TOP-DOWN PROTEOMICS USING SHEATH-FLOW CAPILLARY ELECTROPHORESIS
COUPLED TO MASS SPECTROMETRY

A Dissertation

Submitted to the Graduate School
of the University of Notre Dame
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

by

Yimeng Zhao

______________________________
Norman J. Dovichi, Director

Graduate Program in Chemistry and Biochemistry

Notre Dame, Indiana

January 2017
TOP-DOWN PROTEOMICS USING SHEATH-FLOW CHAPILLARY ELECTROPHORESIS

COUPLED TO MASS SPECTROMETRY

Abstract

by

Yimeng Zhao

While liquid chromatography dominates in the separation stage of proteomics studies, capillary electrophoresis is a valuable alternative and provides an orthogonal separation mode. It has proved its usefulness in bottom-up proteomics by generating complementary peptide identities. In this thesis, I present the applications of CZE-ESI-MS/MS in top-down proteomics. Capillary zone electrophoresis lends good resolving power to proteoforms with minor sequence variations or PTMs, which often have similar hydrophobicities but different charges. Here I apply this system to characterize secretome from M. marinum, demonstrate improve sequence coverage with combined fragmentation of HCD and Al-ETD, separate the heavy chains and light chains from reduced mAbs, and finally characterize proteins and proteoforms from yeast with prefractionation by RPLC. This work demonstrates the ability of CZE separating proteins and the potential for CZE-ESI-MS/MS as a platform for characterizing complex biological samples in top-down proteomics.
This thesis is dedicated to my beloved husband and my parents.
CONTENTS

Figures ................................................................................................................................. vi

Tables ................................................................................................................................. x

Acknowledgments ............................................................................................................... xi

Chapter 1: Introduction ....................................................................................................... 1
  1.1 Proteomics .................................................................................................................... 1
    1.1.1 Bottom-up proteomics ......................................................................................... 2
    1.1.2 Top-down proteomics ......................................................................................... 2
    1.1.3 Middle-down proteomics .................................................................................... 4
  1.2 Separation technologies .............................................................................................. 4
    1.2.1 High performance liquid chromatography (HPLC) ............................................. 5
  1.3 Reverse phase liquid chromatography (RPLC) ........................................................ 5
    1.3.1.1 Ion-exchange chromatography ....................................................................... 6
    1.3.1.2 Affinity chromatography ................................................................................. 7
  1.3.2 Capillary electrophoresis (CE) .............................................................................. 7
    1.3.2.1 Theory of capillary electrophoresis ................................................................. 8
    1.3.2.2 Capillary zone electrophoresis (CZE) ............................................................ 9
    1.3.2.3 Capillary isoelectric focusing (cIEF) ............................................................... 10
    1.3.2.4 Capillary sieving electrophoresis (CSE) ....................................................... 11
    1.3.2.5 Micellar electrokinetic capillary chromatography (MECC) ............................ 12
  1.4 Mass spectrometry (MS) ............................................................................................ 13
    1.4.1 Ionization ............................................................................................................ 13
      1.4.1.1 Electrospray ionization (ESI) ....................................................................... 14
      1.4.1.2 Matrix-assisted laser desorption/ionization (MALDI) ............................... 15
    1.4.2 Mass analyzer ..................................................................................................... 16
      1.4.2.1 Time-of-flight (TOF) .................................................................................... 16
      1.4.2.2 Quadrupole .................................................................................................. 17
      1.4.2.3 Ion trap .......................................................................................................... 18
      1.4.2.4 Fourier transform mass analyzer (FTMS) .................................................... 18
    1.4.3 Tandem mass spectrometry .................................................................................. 19
1.5 Capillary electrophoresis with mass spectrometry interface .................................. 20
  1.5.1 Sheath-flow interface ......................................................................................... 20
  1.5.2 Sheathless interface ........................................................................................... 21

Chapter 2: Top-down Characterization of the *Mycobacterium marinum* Secretome ....... 23
  2.1 Introduction ........................................................................................................... 23
  2.2 Experimental section ........................................................................................... 25
    2.2.1 Materials and reagents .................................................................................... 25
    2.2.2 Sample preparation ......................................................................................... 25
    2.2.3 CZE-ESI-MS/MS analysis .............................................................................. 25
    2.2.4 Mass spectrometer operating parameters ...................................................... 26
    2.2.5 Data analysis ..................................................................................................... 26
  2.3 Results and discussion .......................................................................................... 27
    2.3.1 Sample .............................................................................................................. 27
    2.3.2 Conductivity of acetic and formic acids ......................................................... 28
    2.3.3 Analysis of secretome from *Mycobacterium marinum* .................................. 30
  2.4 Conclusion ............................................................................................................ 32

Chapter 3: Coupling CZE with electron transfer dissociation and activated ion electron
transfer dissociation ............................................................................................... 39
  3.1 Introduction ........................................................................................................... 39
  3.2 Experimental section ........................................................................................... 42
    3.2.1 Materials and reagents .................................................................................... 42
    3.2.2 Sample preparation ......................................................................................... 42
    3.2.3 CZE-ESI-MS/MS analysis .............................................................................. 42
    3.2.4 Data analysis ..................................................................................................... 44
  3.3 Results and discussion .......................................................................................... 45
    3.3.1 Benchmarking the compatibility of ion-ion reactions with CZE-MS/MS .......... 45
    3.3.2 Complementary fragmentation methods for top-down CZE-MS/MS ............... 47
    3.3.3 Evaluating the combination of ETD and HCD for top-down proteomics on the *M. marinum* secretome ............................................................... 48
    3.3.4 Advantages of AI-ETD for top-down CZE-MS/MS on complex proteins mixtures ........................................................................................................... 50
  3.4 Conclusion ............................................................................................................ 52
  3.5 Acknowledgments .................................................................................................. 53

Chapter 4: Fast separation and analysis of reduced monoclonal antibodies ................. 60
FIGURES

Figure 2.1: Conductivity of aqueous solutions of acetic and formic acids at 25 °C. Conductivity was determined from the current generated when applying 6 kV voltage across a 60 cm long, 20 μm ID capillary. Both capillary ends were immersed in 0.1% FA during electrophoresis. To produce a stable reading, current was recorded 10 s after applying voltage. Uncertainties in data are ~5%. Data points are connected by straight lines. .......................................................... 33

Figure 2.2: Electrical resistance across a 40-cm long, 50 μm ID capillary filled with plugs of 70% acetic acid. The running buffer was 0.25% formic acid. Both capillary ends were immersed in 0.25% FA during electrophoresis after the acetic acid solution was injected. To produce a stable reading, current was measured 10 s after applying a 16 kV across the capillary. .......................................................... 34

Figure 2.3: Base peak electropherogram of the secreted proteins analyzed by the CZE−ESI-MS/MS system. Selected peaks were labeled with identified protein spectra. Superscript numbers indicate the protein rank in Table 1. The voltage applied was 15 kV for CE separation and 1.2 kV for electrospray. Inserts show parent ion spectra for proteins centered at the indicated m/z values. .............. 35

Figure 2.4: Spectrum of both unacetylated ESAT-6 and acetylated ESAT-6............. 36

Figure 2.5: HCD fragmentation of the 10-kDa culture filtrate antigen EsxB. (A) Fragmentation spectra of the [M + 7H]^{7+} charge state with HCD (normalized collision energy was 28%). (B) Sequence of this protein and the fragmentation patterns observed with HCD.............................................................. 37

Figure 3.1: Capillary zone electrophoresis on an ETD-enabled Orbitrap system. (A) Base peak electropherogram of four standard proteins separated by the CZE-ESI-MS/MS system, including three proteoforms of β-casein (labeled a-c). Colored MS1 spectra correspond to elution profiles of the designated proteins. (B) Comparison of number of tandem mass spectra acquired for each protein standard using either HCD or ETD, with average time per MS/MS scan provided. .............................................................. 54
Figure 3.2: Overlap of fragmentation sites from AI-ETD and HCD spectra for cytochrome c (A), carbonic anhydrase (B), β-casein (C), and myoglobin (D). ........................................... 55

Figure 3.3: Sequence maps show combined fragmentation patterns of AI-ETD and HCD (A), and ETD and HCD (B) for cytochrome c (top left), carbonic anhydrase (bottom left), β-casein (top right) and myoglobin (bottom right). *Number of fragmentation sites: total fragmentation sites/overlapped fragmentation sites. 56

Figure 3.4: Analysis of *M. marinum* secretome with ETD and HCD, Run 1. (A) The overlap in proteoforms identified using back-to-back ETD/HCD scans in a single CZE-MS/MS run, mapping to a total of 21 gene products. (B) Identification score (E-value) versus protein molecular weight for proteoforms identified with ETD and HCD. (C) Gain in percent sequence coverage when using a combination of fragments from an ETD scan and an HCD scan together over the sequence coverage achieved with HCD alone. Proteins on the y-axis are identified by their accession number. ................................................................. 57

Figure 3.5: Comparison of fragmentation quality using HCD, ETD, and AI-ETD for the characterization of 10 kDa culture filtrate antigen EsxB_1, a protein known to play a role in virulence in *M. marinum*. (A) The number of matching fragments generated by the three fragmentation methods at two different precursors of antigen EsxB_1. HCD results from both Run 1 (which included ETD scans) and Run 2 (which included AI-ETD scans) are included. (B) Overlap of fragmentation sites using ETD and HCD spectra from Run 1, and using AI-ETD and HCD spectra from Run 2. (C) Sequence map depicting the bonds cleaved by AI-ETD alone in a single MS/MS scan. (D) Sequence map depicting the bonds cleaved by a combination fragments from an ETD scan and HCD scan. *total bonds cleaved/shared cleavage sites ................................................................. 58

Figure 3.6: Analysis of *M. marinum* secretome with AI-ETD and HCD, Run 2. (A) The overlap in proteoforms identified using back-to-back AI-ETD/HCD scans in a single CZE-MS/MS run, mapping to a total of 19 gene products. (B) Identification score (E-value) versus protein molecular weight for proteoforms identified with AI-ETD and HCD. (C) Gain in percent sequence coverage when using a combination of fragments from an AI-ETD scan and an HCD scan together over the sequence coverage achieved with HCD alone. Proteins on the y-axis are identified by their accession number. (D) Sequence map depicting the number of bonds cleaved for Transcription Elongation Factor GreA (accession number: vii
Figure 4.1: Base peak electropherograms of a four-standard protein mixture (cytochrome c - 0.1 mg/mL, myoglobin - 0.1 mg/mL, carbonic anhydrase - 0.2 mg/mL and β-casein - 0.8 mg/mL in 0.04% FA, 20% ACN in water). Blue trace: before capillary treatment; this trace was multiplied by 100 before plotting. Orange trace: after capillary treatment. Data were treated with a Lowess filter with five-point span before they were plotted. LTQ-XL mass spectrometer was used for detection. 

Figure 4.2: Base and extracted ion electropherograms of reduced mAbs. (A) Duplicate base peak electropherogram of CZE-ESI-MS analysis of a reduced mAb (mouse anti-human IgG4 Fc-UNLB), including two isoforms of heavy chain and light chain. Mass spectra correspond to elution profiles of the designated polypeptide chains. (B) Extracted ion electropherograms of reduced mouse anti-human IgG4 Fc-UNLB dissolved in 0.01%OG and 50% ACN in water. Top, middle, bottom electropherograms correspond to 10 min, 1 hour and 1.5 hours after desalting of sample. LTQ-XL mass spectrometer was used for detection. 

Figure 4.3: Base peak electropherograms of CZE-ESI-MS analysis of reduced antibody A, B and C. Data treated with a Lowess filter with five-point span before plotting. 50 cm long, 50 µm I.D. capillary was used for separation. About 70 nL sample was injected for each analysis. An LTQ-XL mass spectrometer was used for detection. 

