19, 49, and 32, respectively. The lists of proteins identified from each of the four analyses are provided in the Appendix Tables 9-12.

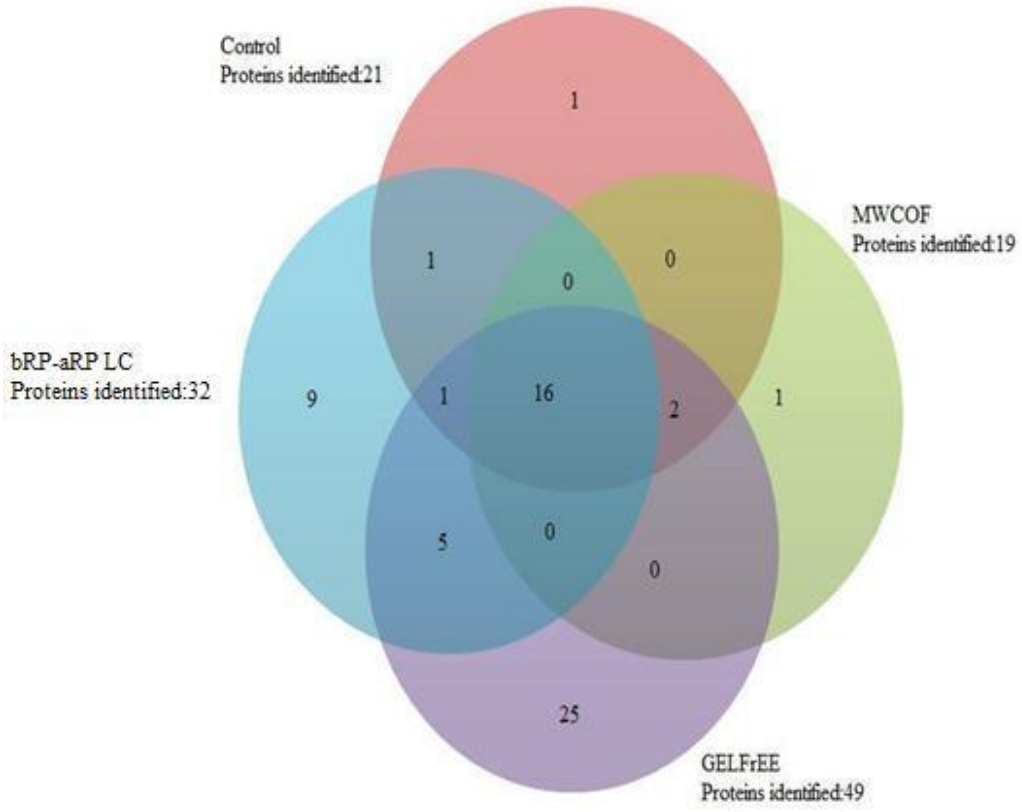

Figure 30: Unique protein identifications and overlapping proteins from *E. coli* lysate for all three fractionation methods.

As summarized in Figure 30, there were 16 proteins identified in all four analyses. The least number of proteins identified was from the MWCOF method. All but one protein identified from this analysis was identified in other fractionation methods. I suggest the low number of identifications is due to protein adsorption on the membrane. The number of proteins identified uniquely in the GELFrEE analysis was 25, while the number of proteins unique to the orthogonal bRP-aRP-LC was 9. Overall, the experiment shows the advantage of fractionating complex mixtures as the number of proteins identified increased 2 fold in the

GELFrEE analysis and 1.5 fold in the bRP-aRP-LC analysis in comparison to the control.

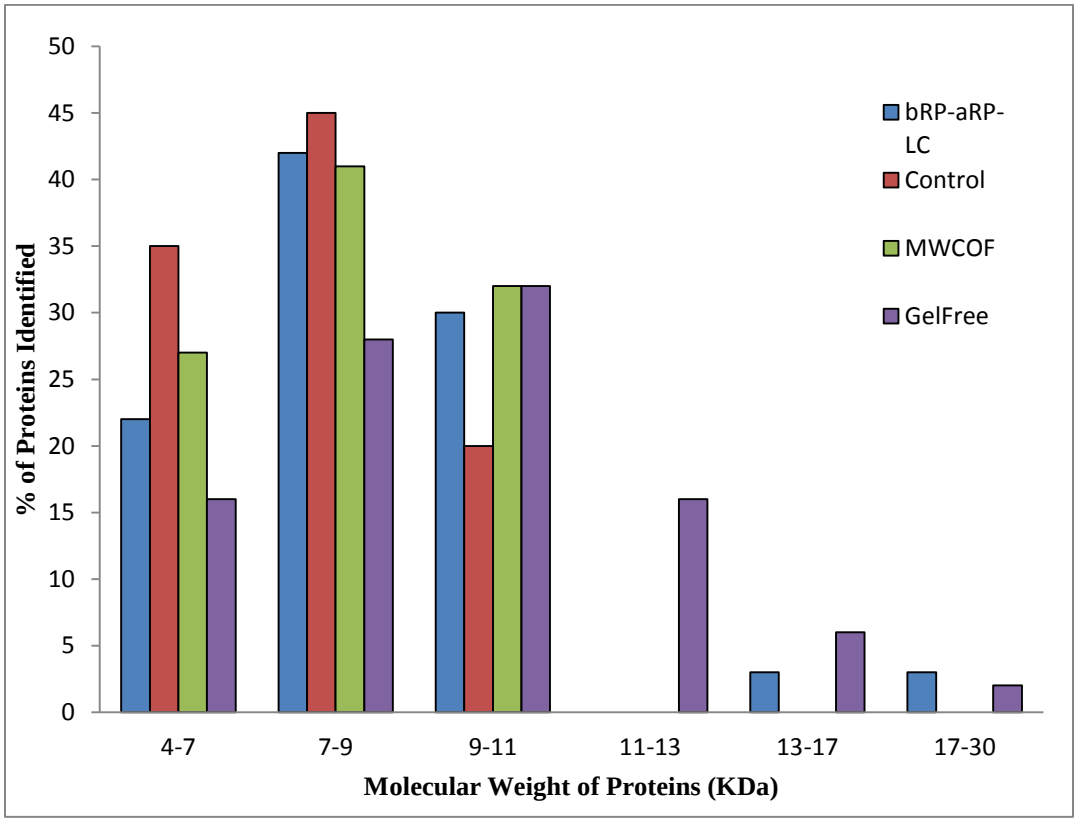

Figure 31: Molecular weight of proteins identified from *E. coli* lysate with all three fractionation methods.

The mass range of proteins identified is shown in Figure 31. The control sample includes proteins from 4.3 KDa to 10.7 KDa as well as 4.3 KDa to 9.6 KDa for the MWCOF method. Analysis with orthogonal RP-RP HPLC allowed the identification of unique proteins between 4.3 KDa and 17.1 KDa; whereas, the study using GELFrEE method identified proteins with a mass range of 4.2 KDa to 20.8 KDa. Twenty five percent of the proteins weighed more than 11 KDa in the GELFrEE analysis. Therefore, in addition to the increasing the number of

Identifications, the GELFrEE fractionation method also enabled increasing the mass range of proteins identified.

	bRP-aRP-LC Fraction 1	bRP-aRP-LC Fraction 3	bRP-aRP-LC Fraction 4	bRP-aRP-LC Fraction 5
bRP-aRP-LC Fraction 1 (19)				
bRP-aRP-LC Fraction 3 (13)	10			
bRP-aRP-LC Fraction 4 (14)	7	8		
bRP-aRP-LC Fraction 5 (16)	8	6	8	
bRP-aRP-LC Fraction 6 (15)	8	6	9	13

Table 2: Unique protein identifications between bRP-aRP-LC fractions for analysis of *E. coli* lysate.

Due to the low resolving power of fractionation methods, a given protein may be identified in more than one fraction. In the analysis using bRP-aRP-LC, a total of 70 proteins were identified from all five fractions. In Table 2, the total number of proteins identified in each fraction is indicated in parenthesis, and the number of overlapping proteins is indicated in the corresponding boxes. Thus, fifty-eight percent of the identified proteins were present in more than one fraction

	GF fraction 4	GF fraction 5	GF fraction 6	GF fraction 7	GF fraction 8	GF fraction 9	GF fraction 10	GF fraction 11
GF Fraction 4(24)								
GF fraction 5 (24)	21							
GF fraction 6 (26)	19	21						
GF fraction 7 (35)	23	23	23					
GF fraction 8 (29)	21	30	21	25				
GF fraction 9 (18)	15	14	15	15	15			
GF fraction 10 (23)	15	13	15	18	17	15		
GF fraction 11 (22)	16	14	18	19	18	15	17	
GF fraction 12 (26)	6	17	18	20	20	14	18	16

Table 3: Unique protein identifications between fractions using GELFrEE for analysis of *E. coli* lysate.

The number of proteins identified in each fraction from the GELFrEE is indicated in parenthesis in Table 3, and the number of overlapping proteins between fractions is indicated in the corresponding box. A total of 227 proteins were identified in the 12 fractions and seventy-eight percent of the identified proteins were present in more than one fraction. While there is overlap between

fractions for both bRP-aRP-LC and GELFrEE, there is a higher overlap of proteins between fractions for GELFrEE.

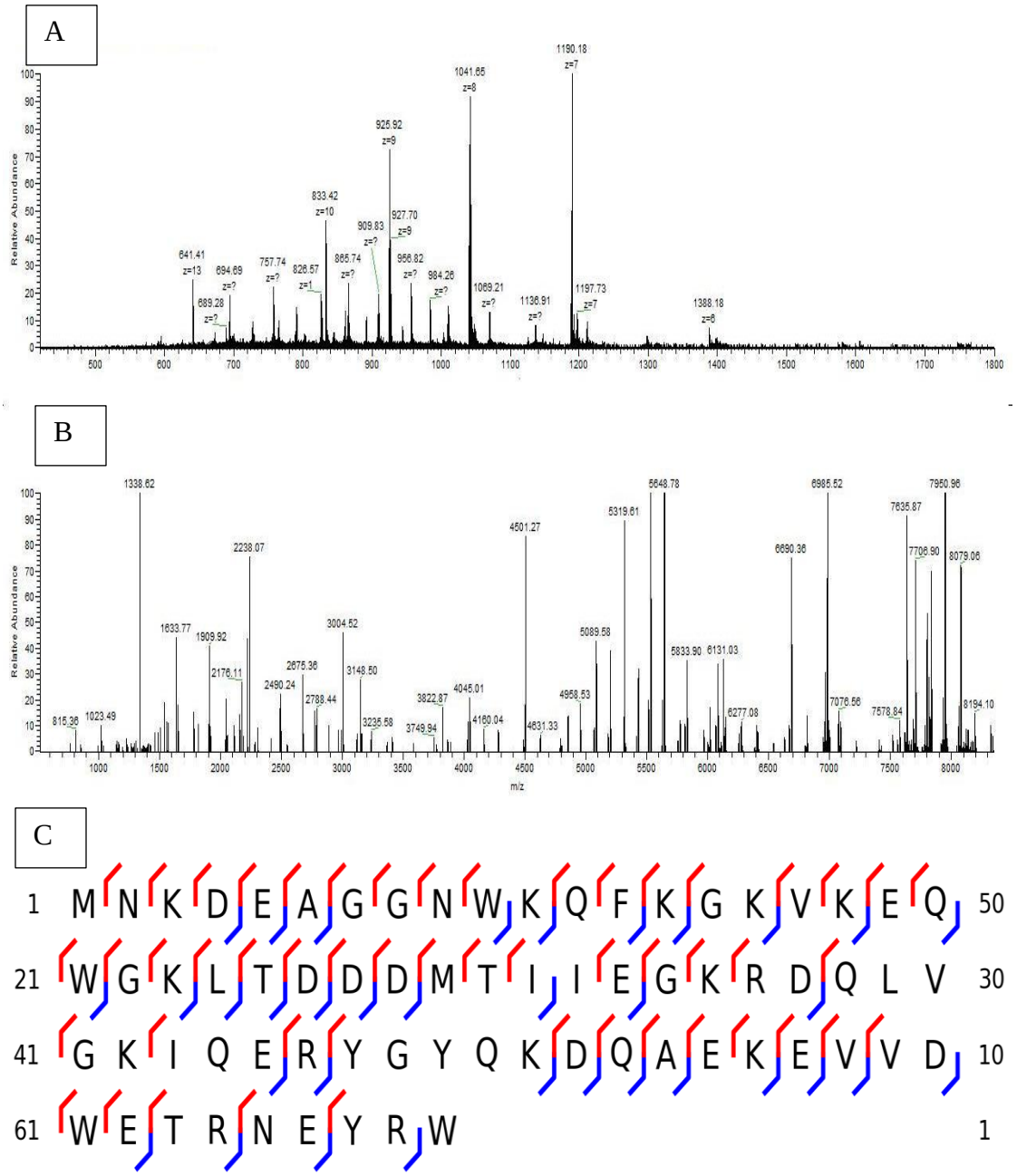

Figure 32: (A) Precursor Ion spectrum at retention time 107.36 minutes from fraction 10 of GELFrEE analysis, (B) Decharged product ions from m/z 1170.36 with charge of +7, (C) Protein sequence and fragments observed for UPF0337 protein yjbJ with a theoretical mass of 8,320.11 Da.

UPF0337 protein yjbJ was identified with the lowest (best) E-value from all four analyses. In fraction 10 of the GELFrEE analysis, the protein eluted at 107.36 minutes from the C3 HPLC column. The precursor ion spectrum with different charge states of the protein is shown in Figure 32A. The precursor ion of 1170.36 m/z with charge state of +7 was selected for fragmentation. The decharged spectrum of the product ions is shown in Figure 32B. The E-value of the protein identified was 1.05E-53 with 36 b-ions and 51 y-ions. The experimental mass of the protein was 8320.1 Da, which matched the theoretical mass of the protein predicted from the gene sequence.

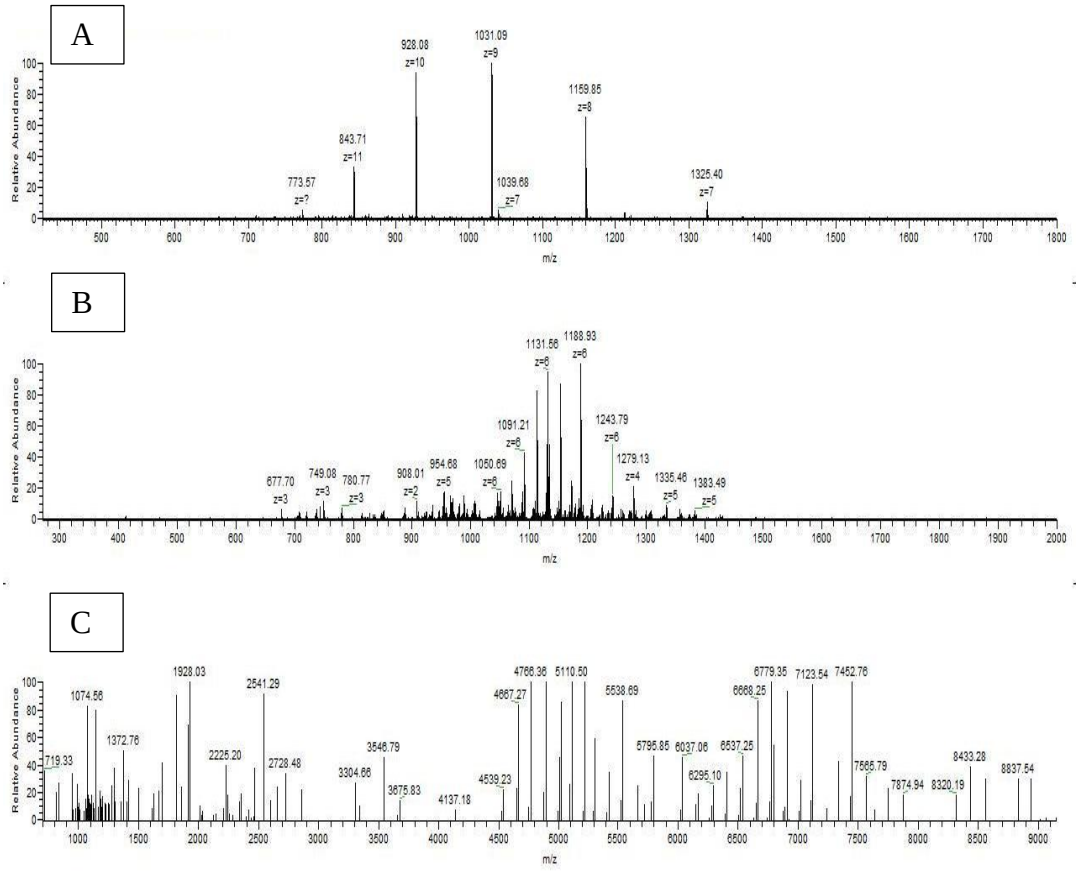

Figure 33 (A): Precursor ion spectrum at retention time 153.06 from fraction 10 of GELFrEE analysis for cell division protein ZapB (B) Product ions from m/z 1031.09 with charge of +9 (C) Decharged product ion spectrum.

