Title
Developing Top-Down Mass Spectrometry for Intact Protein Identification in the Chromatographic Timescale

Permalink
https://escholarship.org/uc/item/16f1599t

Author
Lakshmanan, Rajeswari

Publication Date
2012

Peer reviewed|Thesis/dissertation
Developing Top-Down Mass Spectrometry for Intact Protein Identification in the Chromatographic Timescale

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Biochemistry and Molecular Biology

by

Rajeswari Lakshmanan

2012
ABSTRACT OF THE DISSERTATION

Developing Top-Down Mass Spectrometry for Intact Protein Identification in the Chromatographic Timescale

by

Rajeswari Lakshmanan

Doctor of Philosophy in Biochemistry and Molecular Biology
University of California, Los Angeles, 2012
Professor Joseph A. Loo, Chair

Protein identification by top-down mass spectrometry based methods yield intact mass of the proteins and indicate the presence of post-translational modifications (PTMs) and/or isoforms. Currently, the methods employed for top-down protein identification are performed using instruments with dual mass analyzers and are based on fragmenting isolated charge states, which greatly reduces the duty cycle of the instrument. High throughput top-down methods are required for protein identification in complex sample mixtures. We demonstrate the capability to perform intact protein identifications in a single-stage time-of-flight mass spectrometer during protein elution from a liquid chromatography (LC) column. In addition, we have developed a new data-independent fragmentation method known as ‘Continuous Accumulation of Selected Ions–Collisionally Activated Dissociation’ (CASI-CAD) to fragment multiple charge states of the protein simultaneously for the purpose of identification in the LC timescale. CASI-CAD is performed without any precursor selection and thus, the duty-cycle of the instrument is not
lowered. Both these methods unambiguously identified all the proteins in the human proteasome complex used for method development. The presence of PTMs and N-terminal modifications were also characterized for the proteins in this complex.

Supercharging reagents are known for their ability to enhance the multiple charging of proteins during electrospray ionization (ESI). This improves the mass measurement accuracy and fragmentation efficiency of proteins during ESI-MS. Currently, the mechanism behind supercharging is unknown. We have analyzed different supercharging reagents under a variety of solvent conditions to probe the mechanisms behind supercharging. In addition, the supercharging ability of sulfolane was utilized for proteins eluting from a column by adding the reagent to the LC solvents. Furthermore, reagent introduction in the vapor phase increased the signal intensity for intact proteins eluting from a column when compared to experiments performed without the reagent. These methods presented here are efficient top-down means to address complex samples in the chromatographic timescale.
The dissertation of Rajeswari Lakshmanan is approved.

Catherine F. Clarke
Julian P. Whitelegge
Steven G. Clarke
Thomas G. Graeber
Joseph A. Loo, Committee Chair

University of California, Los Angeles
2012
DEDICATION

To my wonderful parents and loving husband
# TABLE OF CONTENTS

LIST OF FIGURES ................................................................................................................ xi
LIST OF TABLES .................................................................................................................. xiv
LIST OF SUPPLEMENTAL FIGURES .................................................................................. xv
ACKNOWLEDGMENTS ....................................................................................................... xvi
VITA ..................................................................................................................................... xx

CHAPTER 1: Introduction ....................................................................................................... 1

1.1. Mass spectrometry based proteomics ......................................................................... 1
1.2. Top-Down and Bottom-Up MS .................................................................................... 1
1.3. Protein identification and characterization by TD ......................................................... 3
1.4. A short chronological history on the development of TD methods ......................... 4
1.5. Mass analyzers for TD proteomics .............................................................................. 6
   1.5.1. FT-ICR MS ........................................................................................................ 6
   1.5.2. Orbitrap ............................................................................................................. 8
   1.5.3. Quadrupole time-of-flight .................................................................................. 9
   1.5.4. Ion trap ........................................................................................................... 12
1.6. Intact protein fragmentation schemes ....................................................................... 13
   1.6.1. Collisionally activated dissociation .................................................................... 13
   1.6.2. Higher energy collision dissociation ................................................................. 14
   1.6.3. Electron capture dissociation ........................................................................... 14
   1.6.4. Electron transfer dissociation ........................................................................... 15
   1.6.5. Other miscellaneous fragmentation methods .................................................... 17
1.7. Intact protein separation techniques ......................................................................... 17
   1.7.1. Electrophoresis ............................................................................................... 17
CHAPTER 3: Top-Down Protein Identification of Proteasome Complex Proteins with nano LC FT-ICR MS Employing Data-Independent Fragmentation Methods

3.1. Abstract

3.2. Introduction

3.3. Experimental section

3.4. Results and discussion

3.5. Conclusions

3.6. References
CHAPTER 4: Reagents for Enhancing Protein Charging for Electrospray Ionization Mass Spectrometry from Organic/Aqueous Solvent Systems

4.1. Abstract

4.2. Introduction

4.3. Experimental section

4.3.1. Sample preparation

4.3.2. Mass spectrometry

4.3.3. Calculation of average and maximum charges

4.4. Results and discussion

4.4.1. Reagents for supercharging under denaturing conditions

4.4.2. Compounds effective only in denaturing conditions

4.4.3. Effect of different solvents in the supercharging efficiency of m-nitrophenol

4.4.4. Supercharging disulfide bonded proteins

4.4.5. Supercharging high MW proteins

4.4.6. Supercharging membrane proteins

4.4.7. Supercharging peptides

4.4.8. Reagent concentration versus supercharging

4.4.9. Protein conformation during supercharging

4.5. Conclusions

4.6. References

CHAPTER 5: Enhanced Multiple Charging for Protein LC-MS in the Chromatographic Timescale Utilizing Supercharging Reagents
5.3.2. LC conditions and MS parameters ................................................................. 182
  5.3.2.1. Supercharging during LC-MS by adding reagents directly to LC solvents 182
  5.3.2.2. Supercharging during LC-MS by delivering reagent vapors ................. 183
  5.3.3. Calculation of average and maximum charges ........................................ 184
5.4. Results and discussion ...................................................................................... 185
  5.4.1. Supercharging proteins during LC-MS with sulfolane in the LC solvent .. 185
  5.4.2. Supercharging proteins by spotting reagents in the CaptiveSpray air filter .. 187
  5.4.3. Supercharging during LC timescale using reagent vapors ....................... 188
5.4. Conclusions ..................................................................................................... 190
5.5. References ...................................................................................................... 191
CHAPTER 6: Conclusions and Future Directions......................................................... 203
  6.1. Overview ...................................................................................................... 203
  6.2. Top-Down protein identification using a single-stage LC-ESI-TOF MS .............. 204
  6.3. Top-Down protein identification using data-independent fragmentation in FT-ICR MS 205
  6.4. Supercharging reagents for enhancing charge during protein elution from HPLC .. 209
  6.5. Concluding remarks ................................................................................... 211
APPENDIX TO THE DISSERTATION ........................................................................ 213
LIST OF FIGURES

Figure 2.1. Schematic representation of the top-down identification of proteasome complex proteins using LC-ESI-TOF MS ................................................................. 70

Figure 2.2. Mass spectra of BSA samples at different fragmentor voltages to compare the efficiency of fragmentation.............................................................. 71

Figure 2.3. Homology between the N-terminal 40 amino acid residues in albumins from different species............................................................... 72

Figure 2.4. Nozzle-skimmer dissociation of albumins from different species ........................................... 73

Figure 2.5. Fragment ion assignments in mass spectra of albumins from different species .... 74

Figure 2.6. TIC of a standard 7 protein mix run through C3 column ........................................... 75

Figure 2.7. Fragmentation of proteins in the standard 7 protein mix at high NS voltage during elution from the column ........................................... 76

Figure 2.8. Nozzle-skimmer dissociation of superoxide dismutase C during its elution from the LC column ........................................... 77

Figure 2.9. TIC of h20S sample run through C3 column ........................................................... 78

Figure 2.10. Nozzle-skimmer dissociation of β6 subunit during its elution from the LC column 79

Figure 2.11. TIC of h20S sample with the peaks labeled with the corresponding h20S subunits eluting in it ........................................................................... 80

Figure 2.12. Intact mass spectra of all the proteins in the h20S complex identified by LC separation followed by NSD........................................................... 81

Figure 2.13. Deconvoluted mass spectrum of the α3 subunit of h20S proteasome .................. 82

Figure 3.1. Schematic representation of the top-down protein identification using nano LC FT-ICR MS.......................................................... 110

Figure 3.2. TIC of h20S complex run through RP monolithic column ........................................... 111

Figure 3.3. Intact mass spectra of h20S proteins averaged from peaks in the TIC in Figure 3.2 ........................................................................................................ 112

Figure 3.4. Schematic representation of the Bruker 15T FT-ICR MS ........................................... 113

Figure 3.5. FS-CAD of α4 protein....................................................................................... 114

Figure 3.6. CASI-CAD of ubiquitin....................................................................................... 115

Figure 3.7. Fragmentation map of α5 protein identified with CASI-CAD using different ion
accumulation times. .......................................................................................... 116

Figure 3.8. CASI-CAD mass spectra of α5 subunit obtained with different ion accumulation times ............................................................................................................. 117

Figure 3.9. CASI-CAD of α4 protein .................................................................. 118

Figure 3.10. MaxEnt deconvoluted zero-charge mass spectrum of α3 protein ....... 119

Figure 3.11. Comparison of sequence coverage for α4 protein by FS-CAD and FS-CAD with CASI ................................................................................................................ 120

Figure 3.12. Intact mass spectrum of Rpn1, a 100.2 kDa protein in the h19S complex ......... 121

Figure 4.1. Supercharging insulin with different reagents in denaturing conditions ............. 149

Figure 4.2. Supercharging disulfide-intact lysozyme with different reagents in denaturing conditions ............................................................................................................. 150

Figure 4.3. Supercharging disulfide-reduced lysozyme with sulfolane in denaturing conditions ................................................................................................................ 151

Figure 4.4. Supercharging disulfide-intact lysozyme with m-nitrophenol in denaturing and native solvent conditions ........................................................................................................ 152

Figure 4.5. Supercharging disulfide-intact lysozyme with m-nitrophenol in different solvent systems ................................................................................................................ 153

Figure 4.6. Supercharging β-lactoglobulin A and α-lactalbumin with sulfolane ................. 154

Figure 4.7. Supercharging RNase A and RNase B in positive and negative modes with sulfolane ................................................................................................................ 155

Figure 4.8. Supercharging transferrin with different concentrations of sulfolane ............... 156

Figure 4.9. Comparing nozzle-skimmer dissociation of transferrin with and without sulfolane. 157

Figure 4.10. Supercharging BSA with sulfolane ...................................................... 158

Figure 4.11. Supercharging IgG1 with m-nitrophenyl ethanol ........................................ 159

Figure 4.12. Supercharging bacterioopsin with sulfolane in aqueous conditions .................. 160

Figure 4.13. Supercharging bacterioopsin with sulfolane in denaturing conditions ............. 161

Figure 4.14. Supercharging bacterioopsin with sulfolane in chloroform/methanol/1% FA .... 162

Figure 4.15. Supercharging bradykinin with sulfolane in denaturing conditions ................. 163

Figure 4.16. Plots of reagent concentration versus average and maximum charge ............. 164
Figure 4.17. Ion mobility spectrometry analysis of transferrin in denaturing condition ..........165
Figure 4.18. Ion mobility spectrometry analysis of lysozyme in denaturing conditions...........166
Figure 4.19. Ion mobility spectrometry analysis of lysozyme in native solvent conditions ....167
Figure 4.20. Mass spectra of lysozyme acquired during drift time analysis......................168
Figure 4.21. Ion mobility arrival time distributions for selected charge states of lysozyme....169
Figure 5.1. Supercharging reagent delivery through CaptiveSpray air filter .......................193
Figure 5.2. Supercharging reagent delivery in the vapor phase using 100% ACN .................194
Figure 5.3. Supercharging during LC-MS with sulfolane in the LC solvent.......................195
Figure 5.4. Mass spectra of RNase A with sulfolane in the LC solvent ..........................196
Figure 5.5. Mass spectra of BSA with sulfolane in the LC solvent ................................197
Figure 5.6. Mass spectra of lysozyme with supercharging reagents spotted in the air filters...198
Figure 5.7. Supercharging during LC-MS with sulfolane delivery in the vapor phase. ..........199
Figure 5.8. Mass spectra of RNase A with sulfolane delivery in the vapor phase ..............200
Figure 5.9. Mass spectra of ubiquitin with sulfolane delivery in the vapor phase .............201
Figure 5.10. Mass spectra of myoglobin with sulfolane delivery in the vapor phase ..........202
LIST OF TABLES

Table 2.1. ESI $b_n$ product ion $m/z$ for sheep albumin ............................................................. 83

Table 2.2. Summary of h20S proteasome subunits identified by LC separation followed by NSD ............................................................................................................... 84

Table 3.1. Summary of h20S subunits identified by FS-CAD, CASI-CAD, and FS-CAD with CASI techniques ................................................................................................ 122

Table 3.2. Summary of mass accuracies for the proteins in the h20S complex ......................... 123

Table 4.1. Summary of average and maximum charges acquired by supercharging different proteins with various reagents ................................................................. 170

Table 4.2. Summary of average and maximum charges acquired by supercharging high MW proteins and peptides with different reagents ................................................. 171
LIST OF SUPPLEMENTAL FIGURES

Supplemental Figure 4.1. Supercharging calmodulin with sulfolane ........................................ 172
Supplemental Figure 4.2. Supercharging actin with sulfolane ................................................. 173
Supplemental Figure 4.3. Supercharging lysozyme with increasing concentrations of sulfolane in denaturing conditions ................................................................. 174
Supplemental Figure 4.4. Supercharging lysozyme with increasing concentrations of \textit{m}-NBA in denaturing conditions ................................................................. 175
Supplemental Figure 4.5. Supercharging lysozyme with increasing concentrations of \textit{m}-nitrophenyl ethanol in denaturing conditions .......................................... 176
Supplemental Figure 4.6. Supercharging lysozyme with increasing concentrations of \textit{m}-nitrophenol in denaturing conditions ......................................................... 177
I would first like to thank my mentor, Dr. Joe Loo, for the support and encouragement through these years. Joe - The passion I have for mass spectrometry today is truly because of the opportunity to do graduate research under your mentorship. I have learnt a lot from the one-on-one meetings with you. I have always felt more focused, confident, and encouraged after those meetings. Thank you for giving me an opportunity to maintain the Agilent ESI-TOF instrument. It gave me a wonderful opportunity to network with many Agilent Scientists and learn from working with them. I thoroughly enjoyed working with you and Rachel on the top-down and supercharging projects.

Many thanks to Dr. Rachel Loo – It was a great learning experience when I worked on the supercharging project with you. Thank you for all the time you spent to go over (and over) the theories behind supercharging, and explaining the chemistry behind everything. I'm really looking forward to all your supercharging theories and its impact in this field.

I would like to thank my former and current committee members: Dr. James Gober, Dr. Jon Fukuto, Dr. Catherine Clarke, and Dr. Thomas Graeber. Thank you for providing wonderful comments during my orals and midstream. Thank you to Dr. Steve Clarke for giving me an opportunity to rotate in his lab and also for kindly accepting to serve in my committee.

To Dr. Julian Whitelegge – Thank you for providing guidance when I was working with membrane proteins and more recently, for the top-down project. All your suggestions were very valuable and really helpful. And thank you for giving access to ProSight server and also for gladly accepting to serve in my committee when I approached you recently.
Many thanks to Agilent scientist, Dr. Andy Gieschen for help with ESI-TOF and chromatography. Thank you to Dr. Jeremy Wolff and Dr. Guillaume Tremintin at Bruker for answering my questions during the method development in the FT-ICR instrument.

Thank you to Dr. Tom Drake and Dr. James LeBlanc for all the help and support when we collaborated for the brain injury project.

Thank you to Dr. Jane Strouse for all the help during Agilent instrument installation and when I was teaching Chem184.

‘MIC’ family: To Dr. Melissa Sondej – Thank you for both scientific and non-scientific advice, for listening to all my frustrations and making me feel better in the end, for helping network with people in the industry, and for encouraging me constantly. To Rudy Alvarado – Working with you was always fun and enjoyable, and thank you for spending your time in troubleshooting the Dionex nanoLC with me. To Dr. Greg Khitrov – Thank you for answering all the quick questions about the LC columns and general instrument related help.

‘Loo Lab’ family: To Dr. Jimmy Ytterberg – Thank you for teaching me the basics of mass spectrometry. The brain injury project went up to the point of publication just because of the sample prep techniques you had taught me. To Dr. Sabrina Benchaar – Top-down and supercharging were the passions we shared when we worked together. I can’t thank you enough for the time you spent during ASMS when I was practicing for my oral presentation. To Dr. Jonathan Erde – Thank you for answering my questions related to bio-informatics and databases.
Heartfelt thanks to all the great women scientists I made friends with at UCLA: Dr. Cecilia Zurita Lopez, Dr. Deborah Francoleon, Dr. Cathy Kaddis, Dr. Shirley Lomeli, and Hong Nguyen. You gals are wonderful in the perspective you have for life and science.

To my lifetime friends Charumathi Anbalagan and Porkodi Muthu – We made a great ‘trio’ and shared a wonderful friendship during our Masters degree. You gals were the sole reason for my decision to go into research. I can't forget our group studies and the prep for GRE that we did in rain and shine.

Thank you to Kay Andiappan and Jeya Sridharan for being very supportive and encouraging. Thanks for calling/emailing and telling you were thinking about me. You both have inspired me in various ways.

All I'm today is because of my mom and dad’s love and support. Amma – You are my greatest inspiration. The interest I have towards learning is because of the seeds that you sowed during my childhood. Daddy – Thank you for all the help and for financially supporting for everything until my Masters degree. You never said a 'no' for anything I have asked you. You both have always celebrated my academic achievements so much that it made me progress more. To my sister, Seva Lakshmanan – Thank you so much for your love and ways you show it.

To my husband, Athappan – Thanks for all the love, care, support, and most of all, having the same amount of interest as I have towards my Ph.D. I’m indebted to all the things that you did during these 6 years, especially for taking care of the kids single handedly so many times during ASMS and recently, when I wrote my thesis.
To my kids – Though you guys made this Ph.D an uphill journey, I can’t imagine this without you two. Thank you for making me feel that being loved is the most wonderful thing in the world!

Finally, I would like to formally acknowledge Springer Publishing Group for providing copyright access to reprint the Neurocritical Care article, “Metabolic Crisis After Traumatic Brain Injury is Associated with a Novel Microdialysis Proteome”, published on Jan. 2010 in the Appendix to my dissertation.
VITA

1999  Bachelor of Science in Biochemistry  
      University of Madras  
      Chennai, India  

2001  Master of Science in Biochemistry  
      University of Madras  
      Chennai, India  

2007-2011  Teaching Assistant  
            Department of Chemistry & Biochemistry  
            University of California, Los Angeles  
            Los Angeles, CA  

2007 – 2012  Graduate Research Assistant  
              Department of Chemistry & Biochemistry  
              University of California, Los Angeles  
              Los Angeles, CA  

SCHOLARSHIPS AND AWARDS

2001  Received two national level merit scholarships (Graduate Aptitude Test in Engineering and Indian Council of Medical Research – Junior Research Fellowship)  

1999  Received award for academic excellence in Bachelor’s degree in Biochemistry  

PUBLICATIONS AND PRESENTATIONS

Journal Articles


Oral Presentations


Poster Presentations


4. Top-Down MS of Protein Complexes using In-Source Fragmentation and Supercharging: LC-MS$^\text{E}$ Approach using TOF MS. **Lakshmanan, R.**; Peng, I. X; Ferguson, C. N.; Loo, R. O.; Loo, J. A. 58th Annual ASMS Conference on Mass Spectrometry and Allied Topics, Salt Lake City, UT, May 2010.


CHAPTER 1

Introduction

1.1. Mass spectrometry based proteomics

Proteins play a crucial role in defining the cellular structure, assembly, and function in all organisms. The complexity at the protein level is increased due to the co- and post-translational events\(^1\) and the chemical modifications that take place in proteins are not directly evident at the gene level.\(^2\) Additionally, the protein milieu of a cell varies in different cell types and it is also dynamic since it responds to the environmental changes.\(^3, 4\) As a result, the analysis of proteins is more challenging and necessitates technologies with high throughput, specificity, and dynamic range. Mass spectrometry (MS) is one such promising tool and its tremendous growth in the field of proteomics in the last decade confirms its significance. The vast applications of MS from the well-known biomarker field to remote areas such as bio-defense\(^5\) and phylogenetic characterization of microbes\(^6\) portrays the versatility and maturity of the technique. When compared to the other protein analysis techniques, the resolution available in the MS based analysis down to the amino acid level makes it an efficient tool for protein characterization.\(^7\)

1.2. Top-Down and Bottom-Up MS

Depending on whether intact proteins or the peptides that constitute the proteins are being dealt, MS centered proteomic approaches can be categorized as Top-Down (TD) and Bottom-Up (BU) methodologies. In the TD technique, proteins are directly analyzed by MS without any experimental proteolytic processing and this workflow ensures that intact mass of the molecule being studied is obtained.\(^8\) BU methods incorporate a highly specific enzymatic cleavage step prior to MS analyses to break the proteins into its peptide components.\(^9\) The peptides generated can be analyzed by either ‘Peptide Mass Fingerprinting’ (PMF), where only
the mass of the peptides are measured and matched to in-silico peptide masses generated with the same cleavage specificity or both the mass of the peptides and the fragments generated by dissociating the peptides in the gas phase are measured and searched against the database to find the proteins as in ‘shotgun proteomics’. While PMF is suitable for targeted samples and is performed in conventional instruments, the shotgun methodology is more widely applied for complex mixtures and is performed using instruments with high sensitivity and good mass accuracy.

A more popular version of shotgun is ‘MudPIT’ (Multi-Dimensional Protein Identification Technology), where peptides are fractionated by selectivity to two orthogonal stationary phases. This is currently the method of choice and is replacing the traditional ‘GeLC-MS’, in which intact proteins are separated by gel electrophoresis and the ensuing peptides recovered from the gel spots by in-gel digestion are analyzed by tandem MS. The general limitations of GeLC-MS are low recovery of peptides from the gel, multiple proteins being present in one spot, and limited dynamic range.

Ultimately, the proteotypic peptides detected in the BU experiments are searched against the proteomic databases to identify the corresponding proteins. The sequence coverage is seldom 100%, since not all peptides are detected equally well and information about labile post-translational modifications (PTMs) is not preserved. Moreover, isoforms that share the same peptides cannot be differentiated by the BU methods. Though the term ‘isoforms’ has been at times inadvertently used in the literature to refer PTMs, a recent review of the nomenclature has clearly defined that isoforms refer to similar proteins originating from different genes such as, allelic genes giving rise to similar proteins. The terms ‘proteoforms’ or ‘protein species’ have been proposed to indicate chemical changes resulting from alternative splicing, proteolytic processing, and PTMs. Single nucleotide polymorphisms (SNPs) are also
categorized as proteoforms since it is technically a chemical change that takes place within the same gene.\textsuperscript{16}

In spite of the drawbacks that exist in BU proteomics, it is well-suited for cataloguing all the proteins in an organism\textsuperscript{18, 19} and quantitative proteomics approaches are mostly based on BU workflow.\textsuperscript{20-23} In addition, BU is heavily used in the biomarker field where it is more important to compare disease and healthy states, and identify the differentially expressed proteins to trace back the gene that caused a change rather than knowing more about the protein itself.\textsuperscript{24} Shotgun proteomics is commonly applied for such biomarker discovery pipelines but suffers from the inherent disadvantage that isoforms and proteoforms are not efficiently identified.\textsuperscript{25} Even if PTMs are identified by using enrichment techniques in BU experiments, the information about localization of combinatorial modifications and which PTMs belong to which isoforms cannot be envisioned using BU.\textsuperscript{26}

Detecting the presence of the above-mentioned altered forms of proteins is critical since there is mounting evidence about the roles played by them as biomarkers and therapeutic targets.\textsuperscript{27} Switching to TD techniques will overcome these problems since the intact mass obtained in the TD workflow indicates the presence of any altered forms if the observed and theoretical intact masses are in disagreement. Moreover, TD methods also incorporate a gas phase fragmentation technique to interrogate the protein sequence and the information acquired helps to identify and characterize the protein.

1.3. Protein identification and characterization by TD

Though TD methods are generally looked as a special workflow for studying pure proteins for the purpose of characterization, it is being increasingly used for large-scale protein identification.\textsuperscript{28} ‘Identification’ and ‘characterization’ are different types of analysis that can be performed by TD MS. The former refers to finding the comprehensive set of proteins present in
a sample but the sequence coverage obtained for each protein identified remains low while the latter represents in-depth analysis yielding more sequence coverage.

This review will outline the development of TD technique as a tool for targeted protein characterization and also highlight the recent improvements in the area of protein identification for samples of higher complexity. Emphasis will be given to the instrumentation, fragmentation, and intact protein separation methods which are compatible with electrospray ionization (ESI) MS. Finally, the current limitations are stated and future directions are discussed.

1.4. A Short chronological history on the development of TD methods

With the development of ESI, a soft ionization technique that can ionize large, complex, and fragile entities or simply “flying elephants” in John Fenn’s words, intact protein measurement was made possible and opened the door for TD mass spectrometry.\textsuperscript{29, 30} Since then numerous reports on the accurate mass measurements for species as small as tripeptides to macromolecules as large as bovine albumin dimer (133 kDa) using ESI coupled to different types of mass analyzers were reported.\textsuperscript{31} Utilization of MS for examining intact proteins advanced to the next stage when polypeptides were shown to fragment inside the mass spectrometer and primary sequence information could be deduced by measuring the resulting fragments.\textsuperscript{32} In this first “top-down” experiment in 1990 by Loo \textit{et al.}, the multiply charged ions from ribonuclease A (14 kDa) were fragmented by Collisionally Activated Dissociation (CAD) in a triple quadrupole instrument and this approach was later extended to high molecular weight (MW) proteins such as bovine serum albumin (66 kDa).\textsuperscript{33} Feng and Konishi showed that even larger proteins such as antibodies (150 kDa) and complement component (200 kDa) can be analyzed by ESI MS\textsuperscript{34} and they also demonstrated that these multiply charged macrosized antibody ions can be fragmented by CAD in a triple quadrupole instrument.\textsuperscript{35}
Following the TD demonstrations in triple quadrupole instruments, a significant improvement in resolution and mass accuracy was achieved when Fourier Transform Ion Cyclotron Resonance (FT-ICR) instruments were employed.\textsuperscript{36, 37} High resolution is important for the accurate determination of MW of intact proteins from the ESI generated multiple charge states and also for unambiguous assignment of fragment ions during protein characterization. The initial work utilizing a 2.8 Tesla (T) magnet FT instrument yielded unit mass resolution for myoglobin (17 kDa),\textsuperscript{38} followed by a 6.2 T magnet which extended the isotopic resolution to carbonic anhydrase (29 kDa).\textsuperscript{39, 40} In addition to the ultra-high resolving power, the multi-stage tandem MS capability available in the single mass analyzer FT-ICR instrument allowed temporal mass selection\textsuperscript{41} and detected sequence variations as low as single amino acid replacement.\textsuperscript{42} Subsequently, it was shown that the intact mass and sequence tags obtained from tandem MS experiments can be searched against the sequence databases and protein identifications were made possible.\textsuperscript{43} With further advancements in the magnetic field strength and data sampling techniques, unit mass resolution for chondroitinase I (112 kDa) enzyme was obtained with a 9.4 T FT-ICR instrument.\textsuperscript{44}

The term “top-down” was put forth by the McLafferty group in 1999 (though it appears a few times in earlier publications by the McLafferty group\textsuperscript{45, 46}) and they demonstrated that 100% sequence coverage can be attained by TD characterization.\textsuperscript{47} Other milestones include development of dissociation techniques such as IRMPD,\textsuperscript{46} ECD,\textsuperscript{48} and ETD\textsuperscript{49} (discussed more under fragmentation techniques section) and these methods prove to be valuable for characterizing sequence conflicts, disulfide bonds, and PTMs at the intact protein level.\textsuperscript{7, 50-52} Other applications of TD MS include, but not limited to, studies to examine non-covalent protein complexes, protein-ligand complexes, and protein conformation.\textsuperscript{53, 54} In addition, TD characterization of membrane proteins has proved to be highly beneficial since the hydrophobic regions in the membrane proteins do not ionize efficiently by BU methods.\textsuperscript{55} Furthermore, TD
quantitation of proteins provides information about the relative amounts of different isoforms/proteoforms present and protocols are emerging in this area.\textsuperscript{56}

1.5. Mass analyzers for TD proteomics

This section will give a brief overview of the different mass analyzers that are being used in the TD field.

1.5.1. FT-ICR MS

The instruments based on FT-ICR technology are highly sought for TD work since the current capabilities include resolving power greater than $10^6$, subattomole sensitivity, broad $m/z$ range, and accurate mass measurement with less than 1 ppm deviation.\textsuperscript{57} In addition, a variety of tandem MS techniques are also available and a combination of different fragmentation methods can be applied in series.\textsuperscript{58} High MW proteins often tend to have more charges and as the spacing between the isotopes decreases with increasing charge, high resolution is required to unambiguously resolve these high charge states. This is also true for fragment ions at higher charge states and assignment of overlapping fragment ions. The combined benefit of high resolution and mass accuracy makes FT-ICR a popular choice for examining high MW proteins and TD analyses of proteins over 200 kDa using FT instruments have been reported.\textsuperscript{59} Moreover, accurate mass measurements are required to distinguish PTMs (trimethylation versus acetylation) and amino acid residues (lysine versus glutamine) that have small mass differences which becomes significant in high MW proteins.\textsuperscript{26}

Early studies for protein characterization relied heavily on custom FT instruments and even to date, the highest mass protein to be isotopically resolved is IgG1 (148 kDa) by the Marshall group using a custom-built 9.4 T FT-ICR instrument.\textsuperscript{60} With new customized ion trapping cell in a 12 T FT-ICR instrument, Smith’s group demonstrated sub-ppm mass
accuracies for intact proteins.\textsuperscript{51} Agar’s research group has demonstrated that proteins up to 669 kDa (thyroglobulin dimer) can be fragmented by in-source dissociation in a hybrid quadrupole FT-ICR instrument and identified by database searching.\textsuperscript{62} This instrument has a quadrupole mass filter which can be used to select the ions before CAD in the collision cell.

Another widely used type of FT-ICR is the linear ion trap-FT-ICR (LTQ-FT-ICR) hybrid instrument which has the LTQ mass analyzer in addition to the FT-ICR. Ions pass from the ESI source into the LTQ where they are stored, then passed down to the FT-ICR cell for analysis. With LTQ, ion accumulation, mass-selection, and dissociation can be performed within the LTQ instead of the FT-ICR cell. CAD is typically performed within the LTQ and LTQ can also be used for ion detection and used for rapid, low resolution analysis. The Kelleher group first utilized this type of instrument to perform intact protein identifications during protein elution from a liquid chromatography (LC) column and identified 22 yeast proteins ranging from 14 to 35 kDa in a single LC-MS/MS run.\textsuperscript{63}

To the biomedical research community, it is crucial to identify the isoforms and/or proteoforms in proteins related to health and disease. Ge’s group has utilized the FT-ICR instruments to characterize phosphorylations in various cardiac proteins\textsuperscript{64} and Whitelegge’s group detected various isoforms and SNPs in salivary proteins using a similar instrument.\textsuperscript{65} Furthermore, single amino acid polymorphism in the amyloidosis causing protein, transthyretin was accurately identified using the monoisotopic mass obtained from the FT-ICR instrument.\textsuperscript{66} The single amino acid changes in proteins can be identified only if the experimental and theoretical monoisotopic masses are compared, since the average mass calculated from the most abundant isotopes can vary depending on the experimental ion abundances.

In spite of these exceptional capabilities, the FT-ICR instruments are very expensive, slow, and not very user-friendly. Due to these drawbacks, instruments with other types of mass analyzers are also being actively explored for TD studies.
1.5.2. Orbitrap

The orbitrap works under the same principle as FT-ICR, but the ions are oscillating back and forth within an electrostatic field which eliminates the need for superconducting magnets.\textsuperscript{67} Hence, the cost and maintenance of orbitrap is less when compared FT-ICR and this makes it more affordable. The hybrid version of the orbitrap has a LTQ in the frontend and it combines the resolving power and mass accuracy of orbitrap with the speed and sensitivity of LTQ which makes it more amenable to TD proteins analysis. In addition, the resolving power in orbitrap decreases as the square root of $m/z$ as opposed to FT where the resolving power is very sensitive to $m/z$ and decreases as its first power.\textsuperscript{68}

In 2006, Mann and coworkers first reported the utility of LTQ-orbitrap for TD analysis by demonstrating that proteins up to 25 kDa can be isotopically resolved and measured within 3 ppm deviation using the “lock-mass” approach.\textsuperscript{69} In their study, standard proteins were fragmented in the LTQ and mass analyzed in the orbitrap followed by database searching to identify it. These authors also showed the ability to perform MS$^3$ in the ion trap for detailed structural analysis of intact proteins. The mass limit for analyzing intact proteins in LTQ-orbitrap was greatly extended by Bondarenko \textit{et al.} when they showed that intact monoclonal antibodies (150 kDa) can be analyzed in it.\textsuperscript{70} In addition, they characterized isoforms differing in glycosylation and disulfide bonding, but the isotopic resolution could be obtained only up to 23 kDa for the light chain component after disulfide reduction of the antibody. Following this demonstration, there are few other reports that describe the feasibility of using LTQ-orbitrap for TD analysis of high MW standard proteins and antibody components.\textsuperscript{71, 72} Huber’s group investigated a real complex sample mixture from wheat extract and showed isotopic resolution for proteins up to 42 kDa but tandem MS was not performed in their study.\textsuperscript{73}

The Costello group has utilized the potential of LTQ-orbitrap for the biomedical research field by developing a clinical assay for detecting variants in transthyretin.\textsuperscript{74} In their approach,
both intact mass and the fragments from in-source dissociation were used for initial screening of
the mass shift corresponding to the variant. This was followed by selection of the mass shifted
fragment ion (from in-source dissociation) in the ion trap for further MS/MS and detection by the
orbitrap.

Currently, the two newer versions of the LTQ-orbitrap, hybrid orbitrap Velos Pro and
orbitrap Elite are beginning to be explored for TD proteomics. For the TD analysis of histone
proteins which are heavily modified low MW proteins, the fast scan speed and efficient
fragmentation available in the Velos LTQ-orbitrap shows promising results with characterizing
isoforms and proteoforms. Orbitrap Elite has the combined features of Velos and high-field
orbitrap technology, and offers a resolution of 240,000 at an \( m/z \) of 400, which has quadrupled
since its earlier version. Makarov’s group has demonstrated that proteins up to 47 kDa can be
isotopically resolved and identified by combining high resolution MS scan of a selected charge
state and 5-6 high resolution MS/MS scans. Furthermore, Kelleher’s group performed a pilot
study with HeLa cell lysates and showed that high throughput protein identifications can be
effectively carried out using the various fragmentation techniques available in the
chromatographic timescale in the orbitrap Elite.

