INTEGRATION OF TANDEM MASS SPECTROMETRY AND ION MOBILITY SPECTROMETRY FOR PROTEIN CHARACTERIZATION AND STRUCTURAL ANALYSIS

by

Deepali Rathore

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Chemistry

Under the Supervision of Professor Eric D. Dodds

Lincoln, Nebraska

May, 2016

INTEGRATION OF TANDEM MASS SPECTROMETRY AND ION MOBILITY SPECTROMETRY FOR PROTEIN CHARACTERIZATION AND STRUCTURAL ANALYSIS

Deepali Rathore, Ph.D.

University of Nebraska, 2016

Advisor: Eric D. Dodds

Mass spectrometry (MS) based proteomics and intact protein analyses are important tools for the structural study of proteins and provide powerful methods for solving biochemical puzzles involving proteins. The work described in this dissertation is aimed at the development of novel, efficient, and information rich strategies for protein structure and sequence analysis. The approaches developed have been applied to analytes ranging from proteolytic peptides to large non-covalent protein complexes.

Chapter I provides a comprehensive comparison of the MS approaches in existence in present day proteomics along with their implementation, suitability, and complementarity for effective proteome analysis. Ion dissociation methods, which serve as the foundation for tandem mass spectrometry (MS/MS) experiments, are next detailed in Chapter II. This chapter addresses the strengths and limitations of each ion dissociation method for MS/MS in light of the fundamentals of the dissociation processes involved. Mechanisms of peptide fragmentation chemistry and various combinations of these techniques with instrumental designs are also described.

MS/MS, despite being a broadly applied tool for protein characterization, often affords only a portion of the extractable information. Chapter III focuses on the orthogonal application of two complementary ion dissociation methods, *viz.* collision induced dissociation (CID) and electron transfer dissociation (ETD), coupled in conjunction with ion mobility spectrometry (IMS), a rapid gas-phase separation technique compatible with MS and MS/MS experiments. The approach has dual benefits of enhanced sequence coverage and more efficient use of the available ion population. The strategy is then escalated to a multidimensional protein complex analysis by merging the tools of native MS, top down MS/MS, and proteomics. Chapter IV demonstrates a method which combines multiple stages of CID bridged by IMS and targets non-covalent protein assemblies in a manner that allows subunits to be released and activated in a controlled stepwise fashion, to be separated based on their size and charge, and finally to yield subunit identities a single gas-phase analysis. In addition, IMS opens avenues for measurement of ion-neutral collisional cross sections of the analyte at each stage, allowing higher order structural analysis. This uniquely information-rich analysis method is the subject of Chapter V.

PERTINENT PUBLICATIONS

This dissertation is based in part upon the following publications, which have been reproduced with permission from the appropriate copyright holders, where applicable:

Chapter II

Rathore, D.; Aboufazeli, F.; Huang, Y.; Kolli, V.; Fernando, G. S.; Dodds, E. D.; Ion dissociation Methods in Proteomics. *Encyclopedia of Analytical Chemistry* 2015, 1-26.

Chapter III

Rathore, D.; Aboufazeli, F.; Dodds, E. D.; Obtaining Complementary Polypeptide Sequence Information from a Single Precursor Ion Packet via Sequential Ion Mobility-Resolved Electron Transfer and Vibrational Activation. *Analyst* 2015, 140, 7175-7183.

Chapter IV

Rathore, D.; Dodds, E. D.; Collision-Induced Release, Ion Mobility Separation, and Amino Acid Sequence Analysis of Subunits from Mass-Selected Noncovalent Protein Complexes. *Journal of The American Society for Mass Spectrometry* 2014, 25, 1600-1609.

Chapter V

Rathore, D.; Aboufazeli, F.; Dodds, E. D.; Collision Cross Sections and Primary Sequence Information of Noncovalent Protein Complexes Studied by Integrated Collision–Induced Dissociation and Ion Mobility. *In preparation.*

ACKNOWLEDGEMENTS

There are many individuals I'd like to thank for making possible my journey through the graduate school. First and foremost I express my heartfelt thanks to my advisor, Dr. Eric D. Dodds. It has been an honor to be his first Ph.D. student. His consistent and patient guidance, motivation for research and contagiously positive outlook during tough times in my Ph.D. pursuit have had immense impact on me. He has provided me opportunities and resources to grow as a researcher, generously sent me to several conferences to present my results and has always been supportive of my endeavors to pursue various projects. From him, I have learned a lot and deeply appreciate his mentorship. I'd also like to thank my supervisory committee members: Prof. Robert Powers, Prof. David Hage, Dr. Jiantao Guo and Dr. Mark Wilson for their guidance throughout my graduate career. The insightful discussions at committee meetings have certainly broadened my horizon in the field of bio-analytical chemistry.