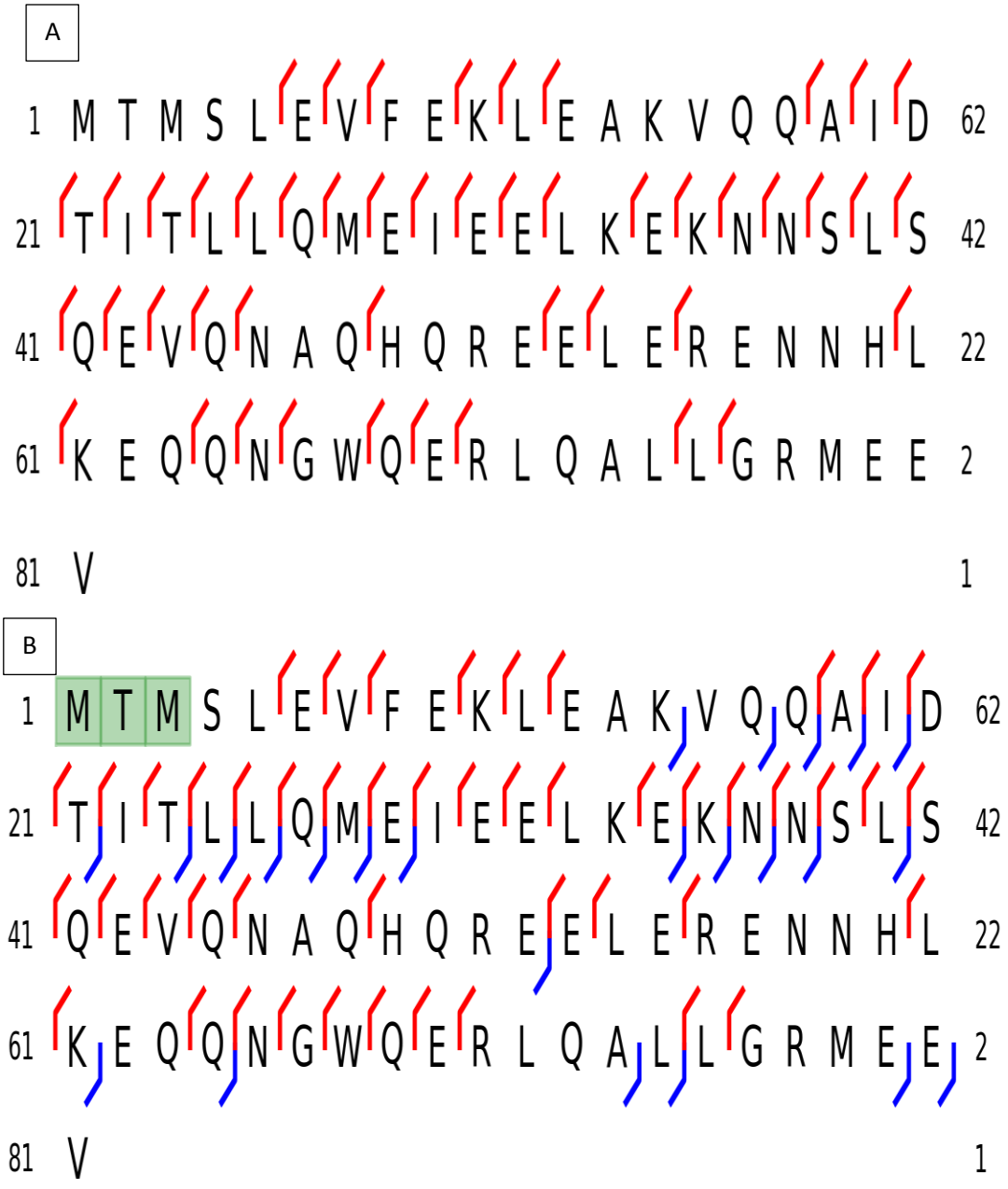

Figure 34: (A) Protein sequence and fragments observed from cell division protein ZapB identified with ProSightPC and E-value of 7.05E-45 and mass difference of 363 Da compared to the theoretical mass, (B) Modifications are localized and highlighted in green using Sequence Gazer in ProSightPC with a recalculated E-value of 3.49E-75.

Cell division protein ZapB was identified in fraction 6 from the GELFrEE analysis of the *E. coli* lysate. The precursor ion spectrum at 153.06 minutes and the product ion spectrum for ions of m/z 1031.09 with a charge state of +9 are shown in Figure 33. The protein was identfied with a 48 y-ions and a mass of 9265.72 Da and with an E-value of 7.05E-45 A mass difference of -363 Da existed between the observed mass and the theortical mass of 9628.81 Da. This mass difference corresponds to the mass of the first three amino acid residues, two methionine (131.09 Da each) and threonine (101 Da). If this change is applied to the protein using Sequence Gazer, the E-value drops to 3.49E-75. The number of matched fragments increases, with 24 matched b-ions and 48 y-ions, making this a stronger identification.

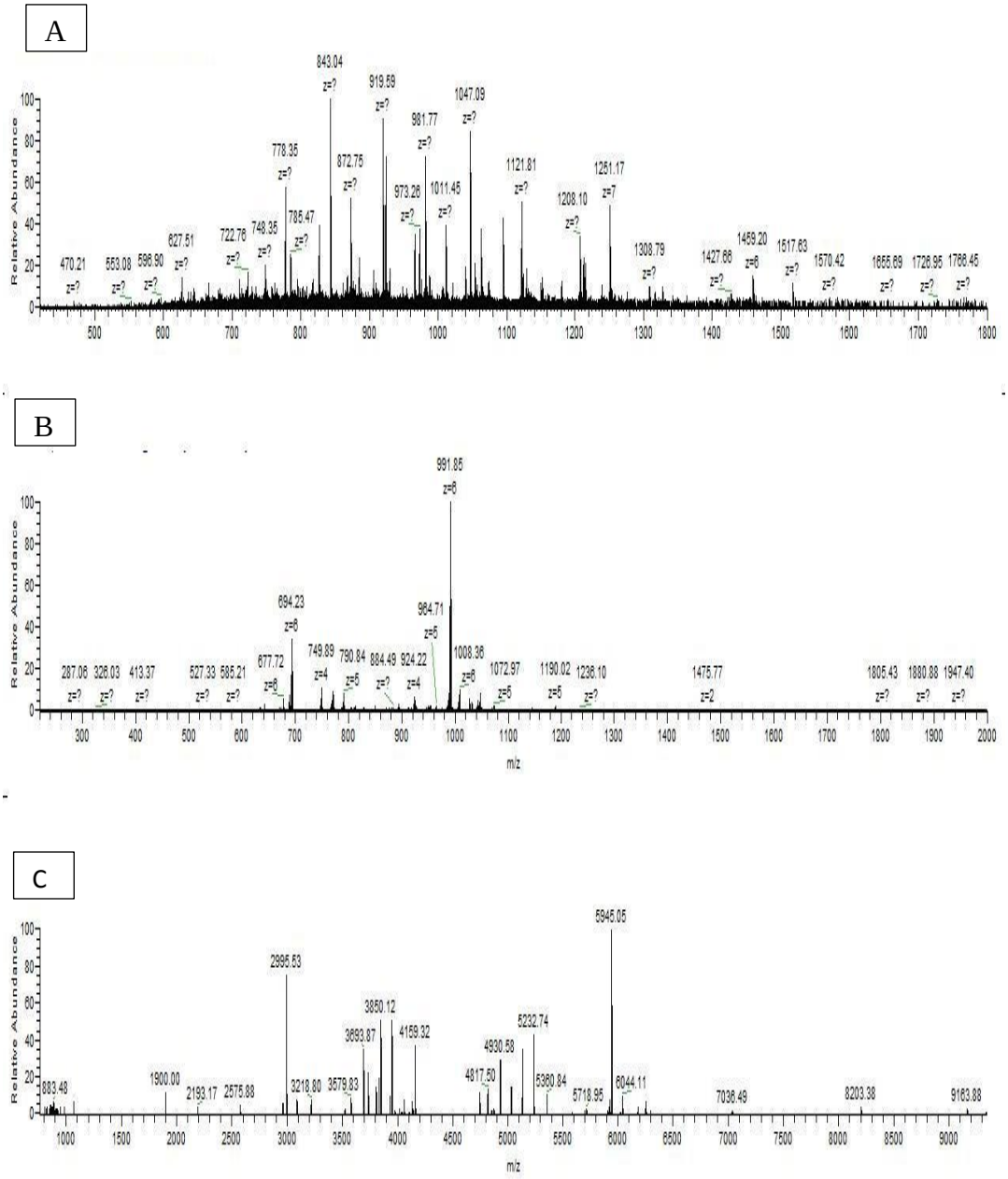

Figure 35: (A) Precursor ion spectrum at retention time 131.45 from fraction 1 of bRP-aRP-LC analysis of DNA-directed RNA polymerase subunit omega (B) Product ions from m/z 843.13 with charge of +12 (C) Decharged product ion spectrum.

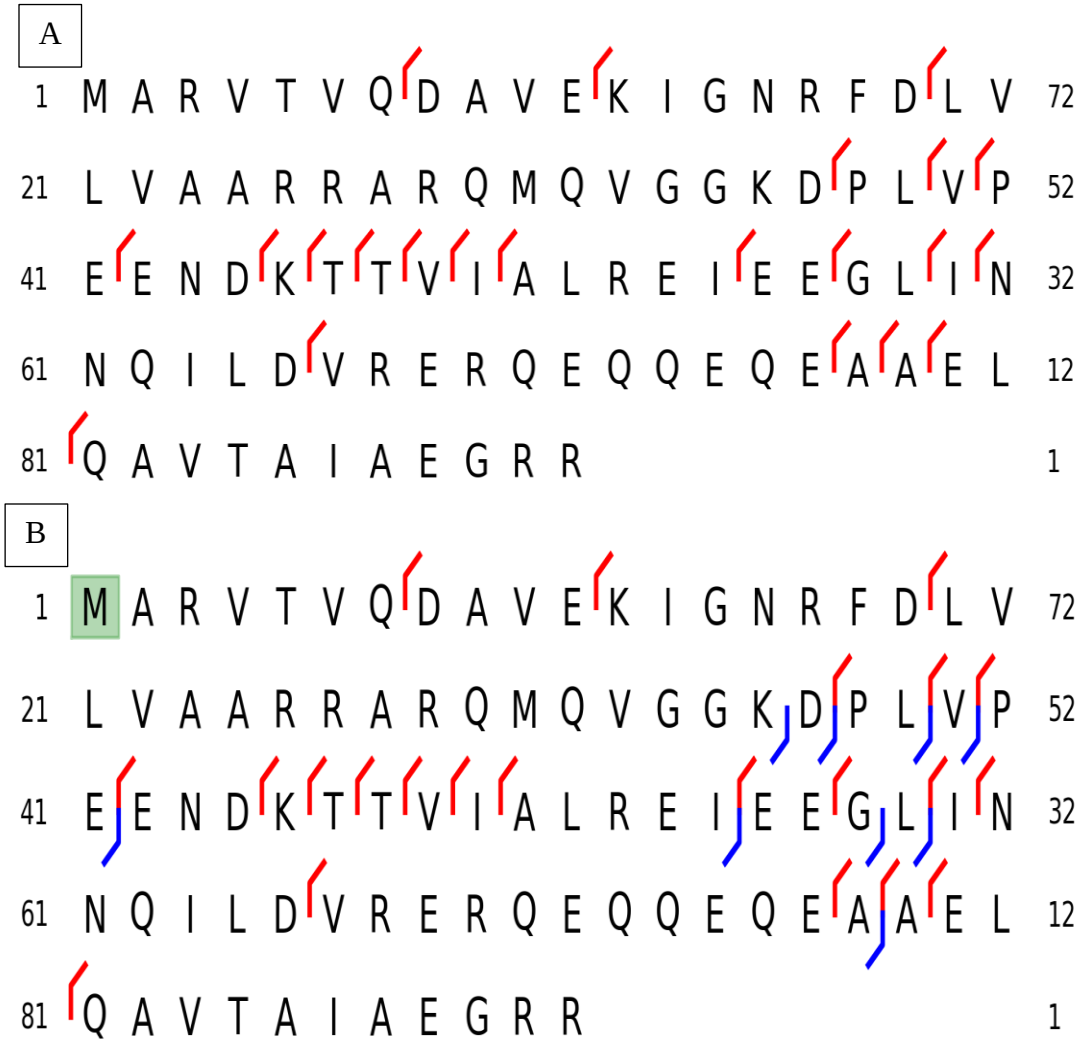

Figure 36: (A) Protein sequence and fragments observed from DNA-directed RNA polymerase subunit omega identified by ProSightPC with an E-value of 5.66E-21 and mass difference of -131 Da compared to the theoretical mass, (B) PTMs localized and highlighted in green using Sequence Gazer in ProSightPC with a recalculated E-value of 2.5E-34.

While there were many overlapping proteins between samples, some proteins identified were unique to a given fractionation method. DNA-directed RNA polymerase subunit was identified only from bRP-aRP-LC analysis. The protein eluted at 131.45 minutes in fraction 1. The precursor ion spectrum obtained

at this retention time is shown in Figure 35A. The ion with m/z 843.12 with a charge state of +12 was fragmented, the product ions and the decharged spectrum are shown in Figure 35B and Figure 35C, respectively. The observed mass for this protein was 1099.33 Da which is 131 Da less than the theoretical mass of 10230.4 Da. Twenty–three y-ions was assigned using ProSightPC. The E-value was 5.66E-21 for the protein identification. The mass difference between the theoretical and observed mass was -131 Da, which is assigned to the truncation of the initial methionine. When this change in mass was applied to the N-terminus of the protein, an addition of 9 b-ions were matched and the recalculated E-value was 2.5E-35 making this a stronger identification.

Conclusion

Fractionation of complex mixtures reduces the sample complexity and increases the number of identifications and mass range of the proteins. Comparing the three fractionation methods, GELFrEE resulted in the most identifications and the highest mass range. In the case of MWCOF method, the low recovery of proteins from the membrane leads to fewer protein identifications. Seventy-eight percent of proteins were identified in more than one fraction in the case of GELFrEE method and fifty-eight percent in the case of the bRP-aRP-LC method. Both the GELFrEE and bRP-aRP-LC workflows allow larger number of sample to be fractionated, up to 500 µg. Additionally, it is easier to dissolve proteins using GELFrEE fractionation due the detergents present in the buffer. A total of 61unique proteins were identified in the control and the three pre-fractionation workflows. Of these, only 12 proteins were found not to be modified and 48 proteins were found to be modified

Chapter 5: Top-Down Analysis of Intact proteins of Extracellular Vesicles Shed by Myeloid-Derived Suppressor Cells

Introduction

Extracellular vesicles (EVs) are small membranous vesicles released from cells under normal and pathological conditions and range from 10 nm-1000 nm in diameter. The EVs have many biological functions, including immune response, antigen presentation, and intercellular communication.[64] EVs are of interest because of the specific mechanisms whereby they are actively released from cells, their involvement in cell-to-cell signaling and their utility as markers of disease.[65] The composition of these vesicles reflects the origin parent cell and usually contains membrane, nuclear and cytosolic proteins.[64] Myeloid-derived suppressor cells (MDSC) accumulate in tumor microenvironment during tumor growth in lymphoid organs, blood and tumor tissue.[66] It has been shown experimentally that the MDSC cells contribute to suppression of T-cell activation. MDSC suppress both innate and adaptive immune responses to tumor growth and prevent effective immunotherapy.[67] They shed extracellular vesicles, which we hypothesize act as intercellular communicators in the tumor microenvironment.

In this experiment we apply the developed effective top-down workflow, including optimized fractionation, front end chromatographic separation and activation, discussed in previous chapters for the analysis of intact proteins in the

EVs shed by MDSC. EVs are shed from MDSC cells that are isolated from the blood of BALB/c mice that are carrying 4T1/IL-1β mammary carcinoma cells. Top-down analysis of intact proteins from these extracellular vesicles was undertaken to identify low mass protein cargo and to characterize post-translational modifications. Current understanding of EV subtypes, biogenesis, cargo and mechanisms of transporting proteins is incomplete. Thus, intact protein analysis of the EVs is useful for determining whether post translational modifications may play a role in the formation or function in the tumor microenvironment.