1.5.3. Quadrupole time-of-flight

The time-of-flight (TOF) mass analyzer measures the masses simultaneously as
opposed to scanning instruments which sequentially focus only one mass at a time. This
makes the TOF inherently fast and sensitive, and when combined with a quadrupole mass filter,
the hybrid QTOF instrument also provides the benefits of MS/MS in space. The deviation in
mass measurement is usually 10 ppm, however, instruments with ‘on-the-fly’ calibration mode
routinely achieve within 2-5 ppm. In addition, even though the current resolution is limited to
10,000, newer generation instruments hold much promise with resolution up to 50,000 which makes it ideal for high throughput TD analysis of proteins.\textsuperscript{27,78}

Almost a decade ago, Nemeth-Cawley et al., first demonstrated the utility of QTOF instruments for TD protein analysis, by investigating 8 standard proteins of a wide mass range (5 to 66 kDa).\textsuperscript{79} They isolated individual charge states from the ESI generated multiple charge state distribution of the protein and fragmented it in the collision cell followed by mass analysis using the TOF. The tandem MS data correctly identified all the proteins during database searches and the same approach was employed by the authors again for the characterization of recombinant IgG1 fusion protein products.\textsuperscript{80} Similar TD experiments in QTOF instruments were reported in a few other studies, but the maximum resolution of 10,000 at a \textit{m/z} of 1000 limited the charge state assignment of ions above 5+.\textsuperscript{81}

Following the early reports on in-source fragmentation mediated TD analysis on a variety of proteins in quadrupole mass analyzers, Thevis et al., employed a similar approach in a hybrid QTOF instrument to characterize the high MW intact protein transferrin (78 kDa).\textsuperscript{82} These authors employed both in-source and conventional MS/MS, and obtained comparable results. In addition, they found a novel C-terminal fragment linked by disulfide bonds in mammalian transferrins using a combination of protein chemistry and MS/MS techniques.

Another strategy that can be employed for intact protein analysis in the QTOF instruments is a pseudo MS\textsuperscript{3} approach by combining in-source fragmentation with conventional dissociation in a dedicated collision cell. Ginter \textit{et al.} applied this dual fragmentation approach to characterize standard proteins up to 66 kDa in a hybrid TOF instrument.\textsuperscript{83} First they fragmented intact proteins in the source region by raising the cone voltage which was followed by isolating one of the fragments for tandem MS in the collision cell and this sequential dissociation yielded sequence tags for unambiguous protein identification.
Johnson et al. reported the feasibility of identifying proteins fragments in recombinant samples by increasing only the cone voltage in a hybrid QTOF instrument.\textsuperscript{84} However, for such an experimental strategy, a stand-alone TOF analyzer would be sufficient to measure the fragment ion masses generated by in-source dissociation. Recently, Ren et al. demonstrated this possibility by characterizing light and heavy chains in antibodies using a single-stage TOF mass spectrometer by inducing fragmentation in the source region.\textsuperscript{85}

The ion mobility separation available in QTOF instrument designs from Waters separates ions based on differences in shape and this adds another dimension to the existing \textit{m/z} based separation. This has already shown promise in the BU area and is being explored for TD research. The Clemmer lab has incorporated ion mobility based separation for intact proteins eluting from a column and the charge states of each intact protein gets separated based on the drift time.\textsuperscript{86} This helps with separating partially folded and unfolded structures of a protein which facilitates protein characterization to some extent during the chromatographic timescale. Fragmentation of each of the separated intact protein charge states yielded sequence information by measuring the \textit{m/z} ratio of the fragments. However, the dissociation data from different charge states of the same protein were not compared to find the value of separating the intact protein charge states based on drift time. Nevertheless, this technique resolves proteins that co-elute with other contaminants, and hence, could potentially resolve co-eluting proteins by generating different ion mobility profiles. Recently, Zinner et al. has further enhanced this technique for TD studies by first isolating and dissociating a charge state, and then recording both the drift time and \textit{m/z} of the fragment ions.\textsuperscript{87} By recording the drift time of the fragment ions instead of the intact protein charge states, they were able to add another layer of separation in the tandem MS spectrum. Similar strategy was used by Halgand et al. to analyze a recombinant measles virus phosphoprotein, and the ion mobility based separation followed by TD MS/MS revealed novel microheterogenities in this protein.\textsuperscript{88}
1.5.4. Ion trap

Though high scanning speed and sensitivity are the inherent advantages that make ion traps high throughput workhorses in BU proteomics, the low resolution and deviation up to 100 ppm in mass accuracy are the major factors that limit its usage in TD analysis.\textsuperscript{10} In spite of these shortcomings, the efficient ion dissociation characteristics and multistage tandem MS in time available in ion trap instruments are advantageous for intact protein characterization.\textsuperscript{89}

McLuckey’s group pioneered in adapting ion trap instruments for targeted TD protein analysis by developing methods to reduce the charge states of multiply charged fragment ions which eliminates the necessity for high resolution.\textsuperscript{90} Intact protein charge states are isolated and fragmented in the ion trap which is followed by gas-phase ion/ion proton transfer reaction (PTR) using an anionic reagent to reduce the multiply charged product ions to mostly 1+ charge state. This charge state reduction results in products with higher \textit{m/z} which necessitates mass analyzers with a broad mass range and so, it is mostly performed in custom-built ion traps with extended mass analyzing capabilities.

Quadrupole ion trap (QIT) instruments with charge state manipulation using PTR have been mostly used for targeted TD analysis of a variety of proteins.\textsuperscript{91-95} These studies provide a wealth of information to understand the fragmentation mechanisms of intact proteins and also demonstrate that different intact protein charge states yield varied amounts of sequence coverage. However, the dynamic range in QITs is limited due to space charge effects but this has been overcome by employing linear ion traps (LTQ, also known as LIT) which have greater ion storage capacity and improved dynamic range.\textsuperscript{96} The tandem MS in space capability in the LIT can be used for TD characterization and Scherperel \textit{et al.} demonstrated this by characterizing recombinant proteins with site-directed mutagenesis in this instrument type.\textsuperscript{97}
1.6. Intact protein fragmentation schemes

This section will briefly cover the fragmentation methods for TD sequence analysis with emphasis on the mechanism, sequence coverage obtained, availability in commercially made instruments, and feasibility of combining it with on-line protein separations.

1.6.1. Collisionally activated dissociation

Collisionally activated dissociation (CAD) (also known as collision induced dissociation) is one of the earliest fragmentation methods to be developed and it was primarily employed for characterizing small molecules and peptides.\textsuperscript{98-100} After the introduction of ESI to ionize proteins, CAD was extended to fragment isolated intact protein charge states\textsuperscript{32} and is still widely used for characterizing proteins. Intact protein ions are collided with inert gas molecules (helium, argon, xenon, or nitrogen) which raise the internal energy above the dissociation threshold for various bonds and leads to fragmentation of the C-N bonds in the peptide backbone. This produces b- and y- series fragment ions and the fragmentation is dependent on the amino acid sequence and location, as explained by the mobile proton model of fragmentation.\textsuperscript{101}

CAD in the quadrupole and ion trap mass spectrometers are low-energy collisions in contrast to the high-energy type that is associated with TOF/TOF analyzers. Due to the low energy associated, these collisions lead to successive fragmentations resulting in internal fragmentation, losses of neutral molecules such as water and ammonia, and do not preserve labile PTMs such as glycosylations and phosphorylations. However, for intact protein separations combined with mass spectrometry based analysis, CAD is compatible within the protein separation timescale and generally result in 7-10 long stretches of amino acids (commonly known as sequence tags) which are then used to identify the protein by database searches.
CAD can also be carried out in the source region of the mass spectrometer by varying the voltage of the sampling cone or skimmer resulting in a similar type of fragmentation known as ‘Nozzle-Skimmer Dissociation’ (NSD).\textsuperscript{102} NSD or in-source dissociation eliminates the necessity of a dedicated collision cell to perform tandem MS and hence, isolation of intact protein charge states is not a prerequisite.\textsuperscript{103} NSD is regarded as a cost-effective way to induce fragmentation and due to its data-independent nature, current on-line protein identification methods utilize it as a way to fragment multiple protein charge states within the duration of protein separation.\textsuperscript{104,105}

**1.6.2. Higher energy collision dissociation**

Higher energy collision dissociation (HCD) is similar to CAD except it is a more energetic type of fragmentation. This was introduced by Makarov’s group when they upgraded to a curved linear trap (C-trap) to eliminate the low mass cutoff loss of fragment ions that is prevalent in the ion trap.\textsuperscript{106} The product ions generated by HCD are similar to CAD in triple quadrupole instruments with prevalence of low mass ions such as immonium ions, and $y_1$ and $y_2$ ions. HCD is now available only in orbitrap and the product ions are detected by the high resolution orbitrap mass analyzer which makes it suitable for intact protein analysis.\textsuperscript{107} More recently, Kelleher’s group has applied HCD in high throughput protein identifications by dissociating intact proteins eluting from a column and their results show most number of unique identifications by HCD among all the fragmentation types compared.\textsuperscript{77}

**1.6.3. Electron capture dissociation**

Electron capture dissociation (ECD), a serendipitous achievement by McLafferty and colleagues,\textsuperscript{48} is a complementary dissociation technique to CAD. Currently, it is performed only in FT-ICR mass spectrometers and it involves capture of a thermal electron by multiply charged
protein cations which results in dissociation of covalent bonds in the polypeptide backbone predominantly. ECD is a non-ergodic process, meaning, fragmentation is independent of bond dissociation energies and thus, it is sequence independent.\textsuperscript{108} Due to this much randomized dissociation, ECD is a preferred method for \textit{de novo} protein sequencing and for structural analysis of protein-protein and protein-ligand non-covalent associations in native pH conditions.\textsuperscript{109} The products of ECD are mostly c- and z- series ions, and labile PTMs are well-preserved during this non-ergodic ECD process.\textsuperscript{110}

Though a few reports demonstrate the applicability of ECD for on-line peptide separations with MS,\textsuperscript{111} it has not been demonstrated for on-line intact protein analysis due to two reasons. Firstly, the product ions formed from ECD are very low in intensity due to the random fragmentation scheme it follows and data has to be averaged for longer durations which prevent on-line usage.\textsuperscript{112} Secondly, for proteins greater than 20 kDa, fragments dissociated by ECD can still be held together by the intramolecular non-covalent bonds existing between amino acid residues and this leads to less sequence coverage. But this has been overcome with the development of ‘activated-ion ECD’ (AI-ECD), where protein ions are preheated by collisions with inert gas to break the intramolecular interactions and then ECD is applied.\textsuperscript{58, 113} AI-ECD has been shown to separate the ECD fragments effectively for proteins up to 200 kDa.\textsuperscript{59}

\subsection*{1.6.4. Electron transfer dissociation}

Electron transfer dissociation (ETD), an off-shoot of ECD, developed by the Hunt lab made it possible to perform ECD type fragmentation in bench top mass spectrometers.\textsuperscript{49} ETD is mediated by transfer of an electron from a low electron affinity anionic reagent to multiply charged protein cations resulting in formation of a radical which induces fragmentation.\textsuperscript{114} The outcome of ETD is similar to ECD, in that, the product ions are of c- and z- series and it preserves PTMs. Due to its availability in ion trap instruments, ETD can be performed in
addition to CAD since the fragmentation data derived from these two methods are complementary. This is advantageous in BU protein identifications for peptides eluting from a column, since both CAD and ETD can be performed or a decision tree based approach where low charge states are selected for CAD and higher charge states are chosen for ETD.\textsuperscript{115,116}

ETD can be combined with other gas-phase reactions such as PTR to reduce the charge state of the product ions which enables intact protein analysis in bench top instruments.\textsuperscript{117} Coon \textit{et al.} demonstrated this applicability in an ion trap instrument by sequentially doing ETD and PTR which allowed them to analyze intact proteins up to 66 kDa (bovine serum albumin) and also characterize variations in histone proteins.\textsuperscript{118} In addition, Chi \textit{et al.} showed that ETD and PTR are fast and efficient reactions, and it can be combined with on-line chromatographic separations for intact proteins under 20 kDa in the 70S ribosomal protein complex.\textsuperscript{119}

High resolution instruments such as orbitraps also have ETD capability which makes it ideal for intact protein analysis. ETD capability in combination with higher mass accuracy and resolution is advantageous in unambiguous assignment of PTMs such as trimethylation \textit{versus} acetylation in small histone proteins. Though ETD has been mostly demonstrated on low MW proteins under 20 kDa, its availability in high resolution instruments is promising for future applications. Recently, Tsybin’s group has characterized IgG1 using ETD in orbitrap by signal averaging time-domain transients from different on-line chromatographic experiments before FT processing to increase the signal intensity and this has almost doubled the sequence coverage from their earlier analysis of the same antibody in the QTOF instrument by signal averaging different ETD experiments.\textsuperscript{120,121} Another important advantage of using ETD is that the sequence coverage it yields is greater than other fragmentation types for high throughput protein identifications performed by combining protein separation with MS analysis.\textsuperscript{77}
1.6.5. Other miscellaneous fragmentation methods

In addition to the above discussed fragmentation methods, other techniques such as IRMPD, activated ion electron detachment dissociation (AI-EDD), sustained off-resonance-CAD (SORI-CAD), negative ion IRMPD, and blackbody infrared radiative dissociation (BIRD) have been applied either solely or in addition to other widely used fragmentation methods for intact protein analysis. Most of these methods are still at initial stages of development and further explorations will reveal future applications in the TD area.

1.7. Intact protein separation techniques

This section briefly reviews the intact protein separation and fractionation techniques available for analyzing large number of proteins that constitute complex sample mixtures, with the ultimate goal of identifying the proteins by ESI-MS. Protein purification by classical biochemical methods and immunoprecipitation protocols are targeted approaches, and so these are not discussed here.

1.7.1. Electrophoresis

Electrophoretic methods separate proteins based on properties such as size, charge, or both of the molecule and the order of separation in these methods is to some extent predictable.

1.7.1.1. Gel-based techniques

One dimensional SDS-PAGE is the most convenient and widely adopted MW based sieving technique to separate proteins. It is based on negatively charged SDS molecules binding proteins to yield a uniform charge/mass ratio and a generic unfolded shape when bound by SDS. This causes the electrophoretic migration to be dependent only on the protein’s MW. Combining SDS-PAGE with isoelectric focusing (IEF) results in a two dimensional separation
(2D PAGE) technique. The separated proteins can be recovered as peptides by digestion with a proteolytic enzyme and this forms the basis of BU proteomic experiments. Though the orthogonal nature of separation of 2D PAGE results in very high resolution, it is seldom utilized in TD proteomics due to the inherent problem of proteins being fixed inside the gel matrix and elution of intact proteins from gels is cumbersome. Methods based on passive\textsuperscript{123} and electroelution are available but these are time-consuming and are ideal only for targeted proteins.

Continuous elution of intact proteins from gels was made possible by applying an electric field to elute proteins at the end of the gel column.\textsuperscript{124} A modern and much improved version of continuous elution tube gel electrophoresis is the ‘GELFrEE’ (Gel-Eluted Liquid Fraction Entrapment Electrophoresis) system by Tran and Doucette.\textsuperscript{125} The GELFrEE device consists of tube gels to separate the proteins based on MW followed by liquid phase recovery of the separated proteins. Though liquid phase recovery is highly attractive, the high concentration of SDS in the fractions has to be removed before MS analysis. Protein precipitation is usually performed to remove SDS followed by resolubilization, but it comes with additional steps during which sample loss is inevitable. Moreover, this technique also requires high amounts of samples to be loaded due to inherent losses of protein adsorption to the walls of the chamber and buffer washes has to be performed to minimize overlapping elutions.

Alternatively, SDS can be replaced with an acid-labile surfactant (ALS) which degrades at low pH and thus, the interference of the surfactant during mass spectrometric analysis can be minimized.\textsuperscript{126} Meng \textit{et al.} utilized this approach and combined continuous-elution gel electrophoresis using ALS with LC to fractionate yeast cell lysate followed by high resolution mass spectrometry to detect intact proteins.\textsuperscript{127}
1.7.1.2. Isoelectric focusing

IEF is based on migration of proteins through a continuous pH gradient until it approaches a pH at which the net charge of the protein is zero and there is no further movement. It can be performed either using carrier ampholytes in tube gel formats or immobilized pH gradient (IPG) in strips. While IPG strips are the preferred method of first dimension separation in 2D PAGE for BU experiments, it can also be directly visualized by MS based methods as reported by Loo et al.\textsuperscript{128,129}

IEF can also be performed in solution (sIEF) and it is a popular orthogonal strategy to MW based separation for TD workflows. The custom sIEF design by Tran et al. has overcome some of the problems that existed in earlier commercially available sIEF devices.\textsuperscript{125} This in conjunction with GELFrEE is a popular method to fractionate complex samples in TD applications.\textsuperscript{130}

Capillary IEF (cIEF) is based on focusing proteins at their isoelectric point by using a polyampholyte mixture in solution inside a capillary and this can be directly coupled to ESI-MS for further analysis of the separated proteins.\textsuperscript{131,132} Though the on-line coupling of cIEF and ESI-MS proved to be high throughput, it has limited capability due to the adverse effects of ampholytes and detergents used for increasing protein solubility during cIEF. Alternatively, reversed-phase liquid chromatography (RPLC) was used as an intermediate step to improve the cIEF-ESI-MS setup.\textsuperscript{133,134}

Free-flow electrophoresis (FFE) is another liquid-based prefractionation technique in a matrix-free environment and can be performed in preparative scale.\textsuperscript{135} Samples can be continuously injected into a carrier ampholyte solution followed by IEF and fraction collection for downstream ESI-MS either directly\textsuperscript{136,137} or in combination with RPLC.\textsuperscript{138}
1.7.1.3. Capillary electrophoresis

Capillary electrophoresis (CE) is based on the migration of solutes inside narrow-bore fused-silica capillaries under the influence of high electric fields. CE includes a wide variety of ancillary methods which are based on the properties of proteins (charge, size, conformation, and hydrophobicity) during electrophoretic separation.\textsuperscript{139} The most common method is capillary zone electrophoresis (CZE) which is based on the differences in the electrophoretic mobilities of the charged species. In addition, electro-osmotic flow (movement of fluid in the capillary under the applied electric field) also contributes to the direction of net flow during CZE.

Though CE was a prevalent separation technique for more than two decades\textsuperscript{140}, it was only in the late 80’s when the Smith group first demonstrated that CE effluent can be interfaced to ESI-MS by using a liquid sheath junction.\textsuperscript{141} Subsequently, Loo \textit{et al.} showed that mixtures of intact proteins can be separated and mass analyzed by CE-ESI-MS.\textsuperscript{142} Recent advancements include sheathless interfaces, where the polyimide coating at the capillary outlet is removed to create a porous tip and this is inserted into the ESI needle and filled with background electrolyte to establish electrical connection.\textsuperscript{143} Using this porous tip sprayer design and nanoESI-MS, Haselberg \textit{et al.} demonstrated low nM limit of detection for intact proteins in comparison to using sheath liquid.\textsuperscript{144} The droplets formed during the electrospray process in the sheathless design are small and this results in more efficient ionization.\textsuperscript{145}

Narrow peak widths, high resolution and sensitivity are the inherent benefits of CE-MS and it has been mostly used for analyzing glycosylations in pharmaceutical proteins.\textsuperscript{145, 146} Though high speed separations reduce the analysis time, it also necessitates the use of instruments such as ion traps and TOF mass analyzers to preserve the separation efficiency. In addition, only very low volumes of samples can be injected and preconcentration techniques have to be performed before sample is injected.\textsuperscript{147} Due to these disadvantages, CE is not yet widely used for large-scale intact protein separations.
1.7.2. Liquid chromatography

Liquid chromatography (LC) is currently the method of choice for separating complex samples before mass spectrometric analysis. The different types of phases commonly used for protein separations are discussed here. In addition, some of the methods discussed below can be directly coupled to mass spectrometric analysis (on-line) while others require fraction collection and further purification (off-line) before MS.

1.7.2.1. Reversed phase chromatography

Reversed phase (RP) is the dominant stationary phase used mainly due to the compatibility of the salt-free aqueous/organic solvents with ESI-MS. Shorter alkyl chains such as C3 and C4 attached to silica based supports and polymeric stationary phases (PLRP material) are the most common types used for intact protein separation. Although Wang et al. have demonstrated that resolution is better with longer alkyl chains, like C18, it is better to use C3/C4 chain lengths to prevent irreversible adsorption of proteins to the column. Vellaichamy et al. have shown that the PLRP columns yield better separation efficiency for high MW proteins and the number of proteins detected was more using PLRP when compared to silica based columns.

In addition to these conventional bead based columns, RP monolithic columns that contain porous polymeric stationary phase are also being explored for intact protein separations. Eeltink et al. have shown that for intact proteins, peak capacities greater than 600 can be achieved using RP monolithic capillary columns with a two hour linear gradients and formic acid as a modifier. In addition, these columns tend to be very sensitive since it lacks the stagnant zones inside the mesopores and hence, there is no intra-particular mass transfer of solutes.
1.7.2.2. Ion-exchange chromatography

In ion-exchange based chromatography, the acidic or basic functional groups in the stationary phase interacts with the corresponding oppositely charged groups in proteins. Increasing concentrations of salt is then used to elute the protein from the column which makes it incompatible for MS analysis. However, it can be combined with a RP column and used as an orthogonal setup for the identification of proteins. Sharma et al. demonstrated this strategy and detected 715 intact proteins in a microbial cell lysate by fractionating the sample using weak anion exchange (WAX) in the first dimension followed by analyzing the individual fractions using on-line RP separation and mass spectrometric analysis using FT-ICR MS.\textsuperscript{151}

1.7.2.3. Size-exclusion chromatography

Size-exclusion chromatography (SEC) has been combined with MS for the TD analysis of membrane proteins. Whitelegge's group have reported the use of SEC columns followed by ESI-MS for the analysis of membrane proteins such as bacteriorhodopsin, proteins in the photosystem 1 complex, and cytochrome b_{6f} complex.\textsuperscript{152-154} SEC was used since the elution of these integral membrane proteins from a RP column is less efficient and often RP columns have to be regenerated after the analyzing these proteins.

1.7.2.4. Multi-dimensional separation

Following the success of MudPIT technology for peptide separations, multi-dimensional separations are gaining popularity for intact proteins. However, it requires special plumbing requirements to connect the two different stationary phases in-line and the fractions separated from one column have to be stored in loops or capillaries until it can be analyzed by the second dimension. Two recent reviews have excellently outlined the various multi-dimensional separations for intact proteins followed by MS.\textsuperscript{155, 156}
1.8. TD data analyses

The data analysis tools available for TD protein analysis are only a handful in contrast to the varied options that exist for BU approaches. This is, in part, because early TD analyses were confined to targeted protein samples and did not demand any database searches. With the advent of high throughput instrumentation and protein identification type experiments in samples of higher complexity, it is necessary to build efficient data processing platforms.

TD data contain information about intact mass of the protein and the products obtained by fragmenting the protein. The N-terminus of the eukaryotic proteins can be acetylated with or without removal of the N-terminal methionine. In addition, in some proteins, signal peptide is cleaved which will make a substantial difference in mass if it is not annotated in the database. So, when complex samples with a wide variety of proteins are searched, appropriate intact mass tolerance window has to be chosen to accommodate these variations. BU analyses are restricted to peptides which fall within a narrow MW range and are not susceptible to these limitations.

TD analysis is more confounded due to the presence of isoforms and/or proteoforms for each protein. These could have a high degree of sequence similarity and unambiguous assignments can be made only if fragment ions that represent the unique region or PTMs are present in the data. This is especially true when a protein is reported to be modified by the same modifying group at different sites. Matching the modified intact mass as reported in the database to the experimental mass is sufficient to bring up the protein as a possible hit but the fragment ions mapped to the protein has to be perused to find if the experimental data can explain the modification identified.

Two different fragmentation types are generally performed in high throughput TD experiments. Either all the charge states of the intact protein are fragmented by NSD or three most abundant charge states are isolated separately and fragmented by CAD followed by
merging the data from these three fragmentations during post-processing. In CAD, when the charge states are isolated, there is a precursor mass that represents the mass of the protein. However, when the all the charge states are dissociated as in NSD, the information about the intact mass of the protein is not present and has to be manually added to the file. This can be overcome by alternating the voltage high/low during data acquisition to collect both intact and fragment masses in the same LC-MS run and determining both the masses by deconvolution during data processing. Vellaichamy et al. had followed this scheme to identify proteins in complex samples, but this is currently limited to instruments from Thermo Scientific.¹⁴⁹

Following is a brief description of TD data analysis options. Each tool was specifically developed to the instrument platform it uses for data acquisition and file format conversions are necessary for navigating cross platforms.

1.8.1. ProSight

ProSight, currently the most widely used tool for TD analysis, searches the data against shotgun annotated protein databases.¹⁵⁷ Three different modes such as absolute mass, biomarker, and sequence tag are available to perform searches. Absolute mass search matches the experimental intact mass to the theoretical intact masses in the database and creates a list of theoretically possible fragment ions and compares it to the experimental fragment mass list. The biomarker mode is designed to search endogenously formed peptides and searches every protein subsequence in the database to match the experimental intact mass given, followed by matching the fragment ions. Finally if these two modes fail to identify the protein, the sequence tag approach is taken, where sequence tags in the fragment ion data are compiled and queried against the database.

The disadvantage of ProSight is that it searches against a virtual database and considers far too many variants during a search which brings false positives and a well-defined
scoring system is necessary to address this disadvantage. In addition, it does not identify truncated proteins with unexpected PTMs that are not in the shotgun annotated databases. Also, internal fragments and neutral losses in the data are not taken into account.

1.8.2. BIG Mascot

BIG Mascot is the TD version of the well-known Mascot search engine for BU analyses. TD searches are enabled by increasing the 16 kDa limit for precursor mass to 110 kDa. Intact and fragment ion data are deconvoluted to generate monoisotopic masses and Mascot generic files are created which are then searched against the databases available. Searches can be performed in MS only, MS/MS only or using both. Internal fragments in the data can also be searched by adding user-defined instrument definitions. Its use in the TD field is currently limited due to various programming tasks that need to be performed to enable these functionalities.

1.8.3. MS-Align+

MS-Align+ is based on spectral alignment between the experimental and theoretically generated spectra for proteins in the database. It uses the deconvolution tool, MS-Deconv to group the isotopomer envelopes from the tandem MS data followed by database searching using MS-Align+. It can also search for unexpected PTMs and these authors also that showed it can outperform ProSight on complex datasets.

1.8.4. Precursor ion independent top-down algorithm

Precursor ion independent top-down algorithm (PIITA) uses only the fragment ion data to match the genomic sequences and makes initial identifications from these searches. This is followed by matching the intact mass and mapping the observed mass shifts due to additions
of PTMs or deletions of N-terminal signal sequences. The main advantage of this method is that precursor ion mass and PTMs are not searched in the beginning which shortens the search time and gives a wider search space respectively, and tries to find as many protein identifications as possible. As in BU analysis, PIITA also considers neutral losses of water and ammonia.

1.9. Supercharging

The efficiency of fragmentation during ECD and ETD depend on the charge state of the precursor selected for dissociation. Zubarev and coworkers demonstrated that the electron capture cross section is proportional to the square of the charge.\(^{163}\) In addition, several studies have shown that the sequence coverage obtained from higher charge state precursors is more when compared to the lower charge states during ETD and ECD.\(^{164}\) Recently, reagents that enhance ESI multiple charging in both aqueous/organic and native pH solvent systems were reported by different groups including ours. These ‘supercharging’ reagents are beneficial for ETD/ECD based tandem MS studies and it is reviewed here briefly.

The multiple charging phenomena observed during ESI-MS generates extensively charged gas-phase ions which yield a bell-shaped charge state envelope and this greatly reduces the mass-to-charge limits of the mass analyzer for analyzing intact proteins. The measurements of intact proteins and protein complexes that significantly exceed the upper limit of the mass analyzer (2000 \(m/z\)) are enabled by the merits of multiple charging.\(^{165}\) There are various factors that influence multiple charging such as molecular conformation, pH, and instrument parameters. In general, when proteins are sprayed in aqueous buffers (native conditions, pH 7), the charges acquired by the protein are low and are in the high \(m/z\) region of the mass spectrum. On the other hand, when proteins are in organic/aqueous solvents or in the presence of acids (denaturing conditions, pH 4 or less), the protein is more unfolded and can
acquire more charges. Consequently, the charge state envelope of the protein is in the low m/z region of the mass spectrum.

In 2001, Iavarone and Williams demonstrated that the multiple charging of proteins can further be increased by adding small molecules such as glycerol and m-nitrobenzyl alcohol (m-NBA). They showed that these reagents can increase the maximum charge and the most abundant charge gained by cytochrome c in solutions containing water/methanol/acetic acid (47/50/3, v/v/v). They proposed that in the absence of other factors that contribute to ESI charging, the increase in charge by adding m-NBA was due to the increased surface tension of the electrospray droplets. To support this theory, they suggest that when evaporation occurs during ESI droplet formation, supercharging reagents get concentrated in the droplet due to its high boiling points (the boiling point of m-NBA is 177°C at 3 Torr) relative to water/organic solvents. This causes the surface tension of the droplet to increase when the more volatile components evaporate, since m-NBA has a higher surface tension than methanol (50 mN/m versus 25 mN/m), and this leads to droplet heating and unfolding of the protein. In summary, their theory supports the idea that ‘chemical/thermal’ unfolding or simply, denaturation is the mechanism behind supercharging.

In contrast to this ‘denaturing’ mechanism, the Loo group demonstrated that proteins can be supercharged in the native state. They demonstrated that in the presence of 1% m-NBA, Zn bound carbonic anhydrase II can be charge enhanced which in turn showed that protein-ligand interactions can be maintained during supercharging. They also suggested that aqueous solvents have higher surface tension (the surface tension of water is 72 mN/m) and adding m-NBA to aqueous solvents decreased the surface tension in the droplet after the evaporation of water. In addition, they suggest the role played by ‘reagent basicity’ as a possible mechanism behind supercharging. Furthermore, the new supercharging reagent introduced by the Loo group, sulfolane (a clear colorless liquid, also known as tetra methylene sulfone, IUPAC name:...
2,3,4,5-tetrahydrothiophene-1,1-dioxide), was weakly basic than water and had a low gas-phase basicity. In addition, they also noticed the presence of ‘protein-sulfolane’ adducts at higher charge states which is contrasting to the salt adducts present at lower charge states that are formed due to incomplete desolvation. Recently, Douglass and Venter also reported the presence of reagent adducts at higher charge states and a direct interaction between the analyte and reagent is considered as one of the mechanisms behind supercharging.

On the contrary, the Williams group provided evidences from ion mobility experiments to show that protein unfolding occurs during reagent addition. An ion in a compact-form has a higher mobility, hence a shorter drift time than the same ion in a more open conformation as it travels through the ion mobility separation device. The drift times of the higher charge state ions of myoglobin in the presence of \textit{m}-NBA was more (i.e., the mobility is decreased), suggesting that it is in the unfolded state and has a greater collision cross section. However, the same charge states with and without supercharging were not compared when they reported this arrival time distribution data. For the 7+, 8+, and 9+ charges which were present with and without supercharging, the overlaid arrival time distributions does not show a significant difference in the mobility plot to conclude that a conformational change occurs. For the higher charge states such as 10+ to 19+ which are present only when \textit{m}-NBA is added, they report that the relative increase in the collision cross-sections indicate unfolding of the protein. To make the argument more compelling, the collision cross-section has to be calculated for the same charge state with and without supercharging.

Hogan \textit{et al.} also employed ion mobility MS using a differential mobility analyzer to probe the conformation of supercharged phosphorylase b ions. Immediately after solvent evaporation, the supercharged protein ions showed a small decrease (6-10%) in the mobility when compared to the one without any reagent added. But this decrease in mobility was not
high enough to conclude the transition of folded to unfolded conformation (~60% decrease). The authors attributed this to the presence of sulfolane adducts in the protein.

In addition to the above evidences, the Williams group employed circular dichroism (CD) to monitor the change in protein conformation upon adding supercharging reagents. Results from CD experiments showed that sulfolane unfolded proteins only at high concentrations and they extended these results to the situation present during ESI droplet evaporation.\textsuperscript{171} Furthermore, dimethyl sulfoxide (DMSO) was also shown to charge enhance by the same mechanism, i.e., the high boiling point of DMSO caused it to enrich in the droplet and denature the protein.\textsuperscript{173} However, DMSO is highly basic and decreases charge at low concentrations (1%) and induces denaturation at high conformation (> 5%). Marshall’s group also noticed charge enhancements with 5% DMSO\textsuperscript{174} but these results cannot be categorized as supercharging since DMSO acts as a denaturant and the charge enhancing effects are not independent from the conformational changes it can cause.

Finally, the recent paper from the Williams group discusses the effects of conformation in supercharging in disulfide bonded proteins.\textsuperscript{175} First, they show that chemical crosslinking in proteins decreases the ability to get supercharged with \textit{m}-NBA in aqueous solution conditions. These results have to be explored with other supercharging reagents using proteins that have naturally occurring disulfide bonds in aqueous conditions. In denaturing conditions, disulfide bonded proteins can be charge enhanced greatly and the only limitation to this is the number of basic groups in the protein that can be protonated.\textsuperscript{176} Chemical crosslinking might create an artificial constraint in the protein that causes an increase in the charge density during supercharging and so it could lead to a decrease in the number of charges observed. Next, supercharging the reduced/alkylated RNase A did not show any further supercharging in aqueous solutions. This effect is also questionable since reducing the disulfide bonds in the protein creates a more unfolded conformation and increases the ESI multiple charging
regardless of supercharging. Loo et al., have reported that the number of charges for disulfide-reduced proteins increases above the number of basic sites\textsuperscript{177} and Smith et al., have tabulated the same observation in different disulfide-reduced proteins.\textsuperscript{31} Supercharging the reduced form might not be expected to generate a substantial gain since the number of sites where a charge can potentially reside is already occupied.

In light of the above discussion about the mechanisms behind supercharging, the following observations should be taken into consideration:

1. Recent evidence by different groups have illustrated that in the negative mode, the increase in charge is not equivalent to the one observed in positive mode.\textsuperscript{170, 176}

2. Peptides which lack secondary structures can be supercharged\textsuperscript{178, 179} and the number of charges acquired exceeds the basic sites.\textsuperscript{176}

3. Non-ionic saccharide detergents known for their ability to solubilize proteins by reducing the surface tension can increase charge for peptides in both aqueous and denaturing conditions.\textsuperscript{180}

4. Not all proteins can be equally charge enhanced. Size and pi of the protein play an important role in the ability to be supercharged.\textsuperscript{176}

Any one mechanism cannot support the supercharging effect observed during the above-mentioned situations. Though it is currently not evident whether unfolding occurs during supercharging, the other factors such as the basicity of the reagent, direct interaction between reagent and protein, dipole moment of the reagent, and the role of non-basic amino acids in proteins can be explored in the future to find the mechanism(s) behind supercharging.