Figure 4.4: Base peak electropherograms for a two-antibody mixture (mouse anti-human IgG4 Fc-UNLB and human/mouse/rat activin A beta A subunit antibody), including migration times for each heavy chain and light chain. Major peaks were assigned in the 1st run. (A) Quintuplicate CZE-ESI-MS analysis of a two-antibody mixture. 50 cm long, 30 µm I.D. capillary was used for separation. About 18 nL liquid aliquot was injected for each analysis. Voltages: 21.3 kV/1.3 kV. LTQ-Orbitrap Velos mass spectrometer was used for detection with Orbitrap as detector. (B) Triplicate CZE-ESI-MS analysis of a two-antibody mixture. 50 cm long, 30 µm I.D. capillary was used for separation. About 30 nL liquid aliquot was injected for each analysis. Sheath liquid: 0.5% FA, 10% methanol. Voltages: 23.7 kV/1.2 kV. LTQ-XL mass spectrometer was used for detection. NL is the normalization level.
Figure 5.1: A. Extracted ion electropherogram for the seven-protein mixture. The signal for each protein was amplified to the same height as shown in the graph. The inserted graphs represent the averaged mass spectra for each protein. Top spectrum: intact mode with pressure at 0.2. Bottom spectrum: non-intact mode. B. The signal intensities for each protein with different instrument settings. Non-intact mode with trapping gas pressure at 1.0 and intact protein mode with trapping gas pressure at 0.1, 0.2, 0.3, and 0.5 were listed here. Duplicate runs were performed for all conditions, except non-intact mode where triplicate runs were performed.

Figure 5.2: Summary of the identification results from CZE-ESI-MS/MS top-down analysis of 23 fractions isolated from the yeast proteome. The number of mass species detected only includes the mass species greater than 5 kDa. The number of total identifications is the cumulative number of produced by combining the new fraction to all the previous ones.

Figure 5.3: Summary of identifications. (A) Molecular weight distributions of yeast proteoforms identifications and mass species detected (> 5 kDa) in MS1 combined from 23 CZE-ESI-MS/MS experiments. (B) Protein abundance distributions of identified protein groups. X-axis is plotted on a log_{10} scale.
TABLES

Table 2.1 Identified proteins in a single shot top-down CZE analysis of the *M. marinum* secretome ....................................................................................................................... 38

Table 5.1 The standard protein identifications with different modes and NCE values shown in the bracket. Non-intact mode was performed with NCE of 20%. 89

Table 5.2 Mass species detected from MS1 spectra (> 5 kDa), and fraction and total proteoforms identifications for each fraction corresponding to the values shown in Figure 5.2 ................................................................................................................................. 90
ACKNOWLEDGMENTS

I would like to acknowledge my advisor, Professor Norman J. Dovichi, who has been an incredible mentor for me. I thank him for supporting my research and for encouraging me to grow as a researcher.

I would like to thank my committee members, Professor Amanda Hummon, and Professor Zachary Schultz, for their brilliant comments, continuous support and guidance throughout my time in graduate school.

In addition, I would like to express my appreciation to all the Dovichi group colleagues, particularly Dr. Liangliang Sun and Dr. Guijie Zhu, for their help and inspiration throughout my graduate research.

Finally, I would like to thank the National Institutes of Health (Grant R01GM096767), and the Eli Lilly LIFT program for the financial support for my research work.
1.1 Proteomics

Proteomics\textsuperscript{1}, or mass spectrometry-based proteomics, is in general the study of cellular function at the protein level. Proteins are important because they directly control the biological functions of a living organism. While genomics provides the sequence information of genes\textsuperscript{2,3}, proteomics, as a postgenomic discipline, measures the qualitative and quantitative diversity of proteins as a result of expression, posttranslational modifications (PTMs), localization, interaction, as well as protein turnover. Understanding the protein identity and quantity greatly facilitates the subsequent research in cell biology and drug discovery. However, due to the complexity of proteome, full investigation of proteome involves various analytical techniques. Typically, a proteomic study\textsuperscript{4} begins with the protein extraction from cell or tissue lysate, followed by a fractionation step, usually a gel-based separation on protein level to simplify the protein mixture. Three approaches can be taken after that: bottom-up, top-down and middle-down approach. The resulting peptide or protein mixtures are subjected to one or more stages of chromatographic or electrophoretic separation with the last one coupled to a mass spectrometer. The mass spectrometric data are searched against a protein database generated from the genome sequence to confirm the
identity of protein. The quantitative information of the protein of interests can be obtained by implementing appropriate labeling techniques, such as SILAC and iTRAQ.

1.1.1 Bottom-up proteomics

Bottom-up approach characterizes proteins by analyzing peptides proteolytically digested from proteins. The term shotgun proteomics is referred to bottom-up analysis of a mixture of proteins because of its analogy to shotgun genomic sequencing. It is the most popular method when performing large-scale and in-depth proteomic research. After protein fractionation, the fractions are digested with an enzyme such as trypsin, and then subjected to, in most cases, LC-MS/MS analysis. Peptides are identified by comparing their tandem mass spectra derived from peptide fragmentation with in silico-generated theoretical tandem mass spectra. Protein identification is achieved by assigning peptide sequences to proteins from a protein database. The successful protein identification requires a probabilistic scoring schemes. Many software tools have been developed to address the computational challenges associated with the large datasets generated by bottom-up proteomics, extending its application to proteome profiling, quantification, modification and protein-protein interaction.

1.1.2 Top-down proteomics

In contrast to bottom-up proteomics, top-down proteomics directly characterizes intact proteins and their fragments. The major advantage of top-down approach includes sequence coverage, identification of PTMs, and less ambiguity
of the peptide-to-protein mapping, which makes it possible to identify protein isoforms within a specific protein group.\textsuperscript{20,21} Ideally, a top-down analysis can provide the full primary structure, all localized PTMs and any correlations between them. However, the current top-down approach has significant limitations due to challenging front-end separation and difficulties with protein ionization, detection and fragmentation, especially for large proteins.

Top-down proteomics requires sophisticated front-end separations and extremely high-resolution mass spectrometers. High-resolution Fourier transform ion cyclotron resonance (FTICR) mass spectrometry was first employed in top-down protein analysis by McLafferty’s group.\textsuperscript{17,22,23} That group later demonstrated the successful characterization of proteins with masses greater than 200 kDa.\textsuperscript{24} One of the most impressive demonstrations of top-down proteomics for complex sample was reported by Tran et al.,\textsuperscript{25} wherein 1,043 gene products and over 3,000 protein species were identified from a human cell lysate with a three-stage separation system; that analysis required roughly 45 hours of analysis time using a FTICR mass spectrometer. With a similar separation system and an Orbitrap mass spectrometer, 1,220 proteins including 347 human mitochondrial proteins were identified for human proteome and over 5,000 proteoforms were observed.\textsuperscript{26} In another study, Ansong and colleagues employed a four-hour UPLC separation of intact proteins from Salmonella typhimurium. This top-down analysis identified 563 unique proteins and 1,665 proteoforms.\textsuperscript{27} The information gained in top-down proteomics is still limited comparing to bottom-up approach even with much more sophisticated separation system and longer analysis time.
1.1.3 Middle-down proteomics

As a hybrid of bottom-up and top-down proteomics, middle-down approach analyzes large peptide fragments usually generated by restricted proteolysis of proteins.\textsuperscript{28} It takes advantage of both bottom-up approach, which is less analytically challenging than intact proteins, and top-down proteomics, which provides further insight into sequence variants and PTMs. Middle-down approaches have been mainly applied in characterization of monoclonal antibodies (mAbs),\textsuperscript{29-31} because it is more feasible to perform restricted proteolysis on mAbs, and comprehensive understanding of mAbs at different molecular levels is necessary for regulation of the biopharmaceutical industry.

1.2 Separation technologies

MS-based proteomics is highly dependent on separation technologies due to the complexity of biological samples. Proper separation methods should be introduced prior to mass analysis in order to provide unambiguous identifications and detection for low-abundance species that may be overwhelmed by a high abundant signal. Selection of appropriate separation methods based on the nature of the sample is important for the accuracy and sensitivity for a proteomic study. Conventional gel-based separation, such as two-dimensional polyacrylamide gel electrophoresis (2D PAGE), delivers high resolution and wide dynamic range\textsuperscript{32}, but generally suffers from laborious sample preparation and low recovery. In contrast, high performance liquid chromatography (HPLC) and capillary electrophoresis (CE) can be directly coupled to mass spectrometer.
Thus, they have better recovery and higher throughput, and are widely used in many biological applications.

1.2.1 High performance liquid chromatography (HPLC)

While gel-based methods excise, digest and off-line prepare the sample for MALDI instruments, HPLC separates peptides and proteins on-line and can be coupled with an ESI source. It has become a standard front-end separation in MS-based proteomics, which is called LC-MS/MS. HPLC is a separation based on column chromatography that pumps a sample mixture in a solvent called mobile phase with high pressure through a column filled with a solid adsorbent material called stationary phase. The different interaction for different components with stationary phase leads to the separation as they pass through the column. Reverse phase liquid chromatography (RPLC) is the most commonly used chromatographic material in LC-MS/MS setups. Other types of HPLC include size exclusion (SEC), ion exchange (IEX), hydrophilic interaction chromatography (HILIC) and affinity chromatography.

1.3 Reverse phase liquid chromatography (RPLC)

RPLC separates compounds based on their hydrophobicity. It can be either used as the single phase or the last dimension of a multidimensional separation in LC-MS/MS setups because the mobile phases used in RPLC are MS compatible. It has high resolution, efficiency and reproducibility, and longer and narrower capillary RP column with small particle size (< 2 μm) further improves the loading capacity, sensitivity and dynamic range of RPLC. More than 2000 proteins were identified in a single LC-
MS/MS analysis with dynamic range spanning six order of magnitude from human plasma using a long, small-particle-size (1.4 μm) RP column operated under ultrahigh pressure (20 kpsi). The implement of small particle size and elevated temperature in RPLC is also called ultra-performance liquid chromatography (UPLC). It is shown to significantly improve the resolution and sensitivity and reduce the analysis time.

1.3.1.1 Ion-exchange chromatography

Another approach to improve the separation performance is to introduce multidimensional separation that combines several orthogonal separation techniques. A biological sample can contain thousands of proteins and digestion to peptides makes it even more complex. Multidimensional separation is designed to address extremely high sample complexity. Ion-exchange chromatography separates ions and polar molecules based on their affinity to the ion exchanger. For example, in strong cation exchange chromatography (SCX), positively charged molecules are attracted to a negatively charged solid phase and eluted with a salt gradient. 2D SCX chromatography followed by RPLC, known as multidimensional protein identification technology (MudPIT), has become a popular approach in shotgun proteomics. Anion-exchange chromatography and a mixed-bed approach has also been employed as the first dimension in MudPIT setup.
1.3.1.2 Affinity chromatography

Affinity chromatography is a separation method based on specific binding and interactions such as that between antibody and antigen, enzyme and substrate or receptor and ligand. It is generally used as the first step in the separation or purification process to enrich a certain type of protein, such as antibodies, or proteins with specific PTMs, such as phosphorylation or glycosylation.

1.3.2 Capillary electrophoresis (CE)

Electrophoresis was developed as a separation technique in the 1930s by Arne Tiselius who received the Nobel Prize in Chemistry for it. Capillary electrophoresis is a format of electrophoretic separation that provides high efficiency separation for both large and small molecules inside a narrow (20-200 μm i.d.) capillary. The separation is driven by applying high voltage, which generates electroosmotic and electrophoretic flow of electrolyte solutions and ionic species, respectively. The high surface area-to-volume ratio of a narrow capillary allows rapid heat dissipation under high electric field and thus generates high efficiency separation. CE provides orthogonal separation to RPLC by separating proteins and peptides by their charge differences. As a result, it has drawn increasing interests in the proteomics research. There are several different type of CE, including capillary zone electrophoresis (CZE), capillary isoelectric focusing (cIEF), capillary sieving electrophoresis (CSE) and micellar electrokinetic chromatography (MEKC), where the first three are commonly used in protein and peptide separations.
1.3.2.1 Theory of capillary electrophoresis

Electroosmosis is one of the fundamental processes in CE, working when an electric field is applied and forms electroosmotic flow (EOF). In bare fused silica capillaries that are typically used in CE separations, the surface of the capillary inner wall is negatively charged under buffer conditions with pH greater than three.\textsuperscript{51-53} Cations are attracted to the silica-solution interface and form an electric double layer. When a voltage is applied across the capillary, those cations migrate towards the cathode and carry all molecules in the capillary with them due to their solvation within the background electrolyte. It is called electroosmotic flow, and is defined by

\[
\mu_{\text{EOF}} = \frac{\varepsilon \zeta}{4\pi \eta} E \tag{Eq 1.1}
\]

where \(\varepsilon\) is the dielectric constant, \(\eta\) is the viscosity of the buffer, \(\zeta\) is the zeta potential measured at the plane of shear close to the liquid-solid interface, and \(E\) is the electric field. The velocity of electroosmotic flow (EOF) is given by

\[
\nu_{\text{EOF}} = \mu_{\text{EOF}} E \tag{Eq 1.2}
\]

An important feature for EOF is its flat flow profile, whereas the flow profile is laminar (or parabolic) for a liquid under pressure. One of the reasons that CE generally has high separation efficiency is that EOF does not contribute to band broadening caused by laminar flow as in liquid chromatography.