I express my gratitude to Dr. Kurt Wulser for providing considerable assistance in the laboratory with his expertise in instrumentation. When things went wrong and I was clueless as to why, he made sure that my experiments ran smoothly. I also thank the Department of Chemistry for awarding me travel fellowships and the very supportive staff for helping me along the way.

I'd like to acknowledge members of Dodds research group: Venkata Karthik Kolli, Yuting Huang, Forouzan Aboufazeli, Katie Schumacher and Abby Gelb. You have not only been great coworkers but wonderful friends. The fun times we had chatting in the office or working collaboratively on projects made my Ph.D. journey memorable and enjoyable. Your support, good advice, and camaraderie will always be cherished. I'd also

like to thank the undergraduate students I had the opportunity to mentor and work with: John-Emmett Mahon and Becca Jarrett for their help and keeping the laboratory atmosphere lively with their enthusiasm. My stay at UNL was also enriched by the Campus Recreation Center. The discipline and hard work associated with physical fitness has influenced my life greatly.

Last but not the least, I would like to thank my parents: my father Pratap Singh, my mother Murti Devi and my brother Aashish Rathore for their unwavering support throughout my entire life. Papa, I will always be thankful to you for encouraging me to pursue my passions without any boundaries. Ma, your contribution in shaping my life cannot be described in words but I am ever inspired by your indomitable spirit and hard-working attitude. Aashu, having a sibling like you is a blessing in every way. Despite being younger, you never stopped surprising me with your wisdom, always motivated me to keep moving forward and stood by me during hard times. I feel fortunate to have a family like you. This thesis is dedicated to you.

TABLE OF CONTENTS

CHAPTER I

MASS SPECTROMETRIC APPROACHES IN PROTEOMICS AND PROTEIN CHARACTERIZATION

1.1 ABSTRACT

Proteomics play a significant role in our understanding of the molecular basis of biological processes, and thus has contributed immensely to major research areas such as targeted drug discovery, biomarker identification and systems biology. Traditional protein identification and structural characterization approaches have had significant impact, these methods each bear inherent analytical strengths and limitations. In recent years, mass spectrometry (MS) has emerged as an effective complement to classical methods, and has established itself as an integral part of protein science. Several MS based approaches are available and each can be distinguished by the level at which the analysis is carried out (peptide or protein). Selection of an approach should be based on type of information desired. The majority of proteomics experiments rely on the digestion-based bottom up approach due to the ease of implementation and high proteome coverage. Top down experiments enable analysis of intact proteins followed by their fragmentation, providing amino acid sequence and preserving information that is often lost in other methods (*e.g.*, mapping of post-translational modification patterns). Middle down proteomic analysis is a hybrid of the bottom up and top down strategies, and allows study of higher molecular mass analytes than is possible in bottom up methods while offering some advantages of top down approaches. Despite its utility in protein identification, top down analysis does not provide information on non-covalent interactions persisting in the native protein complexes. In that regard, native MS is especially useful as a structural tool in proteomics and permits higher order protein assemblies to be analyzed in a form that more closely relates to their biological state.

1.2 INTRODUCTION

Proteomics refers to the large-scale study of proteins involving analysis of their structures and functions. The field is diverse with several branches such as protein quantification, protein sequence analysis, structural proteomics, interaction studies, cellular proteomics, and experimental bioinformatics. A proteome is the entire set of proteins produced by an organism throughout its life cycle and bears more complexity than a genome. Post-translational modifications (PTMs) (*e.g.*, glycosylation, phosphorylation, methylation, *etc.*) and different patterns of gene expression confer upon proteins even more heterogeneity and dynamism. Since proteins play a central role in the cellular functions in any organism, it becomes of value to probe their structural aspects, expression profiles, and interaction modes. The proteomic complexity found in organisms also involves specific higher order protein structures and networks of protein-protein interactions. Protein-protein interactions in *Saccharomyces cerevisiae* have been estimated to be around 35,000, corresponding to roughly six interactions per protein [1-2]. In humans, the number is estimated to be even higher: 150,000-650,000 interactions from 26,000-38,000 protein coding genes [3-4]. This indicates that, in addition to the sequence and higher order structure, the activity of proteins relies on their interaction networks and how they relate to biological function. Some key characteristics in this regard include interaction stoichiometry, interaction sites, identities of interacting proteins, *etc.*, of the protein assembly.