Material and Methods

Intact protein from extracellular vesicle: Extracellular vesicles were provided by Dr. Suzanne Ostrand-Rosenberg's lab at UMBC. Myeloid-derived suppressor cells (MDSC) were harvested from the blood of BALB/c mice injected with 4T1/IL-1β mammary carcinoma cells. Purified MDSC were maintained overnight at 37 ºC with 5% CO_2 in serum free medium. Extracellular vesicles (EVs) were isolated with a 10 mL 0.25 M to 2.0 M sucrose density gradient, and resuspended in water. EVs were lysed in 8M urea in the presence of a protease inhibitor cocktail (Sigma Aldrich, St. Louis, MO). The protein concentration was determined using the RC/DC assay.

GELFrEE fractionation: Three hundred micrograms of intact protein was precipitated using using $CHCl_3$/MeOH/H_2O (4:1:3).[61] The resulting pellet was re-suspended in 120 µL of water and 30 µL of acetate buffer. The sample was

reduced for 10 minutes at 50 ºC in 53 mM DTT. The sample was fractioned by tubular electrophoresis using a 12% Tris-acetate polyacrylamide cartridge (Expedeon, San Diego, CA) and applying a voltage that increased from 50V to 85V over two hours. A total of 12 fractions were collected. Each fraction was precipitated using $CHCl_3$/MeOH/H_2O (4:1:3)[61] followed by resuspension in solvent A (97.5% water, 2.4% ACN, 0.1% formic acid) for LC-MS/MS analysis. The fractionation was visualized by SDS-PAGE. Ten microliter aliquots from each fraction were electrophoresed on an 8-16% Tris-HCl polyacrylamide gel (BioRad, Hercules, CA). The gel was run at 200 V using a BioRad apparatus and stained using silver stain (Pierce, Rockford, IL).

LC-MS/MS analysis: Each fraction was desalted with solvent A (97.5% Water, 2.5% ACN and 0.1% formic acid) for 15 minutes at 10 µL/min on a Zorbax column (5 µm, 5 mm x 0.3 mm). Reversed phase chromatography was carried out on a Shimadzu Prominence LC system and autosampler using an Agilent C3 column (5 µm, 150 mm x 0.1 mm) at a flow rate of 300 nL/min and a linear gradient of increasing concentration of solvent B (97.5% ACN, 2.4% water and 0.1% formic acid) from 0% to 85% over 200 minutes. The LC was connected in line with an LTQ-Orbitrap-XL, and the precursor scans were recorded in the Orbitrap at a resolution of 60,000 at 400 m/z (5 averaged). The most abundant precursor ions were fragmented using CID with a nominal activation energy of 25 in the LTQ ion trap. MS/MS spectra for the most abundant ion were recorded in the Orbitrap at a resolution of 60,000 at 400 m/z. The automatic gain control (AGC) targets were set to 1E6 for precursor scans and 5E5 for MS/MS scans. Data

dependent analysis was set to isolate precursor ions with either unassigned charges or charges greater than +4. Dynamic exclusion was employed with a repeat count of 2, repeat duration of 240 seconds and exclusion duration of 300 seconds. The isolation width for the precursor ions was set at 10 Da.

Bioinformatics: Database searches were performed using ProSightPC 2.0[55] against a custom UniProt mouse database consisting of proteins with molecular weight less than 30KDa. The precursor and fragment ions were deconvoluted using the THRASH[56] algorithm in the program. The precursor mass tolerance was set to 2500 Da and fragment ions mass tolerance was set to 15 ppm. The ΔM mode on ProSightPC 2.0 was used to evaluate mass shifts. Sequence Gazer tool was used to localize post-translational modifications and mass shifts. The proteins identified were automatically assigned an E-value. Identifications with E-value lower than 10E-4 were considered as a strong identification.

Result and Discussion

A total of forty unique proteins were identified in the bioinformatic analysis. The list of proteins identified from the analysis is listed in Table 6. The mass range of proteins identified using one averaged precursor ion scan is 5 KDa-16 KDa. Modifications were observed on thirty-nine of the forty proteins identified. The protein identified without any modification was one of the S100A8 proteoforms.

Table 4: List of proteins identified from GELFrEE analysis of EVs.

Number of Matching Fragments	Theoretical Mass (Da)	Observed Mass (Da)	Mass Diff (Da)	E -Value	Accession Number	Protein Description
14	10044.3	9955.28	-89.02	5.12E-07	P14069	Protein S100-A6
25	10288.1	10200.06	-88.04	5.77E-14	P27005	Protein S100-A8
14	10288.1	10330.11	42.01	7E-11	P27005	Protein S100-A8
24	10288.1	10571.21	283.11	2.33E-12	P27005	Protein S100-A8
21	10288.1	10139.12	-148.98	1.74E-13	P27005	Protein S100-A8
50	10288.1	10157.03	-131.07	8.2E-45	P27005	Protein S100-A8
33	10288.1	10173.02	-115.08	1.24E-23	P27005	Protein S100-A8
34	10288.1	10288.1	0	1.27E-20	P27005	Protein S100-A8
11	14227.8	12024.6	-2203.2	1.92E-05	P70696	Histone H2B type 1-A
22	13943.6	12228.75	-1714.85	3.57E-10	Q64475	Histone H2B type 1-B
21	13943.6	13796.46	-147.14	2.52E-11	Q64475	Histone H2B type 1-B
17	13897.6	13785.45	-112.15	1.45E-05	Q6ZWY9	Histone H2B type 1-C/E/G
14	13927.6	13785.5	-142.1	9.18E-07	P10854	Histone H2B type 1-M
19	13927.6	13779.63	-147.97	1.08E-09	P10854	Histone H2B type 1-M
14	13911.6	13798.41	-113.19	4.81E-05	Q64525	Histone H2B type 2-B
26	14126.9	14037.93	-88.97	8.54E-13	P22752	Histone H2A type 1
18	14086.9	13998.74	-88.16	1.8E-08	Q6GSS7	Histone H2A type 2-A
27	14112.9	14037.89	-75.01	6.53E-12	Q8BFU2	Histone H2A type 3
17	15133.4	15044.38	-89.02	1.31E-05	P27661	Histone H2A.x

Number of Matching Fragments	Theoretical Mass (Da)	Observed Mass (Da)	Mass Diff (Da)	E -Value	Accession Number	Protein Description
20	13897.6	13766.58	-131.02	3.23E-12	Q6ZWY9	Histone H2B type 1-C/E/G
16	13927.6	13779.56	-148.04	5.18E-08	P10853	Histone H2B type 1-F/J/L
22	13927.6	13796.46	-131.14	4.22E-12	P10853	Histone H2B type 1-F/J/L
20	13911.6	13796.45	-115.15	1.49E-10	Q64478	Histone H2B type 1-H
17	13927.6	13766.61	-160.99	7.35E-06	P10854	Histone H2B type 1-M
24	13983.6	13796.51	-187.09	2.6E-09	Q8CGP2	Histone H2B type 1-P
19	13983.6	13811.44	-172.16	2.77E-06	Q8CGP2	Histone H2B type 1-P
20	13911.6	13766.56	-145.04	5.32E-06	Q64525	Histone H2B type 2-B
12	13911.6	13796.58	-115.02	2.96E-05	Q64525	Histone H2B type 2-B
18	15394.5	15305.58	-88.92	9.43E-06	P68433	Histone H3.1
18	15378.5	13026.14	-2352.36	5E-08	P84228	Histone H3.2
20	15318.5	13026.22	-2292.28	5.08E-16	P84244	Histone H3.3
12	15318.5	15291.43	-27.07	2.07E-07	P84244	Histone H3.3
15	15348.5	13025.14	-2323.36	1.52E-06	Q9FX60	Histone H3-like 1
19	15348.5	13807.66	-1540.84	6.22E-09	Q9FX60	Histone H3-like 1
25	11360.4	11301.34	-59.06	3.66E-15	P62806	Histone H4
49	11360.4	9044.04	-2316.36	2.92E-44	P62806	Histone H4
13	11360.4	9705.39	-1655.01	8.98E-06	P62806	Histone H4
19	11360.4	11329.38	-31.02	1.65E-05	P62806	Histone H4
12	6269.31	5621.02	-648.29	3.6E-05	P47945	Metallothionein-4
18	9417.07	9285.97	-131.1	2.82E-07	P09602	Non-histone chromosomal protein HMG-17

The S100 family of proteins is expressed in immature cells of myeloid lineage including myeloid-derived suppressor cells. The members of the S100 family are calcium binding proteins and act as inflammatory mediators when released by the cells of myeloid origin. These proteins are elevated in inflammatory conditions and are chemotactic and chemokines.[68] In addition, they are unregulated in many tumors and serve as markers in immune cells within the tumor microenvironment.[69] The top-down analysis of these vesicles allows the identifications of various proteoforms of proteins from the S100 family. In our analysis the two proteins identified from this family are S100 A6 and S100 A8.

S100 A6 was identified at 100.39 minutes in fraction 2. The intact mass of the protein was observed to be 9,955.28 Da which is 89.02 Da less than the theoretical mass of 10,044.3 Da. Using the Sequence Gazer tool in Prosigh PC this mass difference was localized to truncation of the N-terminal methionine and acetylation of the alanine residue. The protein was matched with an E-value of 5.12E-07. The precursor and the product ion spectrum, as well as the protein sequence with the fragments matched and the modifications observed are showin in Figure 37.

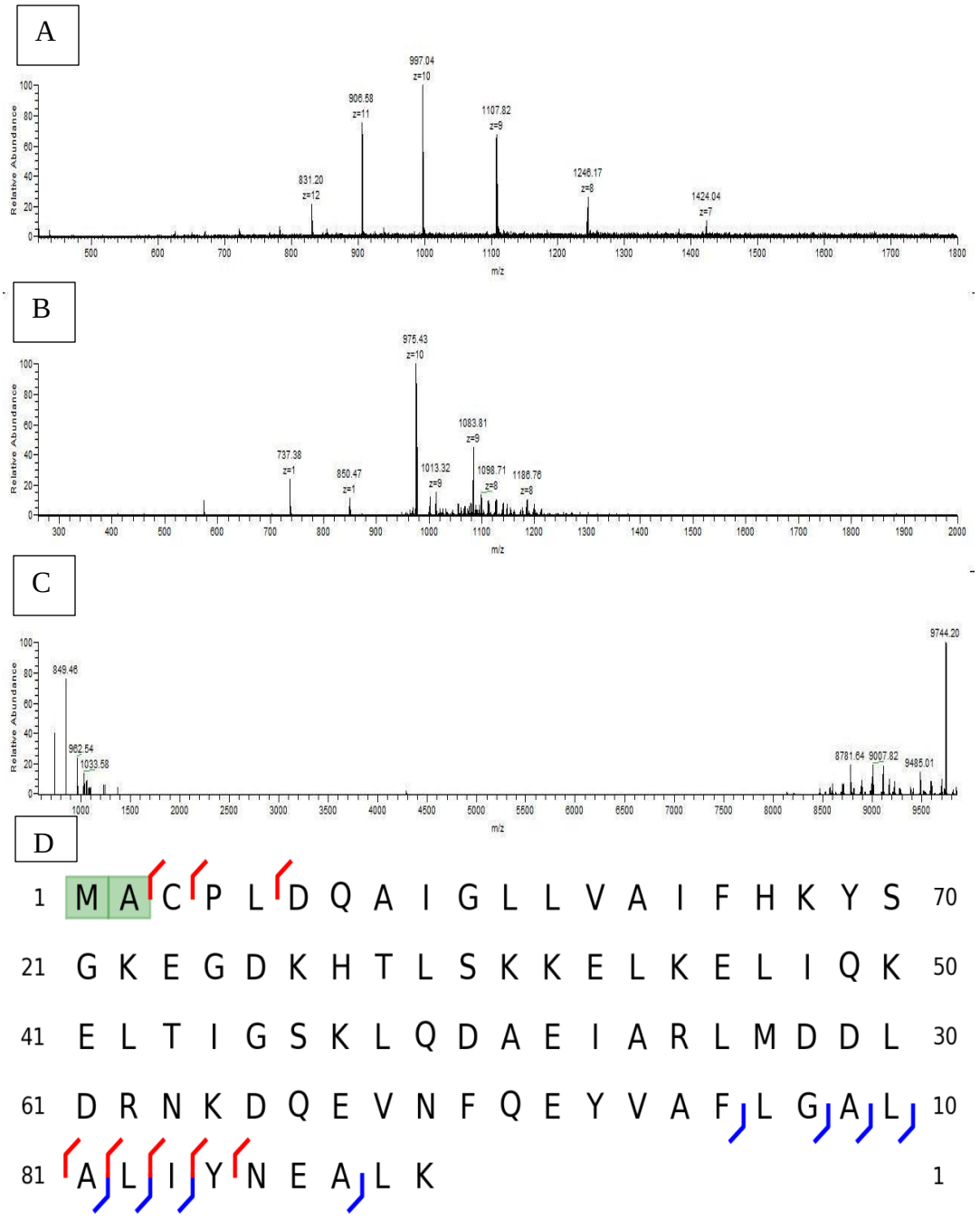

Figure 37: (A) Precursor ion spectrum at retention time 100.39 minutes from fraction 2 of protein S100 A6 (B) Product ions from m/z 997.14 with charge of +10 (C) Decharged product ion scan. (D) Protein sequence, fragments identified and modification localized using ProSightPC with an E-value of 5.12E-07. The mass protein observed was 89.02 Da less compared to the theoretical mass.

There were 6 different proteoforms of S100 A8 identified from fraction 2, 3 and 4. The first proteoform identified in fraction two and the precursor and product ion spectrum is shown in Figure 38A-C. The protein was identified with an E-value of 1.27E-20 and 34 fragments matched. There was no mass difference between the observed and theoretical mass of the protein. The protein sequence and the fragments matched are shown in Figure 38D.

The second proteoform of S100 A8 was identified in fraction 2. The mass difference between the observed and theoretical mass was -88.04 Da. This mass difference was localized using Sequence Gazer by truncation of the initial methionine and acetylation of the serine residue. There were 25 fragments matched, and the protein was identified with an E-value of 5.77E-14. The precursor ion spectrum, product ion spectrum as well as the modification and the fragments matched are shown in Figure 39.

A third proteoform of S100 A8 was also identified in fraction 2. The observed mass of the protein was 10,330.11 Da which is 42.01 Da higher than the theoretical mass of 10288.1 Da. The modification on this protein was localized as an acetylation on the N-terminal methionine residue. The proteoform was identified with an E-value of 7E-11 and 14 matched fragments. The precursor ion spectrum, product ion spectrum as well as the modification and the fragments matched are shown in Figure 40.

A fourth proteoform of S100 A8 was identified in fraction 3, with a mass of 10,139.12 Da which is 148.98 less than the theoretical mass. The precursor ion

spectrum, product ion spectrum, as well as the modification and the fragments matched are shown in Figure 41. The modification was localized using Sequence Gazer in ProSightPC at the first three amino acid residues. There are two possible modifications for this proteoform of S100 A8. The first is removal of the initial methionine and proline residue, and phosphorylation of the third serine residue. A second possible of this modification is loss of the initial methionine residue and neutral loss of water. The protein was identified with an E-value of 1.74E-13 and 21 matched fragments.

The fifth proteoform of S100 A8 was identified in fraction 4. The mass of the protein observed was 10,157.03 Da which is 131.07 Da less than the theoretical mass of 10,288.1 Da. The removal of the initial methionine from the protein was localized using Sequence Gazer in ProSightPC; the protein was identified with an E-value of 8.2E-45 and 50 matched fragments. The precursor ion spectrum, product ion spectrum as well as the modification and the fragments matched for this proteoform are shown in Figure 42.

Protein S100 A8 with an observed mass of 10173.02 was identified in fraction 3. The mass difference between the theoretical mass and observed mass was -115.08 Da. This mass difference is interpreted as resulting from the removal of the initial methionine and oxidation of 37^{th} methionine residue. The localization of these modifications using ProSightPC yielded 30 matched fragments and an E-Value of 1.24E-23. The precursor ion spectrum, product ion spectrum as well as the modification and the fragments matched are shown in Figure 43.