The basis of the CE separation is electrophoretic separation. An ion in an electric field will have a migration velocity, \(\nu\), that is equal to the product of the field strength, \(E\), and the electrophoretic mobility \(\mu_e\).

\[
\nu = \mu_e E \tag{Eq 1.3}
\]
where $\mu_e$ is directly proportional to the charge ($q$) to size ($r$) ratio of the analyte and inversely proportional to the frictional forces resulted from its moving through the separation buffer.

$$\mu_e = \frac{q}{6\pi \eta r} \quad \text{(Eq 1.4)}$$

The velocity of an ion when taking into account the forces of both electrophoresis and electroosmotic flow is given by the sum of Eq 1.2 and 1.3

$$v = (\mu_e + \mu_{EOF})E \quad \text{(Eq 1.5)}$$

which resulted in separation of analytes while all the analytes moving towards the detector. The migration time ($t$) of the analyte through a capillary of length $L$ is given by

$$t = \frac{L}{E (\mu_e + \mu_{EOF})} \quad \text{(Eq 1.6)}$$

1.3.2.2 Capillary zone electrophoresis (CZE)

In capillary zone electrophoresis $^{54}$, analytes are separated in low viscosity free solutions and open tube capillaries. It is the simplest mode of CE. The solution is carried by EOF along with all the analytes. The migration order of the analytes simply follows the charge-to-size ratio. Generally, analytes with the highest ratio (positive) migrate the fastest with electrophoretic mobility at the same direction to EOF towards the cathode, while analytes with the lowest ratio (negative) migrate the slowest with electrophoretic mobility at the opposite direction to EOF. Since EOF is typically faster than electrophoretic velocity, all analytes will migrate toward the cathode where they are detected. However, if one wishes to separate neutral species, MECC should be used.
instead because there is no difference in charge-to-size ratio between neutral species, and they are not separated in CZE.

With the absence of gel and salt, CZE can be directly coupled to mass spectrometry as the front-end separation method for proteins and peptides. It has proved to be an orthogonal separation mode in bottom-up proteomics, which provides complementary information to RPLC.\textsuperscript{55-57} Expanded protein sequence coverage can be observed by identifying different peptides within the same protein. In addition, CZE provides fast separations by elimination of column regeneration in LC methods. In top-down proteomics, by offering a separation based on size-to-charge rather than hydrophobic interactions, CZE has the potential to resolve a different set of proteoforms than RPLC because proteins with minor sequence variations or PTMs often have similar hydrophobicities but different charges.\textsuperscript{58-61} CZE, as an open tubular separation, also facilitates the recovery of hydrophobic peptides and proteins comparing to RP columns.

1.3.2.3 Capillary isoelectric focusing (cIEF)

Separation in cIEF\textsuperscript{62} is based on the pI values of the analytes. A pH gradient is generated by ampholytes in a capillary filled with sample solution. The analyte will be positively or negatively charged depending on the difference between the pH where it is in and its pI value. The analyte migrates in the capillary until it is in the position where the ampholyte pH equals to its pI. In this position, the molecule has a zero net charge and its migration ceases. As a result, every analyte is focused at the pH position that equals to its pI. Any diffusion of analytes causing band broadening is immediately re-
focused because the diffused analyte will be charged and move back towards the zone center. cIEF provides powerful enrichment and high resolution for protein and peptide separation.\textsuperscript{62} Compared to CZE, a much larger sample injection volume can be used in cIEF, which facilitates the analysis of low concentration samples.\textsuperscript{63-65}

Efforts have been done to couple cIEF to mass spectrometer for proteomic study. It is challenging especially for bottom-up proteomics due to the presence of high concentration ampholytes that generate strong signals at the same m/z range as tryptic digests and interfere with peptide signals.\textsuperscript{66} Zhu \textit{et al.} utilized amino acids instead of ampholytes to generate pH gradient that greatly reduced the background signals\textsuperscript{67,68}, and applied it to bottom-up analysis of host cell proteins.\textsuperscript{69} cIEF coupled with FTICR mass spectrometry was applied to top-down analysis of the \textit{Escherichia coli} proteome by Smith’s group; that study generated parent ion mass information for 400-1,000 putative proteins in a single run.\textsuperscript{70}

1.3.2.4 Capillary sieving electrophoresis (CSE)

CSE, also known as capillary gel electrophoresis, is analogous to conventional slab-gel electrophoresis, but with faster separation speed and higher efficiency. Different from other types of CE, a noncon vective polymeric media is introduced to the capillary with the background electrolyte (BGE). CSE separates macromolecules by their size in a sieving matrix, or gel, that forms small pores and serves as a physical barrier to the migration of analytes through the capillary. It is particularly valuable for analysis of DNA fragments or oligonucleotides because they possess the same charge-to-size ratio
and are separated solely according to their sizes with exceptional resolution.\textsuperscript{71} When separating proteins\textsuperscript{72}, sodium dodecyl sulfate (SDS) is introduced and complexed with proteins as in SDS-PAGE.\textsuperscript{73} Both non-crosslinked and crosslinked polymers can be used as the separation media, such as linear polyacrylamide, polyacrylamide and hydroxypropyl methylcellulose. Although CSE is not compatible with MS for proteomics, it is widely used in mAb characterization with LIF or UV detection to determine its size heterogeneity.\textsuperscript{74}

1.3.2.5 Micellar electrokinetic capillary chromatography (MECC)

As discussed above, CE generally separates analytes by their charge-to-size ratio. MECC extends the application of CE to the separation of both neutral and charged molecules.\textsuperscript{75} Micelles are spherical aggregates of amphiphilic monomers called surfactants, which consist of a hydrophilic head and a hydrophobic tail. When dissolved in aqueous solvent with a concentration called the critical micelle concentration, the surfactant will form micelles with hydrocarbon tails facing the core and the polar head facing the solvent. The micelles serve as a pseudostationary phase that can incorporate nonpolar molecules in the hydrophobic core. Since most surfactants used in MECC are highly charged, the free neutral analytes can be separated from the micelle partitioned neutral analytes. The combination of hydrophobic, electrostatic and hydrogen-bonding interactions determines the partition equilibrium between the analytes and micelles, and thus results in the separation of neutral analytes.
1.4 Mass spectrometry (MS)

Mass spectrometry (MS) provides the most comprehensive characterization for biomolecules.\textsuperscript{76} By analyzing the masses of the proteins and their corresponding peptides, MS provides the ability to understand proteins from their primary sequences, PTMs, to higher order structures.\textsuperscript{77}

A mass spectrometer consists of the following parts: an ion source that generates ionized analytes in gas phase, optics that focus and transfer the ions, a mass analyzer that measures the mass-to-charge ratio ($m/z$) of the ionized analytes, and the data processing electronics. The ion source and mass analyzer are two critical parts for a mass spectrometer. The 2002 Nobel prize in chemistry was awarded to John Fenn and Koichi Tanaka for their contribution to the development of protein ionization methods\textsuperscript{78,79}, which were known as the soft ionization techniques and serve as the foundation of MS-based proteomics. Also, the development of mass analyzer with improved sensitivity, resolution, mass accuracy and tandem mass generation allows the characterization of complex biological samples. Together with other advances in experimental method designs and data analysis approaches over the past two decades, modern bench-top MS proteomics has enjoyed rapid growth and is recognized as the indispensable technology to understand the cellular functions.

1.4.1 Ionization

Proteins and peptides are thermally unstable species that are prone to degradation under harsh conditions used in traditional ionization methods, so soft
ionization methods are required to volatize and ionize these biomolecules. Electrospray ionization (ESI)\textsuperscript{79} and matrix-assisted laser desorption/ionization (MALDI)\textsuperscript{78,80-82} are two common techniques used to generate gas phase protein or peptide ions and allow them to be analyzed by MS. ESI ionized the analytes from solution phase, so it can be readily coupled to liquid-based separations, such as LC and CE. MALDI requires the sample to be ionized from a dry and crystalline matrix. The analytes are sublimated and ionized with the matrix via laser pulses. ESI-MS can be employed to both simple and complex samples, whereas MALDI-MS is usually used in analyzing simple mixtures and generate spatial information for interested analytes.

1.4.1.1 Electrospray ionization (ESI)

ESI is driven by applying high voltage between the emitter tip at the end of liquid phase separation and the mass spectrometer inlet. The ESI formation is initiated by the creation of Taylor cone\textsuperscript{83} at the end of emitter tip, followed by the formation of analyte-solvent droplets. The charged droplets then undergo desolvation aided by heated capillary and sheath gas. As the solvent evaporates, the surface charge of the droplets increases and results in Coulombic explosion and forms smaller droplets. The subsequent Coulomb fission leads to finer droplets and the analytes finally exist as gas phase charged ions. While the inlet of the mass spectrometer is at atmosphere pressure, the mass analyzer is consistently at a vacuum. The pressure difference between them directs the ions entering the mass spectrometer, passing the optics and finally being analyzed in the mass analyzed.
Several physical models were established to explain the ion formation in ESI.\textsuperscript{84-86} The common practical features including multiply charged species and that the ion formation is highly sensitive to analyte concentration and flow rate. Gale and Smith developed a new electrospray ionization source (emitter) that allows the effective utilization of very small sample volumes at low flow rate (200 nL/min) to produce high signal intensities and stable signals.\textsuperscript{87} The nano-ESI technique greatly improved the sensitivity of ESI by lowering the flow rate to several nL/min,\textsuperscript{36,88} and enable its implement to capillary RP columns.\textsuperscript{36}

1.4.1.2 Matrix-assisted laser desorption/ionization (MALDI)

Different from ESI, which generates ions from continuous liquid phase, MALDI samples are prepared with a matrix solution and applied to a metal plate. The solvent is then evaporated and the analyte molecules are incorporated in crystalline lattices. Upon irradiation with a laser pulse, the matrix absorbs energy and transfers it to the analyte while the matrix and analytes desorb from the plate to the mass spectrometer with the rapid laser heating. Acetified analytes forms predominately singly protonated ions, which makes MALDI suitable for top-down analysis of large proteins. The main drawbacks are that hundreds of laser shots are required to achieve acceptable signals\textsuperscript{89} and that the shot-to-shot reproducibility is low and highly dependent on the sample preparation and the matrix.\textsuperscript{90,91}
1.4.2 Mass analyzer

The mass analyzer is the central component of a mass spectrometer. It measures the m/z of a peptide or protein as well as determining structural information on primary sequence and PTMs by tandem mass spectra.\textsuperscript{77,92,93} Mass analyzers can be classified as two broad categories: the scanning and ion-beam type, such as ion time-of-flight (TOF) and quadrupole (Q); and the trapping type, such as ion trap (IT) and Orbitrap. The former ones are usually coupled with MALDI while the later ones are often coupled to ESI source. Four basic types of mass analyzers are frequently used in proteomics study and will be introduced here, including TOF, Q, IT, and Fourier transform (FT)-MS. They have very different in designs and properties, and are used for different experimental purposes. Their strength and weakness should be considered in terms of their scan range and speed, resolution, sensitivity and dynamic range. Depending on the purpose of the proteomics research, mass analyzers can be used alone or combined with one or more other mass analyzers known as hybrid mass spectrometer.