Various approaches have been implemented for the study of protein complexes and for characterizing their interactions. A few traditional methods are X-ray crystallography

[5-8], nuclear magnetic resonance (NMR) spectroscopy [9-12], and electron microscopy (EM) [13-14]. Each of these methods are capable of probing protein structure at the atomic level owing to their high level of resolution. Techniques like ultracentrifugation, gel electrophoresis and size exclusion chromatography can provide molecular mass, equilibrium constants, and stoichiometry. However, each of these techniques has specific limitations. For example, NMR and X-ray crystallography require significant quantities (milligram scale) of pure proteins and require substantial preparation and analysis time. Although the secondary structure of a protein as large as 110 kDa has been determined by NMR [15], the technique is mostly limited to ~40 kDa [16-17]. Similarly, not all proteins are able to be crystallized for X-ray crystallography.

Mass spectrometry (MS) has emerged as a powerful complement to more traditional methods of protein structure characterization, and is increasingly becoming the preferred method for the study of proteomes. While MS does not provide atomic-level structural information for intact proteins, MS can provide important stoichiometry information (even in heterogeneous systems) and accurate mass measurement within relatively short time frames and while utilizing minute sample amounts (as low as attomoles) [18]. Development of new MS based experimental approaches has made it possible to analyze a large number of proteins, in line with the ultimate goal of many experiments in proteomics [19-20]. Complex protein mixtures are now routinely characterized using LC (liquid chromatography) coupled with tandem mass spectrometry (MS/MS) [21-26]. Although a number of technical challenges persist in the field of high-throughput proteomics, advances in informatics, hyphenated analytical approaches, and the continued evolution of high performance mass spectrometers are laying the path towards

increasingly high-quality proteomic data that will increase our comprehension of complex biological processes [27-31]. Many biological processes of interest are realized by way of protein-protein or protein-ligand interactions, thus making a number of questions highly relevant and worthy of answering in order to gain insight into biochemical systems. These questions include, for instance, what other proteins bind to a newly expressed protein to make it operative? What other substance does a protein interact with in order to carry out a given function? What is the molecular nature or these protein-protein and protein-ligand interactions? What are the binding affinities between the ligand and target protein in a functional assembly?

Currently, there are two main approaches for proteome investigation by MS that are fairly complementary to each other: the "top down" approach and the "bottom up" approach. Each has established its own realm and deserves a detailed discussion. Besides these two, a lesser known but evolving, intermediate approach known as "middle down" proteomics also merits mention. These approaches are summarized in **Figures 1.1-1.2** [32-33], and are further discussed in the succeeding sections.

1.3 BOTTOM UP PROTEOMICS

Traditionally, MS based proteomics has been practiced in a manner known as "bottom up" proteomics [20, 34]. The approach requires protein of interest to be chemically cleaved (*e.g.* disulfide bond reduction and alkylation; CNBr or formic acid treatment) and / or enzymatically digested (*e.g.* using trypsin, chymotrypsin, LysC, AspN, etc.) before their introduction for MS analysis. The released peptides are then analyzed by MS using