Despite the different theories behind supercharging, the benefits of supercharging are remarkable to the field of TD protein characterization. For targeted protein characterization, selecting the higher charge states obtained during supercharging has proved to be more efficient for ETD and ECD. Sterling et al. compared the ETD of ubiquitin by selecting the most abundant charge state with and without supercharging, and found a 61% increase in the
sequence coverage with supercharging.\textsuperscript{181} Madsen and Brodbelt demonstrated that as infrared multiphoton dissociation (IRMPD) of highly charged peptides obtained by supercharging resulted in increased sequence coverage and yielded more diagnostic ions that help with characterizing the modifications present.\textsuperscript{182} For characterizing protein-ligand complexes, Yin and Loo reported sequence-specific product ions originating from the ligand-bound form of the protein by performing ECD on the higher charge states obtained only during supercharging.\textsuperscript{183}

With the targeted studies highlighting the importance of supercharging, these advantages have been extended to large-scale protein/peptide analysis by adding the supercharging reagents to the solvents used in liquid LC to facilitate increased charging during elution from the column. For peptide analysis, Jensen’s group has demonstrated that ETD fragmentation efficiency and hence, sequence coverage of phosphopeptides and tryptic peptides increased when $m$-NBA was added to the LC solvent.\textsuperscript{178} Valeja \textit{et al.} extended this approach to intact proteins eluting from a column and demonstrated charge increment for transferrin (78 kDa) with reagents such as DMSO and DMF.\textsuperscript{174} However, these reagents are highly basic and there are other reports on the ‘subcharging’ effects of DMSO. For instance, Meyer and Komives reported that addition of 5% DMSO to the LC solvent for peptide identification, narrowed the charge state distribution to a few intense charge states.\textsuperscript{179} They attributed this ‘charge state coalescence’ effect to the improved spray stability and desolvation efficiency, since more uniform droplet sizes resulting from spray stability could result in uniform charge states. In addition, Tsybin’s group also noticed a bimodal charge distribution for cytochrome $c$ when 5% DMSO was used as a supercharging reagent in a dual electrospray tip to deliver the protein solution and supercharging reagent separately into the ESI plume.\textsuperscript{184} In the same report, 5% DMSO did not efficiently supercharge substance P (a peptide, with MW 1347 Da) when compared to the other supercharging reagents used.\textsuperscript{184}
In addition to the charge enhancing effect, supercharging reagents have been used for promoting signal intensity under circumstances where signal suppression from other LC modifiers such as trifluoro acetic acid (TFA) exists.\textsuperscript{185} TFA is the most common organic modifier used for reversed phase LC separations utilizing UV based detectors. TFA works as an ion-pairing agent and it complexes with the basic groups present in the proteins/peptides, decreasing the hydrophilicity and thereby increasing the retention of analytes in a reversed phase column. Yet another mechanism that is favored by addition of TFA is that the pH of the buffer is maintained below 2 which is essential to keep the free silanol groups in the column from binding to the basic groups of the protein, thereby avoiding the second mode of retention which causes peak tailing. Though TFA has lots of properties that yields good resolution, increases peak capacity and lowers peak tailing, TFA is not used for MS because of the signal suppression that happens even at very low concentrations.\textsuperscript{186} TFA is a strong acid and the acetate anions ion-pair with the basic groups of the protein preventing the ionization of the analyte in the gas phase. Chen’s group has reported an increase in the signal intensity in the presence of sulfolane and \textit{m}-NBA when 0.1% TFA was used as a LC modifier.\textsuperscript{185} In their experimental setup, supercharging reagents were introduced by desorption electrospray ionization (DESI) and it gets mixed with the LC effluent containing TFA. They postulate that the supercharging reagents might interact with TFA anions and limit the association of TFA anions with the protein ions, which might result in enhancing the signal in the presence of TFA.

1.10. Current limitations and future directions

The field of TD mass spectrometry has grown tremendously in the last decade with major improvements in mass measurement accuracy, resolving power, and fragmentation techniques for intact proteins. However, these improvements are currently utilized mostly for targeted studies of intact proteins. Large-scale TD proteomics is still under development. Most
high throughput analysis of complex samples by TD methods are aimed at finding merely high numbers of proteins and the information about PTMs/isoforms are seldom obtained. Given the distinct biological roles of different isoforms and/or proteoforms in biomedical and clinical settings, it is important to move towards experimental strategies that are aimed at identifying these species.

Though RPLC is the most conveniently interfaced protein separation method with MS, the resolution obtained is not sufficient to differentiate the modified forms of proteins. When modified and unmodified forms of the protein co-elute, isolation/fragmentation of either form is difficult. Alternative strategies such as pI or charge based separation can be combined with MS since PTMs often change the pI of the protein. In addition, methods such as CE are promising, and with further improvements in sample loading capacity and by integrating with the high speed TOF instruments, it is possible to utilize the high resolution offered by this technique.

While improving the front-end separation technique helps with sorting the proteins for MS analysis, identification and characterization of the intact proteins are solely dependent on the developments in MS technology. Advancements in instrumentation have to be made to surpass the current 30 kDa limit for high throughput intact protein identification. High resolution FT-ICR instruments are ideal for studying high MW proteins and with improved ICR cell design to prolong the transients, the resolving power can be doubled. In addition, with the increase in the magnetic field strength, such as the 21 T instrument that is now under construction, routine characterization of intact proteins as high as 150 kDa are possible.

Though FT-ICR instruments yield unparalleled resolution, it also comes with an expensive price tag and it is not commonly used. Orbitrap MS is comparatively less expensive and recent improvements such as shorter data acquisition time and extended m/z range are attractive features to be employed for intact protein analysis. The current scan speed of the orbitrap instrument is at 1 scan/second and if this is doubled, the resolution gets decreased. In
situations where higher scanning speeds are required to analyze large number of samples, the high scan speed and resolution of TOF instruments (currently 50,000 on the new design instruments) can be utilized.\textsuperscript{190}

In addition to the complementary TD and BU methods, an intermediate area known as 'Middle-Down approach' is being slowly explored as an alternative to overcome the issues encountered with the former technologies. In middle-down, enzymatic or chemical cleavage is performed to generate peptides which are much higher in MW than the tryptic peptides.\textsuperscript{191} For example, ompT cleaves at dibasic amino acids and generates peptides that are 10 kDa. Akin to the BU technique, the sensitivity, chromatographic performance, ionization/fragmentation efficiency are better in the middle-down approach.\textsuperscript{28} More importantly, it preserves the combined modifications which are present in relatively close regions of the protein and the issue of common peptides being identified (as in the BU) is not present.\textsuperscript{191} Wu \textit{et al.} explored this approach for studying the EGFR receptor and they obtained 96\% sequence coverage along with characterizing 13 phosphorylations and 10 N-linked glycan modifications.\textsuperscript{192} Recently, Kelleher’s group also followed the same approach and demonstrated that closely related isoforms/PTMs can be differentiated in a HeLa cell lysate with the ompT generated peptides.\textsuperscript{193}

With continuing developments in separation strategies and MS instrumentation, TD will be the method of choice for protein identification and characterization in complex samples. In addition, with the high throughput characterization of isoforms/PTMs in complex samples, many breakthrough discoveries in the areas of biomarker discovery and protein regulation by PTMs will also be possible by the TD technique.
1.11. Summary

Mass spectrometry based proteomics is one of the powerful tools to identify the comprehensive set of proteins in simple to complex sample mixtures and can be performed using either top-down (TD) or bottom-up (BU) approaches. While proteins analyzed by cleaving into peptides in BU methods, TD approaches are aimed at finding intact proteins. By choosing the TD methods, intact mass of the protein, short stretches of sequence information to identify the protein and information about the presence of isoforms and/or proteoforms are obtained. Current TD methods for protein identification are performed on instruments with dual mass analyzers and employ data-dependent dissociation schemes to fragment the protein. In these methods, three most abundant charge states of the protein ions are selected and fragmented one at a time followed by post-processing to merge the data. Consequently, these methods are time-consuming and reduce the duty-cycle of the instrument. Therefore, it is crucial to develop robust methods for high throughput TD protein identification. The goal of my dissertation is to develop top-down methods for high throughput analysis of proteins in complex samples.

In Chapter 2, I describe a high throughput platform to perform intact protein identification in complex mixtures using a single-stage TOF mass spectrometer. This method obviates the need for expensive instruments with dual mass analyzers to perform TD protein identification.

In Chapter 3, I present a new data-independent fragmentation technique to identify intact proteins in complex samples. In this method, multiple charge states are dissociated all at once inside the collision cell which increases sequence coverage and the duty cycle of the instrument is also not lowered in this method.

In Chapter 4, I investigate the efficiency of various supercharging reagents in different solvent conditions to enhance the multiple charging of proteins. This is performed in order to utilize these reagents to increase the efficiency of mass measurement and fragmentation for proteins eluting from a column.
In Chapter 5, I employ the supercharging reagents to enhance the multiple charging of proteins when eluting from a column. Enhancing charge while the proteins are getting eluted in the chromatographic timescale is beneficial for high throughput analysis of complex samples.

Finally, in Chapter 6, I discuss the major findings from my dissertation and provide future directions for the field.

In the Appendix, I have examined the metabolic distress after traumatic brain injury (TBI) by studying the differential proteome in patients who were critically injured and patients who recovered after TBI. I employed top-down and bottom-up MS based methods to profile both groups of samples. Included here is a reprint of the manuscript: Lakshmanan, R., Loo, J.A., Drake, T., LeBlanc, J., Ytterberg, A.J., McArthur, D.L., Etchepare, M., Vespa, P.M. Metabolic Crisis After Traumatic Brain Injury is Associated with a Novel Microdialysis Proteome. *Neurocritical Care* **2010**, *12*(3), 324-336.
1.12. References


70. Bondarenko, P. V.; Second, T. P.; Zabrouskov, V.; Makarov, A. A.; Zhang, Z., Mass measurement and top-down HPLC/MS analysis of intact monoclonal antibodies on a hybrid


113. Horn, D. M.; Ge, Y.; McLafferty, F. W., Activated ion electron capture dissociation for mass spectral sequencing of larger (42 kDa) proteins. *Anal Chem* 2000, 72, (20), 4778-4784.


CHAPTER 2

Top-Down Protein Identification using a Single-Stage LC-ESI-TOF Mass Spectrometer

2.1. Abstract

Direct dissociation of gas phase intact proteins has the potential for identifying proteins from complex mixtures and elucidating post-translational modifications. Fragmentation of proteins in the nozzle-skimmer (NS) region of the mass spectrometer is an effective dissociation technique that can be utilized for top-down analysis. We demonstrate the capability to perform intact protein identifications in a single-stage time-of-flight (TOF) mass spectrometer by rapidly switching the voltage in the NS region during protein elution from a liquid chromatography (LC) column. The intact protein mass and fragment ion masses obtained at low and high NS voltages respectively were both measured using a stand-alone TOF mass analyzer. With the higher resolution TOF mass analyzer, sequence determination from multiply charged products is facilitated. By coupling on-line protein separations to dissociation in the atmospheric pressure/vacuum interface region of the mass spectrometer, we have identified the proteins in a standard mixture. Furthermore, this technique unambiguously identified all the proteins in the human 20S proteasome complex and characterized their N-terminal modifications. This data-independent pseudo-MS/MS top-down method is attractive for high throughput top-down protein identification since the relatively less expensive TOF instrument can provide highly accurate intact mass measurements and generate sequence tags that are typically 6-7 amino acids long.
2.2. Introduction

Mass spectrometry (MS) based proteomics is one of the powerful tools available to analyze the comprehensive set of proteins in simple to complex sample mixtures. Two major routes that are commonly used in MS based analysis are top-down and bottom-up methodologies. While the proteins are analyzed intact in the top-down methods, the bottom-up methods utilize proteases that cleave the protein at specific sites to generate smaller fragments that are then searched against the database to identify the protein.

Though the choice of the method depends on the application and both methods have its own advantages and disadvantages,\(^1\) top-down methods provide the intact mass of the molecule being studied and indicate the presence of any isoforms and/or proteoforms.\(^2\) Bottom-up methods, on the other hand, are based on analyzing the proteotypic peptides (peptides that represent the protein being identified) and generally do not yield this information.\(^3\) Given the increasing numbers of isoforms/proteoforms being identified in many biological processes and pathologies, and its utilization as biomarkers/therapy targets, it is important to use the top-down methods in order to identify them.\(^4\)

Until recently, top-down methods have mostly been used for ‘characterization’ studies of relatively protein samples\(^5\) which typically yield more data and the sequence coverage obtained is high. With the major improvements in the intact protein separation methods\(^6\) and mass spectrometry instrumentation,\(^7\) top-down technique is starting to be utilized for ‘protein identification’ which serves to find the intact proteins present in the sample, but the sequence coverage is low. The Kelleher group has pioneered intact protein high throughput identification by combining on-line protein separation with mass spectrometry and they identified 22 intact proteins ranging from 14 to 35 kDa from yeast cell lysates in a single liquid chromatography (LC)-tandem mass spectrometry run.\(^8\) In their study, collisionally activated dissociation (CAD)
was used to fragment the proteins eluting from a LC column and, the intact and fragment masses obtained were searched against the database to identify it.

Though many on-line protein identification methods use CAD due to its shorter data averaging time within the protein elution from the column, problems arise during precursor selection for CAD. During electrospray ionization (ESI), proteins are multiply charged and the signal gets split among the ESI generated multiple charge states. Often the three most abundant charge states are chosen by the automated charge state selection algorithm for CAD, since there is no rigorous rule in choosing the precursor ion. But this strategy might not always be effective, since different charge states of the same protein have shown to yield differences in the amount of sequence coverage obtained. Also, the instrument time spent in this charge state selection based CAD lowers the duty cycle of the instrument. Moreover, the collision energy applied for dissociation has to be optimized for each precursor charge state for good sequence coverage. Owing to these problems, CAD is often limited to low MW proteins during on-line protein identification and so, one solution is to choose a dissociation technique that does not require any precursor selection prior to fragmentation. One such technique that is beginning to be employed in transforming top-down methods to high throughput usage is nozzle-skimmer dissociation (NSD).

NSD was first reported by Loo et al. in 1988, when they found that a bias in the nozzle voltage (in the source region of the instrument) dramatically changed the charge distribution of multiply charged proteins. Later this effect was found to be advantageous when it turned out that primary structure of proteins could be obtained by altering this voltage. Since then many research groups have used this technique for characterizing intact proteins by direct infusion. However, it has not been used for proteins eluting from a LC column. The fragmentation obtained by NSD is comparable to CAD and the dissociation in the source region could also be used as an extra level of collision to perform MS/MS/MS in triple quadrupole instruments.
In 2009, Vellaichamy et al. employed NSD in a hybrid linear ion trap (LIT) – Fourier transform (FT) [also known as LTQ-FT] instrument for on-line top-down identification of proteins in the yeast cell lysates. In their method, when intact proteins elute from a polymeric column, the voltage in the nozzle-skimmer (NS) region was switched to aid the detection of intact mass using the ion trap at low voltage and fragment mass using the FT analyzer at high voltage. Since then instruments with dual mass analyzers such as LTQ-FTs and orbitraps have been used for on-line top-down protein identification studies, where typically two mass analyzers are utilized to collect the intact and fragment ion data, and in these experiments, CAD in the ion trap or NSD in the source region was used to fragment the protein.

Though the mass accuracy and the resolution obtained during LC-MS/MS runs in LTQ-FT and orbitrap instruments are extremely high, it also comes with an expensive price tag and the affordability of these instruments is less. Apart from these two platforms, top-down protein identification and characterization by infusing almost pure protein samples have been demonstrated in triple quadrupole, quadrupole ion trap, quadrupole time-of-flight (QTOF), and quadrupole FT-ICR instruments which have tandem mass analyzers for selecting the precursor ion during dissociation and also in single quadrupole instruments using NSD. TOF instruments have only one mass analyzer and tandem MS experiments by isolating intact protein charge states cannot be performed. But the resolution and mass accuracy of this instrument are reasonable to be investigated for top-down studies and by combining a data-independent dissociation method such as NSD, high throughput protein identifications are enabled. When combined with a front-end separation technique such as LC, it makes a cost effective platform to analyze intact protein samples of medium complexity and minimizes the need for tandem MS instruments to obtain some basic structural information for protein identification purposes. So far, the utility of ESI-TOF has been described by only one report where the authors characterized light and heavy chain subunits in antibodies using NSD in a
TOF instrument. More detailed studies on the applicability of ESI-NSD-TOF platform for protein mixtures of higher complexity remain to be explored. To transform it into a high throughput method, switching the NS voltage during the LC-MS run can be implemented.

We demonstrate that when intact proteins separated by LC elute from the column, it can be subjected to NSD in a single-stage TOF mass spectrometer by switching the NS voltage from low to high every second, and both the intact and fragment mass ions obtained at low and high voltages respectively, can be measured using the TOF analyzer. This approach is similar to the label-free quantification strategy commonly known as ‘MS\textsuperscript{E}’ in the QTOF instruments for quantifying absolute amounts of peptides in shotgun proteomics experiments. MS\textsuperscript{E} is performed without any precursor ion selection and the energy in the collision cell continuously alternates between a low-energy mode for peptide precursor identification and an elevated-energy mode for generation of peptide fragmentation data. By applying the MS\textsuperscript{E} approach in the source region during on-line protein elution, we have shown that all the intact proteins in a standard protein cocktail can be identified. In addition, we have used the human 20S (h20S) proteasome as a model to demonstrate that the proteins in a complex mixture could be identified using this top-down approach (Figure 2.1). This is a proof-of-principle study to demonstrate that mass spectrometers with single mass analyzers could be used for top-down studies of moderately complex mixtures.

2.3. Experimental section

2.3.1. Sample preparation

All protein standards were purchased from Sigma-Aldrich (St. Louis, MO) unless otherwise stated. Albumin from bovine, porcine, rabbit, and sheep were made at 1 mg/mL concentration in MS grade water and diluted to 1.5 µM with 0.1% formic acid (FA) before infusion. 10 µL of the sample was injected into the Agilent dual ESI source.
Standard proteins (bradykinin, neurotensin, ubiquitin, ribonuclease A, transferrin, BSA, and carbonic anhydrase) were made at 1 mg/mL concentration in MS grade water. An equimolar mixture of these 7 proteins (10 pmoles each) was prepared by diluting in 0.1% FA. The h20S proteasome complex was purchased from Boston Biochem (Cambridge, MA). 10 pmol of h20S sample was denatured with 0.5% FA and loaded on the column. The LC and MS parameters were the same for the analyses of standard proteins and h20S sample.

2.3.2. LC conditions

5 µL of the standard protein mix was injected into the Agilent C3 column (pore size 300 Å) with dimensions 2.1 mm x 150 mm. LC-MS was performed with an Agilent 1200 LC system connected on-line with an Agilent 6220 ESI-TOF mass spectrometer (Santa Clara, CA). A resolving power of 10,000 at an m/z of 1000 is routinely achieved in this instrument and the mass window can go as high as 3200 m/z in the extended dynamic range mode. Solvent A was 0.1% FA and solvent B was 90% acetonitrile (ACN) / 10% H₂O containing 0.1% FA. The flow rate was 0.2 mL/min and the following gradient was used: 5% B for 3 min, 5%-25% B in 1 min, 25%-55% B in 30 min, 55%-80% B in 2 min, and held at 80% B for 9 min. The column was finally equilibrated with 5% B for 10 min prior to the next run. The column temperature was maintained at 60° C throughout the run.

2.3.3. MS parameters

The source parameters in the MS were as follows: ESI capillary voltage 4300 V, drying gas temperature 350° C, drying gas flow rate 10 L/min, nebulizer pressure 25 psig and fragmentor voltage 250 V. To induce NSD during LC run, the fragmentor voltage was altered between 250 V and 350 V every second during the elution window. The data was acquired at the rate of 1 spectrum / second and the acquisition window was set from 100 to 3000 m/z.
2.3.4. Data analysis

The peaks in the total ion chromatogram (TIC) were manually integrated and the mass spectra at different fragmentor voltages were obtained using the Agilent MassHunter Qualitative Analysis Program. The multiple charge state distributions of intact proteins were deconvoluted using the maximum entropy deconvolution algorithm. The fragment ion spectra were manually sequenced to find sequence tags generated from the protein.

2.4. Results and discussion

2.4.1. Optimization of fragmentor voltage

NSD in the Agilent ESI-TOF instrument is the fragmentation that occurs when the ions are transported from the back end of the capillary into the skimmer. Generally, small molecules and peptides require low fragmentor voltages (175 V) and proteins need higher voltages (250 V) for intact mass analysis. First we optimized the fragmentor voltage for NSD by comparing the efficiency of fragmentation at different voltages ranging from 250-375 V using BSA sample (Figure 2.2). We found that the signal intensity and number of fragment ions observed increases as we increase the voltage up to 350 V and any further increase in the voltage does not generate new fragment ions. Therefore, we decided to use 250 V for intact protein analysis and 350 V for fragmentation.

As a next step, we evaluated the optimized voltage (350 V) by comparing the fragmentation efficiencies of albumin from different species. Sequence alignment of the N-terminal 40 amino acid residues of albumins from bovine, sheep, porcine, and rabbit show that 26 out of 40 residues are conserved among all four species (Figure 2.3). The mass spectra obtained at 350 V from these different albumin species shows differences in the fragment ions obtained (Figure 2.4).
Further mapping these fragments generates sequence tags which are specific for each albumin species (Figure 2.5). Loo et al. have extensively analyzed albumins from a variety of species using NSD and CAD in a triple quadrupole instrument and they reported sequence specific differences from each species.\textsuperscript{13} We did notice many of the findings they reported such as abundant b-series ions (at 2+ to 6+ charge states) originating from the N-terminal region (amino acid position 20-25) from all albumin species and occurrence of y-series ions only for porcine albumin. Moreover, Loo et al. had noticed an error in the published rabbit albumin sequence at position 20 and the fragment ions they observed suggested the presence of isoleucine instead of lysine.\textsuperscript{13} The $b_{19}^{3+}$ and $b_{20}^{3+}$ ions that we observed in our data for rabbit albumin confirm their finding and the sequence conflict is now resolved in the database.

The mass accuracy of the fragment ions is sufficient to distinguish lysine ('K', incremental mass = 128.094) from glutamine ('Q', incremental mass = 128.058). This is illustrated in Table 2.1 for sheep albumin which has a ‘Q’ at position 20 from the N-terminus. The experimental $m/z$ of fragment ions obtained is compared with calculated $m/z$ when ‘Q’ is present and if ‘Q’ is replaced by ‘K’. To distinguish K from Q at an $m/z$ of 2400 (which is the mass of the peptides that originate from residue 20), the error on mass accuracy should be less than 16 ppm. The actual errors are well within 5-6 ppm, which suggests that K and Q can be accurately identified in this instrument. We found that source-induced fragmentation is sufficiently efficient to measure sequence information from the N-terminal 40 residues of all albumins and can distinguish sequence variations among albumins from different species.

\textbf{2.4.2. LC-MS of standard protein mix}

NSD can be used as a protein identification technique to find the proteins in samples of medium complexity. The upper limit of the number of proteins that can be identified depends on the up-front separation technique and the duty cycle of the mass spectrometer. To test this
possibility, we separated a mix of standard proteins on the C3 column and when each protein elutes, the fragmentor voltage was altered from low to high every second. The intact mass spectrum obtained at 250 V and fragmentation spectrum generated at 350 V were analyzed to identify the proteins and each peak in the TIC was mapped with the corresponding protein eluting in it (Figure 2.6). For small peptides like bradykinin, 100% sequence coverage was obtained (Figure 2.7A). For high MW proteins such as BSA and transferrin, 7-8 amino acid sequence tags mostly from the ends of the protein were identified (Figure 2.7B and C). When the sequence coverage of BSA obtained during LC-MS is compared with direct infusion experiment, the efficiency of fragmentation is similar, which demonstrates that there is no loss in fragmentation efficiency at comparatively higher flow rates (100 fold higher) during the LC run.

The bovine carbonic anhydrase 2 sample contained bovine superoxide dismutase C (SOD C) as a contaminant and we found this by using both the intact mass and the mass of fragment ions obtained by NSD (Figure 2.8). This protein is interesting as it has three cysteine residues with one of them in the reduced state (Cys$_6$) and the remaining two are disulfide bonded (Cys$_{56}$ ↔ Cys$_{145}$). The experimental average mass of 15591.34 matches with a deviation of 3 ppm to the theoretical mass of 15591.39 calculated with two cysteines disulfide bonded and one free cysteine. Furthermore, the sequence tag $[^2]TKAVCVL_8$ of singly charged fragment ions from the amino-terminus of SOD C was obtained and this confirms the presence of free cysteine at this position from the amino acid incremental mass obtained. Since the sequence information within the disulfide bonded regions is seldom obtained, the dissociation of backbone carbonyl bonds at position 6 supports that it is a free cysteine. Although the information about the nature of cysteine residues has already been reported in the UniProt database, the intact and fragment ion from SOD C data shows that the state of cysteines can be obtained by this method.
2.4.3. LC-MS of h20S proteasome sample

The top-down method that we developed was applied to a physiological protein complex such as the human proteasome complex with the purpose of protein identification in the LC timescale. The goal was to measure the intact mass of the proteins followed by fragmentation to determine short stretches of amino acid sequence for protein identification. In addition, the observed and theoretical intact masses were compared to identify post-translational modifications (PTMs) if present.

The 26S proteasome complex is a supra-macromolecular assembly of different polypeptides that function to degrade a majority of intracellular proteins. The two main components of the proteasome complex are the core particle or the 20S proteasome complex where the proteolysis occurs and the 19S regulatory particle which controls protein entry and has ATPase subunits that power the protein degradation. These two components function together to perform nonspecific protein degradation by allowing the entry of the substrate protein, proteolytic cleavage of the protein at specific sites followed by exit of the short peptide chains which are then cleaved into constituent amino acids and finally salvaged into various biosynthetic pathways. This mechanism along with ubiquitin as the intra-cellular molecular tag to label proteins for degradation is known as the ‘Ubiquitin-Proteasomal System’ of protein degradation.

The proteins that constitute the 20S complex form a hollow cylinder shaped structure that can be composed of only two distinct subunits as in prokaryotes to 14 different subunits as in eukaryotes. In humans, the 20S proteasome complex is made up of four stacked rings, and each ring is a heptamer of 7 different α subunits or 7 different β subunits. The α subunits form the two outer rings and the β subunits form the two inner rings such that the whole 20S proteasome is arranged as a α7β7α7β7 barrel.
The first detailed study on proteomic characterization of the human 26S proteasome complex was done by Wang et al. using bottom-up proteomic approaches. Though they have listed the subunits in the 19S and 20S particles along with the N-terminal modification on all subunits and post-translational modifications on few of them, the information about the intact mass of these subunits could not be obtained from this experimental approach. Later in 2008, Uttenweiler-Joseph et al. used a combination of proteomic approaches, and measured the intact mass of all the subunits in h20S complex by eluting proteins intact from native two dimensional gels followed infusing the subunits and analysis by high resolution FT-ICR MS. However, this method is time consuming and requires the possibility that intact proteins have to be eluted from the gel in sufficient amounts to be analyzed by top-down methods.

We tried our top-down approach on the proteasome sample by separating the proteins based on their retention on a C3 column and dissociating it in the NS region by applying high voltage. All the proteins eluted between 20 and 32 min as shown in the TIC (Figure 2.9). Alternating the NS voltage during this elution window generated both intact and fragment mass spectra for all the proteins eluting within this timeframe. With the intact mass and the sequence tags generated from fragment mass spectra, the proteins in the h20S complex were identified and one such representative identification is shown for the β6 subunit (Figure 2.10). The list of all the proteins identified along with the sequence tag that was obtained by NSD is shown in Table 2.2.

We found that 6 proteins in the h20S complex elute singly and the rest 8 proteins elute in pairs of two (Figure 2.11). Overlapping charge state distributions in the intact mass spectra can be found for proteins that co-elute (Figure 2.12). By comparing the number of sequence tags obtained for each protein, we can see that the proteins that eluted singly fragmented well than the co-eluted proteins. Except the co-eluting proteins and the β2 subunit, all the other proteins were identified by sequence tags which have 6 to 7 amino acids (Table 2.2). The resolution
(around 10,000 for an ion at 1000 m/z) available in the TOF instrument is efficient to precisely assign the multiply charged fragment ions (up to 6+) during sequencing. The mass accuracy within a deviation of 10 ppm on average for intact and 7 ppm for fragment ions was sufficient to identify the proteins.

2.4.4. Characterization of the N-terminus and PTMs in h20S proteasome subunits

The fragment ion data acquired at high NS voltage showed the presence of fragment ions that indicate the nature of N-terminus for all the subunits in the complex. Out of the 14 proteins, 9 proteins were found to be acetylated at the N-terminus. The rest 5 subunits that were found to be unmodified belonged to the β hepatmeric ring. For the co-eluting proteins, the N-terminal modification was found by correlating the experimental mass obtained with the intact mass with the N-terminal modification suggested in the UniProt database identified from top-down and bottom-up methods of the h20S complex.

Even though many of the subunits in the 20S proteasomal core were reported to be phosphorylated, it was seldom identified since there is no enrichment for the phosphorylated subunits and the extent of ionization of phosphorylated species could be much less than their non-phosphorylated equivalents in positive ESI. Only the α3 subunit was found to be phosphorylated and this was identified by the presence of the unmodified and singly phosphorylated forms of this subunit, differing by 80 Da (which is the mass of the [PO₃]⁻ group) in the deconvoluted mass spectrum (Figure 2.13). The signal from the phosphorylated form was more intense than the non-phosphorylated counterpart as visualized by two dimensional gel electrophoresis by the Monsarrat group. But the site of modification was not identified since this protein co-elutes with β2 subunit and the fragmentation efficiency is usually low during co-elution.
Many research groups have infused targeted proteins and performed NS dissociation in single mass analyzer instruments and employed it as a pseudo tandem MS method. However, when combined with an up-front protein separation technique such as LC, it is an in-expensive solution to identify proteins in complex samples. The method was successful in that it identified all the proteins in the complex and characterized the N-terminal modifications. With the resolving power available in this instrument, fragment ions up to 6+ charges were unambiguously assigned. While this is sufficient for identifying non-homologous proteins, for proteins which have a relatively high degree of sequence similarity in one or both N/C-terminals, higher resolution instruments such as FT-ICR MS or orbitrap are necessary. With higher resolution, the higher m/z fragments originating from cleavages in the middle of the protein can be identified and this yields more sequence coverage to identify homologous proteins.

2.5. Conclusions

The top down strategy that we have described above is well-suited for analyzing samples of medium complexity like physiological protein complexes, proteins eluted off one dimensional gel bands, and protein complexes pulled down by co-immunoprecipitation methods. The factor that limits the number of proteins identified is the separation of proteins by LC, which can be improved by making use of up-front separation techniques like isoelectric focusing by free-flow electrophoresis, GELFrEE, and multidimensional separation using columns connected end-to-end. As shown by the Kelleher group, when coupled to additional orthogonal separation techniques, it is possible to characterize the proteome from cell lysates.

In the bottom-up protein identification methods, the identification is based on matching the precursor and fragment ion spectra of the peptides to the one in the database. In our approach, both the intact mass of the protein and the mass of fragment ions generated by
dissociation are used for protein identification. This narrows down the possible protein hits, since in bottom-up methods, when the number of matching peptides is low for a protein identified, it increases the number of potential proteins it could match to and distinguishing proteins that share same peptides is not possible. Another important advantage of this method is that it is rapid, dependent only on the chromatographic time and does not require any precursor ion selection by the acquisition software.

Data analysis could be improved by automating the proteins identified in each run. The rate limiting step in automation is deconvolution of fragment ion spectra. A typical fragment ion spectrum generated in this method does not have a precursor ion and cannot be deconvoluted by the available deconvolution algorithms, since they require a precursor ion. By utilizing deconvolution methods that does not necessitate specifying the precursor ion and by obtaining the monoisotopic mass of the fragments observed, it is possible to search the databases using top-down search engines such as ProSight\textsuperscript{38} and BIG Mascot,\textsuperscript{39} and thereby progress towards whole proteome applications.
2.6. References


**Figure 2.1.** Schematic representation of the top-down identification of proteasome complex proteins using LC-ESI-TOF MS. When each protein elutes from the column, the voltage in the source region is switched between low and high to obtain intact and fragment ion masses respectively, and both masses were measured using the single-stage TOF mass analyzer.
Figure 2.2. Mass spectra of BSA samples at different fragmentor voltages to compare the efficiency of fragmentation. Both the intensity and number of fragment ions increase with increasing voltage until 350 V after which there is no further increase. The b-ions pertaining to the N-terminal 30 amino acids obtained at different charge states (4+ in green and 3+ in blue) are shown for one representative spectrum.
Figure 2.3. Homology between the N-terminal 40 amino acid residues in albumins from different species. Shown in red are the conserved residues and the residues that are similar in two or three species are in blue. The residues labeled in black are not conserved.
Figure 2.4. Nozzle-skimmer dissociation of albumins from different species. Mass spectra of albumins from (A) bovine, (B) porcine, (C) sheep, and (D) rabbit acquired at high NS voltage. The amino-terminal sequence variations among these albumins were determined using NSD.
**Figure 2.5.** Fragment ion assignments in mass spectra of albumins from different species. Mass spectra of albumins from (A) bovine, (B) porcine, (C) sheep, and (D) rabbit acquired at high NS voltage. The 500-900 m/z region is zoomed-in to show the b- and y-series fragment ion assignments. The b-ions pertaining to the N-terminal 30 amino acids obtained at different charge states (5+ in red, 4+ in green, 3+ in blue, and 2+ in purple) are shown. For porcine albumin, a y-ion sequence tag (shown in black) was also observed. The peaks labeled with an * are contaminant peaks that originated from albumin samples.
Figure 2.6. TIC of a standard 7 protein mix run through C3 column. The peaks are labeled with the corresponding protein eluting in it after analyzing the intact and fragment ions generated from it. Superoxide dismutase C was a contaminant present in carbonic anhydrase 2 sample.
**Figure 2.7.** Fragmentation of proteins in the standard 7 protein mix at high NS voltage during elution from the column. Mass spectrum of (A) bradykinin \([1\text{RPPGFSPFR}]_9\) showing the y-series sequence tag, (B) BSA showing the b-series ions obtained at different charge states 5+ (orange), 4+ (blue), and 3+ (red) from the N-terminal 21-31 amino acids, and (C) human transferrin showing the b-series ions obtained at 6+ (black) and 7+ (purple) charge states from the amino-terminal 56-61 amino acids.
Figure 2.8. Nozzle-skimmer dissociation of superoxide dismutase C during its elution from the LC column. (A) Intact mass spectrum of superoxide dismutase C with the inset showing the deconvoluted intact mass of the same and (B) fragment ion spectrum of superoxide dismutase C showing the fragment ion assignments and the sequence tag obtained.
Figure 2.9. TIC of h20S sample run through C3 column. The peaks obtained are labeled from 1 through 10. The NS voltage was altered from low to high every second during the elution window from 20 to 33 minutes.
Figure 2.10. Nozzle-skimmer dissociation of β6 subunit during its elution from the LC column. (A) Intact mass spectrum of protein eluting in peak 6 in the TIC obtained at low NS voltage and (B) fragment mass spectrum obtained at high NS voltage from peak 6 in the TIC. The deconvoluted mass is shown in the inset of the intact mass spectrum and amino acid sequence tag deduced by mapping the fragment ions is given in the fragment ion spectrum. From the intact and fragment masses, this protein was identified as β6 subunit.
Figure 2.11. TIC of h20S sample with the peaks labeled with the corresponding h20S subunits eluting in it. α and ß subunits are labeled in red and green respectively.
Figure 2.12. Intact mass spectra of all the proteins in the h20S complex identified by LC separation followed by NSD. Co-eluted proteins are shown together in the same spectrum.
Figure 2.13. Deconvoluted mass spectrum of the α3 subunit of h20S proteasome. The singly phosphorylated form is more intense than the unmodified form of the protein.
<table>
<thead>
<tr>
<th>Residue position from N-terminus</th>
<th>Amino acid</th>
<th>Experimental b^{3+}</th>
<th>Calculated^{a} b^{3+}</th>
<th>Calculated^{b} b^{3+}</th>
<th>Error in ppm between Calculated^{a} and Experimental b^{3+}</th>
<th>Error in ppm between Calculated^{b} and Experimental b^{3+}</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>Q</td>
<td>790.3648</td>
<td>790.3686</td>
<td>790.3807</td>
<td>4.7488</td>
<td>20.0578</td>
</tr>
<tr>
<td>21</td>
<td>G</td>
<td>809.3720</td>
<td>809.3757</td>
<td>809.3878</td>
<td>4.5550</td>
<td>19.5456</td>
</tr>
<tr>
<td>22</td>
<td>L</td>
<td>847.0655</td>
<td>847.0704</td>
<td>847.0825</td>
<td>5.7689</td>
<td>20.0925</td>
</tr>
<tr>
<td>23</td>
<td>V</td>
<td>880.0877</td>
<td>880.0932</td>
<td>880.1053</td>
<td>6.2342</td>
<td>20.0203</td>
</tr>
<tr>
<td>24</td>
<td>L</td>
<td>917.7826</td>
<td>917.7879</td>
<td>917.8000</td>
<td>5.7602</td>
<td>18.9439</td>
</tr>
</tbody>
</table>

^{a} b^{3+} if ‘Q’ is present
^{b} b^{3+} if ‘K’ is present

**Table 2.1.** ESI b_{n} product ion m/z for sheep albumin. Error in mass accuracy between calculated and experimental m/z when ‘Q’ is present at position 20 in sheep albumin is compared to the error in accuracy when ‘K’ is supposed to replace ‘Q’. The deviation in the mass at an m/z of 800 is within 6 ppm which is sufficient to differentiate ‘K’ from ‘Q’. 
Table 2.2. Summary of h20S proteasome subunits identified by LC separation followed by NSD.