1.4.2.1 Time-of-flight (TOF)

TOF is conventionally coupled with a MALDI ion source to measure the intact mass of proteins and peptides. It determines the m/z value of an ion by the time needed to fly through a tube with specific length.\textsuperscript{94} Basically, ions are accelerated and separated under an electric field with different velocities, and the detector counts and amplified the arriving ions. In a reflector TOF, the ions turned around in a reflector that compensates for the slight variation in kinetic energy when exiting the ion source, which improves the mass resolution.
Two types of TOF instruments can be employed for tandem MS experiments. A collision cell is incorporated between two TOF mass analyzers or between a quadrupole mass filter and a TOF, known as TOF-TOF\textsuperscript{95} or Q-TOF\textsuperscript{96}, respectively. Ions selected in the first mass analyzer with a specific $m/z$ are fragmented in the collision cell followed by tandem mass spectra generation in the second (TOF) analyzer. With these configurations, mass spectra with high sensitivity, resolution and mass accuracy are provided. In addition, Q-TOF instrument can be coupled to an ESI ion source, which often results in more extensive fragment information than trapping mass analyzers.

1.4.2.2 Quadrupole

Quadrupole selects ions with a specific $m/z$ by varying the electric fields between four parallel, circular metal rods. It only permits a stable trajectory of the ion with the desired $m/z$, while other ions have unstable trajectories and collide with the rods. Quadrupoles are often included in the mass spectrometer as the mass selector, such as in Q-TOF and Q-linear ion trap (LIT/LTQ) instruments. One popular instrumental configuration is called triple quadrupole. It comprises three quadrupoles with the first Q (Q1) selecting ions of a particular $m/z$, the second cell (Q2) fragmenting the ion, and the third one (Q3) measuring the fragments. It is primary designed for multiple reaction monitoring/selected ion monitoring (MRM/SRM), which allows high sensitive quantitative analysis.\textsuperscript{97}
1.4.2.3 Ion trap

An ion trap captures ions for a certain time and then ejects ions based on their \( m/z \) values. Generally, an ion trap is highly used in proteomics due to its robustness, sensitivity and relatively low cost.\(^7,9^8\) The disadvantage of a three-dimensional ion trap analyzer is a comparatively low mass accuracy, which is a result of space-charging distortion caused by large number of ions accumulating at the point-like center. The linear ion trap (LTQ)\(^9^9,1^0^0\) is a linear or two-dimensional ion trap. Increased number of ions can be stored in the trap due to the larger volume with its cylindrical design, which improve the resolution, mass accuracy as well as scan speed.\(^1^0^1\) The LTQ is suitable for bottom-up proteomics for identifying proteins from complex samples such as whole cell lysate. Both the fast scanning rate and the high sensitivity facilitate improved peptide identification and protein sequence coverage. An LTQ can be either used as a stand-alone mass analyzer or as the front end of a hybrid mass spectrometer, such as LTQ-Orbitrap.

1.4.2.4 Fourier transform mass analyzer (FTMS)

Another trapping mass analyzer is called the Fourier transform mass analyzer, including FT-ICR and Orbitrap.\(^1^0^2\) FT-ICR captures ions with high magnet field and measures the frequencies that ions move at and subsequently uses Fourier transformation to convert the time-dome signal to \( m/z \) spectra.\(^6^1,1^0^3-1^0^5\) It is highly sensitive, accurate and provides wide \( m/z \) and dynamic range, but it is also very expensive. The Orbitrap, instead, traps ions in an electrostatic field with the ions \textit{orbiting} around a spindle shaped electrode at the central of it.\(^1^0^6-1^0^9\) Like FT-ICR,
Orbitrap features high resolution, mass accuracy, wide m/z range and dynamic range, while at a relatively lower cost and much smaller in size than an FT-ICR, and the Orbitrap has been used for a large amount of proteomics studies, including both bottom-up\textsuperscript{7,110-113} and top-down\textsuperscript{114,115} approaches. However, FT-ICR is still recognized as the most powerful instrument for top-down proteomics due to its broader m/z range.

1.4.3 Tandem mass spectrometry

MS-based proteomics often utilizes tandem mass spectrometry to confidently identify proteins including their amino acid sequences and PTMs. Generally, mass spectrometric data is acquired in a data-dependent manner, where a full MS scan is followed by several tandem MS scans determined by the results of the full scan.

A few mass spectrometer configurations are available for different applications. Triple quadrupole described above is one example that is used in quantitation for target protein or peptide with high sensitivity.\textsuperscript{116} LTQ-Orbitrap combines fast scan speed of the LTQ and the mass accuracy of the Orbitrap, acting as one of the most versatile mass spectrometers. It can be operated in a parallel manner with the Orbitrap acquiring full MS and the LTQ performing fragmentation. LTQ-Orbitrap has been widely used in large-scale proteomic analysis.\textsuperscript{117} Due to its high mass accuracy, MRM targeted quantification can also be carried out on it with high complexity samples.\textsuperscript{118} It allows the mass analysis of several ions at the same time with the strong resolving power of the Orbitrap.

In addition to LTQ, Orbitrap has also been coupled to quadrupole mass filter, which is marked by Thermo Scientific as Q Exactive.\textsuperscript{119,120} This combination provides
advanced sensitivity and resolution as well as fast scan speed, which enables the largest number of protein identification within a given analysis time.

1.5 Capillary electrophoresis with mass spectrometry interface

The direct coupling of capillary electrophoresis to mass spectrometry through ESI was introduced in 1987 by the Smith group. Over the past twenty years, various CE-ESI-MS interfaces have been developed to explore the potential of CE as a front-end separation method in proteomics and other fields. The challenge for interfacing CE and MS is a result of the nature of CE and ESI, which both require stable electric contact. The electric contact should be available at the capillary outlet with no interference with the CE separation.

CE-ESI-MS interfaces generally fall into two categories: sheathless interface and sheath-flow interface. The use of sheathless interface limits the composition of the BGE used in CE separation and the flow rate to maintain a stable electrospray. Sheath-flow interfaces have more flexibility with the choice of BGE but often suffer from low sensitivity due to dilution of analytes by sheath liquid.

1.5.1 Sheath-flow interface

The first commercialized interface was a coaxial sheath-flow interface that requires a pump to provide the sheath liquid. The sheath liquid mixes with the CE effluent at the exit of the capillary, establishing electrical contact and modifying the BGE to be more MS-compatible. The first sheath-flow interfaces employed a relatively high sheath flow rate, in the range of several microliters per minute, which is 100 times
higher than the nanoliters per minute flow rate of the BGE. The high sheath flow results in significant sample dilution and thus low sensitivity.

To overcome the problems with the existing interfaces, our group developed an electrokinetically pumped sheath-flow nanospray CZE-ESI-MS interface.\textsuperscript{125} The sheath liquid is driven by electroosmosis instead of pressure, which generates much lower flow rate in the nanoliter-per-minute range. The distal end of the separation capillary is inserted in a glass emitter. The analytes are transferred from the separation capillary to the glass emitter and mixed with the sheath liquid. This sheath-flow nanospray interface has been applied to a number of bottom-up proteomics analyses.\textsuperscript{55,56,126-133} Its sensitivity and stability was further improved with the second\textsuperscript{134} and third generation\textsuperscript{135} interfaces by reducing the distance between the capillary end and the emitter tip end and increasing the tip size. This sheath-flow interface has been commercialized recently, and has been applied to many proteomics and biopharmaceutical studies.

1.5.2 Sheathless interface

In sheathless interfaces, the traditional method for establishing electric contact is to coat the capillary end with conductive material or insert a wire electrode into the capillary. Both strategies resulted in unstable interfaces compared to the coaxial sheath-flow interface introduced above. However, as more researches focused on improving the detection limits of CE-ESI-MS systems, sheathless interface has attracted a lot of attention because there is no dilution at the interface.
Moini et al. reported a porous sheathless interface without direct contact of BGE with the metal electrode.\textsuperscript{136} The last several centimeters of the capillary are etched to form an \textasciitilde{}5 \textmu m thick porous wall. This conductive segment is inserted to a stainless steel needle filled with conductive liquid, which establishes the electric contact and forms ESI at the capillary tip. With this design, no dilution or bubble formation is involved with the CE effluent, which greatly improves the sensitivity and stability compared with first-generation sheath flow interfaces. The porous sheathless interface has also been commercialized and is popular especially in the biopharmaceutical industry.
2.1 Introduction

Top-down proteomics has emerged as an interesting alternative to bottom-up proteomics for the identification and localization of post-translational modifications (PTMs) and sequence variants. Recent advances in instrumentation and separation have extended the application of top-down proteomics from analysis of standard proteins or simple protein mixtures to large scale and in depth analysis of complex biological samples. Kelleher’s group has published a series of manuscripts describing the implement of GELFrEE separation and nano LC-MS/MS for large-scale human proteome characterization. 1,043 gene products and over 3,000 protein species were identified from a human cell lysate, while in another experiment, 347 human mitochondrial proteins were identified and over 5,000 proteoforms were observed. Ansong et al. employed a four-hour UPLC separation of intact proteins from Salmonella typhimurium and identified 563 unique proteins and 1,665 proteoforms.

1 A portion of this chapter was published as: Zhao Y, Sun L, Champion MM, Knierman MD, Dovichi NJ. Capillary zone electrophoresis-electrospray ionization-tandem mass spectrometry for top-down characterization of the Mycobacterium marinum secretome. Anal Chem. 2014; 86: 4873-8. doi: 10.1021/ac500092q.
Top-down proteomics primarily couples reverse phase liquid chromatography (RPLC) to a mass spectrometer. Capillary zone electrophoresis (CZE) is an orthogonal separation mode to RPLC, offering a separation based on size-to-charge rather than hydrophobic interactions.\textsuperscript{60,61,139} CZE can resolve a different set of proteoforms than RPLC because proteins with minor sequence variations or PTMs often have similar hydrophobicities but different charges. CZE-ESI-MS/MS has been employed to analyze single protein proteoforms\textsuperscript{58,59}, and more recently, complex biological samples.\textsuperscript{140}

Our group has developed an electrokinetically pumped sheath-flow nanospray CE–MS interface that employs electroosmosis to generate very low sheath flow rate.\textsuperscript{125} It was recently demonstrated that this sheath-flow interface could also be applied for top-down protein analysis.\textsuperscript{141} Model proteins and several impurities were separated and analyzed by that system in 12 min. After database searching of the tandem spectra, three proteins, their post-translational modifications, and one impurity were identified. Kelleher’s group has very recently reported the use of this interface and a Q Exactive mass spectrometer to analyze intact proteins from \textit{Pseudomonas aeruginosa}.\textsuperscript{140} Thirty proteins were identified in the mass range of 30-80 kDa during a 25-min CZE separation.

In this work, I coupled CZE to a high-resolution Q Exactive mass spectrometer via the electrokinetically pumped sheath-flow electrospray interface. The \textit{Mycobacterium marinum} secretome was separated and analyzed using this platform. I first evaluated the compatibility of high concentration (70%) acetic acid as sample preparation buffer with the CZE-MS/MS system using bovine heart cytochrome c as a model protein. I then applied this system to the analysis secretome from \textit{M. marinum}. This experiment
requires minimal sample preparation. I identified 22 gene products and 58 proteoforms in a single run from the wildtype secretome.

2.2 Experimental section

2.2.1 Materials and reagents

All reagents were purchased from Sigma-Aldrich (St. Louis, MO, USA), unless stated otherwise. Formic acid (FA) and glacial acetic acid were purchased from Fisher Scientific (Pittsburgh, PA). Methanol was purchased from Honeywell Burdick & Jackson (Wicklow, Ireland). Water was deionized by a NanoPure system from Thermo Scientific (Marietta, OH). Linear polyacrylamide (LPA)-coated fused capillary (50 μm i.d. × 150 μm o.d.) was purchased from Polymicro Technologies (Phoenix, AZ).

2.2.2 Sample preparation

The culturing of *M. marinum* and generation of short-term culture filtrates have been described elsewhere. A secreted protein fraction containing approximately 200 μg of protein, as determined by the bicinchoninic acid assay, was purified by ice-cold acetone precipitation and resuspension in 50 μL 70% acetic acid, followed by sonication for five minutes. The suspension was then centrifuged and the supernatant was taken for CZE-ESI-MS/MS analysis.