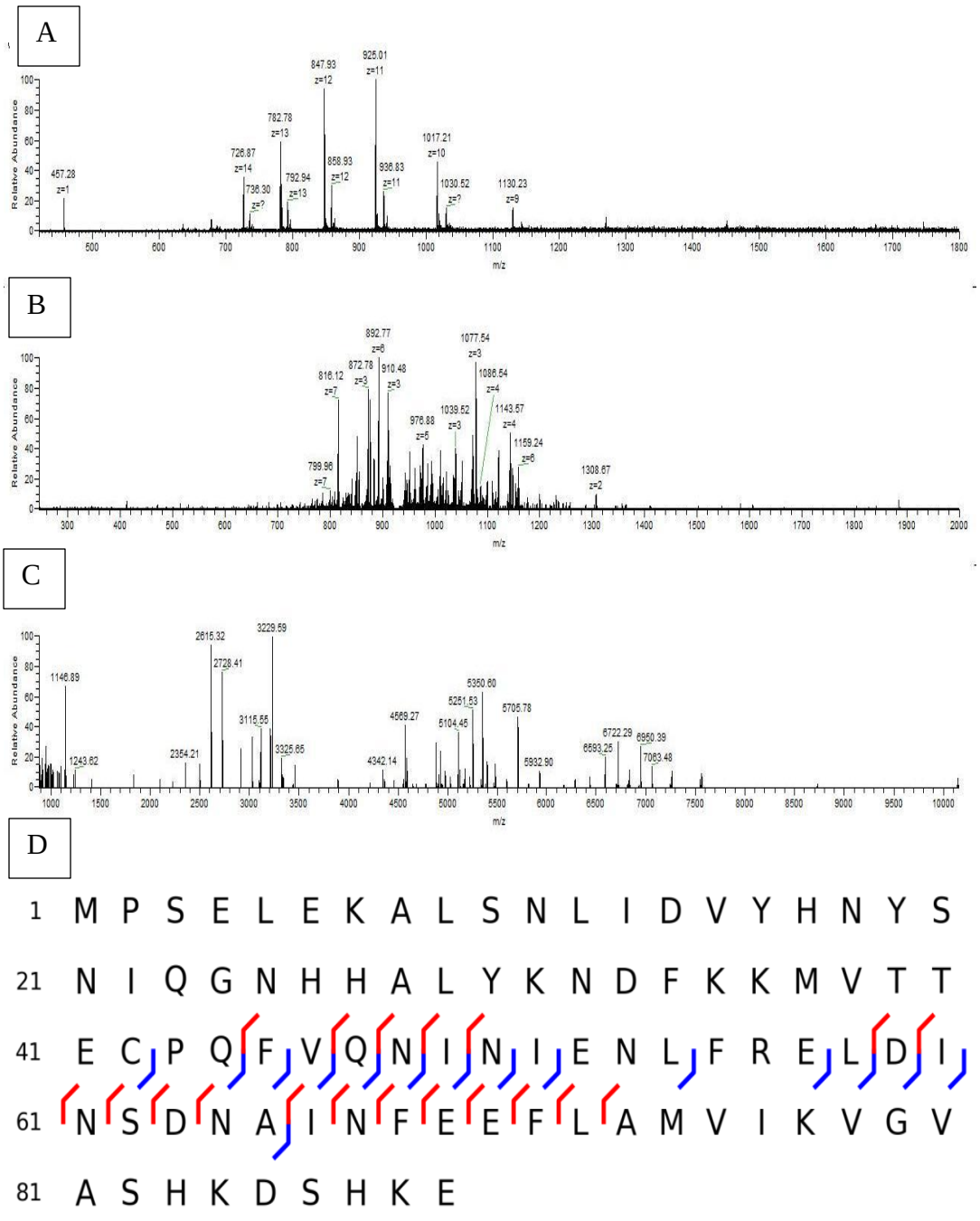

Figure 38: (A) Precursor ion spectrum at retention time 121.74 minutes from fraction 2 assigned to protein S100 A8 (B) Product ions from m/z 936.75 with charge of +11 (C); Decharged product ion scan. (D) Protein sequence assigned (S100-A8) and fragments identified by ProSightPC with an E-value of 1.27E-20

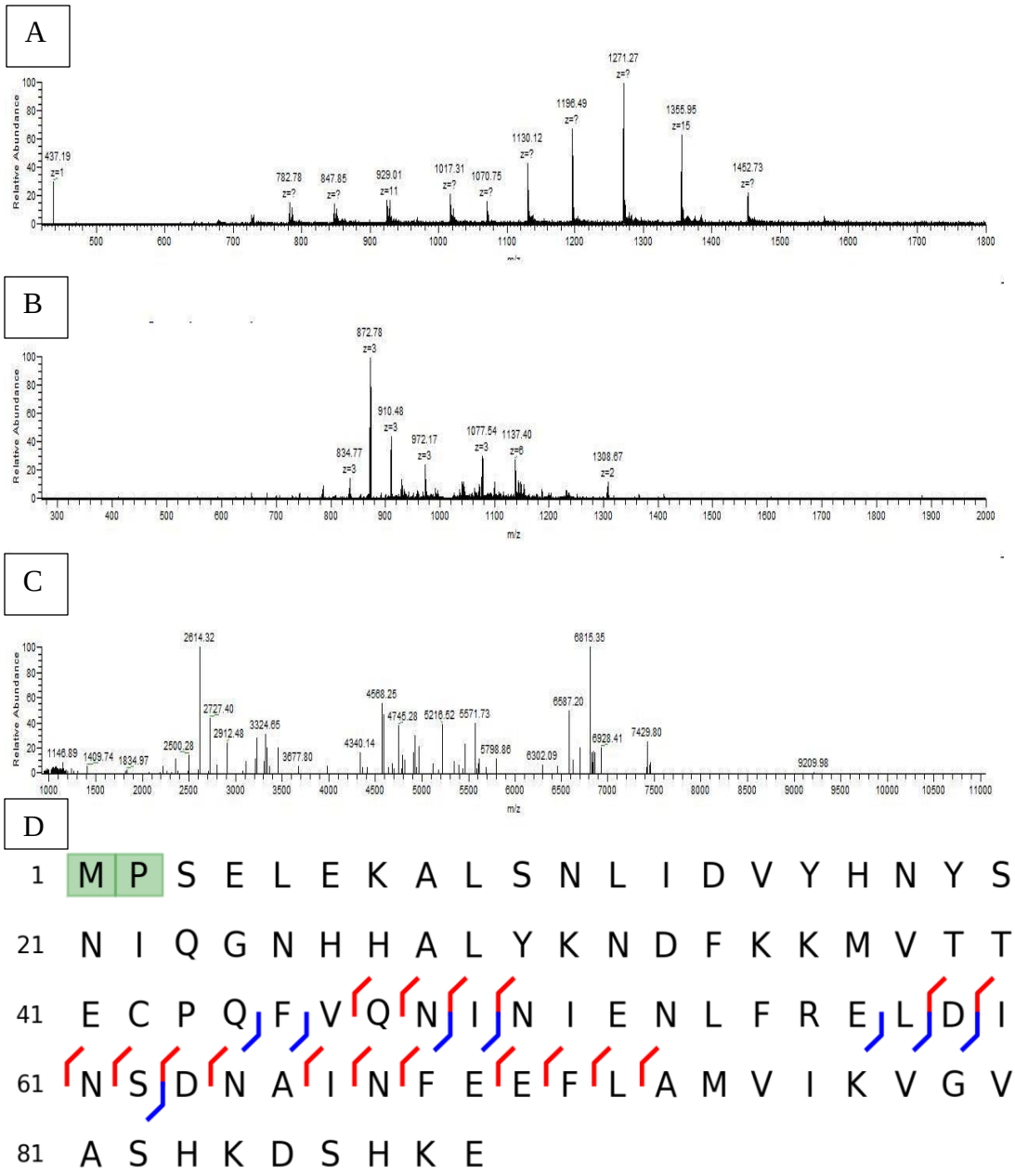

Figure 39: (A) Precursor ion spectrum at retention time 91.27 minutes from fraction 2 assigned as protein S100 A8 (B) Product ions from m/z 1021.60 with charge of +10 (C) Decharged product ion scan. (D) Protein sequence, fragments identified and modification localized using ProSightPC with an E-value of 5.77E-14. The mass protein observed was 88.04 Da less than the theoretical mass.

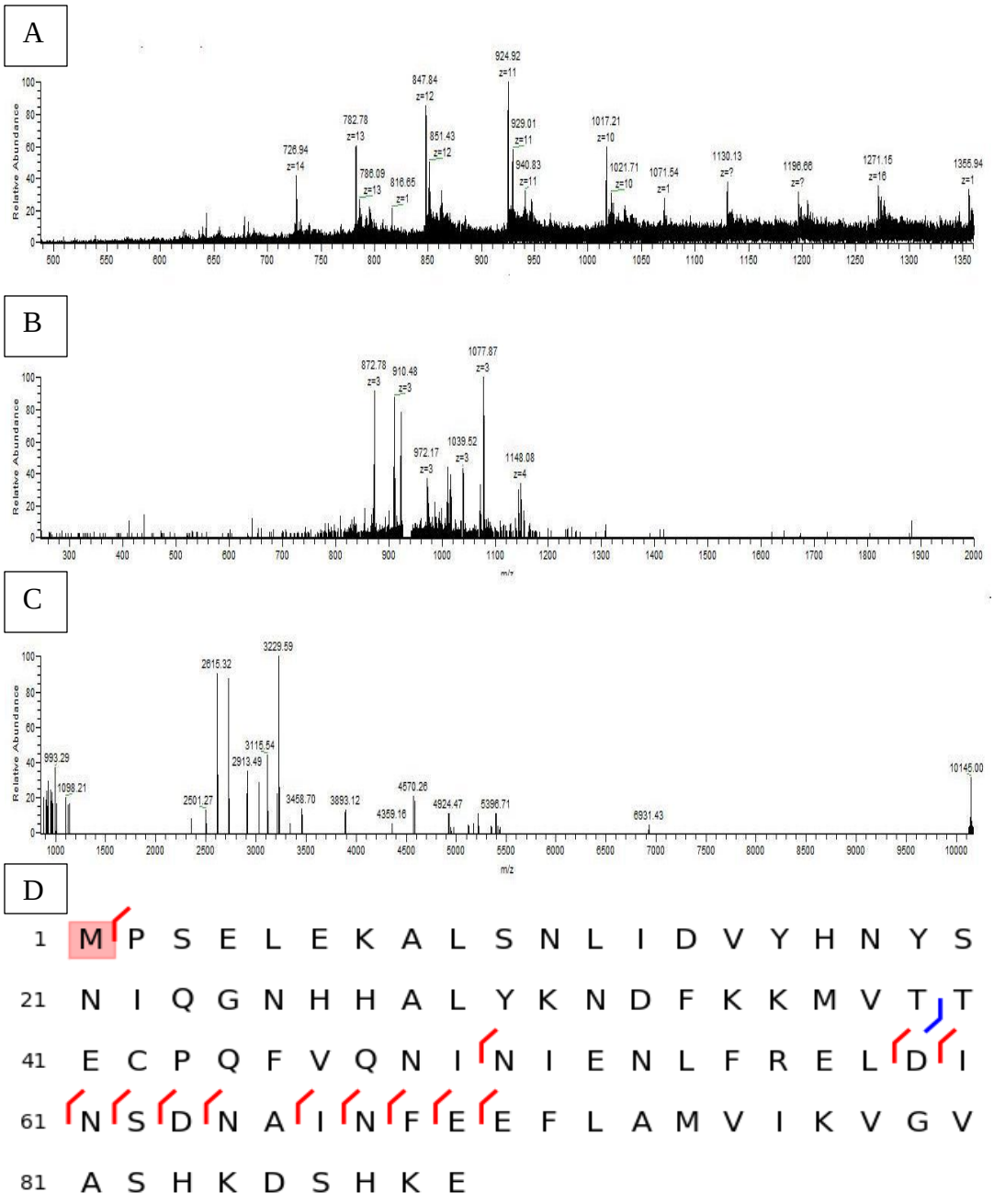

Figure 40: (A) Precursor ion spectrum at retention time 93.66 minutes from fraction 2, assigned as protein S100 A8 (B) Product ions from m/z 940.64 with charge of +11 (C) Decharged product ion scan. (D) Protein sequence, fragments identified and modification localized using ProSightPC with an E-value of 7E-11. The mass protein observed was 42.01 Da more compared to the theoretical mass.

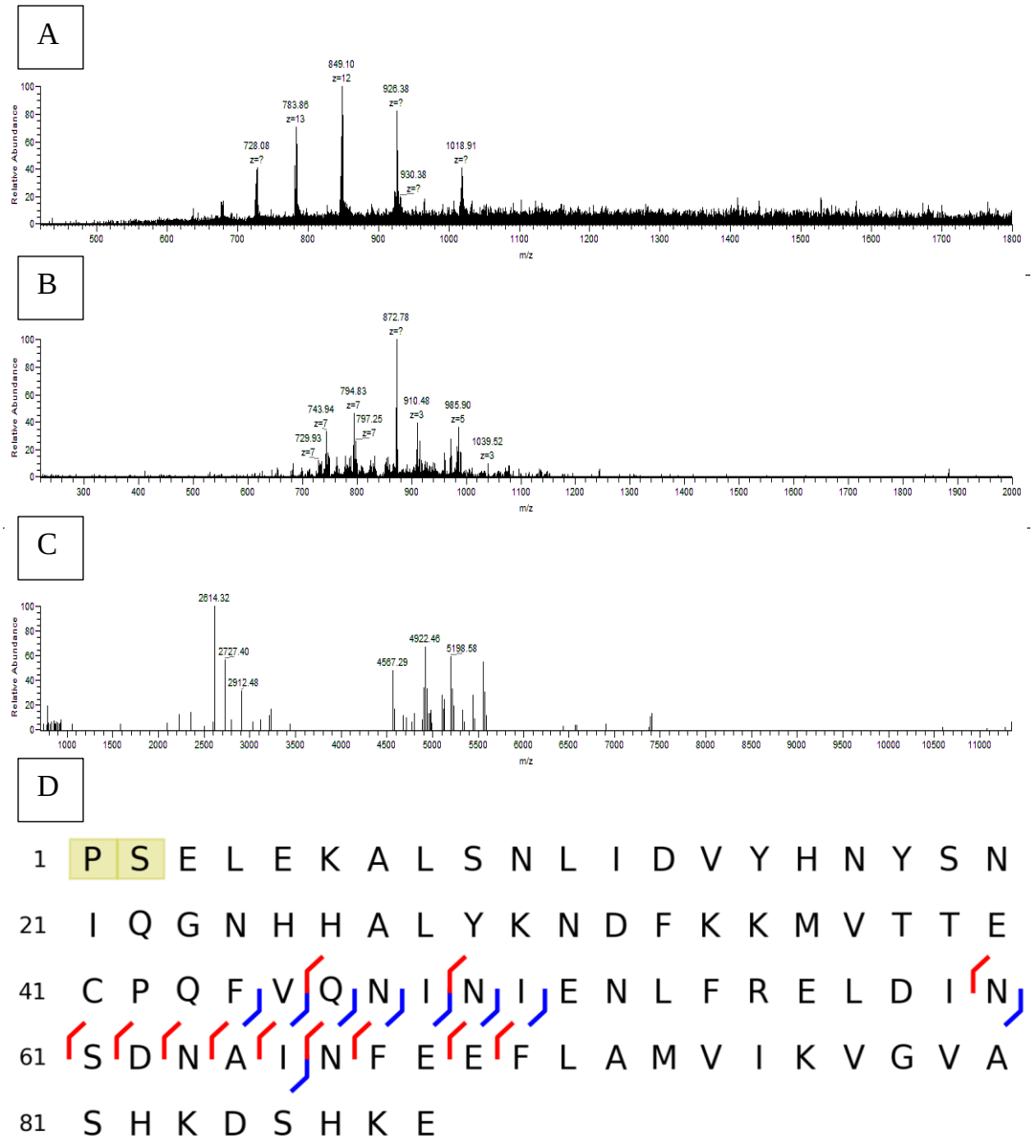

Figure 41: (A) Precursor ion spectrum at retention time 117.63 minutes from fraction 3, assigned as protein S100 A8 (B) Product ions from m/z 846.43 with charge of 12 (C) Decharged product ion scan. (D) Protein sequence, fragments identified and modification localized using ProSightPC with an E-value of 1.74E-13. The mass protein observed was 148.98 Da less compared to the theoretical mass.