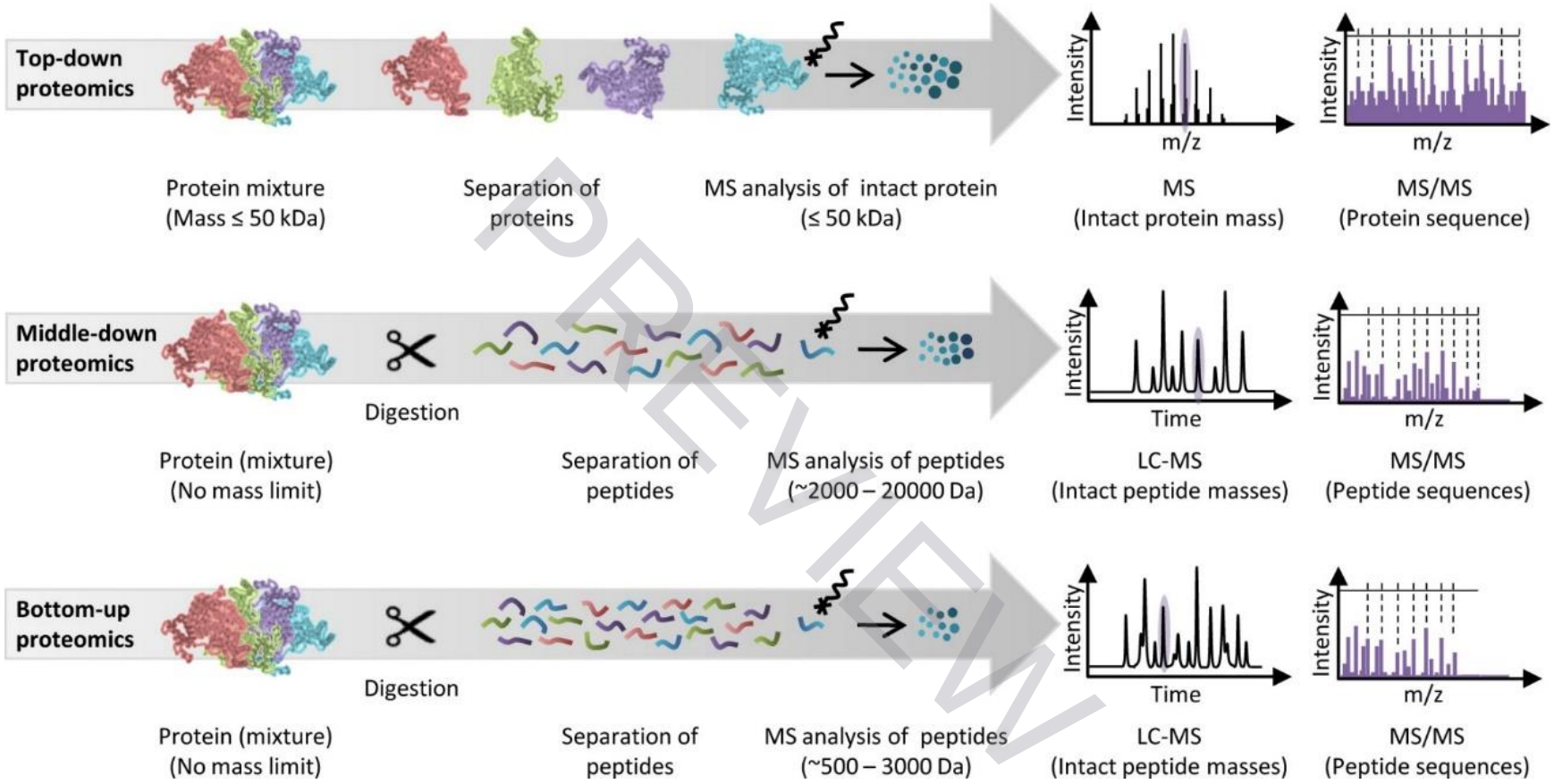

Figure 1.1 Overview of proteomic strategies: top-down *vs.* middle-down *vs.* bottom up. The bottom-up approach analyzes proteolytic peptides. The top-down method measures the intact proteins. The middle-down strategy analyzes larger peptides resulting from limited digestion or more selective proteases. Reproduced with permission from Switzar *et al.* [32], copyright 2013 American Chemical Society.