The average theoretical and experimental MW, deviation in the measured mass, sequence tags obtained, and the modifications observed for each protein are tabulated.
CHAPTER 3
Top-Down Protein Identification of Proteasome Complex Proteins with nano LC FT-ICR MS Employing Data-Independent Fragmentation Methods

3.1. Abstract

Intact protein identification by combining various up-front fractionation techniques followed by LC separation and fragmentation in the gas phase has been presented by different groups to identify proteins in simple standard mixtures to complex samples. We have developed a similar approach using a 15 T FT-ICR mass spectrometer to identify the proteasome complex proteins without any pre-fractionation. For the purpose of identification in the LC timescale, we have employed funnel-skimmer CAD. In addition to this, we have developed a new data-independent fragmentation method called ‘CASI (Continuous Accumulation of Selected Ions)-CAD’. During CASI-CAD, several precursor charge states of the multiply charged protein were selected by choosing a mass range in the external quadrupole and the preselected ions were transferred to the collision cell and fragmented at once by CAD. Both these dissociation methods identified 13 out of the 14 proteins in the 20S complex, each with several 4-5 amino acids long sequence tags. From the product ion data, the nature of the N-terminus was characterized for 9 proteins and the α3 protein was found to be phosphorylated. The deviation in the measured monoisotopic mass for these proteins was 1 ppm on average. Furthermore, in the 19S complex, which has proteins extending over a wide range in MW, the intact mass of the proteins were obtained, including the 100.2 kDa protein which was measured with a deviation of 17 ppm. Efforts are being continued for fragmenting and identifying these proteins. The CASI-CAD technique for fragmenting proteins that we have presented here is an efficient addition to the growing inventory of dissociation methods compatible with the on-line separation of proteins. With further up-front separation techniques, this method could be an efficient top-down means to address more complex samples.
3.2. Introduction

Protein identification by top-down MS can be valuable for the analysis of complex samples (e.g., cell lysates) to relatively few proteins such as endogenous protein complexes. Though bottom-up methodologies have advantages, such as well-developed protocols,\(^1\) efficient peptide chromatography,\(^2\) and well-developed bioinformatics support,\(^3\) the concept of analyzing proteins intact is still unmatched in the extent of information it yields potentially. Protein identification by top-down MS yields intact mass of the protein which readily reveals if the molecular mass is in agreement with that predicted from sequence databases. Any discrepancies in the mass hints the presence of ‘protein species’ or ‘proteoforms’ (proteins modified during co- and post-translational events) and ‘isoforms’ (variations at the gene level such as related genes encoding similar protein sequences)\(^4\) which can be further characterized by targeted studies. In-depth characterization of the protein by top-down methods using an assortment of tandem MS techniques leads to unraveling sequence conflicts,\(^5\) incorrect N-terminal modifications,\(^6\) disulfide bridges,\(^7\) combinatorial PTMs,\(^8\) and positional isomers.\(^9\)

To carry out intact protein identification on the high throughput scale along with the required sensitivity and resolution, a diverse array of separation techniques have been developed\(^10\) with the goal of keeping the proteins intact and soluble at the end of every step. Ultimately the proteins are separated by LC (mostly reversed phase) before the entrance into the mass spectrometer, where once again a choice of fragmentation technique awaits the protein for identification and characterization to some extent. The first demonstration of identifying a large sample set of proteins on the LC separation time scale was accomplished by Kelleher’s group.\(^11\) Since then the separation techniques and mass spectrometry instrumentation have been modified and improved for identifying intact proteins in dense sample populations.\(^12\),\(^13\) Furthermore, the recent addition of orbitrap instruments to the existing family of high resolution mass spectrometers channels the top-down protein identification field to much
shorter data acquisition times (1.5 seconds on the orbitrap Elite versus 4-8 seconds on the 12 T FT-ICR MS).\textsuperscript{14,15}

Another significant achievement was incorporating electron transfer dissociation (ETD) during on-line protein chromatographic separation. For example, ribosomal proteins were characterized by LC-ETD-MS/MS on a modified linear ion trap and this led to protein identification by the c- and z-series ions obtained from the ETD reaction.\textsuperscript{16} Currently, the application of ETD and higher energy collision dissociation (HCD) during on-line protein elution in the orbitrap instrument has shown that these fragmentation methods yield the best sequence coverage and the most number of unique identifications respectively amongst all methods compared to top-down identification of proteins in a cancer cell line.\textsuperscript{15}

Though efforts have been made to employ a variety of fragmentation methods for performing top-down protein identification in the LC timescale, nozzle-skimmer dissociation (NSD) and collisionally activated dissociation (CAD) are the two predominant methods due to shorter data averaging time and feasibility of fragmenting proteins, including high MW proteins by NSD during on-line elution.\textsuperscript{17} ETD requires a longer duration to average data from the far more number of product ions obtained from the sequence independent fragmentation scheme it follows\textsuperscript{18} and is currently limited to proteins under 20 kDa when eluted from a column.\textsuperscript{16} HCD and electron capture dissociation (ECD) require dedicated instrument setups and can be performed only in orbitrap and FT-ICR instruments respectively. Akin to ETD, ECD is not used for on-line protein identification experiments.\textsuperscript{19}

Most high throughput top-down protein identifying platforms use NSD and CAD to envelop high and low MW proteins respectively in the proteome being studied. NSD is performed without any precursor ion selection\textsuperscript{20} and can be programmed to switch between high/low nozzle-skimmer voltages during on-line protein elution.\textsuperscript{21} In contrast, CAD is performed in a dedicated collision cell by isolation of the precursor charge state. During on-line protein
identification, low MW proteins are mostly fragmented by CAD by isolating the top ‘n’ (n = 3 to 5, which determines the instrument duty cycle) charge states and fragmentation of each separately by CAD. The dissociation products obtained from different precursor charge states of the same protein are merged during post-processing to identify it by database searching. Most high throughput protein identifying platforms developed until now work on this scheme. Fragmenting multiple charge states of a protein is preferred compared to relying on one predominant charge state dissociation, since many studies to date have shown that the fragment ions obtained from different charge states vary and analyzing different charges states increases the sequence coverage. Therefore, there is a need to develop high throughput top-down workflows that can perform in data-independent modes to fragment multiple charge states all at once in the collision cell by CAD type fragmentation, since this will eliminate the need to spend instrument time on charge state isolation based on isotope spacing and also reduce post processing time.

We have developed a top-down technique based on nanoscale LC separation followed by on-line ultra-high resolution FT-ICR MS to examine complex samples. Funnel-skimmer dissociation (FS-CAD) which is the FT instrument equivalent of the more widely known NSD was employed to fragment proteins in the LC timescale. FS-CAD been successfully used by Agar’s group to fragment proteins up to 669 kDa by infusion experiments. In addition to using FS-CAD, we have adapted a CAD type fragmentation known as Continuous Accumulation of Selected Ions (CASI)-CAD for protein identification. CASI is predominantly known in the small molecule field for its ability to selectively enrich molecules in complex matrices and for tissue imaging. Due to the high background when imaging tissue sections to probe for the presence of drugs and its metabolites, the isolation and accumulation of the compound enriches it from the other molecules and yields an accurate mass which eliminates the need to do MS/MS for validation. Recently in 2011, Zhong et al. used this technology to determine the structure of an unknown degradation product from an antifungal agent, posaconazole. LC-MS of the sample
showed the degradant at m/z 685 with an accurate mass by CASI. This was followed by infusion experiments that allowed enrichment of the degradant by choosing a mass window and accumulation for a preset time followed by transfer to the ICR cell, where further fragmentation led to identification of the molecular composition of the degradant.

Originally, the initiative to selectively accumulate ions in an external quadrupole prior to transferring them to the ICR cell was developed by Marshall and coworkers in the late 90s.\textsuperscript{31} They used it for selectively zooming in regions across the entire m/z window to get high resolution data to identify the compounds in the crude oil samples.\textsuperscript{32} Since then, this technique has been applied in both targeted\textsuperscript{33} and large-scale top-down studies,\textsuperscript{34} to accumulate ions to increase the signal-to-noise ratio and selectively scan various regions of the mass spectrum respectively. Recently, Tipton \textit{et al.} have used this ‘zoom mapping’ approach for on-line top-down protein identification by selectively choosing ions spanning two sequential 80 to 100 m/z windows for high resolution mass measurement of precursor ions followed by a broad range m/z scan for measuring the fragment ions generated by NSD, which resulted in baseline resolution and identification of proteins up to 72 kDa.\textsuperscript{35}

Here, we present the CASI-CAD technique for top-down protein identification by simultaneously accumulating 6-7 precursor charge states and dissociating all at once during on-line elution. The multiple charge states of an intact protein to be fragmented is selected by choosing a mass window (typically 800-1200 m/z) in an external quadrupole and then the ions are transferred to a collision cell where it is accumulated for a preset time and collision energy is applied to fragment it. The fragment ions are then transferred downstream to the ICR cell to facilitate mass measurement. Human 19S and 20S proteasomes were used as model complexes to develop our top-down technique. These two complexes form an integral part of the ‘Ubiquitin-Proteasomal System’ and degrade a majority of intracellular proteins.\textsuperscript{36} The 20S complex (700 kDa), also known as core particle, consists of two alpha and two beta rings, which
are stacked to form a hollow ring shaped structure (Figure 3.1) and contains the active sites responsible for proteolysis.\textsuperscript{37} Each alpha ring is made up of 7 different alpha proteins and each beta ring is made up of 7 different beta proteins in eukaryotes. The 19S complex, also called regulatory particle, contains more than 19 different subunits ranging from 8.2 to 105.8 kDa, and controls protein import and powers the complex for protein degradation.\textsuperscript{38}

Prokaryotic and eukaryotic 20S complexes have so far been analyzed by both top-down\textsuperscript{39,40} and bottom-up approaches,\textsuperscript{41} and the N-terminus/PTMs of the subunits have already been characterized.\textsuperscript{42,43} The intact mass of the human 20S subunits have been deciphered by 2D gel electrophoretic separation of the 20S subunits followed by eluting the proteins intact from the gel for further analysis with FT-ICR MS.\textsuperscript{40} The 19S complex subunits have also been characterized to some extent by bottom-up methods,\textsuperscript{44} but the intact mass of these proteins are yet to be found. We have used FS-CAD (Figure 3.1a) and CASI-CAD (Figure 3.1b) methods for identifying proteins in the human 20S (h20S) complex, and we also measured the intact masses of proteins in the human 19S (h19S) complex, including the 100.2 kDa protein. In addition to these two dissociation methods, we have also explored the possibility of using the CASI principle to enrich for fragment ions after dissociation in the funnel-skimmer region (Figure 3.1c). Intact and product ions obtained from the tandem MS experiments were then searched against the database for protein identification.

3.3. Experimental section

3.3.1. Sample preparation

Bovine ubiquitin was purchased from Sigma-Aldrich (St. Louis, MO). LC-MS grade solvents were from the following vendors: water and isopropanol from Baker (VWR, Radnor, PA) and acetonitrile (ACN) from EMD (Billerica, MA). High purity formic acid (FA) and trifluoroacetic acid (TFA) were from Thermo Scientific Pierce (Waltham, MA). Human 20S and 19S
proteasome complexes were purchased from Boston Biochem (Cambridge, MA). Prior to loading on the column, the proteasome samples were denatured by diluting in 95/5/0.1 H$_2$O/ACN/FA. Ubiquitin was made at 1 mg/mL concentration in MS grade water and then diluted to 500 fmol/µL using 50/50/0.1 H$_2$O /ACN/FA for calibrating the instrument.

### 3.3.2. LC-MS on Dionex nano LC-Bruker FT-ICR MS

Chromatography was performed using a Dionex Ultimate 3000 rapid separation LC system (Thermo Scientific, Waltham, MA), which consists of a binary high-pressure gradient pump and a ternary low-pressure loading pump. 100 fmol of the h20S proteasome sample was injected into a Dionex Pepswift reversed phase monolithic trap column (200 µm x 5 cm) and washed for 10 min with 100% water 0.05% TFA at a flow rate of 5 µL/min using the loading pump. This was followed by gradient elution into the Dionex Pepswift reversed phase monolithic analytical column (100 µm x 25 cm). 100% water 0.1% FA (solvent A) and 100% ACN 0.1% FA (solvent B) were delivered at a flow rate of 1 µL/min and the following gradient was used: 10% B for 10 min, increased to 25% B at 15 min, 50% B at 49 min, 80% B at 54 min, held at 80% B for 5 min, and finally equilibrated at 10% B for 10 min. Both the analytical and trap columns were held at 60° C throughout the run to aid protein separations. For separating the proteins in the 19S complex, the solvent B in the gradient pump was changed to 90/10/0.1 ACN/ isopropanol/ FA and the gradient used was the same as above.

The Dionex nanoLC system was connected in-line with Bruker solariX hybrid Qq-15 T FT-ICR mass spectrometer. The CaptiveSpray source (Bruker Daltonics, Billerica, MA) was used to couple the LC flow directly into the inlet of the mass spectrometer. It consists of a non-tapered 20 µm i.d. silica tip which produces a stable spray with minimized clogging. The MS source parameters were optimized for the LC flow rate of 1 µL/min and were as follows: drying gas flow of 2.5 L/min at a temperature of 180°C and a capillary voltage of -1.5 kV was applied.
The capillary exit and deflector plate were at 250 V and 230 V respectively. For intact measurements of the proteins in the h20S complex, the funnel 1 and skimmer 1 were held at 140 V and 30 V respectively, whereas for FS-CAD, skimmer 1 potential was increased to 100 V. The potential of funnel 2, which is downstream to skimmer 1, remained at 6 V in all the experiments. For intact measurements of the h19S complex proteins, the voltage in the skimmer 1 was increased to 60 V. The data was acquired in the broadband detection mode, using 1 MHz data points, with a mass range of 150-3000 m/z. The total duration of a MS scan event was 0.63 second, which includes a transient length of 0.34 second. The number of scans averaged for one spectrum was set to one.

For the CASI-CAD, the ions between 800-1200 m/z were selected in the quadrupole by setting the Q1 mass to a 1000 m/z and using an isolation window of 400 m/z. For proteins in the h20S complex, the ions were accumulated for 0.2 second in the collision cell and optimized collision energy of 14.5 V was applied for fragmentation.

The instrument was calibrated externally by infusing 500 fmol/µL of ubiquitin at a flow rate of 1 µL/min using a syringe pump through the CaptiveSpray source. The most abundant peaks from the 13+ to 6+ charge states of ubiquitin were used for calibration and a mass accuracy of 0.1-0.3 ppm was routinely obtained for the ubiquitin peaks.

### 3.3.3. Data analysis

Total Ion Chromatogram (TIC) obtained from LC-MS and LC-MS/MS (using FS-CAD and CASI-CAD) runs were smoothed using the Gaussian smoothing algorithm (0.8 second width for 1 cycle) in the DataAnalysis 4.0 (Bruker Daltonics). Mass spectra were obtained by manually integrating at the apex of each peak in the TIC. Charge state distributions of the intact proteins and fragment ions from the tandem MS experiments were deconvoluted using Sophisticated Numerical Annotation Procedure (SNAP) II algorithm in the DataAnalysis, which determines the
monoisotopic MH⁺ for both intact and fragment ions. The mass lists obtained were exported and converted to neutral masses by subtracting the mass of the proton. Since both FS-CAD and CASI-CAD are data-independent fragmentation methods, the MS/MS data files are devoid of a particular precursor ion and so the mass of the protein being fragmented was manually added to each file as the precursor mass. ProSight 2.0 was used to search the human top-down complex database⁴⁵ which is built against UniProt release 2012_06 and can be downloaded as a ProSight warehouse format file from the following site:
 ftp://prosightpc.northwestern.edu/2012_06/Eukaryotes/Homo%20sapiens/

This database contains 1,496,086 basic sequences and supports top-down searches with known N-terminal modifications, PTMs, alternative splice forms, and endogenous cleavage events.¹⁷ The ‘Absolute Mass’ type search,⁴⁶ which matches the observed precursor mass to the one in the database within the specified mass window, followed by comparing the theoretical fragments from all the proteins within the tolerance limit to the observed fragments, was performed to search the database using the following parameters: precursor mass tolerance of ± 2 Da [to account for mass discrepancies from biological events (amidation or deamidation) and/or errors during deconvolution], fragment mass tolerance of ± 10 ppm, and N-terminal acetylation and phosphorylation as modifications.

3.4. Results and discussion

3.4.1. LC-MS of h20S proteasome using nLC-FT-ICR MS

The h20S complex was separated using a reversed phase monolithic column and all the protein subunits elute between 27 to 42 min as illustrated in the TIC (Figure 3.2 bottom panel). The intact masses of all the proteins in the complex were obtained and this data was collected by loading 100 fmol of the complex in the monolithic column. The polymeric (PS-DVB) material in this column forms a continuous rod of interconnected microglobules and contains flow-
through pores, but lacks the interstitial volume inside the microglobules and hence there is no intra-particular mass transfer.\textsuperscript{47} Consequently, we find these columns to be very sensitive than the conventional bead based columns (C\textsubscript{4}/C\textsubscript{8} group bonded silica based columns and the polymeric PLRP columns) and this has been reported by other groups as well.\textsuperscript{48, 49}

In the top panel of Figure 3.2, a simulated gel-view of the same elution window as in the TIC is shown, with the x and y axes representing the time of elution and \textit{m/z} respectively. The protein(s) eluting in each peak in the TIC has a corresponding charge state envelope as depicted in the gel-view. The intact mass spectra of four h20S subunits eluting in the TIC are shown in Figure 3.3, with the inset in each panel displaying the resolution obtained for the charge state highlighted. A resolution of 45,000-50,000 was routinely obtained for all the proteins with the data points averaged from 15-20 seconds peak widths at the apex on the LC timescale.

The peak widths at half maximum (FWHM) was found to be within 15-30 seconds for all the peaks, except peaks 1 and 9 which have peak widths of 1 min and comprise proteins approaching 30 kDa in MW. Increasing peak widths with protein MW is a general trend, with peak widths in the range of 1-2 min for proteins of 35 -50 kDa MW being obtained when PLRP columns were used by the Kelleher group.\textsuperscript{35, 50} For intact protein analysis, when compared to PLRP and silica based columns, monoliths generally give rise to smaller peak widths and shorter retention times,\textsuperscript{51} and so rigorous comparisons can be done to evaluate the efficiency of these columns in top-down LC-MS protein analysis. The reproducibility of the peaks was found to be within ± 20 seconds, for both intra- and inter-day runs.
3.4.2. LC-MS/MS using data-independent fragmentation methods

To identify the proteins in the h20S complex, when each protein elutes, two different data-independent fragmentation methods such as funnel-skimmer CAD (FS-CAD) and CASI-CAD were employed.

3.4.2.1. Protein identification using FS-CAD: The FS-CAD occurs in the wide ion funnel inter-stage region between the first and second funnel and provides higher transfer efficiency for large, multiply charges ions. Fragmentation in the funnel-skimmer region was facilitated by increasing the declustering potential voltage of skimmer 1 in the source region (Figure 3.4). The voltage of skimmer 1 was optimized by setting up several LC-MS/MS runs of the h20S complex with increasing voltages and the optimum voltage was chosen by examining the intensity and number of fragment ions at every voltage (data not shown). Optimization of the voltage is also to ensure that secondary fragmentation, which tends to be uninformative, is avoided. A skimmer 1 voltage of 100 V corresponding to a ΔSF (skimmer 1 voltage – funnel 2 voltage) of 94 V yielded the best coverage and was used for fragmenting the h20S subunits during on-line elution. This is a 70 V increase in the skimmer 1 potential that was used for intact mass measurements. A fragment ion mass list generated for each subunit by averaging the data points from each peak in the TIC was used for database searching. Table 3.1 shows the experimental MW, number of fragment ions matched (both b- and y-ions), expectation value (e value) obtained from ProSight and the nature of the N-terminus characterized from the fragment ion data for each protein in the h20S complex. 13 out of 14 h20S proteins were identified with b- and y-series fragment ions with errors in the mass accuracies less than 10 ppm (Table 3.1). For the intact mass measurements (which were acquired by keeping the skimmer 1 potential at 30 V), the errors in the measured mass were much lower, and we were able to measure the intact masses of all the proteins within a deviation of 1 ppm on average, during the LC-MS runs (Table 3.2). The
observed difference in measuring the intact mass more accurately than the fragment masses could be due to the increase in the number of ions being measured after fragmentation, when compared measuring fewer ions in the intact mass spectrum.

Illustrated in Figure 3.5 is the FS-CAD of a 29 kDa protein. FS-CAD of this protein yields a fragment ion rich spectrum (Figure 3.5A) with several 6-7 amino acids length sequence tags, one of which is zoomed in Figure 3.5B. Database searching identified this protein as the $\alpha_4$ subunit of the h20S complex, with an e value of 4.94E-47. Low e values represent better matches and are less likely to be false positives. Of the total 135 fragment ions obtained, 50 (32 b-ions and 18 y-ions) were mapped (Figure 3.5C) and this was from 48 of the possible 259 distinct inter-residue amide bond cleavages in the $\alpha_4$ protein by FS-CAD. This translates to 19% sequence coverage from the 37% of the total fragments matched by the search. The sequence coverage and number of fragment ions matched could be lower owing to the presence of a large number of internal fragments that were not interpreted. Recently, Agar’s group have deciphered the fragmentation pathways observed in intact proteins during FS-CAD at a similar declustering potential voltage in an instrument like the one used in our study and they found that intermediate charge states of proteins preferentially cleave through proline and aspartic acid channels, and generate internal fragments that account for around half of the total fragments obtained. Current top-down search engines are limited in the ability to search for internal fragments, which are formed by two backbone amide bond cleavages in the protein and lack information about either termini. Further improvement in the search tools to look for internal fragment ions after generating a list of potential protein hits will be very valuable to interpret more fragment ions and thereby increase the sequence coverage.

3.4.2.2. Protein identification using CASI-CAD: CASI-CAD, which has not yet been explored for proteins, was performed by choosing a mass range such as 800-1200 m/z in the external
quadrupole and all the ions within this range were accumulated for a certain amount of time in the collision cell and fragmented by CAD (Figure 3.4). This is illustrated by infusing a sample of bovine ubiquitin (Figure 3.6). The full m/z range ESI-MS of ubiquitin is shown in Figure 3.6A and only the ions between 800-1200 m/z were selected (Figure 3.6B) in the external quadrupole, by giving a mass window of 1000±400, which predominantly isolates the 10+ to 8+ charge states from the charge state envelope of the protein. These selected ions were then transferred to the collision cell and fragmented by CAD (Figure 3.6C).

We chose to carry out the CASI-CAD type of dissociation for proteins eluting from the column, since, it can be performed rapidly during the course of the experiment. Generally, in data-dependent acquisitions, the time it takes to acquire the fragment ion data from the charge states isolated determines the duty cycle of the instrument. The duty cycle of the instrument is lowered when the program has to automatically select the precursor based on the isotopic distribution followed by acquiring the high resolution fragment ion data for that charge state. This process has to be repeated for at least three most abundant charge states within the timescale the protein elutes from the column in data-dependent top-down protein identification experiments. CASI-CAD is a data-independent method and so it eliminates the need for spending the instrument time on precursor selection. Initial survey experiments with the h20S sample were conducted to determine the mass window or the range of charge states to be selected. All the proteins in the h20S complex fall within a mass range of 20,000 – 30,000 Da and the ESI charge state distributions for these proteins lie within 700-2000 m/z range. Since CAD is much efficient when the intermediate charge states are selected, a mass window of 800-1200 m/z was selected, and this isolates 6-12 precursors varying between 38+ to 17+, depending on the ESI generated charge state distribution of the protein. The collision voltage was optimized and the ion accumulation time in the collision cell was increased from 0.2 second to 0.3 second to find if sequence coverage increases with accumulation time. The sequence
coverage of α5 subunit at 0.2 and 0.3 second ion accumulation times are shown in Figure 3.7A and 3.7B respectively, with the table in the inset showing the total number of ions exported, the number of matching fragment ions, and the e values for this ion accumulation time comparison. For approximately an equal number of ions exported, the number of fragments matched is the same and increasing accumulation time did not yield more sequence coverage. Moreover, the relative intensities of the fragment ions obtained at 0.2 and 0.3 second accumulation times were also equal (Figure 3.8). Therefore, the precursor ions were accumulated for 0.2 second in the collision cell prior to CAD. Generally, higher ion accumulation times have to be avoided since this will increase the total time it takes to acquire a scan, thereby leading to a decrease in the LC resolution.

The MS/MS data obtained from CASI-CAD experiments were also processed and searched with the same parameters as FS-CAD and the results from these identifications are summarized (Table 3.1). Akin to the FS-CAD protein identification results, the CASI-CAD technique also identified 13 out of the total 14 subunits with fragment ions being matched with errors in the mass accuracy under 10 ppm. One example, such as the CASI-CAD fragmentation of the α4 subunit is displayed in Figure 3.9A with the fragment ions obtained after dissociating the charge states that were enriched between 800-1200 m/z and a zoomed-in view is displayed in Figure 3.9B to show the b_{95}^{10+} to b_{99}^{10+} sequence tag that was obtained by mapping the fragment ions. The sequence coverage obtained for this protein is shown in Figure 3.9C.

CASI-CAD produced higher number of matching fragments for three proteins (highlighted in green in Table 3.1) when compared to the other methods used. The β7 subunit was the only protein not identified since it co-eluted with β5 subunit and did not yield any fragment ions.

The three proteins that were identified with more fragments by the CASI-CAD method were α1, α2, and β6 subunits. α1 was found to co-elute with α7 and β6 co-eluted with β4. α1
and β6 subunits were identified with 14 and 13 fragments respectively by FS-CAD. On the other hand, CASI-CAD technique picked up more fragments, and the α1 and β6 subunits were identified with 17 and 23 fragment ions respectively by CASI-CAD. Similarly, α2 subunit co-eluted with α6 and was identified with 48 b- and y-ions by CASI-CAD, whereas FS-CAD identified only 38 ions. Moreover, the analyses of proteins that were identified equally well by both the methods indicate that the data from these two methods are complementary. CASI-CAD could be used as an additional tool to enrich for fragment ions from poorly fragmenting proteins and co-eluting proteins. Though both the methods are based on the same CAD mechanism, the internal energy of the multiply charged protein ions at different stages in the mass spectrometer are different and so further voltage comparisons cannot be made.

3.4.3. Identification of co-eluting proteins

The β7 protein was not identified due to its co-elution with β5 subunit. Apart from this pair, three other pairs of proteins such as α1-α7, α2-α6, and β4-β6 co-eluted, but both the proteins in all these pairs were identified during database searching. This was possible by including both the proteins in the co-eluting pair as precursor masses when performing the search. However, ProSight does not automatically take away the fragments that already matched to one protein. Instead, it matches the complete list of fragments again with the second precursor mass and so the e values which are calculated based on the number of fragments matched versus total number of fragments entered during the search could be artificially high due to the presence of fragment ions from the co-eluting protein.

Co-eluting proteins often present a significant problem during on-line identifications and sometimes result in only one of the two proteins in the co-eluting pair being identified as the β7-β5 pair in our study. This is due to poor fragmentation of any one protein in the pair resulting in a few fragment ion matches and a greater chance of being a random hit as depicted by higher
e values for these proteins (Table 3.1). Efforts to separate these co-eluting pairs by decreasing the rate of change of the solvent B did not result in separating these pairs. Other changes like selecting a different stationary phase and/or mobile phase, changing the conformation of the proteins by altering the salt concentration followed by on-line desalting can be tried in order to separate the co-eluting species.

3.4.4. N-terminal characterization and PTM analyses

Data from both the methods showed the N-terminus to be acetylated for 9 proteins and the rest 5 proteins that were unmodified were found to be proteins in the beta ring. Only the α3 subunit was found to be phosphorylated and this was confirmed by matching the intact molecular mass obtained to the phosphorylated form of this protein reported in the database by bottom-up methods. But due to poor fragmentation efficiency of phosphorylated proteins in positive mode ESI, this protein was not effectively identified by both the methods and the site of modification could be Tyr_{160}, Ser_{242}, or Ser_{249} as indicated in the UniProt database. The y-ions obtained for this protein from Gln_{102} to Pro_{108} were too far from these sites that are known to be phosphorylated and were unable to locate the site of modification. Targeted studies on characterization of the h20S complex using bottom-up and top-down methods have shown only Ser_{249} to be phosphorylated. But large-scale phosphopeptide profiling by ETD and immunoaffinity based profiling using tyrosine specific antibodies have shown Ser_{242} and Tyr_{160} respectively to be phosphorylated too. This variation could be due to the origin of the α3 protein being from different cell lines in these studies and different enrichment methods used could also cause variable results.

The MaxEnt deconvolution of this protein shows the presence of unmodified form at 5 times lower in intensity than the singly phosphorylated form (Figure 3.10). This correlates well with a darker 2D gel spot being obtained for the singly phosphorylated form at Ser_{249} in a trail of
three gel spots for this protein in the proteomic analyses of the h20S complex by the Monsarrat group.\textsuperscript{40, 57} The presence of the phosphorylated form of this protein at a higher intensity when compared to unmodified form of this protein and also to the other phosphorylated subunits in the proteasome complex has been observed by other groups too\textsuperscript{58} and it has been suggested to stabilize the association of 19S regulatory complex to the 20S catalytic core complex to form the 26S functional proteasome.\textsuperscript{59}

3.4.5. FS-CAD followed by ion enrichment using CASI

In addition to using the above described dissociation methods, we have also enriched for the fragment ions generated by FS-CAD by choosing a mass range in the quadrupole and accumulating the fragment ions generated from FS-CAD. Applying the CASI technique after FS-CAD selectively enriches only the fragments which are within 800-1200 \textit{m/z} mass range from the proteins in the h20S complex, as opposed to scanning the entire mass window after the FS-CAD. When the h20S subunits elute from the column, the voltage in the skimmer 1 was increased to 100 V to induce fragmentation and the fragment ions were enriched by using the mass window in the quadrupole, but the collision energy was not applied, which would otherwise initiate further fragmentation of the fragments. An ion accumulation time of 0.2 second in the quadrupole was used for this method.

The data collected with FS-CAD alone and with enriching the FS-CAD fragments by using CASI were compared, and we find that using the latter approach produces more matching fragment ions than the former technique alone. This is shown for \(\alpha_4\) protein (Figure 3.11) which was identified with 50 b- and \(y\)-type ions by FS-CAD alone (Figure 3.11A), whereas the FS-CAD with CASI identified this protein with a total of 62 b- and \(y\)-ions (Figure 3.11B). Enriching for the fragments by using the CASI technique increases the total number of ions exported for the search (Figure 3.11 inset) and the e values obtained are lower for the CASI enriched data.
indicating a much confident identification, with the higher number of fragment ions being mapped. Similar results with higher total number of matching ions and lower e values were obtained for 7 other proteins in the h20S complex (highlighted in blue in Table 4.1).

We also used longer ion accumulation time such as 0.4 and 0.6 second to find if more sequence coverage could be obtained. However, the 0.4 second accumulation time did not yield a significant increase in the number of matching fragment ions and with 0.6 second, the number of matching fragments decreased (data not shown). This result confirms our previous observation of using longer ion accumulation time during CASI-CAD (Figure 3.7). The apparent decrease in the sequence coverage at higher ion accumulation times could be due to the space-charge effect in the collision cell at longer accumulation time. A decrease in the chromatographic resolution was also noticed at longer ion accumulation time.

3.4.6. LC-MS of h19S complex

The 19S complex is composed of at least 19 different canonical subunits of theoretical MWs ranging from 8.2 kDa to 105.8 kDa (obtained from primary sequence information with N-terminal modification reported in the UniProt database) and is relatively more dynamic when compared to the 20S complex. For the efficient recovery of the high MW proteins, the LC-MS of the h19S complex was performed by changing the solvent B to 90/10/0.1 ACN/IPA/FA. Though IPA works efficiently in eluting high MW species from the column, higher amounts such as 20%, resulted in higher back pressure and unstable spray, and so 10% IPA was used for h19S separations.

With 1 pmol of the h19S complex loaded on the monolithic column, the intact mass of the proteins were measured including Rpn1 (Figure 3.12), which is one of the two largest subunits in the 19S complex. Rpn1 and Rpn2 are suggested to have a MW of 110 and 104 kDa respectively, and are explored by different research groups to understand the topology and
role played by these high MW proteins in the protein destruction machinery. The intact mass measurement by LC-MS yielded an average MW of 100.2392 kDa which is 0.0017 kDa (17 ppm at a MW of 100 kDa) less than the theoretical average MW of 100.2409 kDa calculated with a N-terminal acetylation for the Rpn1 subunit as identified by bottom-up studies. The approximately 10 kDa difference in the suggested and experimental mass for the Rpn1 subunit cannot be explained by the observed PTMs such as phosphorylations in the UniProt database, and so further targeted top-down characterization is required to find the intact mass of such high MW proteins. Recently, these proteins have been suggested to act as docking sites for proteins conjugated with ubiquitin for degradation, to coordinate deubiquitination, and movement of the unfolded protein to the catalytic core. Any novel PTMs identified could shed more light the additional roles played by these high MW proteins.