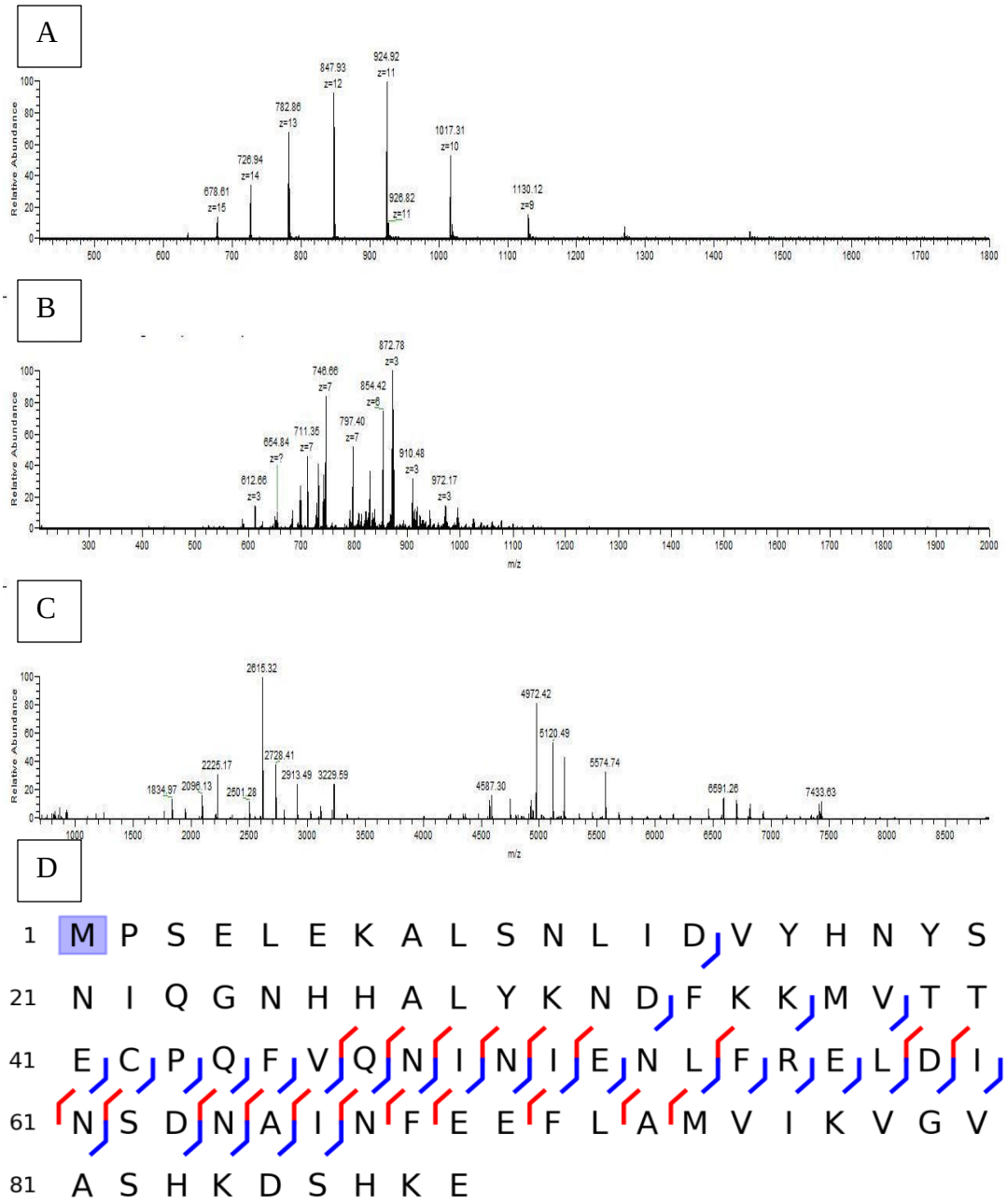

Figure 42: (A) Precursor ion spectrum at retention time 119.28 minutes from fraction 4 assigned as S100 A8 (B) Product ions from m/z 783.78 with charge of +13 (C) Decharged product ion scan. (D) Protein sequence, fragments identified and modification localized using ProSightPC with an E-value of 8.2E-45. The mass difference of protein observed was 131.04 Da less compared to the theoretical mass.

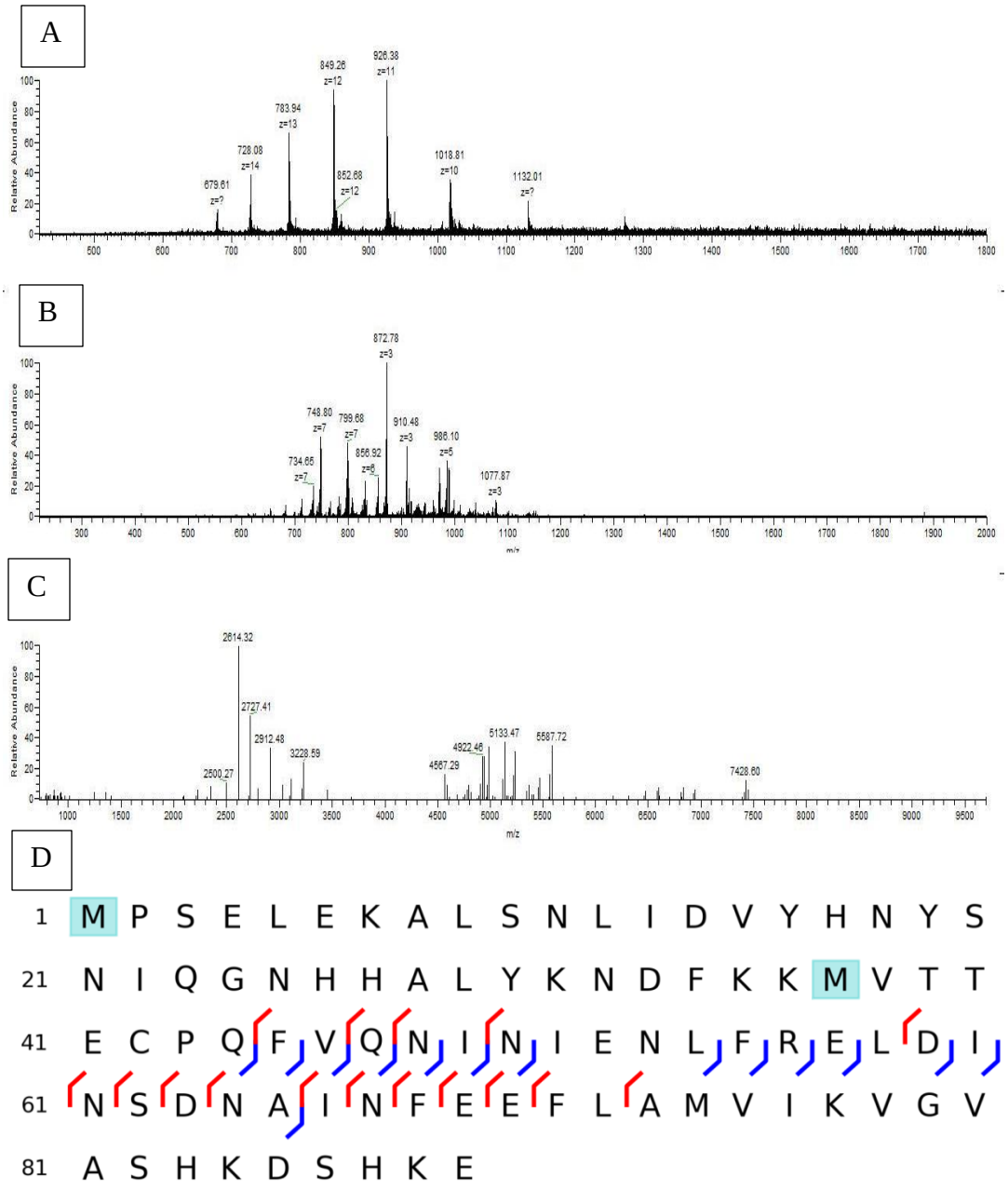

Figure 43: (A) Precursor ion spectrum at retention time 118.00 minutes from fraction 3 assigned as S100 A8 (B) Product ions from m/z 849.26 with charge of +12 (C) Decharged product ion scan. (D) Protein sequence, fragments identified and modification localized using ProSightPC with an E-value of 1.24E-23. The mass difference of protein observed was 115.08 Da less compared to the theoretical mass.

The other family of proteins that were identified in this analysis was histones. They are important chromatin proteins that are associated with the DNA, which package and order the DNA in to nucleosomes. The histone proteins have a wide variety of post-translational modifications, including acetylation, methylation phosphorylation and ubiquitylation. Although there only 4 core histones H4, H2B, H2A and H3 each family has many potential isoforms due to sequence variation and observed post-translational modifications.[70] The analyses of the intact proteins of EVs by high resolution mass spectrometry allow the identifications of these intact proteins as well as their post translational modifications.

The three proteoforms identified for histone H2A were histone 2A type 1, histone 2A type 2-A and histone 2A.x. The three proteins have very similar sequences and differ only by a few amino acid residues. Histone 2A type 1 and histone 2A type 2-A were identified in fraction 6 of the tubular electrophoresis, and histone 2A.x was identified in fraction 7. In all three cases, the modifications were observed on the N-terminus of the protein, with removal of the initial methionine and acetylation of the serine residue. The precursor ion spectrum, product ion spectrum as well as the modification and the fragments matched are for histone 2A type 1, histone 2A type 2-A and histone 2A.x in Figure 44, 45 and 46, respectively.

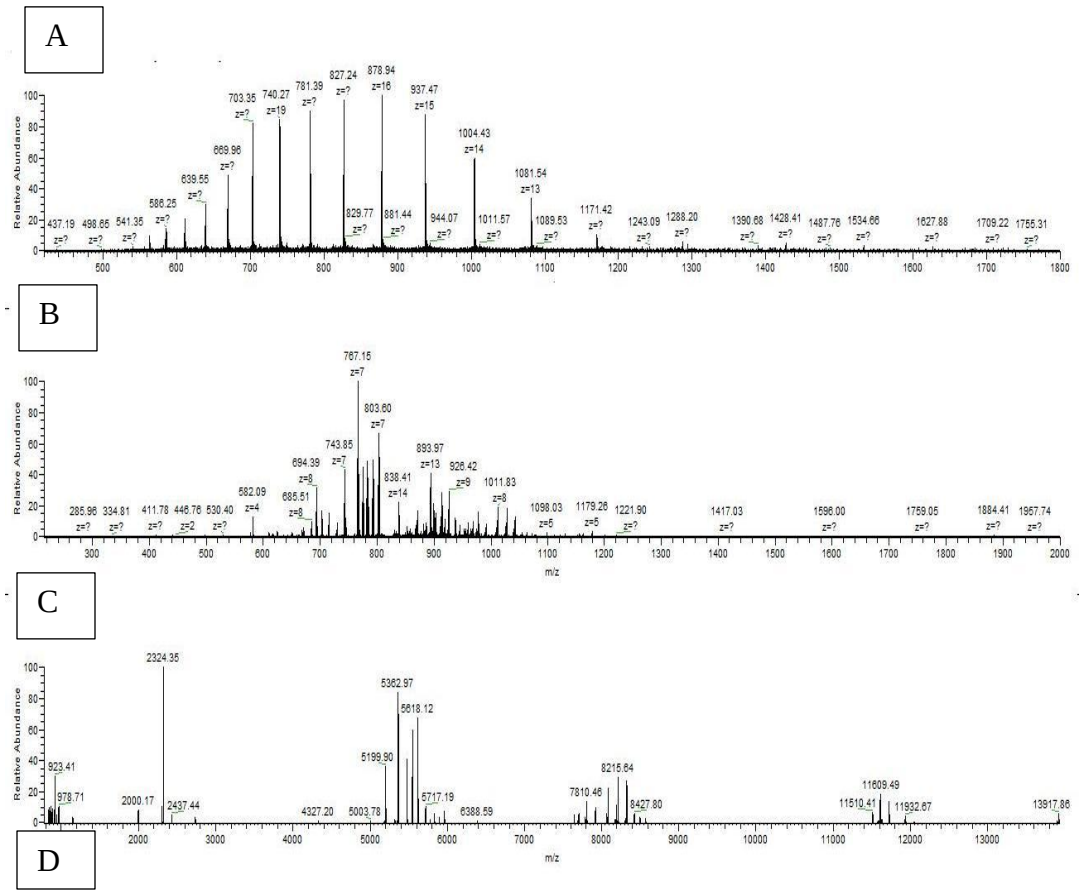

Figure 44: (A) Precursor ion spectrum at retention time 104.37 minutes from fraction 6 assigned to histone 2A type 1 (B) Product ions from m/z 827.24 with charge of +17 (C) Decharged product ion scan. (D) Protein sequence, fragments identified and modification localized using ProSightPC with an E-value of 8.54E-13. The mass difference of protein observed was 88.97 Da less than the theoretical mass.

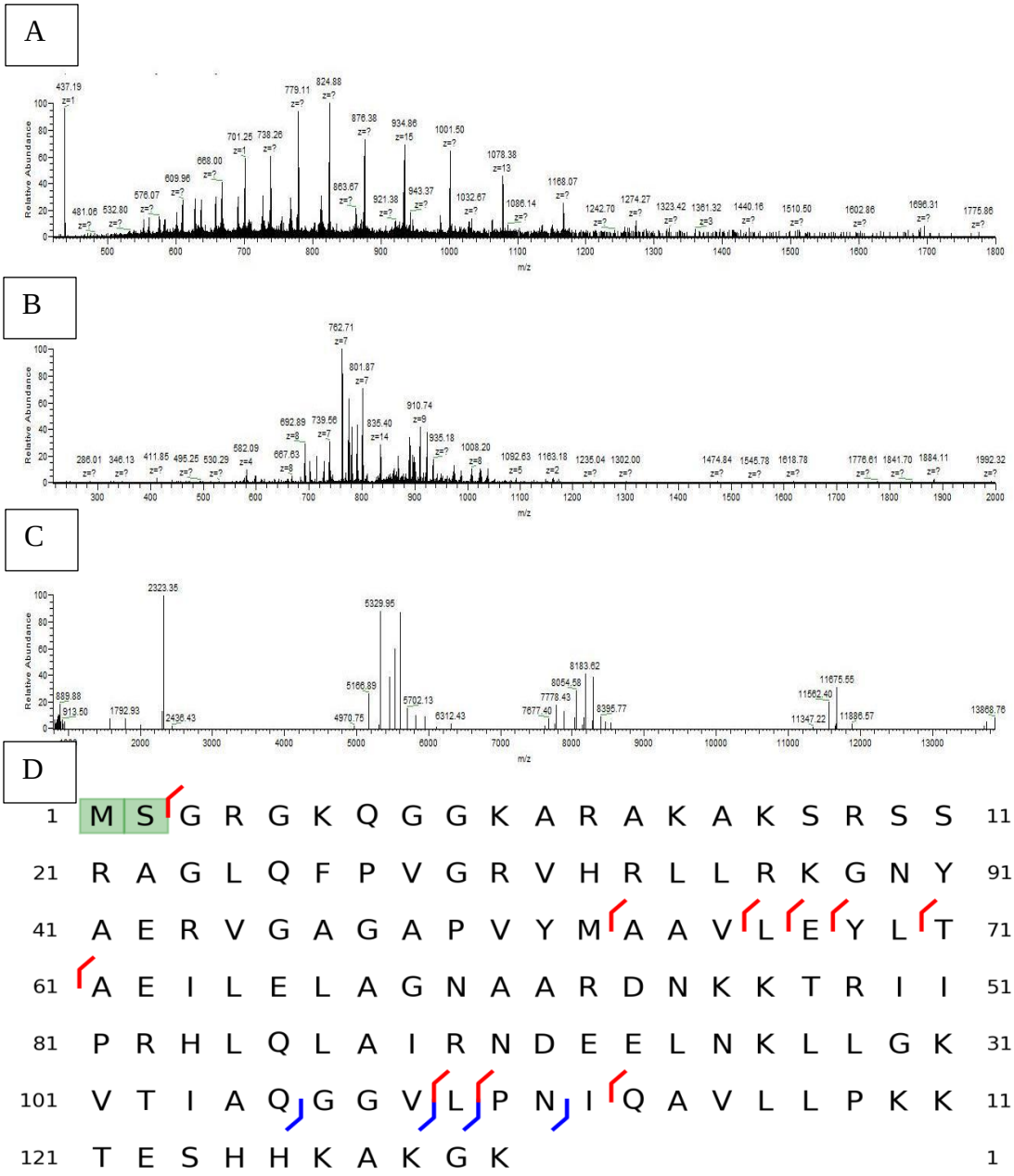

Figure 45: (A) Precursor ion spectrum at retention time 101.56 minutes from fraction 6 histone assigned to 2A type 2-A (B) Product ions from m/z 824.89 with charge of +17 (C) Decharged product ion scan. (D) Protein sequence, fragments identified and modification localized using ProSightPC with an E-value of 1.8E-08. The mass difference of protein observed was 88.16 Da less than theoretical mass.