2.2.3 CZE-ESI-MS/MS analysis

CZE was coupled to a Q Exactive mass spectrometer for secretome characterization. Electrospray was generated using an electrokinetically pumped sheath flow through a nanospray emitter. The borosilicate glass emitter (1.0 mm o.d. × 0.75
mm i.d., 10-cm length) was pulled with a Sutter instrument P-1000 flaming/brown micropipet puller. The emitter inner diameter was 7–12 µm. Separation was performed in a 50 cm long, 50-µm ID, 150-µm OD LPA-coated fused capillary. The separation buffer was 0.25% (v/v) FA. The electrospray sheath liquid was 10% (v/v) methanol and 0.1% (v/v) FA. A ~500 ng protein aliquot (~6 cm in length) was injected into the separation capillary by pressure. The separation voltage was 15 kV and the electrospray voltage was 1.2 kV.

2.2.4 Mass spectrometer operating parameters

A Q Exactive mass spectrometer (Thermo Fisher Scientific) was operated with the S-lens rf level set at 50% and the ion transfer tube temperature at 280 °C. Full MS scans were acquired in the Orbitrap over the m/z 600–2000 range with resolution of 140,000 at m/z 200. The three most intense peaks with charge state ≥2 were selected for fragmentation in the higher energy collisional dissociation (HCD) cell and detection in the Orbitrap with resolution of 70,000 at m/z 200. The target value for MS and MS/MS acquisition were 3.00×10⁶ and 1.00×10⁶, respectively. One microscan was used. The maximum injection times for MS and MS/MS were both 500 ms. Dynamic exclusion was 60 s.

2.2.5 Data analysis

The tandem spectra were decharged and deisotoped by MS-Deconv (version 0.8.0.7370),¹⁴³ followed by database searching with MS-Align+ software (version 0.7.1.7143).¹⁴⁴ Raw files from Q Exactive were first converted to mzXML files with
ReAdW (version 4.3.1). Then, MS-Deconv (v 0.8.0.7370) was used to generate msalign files with mzXML files as the input. Finally, the MS-Align+ software (http://bix.ucsd.edu/projects/msalign/) was used for database searching with msalign files as the input. NCBI protein database for *M. marinum* including common contaminates (5,583 protein sequences) was used for database searching. The parameters for database searching included maximum number of modifications (shift number) as 2, mass error tolerance as 10 ppm, “doOneDaltonCorrection” and “doChargeCorrection” as false, “cutoffType” as EVALUE, and cutoff as 0.01. For protein identification, results were filtered with E-value better than 0.009. E-value describes the possibility of observing a similar matching by chance in a database of the current size.

2.3 Results and discussion

2.3.1 Sample

This study employed the proteins derived from short-term culture filtrates of *M. marinum*. This bacterium is closely related to the causative agent of tuberculosis (*M. tuberculosis*) and is often used as a model system for the study of some aspects of that disease, specifically ESX-1 protein secretion. It has been previously reported the comparison of both CZE and UPLC for the bottom-up analysis of this secretome; CZE identified 140 proteins and UPLC identified 134 proteins. In both cases, analysis required roughly three hours of mass spectrometer time.
2.3.2 Conductivity of acetic and formic acids

Despite the success of CZE in bottom-up proteomics and the top-down analysis of standard proteins, there has been limited work on extension of CZE-ESI-MS/MS for the top-down characterization of proteins from a complex sample. One challenge hindering the application of CZE to top-down proteomics is protein solubilization. A clue to enhanced protein solubilization comes from reports that employ organic acids to solubilize membrane proteins.\textsuperscript{145} As an example, Catherman employed a high concentration of formic acid to solubilize intact proteins for LC-MS analysis.\textsuperscript{146} Unfortunately, high concentrations of formic acid are not compatible with CZE because the high conductivity of formic acid results in high current and band broadening.

Intriguingly, there is a dramatic difference in conductivity between acetic and formic acid solutions at concentrations up to 50% in concentration.\textsuperscript{34} Published data cover a limited concentration range. To extend data to higher concentrations, I determined the conductivity of aqueous acetic acid and formic acid solutions by applying 6 kV across a 60 cm capillary filled with acetic acid and formic acid in water at concentrations ranging from 0.1% to 100% and measuring current. Ohm’s law and the capillary geometry were used to calculate conductivity, Figure 2.1. Across all concentration ranges studied, acetic acid solutions have much lower conductivity than formic acid. Furthermore, this data suggests that very high concentrations of acetic acid (>50%) will have lower conductivity than the 0.25% formic acid running buffer that is commonly used in CZE analysis of proteins.
I also examined the current in a capillary filled with plugs of 70% acetic acid in a capillary filled with 0.25% formic acid running buffer. Plugs of acetic acid between 0 and 27 cm in length were injected into a 40 cm LPA coated capillary by pressure. The resistance of the capillary increased linearly with plug length, Figure 2.2. The resistance across the 40-cm long capillary was 1.4 GΩ when the capillary was filled with formic acid, and the resistance increased at a rate of 96 MΩ per centimeter of injected acetic acid. These resistance values correspond to a conductivity of 1.5 mS/cm for 0.25% formic acid and 0.5 mS/cm for 70% acetic acid; the conductivity of 70% acetic acid is roughly three times lower than the 0.25% formic acid separation buffer. These results suggest that a modest stacking effect can be expected for samples prepared in 70% acetic acid used with a 0.25% formic acid background electrolyte, due to the lower conductivity of the 70% acetic acid sample buffer.

Next, to evaluate the compatibility of 70% acetic acid as sample electrolyte with CZE-MS system, about 30 ng cytochrome c was dissolved in 0.25% FA and in 70% acetic acid solutions, and analyzed the samples by CZE-ESI-MS under the same conditions. Triplicate runs were performed for both sample solutions with a LTQ-XL mass spectrometer. On average, the peak height and widths were the same for the two solutions, although the variance for both peak height and width were larger in 70% acetic acid. The migration time was consistently 20% longer for the sample prepared in 70% acetic acid (p<0.025). Longer migration time in 70% acetic acid was likely due to the higher viscosity of the acetic acid solution compared with water.\textsuperscript{147}
2.3.3 Analysis of secretome from *Mycobacterium marinum*

Normalized collision energy (NCE) was first varied to optimize the number of protein identifications with a *M. marinum* WT secreted protein sample. The number of identifications maximized with NCE near 30%. Lower NCE resulted in poor fragmentation of the selected precursor ion, so fewer product ions were generated, causing poor tandem mass spectra matching. Higher NCE generated tandem spectra that were too complex for identification. It is worth mention that all mass spectrometry parameters used here were generic, and there was no modification made to the commercially available Q Exactive mass spectrometer.

The *M. marinum* WT secreted protein sample was characterized. A ~500 ng protein aliquot was injected. As shown in Figure 2.3, the separation window was about 35 min, and the peak widths were less than 1 min. 22 proteins were identified in a single run when NCE was set to 28% (Table 2.1). The protein identification efficiency (the number of protein IDs per hour instrument time) is similar to those reported by Tran et al., who identified 1,043 proteins in 45 hour-long LC-MS runs. The size of identified proteins ranged from several kDa to over 20 kDa. The high-resolution mass spectrometer resolved isotopic peaks for these relatively low molecular weight proteins (Figure 2.3). Most of these proteins were also identified in our bottom-up study of this secretome. Five of the detected proteins were not present in our earlier bottom-up proteomics study of *M. marinum* secretome; those proteins were all hypothetical proteins.
All of the identified proteins had molecular weights less than 25 kDa. The Q Exactive mass spectrometer has a resolution of 140,000 \((m/z \ 200)\), which limits our ability to identify larger proteins; a mass spectrometer with higher resolving power will be required to extend our top-down analysis to higher molecular weight proteins. This low-molecular weight bias likely accounts for the decreased number of protein identifications compared with our bottom-up analysis of the *M. marinum* secretome. Moreover, the nature of this secretome suggests that large proteins are present in low abundance, which makes their identification difficult. Also, there are several small proteins with extremely high abundance, which can induce ionization suppression of co-migrating proteins. Base on the number and size of identified proteins, our system still has limited separation and identification ability compared to LC-MS system.\(^{27}\) This limitation is caused by the small sample injection amount and the narrow separation window of capillary electrophoresis compared with HPLC. Protein prefractionation should improve the results, which will be addressed in following studies.

Top-down proteomics has a distinct advantage in exploring protein complexity by generating information on proteoforms. 58 proteoforms were observed from 22 gene products, including 16 proteoforms components of the TypeVII ESX-1 protein secretion system (CFP-10 and ESAT-6), which is essential for virulence in pathogenic mycobacteria and conserved in several gram positive pathogens. For CFP-10, protein isomers were also separated and observed from the base peak electropherogram appearing as small peaks (Figure 2.3), from which 15 proteoforms were identified. Post-translational modifications include signal peptide removal, N-terminal methionine
excision, and acetylation. Only the N-terminal acetylation form of ESAT-6 was found in our database search. However, the existence of its unacetylated form was confirmed by manually checking the spectrum (Figure 2.4).

Quality tandem spectra were obtained with the optimized collision energy. An example is shown in Figure 2.5, the best matching spectrum for 10 kDa culture filtrate antigen EsxB (CFP-10) generated 85 matched fragment ions, and 80 of them were of less than 5 ppm mass error. Also, an N-terminal methionine excision was observed from the tandem mass spectrum.

2.4 Conclusion

Capillary zone electrophoresis (CZE) with an electrokinetically-pumped sheath-flow nanospray interface was coupled with a high-resolution Q Exactive mass spectrometer for the analysis of culture filtrates from *Mycobacterium marinum*. 22 gene products were confidently identified from the wildtype *M. marinum* secretome in a single CZE-tandem mass spectrometry (MS/MS) run. A total of 58 proteoforms was observed with post-translational modifications including signal peptide removal, N-terminal methionine excision, and acetylation. The conductivities of aqueous acetic acid and formic acid solutions were measured from 0.1% to 100% concentration (v/v). 70% acetic acid provided lower conductivity than 0.25% formic acid, and was evaluated as low ionic-strength and CZE-MS compatible sample buffer with good protein solubility.
Figure 2.1: Conductivity of aqueous solutions of acetic and formic acids at 25 °C. Conductivity was determined from the current generated when applying 6 kV voltage across a 60 cm long, 20 μm ID capillary. Both capillary ends were immersed in 0.1% FA during electrophoresis. To produce a stable reading, current was recorded 10 s after applying voltage. Uncertainties in data are ~5%. Data points are connected by straight lines.
Figure 2.2: Electrical resistance across a 40-cm long, 50 µm ID capillary filled with plugs of 70% acetic acid. The running buffer was 0.25% formic acid. Both capillary ends were immersed in 0.25% FA during electrophoresis after the acetic acid solution was injected. To produce a stable reading, current was measured 10 s after applying a 16 kV across the capillary.
Figure 2.3: Base peak electropherogram of the secreted proteins analyzed by the CZE–ESI-MS/MS system. Selected peaks were labeled with identified protein spectra. Superscript numbers indicate the protein rank in Table 2.1. The voltage applied was 15 kV for CE separation and 1.2 kV for electrospray. Inserts show parent ion spectra for proteins centered at the indicated m/z values.
Figure 2.4: Spectrum of both unacetylated ESAT-6 and acetylated ESAT-6.
Figure 2.5: HCD fragmentation of the 10-kDa culture filtrate antigen EsxB. (A) Fragmentation spectra of the [M + 7H]^{7+} charge state with HCD (normalized collision energy was 28%). (B) Sequence of this protein and the fragmentation patterns observed with HCD.
### TABLE 2.1
IDENTIFIED PROTEINS IN A SINGLE SHOT TOP-DOWN CZE ANALYSIS OF THE *M. MARINUM* SECRETOME