3.5. Conclusions

The top-down on-line protein identification technique that we have presented here using the proteasome complex in a hybrid Qq-FT-ICR instrument can be applied to samples of medium complexity, such as identifying proteins in endogenous protein complexes. Generally, the factors that limit the number of proteins identified by on-line protein identification methods are the peak capacity of the LC column and the duty cycle of the instrument. Using a data-independent fragmentation method reduced the duty cycle of the instrument to within 0.6 second in our method. This is beneficial since the current data-dependent fragmentation methods employed in FT-ICRs and orbitraps require 2-4 fold more time to acquire high resolution data. By increasing the duty cycle, it gives more room to accommodate high speed and sensitive separation techniques such as capillary electrophoresis with high resolution FT-ICR instruments to attain best performance. Also, with further improvement in chromatographic performance to limit the peak widths to within 10-20 seconds for proteins in 30-50 kDa MW
range, peak capacity can be increased. For samples of higher complexity such as whole cell lysates, further up-front fractionation techniques are required.

When this method is compared to our previous method using the TOF instrument, there are obviously several advantages. First, the mass accuracy of the intact proteins significantly improved and all protein subunits were measured with 1 ppm deviation using the ultra-high resolution of the FT-ICR instrument. This is a significant advantage when complex cell lysates are analyzed since unambiguous identifications during database searches are possible by improved mass accuracy. In addition, as the MW of the protein increases, there is an increased necessity to measure it more accurately. An error of 1 Da relates to a deviation of 10 ppm at a MW of 100 kDa when compared to 20 ppm at a MW of 50 kDa. It becomes difficult to differentiate the biological changes (such as an amidation or deamidation) from instrument performance at these high MWs. This is also true for measuring the high MW b- and y- ions which extend up to hundreds of amino acids in length and are formed due to cleavage of the amide bond close to the middle of the protein. Measuring these high MW fragments accurately leads to deriving more sequence tags and increases the confidence of the identification. Furthermore, in the TOF instrument only ions up to 6+ charge states were assigned and this limited the sequence coverage to the terminal regions of the protein. The resolution available in the FT-ICR instrument allowed assigning high MW fragment ions which originated from cleavages in the middle of the protein.

With new developments in on-line intact protein separation techniques, increased mass accuracy for high MW proteins, and including high efficient fragmentation methods in the LC timescale, top-down protein identification will become a popular choice to study samples of higher complexity.
3.6. References


Figure 3.1. Schematic representation of the top-down protein identification using nano LC FT-ICR MS. The human 20S proteasome complex was separated into its components using a monolithic column followed by intact mass measurement using FT-ICR MS and fragmentation using either (a) FS-CAD, (b) CASI-CAD, or (c) selectively enriching the fragments generated by FS-CAD using CASI technology. The intact and fragment masses of each protein were then searched against the database to aid protein identification.
Figure 3.2. TIC of h20S complex run through RP monolithic column. All the 14 protein subunits elute within 27 to 42 min in the peaks labeled 1 through 11 (bottom panel). A simulated gel-view of the protein(s) eluting in each peak is depicted in the top panel for the corresponding elution window shown in the TIC.
Figure 3.3. Intact mass spectra of h20S proteins averaged from peaks in the TIC in Figure 3.2: (A) peak 9, (B) peak 5, (C) peak 10, and (D) peak 6 in the TIC, with the inset in each panel showing the resolution obtained for the highlighted charge state in the corresponding mass spectrum.
Figure 3.4. Schematic representation of the Bruker 15 T FT-ICR MS. The declustering potential voltage of skimmer 1 was increased to facilitate FS-CAD. During CASI-CAD, a mass range of 800-1200 $m/z$ was chosen in the external quadrupole and all the ions within this range (shown in red) were transferred to the collision cell where it is accumulated for a certain amount of time and dissociated by CAD.
Figure 3.5. FS-CAD of α4 protein: (A) full m/z range mass spectrum obtained after FS-CAD and the inset shows the decharged isotopic distribution of α4 protein, (B) 1000-1100 m/z region zoomed in to show the sequence tag identified from mapping the fragment ions, and (C) amino acid sequence of α4 protein with the b (red) and y (green) ions mapped by database searching. The sequence tag obtained by mapping the fragments in (B) is labeled in blue in (C).
Figure 3.6. CASI-CAD of ubiquitin: (A) full \( m/z \) range mass spectrum of ubiquitin, (B) mass spectrum obtained after accumulating only the ions between 800-1200 \( m/z \), and (C) mass spectrum obtained after applying the collision energy, with the ubiquitin fragment ions mapped.
**Figure 3.7.** Fragmentation map of α5 protein identified with CASI-CAD using different ion accumulation times: (A) 0.2 second ion accumulation time and at (B) 0.3 second accumulation time. The table in the inset provides the total number of fragment ions exported for data analysis, number of matching fragments, and the corresponding e values for accumulation times compared.

<table>
<thead>
<tr>
<th>Ion accumulation time (sec)</th>
<th>Number of matching fragment ions</th>
<th>Total ions exported</th>
<th>e value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>46</td>
<td>154</td>
<td>2.19E-44</td>
</tr>
<tr>
<td>0.3</td>
<td>45</td>
<td>145</td>
<td>2.6E-44</td>
</tr>
</tbody>
</table>
Figure 3.8. CASI-CAD mass spectra of α5 subunit obtained with different ion accumulation times. The relative intensities of fragment ions obtained at 0.2 second (top spectrum) and 0.3 second (bottom spectrum) are similar.
Figure 3.9. CASI-CAD of α4 protein: (A) full m/z range mass spectrum obtained after CASI-CAD and the inset shows the decharged isotopic distribution of α4 protein, (B) 1060-1150 m/z region zoomed in to show the sequence tag identified from mapping the fragment ions, and (C) amino acid sequence of α4 protein with the b (red) and y (green) ions mapped by database searching. The sequence tag obtained by mapping the fragments in (B) is highlighted in a blue box in (C).
Figure 3.10. MaxEnt deconvoluted zero-charge mass spectrum of α3 protein. The singly phosphorylated form is present at 5 times more intensity than the non-phosphorylated form.
Figure 3.11. Comparison of sequence coverage for α4 protein by FS-CAD and FS-CAD with CASI: (A) FS-CAD alone and (B) enriching for the fragments from FS-CAD by using the CASI technique. The table in the inset provides the total number of fragment ions exported for data analysis, number of matching fragments, and the corresponding e values for both the methods.
Figure 3.12. Intact mass spectrum of Rpn1, a 100.2 kDa protein in the h19S complex. The mass spectrum was averaged from the peak obtained in the TIC which is shown in the inset.
<table>
<thead>
<tr>
<th>Uniprot ID</th>
<th>Protein subunit</th>
<th>Expt. MW (Monoisotopic)</th>
<th>Modification</th>
<th>Number of amino acids</th>
<th>FS-CAD</th>
<th>CASI-CAD</th>
<th>FS-CAD with CASI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Total b &amp; y ions</td>
<td>e value</td>
<td>Total b &amp; y ions</td>
</tr>
<tr>
<td>P25786</td>
<td>α1</td>
<td>29578.8498</td>
<td>+Nacetyl Met$_{\text{ini}}$</td>
<td>263</td>
<td>14</td>
<td>2.4E-04</td>
<td>17</td>
</tr>
<tr>
<td>P25787</td>
<td>α2</td>
<td>25793.2227</td>
<td>+Met$_{\text{ini}}$ +Nacetyl Ala</td>
<td>233</td>
<td>38</td>
<td>4.69E-28</td>
<td>48</td>
</tr>
<tr>
<td>P25788</td>
<td>α3</td>
<td>28405.9678</td>
<td>+Met$_{\text{ini}}$ +Nacetyl Ser, +1 PO$_3$</td>
<td>254</td>
<td>7</td>
<td>1.46</td>
<td>15</td>
</tr>
<tr>
<td>P25789</td>
<td>α4</td>
<td>29376.1224</td>
<td>+Met$_{\text{ini}}$ +Nacetyl Ser</td>
<td>260</td>
<td>50</td>
<td>4.94E-47</td>
<td>32</td>
</tr>
<tr>
<td>P28066</td>
<td>α5</td>
<td>26436.2068</td>
<td>+Nacetyl Met$_{\text{ini}}$</td>
<td>241</td>
<td>53</td>
<td>4.01E-48</td>
<td>42</td>
</tr>
<tr>
<td>P60900</td>
<td>α6</td>
<td>27292.7554</td>
<td>+Met$_{\text{ini}}$ +Nacetyl Ser,</td>
<td>245</td>
<td>67</td>
<td>4.18E-79</td>
<td>42</td>
</tr>
<tr>
<td>O14818</td>
<td>α7</td>
<td>27780.5607</td>
<td>+Met$_{\text{ini}}$ +Nacetyl Ser</td>
<td>247</td>
<td>39</td>
<td>9.12E-36</td>
<td>18</td>
</tr>
<tr>
<td>P20618</td>
<td>β1</td>
<td>23533.9295</td>
<td>No modification</td>
<td>213</td>
<td>39</td>
<td>9.13E-31</td>
<td>27</td>
</tr>
<tr>
<td>P49721</td>
<td>β2</td>
<td>22863.6804</td>
<td>+Nacetyl Met$_{\text{ini}}$</td>
<td>201</td>
<td>28</td>
<td>1.25E-20</td>
<td>21</td>
</tr>
<tr>
<td>P49720</td>
<td>β3</td>
<td>22845.3970</td>
<td>+Met$_{\text{ini}}$ +Nacetyl Ser</td>
<td>204</td>
<td>63</td>
<td>1.09E-70</td>
<td>56</td>
</tr>
<tr>
<td>P28070</td>
<td>β4</td>
<td>24376.1058</td>
<td>No modification</td>
<td>219</td>
<td>49</td>
<td>1.75E-57</td>
<td>48</td>
</tr>
<tr>
<td>P28074</td>
<td>β5</td>
<td>22444.0922</td>
<td>No modification</td>
<td>204</td>
<td>40</td>
<td>1.12E-38</td>
<td>25</td>
</tr>
<tr>
<td>P28072</td>
<td>β6</td>
<td>21889.7342</td>
<td>No modification</td>
<td>205</td>
<td>13</td>
<td>2.25E-10</td>
<td>27</td>
</tr>
<tr>
<td>Q99436</td>
<td>β7</td>
<td>25278.7467</td>
<td>No modification</td>
<td>234</td>
<td>-</td>
<td>Co-elutes with β5</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 3.1. Summary of h20S subunits identified by FS-CAD, CASI-CAD, and FS-CAD with CASI techniques. The total number of ions matched and the corresponding e values obtained by all three fragmentation methods are given. The highest number of ions matched by FS-CAD, CASI-CAD, and FS-CAD with CASI are in order highlighted in red, green, and blue. Except the β7 subunit, all the other proteins were identified by both the methods.
Table 3.2. Summary of mass accuracies for the proteins in the h20S complex. Monoisotopic theoretical and experimental MWs of all the proteins in h20S complex, along with the errors in the measured mass in Da and ppm are tabulated.

<table>
<thead>
<tr>
<th>Uniprot ID</th>
<th>Protein subunit</th>
<th>Thr. MW (monoisotopic)</th>
<th>Expt. MW (monoisotopic)</th>
<th>Error Da</th>
<th>Error ppm</th>
</tr>
</thead>
<tbody>
<tr>
<td>P25786</td>
<td>α1</td>
<td>29578.8710</td>
<td>29578.8498</td>
<td>0.0212</td>
<td>0.717</td>
</tr>
<tr>
<td>P25787</td>
<td>α2</td>
<td>25793.2400</td>
<td>25793.2227</td>
<td>0.0173</td>
<td>0.672</td>
</tr>
<tr>
<td>P25788</td>
<td>α3</td>
<td>28406.0180</td>
<td>28405.9878</td>
<td>0.0302</td>
<td>1.064</td>
</tr>
<tr>
<td>P25789</td>
<td>α4</td>
<td>29376.1500</td>
<td>29376.1224</td>
<td>0.0276</td>
<td>0.940</td>
</tr>
<tr>
<td>P28066</td>
<td>α5</td>
<td>26436.2160</td>
<td>26436.2068</td>
<td>0.0092</td>
<td>0.349</td>
</tr>
<tr>
<td>P60900</td>
<td>α6</td>
<td>27292.7790</td>
<td>27292.7554</td>
<td>0.0236</td>
<td>0.865</td>
</tr>
<tr>
<td>O14818</td>
<td>α7</td>
<td>27780.5847</td>
<td>27780.5607</td>
<td>0.0240</td>
<td>0.865</td>
</tr>
<tr>
<td>P20618</td>
<td>β1</td>
<td>23533.9450</td>
<td>23533.9295</td>
<td>0.0155</td>
<td>0.660</td>
</tr>
<tr>
<td>P49721</td>
<td>β2</td>
<td>22863.6890</td>
<td>22863.6804</td>
<td>0.0086</td>
<td>0.377</td>
</tr>
<tr>
<td>P49720</td>
<td>β3</td>
<td>22844.4210</td>
<td>22844.3789</td>
<td>0.0421</td>
<td>1.842</td>
</tr>
<tr>
<td>P28070</td>
<td>β4</td>
<td>24376.1190</td>
<td>24376.1058</td>
<td>0.0132</td>
<td>0.543</td>
</tr>
<tr>
<td>P28074</td>
<td>β5</td>
<td>22444.1170</td>
<td>22444.0922</td>
<td>0.0248</td>
<td>1.106</td>
</tr>
<tr>
<td>P28072</td>
<td>β6</td>
<td>21889.7860</td>
<td>21889.7342</td>
<td>0.0518</td>
<td>2.368</td>
</tr>
<tr>
<td>Q99436</td>
<td>β7</td>
<td>25278.7900</td>
<td>25278.7467</td>
<td>0.0433</td>
<td>1.713</td>
</tr>
</tbody>
</table>

Average | 0.0252 Da | 1.006 ppm |

Table 3.2. Summary of mass accuracies for the proteins in the h20S complex. Monoisotopic theoretical and experimental MWs of all the proteins in h20S complex, along with the errors in the measured mass in Da and ppm are tabulated.
CHAPTER 4
Reagents for Enhancing Protein Charging for Electrospray Ionization Mass Spectrometry from Organic/Aqueous Solvent Systems

4.1. Abstract

Reagents increasing charge from acidic organic/aqueous solvents (e.g., \textit{m}-nitrobenzyl alcohol) and neutral pH aqueous solutions (e.g., sulfolane, \textit{m}-nitrophenyl ethanol) have been reported by different groups including ours. Here, we demonstrate that reagents that enhance charge for proteins in aqueous solutions (in native pH) can also increase charge in solutions that tend to denature proteins (low pH/organic solvents). Some compounds ineffective at neutral pH (e.g., \textit{m}-nitrophenol) can increase charging from denaturing solutions. Extent of charge-shifting is protein-dependent, as some proteins do not exhibit the same proportion of increased charging as other, similarly-sized proteins. Increase in charge by adding supercharging reagents was higher for disulfide bonded proteins than for high MW proteins. Effective charging is important, since it is difficult to add more charges due to the compactly folded nature of the former and limitations in the MW for the latter. Increase in the number of charges acquired moves the charge state distribution of these proteins into the lower mass/charge region where it can be effectively analyzed. Furthermore, we have tested the supercharging potential of these reagents to increase the number of charges in peptides and we found that peptides which lack secondary structures can also be supercharged. Our analysis of these reagents with different classes of proteins and peptides under denaturing solvent conditions helps with understanding some of the mechanisms behind supercharging and showed some interesting observations which encourage further research in this area.
4.2. Introduction

The hallmark of electrospray ionization (ESI) as a technique to generate and introduce gas phase molecules into a mass spectrometer is its capability of generating multiple charging. Before John Fenn's landmark ESI-MS study of proteins, he had noted in an earlier 1985 publication that small peptides such as gramicidin S (MW 1141) and substance P (MW 1347) yielded abundant doubly charged ions, and if they were able to determine how multiple charging occurs, then it would "extend substantially the effective mass range, not only of the (ESI) source but of available mass analyzers as well."¹ Nearly at the same time, similar observations from an atmospheric pressure ionization source were noted by former Soviet Union researchers.² The efficiency of tandem mass spectrometry (MS/MS) was also predicted by Fenn to benefit from the merits of multiple charging,³ and this feature was later demonstrated to be correct, as deriving sequence-informative product ions from peptides to large proteins by collisionally activated dissociation (CAD) and other activation/dissociation methods was shown to be possible.⁴

There has been a large body of work investigating the factors that govern the extent of ESI multiple charging for a given biomolecule.⁵ The role of solvent properties, solution pH, peptide/protein sequence, and size has been explored. However, enhanced multiple charging for peptides and proteins was shown by Iavorone and Williams with compounds such as \textit{m}-nitrobenzyl alcohol (\textit{m}-NBA) in solutions containing a high proportion of organic solvents and at low pH, i.e., solvent systems that are typical for ESI-MS to promote high sensitivity analysis.⁶ Marshall's group reported a similar "supercharging" effect in low pH, organic/aqueous solutions for proteins containing small amounts of DMSO and DMF.⁷ Very recently, Douglass and Venter have reported on the supercharging ability of sulfolane in a denaturing environment using 50/50/0.1 methanol/water/formic acid.⁸
Our laboratory first reported the supercharging effect by $m$-NBA in aqueous solutions (i.e., without organic solvent) at near neutral pH for the measurement of non-covalently bound protein complexes. Native protein complexes can remain intact, albeit with increased charging, for their measurement by ESI-MS. Newly discovered reagents, such as $m$-nitrophenyl ethanol, sulfolane (tetramethylene sulfone), and $m$-(trifluoromethyl)-benzyl alcohol were reported subsequently by our laboratory to enhance protein charge for ESI-MS from aqueous neutral pH solution; these reagents promote charging to an equal or higher level than that observed for $m$-NBA. Regardless of the mechanism of how these compounds are able to enhance ESI charging (*vide infra*), we demonstrate that sulfolane and all other reagents that enhance protein charge from neutral pH solutions have similar effects in solutions containing organic solvents and/or low pH. We also report that some compounds that do not supercharge native proteins (i.e., neutral pH) can increase charge in denaturing solutions. Moreover, we have investigated different classes of proteins such as disulfide bonded proteins, high MW proteins, and membrane proteins, and found that the extent of enhanced charging is protein dependant, as some proteins do not exhibit the same proportion of increased charging as another similar sized protein. Finally, we compared the maximum charge (highest observed charge state) and average charge attained by a model protein, lysozyme by adding different supercharging reagents to find the lowest concentration that is required to obtain the supercharging effect.

4.3. Experimental section

4.3.1. Sample preparation

Bovine proteins (insulin, β-lactoglobulin A, α-lactalbumin, ribonuclease A, ribonuclease B, and albumin), chicken lysozyme, human transferrin, antibody IgG1 kappa murine myeloma (IgG1), bradykinin, and substance P were purchased from Sigma-Aldrich (St. Louis, MO), as
were $m$-NBA, $m$-nitrophenyl ethanol, nitrobenzene, $m$-nitrobenzonitrile, $m$-nitrophenol, and sulfolane.

Protein samples were diluted to a final protein concentration of approximately 5 - 10 $\mu$M for ESI-MS measurements in the solvent systems described in the text and the pH adjusted by the addition of formic acid (FA). IgG1 sample was desalted and concentrated using centrifugal filter devices and the final concentration was 6 $\mu$M in ACN/H$_2$O/FA 50/49.9/0.1. Bacteriorhodopsin from the purple membrane of *Halobacter salinarium* (from Prof. Jim Bowie's lab) was isolated and precipitated by methanol-chloroform precipitation protocol as described by Whitelegge *et al.* The precipitated protein was resuspended by adding a low volume of 100% FA followed by diluting the FA 25 fold using solvents described in the text. Increased ESI-generated charge states for peptides and proteins were enabled by adding the reagents to analyte solutions either directly (for reagents that were liquids at room temperature) or by dissolving the reagents in solvents described in the text (for reagents that were in solid state).

4.3.2. Mass spectrometry

Positive ion ESI mass spectra were acquired with three different systems: ESI-TOF mass spectrometer with a dual electrospray source (Agilent ESI-TOF MS, Santa Clara, CA), a LTQ linear ion trap system (Thermo Fischer Scientific, Waltham, MA) with a nanoESI source (New Objective, Woburn, MA), and a hybrid quadrupole time-of-flight (QTOF) mass spectrometer with a Triwave ion mobility (IM) separator (QTOF/IM; Waters Synapt HDMS, Manchester, UK). The nanoESI source using borosilicate glass capillaries with Au/Pd coatings (Proxeon Biosystems, Odense, Denmark) was operated at low analyte flow conditions (50 nL/min). For a given comparison, instrument settings are held constant as reagent concentration is varied, to ensure that changes to the stability and charging of protein complexes may properly be attributed to changes in solution conditions.
4.3.3. Calculation of average and maximum charges

Charge state distribution (CSD) refers to the observed set of charges acquired by a protein during ESI-MS. From the CSD, the average charge was calculated using the formula: 
\[
\frac{\sum q_i I_i}{\sum I_i}
\]
from i to N, where \(q_i\) is the charge of the i-th charge state, \(I_i\) is the intensity of the i-th charge state and N is the number of measured charge states in a mass spectrum. Average charge can be simply correlated to the charge state with the highest intensity in the CSD obtained for a given protein. When comparing the supercharging effect, the maximum charge state or the highest observed charge state was also taken into consideration. While average state gives an estimate for the overall charge distribution upon supercharging, the maximum charge state is an indication of the highest number of charges that can be attained without limitations due to electrostatic repulsion from the existing charges.

4.4. Results and discussion

4.4.1. Reagents for supercharging under denaturing conditions

In our previous work, we disclosed the following 9 reagents that increased ESI positive charging for our test protein (myoglobin) when added to the analyte aqueous solution buffered to pH 6.8 with ammonium acetate: benzyl alcohol, \(m\)-nitroacetophenone, \(m\)-nitrobenzonitrile, \(o\)-NBA, \(m\)-NBA, \(p\)-NBA, \(m\)-nitrophenyl ethanol, sulfolane, and \(m\)-(trifluoromethyl)-benzyl alcohol. Based on the average charge state, sulfolane displayed a greater charge increase (61%) than \(m\)-NBA (21%) for myoglobin in aqueous solutions.

These new reagents also increased protein charge from denaturing solutions, i.e., solutions containing a high proportion of organic solvent and/or low pH that tend to induce protein unfolding. Chicken lysozyme (14.3 kDa) and insulin (5.8 kDa) were used as test proteins to measure the efficiency of supercharging reagents under denaturing conditions. \(m\)-Nitrobenzonitrile (22 mM) and sulfolane (300 mM) potently supercharged insulin when sprayed
from the denaturing solvent system, ACN/H₂O/FA 50/49.9/0.1 (v/v/v), shifting insulin’s average charge from 4.4 to 5.4 and 5.7, respectively (Figure 4.1). 1% m-NBA also has a similar effect as the other two reagents, shifting insulin’s average charge to 5.8 and the maximum charge observed is 7+ in the presence of the either of these supercharging reagents (Table 4.1). This is a substantial increase in charging for such a small protein like insulin with three disulfide bonds.

In the presence of the above discussed denaturing solvent system (ACN/H₂O/FA 50/49.9/0.1), 50 mM sulfolane potently supercharges another disulfide-intact protein, lysozyme by shifting the average charge from 11.1 to 14.4 (Figure 4.2A and B). Equivalent supercharging effect was observed when 50 mM m-nitrophenyl ethanol was used as the supercharging reagent (average charge 14.1) (Figure 4.2C). Also, the maximum number of charges obtained by lysozyme increases from 15+ to 19+ in the presence 100 mM sulfolane (Table 4.1), which happens to correlate to the number of basic groups in lysozyme (11 Arg, 6 Lys, 1 His, and N-terminus). Earlier studies have shown that in small proteins the number of basic residues can be directly correlated to the maximum charge state observed whereas in disulfide bonded proteins, the maximum charge state observed is less than the number of basic groups due to the higher order structure that exists from disulfide bridges.¹² Loo et al. have shown that under denaturing conditions, 14+ is the most highly charged species observed for disulfide-intact lysozyme and it increases dramatically to 20+ by cleaving the cysteine-cysteine bond using DTT.¹³ Supercharging lysozyme increases the maximum charge to an extent formerly attainable by reducing the disulfides. This is advantageous since disulfide containing proteins can be characterized without the necessity to reduce with DTT and it also removes the additional steps required to alkylate the protein to prevent disulfide bond reformation.

Reducing the disulfide linkages generally allows more multiple charging to be observed by ESI-MS. In order to see if more charges can be added, the four disulfide bonds in the lysozyme were cleaved by adding 10 mM DTT. The increase in charge is less dramatic for the
disulfide-reduced form of lysozyme, as the average charging without and with sulfolane is fairly similar (16.5 and 18.0, respectively) (Figure 4.3A and B) and a maximum charge of 22+ is observed for disulfide-reduced lysozyme regardless of the addition of sulfolane. Supercharging the reduced form of lysozyme would not be expected to generate a substantial gain in ESI charging, as the maximum charge observed for reduced-lysozyme is beyond the number of basic sites.

Smith et al. have tabulated the increase in ESI charging beyond the number of basic sites in disulfide-reduced proteins such as proinsulin, ribonuclease A, interferon-1α, and lysozyme from studies by different groups. In addition, Loo et al. have also reported that the maximum ESI charge of 20+ for reduced lysozyme was more than the number of basic sites. They attributed this to the presence of glutamine residues which have similar gas-phase proton affinity as lysine as suggested by Bojesen. Lysozyme has three glutamine residues and therefore has a total of 22 sites (19 basic + 3 glutamine residues) where the charge can potentially reside upon extending the conformation by disulfide reduction.

4.4.2. Compounds effective only in denaturing conditions

Some compounds that do not increase ESI charging from aqueous neutral pH solutions increase charge from denaturing solutions. Previously, we discussed that 20 mM m-nitrophenol appeared to denature holo-myoglobin in aqueous solutions to generate an ion series for the apo-myoglobin form. However, m-nitrophenol increased charging for lysozyme under denaturing conditions. The average charge state increases to 14.1 upon adding 50 mM m-nitrophenol (Figure 4.4B) in comparison to lysozyme sprayed by itself (Figure 4.4A) with an average charge of 11.1. To make an effective comparison under the same instrumental parameters, we sprayed lysozyme in a native solvent such as 10 mM ammonium acetate. Figure 4.4C and 4.4D display the CSDs of lysozyme in native conditions without and with 10
mM m-nitrophenol respectively. 10 mM concentration was chosen because of the solubility limit of this reagent in aqueous buffers. The apparent no change in the CSD acquired by lysozyme, when m-nitrophenol is added to the native solution conditions can be attributed to the low solubility of the reagent in aqueous buffers and it confirms our previous results obtained in native conditions. 10

We also tested the supercharging ability of m-nitrophenol under denaturing conditions with a smaller size protein like insulin and found that insulin showed increased charging with 20 mM m-nitrophenol (5.3 average charge) (Figure 4.1E). Another reagent that is effective only under denaturing conditions is nitrobenzene. Nitrobenzene additions to 0.5% v/v did not alter protein charge state distributions from aqueous solutions. 10 At 0.9% v/v (88 mM), close to the miscibility limit in water, unresolved peaks tailed to higher m/z, suggesting adduct binding and/or incomplete desolvation. 10 In ACN/H2O/FA 50/49.9/0.1, however, increased charging with 1% nitrobenzene for insulin is evident as it increases the average charge to 5.1 (Figure 4.1F).

4.4.3. Effect of different solvents in the supercharging efficiency of m-nitrophenol

m-Nitrophenol was found to supercharge in denaturing conditions, but not in native solvents. To further investigate the effects of solvent conditions in supercharging, lysozyme was supercharged with m-nitrophenol in each of following solvent concoctions: ACN/H2O/FA 50/49.9/0.1, ACN/water 50/50, methanol/H2O/FA 50/49.9/0.1, methanol/H2O 50/50, H2O/FA 99.9/0.1, and aqueous 10 mM ammonium acetate buffer. 5 µM lysozyme was diluted in each of the solvents separately and tested with/without 10 mM m-nitrophenol, which was dissolved in the same diluent as the protein. 10 mM concentration of m-nitrophenol was chosen because of the limited solubility of the reagent in aqueous buffers above this concentration and so, we decided to test this concentration across all solvent systems studied.
The panel of Figures in 4.5A-L shows the mass spectra acquired from testing lysozyme with \textit{m}-nitrophenol in different solvent conditions. Higher charge state ions were observed for \textit{m}-nitrophenol addition to disulfide-intact lysozyme sprayed from ACN/H\textsubscript{2}O/FA 50/49.9/0.1 and ACN/water 50/50, but not from methanol/H\textsubscript{2}O/FA 50/49.4/0.1, methanol/H\textsubscript{2}O 50/50, H\textsubscript{2}O/FA 99.9/0.1, and aqueous 10 mM ammonium acetate buffer. With ACN/H\textsubscript{2}O/FA 50/49.9/0.1, the maximum degree of charge shift (the average charge shifted from 11.3 to 13.6) among all the solvent systems tested is achieved (Figure 4.5A and B). When the FA is dropped from the above solvent mixture, there is still an appreciable amount of shift in average charge (from 10.6 to 11.9) (Figure 4.5C and D). With 50/49.9/0.1 methanol/water/FA, the average charge increases from 9.6 to 10.4 (Figure 4.5E and F) and in 50/50 methanol/water the average charge remains the same with addition of \textit{m}-nitrophenol (Figure 4.5G and H). In the presence of 99.9/0.1 water/FA, \textit{m}-nitrophenol did not show any supercharging effect, as noticed by the no change in the number of charges (Figure 4.5I and J). Finally, with 10 mM ammonium acetate buffer, there was no supercharging, confirming our previous results that this reagent does not supercharge under aqueous conditions (Figure 4.5K and L).

This behavior contrasts with that of sulfolane and \textit{m}-NBA, i.e., reagents which increased lysozyme charging not only from ACN-containing solvents, but also from 99.9/0.1 water/FA. A key difference between \textit{m}-NBA and \textit{m}-nitrophenol is that the former reagent is both a very weak base and very weak acid (\textit{pK\textsubscript{a}}=14.9),\textsuperscript{15} whereas the latter reagent is acidic (\textit{pK\textsubscript{a}}=8.4).\textsuperscript{16} We predict that the supercharging of \textit{m}-nitrophenol could be attenuated in aqueous solutions by ion-pairing; \textit{i.e.}, nitrophenolate anion-binding to protonated sites on the protein. Such complexes decrease overall charge when present and when collisionally dislodged as a neutral species. This is explained in detail below with the observations from the multiple charging in different solvent conditions.
In aqueous solvents, \( m \)-nitrophenol dissociates into \( m \)-nitrophenyl anion, which shares a proton with the acidic amino acids in the protein and when desolvation occurs during ESI, the \( m \)-nitrophenyl anion may take a proton with it and leave as a neutral, and hence there is no supercharging. In solvents containing a weak base like acetonitrile (50/49.9/0.1 ACN/water/FA), \( m \)-nitrophenol dissociates less and donates the proton to the protein and which results in supercharging. To confirm this theory, the supercharging efficiency was tested in methanol containing solvents (50/49.4/0.1 methanol/water/FA and 50/50 methanol/water) and the absence/negligible amount of supercharging in these solvents supports the above theory. Also, as a control, 99.9/0.1 water/FA was tested and we found that there is no supercharging in this solvent condition, which indicates that this effect is independent of the pH. \( m \)-Nitrophenol does supercharge to some extent in 50/50 ACN/water, which again indicates the necessity of a weak base like acetonitrile. This also explains the individual roles of ACN and formic acid in supercharging. When 50/50 ACN/water and 50/49.9/0.1 ACN/water/FA are compared, the latter solvent has lots of excess protons that can be donated to the protein and hence it supercharges more than 50/50 ACN/water alone.

4.4.4. Supercharging disulfide bonded proteins

\( \beta \)-lactoglobulin A (18.3 kDa) and \( \alpha \)-lactalbumin (14.1 kDa) are disulfide bonded proteins prevalent in milk with two and four disulfide bonds respectively. The isoelectric point of these proteins are equivalent (\( \text{pI}=4.8 \)) and these proteins are considered as acidic proteins. This is in direct contrast to lysozyme, which is a basic protein with a \( \text{pI} \) of 9.32 but shares 40% sequence homology with \( \alpha \)-lactalbumin and the three dimensional structures are also similar.\(^{17} \) As discussed earlier for lysozyme, the ESI charging of disulfide bonded proteins are reduced due to the spatial hinderance from the compactly folded structure and electrostatic repulsion from the existing protonated sites.
We wanted to explore if more charges can be added to these acidic disulfide bonded proteins by supercharging. Figure 4.6A and B are the ESI mass spectra of β-lactoglobulin A without and with 200 mM sulfolane obtained under denaturing conditions with ACN/H$_2$O/FA 50/49.9/0.1. The average charge increases from 15.3 to 19.5 and the maximum charge attained increases from 19+ to 23+ upon addition of sulfolane (Table 4.1). It is important to note that β-lactoglobulin A has 21 basic sites and a maximum of 19+ charges have been reported for the disulfide-intact form of this protein. Figure 4.6C and D show the charge increased mass spectra of α-lactalbumin without and with adding 200 mM sulfolane, with the average charge increasing from 9.3 to 13.5. Upon sulfolane addition, the maximum charge increases from 13+ to 17+ which is equivalent to the total number of basic sites in α-lactalbumin.

While the maximum charge obtained is more than the number of basic sites in β-lactoglobulin A only, α-lactalbumin shows a greater increase (45.2%) than β-lactoglobulin A (27.4%) when the number of average charges acquired is compared. In spite of the extremely different pI, lysozyme and α-lactalbumin show significant amount of supercharging, and this suggests that the supercharging effect in compactly folded proteins is independent of the overall charge of the protein. We also tested if other acidic proteins that do not have any disulfide bonds such as calmodulin (16.7 kDa, pI = 4.09) and actin (42.1 kDa, pI = 5.23) can be charge enhanced, since this will bring to light the supercharging ability of acidic proteins without the influence from the disulfide bonded architecture. Calmodulin and actin did not show any substantial increase in the average and maximum charges acquired in the presence of 200 mM sulfolane (Table 4.1 and Supplemental Figure 4.1 and 4.2).

We also investigated the ability to add more charges in two other disulfide bonded proteins, ribonuclease A (RNase A) and ribonuclease B (RNase B). These two proteins have exactly the same amino acid composition but only RNase B is glycosylated by attachment of N-linked oligosaccharides to Asn$_{34}$. Five predominant types of glycosylations have been reported
in RNase B, which differ in the number of mannose residues attached to the chitobiose core (Man5GlcNAc2, Man6GlcNAc2, Man7GlcNAc2, Man8GlcNAc2 and Man9GlcNAc2).\textsuperscript{18} The presence of these glycan groups imparts negative charges on the surface of RNase B and we were especially interested to find if the sugar residues impact supercharging. Figure 4.7A shows the ESI mass spectra of RNase A sprayed from denaturing solvents where the average charging increases from 11.2 to 15 upon addition of 200 mM sulfolane (Figure 4.7B), which corresponds to a 33.9% increase. RNase B, on the other hand, is not efficiently supercharged under the same conditions, the average charge being shifted from 11.2 to 13.7 which is a 22.3% increase (Figure 4.7C and D). It should also be noticed that due to the heterogeneity in the number of mannose residues present, the signal from the protein gets distributed in a wide range of $m/z$. This less dramatic shift in the number of charges gained can be explained by the bulky glycan groups that are present in the surface of the protein, which may cause a spatial hindrance and also impart negative charge in the surface of the protein which decreases the likelihood of picking more charges.