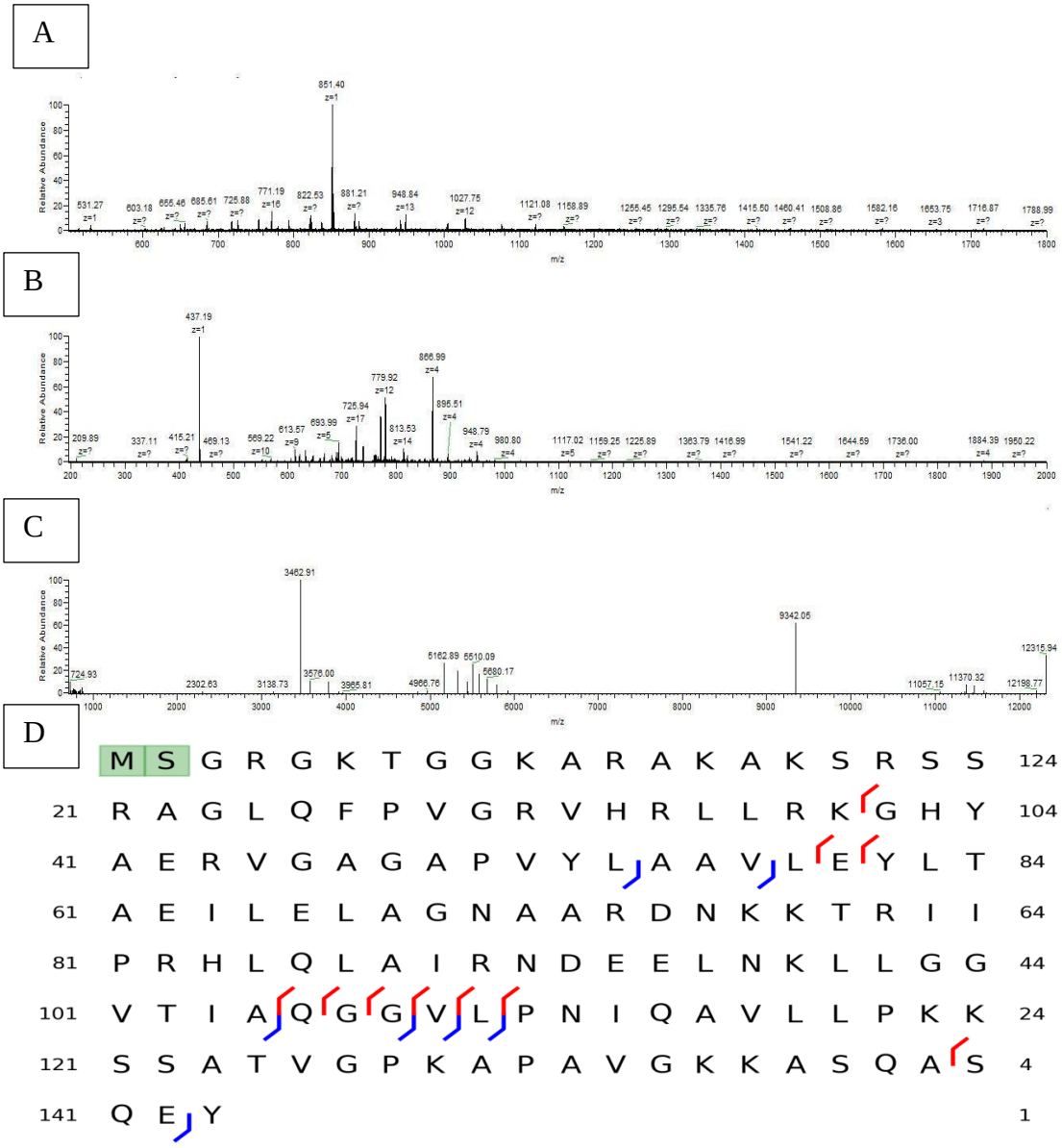

Figure 46: (A) Precursor ion spectrum at retention time 101.56 minutes from fraction 7 assigned to histone 2A.x (B) Product ions from m/z 753.62 with charge of +20 (C) Decharged product ion scan. (D) Protein sequence, fragments identified and modification localized using ProSightPC with an E-value of 1.31E-05. The mass difference of protein observed was 88.02 Da less compared to the theoretical mass.

Conclusion

Application of an effective work flow, including optimized fractionation, front end chromatographic separation and activation, allowed the identification of 40 proteins in EVs shed by MSDC cells. Modifications were observed on 39 of the 40 proteins identified. The two main families of proteins identified in this analysis were histones and S100. Twenty nine proteoforms were identified from the forty proteins were histone proteins. The analysis enabled the identification of six proteoforms of S100 A8. Most of the proteins identified are small basic proteins; the proteins, which ionize well with ESI. The sample preparation for the analysis is also bias to basic proteins.

Chapter 6: Conclusions

The field of proteomics predominantly uses the bottom-up approach for complex mixture analysis. Bottom up strategies identify the presence of many hundreds of proteins. However, top-down strategies provide more information about each protein. Due to the increased availability of robust high resolution mass analyzers such as the Orbitrap and TOF, it is possible for more laboratories to analyze intact protein in complex mixtures. The work in this thesis shows that, on an LC-MS/MS time scale, it is possible to identify up to 50 proteins in a given analysis. The mass range of the proteins identified was 4 KDa to 22 KDa.

Pre-fractionation of complex mixtures prior to intact LC-MS/MS analysis increases the number of identifications. Optimization of experimental conditions for front end fractionation of complex mixtures increases the number of identification. The HPLC columns are not as effective for intact proteins as for peptide analysis. In our evaluation none of the 3 activation methods has a clear advantage. Our Orbitrap has sufficient resolution, but seems to have an upper mass limit of 25 KDa. In addition, only limited choices in software are currently available for intact protein analysis. Specialized software is required to deconvolute the precursor ions and fragment ions. The next iteration for this optimized work flow would be to be able to perform quantitative proteomics analysis on a protein level.

Appendices

Table 1: Proteins identified with 1D LC-MS/MS analysis of *E. coli*. lysate from 1D-LC-MS/MS

# of Matching Fragments	Theoretical Mass (Da)	Observed Mass (Da)	E Value	Mass Diff (Da)	Accession Number	Protein Description
13	4961.72	5092.75	1.88E-17	131.03	253973439	30S ribosomal subunit protein S22
8	6236.58	6250.58	0.000055	14	253975465	50S ribosomal protein L33
13	6311.39	6311.34	1.46E-12	-0.05	254161195	50S ribosomal protein L32
6	6460.54	6591.56	4.78E-06	131.02	253975191	hypothetical protein ECB_03213
6	7137.94	7284.94	6.23E-06	147	253975141	50S ribosomal protein L29
12	7145.6	7145.55	7.4E-13	-0.05	254160455	hypothetical protein ECB_00336
8	7266.72	7266.7	2.42E-08	-0.02	253973783	stress protein, member of the CspA-family
8	8152.25	8153.29	6.33E-10	1.04	253974103	hypothetical protein ECB_02117
6	8189.05	8319.99	3.57E-05	130.94	254163981	putative stress-response protein
10	8747.19	8762.2	3.43E-10	15.01	253973797	hypothetical protein ECB_01807
13	9088.95	9220.01	1.85E-14	131.06	253972411	HU, DNA-binding transcriptional regulator, beta subunit
14	9219.99	9219.89	5.54E-17	-0.1	254160510	transcriptional regulator HU subunit beta
8	9398.15	9529.14	8.42E-07	130.99	253975843	HU, DNA-binding transcriptional

# of Matching Fragments	Theoretical Mass (Da)	Observed Mass (Da)	E Value	Mass Diff (Da)	Accession Number	Protein Description
						regulator, alpha subunit
5	10612.6	10743.54	5.07E-05	130.94	253975046	predicted ribosome-associated, sigma 54 modulation protein
5	11028.5	11028.53	3.49E-05	0.03	253976002	unknown, RNA-binding protein Hfq

Table 2: Proteins identified from bRP-aRP LCMS/MS analysis of *E. coli.* lysate

Number of Matching Fragments	Theoretical Mass(Da)	Observed Mass (Da)	E Value	Mass Diff (Da)	Accession Number	Protein Description
6	4230.41	4230.38	1.53E-07	-0.03	254163227	50S ribosomal protein L36
8	4961.72	5092.7	1.18E-08	130.98	253973439	30S ribosomal subunit protein S22
25	6236.58	6250.57	5.38E-37	13.99	253975465	50S ribosomal protein L33
23	6311.39	6311.37	2.24E-32	-0.02	254161195	50S ribosomal protein L32
10	6406.6	6406.59	6.12E-11	-0.01	253975131	50S ribosomal protein L30
17	6503.26	6541.16	7.20E-18	37.9	254161067	ribosome modulation factor
17	6591.58	6591.56	3.28E-18	-0.02	253975191	hypothetical protein ECB_03213
6	6851.65	6878.67	6.33E-05	27.02	253974528	carbon storage regulator
15	7112.83	7112.79	4.05E-21	-0.04	253973666	hypothetical protein ECB_01674
10	7140.72	7140.7	2.35E-12	-0.02	253973496	hypothetical protein ECB_01500
25	7145.6	7145.58	1.82E-36	-0.02	254160455	hypothetical protein ECB_00336
9	7182.72	7182.73	3.98E-08	0.01	253973496	hypothetical protein ECB_01500
17	7266.72	7266.67	1.67E-21	-0.05	253973783	stress protein, member of the CspA-family
7	7267.59	7265.69	3.48E-06	-1.9	253975380	unknown, major cold shock protein
17	7268.98	7268.89	1.92E-21	-0.09	253975141	50S ribosomal protein L29
12	7859.21	7859.18	1.89E-12	-0.03	254163529	hypothetical protein ECB_03457

Number of Matching Fragments	Theoretical Mass(Da)	Observed Mass (Da)	E Value	Mass Diff (Da)	Accession Number	Protein Description
22	7865.92	7865.89	2.85E-31	-0.03	253975787	50S ribosomal subunit protein L31
9.22	8113.28	8113.19	3.47E-08	-0.09	253972899	translation initiation factor IF-1
11	8152.25	8156.26	3.02E-15	4.01	253974103	hypothetical protein ECB_02117
10	8189.05	8320.06	1.72E-10	131.01	254163981	putative stress-response protein
12	8363.69	8362.64	2.83E-13	-1.05	254163012	30S ribosomal protein S21
6	8709.67	8840.68	6.81E-07	131.01	253973803	DNA polymerase III, theta subunit
10	8741.59	8594.48	3.12E-11	-147.1	254161485	hypothetical protein ECB_01383
18	8808.32	8807.28	9.37E-24	-1.04	254161302	cation transport regulator
16	9000.59	8998.53	1.37E-20	-2.06	253975441	glutaredoxin 3
8	9088.95	9219.87	5.89E-07	130.92	253972411	HU, DNA-binding transcriptional regulator, beta subunit
23	9219.99	9219.91	2.17E-25	-0.08	254160510	transcriptional regulator HU subunit beta
5	9248.77	9379.81	1.62E-05	131.04	253973254	hypothetical protein ECB_01249
25	9529.19	9529.09	5.64E-28	-0.1	253975843	HU, DNA-binding transcriptional regulator, alpha subunit
20	9547.28	9548.25	4.36E-20	0.97	253972050	30S ribosomal protein S20
8	10131.5	10130.5	2.08E-08	-1	254163109	30S ribosomal protein S15
15	10203.3	10203.26	1.08E-16	-0.04	254161177	anti-sigma28 factor FlgM
12	10292.7	10292.65	4.67E-12	-0.05	254163244	30S ribosomal protein S19
9	10394.6	10394.54	2.43E-09	-0.06	254163959	hypothetical protein

Number of Matching Fragments	Theoretical Mass(Da)	Observed Mass (Da)	E Value	Mass Diff (Da)	Accession Number	Protein Description
						ECB_03895
8	10461.5	10509.44	4.34E-07	47.94	254164060	putative acyltransferase with acyl-CoA N-acyltransferase domain
7	10464.4	10594.41	6.72E-07	-0.99	254161330	YciI-like protein
15	11178.3	11178.18	1.63E-19	-0.12	254163237	50S ribosomal protein L24
6	13688.3	13688.25	9.11E-06	-0.05	254163530	hypothetical protein ECB_03458

Table 3: Proteins identified from LC-MS/MS analysis of cytosol from MCF-7 cancer cells with one averaged scan for the precursor ion

Number of Matching Fragments	Theoretical Mass (Da)	Observed Mass (Da)	Mass Diff (Da)	E Value	Accession Number	Protein Description
7	4000.05	4227.15	227.1	5.1E-05	Q05BU3	putative uncharacterized protein LOC100125556
36	5049.52	4960.52	-89	3.7E-54	P62328	Thymosin beta-4
11	6643.82	6643.81	-0.01	1.6E-09	P62861	40S ribosomal protein S30
16	6895.64	6937.63	41.99	2.3E-21	P84101	Small EDRK-rich factor 2
10	7061.93	6972.9	-89.03	7.6E-09	Q9Y2S6	Coiled-coil domain-containing protein 72
6	7396.74	7263.72	-133.02	5.4E-05	O00244	Copper transport protein ATOX1
16	7836.2	7878.24	42.04	5.1E-19	P62857	40S ribosomal protein S28
9	8919.42	7136.57	-1782.85	5.4E-06	Q9NRR8	CDC42 small effector protein 1
9	9065.95	8556.69	-509.26	0.00001	Q15843	NEDD8
21	9387.06	9256.03	-131.03	2E-25	P05204	Non-histone chromosomal protein HMG-17
22	10038	9949.01	-88.99	3.6E-26	P07108	Acyl-CoA-binding protein
25	11523.2	11532.17	8.97	1E-15	P20962	Parathymosin
38	11732.8	11643.81	-88.99	1.2E-49	P31949	Protein S100-A11
47	12196	11978.96	-217.04	2.4E-53	P06454	Prothymosin alpha

Table 4: Proteins identified from LC-MS/MS analysis of cytosol from MCF-7 cancer cells with five averaged scans for the precursor ion

Number of Matching Fragments	Theoretical Mass (Da)	Observed Mass (Da)	Mass Diff (Da)	E-Value	Accession Number	Protein Description
37	5049.52	4960.48	-89.04	6.22E-56	P62328	Thymosin beta-4
12	6643.82	6643.82	0	8.01E-13	P62861	40S ribosomal protein S30
16	6895.64	6937.66	42.02	1.58E-20	P84101	Small EDRK-rich factor 2
8	7061.93	6972.9	-89.03	7.13E-06	Q9Y2S6	Coiled-coil domain-containing protein 72
16	7836.2	7878.22	42.02	7.05E-20	P62857	40S ribosomal protein S28
9	9065.95	8537.64	-528.31	7.61E-05	Q15843	NEDD8
17	9387.06	9256.04	-131.02	2E-18	P05204	Non-histone chromosomal protein HMG-17
12	10038	8440.59	-1597.41	9.25E-12	P07108	Acyl-CoA-binding protein
10	10652.6	10520.62	-131.98	3.03E-07	P05114	Non-histone chromosomal protein HMG-14
10	11523.2	11434.23	-88.97	4.07E-11	P20962	Parathymosin
32	11732.8	11643.8	-89	1.88E-42	P31949	Protein S100-A11
24	12196	11978.95	-217.05	5.93E-31	P06454	Prothymosin alpha
7	22768.5	22678.61	-89.89	7.35E-05	P04792	Heat shock protein beta-1

Table 5: Proteins identified from 1D LC-MS/MS analysis of cytosol from MCF-7 cancer cells with five averaged scans for the precursor ion using ETD fragmentation and reaction time of 5 ms.

Number of Matching Fragments	Theoretical Mass (Da)	Observed Mass (Da)	Mass Diff (Da)	E Value	Accession Number	Protein Description
22	6643.82	6643.82	0	5.59E-11	P62861	40S ribosomal protein S30
18	5049.52	4960.49	-89.03	5.27E-10	P62328	Thymosin beta-4

Table 6: Proteins identified from 1D LC-MS/MS analysis of cytosol from MCF-7 cancer cells with five averaged scans for the precursor ion using ETD fragmentation and reaction time of 10 ms.