<table>
<thead>
<tr>
<th>Rank*</th>
<th>Accession</th>
<th>Name</th>
<th>Size (kDa)</th>
<th>Bottom-up dataset$^{55}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>gi</td>
<td>183980221</td>
<td>10 kDa culture filtrate antigen EsxB</td>
<td>10.6</td>
</tr>
<tr>
<td>2</td>
<td>gi</td>
<td>183985424</td>
<td>hypothetical protein MMAR_5453</td>
<td>5.7</td>
</tr>
<tr>
<td>3</td>
<td>gi</td>
<td>183980745</td>
<td>hypothetical protein MMAR_0722</td>
<td>15.0</td>
</tr>
<tr>
<td>4</td>
<td>gi</td>
<td>183985379</td>
<td>immunogenic protein Mpt64</td>
<td>22.7</td>
</tr>
<tr>
<td>5</td>
<td>gi</td>
<td>183983668</td>
<td>low molecular weight antigen Cfp2</td>
<td>12.2</td>
</tr>
<tr>
<td>6</td>
<td>gi</td>
<td>183984660</td>
<td>hypothetical protein MMAR_4692</td>
<td>12.3</td>
</tr>
<tr>
<td>7</td>
<td>gi</td>
<td>183985108</td>
<td>cold shock protein A CspA_1</td>
<td>7.2</td>
</tr>
<tr>
<td>8</td>
<td>gi</td>
<td>183982932</td>
<td>hypothetical protein MMAR_2929</td>
<td>8.3</td>
</tr>
<tr>
<td>9</td>
<td>gi</td>
<td>183985378</td>
<td>hypothetical protein MMAR_5548</td>
<td>4.2</td>
</tr>
<tr>
<td>10</td>
<td>gi</td>
<td>183982679</td>
<td>hypothetical protein MMAR_2672</td>
<td>8.9</td>
</tr>
<tr>
<td>11</td>
<td>gi</td>
<td>183985410</td>
<td>hypothetical protein MMAR_5439</td>
<td>3.7</td>
</tr>
<tr>
<td>12</td>
<td>gi</td>
<td>183982898</td>
<td>PE family protein</td>
<td>4.5</td>
</tr>
<tr>
<td>13</td>
<td>gi</td>
<td>183984791</td>
<td>cold shock protein a, CspA</td>
<td>7.2</td>
</tr>
<tr>
<td>14</td>
<td>gi</td>
<td>183981569</td>
<td>hypothetical protein MMAR_1553</td>
<td>14.5</td>
</tr>
<tr>
<td>15</td>
<td>gi</td>
<td>183983350</td>
<td>transmembrane protein, MmpS5_2</td>
<td>9.1</td>
</tr>
<tr>
<td>16</td>
<td>gi</td>
<td>183985421</td>
<td>6 kDa early secretory antigenic target EsxA (EsaT-6)</td>
<td>10.0</td>
</tr>
<tr>
<td>17</td>
<td>gi</td>
<td>183980929</td>
<td>hypothetical protein MMAR_0908</td>
<td>9.5</td>
</tr>
<tr>
<td>18</td>
<td>gi</td>
<td>183985025</td>
<td>lipoprotein DsbF</td>
<td>14.6</td>
</tr>
<tr>
<td>19</td>
<td>gi</td>
<td>183980785</td>
<td>PPE family protein, PPE10</td>
<td>8.6</td>
</tr>
<tr>
<td>20</td>
<td>gi</td>
<td>183982895</td>
<td>hypothetical protein MMAR_2891</td>
<td>10.2</td>
</tr>
<tr>
<td>21</td>
<td>gi</td>
<td>183982952</td>
<td>hypothetical protein MMAR_2949</td>
<td>15.3</td>
</tr>
<tr>
<td>22</td>
<td>gi</td>
<td>183983815</td>
<td>hypothetical protein MMAR_3840</td>
<td>9.8</td>
</tr>
</tbody>
</table>

*Rank is based on E-value (E<9e-4)
CHAPTER 3:

COUPLING CZE WITH ELECTRON TRANSFER DISSOCIATION AND ACTIVATED ION ELECTRON TRANSFER DISSOCIATION

3.1 Introduction

Even as CZE technology rapidly improves for proteomic applications, this separation technique has yet to capitalize on the advantages of electron-driven dissociation, remaining one-dimensional in its exclusive use of canonical collision-based dissociation methods for protein characterization.\textsuperscript{56,70,148} Electron-driven dissociation methods, e.g. electron capture dissociation (ECD)\textsuperscript{149} and electron transfer dissociation (ETD),\textsuperscript{150,151} have been a significant boon to intact protein analysis over the past 15 years, providing extensive cleavage of peptide and protein backbone bonds. These dissociation methods leverage electron rearrangements driven by capture of free, low-energy electrons (ECD) or transfer of an electron from radical reagent anions (ETD) to generate sequence-informative c- and z-type fragment ions. Both ECD and ETD retain labile PTMs and promote backbone bond cleavage largely independent of amino acid sequence, addressing several limitations intrinsic to the threshold-type dissociation mechanism of

\textsuperscript{2} A portion of this chapter was published as: Zhao Y, Riley NM, Sun L, Hebert AS, Yan X, Westphall MS, Rush MJ, Zhu G, Champion MM, Mba Medie F, Champion PA, Coon JJ, Dovichi NJ. Coupling capillary zone electrophoresis with electron transfer dissociation and activated ion electron transfer dissociation for top-down proteomics. Anal Chem. 2015;87:5422-9. doi: 10.1021/acs.analchem.5b00883.
collision-based methods like collisionally activated dissociation (CAD)\textsuperscript{152-155} and higher-energy collisional dissociation (HCD)\textsuperscript{156,157}. The transferability of ETD to any mass spectrometry platform with an rf trapping device has made it especially valuable as top-down proteomics continues to advance beyond a few specialized labs to more ubiquitous use in the proteomic community.\textsuperscript{158-161}

The dependence of electron-driven dissociation efficiency on precursor charge density, however, has diminished the extent at which ETD can robustly fragment peptide and protein precursor cations with low charge density (i.e., higher m/z values).\textsuperscript{162-164} As precursor m/z values increase, so increases the likelihood of non-dissociative electron transfer (ETnoD),\textsuperscript{165} a process in which backbone bond cleavage occurs but the fragment ions remain bound together by non-covalent interactions. ETnoD reduces precursor-to-product ion conversion, limiting the sequence information gleaned from an ETD MS/MS event. The secondary gas-phase structure responsible for ETnoD can be disrupted through addition of supplemental energy through resonant excitation, photoactivation, and elevated temperatures (collectively termed “activated ion” techniques), effectively increasing product ion yield. Inspired by successes of activated ion ECD (AI-ECD) methods\textsuperscript{166-170} that mitigate non-dissociative electron capture events using a variety of strategies, tools have been developed to increase peptide identifications with ETD using collisional activation of electron transfer products\textsuperscript{171,172} as well as concurrent infrared photoactivation during ETD reactions.\textsuperscript{173-175} The latter of these techniques, termed activated ion ETD (AI-ETD), has shown the greatest potential for increasing sequence-informative product ion generation because
it not only minimizes problematic hydrogen abstraction but also incurs no additional time to the MS/MS scan event. Recently, the substantial benefits AI-ETD can offer for intact protein characterization were described, but the boosts provided by AI-ETD have yet to be coupled with online separations for top-down proteomics.

In this chapter, I will present the first study that demonstrates the compatibility of online CZE separations with ETD and AI-ETD methodologies. First, the performance of HCD, conventional ETD, and AI-ETD were compared for CZE-MS/MS experiments to analyze a mixture of four standard proteins, with molecular weights ranging from ~12 to ~29 kiloDaltons (kDa). ETD generated either comparable or greater numbers of matched fragments than HCD for three of the protein standards while maintaining higher spectra quality than HCD for all four proteins. AI-ETD extended the performance of ETD in several cases, generating more matching fragments and higher spectral quality than ETD and HCD. Following these studies, I analyzed complex protein mixtures from the secretome of *M. marinum*. Here it is showed that ETD and HCD can be performed in consecutive scans on the electrophoretic timescale, affording more than double the protein sequence coverage achieved with HCD fragmentation alone. Furthermore, AI-ETD outperformed the other fragmentation methods, identifying more proteins than ETD and generating higher quality tandem mass spectra than both ETD and HCD — sometimes accounting for even higher protein sequence coverage than ETD and HCD combined. In addition, the combination of fragments generated from AI-ETD and HCD scans improved sequence coverage by more than three-fold over HCD alone. This work not only demonstrates the amenability of ion-ion reactions to the electrophoretic
timescale, but also illustrates that ETD technology can greatly extend the analytical power CZE provides for the top-down approach.

3.2 Experimental section

3.2.1 Materials and reagents

All reagents were purchased from Sigma-Aldrich (St. Louis, MO, USA), unless stated otherwise. Formic acid (FA) and glacial acetic acid were purchased from Fisher Scientific (Pittsburgh, PA). Methanol was purchased from Honeywell Burdick & Jackson (Wicklow, Ireland). Water was deionized by a NanoPure system from Thermo Scientific (Marietta, OH). Linear polyacrylamide (LPA)-coated fused capillary (50 µm i.d. × 150 µm o.d.) was purchased from Polymicro Technologies (Phoenix, AZ).

3.2.2 Sample preparation

A standard protein mixture was prepared by dissolving cytochrome c, myoglobin, carbonic anhydrase and β-casein in 0.08% FA, 20% acetonitrile in water with concentration of 0.06 mg/mL, 0.06 mg/mL, 0.12 mg/mL and 0.4 mg/mL, respectively. The culturing of *M. marinum* and generation of short-term culture filtrates have been described elsewhere. A secreted protein fraction containing approximately 100 µg of protein, as determined by the bicinchoninic acid assay, was purified by ice-cold acetone precipitation and resuspended in 30 µL 30% acetic acid, 30% acetonitrile in water.

3.2.3 CZE-ESI-MS/MS analysis

CZE was coupled to an ETD-enabled Orbitrap Elite mass spectrometer (Thermo Fisher Scientific, San Jose, CA) that included a multipurpose dissociation cell (MDC),
permitting HCD, ETD, and AI-ETD to occur all in the same reaction vessel. The nanospray emitter inner diameter was 10–12 µm. Separation was performed in a 50 cm long, 50 µm ID, 150 µm OD LPA-coated fused capillary. The separation buffer was 0.1% (v/v) FA. The electrospray sheath liquid was 10% (v/v) methanol and 0.1% (v/v) FA. About 35 nL standard protein solution or 70 nL secretome solution was injected for each experiment. The separation voltages were 26 kV and 14 kV for standard proteins and secretome, respectively.

The inlet ion transfer tube was held at 275°C and the S-lens rf level was set at 60%. Full MS scans were acquired in the Orbitrap over the m/z 500-1500 range with a resolving power of 60,000 at m/z 400. The three most intense peaks with charge state ≥4 were selected in data-dependent fashion for fragmentation. A normalized collision energy of 25 was used for HCD, while ETD and AI-ETD reaction times were set to 10 ms and fluoranthene reagent anions were accumulated for 15 ms prior to the ion-ion reactions. During AI-ETD reactions, an external Firestar T-100 Synrad 100-W CO2 continuous wave laser (Mukilteo, WA) was triggered using instrument firmware and modification to instrument code in conjunction with a gated laser controller. The laser irradiated the trapping volume of the MDC during the entirety of the ETD reaction and 70% total output. For all AI-ETD experiments, the nitrogen pressure in the MDC was lowered to a ∆N2 pressure of ~0.1 x 10^{-10} Torr, as measured by the Penning gauge in the Orbitrap chamber, to prevent collisional cooling that negates the additional energy supplied by the infrared laser. Gas pressure was left at normal operating levels for HCD and ETD scans to ensure proper collisional dissociation (∆N2 of ~0.3 x 10-10 Torr).
Detection for all tandem mass spectra was performed in the Orbitrap with a resolving power of 30,000 at m/z 400. The MS1 AGC target value was 1,000,000 with a maximum injection time of 100 ms, while MS/MS scans has an AGC target value of 500,000 and a maximum injection time of 500 ms. Three microscans was used in both MS and MS/MS scans. An exclusion window of ±10 ppm was constructed around the monoisotopic peak of each selected precursor for 30 seconds.