Supercharging reagents can also be used in the negative mode to abstract protons from the protein and hence being able to ionize with more negative charges in the negative mode. Sulfolane is both a very weak base and a weak acid, and has been reported to supercharge oligosaccharides in the negative mode.\textsuperscript{19} We wanted to see if the negative charges in the glycans in RNase B can be supercharged in the negative mode. RNase A and RNase B were diluted in ACN/H$_2$O 50/50 containing 10 mM ammonium acetate to facilitate charging in the negative mode. In RNase A, the base peak shifted by one charge state from 6- to 7- (Figure 4.7E and F), while there was no change in the base peak in RNase B upon supercharging (Figure 4.7G and H).

For RNase A which does not have any glycans to be ionized in the negative mode, our results are consistent with the results published recently by Douglass and Venter.\textsuperscript{8} These
authors supercharged cytochrome c with 100 mM sulfolane in the positive and negative modes in the presence of methanol/H$_2$O (96/4) as solvent. While the positive mode showed a dramatic increase in both average and maximum charges, there was no supercharging in the negative mode.$^8$ On the other hand, the observation that RNase B does not supercharge in the negative mode in spite of the presence of glycans is in contrast to the report by Huang et al., where they showed that heparan sulfate oligosaccharides can be supercharged with sulfolane.$^{19}$ While heparan sulfate is a linear acidic polysaccharide, RNase B is a polypeptide with both acidic and basic groups. More experiments are required to determine if supercharging in acidic proteins and glycosylated proteins is compromised due to the presence of competing acidic and basic groups.

4.4.5. Supercharging high MW proteins

Supercharging high MW proteins is advantageous since the increased charging shifts the mass spectrum into the lower $m/z$ region from the typical higher $m/z$ area where most proteins with a significant size ionize. We explored the multiple charging effects of high MW proteins such as bovine serum albumin (BSA, 66.5 kDa), transferrin (78 kDa), and IgG1 (148.5 kDa) by adding supercharging reagents. This will also help with answering the underlying question if high MW proteins can be charge shifted, since the charge density and size of the molecule play important roles in the supercharging phenomenon. 5-10 µM concentration of these proteins in ACN/H$_2$O/FA 50/49.9/0.1 was analyzed with and without the addition of reagents.

Figure 4.8A-F shows the MS profiles of transferrin with increasing concentrations of sulfolane ranging from 25 mM to 200 mM. The average charge increases from 38 without any sulfolane to 43.4 at 75 mM sulfolane, with intermediate concentrations showing proportional increase in average charge state (Table 4.2). Further increase in the amount of sulfolane to 100
mM and 200 mM did shift the charge state envelope of transferrin, but the protein-sulfolane adducts also increases and the activation energy applied is not enough to dissociate the adducts. Higher resolution is required to assign the charge state distributions obtained at high concentrations of the reagent. We also noticed that the product ions formed by applying higher voltages in the nozzle-skimmer region can be supercharged when higher concentration of sulfolane is added. The b-ions obtained are at 6+ and 7+ charge states without any reagent (Figure 4.9A), but addition of 200 mM sulfolane yields the same b-ions at 8+ and 9+ charges (Figure 4.9B), which becomes very useful in subsequent MS techniques to obtain more information.

Transferrin also showed similar supercharging effects with $m$-nitrophenyl ethanol in denaturing conditions (Table 4.2). Previously, Marshall’s group has shown charge increases for transferrin eluting from a reversed phase column by adding 5% DMSO $v/v$ in the LC solvent and they reported that a dramatic shift in the highest charge state from 67+ to 77+. The maximum charge attained by transferrin (52+) in our experiments was less when compared to their LC results, but this can be attributed to the high percentages of ACN that exists when transferrin elutes from the column. Moreover, regardless of the efficiency of supercharging, DMSO has additional conformational effects and proteins can be denatured at high DMSO concentrations.

Yet another similarly-sized protein as transferrin is BSA and addition of 200 mM sulfolane shifted the average charge of BSA from 51 to 59 (Figure 4.10) and the maximum charge increased from 68+ to 73+. With 22.5 mM $m$-nitrophenyl ethanol, similar results were obtained (Table 4.2).

IgG1 is almost twice the size of transferrin and we were interested to find if such large proteins can be charge enhanced. Mass spectra were obtained from desalted IgG1 samples by adding different supercharging reagents, since with sizeable proteins, it is necessary to try different reagents to make sure that the limited supercharging is not due to the reagent itself.
The average charge without any reagent was 50.7 and it shifted to 60.9 with the addition of \textit{m}-nitrophenyl ethanol (Figure 4.11). \textit{m}-NBA and sulfolane also increased the number of average charges to 64.2 and 62.7 respectively (Table 4.2). By increasing the charge on such large proteins like antibodies, it helps with top-down characterization by utilizing the charge increase in the precursor and subsequent product ions. It will be interesting to compare the sequence coverage obtained in disulfide-intact IgG1 with and without supercharging. The limits on the size and disulfide bonded architecture creates a sufficient level of hindrance to fragmenting such proteins and if supercharging can circumvent this requirement without the need to reduce the disulfide bonds, larger proteins can be characterized intact.

4.4.6. Supercharging membrane proteins

Membrane protein analysis is challenging due to the precipitation/aggregation issues encountered during sample preparation for proteomic investigations.\textsuperscript{21} The hydrophobic stretches of amino acids referred to as the ‘transmembrane domains’ require special solvent conditions to keep the protein soluble for analysis by mass spectrometry.\textsuperscript{22} We were interested to find if the membrane protein can be solubilized by adding supercharging reagents, since this will eliminate the requirement of special solvents for ESI-MS analysis. In addition, we wanted to explore the supercharging potential of membrane proteins since this will help with their characterization by analyzing the higher charge states formed during supercharging.

Bacteriorhodopsin (BR) is a classic example of an integral membrane protein that constitutes 75\% of the membrane of purple bacterium \textit{Halobacter salinarium}. Different groups have characterized this protein by precipitating it out of the membrane.\textsuperscript{23-26} After precipitation, the rhodopsin moiety falls off from BR and the \textit{apo}-protein is termed bacterioopsin (BO). We precipitated BO from the membrane using methanol-chloroform protocol\textsuperscript{11} and resolubilized the protein using a small amount of neat FA followed by diluting the FA about 25 fold by adding
Aliquots were taken from this preparation and different concentrations of sulfolane starting from 300 mM to 3 M were added to find if adding the supercharging reagent will yield a signal from BO under conditions that are devoid of denaturing solvents. Addition of 300 mM sulfolane to BO dissolved in water gave only negligible amount of signal in ESI-MS, and the signal intensity and CSD of BO was the same with and without adding sulfolane (Figure 4.12). There was no improvement in signal intensity with increasing concentrations of sulfolane. This can be attributed to the low solubility of BO in a polar solvent such as water and addition of sulfolane did not increase the solubility of BO in water. There is an older report on utilizing sulfolane as a solubilizing agent for hydrophobic proteins such as zein and globins, and the researchers employed circular dichroism to show that sulfolane is an effective solubilizing agent for water-insoluble proteins. They also demonstrated various properties of sulfolane such as stability of sulfolane in aqueous solutions, does not chemically modify proteins in comparison to other denaturants like urea, and its compatibility with electrophoretic and chromatographic techniques. Though our results contradict with their findings, similar proteins and experiments types have to be compared before making any further conclusions.

To confirm that solubility of BO is low in the absence of denaturing solvents, BO was resuspended in ACN/H$_2$O 50/50 (this step was done after solubilizing the precipitated protein pellet in neat FA as in earlier experiment). In the presence of 50% ACN, ESI-MS signal was much better and the maximum charge attained was 24+, with the average charge being 16.9 (Figure 4.13A). When 300 mM sulfolane was added to the BO in 50% ACN, the CSD narrowed to a few predominant charge states. The maximum charge attained did not increase and the average charge was 18.1 (Figure 4.13B).

The narrowing of the CSD upon addition of sulfolane was contrasting to the results obtained from all other proteins. To investigate more on this observation, different denaturing solvent conditions were tested to find out if this was related to the nature of the solvent. BO was
precipitated and solubilized in neat FA in the above-mentioned procedure and diluted in each of
the following solvents: chloroform/methanol/1%FA (4/4/1), ACN/H₂O 80/20, and
isopropanol/H₂O 50/50. In the presence of chloroform/methanol/1% FA (4/4/1), the classical
solvent for membrane protein analysis, the maximum and average charges decreased in the
presence of 300 mM sulfolane (Figure 4.14 and Table 4.2). All the other solvent conditions
tested also showed similar results (Table 4.2) and the general observation was that adding
sulfolane lead to charge state coalescence, which results in narrowing the signal to a few
predominant charge states. Recently, Meyer and Komives reported a similar effect for peptides
when DMSO was added as a supercharging reagent. They proposed that DMSO improves
spray stability and desolvation efficiency, which might result in increasing the total signal to a
few predominant charge states.

From our experiments on adding sulfolane to different classes of proteins in denaturing
and native conditions, we did not notice this kind of effect. To test if this effect is specific to
hydrophobic class of proteins such as BO, we tested BO with m-NBA. Addition of 0.5% m-NBA
did not narrow the CSD of BO. The maximum charge and average charge obtained remained
the same after adding m-NBA (Table 4.2). These results suggest that the charge state
coalescence effect was specific to BO and sulfolane co-addition, and is independent of the
choice of denaturing solvent, supercharging reagent used, or nature of the protein.

4.4.7. Supercharging peptides

Previously Jensen’s group have explored the supercharging effects of m-NBA on tryptic
peptides and phosphopeptides, and they have shown that the predominant charge increases for
peptides by adding 0.1% m-NBA in the LC solvent during reversed phase separations. We
chose to explore this multiple charging phenomenon of peptides by testing a few standard
peptides with supercharging compounds, since this will answer basic questions such as conformational effects during supercharging.

Reagents such as sulfolane, m-NBA, and m-nitrophenyl ethanol have been shown to increase charging across all sizes of proteins. When extending the same approach to peptides in denaturing solvent conditions, we found the ESI charging of small peptides can also be increased. The most abundant charge state of bradykinin (RPPGFSPFR) under denaturing conditions was 2+ (Figure 4.15A). In the presence of 25 mM sulfolane the most abundant charge shifts to 3+ (Figure 4.15B). Moreover, the maximum charge also increases to 4+ upon addition of sulfolane (maximum charge without sulfolane is 3+), despite the fact that there are only 3 potentially basic sites (two Arg residues and the N-terminus). Another aspect that is evident from supercharging peptides is that, the mechanism of supercharging may not based only on denaturation, since in biomolecules like peptides which do not have a significant amount of secondary structure, denaturation alone could not be the cause of supercharging. The roles played by other characteristics such as reagent basicity can explain the supercharging of peptides.9

The same effect was observed for substance P (RPKPQQFFGLM) where the most abundant charge state shifts from 2+ to 3+ with 25 mM sulfolane and a 4+ charge state was observed (Table 4.2). Substance P also has only three basic sites in the sequence where the charge could potentially reside. The presence of 4th charge in both bradykinin and substance P shows that the other residues such as glutamine or proline could potentially be the sites where an extra charge could remain under supercharging conditions.

4.4.8. Reagent concentration versus supercharging

To find the concentrations of various supercharging reagents necessary to enhance charging without losing signal intensity, lysozyme was tested with increasing concentrations of
supercharging reagents to find the concentration at which there is no further increase in charge. All the experiments were performed in denaturing conditions with 50/49.9/0.1 ACN/H₂O/ FA as the solvent to dilute the protein and to make appropriate concentrations of the supercharging reagents. The plots of reagent concentration against average charge and maximum charge are shown in Figure 4.16A-D and the corresponding mass spectra acquired at increasing concentrations of each supercharging reagent are displayed in Supplemental Figures 4.3-4.6.

The average charge acquired by lysozyme increases significantly upon addition of 50 mM sulfolane (equivalent to 0.47%) and ascends steadily until 250 mM concentration where an average charge of 16.4 is obtained and does not increase further at 300 mM concentration (Figure 4.16A and Supplemental Figure 4.3). The maximum charge acquired without any sulfolane is 15+ and at 150 mM sulfolane (equivalent to 1.41%), the maximum charge observed increases to 20+, which exceeds the number of basic sites and does not increase further with higher concentration.

With m-NBA, the average charge increases from +11.2 to +14.1 at 50 mM m-NBA (which is equivalent to 0.6% m-NBA) and at 100 mM m-NBA (equivalent to 1.19%), there is a very slight increase to 14.2, after which it does not increase (Figure 4.16B and Supplemental Figure 4.4). A maximum of 18 charges is attained with adding 100 mM m-NBA and there is no additional increase at 200 mM concentration (equivalent to 2.3%).

m-Nitrophenyl ethanol is slightly distinct, since the average charge of 14.1 acquired with 50 mM m-nitrophenyl ethanol drops to 14 and 13.3 at 100 and 200 mM concentrations respectively, while the maximum charge remains stable at 17+ at these concentrations (Figure 4.16C and Supplemental Figure 4.5). This drop in the supercharging trend is unexplainable at this point, but it is important to note that with some reagents there is not always a sigmoidal trend in the degree of supercharging. m-Nitrophenol shows the same amount of supercharging as m-NBA, with the average charge increased to a maximum of 14.4 at 200 mM concentration.
and the maximum charge does not increase above 18+ (Figure 4.16D and Supplemental Figure 4.6).

When equivalent concentrations of all the reagents are compared, sulfolane shows the highest amount of charge increase at the lowest possible concentration (50 mM). These results also indicate that, in terms of maximum charge, other reagents reach up to 18+, whereas sulfolane at 150 mM concentration generates 20+ charges on lysozyme. In addition, when the supercharging efficiency of sulfolane and m-NBA are compared, 100 mM sulfolane (equivalent to 0.94%) yields an average charge of 15.3 and a maximum charge of 19+ whereas 100 mM m-NBA (equivalent to 1.19%) generates an average charge of 14.2 and a maximum charge of 18+. These results are in contrast to the recent study by the Williams group where they reported that the supercharging efficiency of sulfolane is lower than m-NBA on a per mole basis comparison. However, the dissimilarity between the results could be due to the solvent conditions (denaturing in our study versus native pH in their report) to test the supercharging ability.

4.4.9. Protein conformation during supercharging

To investigate the effects of protein conformation on supercharging, we analyzed the ion mobility of proteins with and without supercharging. An ion in a compact-form has a higher mobility, hence a shorter drift time than the same ion in a more open conformation as it travels through the ion mobility separation device. The drift time distributions for multiply charged ions of transferrin before and after adding 50 mM sulfolane (Figure 4.17A and B) display the arrival times of different charge states. However, the ion mobility arrival time distributions for the same charge state of transferrin with and without supercharging could not be compared because the peak width of different charge states of transferrin increase upon supercharging (data not shown) due to formation of protein-sulfolane adducts and the charge states are also not
isotopically resolved. Therefore, we analyzed the ion mobility profile of a low MW protein such as lysozyme, and compared the drift time of a particular charge state with and without supercharging since the charge states can be isotopically resolved and a more accurate comparison of drift time can be performed.

The two dimensional plots of $m/z$ against drift time for lysozyme with and without 200 mM sulfolane in denaturing conditions (Figure 4.18A and B) and native solvent conditions (Figure 4.19A and B) are displayed and the corresponding mass spectra are shown in Figure 4.20A-D. Although it is ideal to compare the same charge state of lysozyme with and without supercharging in both denaturing and native solvent conditions, there is no common charge state that is present in the charge state distributions obtained for lysozyme at all the different conditions (Figure 4.20A-D). When the 9+ charge state of lysozyme in native solvent conditions is compared with and without supercharging (Figure 4.21A and B), there is no difference in the drift time obtained. Similar results were obtained when the 12+ charge state is compared in denaturing solvent conditions (Figure 4.21C and D). However, the Williams group had noticed that in native conditions when different charge states of lysozyme with and without reagent were compared, the collision cross sections of higher charge state ions were higher and they had concluded that supercharging denatures the protein and tends to cause more open conformation of the protein which in turn reduces the mobility. But from our data, there is a clear trend across all the charge states of lysozyme that the addition of the reagent does not cause any difference in the drift time in both native and denaturing conditions and hence our data suggests that the conformation of the protein remains unchanged up on supercharging.
4.5. Conclusions

Compounds such as sulfolane, nitrobenzonitrile, and \( m \)-nitrophenyl ethanol were found to be effective in enhancing charge from both native and denaturing environments. \( m \)-Nitrophenol and nitrobenzene were limited in solubility in aqueous solvents and did not affect the CSD of proteins. However, these two reagents were found to be potent supercharging agents in denaturing conditions. For \( m \)-nitrophenol, the supercharging ability was found to be greater in a weak base such as ACN when compared to other solvents studied. This highlights the importance of investigating the chemical properties of the supercharging reagents when probing the mechanisms behind supercharging. Disulfide bonded proteins displayed a significant amount of supercharging than the other classes of proteins. This aspect of increasing the number of charges acquired without the need for disulfide reduction is beneficial for characterizing disulfide bonded proteins. High MW proteins did not show the same extent of supercharging as the other classes of proteins and the charge density might play a role in limiting the number of charges in this class. In addition, the gas-phase electrophoretic mobility diameters of high MW proteins can be studied to find if any correlation exists between size and supercharging. Supercharging peptides showed that species that lack secondary structure can be supercharged and future investigations with synthetic peptides will be beneficial to find the roles played by non-basic amino acids in supercharging. Finally, the concentration of the supercharging reagent chosen is important since the loss of signal intensity and adduct formation can be limited by selecting appropriate concentration of the reagents.
4.6. References


4. Loo, J. A.; Edmonds, C. G.; Smith, R. D., Primary sequence information from intact proteins by electrospray ionization tandem mass spectrometry. Science 1990, 248, (4952), 201-204.


Figure 4.1. Supercharging insulin with different reagents in denaturing conditions. ESI-MS of insulin sprayed with (A) no reagent, (B) m-NBA, (C) sulfolane, (D) m-nitrobenzonitrile, (E) m-nitrophenol, and (F) nitrobenzene in denaturing solvent conditions.
Figure 4.2. Supercharging disulfide-intact lysozyme with different reagents in denaturing conditions. ESI-MS of disulfide-intact lysozyme in ACN/H₂O/FA 50/49.9/0.1 sprayed without any reagent (A), with 50 mM sulfolane (B), and 50 mM m-nitrophenyl ethanol (C).
Figure 4.3. Supercharging disulfide-reduced lysozyme with sulfolane in denaturing conditions. ESI-MS of lysozyme reduced with DTT and sprayed without any reagent (A) and with 200 mM sulfolane (B) in denaturing solvent conditions.
Figure 4.4. Supercharging disulfide-intact lysozyme with m-nitrophenol in denaturing and native solvent conditions. ESI-mass spectra of lysozyme sprayed by itself (A) and with addition of 50 mM m-nitrophenol (B) in denaturing conditions. (C) and (D) are the charge state distributions of lysozyme sprayed without and with 10 mM m-nitrophenol respectively, in 10 mM ammonium acetate buffer (native conditions).
Figure 4.5. Supercharging disulfide-intact lysozyme with \textit{m}-nitrophenol in different solvent systems. ESI-mass spectra of disulfide-intact lysozyme sprayed in different solvent conditions without and with 10 mM \textit{m}-nitrophenol (\textit{m}-NP). (A) 50/49.9/0.1 ACN/H\textsubscript{2}O/FA, (B) 50/49.9/0.1 ACN/H\textsubscript{2}O/FA with \textit{m}-NP, (C) 50/50 ACN/water, (D) 50/50 ACN/water with \textit{m}-NP, (E) 50/49.9/0.1 methanol/H\textsubscript{2}O/FA, (F) 50/49.9/0.1 methanol/H\textsubscript{2}O/FA with \textit{m}-NP, (G) 50/50 methanol/H\textsubscript{2}O, (H) 50/50 methanol/H\textsubscript{2}O with \textit{m}-NP, (I) 99.9/0.1 H\textsubscript{2}O/FA, (J) 99.9/0.1 H\textsubscript{2}O/FA with \textit{m}-NP, (K) 10 mM ammonium acetate buffer, and (L) 10 mM ammonium acetate buffer with \textit{m}-NP.
Figure 4.6. Supercharging β-lactoglobulin A and α-lactalbumin with sulfolane. (A) and (B) are ESI-MS of β-lactoglobulin A without and with 200 mM sulfolane respectively. (C) and (D) are the ESI-MS of α-lactalbumin without and with 200 mM sulfolane respectively. All spectra were obtained under denaturing conditions.
Figure 4.7. Supercharging RNase A and RNase B in positive and negative modes with sulfolane. RNase A and RNase B were sprayed without and with 200 mM sulfolane in positive and negative modes: (A) +ESI-MS RNase A no reagent, (B) +ESI-MS RNase A with sulfolane, (C) +ESI-MS RNase B no reagent, (D) +ESI-MS RNase B with sulfolane, (E) -ESI-MS of RNase A no reagent, (F) -ESI-MS of RNase A with sulfolane, (G) -ESI-MS of RNase B no reagent, and (H) -ESI-MS of RNase B with sulfolane. ACN/H$_2$O/FA 50/49.9/0.1 and ACN/water 50/50 with 10 mM ammonium acetate were used as ESI solvents for collecting the data in positive and negative modes respectively.
Figure 4.8. Supercharging transferrin with different concentrations of sulfolane. ESI-MS of transferrin acquired under denaturing solvent conditions with (A) no reagent, (B) 25 mM sulfolane, (C) 50 mM sulfolane, (D) 75 mM sulfolane, (E) 100 mM sulfolane, and (F) 200 mM sulfolane.
Figure 4.9. Comparing nozzle-skimmer dissociation of transferrin with and without sulfolane. Nozzle-skimmer dissociation of transferrin obtained without (A) and with 200 mM sulfolane (B). The b-ion assignments from N-terminal 56-61 amino acids at different charge states are labeled in different colors: 6+ (red), 7+ (green), 8+ (blue), and 9+ (black).
**Figure 4.10.** Supercharging BSA with sulfolane. ESI-MS of BSA in denaturing solvents acquired in (A) without any reagent and (B) with 200 mM sulfolane.
Figure 4.11. Supercharging IgG1 with $m$-nitrophenyl ethanol. ESI-MS of IgG1 in denaturing solvents acquired with (A) no reagent and (B) 22.5 mM $m$-nitrophenyl ethanol.
Figure 4.12. Supercharging bacterioopsin with sulfolane in aqueous conditions. ESI-MS of precipitated BO resuspended in a small volume of neat formic acid followed by 25 fold dilution in 100% water (A) and with addition of 300 mM sulfolane (B).
Figure 4.13. Supercharging bacterioopsin with sulfolane in denaturing conditions. ESI-MS of precipitated BO resuspended in a small volume of neat formic acid followed by 25 fold dilution in ACN/water 50/50 with (A) no reagent and (B) 300 mM sulfolane.
Figure 4.14. Supercharging bacterioopsin with sulfolane in chloroform/methanol/1% FA. ESI-MS of precipitated BO resuspended in a small volume of neat formic acid followed by 25 fold dilution in chloroform/methanol/1% FA (4/4/1) solvent with (A) no reagent and (B) 300 mM sulfolane.
**Figure 4.15.** Supercharging bradykinin with sulfolane in denaturing conditions. Nanoelectrospray mass spectra of bradykinin in ACN/H₂O/FA 50/49.9/0.1 solvent with no reagent added (A) and with addition of 25 mM sulfolane (B). The zoomed-in view shows the 4+ molecular ion of bradykinin which is present only when sulfolane is added.
Figure 4.16. Plots of reagent concentration versus average and maximum charge. Trends for average charge (blue) and maximum charge (red) as a function of concentration for (A) sulfolane, (B) m-NBA, (C) m-nitrophenyl ethanol, and (D) m-nitrophenol.
Figure 4.17. Ion mobility spectrometry analysis of transferrin in denaturing conditions. Plots of drift time versus m/z for transferrin with (A) no reagent and (B) 50 mM sulfolane.
Figure 4.18. Ion mobility spectrometry analysis of lysozyme in denaturing conditions. Plots of drift time versus \( m/z \) for lysozyme with (A) no reagent and (B) 200 mM sulfolane.
Figure 4.19. Ion mobility spectrometry analysis of lysozyme in native solvent conditions. Plots of drift time versus m/z for lysozyme with (A) no reagent and (B) 200 mM sulfolane.
Figure 4.20. Mass spectra of lysozyme acquired during drift time analysis. ESI-MS of lysozyme in denaturing solvent conditions without (A) and with 200 mM sulfolane (B). The mass spectra in (C) and (D) were acquired in native solvents without and with 200 mM sulfolane respectively. To compare the drift times with and without supercharging, the 12+ charge state was selected in denaturing conditions and the 9+ charge state was selected in native conditions.
Figure 4.21. Ion mobility arrival time distributions for selected charge states of lysozyme with and without supercharging. Plots of drift time for the 9+ charge state of lysozyme obtained in native solvent conditions without (A) and with 200 mM sulfolane (B). (C) and (D) are the drift time plots for the 12+ charge state of lysozyme acquired in denaturing solvent conditions without and with 200 mM sulfolane respectively.
Table 4.1. Summary of average and maximum charges acquired by supercharging different proteins with various reagents. Average and maximum number of charges obtained with and without adding supercharging reagents for insulin, lysozyme, β-lactoglobulin A, α-lactalbumin, RNase A, RNase B, calmodulin, and actin are tabulated. The solvent condition and concentration of reagent tested are also provided.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Solvent conditions</th>
<th>Number of charges without any reagent</th>
<th>Supercharging reagent with concentration</th>
<th>Number of charges with supercharging reagent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Average charge</td>
<td>Maximum charge</td>
<td>Average charge</td>
</tr>
<tr>
<td>Insulin</td>
<td>ACN/H2O/FA 50/49.9/0.1</td>
<td>4.4</td>
<td>6+</td>
<td>22 mM m-nitrobenzonitrile</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>300 mM sulfolane</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1% m-NBA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20 mM m-nitrophenol</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1% nitrobenzene</td>
</tr>
<tr>
<td>Lysozyme disulfide-intact</td>
<td>ACN/H2O/FA 50/49.9/0.1</td>
<td>11.1</td>
<td>15+</td>
<td>50 mM sulfolane</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100 mM sulfolane</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50 mM m-nitrophenyl ethanol</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50 mM m-nitrophenol</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10 mM Amm. Acetate</td>
</tr>
<tr>
<td>Lysozyme disulfide-reduced</td>
<td>ACN/H2O/FA 50/49.9/0.1</td>
<td>16.5</td>
<td>22+</td>
<td>200 mM sulfolane</td>
</tr>
<tr>
<td>β-lactoglobulin A</td>
<td>ACN/H2O/FA 50/49.9/0.1</td>
<td>15.3</td>
<td>19+</td>
<td>200 mM sulfolane</td>
</tr>
<tr>
<td>α-lactalbumin</td>
<td>ACN/H2O/FA 50/49.9/0.1</td>
<td>9.3</td>
<td>13+</td>
<td>200 mM sulfolane</td>
</tr>
<tr>
<td>RNase A positive ESI</td>
<td>ACN/H2O/FA 50/49.9/0.1</td>
<td>11.2</td>
<td>15+</td>
<td>200 mM sulfolane</td>
</tr>
<tr>
<td>RNase A negative ESI</td>
<td>ACN/H2O/50/50 10 mM Amm. Acet.</td>
<td>6</td>
<td>7-</td>
<td>200 mM sulfolane</td>
</tr>
<tr>
<td>RNase B positive ESI</td>
<td>ACN/H2O/FA 50/49.9/0.1</td>
<td>11.2</td>
<td>15+</td>
<td>200 mM sulfolane</td>
</tr>
<tr>
<td>RNase B negative ESI</td>
<td>ACN/H2O/50/50 10 mM Amm. Acet.</td>
<td>6.3</td>
<td>7-</td>
<td>200 mM sulfolane</td>
</tr>
<tr>
<td>Calmodulin</td>
<td>ACN/H2O/FA 50/49.9/0.1</td>
<td>13.7</td>
<td>19+</td>
<td>200 mM sulfolane</td>
</tr>
<tr>
<td>Actin</td>
<td>ACN/H2O/FA 50/49.9/0.1</td>
<td>38.2</td>
<td>53+</td>
<td>200 mM sulfolane</td>
</tr>
<tr>
<td>Protein</td>
<td>Solvent conditions</td>
<td>Number of charges without any reagent</td>
<td>Supercharging reagent with concentration</td>
<td>Number of charges with supercharging reagent</td>
</tr>
<tr>
<td>---------</td>
<td>--------------------</td>
<td>---------------------------------------</td>
<td>----------------------------------------</td>
<td>------------------------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Average charge</td>
<td>Maximum charge</td>
<td>Average charge</td>
</tr>
<tr>
<td>BSA</td>
<td>ACN/H$_2$O/FA 50/49.9/0.1</td>
<td>51</td>
<td>68+</td>
<td>200 mM sulfolane</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>22.5 mM m-nitrophenyl ethanol</td>
</tr>
<tr>
<td>Transferrin</td>
<td>ACN/H$_2$O/FA 50/49.9/0.1</td>
<td>38</td>
<td>46+</td>
<td>25 mM sulfolane</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50 mM sulfolane</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>75 mM sulfolane</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100 mM sulfolane</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>200 mM sulfolane</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>22.5 mM m-nitrophenyl ethanol</td>
</tr>
<tr>
<td>IgG1</td>
<td>ACN/H$_2$O/FA 50/49.9/0.1</td>
<td>50.7</td>
<td>62+</td>
<td>50 mM sulfolane</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>22.5 mM m-nitrophenyl ethanol</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.5 % m-NBA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10 mM m-nitrophenol</td>
</tr>
<tr>
<td>BO</td>
<td>ACN/H$_2$O 50/50</td>
<td>16.9</td>
<td>24+</td>
<td>300 mM sulfolane</td>
</tr>
<tr>
<td></td>
<td>CHCl$_3$/MeOH/1%FA (4/4/1)</td>
<td>20</td>
<td>31+</td>
<td>200 mM sulfolane</td>
</tr>
<tr>
<td></td>
<td>CHCl$_3$/MeOH/1%FA (4/4/1)</td>
<td>20</td>
<td>31+</td>
<td>0.5 % m-NBA</td>
</tr>
<tr>
<td></td>
<td>ACN/H$_2$O 80/20</td>
<td>20.2</td>
<td>31+</td>
<td>200 mM sulfolane</td>
</tr>
<tr>
<td></td>
<td>IPA/H$_2$O 50/50</td>
<td>21.2</td>
<td>31+</td>
<td>200 mM sulfolane</td>
</tr>
<tr>
<td>Bradykinin</td>
<td>ACN/H$_2$O/FA 50/49.9/0.1</td>
<td>2+</td>
<td>3+</td>
<td>25 mM sulfolane</td>
</tr>
<tr>
<td>Substance P</td>
<td>ACN/H$_2$O/FA 50/49.9/0.1</td>
<td>2+</td>
<td>3+</td>
<td>25 mM sulfolane</td>
</tr>
</tbody>
</table>

**Table 4.2.** Summary of average and maximum charges acquired by supercharging high MW proteins and peptides with different reagents. Average and maximum number of charges obtained with and without adding supercharging reagents for high MW proteins (BSA, transferrin, and IgG1), BO, and peptides (bradykinin and substance P) are tabulated. The solvent condition and concentration of reagent tested are also provided.
**Supplemental Figure 4.1.** Supercharging calmodulin with sulfolane. ESI-MS of calmodulin in denaturing solvents acquired with (A) no reagent and (B) 200 mM sulfolane.
Supplemental Figure 4.2. Supercharging actin with sulfolane. ESI-MS of actin in denaturing solvents acquired with (A) no reagent and (B) 200 mM sulfolane.
**Supplemental Figure 4.3.** Supercharging lysozyme with increasing concentrations of sulfolane in denaturing conditions.
Supplemental Figure 4.4. Supercharging lysozyme with increasing concentrations of \( m\)-NBA in denaturing conditions.
Supplemental Figure 4.5. Supercharging lysozyme with increasing concentrations of \( m \)-nitrophenyl ethanol in denaturing conditions.
Supplemental Figure 4.6. Supercharging lysozyme with increasing concentrations of \textit{m}-nitrophenol in denaturing conditions.
5.1. Abstract

Supercharging reagents enhance the ESI multiple charging of proteins and peptides, and this shifts the charge state envelope acquired by the protein to a lower mass-to-charge region where it can be analyzed at a high resolution by most mass spectrometers. In addition, selecting the most abundant charge state upon supercharging yields higher fragmentation efficiency and hence, sequence-informative product ions are obtained in contrast to selecting the most abundant charge state without supercharging. So far, supercharging has been mostly studied by infusing targeted proteins/peptides with the reagent being added in the solvent used for ESI. Only a few groups have demonstrated charge enhancements for proteins/peptides eluting from a liquid chromatography (LC) column by adding the supercharging reagents to the LC solvents. The compatibility and efficiency of other potent reagents such as sulfolane remains to be explored. Here, we report the supercharging ability of sulfolane for intact proteins of a wide MW range during HPLC-MS. Other new methods of reagent introduction such as in the vapor phase during LC-MS were also explored. In addition to the charge enhancing effect, the introduction of the reagent in the vapor phase shows increased signal intensity when compared to experiments performed without the reagent and thus can be utilized as a way to enhance signal during circumstances where the signal intensity is low.
5.2. Introduction

The multiple charging phenomena observed during electrospray ionization (ESI) for mass spectrometry extends the typical limits of the common mass analyzers (2000 m/z) to measure high MW species.1 Protein-protein, protein-ligand, and covalent modifications also can greatly increase the MW of the complex/molecule and are efficiently analyzed by the merits of the multiple charging. In addition to the effective mass measurement, probing the primary sequence of the biomolecule by tandem MS methods are also dependent on the charge state chosen for dissociation. Several studies have reported the benefits of selecting highly charged species for fragmentation schemes based on electron transfer/capture.2-4

Enhanced multiple charging was introduced by the Williams group when they demonstrated that the extent of multiple charging observed for proteins and peptides during ESI can further be increased by adding reagents such as m-nitrobenzyl alcohol (m-NBA).5 This finding stirred further research in this area and lead to discovering new reagents by different groups.6, 7 Currently, the mechanism(s) by which these reagents increase charge are debatable.8, 9 However, it is evident that the supercharging effect varies with solvent conditions (native/denaturing), chemical nature of the reagent (weakly basic/weakly acidic/denaturant), and also on the pl (acidic/basic) and MW of the protein. Regardless of these variations, these charge enhancing reagents have a wide variety of applications for both targeted and complex proteomic studies.

Increased multiple charging in native solvent conditions (at near neutral pH, without any organic solvent) is beneficial for targeted analysis of noncovalent interactions. Yin and Loo demonstrated that characterizing protein-ligand complexes can be improved by adding supercharging reagents. When the higher charge states obtained by supercharging were interrogated by CAD and ECD, it yielded information about the ligand-bound form of the protein in contrast to fragmenting the charge states from the non-supercharged form of the complex
that yielded product ions pertaining to the apo-form.\textsuperscript{10} Brodbelt’s group showed that IRMPD of highly charged peptides obtained by supercharging resulted in increased sequence coverage and yielded more diagnostic ions which help with characterizing the modifications present.\textsuperscript{11} These reports have highlighted the advantages of dissociating the higher charge state ions obtained by supercharging by different fragmentation techniques and have ultimately laid the foundation for its application in high throughput complex analysis.