Number of Matching Fragments	Theoretical Mass(Da)	Observed Mass (Da)	Mass Diff Da	E Value	Accession Number	Protein Description
22	5049.52	4960.49	-89.03	7.67E-27	P62328	Thymosin beta-4
35	6643.82	6643.82	0	1.44E-36	P62861	40S ribosomal protein S30
31	6895.64	6937.64	42	7E-17	P84101	Small EDRK-rich factor 2
22	9387.06	9256.01	-131.05	7.53E-25	P05204	Non-histone chromosomal protein HMG-17
7	10652.6	10521.54	-131.06	1.37E-05	P05114	Non-histone chromosomal protein HMG-14
19	14719	15918.5	1199.5	1.04E-09	P62987	Ubiquitin-60S ribosomal protein L40

Table 7: Proteins identified from 1D LC-MS/MS analysis of cytosol from MCF-7 cancer cells with five averaged scans for the precursor ion using ETD fragmentation and reaction time of 20 ms.

Number of Matching Fragments	Theoretical Mass (Da)	Observed Mass (Da)	Mass Diff (Da)	E Value	Accession Number	Protein Description
11	5049.52	4072.04	-977.48	3.04E-11	P62328	Thymosin beta-4
24	6643.82	6643.78	-0.04	4.99E-32	P62861	40S ribosomal protein S30
18	6895.64	6936.64	41	3.58E-20	P84101	Small EDRK-rich facto 2
10	7836.2	7878.21	42.01	2.38E-10	P62857	40S ribosomal protein S28
9	9387.06	9256.03	-131.03	8.01E-11	P05204	Non-histone chromosomal protein HMG-17
5	10652.6	10521.56	-131.04	4.63E-06	\P05114	Non-histone chromosomal protein HMG-14
8	11732.8	11643.78	-89.02	2.57E-06	P31949	Protein S100-A11
10	12196	11977.92	-218.08	2.71E-11	P06454	Prothymosin alpha

Table 8: Proteins identified from 1D LC-MS/MS analysis of cytosol from MCF-7 cancer cells with five averaged scans for the precursor ion and CID fragmentation.

Number of Matching Fragments	Theoretical Mass (Da)	Observed Mass (Da)	Mass Diff (Da)	E-Value	Accession Number	Protein Description
35	5049.52	4976.49	-73.03	2.55E-20	P62328	Thymosin beta-4
12	6643.82	6643.82	0	8.01E-13	P62861	40S ribosomal protein S30
11	6643.82	6643.8	-0.02	1.78E-06	P84101	Small EDRK-rich factor 2
16	7836.2	7878.22	42.02	7.05E-20	P62857	40S ribosomal protein S28
13	7836.2	7878.18	41.98	6.75E-08	P62857	40S ribosomal protein S28
9	9065.95	8537.64	-528.31	7.54E-05	Q15843	NEDD8
12	9387.06	9256.02	-131.02	5.34E-10	P05204	Non-histone chromosomal protein HMG-17
12	10038	8440.59	-1597.41	9.25E-12	P07108	Acyl-CoA-binding protein
10	10652.6	10520.62	-131.98	3.03E-07	P05114	Non-histone chromosomal protein HMG-14
10	11523.2	11434.23	-88.97	4.07E-11	P20962	Parathymosin
32	11732.8	11643.8	-89	1.88E-42	P31949	Protein S100-A11
28	11732.8	11643.77	-89.03	1.1E-30	P31949	Protein S100-A11
24	12196	11978.95	-217.05	5.93E-31	P06454	Prothymosin alpha
6	22768.5	22678.61	-89.89	6.28E-05	P04792	Heat shock protein beta-1

Table 9: Proteins identified from control sample without fractionation from *E. coli.* lysate

Number of Matching Fragments	Theoretical Mass (Da)	Observed Mass (Da)	Mass Diff Da	E Value	Accession Number	Protein Description
17	4361.45	4359.42	-2.03	5.93E-08	B1X6F1	50S ribosomal protein L36
36	5092.77	5092.76	-0.01	5.65E-29	P68191	Stationary-phase-induced ribosome-associated protein
8	5879.77	5748.79	-130.98	7.35E-06	P56614	Uncharacterized protein ymdF
53	6367.62	6250.61	-117.01	6.74E-33	P0A7N9	50S ribosomal protein L33
41	6442.43	6327.39	-115.04	1.56E-30	B7NAW5	50S ribosomal protein L32
20	6537.65	6406.63	-131.02	4.67E-13	B1IPZ7	50S ribosomal protein L30
46	6851.65	6878.74	27.09	3.77E-31	B1IUY3	Carbon storage regulator
22	7268.98	7268.08	-0.9	2.6E-13	P0A7M7	50S ribosomal protein L29
39	7276.64	7145.62	-131.02	4.11E-29	P0AAN7	Uncharacterized protein yaiA
35	7397.76	7266.76	-131	6.15E-20	P0A9Y7	Cold shock-like protein CspC
24	7865.92	7861.86	-4.06	1.13E-12	B7N2S7	50S ribosomal protein L31
25	7990.25	7859.3	-130.95	1.66E-07	P0C265	Uncharacterized protein yibT
43	8283.29	8152.24	-131.05	2.21E-40	B7MXK3	UPF0352 protein YejL
72	8320.09	8320.07	-0.02	5.96E-59	P68206	UPF0337 protein yjbJ
24	8872.63	8872.63	0	2.63E-11	P0ACW6	Uncharacterized protein ydcH
18	8939.36	8807.4	-131.96	2.02E-07	P0AE64	Cation transport regulator ChaB
35	9184.95	9203.9	18.95	8.34E-26	P0A7T4	30S ribosomal protein S16

Number of Matching Fragments	Theoretical Mass (Da)	Observed Mass (Da)	Mass Diff Da	E Value	Accession Number	Protein Description
67	9219.99	9219.07	-0.92	4.19E-57	P0ACF4	DNA-binding protein HU-beta
70	9396.72	9265.71	-131.01	2.05E-41	B6I4S1	Cell division protein ZapB
61	9529.19	9529.22	0.03	4.17E-38	P0ACF2	DNA-binding protein HU-alpha
23	9889.14	11177.38	1288.24	1.04E-14	P75694	Uncharacterized protein YahO
33	10743.6	10742.7	-0.9	1.33E-24	P0AFX2	Ribosome hibernation promoting factor

Table 10: Proteins identified from MWCOF fractionation workflow of *E. coli*. lysate

Number of Matching Fragment	Theoretical Mass (Da)	Observed Mass Da	Mass Diff (Da)	E -Value	Accession Number	Protein Description
18	4361.45	2959.59	-1401.86	1.65E-10	B1X6F1	50S ribosomal protein L36
35	5092.77	5092.74	-0.03	1.6E-29	P68191	Stationary-phase-induced ribosome-associated protein
50	6367.62	6250.61	-117.01	3.05E-33	P0A7N9	50S ribosomal protein L33
34	6442.43	6327.36	-115.07	1.82E-31	B7NAW5	50S ribosomal protein L32
33	6537.65	6406.61	-131.04	5.68E-18	B1IPZ7	50S ribosomal protein L30
23	6594.53	6591.57	-2.96	3.68E-11	P0ADX6	Uncharacterized protein yhfG
49	6851.65	6878.74	27.09	1.98E-34	B1IUY3	Carbon storage regulator
38	7268.98	7268.97	-0.01	7.24E-33	P0A7M7	50S ribosomal protein L29 O
47	7276.64	7145.62	-131.02	1.71E-42	P0AAN7	Uncharacterized protein yaiA
21	7397.76	7266.71	-131.05	2.07E-11	P0A9Y7	Cold shock-like protein CspC
19	7865.92	7861.91	-4.01	1.09E-10	B7N2S7	50S ribosomal protein L31
48	7990.25	7859.26	-130.99	6.59E-33	P0C265	Uncharacterized protein yibT
26	8283.29	8152.21	-131.08	2.84E-15	B7MXK3	UPF0352 protein YejL
74	8320.09	8320.07	-0.02	3.83E-61	P68206	UPF0337 protein yjbJ
31	8872.63	8872.63	0	6.47E-17	P0ACW6	Uncharacterized protein ydcH
28	9184.95	10643.45	1458.5	5.84E-19	P0A7T4	30S ribosomal protein S16
28	9219.99	10019.22	799.23	2E-16	P0ACF4	DNA-binding protein HU-beta
58	9396.72	9265.67	-131.05	5.31E-40	B6I4S1	Cell division protein ZapB
57	9529.19	9529.22	0.03	3.84E-38	P0ACF2	DNA-binding protein HU-alpha

Table 11: Proteins identified from bRP-aRP fractionation workflow of *E. coli* lysate

Number of Matching Fragment	Theortical Mass (Da)	Observed Mass (Da)	Mass Diff (Da)	E-value	Accession Number	Protein Description
26	5092.77	5107.77	15	2.37E-31	P68191	Stationary-phase-induced ribosome-associated protein
17	5879.77	5748.79	-130.98	5.41E-15	P56614	Uncharacterized protein ymdF
27	6367.62	6250.64	-116.98	9.52E-31	P0A7N9	50S ribosomal protein L33
23	6442.43	6311.43	-131	6.99E-21	B7NAW5	50S ribosomal protein L32
25	6537.65	6406.6	-131.05	6.46E-10	B1IPZ7	50S ribosomal protein L30
42	6550.21	6679.26	129.05	5.72E-16	Q1RAY1	UPF0181 protein yoaH
51	6851.65	6878.74	27.09	1.42E-37	B1IUY3	Carbon storage regulator
29	7268.98	7266.76	-2.22	9.74E-08	P0A7M7	50S ribosomal protein L29
50	7276.64	7145.59	-131.05	5.02E-38	P0AAN7	Uncharacterized protein yaiA
47	7397.76	7266.72	-131.04	1.55E-50	P0A9Y7	Cold shock-like protein CspC
24	7398.63	7267.54	-131.09	8.2E-19	P0A9Y0	Cold shock protein CspA
50	7990.25	7859.26	-130.99	8.52E-42	P0C265	Uncharacterized protein yibT
34	8283.29	8152.26	-131.03	1.25E-25	B7MXK3	UPF0352 protein YejL OS=Escherichia coli O81 (strain
77	8320.09	8320.11	0.02	4.75E-66	P68206	UPF0337 protein yjbJ
31	8409.62	8409.68	0.06	4.85E-23	P0AD41	Uncharacterized protein ypeB
29	8634.22	8503.22	-131	1.34E-20	B7LG27	Acyl carrier protein

Number of Matching Fragment	Theortical Mass (Da)	Observed Mass (Da)	Mass Diff (Da)	E-value	Accession Number	Protein Description
66	8747.19	8745.13	-2.06	1.93E-50	P64503	Uncharacterized protein yebV
33	8872.63	8872.62	-0.01	1.08E-23	P0ACW6	Uncharacterized protein ydcH
17	8890.53	7702.93	-1187.6	3.19E-09	P0AAX7	Uncharacterized protein McbA
16	8939.36	8808.31	-131.05	9.76E-07	P0AE64	Cation transport regulator
39	9113.74	9113.77	0.03	1.1E-20	P0AA06	Phosphocarrier protein HPr
24	9184.95	9184.91	-0.04	0.000047	P0A7T4	30S ribosomal protein S16
48	9219.99	9219.98	-0.01	4.37E-28	P0ACF4	DNA-binding protein HU-beta
37	9379.81	9400.71	20.9	1.37E-20	P0AB63	Protein yciN
71	9396.72	9265.71	-131.01	3.17E-59	B6I4S1	Cell division protein ZapB
54	9529.19	9529.17	-0.02	4.24E-34	P0ACF2	DNA-binding protein HU-alpha
37	10230.4	10099.33	-131.07	5.59E-24	B7NEV3	DNA-directed RNA polymerase subunit omega
52	10595.4	10617.4	22	3.31E-45	P0AB55	Protein yciI
36	10644.5	10569.44	-75.06	6.83E-16	P0A6Y1	Integration host factor subunit beta
55	10743.6	10765.57	21.97	2.03E-46	P0AFX2	Ribosome hibernation promoting factor
14	15207.4	15550.46	343.06	3.95E-06	B7NGD4	30S ribosomal protein S6
14	17677.3	16146.67	-1530.63	3.02E-10	P0AEU9	Chaperone protein skp

Table 12: Proteins identified from GELFrEE fractionation workflow of *E. coli*. lysate

Number of Matching Fragment	Theortical Mass (Da)	Observed Mass (Da)	Mass Diff (Da)	E-value	Accession Number	Protein Description
16	4361.45	4359.45	-2	2.39E-16	B1X6F1	50S ribosomal protein L36
33	5092.77	5092.74	-0.03	8.94E-36	P68191	Stationary-phase-induced ribosome-associated protein
14	5377.08	5377.07	-0.01	3.22E-05	B1LL30	50S ribosomal protein L34
39	6367.62	6250.61	-117.01	7.34E-36	P0A7N9	50S ribosomal protein L33
46	6442.43	6311.4	-131.03	2.35E-40	B7NAW5	50S ribosomal protein L32
35	6537.65	6406.63	-131.02	1.71E-21	B1IPZ7	50S ribosomal protein L30
22	6571.44	6571.4	-0.04	8.29E-15	P0AC92	Protein gnsA
52	6851.65	6878.74	27.09	4.52E-39	B1IUY3	Carbon storage regulator
15	7261.83	7130.83	-131	2.9E-09	P64476	Uncharacterized protein ydiH
40	7268.98	7268.97	-0.01	3.5E-21	P0A7M7	50S ribosomal protein L29
44	7276.64	7145.57	-131.07	1.13E-33	P0AAN7	Uncharacterized protein yaiA
41	7397.76	7266.77	-130.99	6.78E-40	P0A9Y7	Cold shock-like protein CspC
16	7458.81	7327.73	-131.08	1.75E-08	P0A974	Cold shock-like protein CspE
17	7865.92	7861.94	-3.98	2.28E-13	B7N2S7	50S ribosomal protein L31
15	7990.25	7859.21	-131.04	7.17E-07	P0C265	Uncharacterized protein yibT
57	8244.32	8113.24	-131.08	8.81E-49	P69223	Translation initiation factor IF-1
35	8283.29	8152.29	-131	5.76E-	B7MXK3	UPF0352

Number of Matching Fragment	Theortical Mass (Da)	Observed Mass (Da)	Mass Diff (Da)	E-value	Accession Number	Protein Description
				34		protein YejL
37	8318.3	7195.88	-1122.42	1.47E-33	P69777	Major outer membrane lipoprotein
85	8320.09	8320.16	0.07	5.63E-77	P68206	UPF0337 protein yjbJ
44	8634.22	8503.16	-131.06	2.98E-30	B7LG27	Acyl carrier protein
54	8747.19	8745.13	-2.06	1.63E-49	P64503	Uncharacterized protein yebV
15	8872.63	8888.63	16	5.16E-08	P0ACW6	Uncharacterized protein ydcH
65	9113.74	9113.77	0.03	2.76E-48	P0AA06	Phosphocarrier protein HPr
9	9118.88	8987.81	-131.07	5.52E-05	P0A7L8	50S ribosomal protein L27
36	9184.95	9184.96	0.01	4.01E-20	P0A7T4	30S ribosomal protein S16
61	9219.99	9218.97	-1.02	2.84E-41	P0ACF4	DNA-binding protein HU-beta
59	9379.81	9395.79	15.98	2.93E-52	P0AB63	Protein yciN
74	9396.72	9265.71	-131.01	6.35E-61	B6I4S1	Cell division protein ZapB
69	9529.19	9529.22	0.03	1.01E-55	P0ACF2	DNA-binding protein HU-alpha
37	9628.81	9313.62	-315.19	1.8E-23	C5A095	Cell division protein ZapB
30	9678.32	9547.28	-131.04	3.37E-22	B5YYB6	30S ribosomal protein S20
17	9889.14	10291.63	402.49	5.5E-15	P75694	Uncharacterized protein YahO
49	10380.6	10395.51	14.91	1.85E-39	A8A7N8	10 kDa chaperonin
45	10423.7	10292.64	-131.06	8.07E-31	B7M1N0	30S ribosomal protein S19
57	10595.4	10595.4	0	1.24E-47	P0AB55	Protein yciI
37	10630.5	10644.57	14.07	4.69E-22	B7UMZ8	Integration host factor subunit beta
42	10743.6	10743.61	0.01	1.13E-27	P0AFX2	Ribosome hibernation promoting

Number of Matching Fragment	Theortical Mass (Da)	Observed Mass (Da)	Mass Diff (Da)	E-value	Accession Number	Protein Description
						factor
18	11044.9	10912.85	-132.05	1.54E-09	P64583	Uncharacterized protein yqjD
18	11159.6	11028.56	-131.04	4.29E-14	C4ZR49	Protein hfq
57	11280.8	11195.76	-85.04	2.78E-54	P0ACX3	Putative monooxygenase ydhR
21	11309.3	11177.22	-132.08	1.76E-12	B1LHC3	50S ribosomal protein L24
37	11799.1	11666.07	-133.03	1.32E-24	P0AA25	Thioredoxin-1
50	11960	11855.95	-104.05	3.76E-35	Q8X7I0	UPF0339 protein yegP
56	12287.5	12198.47	-89.03	2.44E-43	B1XBY8	50S ribosomal protein L7/L12
21	12776.6	12644.45	-132.15	1.88E-07	P0AD50	Ribosome-associated inhibitor
38	14002.1	12021.95	-1980.15	5.78E-22	P0ADU5	Protein ygiW
21	16160.6	16146.7	-13.9	4.68E-07	B7MD73	6,7-dimethyl-8-ribityllumazine synthase
35	17677.3	15681.18	-1996.12	3.18E-16	P0AEU9	Chaperone protein skp
18	20832.4	20701.44	-130.96	3.33E-05	A7ZYV7	Flavoprotein

Bibliography

1. Dass, C., *Fundamentals of Comtemporary Mass Spectrometry*. Wiley-Interscience: 2007.

2. Anderson, N. L.; Anderson, N. G., Proteome and proteomics: New technologies, new concepts, and new words. *ELECTROPHORESIS* **1998,** 19, 1853-1861.