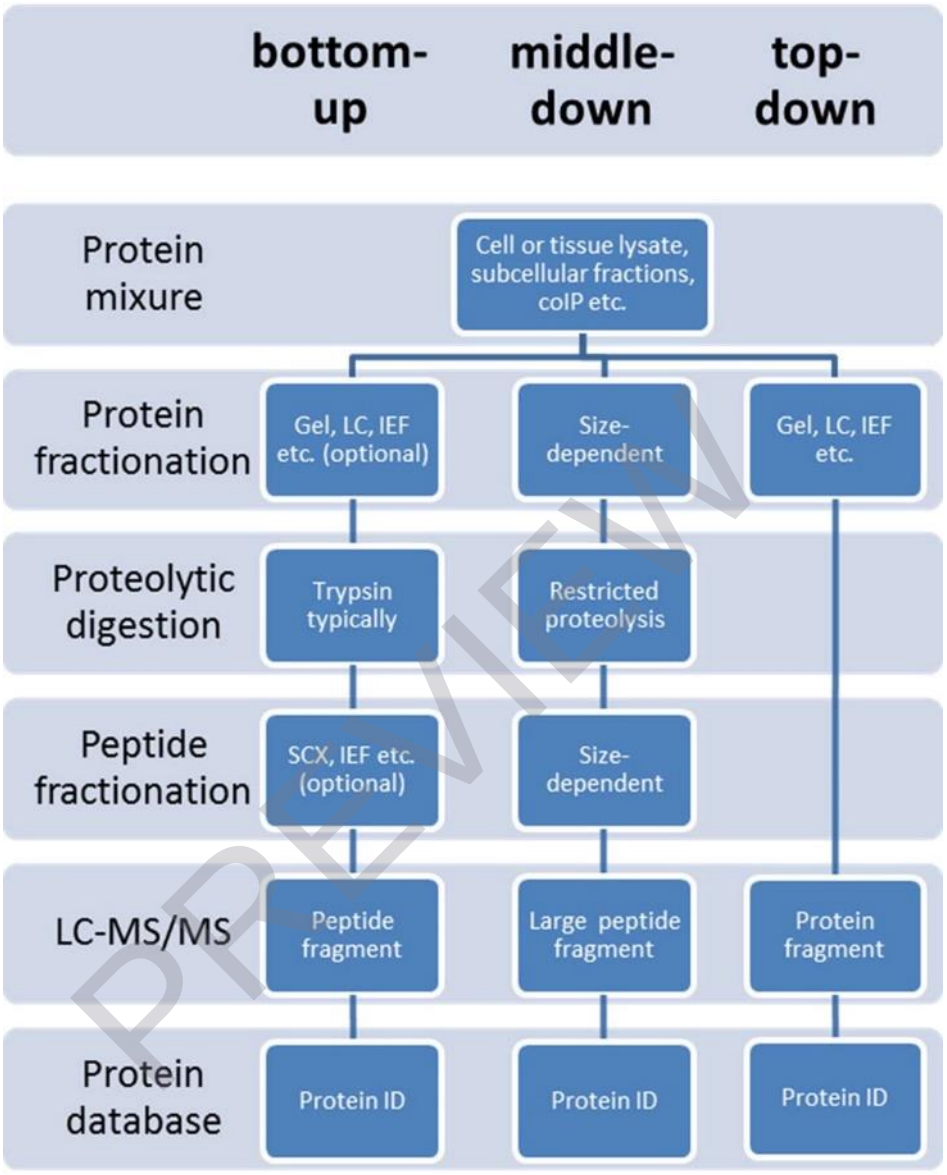

Figure 1.2 Sample preparation workflows for different proteomic strategies: bottom-up *vs.* middle-down *vs.* top-down. One or more protein or peptide fractionation techniques can be applied prior to MS analysis and database searching. Reproduced with permission from Zhang *et al.* [33], copyright 2013 American Chemical Society.

ionization methods such as electrospray ionization (ESI) and matrix-assisted laser desorption / ionization (MALDI). These allow peptide ions to be transferred in to the gas phase without unintentional fragmentation. The digested peptides can be analyzed in intact form and can also be subjected to further fragmentation in MS/MS experiments to yield data on their sequence as well as modifications. These two lines of information furnish sufficient inferences for the identification of the parent protein. The method is especially useful because proteolytic peptides are generally more convenient analytes as opposed to the intact proteins, which can pose greater challenges in sample preparation, ionization, and may not be accessible to some types of MS instruments that lack sufficient resolution or upper mass limit for detection. A term has been coined for when the bottom up approach is performed on a mixture of proteins, *viz.* 'shotgun proteomics' by Yates lab [35-37]. A typical shotgun proteomics experiment is accomplished by fractionating the peptide mixture followed by the LC/MS-MS analysis. Tandem mass spectra can be collected for as many peptides as possible which are then compared with theoretically generated spectra. Matching of peptide sequences against protein sequence database is done for their assignment to a certain protein. While shotgun proteomics is a sensitive method for identifying parts of protein, it poses ambiguity in identification of the protein itself. Since small pieces of peptides are not very specific to an individual protein and can be shared by more than one protein, it may lead to the inaccurate grouping and scoring. Large portions of a protein could remain unidentified if they contain post translational modifications, since chances are that they are already lost during digestion process, leaving behind important information. The bottom up approach, therefore, may not prove optimum for determining specific modification patterns and unique sequence variations due to either an inability to

cover the complete protein sequence, or an inability to reconstruct individual protein isoforms from the pool of proteolytic fragments. The missed pieces of information may be critical to a complete molecular understanding of protein structure and function, hence the need for an alternate approach in order to alleviate those limitations. Nonetheless, bottom up proteomics remains a widely used method for protein characterization by allowing high-throughput, sensitive, and efficient analyses of complex mixtures along with high proteome coverage.