For large-scale protein and peptide analysis, supercharging reagents can be added as modifiers to LC solvents to facilitate increased charging during elution from the column. Jensen’s group first implemented this approach by adding \textit{m}-NBA to the LC solvents and showed that ETD fragmentation efficiency and hence, sequence coverage increased for tryptic peptides and phosphopeptides during on-line LC-MS/MS.\textsuperscript{12} The Marshall lab extended this idea to proteins and they showed that intact proteins eluting from a column can be efficiently supercharged by adding reagents such as DMSO and DMF.\textsuperscript{7}

During LC-MS analysis, the signal from the protein is constricted to the elution time window (usually 20-45 seconds for medium to high MW proteins) and medium to high MW proteins tend to have broad charge state envelopes that often stretches to the higher \textit{m/z} region. High resolution instruments such as FT-ICR are often employed for analyzing high MW proteins and the resolving power of this instrument is inversely proportional to \textit{m/z}.\textsuperscript{13} The resolving power achievable is also directly proportional to the time and frequency at which the data is acquired. Higher \textit{m/z} ions have lower frequency and their resolving power is lower compared to their lower \textit{m/z} (i.e., higher frequency) counterparts. Because the resolving power is directly related to the length of the acquired transient, if higher resolving power is required for high \textit{m/z} ions, a longer transient must be acquired. Therefore, experiments which require high resolving powers will have a much lower duty cycle. From an LC perspective, this means that
LC experiments in which the MS is acquired at higher resolving powers will have fewer data points across the LC run, and therefore a lower LC resolution.

On-line supercharging with HPLC has significant advantages since enhanced charging during the elution timescale shifts the charge state envelope to a lower m/z region where it can be efficiently analyzed at high resolution without compromising the scan speed. At higher resolution, the mass measurement accuracy of these high MW species increases which becomes important for unambiguous protein identification in complex samples. Furthermore, during LC-MS/MS, the mass spectrometers are programmed to select the ‘most abundant’ charge states for fragmentation and the intense charge states chosen are not usually the highly charged ones. Many supercharging reagents shift the charge state envelope such that the maximum charge obtained without supercharging becomes the most abundant charge state with the addition of the reagent. When the most abundant charge state obtained during supercharging is selected, it has the potential to increase sequence coverage during on-line LC-MS/MS. In addition, the product ions formed during supercharging could in turn be highly charged which makes its mass measurement more efficient.

Since the LC-MS analysis takes place in denaturing conditions, it is worthwhile to explore the capabilities of reagents which have shown promise in denaturing conditions. Sulfolane is one such reagent and its charge enhancing efficiency in denaturing solvents is found to be higher than some of the other reagents compared.¹⁴ We have explored the efficiency of sulfolane for on-line LC-MS analysis of intact proteins by adding it to the LC solvent and we discuss the solvent compatibility, column performance, and supercharging ability when sulfolane is added to the LC solvent. In addition, we have also described methods by which supercharging reagents can be delivered in the vapor phase during ESI and this stimulates further research in this area to find novel ways by which supercharging reagents can be introduced.
5.3. Experimental section

5.3.1. Sample preparation

All protein standards and supercharging reagents (sulfolane and m-NBA) were purchased from Sigma-Aldrich (St. Louis, MO) unless otherwise stated. Standard proteins (bradykinin, neurotensin, ubiquitin, ribonuclease A, transferrin, BSA, and carbonic anhydrase) were made at 1 mg/mL concentration in MS grade water. An equimolar mixture of these 7 proteins (10 pmoles each) was prepared by diluting in 0.1% formic acid (FA) and used for experiments with the Agilent column. For experiments with the Dionex column, an equimolar mixture of ubiquitin, RNase A, lysozyme, and myoglobin was prepared and diluted to 1 pmol of each protein in 5% ACN 0.1% FA.

5.3.2. LC conditions and MS parameters

5.3.2.1. Supercharging during LC-MS by adding reagents directly to LC solvents

5 µL of the standard protein mix containing 10 pmoles of each protein was injected into the Agilent C3 column (pore size 300 Å) with dimensions 2.1 mm x 150 mm. LC-MS was performed with an Agilent 1200 LC system connected on-line with an Agilent 6220 ESI-TOF mass spectrometer (Santa Clara, CA). Solvent A was 0.1% FA and solvent B was 90% ACN/10% H₂O containing 0.1% FA. For on-line supercharging experiments, 100 mM sulfolane was added to solvent A and B. The flow rate was 0.2 mL/min and the following gradient was used: 5% B for 3 min, 5%-25% B in 1 min, 25%-55% B in 30 min, 55%-80% B in 2 min, and held at 80% B for 9 min. The column was finally equilibrated with 5% B for 10 min prior to next run. The column temperature was maintained at 60° C throughout the run.

The source parameters in the MS were as follows: ESI capillary voltage 4300 V, drying gas temperature 350° C, drying gas flow rate 10 L/min, nebulizer pressure 25 psig, and
fragmentor voltage 250 V. The data was acquired at the rate of 1 spectrum / second and the acquisition window was set from 100 to 3000 \( m/z \).

5.3.2.2. Supercharging during LC-MS by delivering reagent vapors

Chromatography was performed using a Dionex Ultimate 3000 rapid separation LC system (Thermo Scientific, Waltham, MA), which consists of a binary high-pressure gradient pump and a ternary low-pressure loading pump. 5 \( \mu \)L of the standard protein mix containing 1 pmole of each protein was injected into a Dionex Pepswift reversed phase monolithic trap column (200 \( \mu \)m x 5 cm) and washed for 10 min with 100% water 0.05% TFA at a flow rate of 5 \( \mu \)L/min using the loading pump. This was followed by gradient elution into the Dionex Pepswift reversed phase monolithic analytical column (100 \( \mu \)m x 25 cm). 100% water 0.1% FA (solvent A) and 100% ACN 0.1% FA (solvent B) were delivered at a flow rate of 1 \( \mu \)L/min and the following gradient was used: 10% B for 10 min, increased to 25% B at 15 min, 50% B at 49 min, 80% B at 54 min, held at 80% B for 5 min, and finally equilibrated at 10% B for 10 min. Both the analytical and trap columns were held at 60° C throughout the run to aid protein separations.

The Dionex nanoLC system was connected in-line with Bruker solariX hybrid Qq-15 T FT-ICR mass spectrometer (Bruker Daltonics, Billerica, MA). The CaptiveSpray source (Bruker Daltonics, Billerica, MA) was used to couple the LC flow directly into the inlet of the mass spectrometer. It consists of a non-tapered 20 \( \mu \)m i.d. silica tip which produces a stable spray with minimized clogging. The MS source parameters were optimized for the LC flow rate of 1 \( \mu \)L/min and were as follows: drying gas flow of 4 L/min at a temperature of 250°C and a capillary voltage of -1.5 kV was applied. The capillary exit and deflector plate were at 250 V and 230 V respectively. The funnel 1 and skimmer 1 were held at 140 V and 30 V respectively. The data was acquired in the broadband detection mode, using 1 MHz data points, with a mass range of 150 – 3000 \( m/z \). The total duration of a MS scan event was 0.63 second, which includes a
transient length of 0.34 second. The number of scans averaged for one spectrum was set to one.

Sulfolane and m-NBA were diluted in 1:4 ratio in a volatile solvent such as isopropanol. Supercharging experiments were performed by applying the reagents either directly to the syringe filter (Figure 5.1A) or by spotting it in the inside walls of the Luer-lock filter adapter (Figure 5.1B) which connects the filter to the CaptiveSpray source (Figure 5.1C). In addition to using the air filter, another setup as described below was used to deliver the vapors generated from adding the supercharging reagents to a volatile solvent. 300 mM sulfolane and 0.3% m-NBA were each solubilized separately in 100% ACN in a solvent bottle. The vapors from this bottle were sent inside the CaptiveSpray source by attaching a piece of tubing from the air filter inlet in the source to the headspace of the solvent bottle (Figure 5.2). The tubing that goes inside the bottle was monitored to make sure that it does not get in touch with the solvent in the bottle which would otherwise pull the solvent by the suction created by the vacuum in the source region of the instrument. Another piece of tubing was immersed in the solvent with the other end left open to keep it vented. During on-line supercharging, this setup was maintained throughout the run time.

5.3.3. Calculation of average and maximum charges

Charge state distribution (CSD) refers to the observed set of charges acquired by a protein during ESI-MS. From the CSD, the average charge was calculated using the formula: \( \frac{\sum q_i I_i}{\sum I_i} \) from i to N, where \( q_i \) is the charge of the i-th charge state, \( I_i \) is the intensity of the i-th charge state and N is the number of measured charge states in a mass spectrum. Average charge can be simply correlated to the charge state with the highest intensity in the CSD obtained for a given protein. When comparing the supercharging effect, the maximum charge state or the highest observed charge state was also taken into consideration. While average
state gives an estimate for the overall charge distribution upon supercharging, the maximum charge state is an indication of the highest number of charges that can be attained without limitations due to electrostatic repulsion from the existing charges.

5.4. Results and discussion

5.4.1. Supercharging proteins during LC-MS with sulfolane in the LC solvent

Previously, sulfolane has been shown to charge enhance proteins in native conditions. Recent results by infusion experiments from our group and Venter’s group have shown that this reagent also works well under denaturing conditions. We wanted to extend this charge enhancement to proteins eluting from a column during LC since, increasing charge is beneficial to identify and characterize the proteins in a complex mixture when analyzed by LC-MS.

100 mM sulfolane was added to both solvent A and B to keep the amount of sulfolane uniform throughout the LC run. It is important to keep the concentration of sulfolane low since, sulfolane adducts are observed at higher charge states at increasing sulfolane concentrations. These adducts are resolved well in low MW proteins, but for high MW proteins, the activation energy applied in the source region is not sufficient to dissociate these adducts and so, the charge state envelope of protein appears as an unresolved hump. The 100 mM concentration was chosen since during infusion experiments in denaturing conditions with high MW proteins such as BSA and transferrin, it showed the maximum amount of supercharging without forming unresolved charge state envelopes.

In the total ion chromatograms (TIC) obtained without and with 100 mM sulfolane in the LC solvents (Figure 5.3A and B), the order of elution of all the proteins in the standard protein mix is preserved. However, retention time ($t_R$) of the proteins get shorter with sulfolane addition. This decrease in $t_R$ upon addition of sulfolane can be attributed to the increase in the non-polar nature of the solvent. Bradykinin, RNase A, and neurotensin (denoted as peaks 1, 2, and 3 in
Figure 5.3A) elute 0.4 min early when sulfolane was added (Figure 5.3B). Shift in the $t_R$ was highest for ubiquitin and hemoglobin β chain (labeled as peaks 4 and 6 respectively), which elute 2.2 min and 1.2 min early respectively. Hemoglobin β chain was found to be a contaminant from bovine carbonic anhydrase sample. High MW proteins (peaks 7, 8, and 9) elute 1.5 min early on average with solvents containing sulfolane. Previously, different groups have reported such a decrease in $t_R$ for peptides and proteins when $m$-NBA and DMSO were added to LC solvents.\textsuperscript{7,12} In addition, Jensen \textit{et al.} had noticed an increase in peak width for peptides eluting from the column with increasing concentrations of $m$-NBA.\textsuperscript{12}

Also, a small peak labeled 5 in Figure 5.3B appears only during sulfolane addition and it corresponds to a MW of 8591.28 Da which is 26.22 Da more than the MW of ubiquitin, 8565.06 Da. While the chromatographic resolution is high enough to separate this species from ubiquitin, this peak is unique to sulfolane addition and does not correspond to the any of the well-known protein-salt adducts.

Regarding the multiple charging effect, for small proteins such as RNase A, the average charge increases from 9.6 to 14.4 when sulfolane was present in the solvent (Figure 5.4A and B). For the high MW proteins, BSA and transferrin, in spite of choosing a concentration that showed well-resolved charge states during infusion experiments, the protein-sulfolane adducts predominated and resulted in unresolved hump-like distribution (Figure 5.5A and B). The concentration of sulfolane in the solvents was dropped to 20 mM to find if resolved CSD can be obtained for high MW proteins. While the low sulfolane concentration restored the $t_R$ close to that of no sulfolane addition in the TIC (Figure 5.3C), the unresolved CSD of high MW proteins still existed (Figure 5.5C). On the low MW proteins such as RNase A, the 20 mM sulfolane still showed effective multiple charging similar to 100 mM sulfolane (Figure 5.4C). Furthermore, the peak 5 which corresponded to MW of 8591 Da was still present at 20 mM concentration of sulfolane (Figure 5.3C).
The addition of sulfolane to chromatographic solvents also resulted in other undesirable changes such as irreproducible retention times between different runs setup with sulfolane. Furthermore, when the column was run without any sulfolane, the \( t_R \) did not revert to the original values as in the TIC in Figure 5.3A. Sulfolane permanently changes the active sites in the column and efforts to regenerate the column did not alter this situation.

5.4.2. Supercharging proteins by spotting reagents in the CaptiveSpray air filter

Enhanced multiple charging for proteins eluting on-line from a column by adding reagents to the LC solvents is advantageous, however, the chromatographic effects and column performance are key characteristics to be kept under optimum conditions. Retention time changes, loss of column integrity, and inability to revert to the conventional solvent system are all impacted during reagent addition to LC solvents. In order to overcome these unfavorable chromatographic issues, other methods to deliver reagents to proteins eluting on-line can be considered.

The CaptiveSpray source couples an LC column through a capillary spray tip directly into the inlet of the mass spectrometer. A vacuum seal is then created and ambient air is pulled in and around the spray tip creating an air funnel that sweeps the spray tip and desolvates the sample ions as they are pulled into the instrument. A syringe air filter removes the contaminants present in the ambient air environment and thus reduces the background ions. These air filters can be adapted to spray lock-mass calibrants for on-line LC-MS experiments and the calibrant delivery can be influenced by increasing the source parameters such as drying gas flow and temperature. We extended the use of air filters to deliver the supercharging reagent during ESI by spotting the reagent in it.

This method of applying the reagent was tested by infusing a sample of lysozyme in denaturing solvent conditions and installing different filters with sulfolane or \( m \)-NBA spotted. As
a control, a filter with no reagent applied was used. With a filter that had no reagent, the average charge obtained by lysozyme was 10.5 (Figure 5.6A) and this increased to 10.8 with the filter that had \textit{m}-NBA (Figure 5.6B). When the air filter supplemented with sulfolane was placed in the source, the average charge increased to 10.9 (Figure 5.6C). We also noticed a slight increase in the signal intensity in the presence of filters with supercharging reagents. When the relative ion intensities with and without the filter are compared, there is a 2.5 fold increase in the presence of either of the supercharging reagents. Other adjustments such as increasing the drying gas flow and temperature were tried, since the drying gas heats the source block which might enhance the reagent to vaporize more. However, the increase in charge acquired was still the same. It is possible that the concentration of supercharging reagent in the vapor phase is low. Sulfolane and \textit{m}-NBA have high boiling points and consequently the vapor pressure of these reagents is low. For instance, the boiling point of sulfolane is \(287^\circ\text{C}\)\(^{16}\) and the vapor pressure exerted by sulfolane is \(6.15\times10^{-4}\) mm Hg.\(^{17}\) For a comparison, the vapor pressure of ACN is \(90\) mm Hg.\(^{18}\) The relatively low concentration of sulfolane in the vapor phase could be limiting the charge enhancement in this experimental setting.

\textbf{5.4.3. Supercharging during LC timescale using reagent vapors}

Since spotting the reagents in the air filter showed only modest increments in the average and maximum charges, other methods such as adding the reagents in a volatile solvent such as ACN followed by delivering the vapors generated was explored. At room temperature, ACN is highly volatile and when the ACN vapors transported into the source, the vapors from the supercharging reagents can also be delivered. A standard protein cocktail was analyzed by LC-MS without and with solvent mediated reagent delivery as shown in the Figure 5.2 and the TICs obtained are displayed (Figure 5.7A and B). This method of reagent delivery does not
impact column performance and retention time which are the main advantages when compared to adding the reagents to LC solvents.

Figure 5.8A and B are the mass spectra of RNase A obtained without and with 300 mM sulfolane in the delivery solvent. The average charge shifted from 9.2 to 9.8 upon transfer of sulfolane in the vapor phase and a maximum charge of 14+ is attained only when sulfolane was present. Similar results were obtained when 0.3% \textit{m}-NBA was added to ACN (data not shown). Other proteins in the chromatogram also showed similar slight charge increases as shown in Figures 5.8 and 5.9 for ubiquitin and myoglobin respectively. The average charge for ubiquitin increased from 7.8 to 8.8 and for myoglobin it increased from 17.6 to 18.2. Similar to spotting the reagents in the air filter, this method also showed a 2-3 fold increase in the signal intensity for RNase A and ubiquitin, while the intensity of myoglobin remained the same with and without sulfolane vapor.

Signal enhancement is interesting since most supercharging reagents decrease the signal intensity while increasing charge, mostly due to adduct formation which results in splitting the signal among many peaks. Recently, our group and Chen’s group have noticed the similar signal enhancing effects with the addition of sulfolane and \textit{m}-NBA in solvents containing 0.1% trifluoro acetic acid (TFA) as the modifier.\textsuperscript{19} In both infusion and LC-MS experiments, TFA was used as a modifier and supercharging reagents were introduced in the plume by desorption electrospray ionization (DESI). The authors postulate that the supercharging reagents might interact with TFA anions and limit the association of TFA anions with the protein ions, which might result in enhancing the signal in the presence of TFA. While TFA has excellent benefits by working as an ion-pairing agent during LC, it causes signal suppression in MS even at very low concentrations. Future experiments can be devised to get the benefits of using TFA in the LC, while supercharging reagents can be delivered in the vapor phase using the CaptiveSpray air filters to enhance the signal intensity in the MS.
5.5. Conclusions

Supercharging during on-line LC-MS/MS has tremendous advantages for both bottom-up and top-down proteomic studies. Enhancing charge while the proteins/peptides are getting eluted in the chromatographic timescale is beneficial for both MS and tandem MS. While addition of sulfolane to the LC solvents enhances the charge for both medium and high MW proteins, unfavorable chromatographic effects are also associated with it. Reagent delivery in the vapor phase overcomes the disadvantages encountered during LC solvent addition, but the charge enhancing effect is moderate. However, the increase in the signal intensity obtained during reagent delivery in the vapor phase is beneficial and can be considered for proteins with low ionization efficiency, when sample amounts loaded on the column is low or when TFA is used for performing chromatography. Very recently, Tsybin’s group utilized a dual spray tip for ESI to infuse the supercharging reagents such as m-NBA or sulfolane through one spray tip and protein was infused through the other tip separately, and the supercharging reaction takes place just before ion liberation from the ESI droplet.\textsuperscript{20} With the development of innovative methods to introduce the reagent either in solution/vapor phase just before the formation of electrospray plume and with testing other supercharging reagents, the benefits of charge enhancement during LC-MS/MS can be achieved.
5.6. References


Figure 5.1. Supercharging reagent delivery through CaptiveSpray air filter. Sulfolane or m-NBA was spotted either directly on the syringe filter (A) or on the Luer-lock adapter (B). The syringe filter and adapter were re-installed and fixed on the CaptiveSpray source (C).
Figure 5.2. Supercharging reagent delivery in the vapor phase using 100% ACN. 300 mM sulfolane or 0.3% m-NBA was solubilized in 100% ACN in a solvent bottle and the vapors from this bottle were sent inside the CaptiveSpray source by attaching a piece of tubing (A) from the headspace of the solvent bottle to the air filter inlet in the source. Another piece of tubing (B) was immersed in the solvent with the other end left open in order to keep it vented.
Figure 5.3. Supercharging during LC-MS with sulfolane in the LC solvent. Total ion chromatograms of a standard 7 protein mix run through Agilent C3 column with LC solvents containing (A) no sulfolane, (B) 100 mM sulfolane, and (C) 20 mM sulfolane. The peaks labeled 1 through 9 represent the following proteins: bradykinin (1), RNase A (2), neurotensin (3), ubiquitin (4), unknown protein with MW 8591 Da (5), hemoglobin chain (6), transferrin (7), BSA (8), and carbonic anhydrase (9).
Figure 5.4. Mass spectra of RNase A with sulfolane in the LC solvent. Mass spectra of RNase A (extracted from peak 2 in the TIC in Figure 5.3) eluting from the C3 column with LC solvents containing (A) no sulfolane, (B) 100 mM sulfolane, and (C) 20 mM sulfolane. The average and maximum charge acquired increases upon sulfolane addition to the LC solvent.
Figure 5.5. Mass spectra of BSA with sulfolane in the LC solvent. Mass spectra of BSA (extracted from peak 8 in the TIC in Figure 5.3) eluting from the C3 column with LC solvents containing (A) no sulfolane, (B) 100 mM sulfolane, and (C) 20 mM sulfolane. The average and maximum charge increased upon adding 100 mM sulfolane to the LC solvent, but the CSD is unresolved due to the presence of sulfolane adducts. Decreasing the sulfolane concentration to 20 mM reduced the adduct formation and resulted in a modest resolution of the charge states.
Figure 5.6. Mass spectra of lysozyme with supercharging reagents spotted in the air filters. Mass spectra of lysozyme obtained with air filters containing (A) no reagent, (B) \textit{m}-NBA, and (C) sulfolane. Though the increase in the average charge is modest upon adding reagents, the relative intensities of the peaks in the CSD increase during supercharging when compared to no reagent addition.
Figure 5.7. Supercharging during LC-MS with sulfolane delivery in the vapor phase. Total ion chromatograms of a standard protein mix run through Dionex monolithic column with (A) no reagent delivery and (B) vapors from 300 mM sulfolane mixed in 100% ACN. The peaks labeled 1 through 4 represent the following proteins: RNase A (1), ubiquitin (2), lysozyme (3), and myoglobin (4).
Figure 5.8. Mass spectra of RNase A with sulfolane delivery in the vapor phase. Mass spectra of RNase A obtained with (A) no reagent and (B) 300 mM sulfolane delivered in vapor phase. In addition to charge enhancement, there is a 3 fold increase in the signal intensity during sulfolane addition.
Figure 5.9. Mass spectra of ubiquitin with sulfolane delivery in the vapor phase. Mass spectra of ubiquitin obtained with (A) no reagent and (B) 300 mM sulfolane delivered in vapor phase. In addition to charge enhancement, the signal intensity doubles during sulfolane addition.
Figure 5.10. Mass spectra of myoglobin with sulfolane delivery in the vapor phase. Mass spectra of myoglobin obtained with (A) no reagent and (B) 300 mM sulfolane delivered in vapor phase. There is a slight increase in the average charge from 17.6 to 18.2 upon sulfolane addition.
CHAPTER 6

Conclusions and Future Directions

6.1. Overview

The top-down method developments I have presented in this thesis are efficient methods for high throughput protein identification in the chromatographic timescale. In the current methods used for top-down protein identification, instruments with dual mass analyzers are used to measure the intact and fragment ions separately. By taking advantage of the resolution and mass accuracy available in the TOF mass analyzer, we have combined it with in-source dissociation for proteins eluting on-line from a LC column and demonstrated that it is possible to utilize a stand-alone TOF instrument for the purpose of protein identification.

Additionally, in the current top-down methods, the instrument time spent in isolating and fragmenting intact protein charge states decreases the duty cycle of the instrument. To overcome this disadvantage, we have developed a new data-independent protein identification method known as CASI-CAD. In this technique, multiple intact protein charge states are selected and dissociated simultaneously inside the collision cell. This greatly reduces the time it takes to acquire the fragment ion data from the multiple charge states selected.

To increase the efficiency of intact protein measurement and also fragmentation of protein precursor charge states in the chromatographic timescale, we have developed techniques to introduce the reagent in the liquid and vapor phase in the electrospray plume. In addition to enhancing charge, these reagents are promising in that it can increase the signal intensity for proteins eluting from a column.
6.2. Top-Down protein identification using a single-stage LC-ESI-TOF MS

In Chapter 2, I describe a high throughput platform for protein identification using a stand-alone TOF instrument and the human 20S proteasome as a model complex. The protein subunits in the proteasome complex elute on-line from a LC column and undergo fragmentation in the nozzle-skimmer (NS) region. Both the intact and fragment ion masses are measured using the TOF mass analyzer.

Many research groups have infused targeted proteins and performed NS dissociation in instruments with single mass analyzer and employed it as a pseudo tandem MS method. However, when combined with an up-front protein separation technique, it is an in-expensive solution to identify proteins in complex samples. The highlight of the method is that the NS voltage can be switched from low to high every second enabling both intact and fragment ion mass measurements in the same run. In addition, the resolving power of 10,000 for an \( m/z \) of 1000 available in this instrument is sufficient to precisely assign the multiply charged fragment ions up to 6+ during data interpretation. Furthermore, the advantage of using the TOF instrument is that the scan speed is high enough to preserve the chromatographic resolution.

The method was successful in that intact mass of all the subunits were obtained and short stretches of amino acid sequence information were obtained for ten subunits. Four proteins co-eluted with other proteins and the sequence information was not obtained for these co-eluting proteins. In addition, by comparing the intact MWs obtained to the theoretical MW in the database, the presence of a phosphorylation in one of the subunits was identified. Finally, the N-terminal modifications in all the subunits were identified by utilizing the intact mass obtained and fragment ions originating from the N-terminus of the subunits.

While this method is highly useful for identification of proteins in samples with moderate complexity, the upper limit on the number of proteins identified using this method is dependent on the peak capacity of the LC column. Co-eluting proteins often present a significant problem.
in reversed phase LC and when a data-independent technique such as NS dissociation is chosen for fragmentation, fragments from the co-eluting proteins are present together in the mass spectrum, posing additional problems during data interpretation. For these situations where peak capacity is decreased due to co-elution, additional up-front fractionation techniques are required. Secondly, with the resolving power available in this instrument, fragment ions up to 6+ charges can be unambiguously assigned. This translates to identifying the sequence information pertaining to N- and C-terminal regions of the protein. While this is sufficient for identifying non-homologous proteins, for proteins which have a relatively high degree of sequence similarity in one or both terminals, higher resolution instruments such as FT-ICR MS or orbitrap are necessary. With the help of higher resolution, the higher m/z fragments originating from cleavages in the middle of the protein can be identified and this helps with identifying homologous proteins.

6.3. Top-Down protein identification using data-independent fragmentation in FT-ICR MS

In Chapter 3, I describe a new data-independent fragmentation method for selecting multiple charge states of the protein and fragmenting all simultaneously in the collision cell. A majority of current high throughput top-down protein identification methods employ CAD for proteins less than 20 kDa and NSD for proteins with higher MW. While NSD is a data-independent dissociation method, CAD is performed by isolating the precursor charge states and fragmenting each separately in the collision cell, followed by merging the fragment ions during post-processing. There are several disadvantages while employing data-dependent CAD. First, the duty cycle of the instrument is lowered, particularly in high resolution instruments such as orbitraps and LTQ-FTs, when fragment ion data is set to be acquired at a high resolution. As it takes more time to acquire the fragment ion data, the subsequent proteins
eluting from the column are not adequately sampled and this lowers chromatographic resolution.

Next, during data-dependent CAD, usually the top three most abundant charge states of the protein are selected for fragmentation. Many reports have shown that different precursor charge states undergo cleavage through different channels and it is important to fragment multiple charge states to get comprehensive sequence coverage. In the new data-independent fragmentation method, ‘CASI-CAD’, multiple charge states were isolated and fragmented by selecting a broad mass window of 800-1200 m/z in the quadrupole. This eliminates the need to isolate the precursors based on data during acquisition and also yields fragmentation from multiple charge states simultaneously. In addition, the time it takes to acquire a scan is 0.63 second which is two-fold less than the current data-dependent cycle time in the orbitrap set at the same resolution.

When comparing CASI-CAD with the widely adopted NSD, the data from the proteasome complex shows that for proteins which co-elute, CASI-CAD yields more fragment ions than NSD. Typically, when two proteins co-elute from the column, the ionization efficiency of one of the two proteins is less and similarly, during fragmentation, one of the two proteins in the pair yields only few fragment ions. When CASI-CAD is used during this situation, the mass filter set in the quadrupole selects only the ions within the 800-1200 m/z for measurement as opposed to a broadband spectrum. The fragment ions that fall within this mass range are measured more efficiently than the intact protein ions. Consequently, for the co-eluting proteins, CASI-CAD method picked up additional fragment ions.

Furthermore, as an additional data-independent dissociation technique, the fragment ions generated after NSD were enriched by using the CASI technique. By filtering the fragment ions in the 800-1200 m/z range after FS-CAD, more sequence coverage was obtained for 8 subunits in the complex. By combining the fragmentation efficiency of FS-CAD technique with
the fragment ion selectivity of the CASI principle, more sequence coverage was obtained than using either technique alone. This new method is an efficient addition to the top-down data-independent acquisitions, and can be employed in other instrument types such as LTQ-FT-ICR MS by setting a mass range in the LTQ before fragment ion are measured in the FT mass analyzer.

In both CASI-CAD and FS-CAD followed by CASI enrichment, ion accumulation time in the quadrupole was increased from 0.2 second to 0.4 and 0.6 second to find if sequence coverage can be increased by accumulating the fragment ions for a longer time. For small molecule analysis, there are studies where the signal from an isolated small molecule ion is accumulated by CASI to increase the signal intensity before fragmenting by CAD. While this is advantageous for small molecules where only one or a few related ions are isolated, this strategy was not helpful in protein analysis where multiple charge states with the 800-1200 m/z window are selected. By increasing the accumulation time to 0.4 second, the sequence coverage was the same as that of 0.2 second. When 0.6 second accumulation was used, the number of fragment ions decreased and this can be attributed to more number of charges being present in the collision cell for a longer duration and leading to the space charge effect. A decrease in the chromatographic resolution (increase in the peak width and decrease in the total ion current) was also observed at 0.6 second accumulation.

When this method is compared to the NSD in the TOF instrument, there are obviously several advantages. First, the mass accuracy of the intact proteins significantly improved and all the subunits were measured with 1 ppm deviation using the ultra-high resolution of the FT-ICR instrument. This is a significant advantage when complex cell lysates are analyzed since unambiguous identifications during database searches are possible by improved mass accuracy. In addition when high MW proteins are analyzed, increased mass accuracy plays a significant role in differentiating small differences such as amidation/deamidation (mass
difference of 1 Da) and disulfide-intact/reduced states (1 Da difference for every disulfide bond reduced). These small changes are very common in therapeutic proteins such as antibodies and currently, these changes are measured by peptide mapping where the antibody is digested with a proteolytic enzyme and the digested sample is run through a column followed by characterizing the peptides. While this peptide mapping strategy is necessary when characterizing the site of modifications initially in the antibody, for routine stability/efficacy testing, this is not required. By measuring the intact mass accurately, the steps involved in digestion and characterizing the peptides can be avoided.

In the TOF instrument only ions up to 6+ charge states were assigned and this limited the sequence coverage to the terminal regions of the protein. On the other hand, the resolution available in the FT-ICR instrument allowed assigning high MW fragment ions which originated from cleavages in the middle of the protein. This is advantageous in situations where a protein could be potentially modified at one of the several possible sites by the same modification. For instance, if a protein’s intact MW observed and the theoretical MW in the database are different by 42 Da, it is evident that an acetylation is present, but given that there are several sites where this modification could reside (in the N-terminal methionine/alanine residues or any lysine residue), unambiguous characterization is not possible. By obtaining more sequence coverage pertaining to residues in the middle of the protein, it is possible that the site of modification could be identified. However, this also depends on the type of fragmentation used, since in CAD, labile PTMs are not preserved.

In the TOF and FT-ICR data, the presence of phosphorylated α3 subunit was identified by matching the intact MW obtained with the phosphorylation reported in the database for this subunit. This is an important advantage available in the top-down technique. Though the site of modification was not identified in the data from both the instrument platforms, determining the presence of the modification itself is a significant achievement and with further selective
enrichment for phosphoproteins, the site of modification can be identified. In addition, the nature of the N-terminus was identified by both the methods (TOF and FT-ICR) for all the subunits. This was possible by identifying the fragments originating from the N-terminus for proteins that eluted singly. For co-eluting proteins, since the efficiency of fragmentation is low, the N-terminal fragment ions were not present. However, when the accurate intact mass obtained was compared to the theoretical mass in the database, the N-terminal modification was identified in these proteins.

Moreover, the intensity of phosphorylation of the α3 subunit was found to be more than the non-phosphorylated counterpart. While quantifying intact proteins using MS is in the primitive stages of method development currently, it holds promise for the future. Also, by combining MS with CE, instead of RPLC, the modified and unmodified forms of the protein can be separated before MS analysis, which gives a greater chance to identify the modification by fragmenting it separately from the unmodified form.

Finally, in both TOF and FT-ICR data, internal fragment ions were present and these were not interpreted. The current data processing tools do not consider internal fragments when comparing the observed and theoretical fragments. Further improvement in the search tools to look for internal fragment ions after generating a list of potential protein hits will be very valuable to interpret more fragment ions and thereby increase the sequence coverage.

6.4. Supercharging reagents for enhancing charge during protein elution from HPLC

In Chapters 4 and 5, I present the results from supercharging targeted proteins by direct infusion and delivering the reagent for proteins eluting from a LC column. Supercharging is a recent addition to the field of top-down mass spectrometry and only a few groups have tested these reagents for a few standard proteins. In order to better understand the mechanism behind supercharging and to develop methods for utilizing these reagents efficiently in top-down
studies, these reagents have to be tested with various classes of proteins in different solvent conditions.

There are several conclusions from supercharging targeted proteins. First, higher degree of supercharging was observed for medium sized disulfide-intact proteins while disulfide-reduced counterparts did not further increase in charge. This leads to questioning if the number of basic sites and glutamine residues (from supercharging disulfide-reduced lysozyme, the maximum charge happens to correlate to the number of basic sites and glutamine residues present) determines the number of charges that can be attained. In addition, supercharging proteins that are already in the denatured state addresses the important controversy of the mechanism behind supercharging. While it could be possible that proteins in the native state can tend to become denatured while supercharging, charge enhancing the proteins that are already in the denatured state can be explained by examining other factors such as reagent basicity and direct interaction of the reagent with the solvent. Similarly, for higher MW disulfide bonded proteins such as BSA and transferrin, the supercharging effect might be limited by electrostatic repulsion resulting from the size/charge ratio.

Next, acidic disulfide bonded proteins such as β-lactoglobulin A and α-lactalbumin were supercharged efficiently; however, acidic proteins such as calmodulin and actin did not show any supercharging. This brings out the individual roles of acidic amino acids and the presence of disulfide bonds in supercharging. Despite the denaturing solvent conditions, the disulfide-intact architecture might keep the protein in a relatively constrained form and when supercharging reagents are added, enhanced charging is obtained. On the other hand, in acidic proteins without any disulfide bonds, these constraints are not present and so supercharging ability is only limited by the constituent amino acids in the protein. Similarly, in the glycosylated proteins, the net negative charge might impede the supercharging efficiency. For these acidic/glycosylated proteins, sulfolane might not be a suitable reagent. Other reagents such as
m-NBA which is a weak base and weak acid can be used in the negative mode to supercharge these proteins.

Only a few groups have studied the efficiency of supercharging reagents in the LC-MS mode. The results presented in Chapter 5 are exploratory in nature to find the suitable reagent and delivery method for charge enhancing during LC-MS. By studying the efficiency of different reagents, sulfolane was found to be a potent supercharging reagent under denaturing conditions and its efficiency for charge enhancing proteins when eluting from the column was analyzed. By adding it to the LC solvents, the charge enhancing effect of sulfolane was significant, while its reactivity with the active sites in the column limited its usage.