3.2.4 Data analysis

Similar to the data analysis described in Chapter 2, raw files analyzing the standard protein mixture were first converted to mzXML files, while raw files from secretome analysis were separated based on fragmentation method (HCD/ETD) and converted to mzXML files with msConvert. Tandem spectra were deconvoluted and database searched with MS-Align+ software using b- and y-type ions for HCD searches and c- and z-type ions for ETD and AI-ETD. A custom database including four model proteins (bovine cytochrome c, equine myoglobin, bovine carbonic anhydrase, and bovine β-casein) was used for standard protein mixture analysis. An NCBI protein database for *M. marinum* secretome including common contaminates (5,583 protein sequences) was used for secretome analysis. The parameters for both database searches included maximum number of modifications (shift number) as 2, mass error tolerance as 10 ppm, “doOneDaltonCorrection” and “doChargeCorrection” as false, “cutoffType” as EVALUE, and cutoff as 0.001.
3.3 Results and discussion

3.3.1 Benchmarking the compatibility of ion-ion reactions with CZE-MS/MS

Prior to this study, CZE separations have exclusively employed collision-based fragmentation methods for MS/MS analysis. Given the utility of CZE and electron-driven dissociation for intact protein separations and fragmentation, respectively, the potential such a combination could have for top-down proteomics was recognized, presenting herein the first study to couple the two technologies. I first evaluated the performance of ETD on the electrophoretic timescale by analyzing a standard protein mixture, using a CZE-MS/MS analysis of the same mixture but with HCD fragmentation as a reference point for comparison. This protein mixture contained cytochrome c, myoglobin, carbonic anhydrase and β-casein, representing the molecular weight range of proteins seen in typical top-down experiments (< 30 kDa). Similar to our previous results with CZE separations of a standard protein mixture, all four proteins were nearly baseline resolved and three peaks were resolved for β-casein, which indicates that at least three proteoforms were separated (Figure 3.1, panel A). ETD MS/MS events typically require slightly longer scan times than HCD analyses to account for sufficient ETD reaction time. Longer scan times translate to fewer tandem mass spectra collected over the elution profile of a given protein, which can challenge the robustness of a top-down experiment. Panel B of Figure 3.1 shows the number of MS/MS scans that were collected over the elution profile of each of the protein standards. As expected, ETD tandem mass spectra averaged marginally longer scan times than those performed with HCD (1.38 s/scan vs. 1.21 s/scan), and fewer MS/MS scans were acquired for most of the proteins. Despite
this small drop in scan acquisition, all four proteins were identified with both ETD and HCD.

Knowing that scan speed requirements of ETD provide successful top-down analysis with CZE separations, I also analyzed the standard protein mixture using AI-ETD fragmentation, which also identified all proteins in the mixture. Table 1 summarizes the results achieved with all three fragmentation methods (HCD, ETD, and AI-ETD). ETD substantially outperformed HCD in product ion generation for the relatively smaller proteins cytochrome c and myoglobin (both < 17 kDa), while maintaining comparable production of matching fragments for β-casein (~24 kDa) and carbonic anhydrase (~29 kDa). Furthermore, ETD outperformed HCD in spectral quality for these standard proteins, producing higher matching percentages (number of matched fragments divided by total number of ions in the spectrum) for all four proteins. This difference is likely explained by the high-energy deposition in the HCD process, which often generates ammonia and water neutral losses in addition to a-type ions and low mass internal fragments produced by multiple backbone cleavages (all of which are not included in the MS-Align+ database search). AI-ETD expanded the benefits of ETD even further, generating more matching fragments than ETD for cytochrome C and carbonic anhydrase, and also surpassing the matching percentages (i.e., spectral quality) of both ETD and HCD for three of the four proteins. Fragmentation of β-casein with AI-ETD, however, produced suboptimal results, even though HCD and ETD performed similarly to each other. This may be due to several idiosyncratic characteristics of β-casein, including high levels of phosphorylation (up to five modifications within a 21 residue
stretch) and a strikingly proline-rich sequence — both of which can cause unpredicted fragmentation in the presence of IR irradiation. Regardless, these electron-driven dissociation methods provided commensurate, if not superior fragmentation of these four standard proteins compared to HCD. Additionally, AI-ETD often enhanced the results ETD could provide, granted further experiments outside the scope of this study are required to determine the amenability of AI-ETD to proteins with uncommon traits, such as β-casein.

3.3.2 Complementary fragmentation methods for top-down CZE-MS/MS

Practitioners of top-down proteomics often leverage the complementary fragmentation achieved with ETD and HCD to improve protein characterization. When comparing sites of backbone cleavage in the standard protein sequences, relatively few fragmentation sites overlap between ETD and HCD, reflecting the inherent differences in their dissociation mechanisms. The majority of cleavage sites seen with ETD and AI-ETD are the same - unsurprising, as AI-ETD is merely an enhancement of the ETD reaction. Figure 3.2 highlights the amount overlap (or lack thereof) in fragmentation sites between the three methods for each standard protein. Combining the total backbone cleavages achieved separately with HCD and ETD afforded sequence coverage gains of 106% for cytochrome c, 215% for myoglobin, 115% for β-casein, and 72% for carbonic anhydrase over HCD alone, demonstrating the improvements in protein characterization ETD can provide for current CZE-MS/MS approaches. (Note, the protein sequence coverage in this article is backbone bond coverage, defined as a percentage that
represents the number backbone bonds cleaved divided by total number of backbone bonds.)

AI-ETD further bolstered these gains in protein sequence coverage, as shown in Figure 3.3. Cytochrome c achieved near complete coverage across the entire sequence, excluding the region surrounding the heme group binding domain. The combination of AI-ETD/HCD fragments increased sequence coverage of carbonic anhydrase and β-casein by 72% and 60%, respectively. The most remarkable improvement in sequence coverage was achieved with myoglobin, as the combination of AI-ETD and HCD increased sequence coverage greater than three-fold over fragmentation solely with HCD (from 13% to 40%). HCD has been shown to provide less extensive fragmentation of myoglobin due to preferential cleavages that arise from the protein’s globular nature with regions of varying disorder, which suggests that the addition of alternative fragmentation methods like ETD and AI-ETD is crucial for sufficient characterization.

Altogether, these improvements in sequence coverage highlight the compatibility of CZE protein separations to novel fragmentation methods like AI-ETD and also demonstrate the extensive protein characterization that can be achieved using electron-driven dissociation techniques with online CZE separations.

3.3.3 Evaluating the combination of ETD and HCD for top-down proteomics on the M. marinum secretome

I followed these standard protein studies with by expanding the experimental scope to a more complex, biologically relevant protein mixture derived from the Mycobacterium marinum secretome. This opportunistic pathogen is regularly used as a
model system for studying more insidious pathogens like *Mycobacterium tuberculosis*, which is the causative agent of most cases of tuberculosis.\textsuperscript{142,177-179} Virulence of these bacterial species is thought to be a function of secreted proteins, although the exact causes are unknown and remain an area of active research. Additionally, the pool of proteins contained in the *M. marinum* secretome span a relatively small range of protein molecular weights (< 30 kDa), providing an ideal system to use as a test bed for our ETD-enabled CZE-MS/MS platform.\textsuperscript{180} Capitalizing on the success I saw above using ETD and HCD as complementary fragmentation techniques, I performed a 60 minute CZE separation using the two dissociation methods in tandem. Precursors were selected in data-dependent fashion from a high-resolution MS1 scan performed in the Orbitrap, followed by three pairs of sequential HCD and ETD tandem mass spectra, for a total of six MS/MS events per cycle. Figure 3.4 summarizes the results of this analysis. Because ETD and HCD generate different product ion types, tandem mass spectra from the two respective dissociation methods were searched separately against a complete *M. marinum* database using MS-Align+. ETD identified 26 total proteoforms, while HCD performed better, characterizing 36 proteoforms. Altogether, ETD and HCD accounted for 41 proteoforms, corresponding to 21 unique gene products (panel A, Figure 3.4). This analysis produced fewer proteoform identifications but nearly the same number of total gene product identifications as our previous CZE-MS/MS analyses using solely HCD. This reduction in number of identified proteoforms is attributed to the redundant precursor sampling of subsequent HCD and ETD scans, although this did not adversely affect the number gene products characterized.
HCD spectra generally earned better identification scores, i.e., E-values, than ETD spectra (panel B, Figure 3.4), although the quality of ETD spectra could be improved in future studies through more judicious selection of ETD reaction parameters, such as reaction duration and reagent anion accumulation time. Regardless, the addition of ETD to the CZE-MS/MS platform still provided considerable advantages. Panel C of Figure 3.4 displays the gain in percent protein sequence coverage that was contributed by the addition of ETD fragmentation over HCD alone for 16 gene products detected with both methods. This value does not represent a fold-change measurement but rather depicts the raw difference in percent sequence coverage observed. For example, iron-regulated Lsr2 protein (a protein involved in the iron-dependent protein secretion implicated in bacterial virulence,181,182, accession number 183982454) had a sequence coverage of only 15.2% with HCD fragmentation. When fragments generated from HCD were combined with fragments from ETD, this sequence coverage improved to 31.1%, a gain in raw present sequence coverage of ~16% (corresponding to a more than two-fold improvement). These results illustrate that ETD can be coupled with HCD in sequential scans on the electrophoretic timescale, improving protein characterization achievable with front-end CZE separations.

3.3.4 Advantages of AI-ETD for top-down CZE-MS/MS on complex proteins mixtures

Spring-boarding from the success of combining subsequent ETD and HCD in a single CZE-MS/MS experiment, I then analyzed the M. marium secretome protein mixture with a combination of AI-ETD and HCD in the same fashion. The CZE separations
were highly reproducible: many proteins that were detected in the first ETD/HCD analysis (Run 1) were also identified in the AI-ETD/HCD experiment (Run 2), enabling the comparison of ETD, AI-ETD, and HCD for shotgun top-down proteomics. Figure 3.5 encapsulates these comparisons, showing results for *M. marinum* protein 10 kDa culture filtrate EsxB_1 (accession number 183980221). This protein is directly connected with *M. marinum* virulence through involvement in a novel secretion apparatus and is in a family of proteins targeted in current tuberculosis research.\(^{142,182}\) Panel A of Figure 3.5 shows that AI-ETD maintained optimal product ion generation over both ETD and HCD, even as precursor charge density decreased. For the +10 precursor, ETD and AI-ETD both provided complementary fragmentation to HCD (Panel B, Figure 3.5), although AI-ETD substantially increased the number of total fragmentation sites over ETD and HCD alone (54 with AI-ETD vs. 19 with ETD, 34 with HCD Run 1, and 35 with HCD Run 2). As described above, the combination of ETD and HCD fragments improved sequence coverage over HCD alone (from \(~35\%) to \(~48\%)\), but AI-ETD still afforded the highest sequence coverage on its own (\(~55\%)\), even over the combination of ETD and HCD (Panels C and D, Figure 3.5).

For a more holistic evaluation of the combination of AI-ETD and HCD in the same experiment, Figure 3.6 summarizes the results of Run 2 and permits straightforward comparisons to ETD/HCD performance in Run 1. AI-ETD identified more proteoforms (29) than ETD, although the overlap in identifications between AI-ETD and HCD in Run 2 was higher than ETD/HCD overlap seen in Run 1 (Panel A, Figure 3.6). As with the standard protein analysis above, AI-ETD outperformed HCD in spectral quality in Run 2, generally
affording better identification scores for proteins across the molecular weight range observed (Panel B, Figure 3.6). AI-ETD also significantly improved the quality of tandem mass spectra acquired over ETD in Run 1, with average E-values of $7.3 \times 10^{-10}$ and $1.0 \times 10^{-4}$, respectively, while HCD identifications score remained relatively constant between the two experiments (Run 1 average of $1.4 \times 10^{-5}$ and Run 2 average of $4.8 \times 10^{-6}$). Combining fragments from AI-ETD and HCD typically increased gains in percent protein sequence coverage (Panel C, Figure 3.6) more than the ETD/HCD combinations from Run 1. Panel D of Figure 3.6 shows the extensive sequence coverage achieved from the AI-ETD/HCD combination for transcription elongation factor GreA (accession number 183984358). Here HCD provided 19 total fragmentation sites, or 11.6% sequence coverage. The addition of AI-ETD fragments contributed a gain in sequence coverage of over 25%, a more than three-fold improvement in sequence coverage. These results show that AI-ETD is a robust fragmentation method for improved ETD performance that can be used in top-down experiments on its own or can be complemented by HCD for even further enhancement in protein characterization.