1.4 TOP DOWN PROTEOMICS

Top down mass spectrometry refers to the intact analysis of proteins and usually begins with upfront separation of complex protein mixtures. Fractionation methods like reversed-phase high-performance liquid chromatography (RP-HPLC) ensure a high level of purification of complex samples prior to their introduction to the mass spectrometer [38-40]. The eluted fractions are then infused into the instrument in either an on-line or off-line manner. ESI, being a soft ionization method is well suited for top down analysis and allows larger proteins to be transferred intact into the gas phase. Next, MS/MS is performed in a data-dependent or targeted manner and spectra are analyzed with the help of specialized software as well as manually. This approach allows both intact and fragment ion masses to be measured and overcomes the limitations inherent to bottom up analysis, in that patterns of PTMs are preserved and protein isoforms can be distinguished. With sufficient number of useful fragments, top down can provide complete primary sequence characterization and at the same time reveal modifications on the protein [41-46]. Intact proteins measuring over 200 kDa have been successfully fragmented to obtain considerable sequence information

[47], although relatively large amounts of energy need to be supplied in such cases (*i.e.*, in order to bring about fragmentation in MS/MS experiments). High-resolution mass analyzers, such as Fourier transform ion cyclotron resonance (FTICR) MS instruments, are especially advantageous in allowing high mass and high charge state precursor and fragment ions to be measured [48-52]. However, these instrument are also more expensive and somewhat less user friendly compared to the bench-top instruments, making top down analyses less routine as compared to bottom up MS.

Top down proteomic analyses have been performed on triple quad (QQQ) and quadrupole time-of-flight (QTOF) type configurations as well [53], with both in-source and collision-induced dissociation (CID) based fragmentations, resulting in interpretable sequence tags [54]. Ion-trap and Orbitrap type instruments have also seen developments done for top down MS analysis [55-56]. Initially, the majority of top down proteomics experiments utilized low-energy CID; however, over the last decade, alternate dissociation techniques have come into light that complement CID in beautiful ways. Electron capture dissociation (ECD), developed by McLafferty and coworkers [57], and electron transfer dissociation (ETD), developed by Hunt laboratory [58], have significantly advanced the capabilities of top down analyses. These ion-electron reaction based methods, when coupled with appropriate instrumentation, allow for the detection for labile PTMs and fragmentation of high molecular weight proteins [59-62]. Ion dissociation methods for MS/MS as applied to problems in proteomics and protein characterization are discussed in depth in Chapter II.

Another area for improvement in top down MS analysis deals with ionization of bulky protein molecules. Higher charge states can be especially beneficial with respect to

fragmentation processes such as ECD and ETD, since these display increased efficiency with more charges on the precursor ion. Ways to manipulate charge state distributions include addition of "supercharging agents" to the sample buffer or solvent systems [63-66]. Reagents such as *m*-nitrobenzyl alcohol and glycerol among many others have been shown to increase the charge state and have been popular supercharging agents [67-68].

Despite the current challenges, the potential to achieve near complete identification of individual proteoforms has distinguished top down MS as an extremely useful method and a powerful complement to digestion-based proteomic approaches [53, 69-71].

1.5 MIDDLE DOWN PROTEOMICS

Middle down proteomics, as the name suggests, refers to a hybrid version of top down and bottom up and is a not an entirely new concept [72-73]. The mass range of analytes is higher than suitable for bottom up (greater than 3000 Da), meaning longer polypeptides and larger regions of the analyte protein can be detected. This can be useful for mapping combinatorial patterns of PTMs, even if the PTMs occur on distant residues. For this purpose, proteins are lysed by enzymes known to target less abundant amino acid residues (*e.g.* AspN and GluC) compared to proteases such as trypsin. The generated polypeptides (3-20 kDa) are then separated and subjected to similar workflow as top down for protein identification. Work by Karger *et al.* shows that this approach has the potential to become a high-throughput platform for proteomics experiments. Middle down gives better sequence coverage than bottom up [74] in addition to being compatible with chromatographic workflows without compromising the sensitivity. Improved confidence in protein identification puts this approach at an advantage over top down. High-resolution