In order to find other efficient methods for reagent delivery, the air filter adapters in the CaptiveSpray source were employed. This is a simple method to deliver the reagent, with only the vapor pressure of the sulfolane limiting the concentration that can be achieved in the spray. Sulfolane has a very high boiling point (287°C at 760 Torr) and the vapor pressure exerted is consequently low. This might be the reason behind the moderate supercharging effect of sulfolane. Furthermore, to enhance the rate at which sulfolane can volatilize, 100% ACN was used as a delivery vehicle to increase the concentration of sulfolane in the spray. While this method also showed only moderate increase in supercharging, it increased the signal intensity of the ESI multiple charge states of several proteins and this could be used in situations where TFA is added as a LC modifier. The signal suppression caused by TFA can be negated by delivering sulfolane in the vapor phase.

6.5. Concluding remarks

The goal of the dissertation was to develop high throughput top-down methods to identify proteins in complex samples during elution from a LC column. The methods presented here are valuable additions to the top-down community and can be used for data-independent
fragmentation of proteins in the LC timescale. In addition, supercharging reagents can be used as a means to enhance top-down protein identification in the chromatographic timescale.
Metabolic Crisis After Traumatic Brain Injury is Associated with a Novel Microdialysis Proteome

R. Lakshmanan · J. A. Loo · T. Drake · J. Leblanc · A. J. Ytterberg · D. L. McArthur · M. Etchepare · P. M. Vespa

Abstract

Background To examine if the metabolic distress after traumatic brain injury (TBI) is associated with a unique proteome.

Methods Patients with severe TBI prospectively underwent cerebral microdialysis for the initial 96 h after injury. Hourly sampling of metabolism was performed and patients were categorized as having normal or abnormal metabolism as evidenced by the lactate/pyruvate ratio (LPR) threshold of 40. The microdialysate was frozen for proteomic batch processing retrospectively. We employed two different routes of proteomic techniques utilizing mass spectrometry (MS) and categorized as diagnostic and biomarker identification approaches. The diagnostic approach was aimed at finding a signature of MS peaks which can differentiate these two groups. We did this by enriching for intact peptides followed by MALDI-MS analysis. For the biomarker identification approach, we applied classical bottom-up (trypsin digestion followed by LC-MS/MS) proteomic methodologies.

Results Five patients were studied, 3 of whom had abnormal metabolism and 2 who had normal metabolism. By comparison, the abnormal group had higher LPR (1609 ± 3691 vs. 15.5 ± 6.8, P < 0.001), higher glutamate (157 ± 84 vs. 1.8 ± 1.4 μM, P < 0.001), and lower glucose (0.27 ± 0.35 vs. 1.8 ± 1.1 mmol/l, P < 0.001). The abnormal group demonstrated 13 unique proteins as compared with the normal group in the microdialysate. These proteins consisted of cytoarchitectural proteins, as well as blood breakdown proteins, and a few mitochondrial proteins. A unique as yet to be characterized peptide was found at m/z (mass/charge) 4733.5, which may represent a novel biomarker of metabolic distress.

Conclusion Metabolic distress after TBI is associated with a differential proteome that indicates cellular destruction during the acute phase of illness. This suggests that metabolic distress has immediate cellular consequences after TBI.

Keywords Traumatic brain injury · Microdialysis · Proteomics · Metabolic crisis · Biomarkers

Introduction

Traumatic brain injury (TBI) results in an immediate cellular injury often consisting of a hemorrhagic injury and/or early cell death in a defined region of the brain. Secondary cellular injury resulting in cell loss occurs in regions that were not directly involved in the primary injury. Brain regions that are in metabolic distress during the early post-injury phase are more prone to long-term tissue atrophy [1]. The mechanisms involved that lead to metabolic distress and eventual tissue loss are as yet unknown. Better characterization of the metabolic distress state is needed in order to understand its significance.
Cerebral microdialysis has recently gained increased popularity as a clinical investigation monitor of brain metabolism after TBI [2–4]. Preliminary studies have outlined several neurochemical markers of metabolic distress after injury, including elevated glutamate, elevated glycerol, and elevated lactate/pyruvate ratio as well as low glucose concentrations. These markers tell us about the metabolic state of the tissue, but tell us little about the mechanisms leading to this distress. Conversely, microdialysis also is capable of measuring other non-metabolic compounds that may reflect important mechanisms of how cells survive or die. Proteomics is one method of analyzing compounds that may reflect important mechanisms of how metabolic state has heretofore not been addressed.

The purpose of this study is to initially characterize the differential proteome after TBI based on the metabolic state of the tissue. The metabolic state of the tissue is defined by the lactate/pyruvate ratio (LPR). In normal metabolic states, the LPR is < 25 and in metabolic distress the LPR is > 40. We have carefully analyzed five subjects in this report, three of whom are in metabolic distress and two of whom are not. Our proteomic methods enable us to evaluate the presence of protein fragments in the extracellular fluid despite the relatively limited membrane size of the microdialysis probe (i.e., 20 kD). We report the novel finding of a differential proteome including a specific peptide under the various pathological conditions leading to metabolic distress.

Materials and Methods

Patient Population

This study was approved by our institutional review board and was conducted as part of the UCLA Brain Injury Research Center in patients with traumatic brain injury. The main inclusion criteria were (1) a Glasgow Coma Scale (GCS) score of ≤8 or a GCS of 9–14 with computerized tomographic (CT) scan to be present in normal appearing white matter (Fig. 1). The microdialysis catheter was tunneled 3 cm under the skin and secured to the scalp with sutures. It was then attached to the perfusion pump (CMA 106 microdialysis pump, CMA microdialysis AB, Solna, Sweden). Normal saline was perfused through the catheter at a rate of 1 µl/min. Fluid in the samples was collected every hour, analyzed immediately on the CMA 600 for glucose, lactate, pyruvate, and glycerol and frozen in dry ice or in a freezer. We collected microdialysis hourly samples starting from post-injury hour 20 through post-injury hour 140. Repeat analysis was conducted within 4 weeks with stability of the results demonstrated for all analytes ($R^2 = 0.92$). The concentration of four analytes (glucose, lactate, pyruvate, and glutamate) was measured in the automated analyzer (CMA 600 microdialysis analyzer, CMA microdialysis AB, Solna, Sweden). The concentration of each analyte was measured twice and the mean value used in the final analysis. Values were considered...
abnormal as follow: glutamate values above 5 μmol/l, lactate/pyruvate ratio (LPR) above 40, glucose values <0.20 mmol/l, and glycerol >50 μmol/l (16). Patients with LPR > 40 were designated as "abnormal" samples (Abn1, Abn2, Abn3) and patients with LPR < 40 were designated as "normal" samples (Nor1 and Nor2). For the proteomic studies outlined below, representative hourly samples of microdialysate from each patient were taken once every 4 h from the start of monitoring up to 84 h after injury or until the end of microdialysis monitoring.

Imaging and Volumetric Analysis

Acute computerized tomographic (CT) imaging was performed using 0.5-cm slice thickness, non-contrast axial images. Acute and magnetic resonance imaging (MRI) was performed using Volumetric T1 Magnetization Prepared Gradient Echo (MPRAGE: TR 1900 ms, TE 3.5 ms, FOV 256 × 256, 1-mm slice thickness), axial fluid inversion recovery (FLAIR: TR 9590 ms, TE 70 ms, FOV 512 × 384, slice thickness 3 mm), axial gradient recalled echo axial (GRE: TR 1500 ms, TE 7 ms, FOV 512 × 384, slice thickness 3 mm). Microdialysis catheter location was verified on CT or MRI images. In normal subjects, the microdialysis catheter was contained within normal appearing white matter on MRI.

Proteomic Methods

Peptide Profiling with Magnetic Beads and Matrix-Assisted Laser Desorption Ionization Time-Of-Flight (MALDI-TOF) MS

Peptides in the hourly microdialysate samples from each patient were enriched by binding to C18 magnetic beads. 5 μl of microdialysate sample was diluted with 65 μl of
binding solution which has 5% acetonitrile (ACN) and 0.1% formic acid (FA), and mixed with 10 μl of 25 mg/ml C18 Dynabeads® (Invitrogen, Carlsbad, CA, USA). The samples were incubated in a shaker for 20 min. The magnetic beads with the bound peptides were focused by placing a magnet externally to the sample tubes, and the supernatant which has the salts and other contaminants was removed. The beads were washed thrice with 80 μl of the binding solution. To elute the bound peptides, 10 μl of 50% ACN was added to the beads and mixed well. Approximately 5 μl of elute was mixed with 5 μl of MALDI matrix (7 mg/ml 2-cyano-4-hydroxycinnamic acid in 70% ACN and 0.1% trifluoroacetic acid), and 2 μl was spotted onto the MALDI target plate in duplicates and analyzed by MALDI-MS (PerkinElmer-Sciex prOTOF MALDI-TOF mass spectrometer, Shelton, CT, USA). The instrument was calibrated externally using a peptide mixture (PepMix1, LaserBio Labs, Sophia-Antipolis Cedex, France) which has five different peptides with mass ranging from m/z 1000 to 2500. The laser intensity was set at 50%, and the acquisition mass window ranged from m/z 1500 to 15000. Spectra were collected using an automated data acquisition tool in the software, which irradiated each sample well with 100 shots of laser at 32 different spots and the data from all these spots were added to generate the accumulated mass spectrum. Mass spectra obtained from both abnormal and normal samples were compared to find peaks that are different between these two sample groups. Progenesis PG600 software (Nonlinear Dynamics, Newcastle upon Tyne, UK) was used to recognize the differential peaks. PG600 aligns all the spectra from both groups and displays a ‘Simulated gel view’.

Pooling Microdialysate Samples Initially, from each patient, 60 μl of microdialysate sample was collected every hour and from this 20 μl was removed for small molecule analysis and profiling with C18 beads. For further identification of proteins and peptides in the microdialysate, all the hourly timepoints were pooled to have one representative sample for each patient.

Tandem MS on Endogenous Peptides To obtain sequence information of the peptides in the microdialysate sample, tandem MS (MS/MS) of the pooled microdialysate sample using a quadrupole time-of-flight (QTOF) mass spectrometer (Waters Synapt HDMS, Manchester, UK) was performed. Pooled samples were concentrated using our in-house made C18 microcolumns according to Gobom et al. [9]. In short, the microcolumns were fabricated with gel loading pipette tips (Eppendorf, Westbury, NY, USA) and Poros R2 (C18) stationary phase (Applied Biosystems, Foster City, CA, USA). After constricting the ends of the gel loading tip, C18 material was packed on these tips to achieve a 3–4-mm column length and equilibrated with 20-μl of 5% FA. 15 μl of each of abnormal subject #3 (Abn3) pool and Nor pool were diluted with 20 μl of 5% FA and loaded slowly on the column. The column was then washed with 20 μl of loading solvent and eluted directly into electrospray ionization (ESI) needles with 1 μl of 50% methanol and 1% FA. The collision energy for collisionally activated dissociation (CAD) MS/MS and the acquisition time were optimized for each targeted peptide to generate MS/MS spectra of sufficient quality. The spectra were processed using both MassLynx and MaxEnt3, and Mascot Distiller (Matrix Science, London, UK) to generate mass lists to search an in-house version of the Mascot search engine (Matrix Science). Additionally, the spectra were interpreted manually and the derived sequences searched against the NCBI database using BLAST.

Protein Digestion with Trypsin To identify all the proteins present in the microdialysate, a classical proteomic bottom-up approach was undertaken. Protein concentration in the pooled samples was determined using MicroBCA protein assay (Pierce, Rockford, IL, USA). The protein concentration ranged from 0.15 to 0.19 μg/μl in the abnormal samples and from 0.08 to 0.13 μg/μl in the normal samples. To 10 μg of protein from each patient, 25 μl of 100% dimethyl sulfoxide (final concentration 10%) was added to denature the proteins. It was then digested with 0.5 μg of trypsin (Promega, Madison, WI, USA) in 50-mM ammonium bicarbonate buffer at 37°C overnight. After digestion, the samples were dried down, resuspended in 10 μl of 5% FA, and desalted using Omix C18 tips (Varian, CA, USA).

Protein Identification by Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS) LC-MS/MS was performed with a QTOF mass spectrometer (QSTAR Pulsar XL, Applied Biosystems, Foster City, CA, USA) equipped with a nanoelectrospray interface (Protana, Odense, Denmark) and a nano-HPLC system (Dionex/LC Packings, Sunnyvale, CA). For each LC-MS/MS run, 2 μg of the trypsin digest was resuspended in 6 μl of 0.1% FA and injected into the nano-LC equipped with a homemade precolumn (150 μm × 5 mm) and an analytical column (75 μm × 150 mm) packed with Jupiter Proteo C12 resin (particle size 4 μm, Phenomenex, Torrance, CA). The precolumn was washed with the loading solvent (0.1% FA) for 4 min before the sample was injected onto the LC column. The eluants used for the LC were 0.1% FA (solvent A) and 95% ACN containing 0.1% FA (solvent B). The flow rate was 200 nl/min, and the following gradient was used: 3–21% B in 36 min, 21–35% B in 14 min, 35–80% B in 4 min, and maintained at 80% B for 10 min. The column was finally equilibrated with 3% B for 16 min prior
to the next run. All five patient samples were analyzed in triplicate. The tandem MS results from both the sample groups were searched against various databases (IPI human version 3.39, MSDB_06 and SwissProt) using MASCOT (Matrix Science) sequence search algorithm allowing for acetyl (N-term), formyl (N-term), oxidized methionine, carboxamidomethylated cysteine, pyro-Glu (cyclization of N-terminal glutamine and glutamate), methyl (C-term), methylated aspartate and glutamate. The mass tolerance was set to 0.3 Da for both precursor and fragment ions. Also, a semi-tryptic search against IPI human database was included to account for endogenous peptides that were cleaved by trypsin. The results from all the searches were merged and recalibrated using our in-house developed Perl scripts. The query lists generated from the search results were converted to peptide/protein lists and grouped according to proteins that share peptides. Finally, the following criteria were used for generating a list of proteins that is present in each sample replicate:

1. The protein should be identified in at least two replicates (out of 3 total) in each patient.
2. If it is a single peptide match in the protein identified, peptide score should be more than 35.
3. If there are two or more peptide matches, peptide score should be more than 20.

**Phosphopeptide Enrichment with TiO2 Microcolumns**

Pooled abnormal and normal samples were enriched for phosphopeptides using in-house made titanium dioxide (TiO2) microcolumns. These microcolumns were made on P10 pipette tips using plugs of 3M Empore C8 extraction disks (3M Bioanalytical Technologies, St. Paul, MN, USA) to close the end of the tips. A 3–4-mm column length of TiO2 stationary phase was made using 4-μm TiO2 (GL Sciences Inc., Torrance, CA, USA). The column was equilibrated with 30 μl of 80% ACN, 3% trifluoroacetic acid. The samples were diluted four times in the same solvent used for equilibration and loaded slowly on the column, followed by a wash with the same solvent. The peptides were eluted step-wise using 2 μl of dd H2O, 10 μl of 1% ammonium hydroxide, and 2 μl of 40% ACN, 0.1% FA, 0.5 μl of elute and 0.5 μl of 2,5-dihydroxybenzoic acid (DHB) matrix (10 mg/ml in 25% ACN, 15% isopropanol, 0.5% phosphoric acid) were spotted using dried droplet method on a MALDI target plate coated with lanolin. The samples were dephosphorylated to confirm the phosphorylation. The peptides were dephosphorylated at pH 7.8 according to Larsen et al. [10]. Briefly, 1 μl of the TiO2 purified phosphopeptides was added into a plug of 20 μl of 0.05-U alkaline phosphatase in 50-mM ammonium bicarbonate in a C18 microcolumn. The top of the column was covered with parafilm and column was incubated at 37°C for 1 h. 10 μl of 5% FA was added to stop the reaction by acidification and the peptides were loaded onto C18 material. After washing with 5% FA, the peptides were eluted using 1 μl of DHB matrix and spotted on the MALDI target plate for analysis.

**Phosphopeptide Identification by Tandem MS Using Synapt HDMS**

To obtain phosphopeptide sequences, they were further purified using R3 oligo resin (Applied Biosystems) and fragmented using a QTOF mass spectrometer (Waters Synapt HDMS). Briefly, 1 μl of phosphopeptides eluted from TiO2 microcolumns was diluted with 20 μl of 5% FA and loaded slowly onto R3 oligo stationary phase, washed with 20 μl of 5% FA and eluted with 1 μl of 50% methanol and 1% formic acid directly into ESI needles. Spectra were acquired manually on the Synapt HDMS as described above.

**Results**

**Clinical Characteristics**

Five TBI patients were studied with microdialysis. Three subjects had metabolic distress and two did not. The injury characteristics are outlined in Table 1. The three patients with metabolic distress had the microdialysis probe in injured tissue, and had elevated ICP, whereas the remainder had microdialysis probe in normal appearing tissue. The injured tissue consisted of a secondary ischemic stroke due to traumatic internal carotid artery dissection, pericontusional tissue near an evacuated temporal contusion and hypoxic ischemic tissue due to hypotensive shock after TBI. Figure 1 outlines the probe location and tissue typing in the five patients.

**Metabolic Profile of Microdialysate**

Three patients had persistent metabolic crisis during the observation period (Abnormal 1–3). The microdialysate data were analyzed using the mixed effects model for group, time, and interaction for group × time. During this period, the abnormal group had higher LPR (1609 ± 3691 vs. 15.5 ± 6.8, \( P < 0.001 \)), higher glutamate (157 ± 84 vs. 1.8 ± 1.4 μM, \( P < 0.001 \)), and lower glucose (0.27 ± 0.35 vs. 1.8 ± 1.1 mmol/l, \( P < 0.001 \)). Glycerol values were similar in the two groups by the mixed model analysis despite higher mean values in the abnormal group (186 ± 120 vs. 21 ± 9, \( P < 0.11 \)). The temporal pattern
of microdialysis changes indicates that the maximum abnormality in the tissue was present within the initial 48 h after injury. Figure 2 shows the time course for the abnormal and normal patients. Table 2 outlines the mean microdialysis results for each group.

Initial Peptide Profiling Using MALDI-TOF MS

Using Progenesis PG600 to generate simulated gel views, the abnormal and normal patient profiles were reviewed. Each horizontal row in this gel view represents a mass spectrum obtained from the fluid collected at a certain time point (hours after injury) and profiled using magnetic beads coupled with MALDI-MS. The mass spectra of individual time points from all abnormal and normal samples are stacked separately to obtain the gel view shown in Fig. 3a, b, respectively. The mass spectra obtained from all three abnormal patients have a peak at \( m/z \) 4733.5. This peak is not present in both the normal LPR patients.

Identification of Proteins Present in Microdialysate Samples

A total of 35 proteins were identified by LC-MS/MS in Abn1, 29 proteins in Abn2, and 26 proteins in Abn3. Compared to this, the number of proteins identified in the normal samples was relatively small; eight proteins were identified from Nor1 and 9 proteins from Nor2. Venn diagrams that illustrate the number of proteins uniquely present and those that are common between patients in both the sample groups were constructed (Fig. 4). Proteins present in at least two out of three patients in abnormal samples are listed in Table 3 and for normal samples in Table 4. All three abnormal samples have 13 proteins in common and there are 9 other proteins that are present in at least two out of three abnormal samples (Table 3). There are six proteins common to both the normal samples (Table 4), and three of these were present in all three abnormals as well. The proteins unique to each patient in both the sample groups were not included in the analysis because these proteins may represent individual variations specific to each patient. As one can see, there are a greater proportion of structural proteins seen in the metabolic distress group.

Endogenous Peptides Identified by Tandem MS

The fragmentation spectra obtained from tandem MS experiments were processed using designated software and interpreted both manually and using the MASCOT search algorithm. Figure 5 shows the full-scan deconvoluted mass spectrum of pooled Abn3 sample eluted from C18 micro-column, with the labeled peaks further identified by tandem MS experiments. The ions at \( m/z \) 997.49, 1195.62, and 1308.71 matched to a part of hemoglobin beta chain protein sequence with N-terminal variations as shown in Fig. 5 (inset). The \( m/z \) 1798.93 ion matched to a different stretch of amino acids within the same protein. The peptide at \( m/z \) 3039.54 matched to 30 amino acids from the N-terminus of the hemoglobin alpha chain and the \( m/z \) 2022.03 ion matched to 20 amino acids close to the C-terminus of the same polypeptide. Another cluster of four peptides (\( m/z \) 3326.67, 3342.66, 3358.64, and 3374.66) differing from one another by 16 Da also matched to first 32 residues from the N-terminus of hemoglobin alpha chain. The peptides at \( m/z \) 3342.66, 3358.64, and 3374.66 are singly, doubly, and triply oxidized forms of the \( m/z \) 3326.67 ion, respectively.

In contrast, the full-scan mass spectra of normal sample (Fig. 6) eluted from C18 microcolumn does not have any of the peptides identified from the abnormal sample.

Phosphopeptide Enrichment, Confirmation, and Identification

Two phosphopeptides at \( m/z \) 1545.6 and 1616.6 were both present in all the pooled abnormal (Abn1, Abn2, and Abn3) samples enriched using TiO2 microcolumns (Fig. 7). None of these two peaks were present in pooled Nor1 sample and there is a small peak at \( m/z \) 1616.6 in the Nor 2 pool (Fig. 7). The dephosphorylation of these two phosphopeptides using alkaline phosphatase showed a 80 Da decrease in mass for both peaks, depicting the loss of the phosphate group (data not shown). The MS/MS spectra of the peptide ion at \( m/z \) 1616.6 matched to the published sequence of fibrinopeptide A and the peptide at \( m/z \) 1545.6

<table>
<thead>
<tr>
<th>Subject</th>
<th>Injury</th>
<th>MD location</th>
<th>MD tissue character</th>
<th>Sampling time</th>
<th>ICP elevation</th>
<th>GOSe 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abn1</td>
<td>C-TBI</td>
<td>R Temporal</td>
<td>Pericontusional</td>
<td>PIH 43–106</td>
<td>Y</td>
<td>4</td>
</tr>
<tr>
<td>Abn2</td>
<td>C-TBI</td>
<td>R Frontal</td>
<td>Ischemia</td>
<td>PIH 29–80</td>
<td>Y</td>
<td>1</td>
</tr>
<tr>
<td>Abn3</td>
<td>GSW</td>
<td>R Frontal</td>
<td>Ischemia</td>
<td>PIH 19–67</td>
<td>Y</td>
<td>1</td>
</tr>
<tr>
<td>Norm 1</td>
<td>C-TBI</td>
<td>R Frontal</td>
<td>NAWM</td>
<td>PIH 60–108</td>
<td>N</td>
<td>8</td>
</tr>
<tr>
<td>Norm 2</td>
<td>C-TBI</td>
<td>R Frontal</td>
<td>NAWM</td>
<td>PIH 23–111</td>
<td>N</td>
<td>8</td>
</tr>
</tbody>
</table>

Table 1 Injury characteristics for each patient. GOSe 12 indicates the Glasgow outcome score extended at 12 months after injury.
Fig. 2 The time plot of hourly microdialysis lactate/pyruvate ratio for all five patients. The abnormal patients are solid and normal patients are open circles. Each patient is represented by a continuous stream of data points. The LPR values have been log transformed to be able to plot the time series.
also matched to fibrinopeptide A missing the N-terminal alanine residue (Fig. 8).

**Discussion**

The principal findings of this study are (1) the extracellular fluid of TBI patients has a representative proteome that can be detected through microdialysis, (2) the proteome appears to be differentially affected with greater numbers of cytoskeletal proteins in those patients exhibiting extreme metabolic distress, and (3) there appears to be a novel peptide that as yet has not been well characterized. These results indicate that metabolic distress is accompanied by various levels of early proteolysis and that this proteolysis

<table>
<thead>
<tr>
<th>Glucose (mmol/l)</th>
<th>Lactate (mmol/l)</th>
<th>Glutamate (µM)</th>
<th>Pyruvate (µM)</th>
<th>L/P ratio</th>
<th>Glycerol (µM)</th>
<th>Urea (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abnl 0.27 ± 0.30</td>
<td>3.25 ± 2.1</td>
<td>161 ± 73</td>
<td>23 ± 29</td>
<td>1530 ± 2141</td>
<td>186 ± 89</td>
<td>1.34 ± 1.2</td>
</tr>
<tr>
<td>Norm 2.3 ± .74</td>
<td>1.1 ± 0.5</td>
<td>1.4 ± 0.05</td>
<td>89 ± 9</td>
<td>13.5 ± 6.4</td>
<td>15 ± 7</td>
<td>1.37 ± 0.56</td>
</tr>
</tbody>
</table>

The values were all significantly different except glycerol and urea.
may be one mechanism through which long-term damage is done by metabolic distress.

Previous Proteomic Studies in TBI

Previously, a few groups have studied the proteome found in human serum [6, 7] or cerebrospinal fluid (CSF) [8, 11, 12] after TBI. These studies have focused on a general description of the proteome or selective study of unique proteins as predictive biomarkers. A variety of cytoskeletal and neuronal proteins such as S100B, neuron-specific enolase (NSE), myelin-basic protein (MBP), and glial fibrillary acidic protein (GFAP) have been detected in these studies [13–15]. Interpretation of proteomic studies on serum is challenging since the proteins or proteolytic fragments from the damaged brain tissue are reabsorbed via the arachnoid villie into the blood [7] and this may lead to additional dilution effects.

Proteomic studies on cerebral microdialysis have been much more limited. The first description of proteomic study of microdialysis was by Maurer et al. [16], and was done in ischemic stroke patients using two dimensional gel electrophoresis followed by mass spectrometry and they found 10 proteins unique to the microdialysate sample. Surprisingly, some proteins identified in Maurer’s study were larger than 20 kD which is the stipulated cut-off of the microdialysate semi-permeable membrane. It could be that the folded three-dimensional structure of some of these

Table 3  List of unique proteins present in at least two out of three abnormal samples

<table>
<thead>
<tr>
<th>Acc. No.</th>
<th>Description</th>
<th>Abn1</th>
<th>Abn2</th>
<th>Abn3</th>
<th>No. of AA</th>
<th>MW</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPI00410714</td>
<td>Hemoglobin subunit alpha</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>142</td>
<td>15258</td>
</tr>
<tr>
<td>IPI00654755</td>
<td>Hemoglobin subunit beta</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>147</td>
<td>15998</td>
</tr>
<tr>
<td>IPI00029717</td>
<td>Isoform 2 of fibrinogen alpha chain</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>644</td>
<td>69757</td>
</tr>
<tr>
<td>IPI00021885</td>
<td>Isoform 1 of fibrinogen alpha chain</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>866</td>
<td>94973</td>
</tr>
<tr>
<td>IPI00021907</td>
<td>Isoform 1 of myelin basic protein</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>304</td>
<td>33117</td>
</tr>
<tr>
<td>IPI00220327</td>
<td>Keratin, type II cytoskeletal 1b</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>644</td>
<td>66039</td>
</tr>
<tr>
<td>IPI00021440</td>
<td>Actin, cytoplasmal 2</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>375</td>
<td>41793</td>
</tr>
<tr>
<td>IPI00798155</td>
<td>Ubiquitin</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>106</td>
<td>12237</td>
</tr>
<tr>
<td>FETUA_HUMAN</td>
<td>Alpha-2-HS-glycoprotein</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>367</td>
<td>39325</td>
</tr>
<tr>
<td>IPI00643115</td>
<td>Stathmin</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>116</td>
<td>13492</td>
</tr>
<tr>
<td>IPI00220828</td>
<td>Thymosin beta-4</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>44</td>
<td>5053</td>
</tr>
<tr>
<td>IPI00220827</td>
<td>Thymosin beta-10</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>44</td>
<td>5026</td>
</tr>
<tr>
<td>IPI0012011</td>
<td>Cofilin-1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>166</td>
<td>18502</td>
</tr>
<tr>
<td>IPI00025363</td>
<td>Isoform 1 of glial fibrillary acidic protein</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>432</td>
<td>49880</td>
</tr>
<tr>
<td>IPI00217465</td>
<td>Histone H1.2</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>213</td>
<td>21365</td>
</tr>
<tr>
<td>IPI00298497</td>
<td>Fibrinogen beta chain</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>491</td>
<td>55928</td>
</tr>
<tr>
<td>IPI0021841</td>
<td>Apolipoprotein A-I</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>267</td>
<td>30778</td>
</tr>
<tr>
<td>IPI00793930</td>
<td>Tubulin alpha</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>335</td>
<td>37218</td>
</tr>
<tr>
<td>IPI00745872</td>
<td>Albumin</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>609</td>
<td>69367</td>
</tr>
<tr>
<td>IPI00883934</td>
<td>Prothymosin alpha</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>73</td>
<td>8161</td>
</tr>
<tr>
<td>IPI00008868</td>
<td>Microtubule-associated protein 1B</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>2468</td>
<td>270620</td>
</tr>
<tr>
<td>IPI00216461</td>
<td>Acylphosphatase-2</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>127</td>
<td>13891</td>
</tr>
</tbody>
</table>

Listed is the number of amino acids (AA) and the molecular weight (MW) of the native, undigested protein. We identified only fragments of the native protein using the methods outlined above

Table 4  List of proteins present in both normal samples

<table>
<thead>
<tr>
<th>Acc. No.</th>
<th>Description</th>
<th>No. of AA</th>
<th>MW</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPI00220327</td>
<td>Keratin, type II cytoskeletal 1b</td>
<td>644</td>
<td>66039</td>
</tr>
<tr>
<td>IPI00029717</td>
<td>Isoform 2 of fibrinogen alpha chain</td>
<td>644</td>
<td>69757</td>
</tr>
<tr>
<td>IPI00021885</td>
<td>Isoform 1 of fibrinogen alpha chain</td>
<td>866</td>
<td>94973</td>
</tr>
<tr>
<td>IPI00607600</td>
<td>Amyloid precursor-like protein 1</td>
<td>651</td>
<td>72247</td>
</tr>
<tr>
<td>IPI00429366</td>
<td>Programmed cell death 1 ligand 2</td>
<td>183</td>
<td>20867</td>
</tr>
<tr>
<td>IPI00470716</td>
<td>Isoform 2 of neuroendocrine protein 7B2</td>
<td>211</td>
<td>23659</td>
</tr>
</tbody>
</table>
proteins may allow for their retainment in the microdialysate sample. A more likely explanation is that fragments <20 kD of larger proteins are being identified using our methods. A more limited analysis of selective whole proteins in microdialysis, namely amyloid proteins have been done by Brody et al. [17] and Marklund et al. [18], using larger membrane (100 kD) catheters. In the Brody report, amyloid proteins appeared to be more prevalent under conditions of normal LPR and increasingly present as a function of time after injury. Marklund reported the presence of Tau and Aβ42 without specific comment about the state of metabolic distress. In neither study was a comprehensive approach performed to determine the range of potential proteins present nor the comparison of various proteins.
metabolic states. In our study, we sought to explore the range of proteins that occur under conditions of metabolic distress rather than focus on one pathway.

Our main findings are that there is a differential set of proteins present in tissue that is under metabolic distress as compared with tissue that is metabolically more...
normal. The types of proteins that were found are variable in class but several of these proteins raise the possibility that particular chemical processes are occurring under conditions of metabolic distress. First, we see a wider array of cytoarchitectural proteins (i.e., keratin, actin, tubulin alpha, cofilin) as compared with the normal metabolic state. Stathmin is a protocongene that is important in the disassembly of microtubules in the cytoskeleton that can become active once upregulation of phosphorylation kinases are activated [19]. Thymosin beta 4 is a regulatory protein of actin, thereby regulating intracellular levels of actin. Presumably, metabolic distress leads to upregulation of these kinases and in turn may activate stathmin, leading to microtubular disassembly. It is interesting to consider the presence of cytoskeletal proteins together with the increased level of extracellular glycerol in these samples. Glycerol is a proposed biomarker of cell membrane integrity and increases under conditions of membrane breakdown. It is possible that the observed proteins reflect loss of cytoarchitecture and subsequent necrosis and exudation of cellular contents into the extracellular space. Secondly, we see the presence of acylphosphatase, which is a 98 amino acid enzyme that has a known role in mediating glycolysis, mediates the endoplasmic reticulum’s function in protein folding and unfolding, and can mediate formation of amyloid under experimental conditions. The presence of acylphosphatase may indicate compensatory responses in glycolysis and/or alteration of the protein folding/unfolding/aggregation functions under conditions of metabolic distress. Activation of proteosome pathways may underlie that presence of multiple structural proteins in these samples. Hence, metabolic distress may trigger cytoskeletal disruption leading to the proteome that we have observed. This finding matches well with experimental reports of cytoskeletal disruption after TBI in which the calpain–calpastatin–caspase pathways are activated leading to cell death [20] and with proteomics performed on fresh whole brain pathological specimens after TBI [21].

In addition to cytoskeletal proteins, we found evidence of hematological peptides (hemoglobin subunit alpha and beta, fibrinogen alpha chain, and fibrinopeptide A) under conditions of metabolic distress. The presence of blood in the extracellular space could well explain this finding. However, this also raises the possibility that the presence of blood may have led to cellular toxicity through oxidative stress. Alternatively, microscopic bleeding could have resulted from the demise of tissue and be a secondary event rather than causative factor. Despite the presence of hematological peptides, the specific proteome in normal and abnormal subgroups were not representative of the blood proteome after TBI [21] hence we do not think this is related to blood brain barrier disruption alone.

Metabolic distress after TBI is characterized by a reduction in energy formation that may or may not occur in the context of ischemia. LPR has been correlated regionally with reduced oxidative metabolism [22]. The LPR is a good biomarker of metabolic distress. The duration of elevated LPR correlates with long-term regional atrophy [1]. Elevation of LPR for several days after TBI is associated with more severe atrophy of the frontal lobe, despite normal appearance of the frontal lobe on acute imaging. This suggests that there may be ongoing damage to normal appearing brain over time after the initial impact. The current proteomic data suggest that under extremely high LPR conditions, there is early loss of cellular integrity and proteolysis. This is a novel finding with regards to LPR and helps to underscore the importance of LPR as a biomarker in critically ill patients.

Limitations

This study is limited by several important features. First, the sample size is quite small but this was truly exploratory and a proof of concept rather than a definitive assessment of the proteome. Second, the small membrane pore size limits the sampling to fragments of larger proteins and/or small proteins. Use of a 100-kD probe may be better, but is also fraught with technical issues of requiring the perfusate fluid to contain macromolecules (dextran or albumin) which may in themselves affect proteomic sampling. Hence, the ideal methodology and catheter size for proteomics is unknown. Given the spectrum of sizes and the apparent proteolysis that occurred, rendering fragments of large proteins capable of being dialyzed via the small pore size, we are not certain that a bigger pore size is really necessary. Thirdly, we were not able to identify the novel protein at m/z 4733.5. This analysis requires more fluid than we had remaining. We anticipate needing an additional 600 µL of fluid to this. Hence, we have more work to do to evaluate if this is truly unique and what potential biological role it plays.

Conclusions

The evaluation of small molecular weight peptides is feasible using cerebral microdialysis. Tissue under metabolic distress displays a reproducible and unique proteome as compared with normal appearing TBI tissue. The proteome suggests ongoing cytoarchitectural disruption that may be mediated by proteosomal activity.
References