3. O'Farrell, P. H., High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* **1975,** 250, 4007-21.

4. Nesvizhskii, A. I.; Aebersold, R., Analysis, statistical validation and dissemination of large-scale proteomics datasets generated by tandem MS. *Drug Discov Today* **2004,** 9, 173-81.

5. Pasa-Tolic, L.; Masselon, C.; Barry, R. C.; Shen, Y.; Smith, R. D., Proteomic analyses using an accurate mass and time tag strategy. *Biotechniques* **2004,** 37, 621-4, 626-33, 636 passim.

6. Zhu, H.; Bilgin, M.; Snyder, M., Proteomics. *Annu Rev Biochem* **2003,** 72, 783-812.

7. Ho, C. S.; Lam, C. W.; Chan, M. H.; Cheung, R. C.; Law, L. K.; Lit, L. C.; Ng, K. F.; Suen, M. W.; Tai, H. L., Electrospray ionisation mass spectrometry: principles and clinical applications. *Clin Biochem Rev* **2003,** 24, 3-12.

8. Hunt, D. F.; Yates, J. R.; Shabanowitz, J.; Winston, S.; Hauer, C. R., Protein sequencing by tandem mass spectrometry. *Proceedings of the National Academy of Sciences* **1986,** 83, 6233-6237.

9. Meng, F.; Cargile, B. J.; Miller, L. M.; Forbes, A. J.; Johnson, J. R.; Kelleher, N. L., Informatics and multiplexing of intact protein identification in bacteria and the archaea. *Nat Biotech* **2001,** 19, 952-957.

10. Kellie, J. F.; Catherman, A. D.; Durbin, K. R.; Tran, J. C.; Tipton, J. D.; Norris, J. L.; Witkowski, C. E., 2nd; Thomas, P. M.; Kelleher, N. L., Robust analysis of the yeast proteome under 50 kDa by molecular-mass-based fractionation and top-down mass spectrometry. *Anal Chem* **2012,** 84, 209-15.

11. Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S., Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrometry Reviews* **1998,** 17, 1-35.

12. Armirotti, A.; Damonte, G., Achievements and perspectives of top-down proteomics. *PROTEOMICS* **2010,** 10, 3566-76.

13. Cannon, J.; Lohnes, K.; Wynne, C.; Wang, Y.; Edwards, N.; Fenselau, C., High-Throughput Middle-Down Analysis Using an Orbitrap. *Journal of Proteome Research* **2010,** 9, 3886-3890.

14. Motoyama, A.; Yates, J. R., Multidimensional LC Separations in Shotgun Proteomics. *Analytical Chemistry* **2008,** 80, 7187-7193.

15. Gao, M.; Deng, C.; Yu, W.; Zhang, Y.; Yang, P.; Zhang, X., Large scale depletion of the high-abundance proteins and analysis of middle- and low-

abundance proteins in human liver proteome by multidimensional liquid chromatography. *PROTEOMICS* **2008,** 8, 939-947.

16. James, P., Protein identification in the post-genome era: the rapid rise of proteomics. *Quarterly Reviews of Biophysics* **1997,** 30, 279-331.

17. Bonvin, G.; Schappler, J.; Rudaz, S., Capillary electrophoresis–electrospray ionization-mass spectrometry interfaces: Fundamental concepts and technical developments. *Journal of Chromatography A* **2012,** 1267, 17-31.

18. Wu, Q.; Yuan, H.; Zhang, L.; Zhang, Y., Recent advances on multidimensional liquid chromatography–mass spectrometry for proteomics: From qualitative to quantitative analysis—A review. *Analytica Chimica Acta* **2012,** 731, 1-10.

19. Tao, D.; Zhang, L.; Shan, Y.; Liang, Z.; Zhang, Y., Recent advances in micro-scale and nano-scale high-performance liquid-phase chromatography for proteome research. *Analytical and Bioanalytical Chemistry* **2011,** 399, 229-241.

20. Nakamura, T.; Kuromitsu, J.; Oda, Y., Evaluation of Comprehensive Multidimensional Separations Using Reversed-Phase, Reversed-Phase Liquid Chromatography/Mass Spectrometry for Shotgun Proteomics. *Journal of Proteome Research* **2008,** 7, 1007-1011.

21. Fournier, M. L.; Gilmore, J. M.; Martin-Brown, S. A.; Washburn, M. P., Multidimensional Separations-Based Shotgun Proteomics. *Chemical Reviews* **2007,** 107, 3654-3686.

22. Tian, Z.; Zhao, R.; Tolić, N.; Moore, R. J.; Stenoien, D. L.; Robinson, E. W.; Smith, R. D.; Paša-Tolić, L., Two-dimensional liquid chromatography system for online top-down mass spectrometry. *PROTEOMICS* **2010,** 10, 3610-3620.

23. Tran, J. C.; Doucette, A. A., Multiplexed Size Separation of Intact Proteins in Solution Phase for Mass Spectrometry. *Analytical Chemistry* **2009,** 81, 6201-6209.

24. Veenstra, T. D. a. Y., J. R, Front Matter. In *Proteomics for Biological Discovery*, John Wiley & Sons, Inc.: 2006; pp i-xvi.

25. Banerjee, S.; Mazumdar, S., Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte. *Int J Anal Chem* **2012,** 282574, 15.

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

27. Przybylski, M.; Glocker, M. O., Electrospray Mass Spectrometry of Biomacromolecular Complexes with Noncovalent Interactions—New Analytical Perspectives for Supramolecular Chemistry and Molecular Recognition Processes. *Angewandte Chemie International Edition in English* **1996,** 35, 806-826.

28. Domon, B.; Aebersold, R., Mass spectrometry and protein analysis. *Science* **2006,** 312, 212-7.

29. Kebarle, P., A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry. *Journal of Mass Spectrometry* **2000,** 35, 804-817.

30. Wilm, M.; Mann, M., Analytical Properties of the Nanoelectrospray Ion Source. *Analytical Chemistry* **1996,** 68, 1-8.

31. Douglas, D. J.; Frank, A. J.; Mao, D., Linear ion traps in mass spectrometry. *Mass Spectrom Rev* **2005,** 24, 1-29.

32. The Nobel Prize in Physics 1989. www.nobelprize.org (March 6, 2013),

33. Hopfgartner, G.; Varesio, E.; Tschappat, V.; Grivet, C.; Bourgogne, E.; Leuthold, L. A., Triple quadrupole linear ion trap mass spectrometer for the analysis of small molecules and macromolecules. *J Mass Spectrom* **2004,** 39, 845-55.

34. March, R. E., An introduction to quadrupole ion trap mass spectrometry. *Journal of Mass Spectrometry* **1997,** 32, 351-369.

35. Makarov, A., Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. *Analytical Chemistry* **2000,** 72, 1156-1162.

36. Makarov, A.; Denisov, E.; Kholomeev, A.; Balschun, W.; Lange, O.; Strupat, K.; Horning, S., Performance Evaluation of a Hybrid Linear Ion Trap/Orbitrap Mass Spectrometer. *Analytical Chemistry* **2006,** 78, 2113-2120.

37. Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Graham Cooks, R., The Orbitrap: a new mass spectrometer. *Journal of Mass Spectrometry* **2005,** 40, 430-443.

38. Perry, R. H.; Cooks, R. G.; Noll, R. J., Orbitrap mass spectrometry: instrumentation, ion motion and applications. *Mass Spectrom Rev* **2008,** 27, 661-99.

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

40. Sadygov, R. G.; Cociorva, D.; Yates, J. R., Large-scale database searching using tandem mass spectra: Looking up the answer in the back of the book. *Nat Meth* **2004,** 1, 195-202.

41. Mitchell Wells, J.; McLuckey, S. A., Collision-Induced Dissociation (CID) of Peptides and Proteins. In *Methods in Enzymology*, Burlingame, A. L., Ed. Academic Press: 2005; Vol. Volume 402, pp 148-185.

42. Frese, C. K.; Altelaar, A. F.; Hennrich, M. L.; Nolting, D.; Zeller, M.; Griep-Raming, J.; Heck, A. J.; Mohammed, S., Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos. *J Proteome Res* **2011,** 10, 2377-88.

43. Mikesh, L. M.; Ueberheide, B.; Chi, A.; Coon, J. J.; Syka, J. E. P.; Shabanowitz, J.; Hunt, D. F., The utility of ETD mass spectrometry in proteomic analysis. *Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics* **2006,** 1764, 1811-1822.

44. Coon, J.; Shabanowitz, J.; Hunt, D.; Syka, J., Electron transfer dissociation of peptide anions. *Journal of The American Society for Mass Spectrometry* **2005,** 16, 880-882.

45. Coon, J. J.; Shabanowitz, J.; Hunt, D. F.; Syka, J. E., Electron transfer dissociation of peptide anions. *J Am Soc Mass Spectrom* **2005,** 16, 880-2.

46. Pekar Second, T.; Blethrow, J. D.; Schwartz, J. C.; Merrihew, G. E.; MacCoss, M. J.; Swaney, D. L.; Russell, J. D.; Coon, J. J.; Zabrouskov, V., Dual-

Pressure Linear Ion Trap Mass Spectrometer Improving the Analysis of Complex Protein Mixtures. *Analytical Chemistry* **2009,** 81, 7757-7765.

47. Horn, D. M.; Zubarev, R. A.; McLafferty, F. W., Automated reduction and interpretation of high resolution electrospray mass spectra of large molecules. *Journal of The American Society for Mass Spectrometry* **2000,** 11, 320-332.

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

49. Yang, F.; Shen, Y.; Camp, D. G.; Smith, R. D., High-pH reversed-phase chromatography with fraction concatenation for 2D proteomic analysis. *Expert Review of Proteomics* **2012,** 9, 129-134.

50. Guryca, V.; Kieffer-Jaquinod, S.; Garin, J.; Masselon, C. D., Prospects for monolithic nano-LC columns in shotgun proteomics. *Anal Bioanal Chem* **2008,** 392, 1291-7.

51. Shen, Y.; Smith, R. D.; Unger, K. K.; Kumar, D.; Lubda, D., Ultrahigh-throughput proteomics using fast RPLC separations with ESI-MS/MS. *Anal Chem* **2005,** 77, 6692-701.

52. Tao, D.; Zhu, G.; Sun, L.; Ma, J.; Liang, Z.; Zhang, W.; Zhang, L.; Zhang, Y., Serially coupled microcolumn reversed phase liquid chromatography for shotgun proteomic analysis. *PROTEOMICS* **2009,** 9, 2029-36.

53. Vellaichamy, A.; Tran, J. C.; Catherman, A. D.; Lee, J. E.; Kellie, J. F.; Sweet, S. M.; Zamdborg, L.; Thomas, P. M.; Ahlf, D. R.; Durbin, K. R.;

Valaskovic, G. A.; Kelleher, N. L., Size-sorting combined with improved nanocapillary liquid chromatography-mass spectrometry for identification of intact proteins up to 80 kDa. *Anal Chem* **2010,** 82, 1234-44.

54. Nakamura, T.; Kuromitsu, J.; Oda, Y., Evaluation of comprehensive multidimensional separations using reversed-phase, reversed-phase liquid chromatography/mass spectrometry for shotgun proteomics. *J Proteome Res* **2008,** 7, 1007-11.

55. Boyne, M. T.; Garcia, B. A.; Li, M.; Zamdborg, L.; Wenger, C. D.; Babai, S.; Kelleher, N. L., Tandem mass spectrometry with ultrahigh mass accuracy clarifies peptide identification by database retrieval. *J Proteome Res* **2009,** 8, 374-9.

56. Horn, D. M.; Zubarev, R. A.; McLafferty, F. W., Automated de novo sequencing of proteins by tandem high-resolution mass spectrometry. *Proc Natl Acad Sci U S A* **2000,** 97, 10313-7.

57. Altelaar, A. F.; Mohammed, S.; Brans, M. A.; Adan, R. A.; Heck, A. J., Improved identification of endogenous peptides from murine nervous tissue by multiplexed peptide extraction methods and multiplexed mass spectrometric analysis. *J Proteome Res* **2009,** 8, 870-6.

58. Swaney, D. L.; McAlister, G. C.; Wirtala, M.; Schwartz, J. C.; Syka, J. E.; Coon, J. J., Supplemental activation method for high-efficiency electron-transfer dissociation of doubly protonated peptide precursors. *Anal Chem* **2007,** 79, 477-85.

59. Good, D. M.; Wirtala, M.; McAlister, G. C.; Coon, J. J., Performance characteristics of electron transfer dissociation mass spectrometry. *Mol Cell Proteomics* **2007,** 6, 1942-51.

60. Ramsby, M. L.; Makowski, G. S., Differential Detergent Fractionation of Eukaryotic Cells. In 1998; Vol. 112, pp 53-66.

61. Wessel, D.; Flugge, U. I., A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal Biochem* **1984,** 138, 141-3.

62. Doucette, A. A.; Tran, J. C.; Wall, M. J.; Fitzsimmons, S., Intact proteome fractionation strategies compatible with mass spectrometry. *Expert Rev Proteomics* **2011,** 8, 787-800.

63. Gygi, S. P.; Aebersold, R., Mass spectrometry and proteomics. *Curr Opin Chem Biol* **2000,** 4, 489-94.

64. van Niel, G.; Porto-Carreiro, I.; Simoes, S.; Raposo, G., Exosomes: a common pathway for a specialized function. *J Biochem* **2006,** 140, (1), 13-21.

65. Witwer, K. W.; Buzás, E. I.; Bemis, L. T.; Bora, A.; Lässer, C.; Lötvall, J.; Nolte-‘t Hoen, E. N.; Piper, M. G.; Sivaraman, S.; Skog, J.; Théry, C.; Wauben, M. H.; Hochberg, F., *Standardization of sample collection, isolation and analysis methods in extracellular vesicle research*. 2013.

66. Xiang, X.; Poliakov, A.; Liu, C.; Liu, Y.; Deng, Z. B.; Wang, J.; Cheng, Z.; Shah, S. V.; Wang, G. J.; Zhang, L.; Grizzle, W. E.; Mobley, J.; Zhang, H. G., Induction of myeloid-derived suppressor cells by tumor exosomes. *Int J Cancer* **2009,** 124, 2621-33.

67. Ostrand-Rosenberg, S.; Sinha, P., Myeloid-derived suppressor cells: linking inflammation and cancer. *J Immunol* **2009,** 182, 4499-506.

68. Sinha, P.; Okoro, C.; Foell, D.; Freeze, H. H.; Ostrand-Rosenberg, S.; Srikrishna, G., Proinflammatory S100 proteins regulate the accumulation of myeloid-derived suppressor cells. *J Immunol* **2008,** 181, 4666-75.

69. Ichikawa, M.; Williams, R.; Wang, L.; Vogl, T.; Srikrishna, G., S100A8/A9 activate key genes and pathways in colon tumor progression. *Mol Cancer Res* **2011,** 9, (2), 133-48.

70. Tian, Z.; Tolic, N.; Zhao, R.; Moore, R. J.; Hengel, S. M.; Robinson, E. W.; Stenoien, D. L.; Wu, S.; Smith, R. D.; Pasa-Tolic, L., Enhanced top-down characterization of histone post-translational modifications. *Genome Biol* **2012,** 13.