3.4 Conclusion

In this study, I have described the first implementation of front-end CZE separations with ETD, leveraging the strengths of both for top-down protein analysis of simple and complex protein mixtures. Furthermore, I have shown that CZE separations are also compatible with novel dissociation techniques like AI-ETD for enhanced intact protein fragmentation, showing the impact AI-ETD can have in improving top-down
results. Importantly, I also demonstrate that electron-driven dissociation methods like ETD and AI-ETD can be coupled in tandem with collision-based fragmentation (e.g., HCD) in the same CZE-MS/MS analysis, offering superior protein characterization via complementary fragmentation mechanisms on the electrophoretic timescale.

Ultimately, this work highlights the compelling potential electron-driven dissociation methods have for top-down proteomics with CZE separations, opening new possibilities for intact protein analysis.

3.5 Acknowledgments

I would like to thank Nicholas Riley and Alexander Hebert in Dr. Joshua Coon’s group at the University of Wisconsin Madison for the collaboration for this work. They generously provided the mass spectrometer, technique and data analysis support, and comments and insight to the work described in this Chapter.
Figure 3.1: Capillary zone electrophoresis on an ETD-enabled Orbitrap system. (A) Base peak electropherogram of four standard proteins separated by the CZE-ESI-MS/MS system, including three proteoforms of β-casein (labeled a-c). Colored MS1 spectra correspond to elution profiles of the designated proteins. (B) Comparison of number of tandem mass spectra acquired for each protein standard using either HCD or ETD, with average time per MS/MS scan provided.
Figure 3.2: Overlap of fragmentation sites from Al-ETD and HCD spectra for cytochrome c (A), carbonic anhydrase (B), β-casein (C), and myoglobin (D).
Figure 3.3: Sequence maps show combined fragmentation patterns of AI-ETD and HCD (A), and ETD and HCD (B) for cytochrome c (top left), carbonic anhydrase (bottom left), β-casein (top right) and myoglobin (bottom right). *Number of fragmentation sites: total fragmentation sites/overlapped fragmentation sites.
Figure 3.4: Analysis of M. marinum secretome with ETD and HCD, Run 1. (A) The overlap in proteoforms identified using back-to-back ETD/HCD scans in a single CZE-MS/MS run, mapping to a total of 21 gene products. (B) Identification score (E-value) versus protein molecular weight for proteoforms identified with ETD and HCD. (C) Gain in percent sequence coverage when using a combination of fragments from an ETD scan and an HCD scan together over the sequence coverage achieved with HCD alone. Proteins on the y-axis are identified by their accession number.
Figure 3.5: Comparison of fragmentation quality using HCD, ETD, and AI-ETD for the characterization of 10 kDa culture filtrate antigen EsxB_1, a protein known to play a role in virulence in M. marium. (A) The number of matching fragments generated by the three fragmentation methods at two different precursors of antigen EsxB_1. HCD results from both Run 1 (which included ETD scans) and Run 2 (which included AI-ETD scans) are included. (B) Overlap of fragmentation sites using ETD and HCD spectra from Run 1, and using AI-ETD and HCD spectra from Run 2. (C) Sequence map depicting the bonds cleaved by AI-ETD alone in a single MS/MS scan. (D) Sequence map depicting the bonds cleaved by a combination fragments from an ETD scan and HCD scan. *total bonds cleaved/shared cleavage sites
Figure 3.6: Analysis of M. marinum secretome with AI-ETD and HCD, Run 2. (A) The overlap in proteoforms identified using back-to-back AI-ETD/HCD scans in a single CZE-MS/MS run, mapping to a total of 19 gene products. (B) Identification score (E-value) versus protein molecular weight for proteoforms identified with AI-ETD and HCD. (C) Gain in percent sequence coverage when using a combination of fragments from an AI-ETD scan and an HCD scan together over the sequence coverage achieved with HCD alone. Proteins on the y-axis are identified by their accession number. (D) Sequence map depicting the number of bonds cleaved for Transcription Elongation Factor GreA (accession number: 183984358), explained by a combination of fragments generated from an AI-ETD scan and an HCD scan. *total bonds cleaved/shared cleavage sites
4.1 Introduction

In this chapter, I extend the application of the CZE-ESI-MS platform to the characterization of monoclonal antibodies. Monoclonal antibodies (mAbs) are tetrameric glycoproteins with molecular weights of approximately 150 kDa. They consist of four polypeptide chains: two heavy chains (HCS) (~50 kDa each) and two light chains (LCS) (~25 kDa each) linked by disulfide bonds. The Fc regions on HCs bear highly conserved N-glycosylation sites that contribute to the mAbs heterogeneity. Other modifications such as deamidation and oxidation can also impact the function of mAbs. As a result, quality control is required to ensure mAb efficiency, consistency, and stability. Conventionally, bottom-up liquid chromatography-mass spectrometry (LC-MS) is used to characterize isoforms of a digested mAb. However, this bottom-up approach suffers from incomplete sequence coverage, the introduction of artificial modifications, and time-consuming data analysis.

\[^{3}\text{A portion of this chapter was published as: Zhao Y, Sun L, Knierman MD, Dovichi NJ. Fast separation and analysis of reduced monoclonal antibodies with capillary zone electrophoresis coupled to mass spectrometry. Talanta 2016; 148: 529-33. doi: 10.1016/j.talanta.2015.11.020.}\]
To overcome these shortcomings, the intact antibody could be directly analyzed by mass spectrometry. However, this top-down analysis requires use of mass spectrometers with extremely high resolving power. Middle-down approaches, on the other hand, analyze subunits of mAbs, including reduced mAbs (HCs and LCs) and large fragments generated by limited proteolysis (Fab and Fc fragments). Middle-down analysis can use mass spectrometers with lower resolution than top-down approaches and introduces fewer artifacts compared to traditional bottom-up approach.

Although liquid chromatography is more often used for mAbs separations, various proteomic studies have investigated CZE-ESI-MS as a promising platform for mAbs characterization, mostly through glycan analysis. In this chapter, I will describe and discuss the use of CZE for separation of light and heavy chains of reduced antibodies with mass spectrometric analysis. I first developed a simple protocol to condition commercial linear-polyacrylamide coated capillaries for use in top-down proteomics. I then suspended reduced antibodies in a solution of 35% acetic acid, 50% acetonitrile in water. Heavy and light chains were baseline resolved within 10 minutes and with 3 to 30 µg/mL detection limits using a 0.1% aqueous formic acid background electrolyte. Quintuplicate runs of a two-antibody mixture produced relative standard deviations of ~1% in migration time and 10% in peak amplitudes. Resolution was further improved for the two-antibody mixture by using 5% acetic acid as the background electrolyte, highlighting the potential of capillary electrophoresis-mass spectrometry for analysis of antibody mixtures.
4.2 Experimental section

4.2.1 Materials and reagents

All reagents were purchased from Sigma-Aldrich (St. Louis, MO, USA), unless stated otherwise. Formic acid (FA) and glacial acetic acid were purchased from Fisher Scientific (Pittsburgh, PA). Methanol was purchased from Honeywell Burdick & Jackson (Wicklow, Ireland). Water was deionized by a NanoPure system from Thermo Scientific (Marietta, OH). TCEP solution was included in an iTRAQ kit from Ab Sciex. C4 Ziptip was purchased from EMD Millipore (Billerica, MA). Mouse Anti-Human IgG4 Fc-UNLB was purchased from SouthernBiotech. Human/Mouse/Rat Activin A beta A subunit Antibody was purchased from R&D System. Anti-Phosphotyrosine Antibody was purchased from EMD Millipore. Linear polyacrylamide (LPA)-coated fused capillary was purchased from Polymicro Technologies (Phoenix, AZ).

4.2.2 Sample preparation

A 10 μg aliquot of antibody solution was denatured and reduced by adding twice the volume of 6 M guanidine HCl dissolved in 100 mM NH₄HCO₃ followed by adding 2 μL of 500 mM TCEP solution. The mixture was incubated at 37°C for 30 min. About half a microliter of solid iodoacetamide (dispensed with a pipette tip) was directly dissolved in the mixture and the mixture was incubated at 37°C in the dark for 15 min. Finally, the antibody solution was desalted with a C4 Ziptip and eluted with 35% acetic acid and 50% acetonitrile in water.
4.2.3 CZE-ESI-MS/MS analysis

The CZE system was coupled to an LTQ XL or a LTQ-Orbitrap Velos mass spectrometer (Thermo Fisher Scientific, San Jose, CA USA). Electrospray was generated using an electrokinetically pumped sheath flow nanospray emitter with an inner diameter at 10–12 µm.

Separation was performed using a commercially coated 50 cm long, 30 or 50 µm I.D., 150 µm O.D. LPA-coated fused capillary (Polymicro Technologies, Phoenix AZ USA). Initial separations of antibodies in these commercially-coated capillaries were unsuccessful, presumably due to irreversible adsorption onto active sites within the capillary. To minimize this adsorption, a standard protein mixture was prepared in 0.04% FA, 20% ACN in water by dissolving cytochrome c, myoglobin, carbonic anhydrase, and β-casein with concentrations of 0.1 mg/mL, 0.1 mg/mL, 0.2 mg/mL, and 0.8 mg/mL, respectively. About 20 nL of this standard protein solution was injected into the capillary. After injection, both voltage and pressure were applied to pump the mixture through the capillary. Once a signal was detected by the mass spectrometer, the capillary was flushed with background electrolyte (0.1% FA) for 10 min. This conditioning step was successful in reducing loss to the inner surface and allowed detection of antibody chains.

The background electrolyte was 0.1% (v/v) FA and the electrospray sheath liquid was 10% (v/v) methanol and 0.1% (v/v) FA, unless stated otherwise. The separation voltage was 20 kV. Full MS scans were acquired over the m/z 600-2000 range in the LTQ-XL mass spectrometer. For LTQ-Orbitrap Velos detection, full MS scans were
acquired in the Orbitrap detector over the m/z 600-2000 range at a resolution of 60,000 at m/z of 400. The c-trap pressure was lowered from 0.5 to a value between 0.1 and 0.2 on the pressure regulator.

4.3 Results and discussion

4.3.1 LPA-coated capillary treatment

Fused silica capillary tends to adsorb intact proteins, which causes protein loss during CZE analysis.\textsuperscript{190} LPA coating is commonly used to minimize adsorption to the capillary wall. Unfortunately, both batch-to-batch and within-batch difference have been observed in the performance of commercially available LPA-coated capillaries. In some cases, severe protein loss is observed during the first few runs, presumably due to adsorption of proteins onto uncoated regions. However, the signal often recovers for later runs. This observation led to a simple strategy to pre-occupy those uncoated spots with sacrificial proteins before injecting analyte. In this protocol, a plug of relatively high concentration standard proteins is passed through the column using a combination of pressure and electrokinetic flow. The capillary is ready for analysis once the proteins have left the capillary.

I evaluated the performance of the treated capillary by performing electrophoresis with the same standard protein sample that was used to treat the capillary. No signal was observed for the untreated capillary, whereas signal was restored following treatment, Figure 4.1. The treatment was also effective after flushing the capillary with 0.1% formic acid.
4.3.2 High concentration acetic acid aids in solubilizing reduced antibodies

Samples were first desalted using a C4 ZipTip and then eluted in the injection electrolyte; desalted samples undergo stacking during injection, allowing use of larger injection volumes. I observed no antibody signal after sample desalting with elution in 50% acetonitrile 0.1% formic acid in our initial experiments. It was suspected that reduced antibodies tend to precipitate when eluted in 50% acetonitrile 0.1% formic acid. High concentration acetic acid (70% (v/v)) has been employed to improve the solubility of secretome from *M. marinum* in top-down CZE analysis. However, inefficient ionization was observed during ESI process with the use of an acetic acid solution. In this experiment, I decreased the acetic acid concentration to 35% (v/v) with no observable decrease in ionization efficiency. 35% acetic acid and 50% acetonitrile (ACN) was used to elute the reduced antibody from a C4 ZipTip column to help retain the denatured antibody chains in solution. The resulting desalted antibody sample can be kept in solution after desalting for months, and can be directly injected for CZE-MS analysis. Duplicate runs were performed for mouse anti-human IgG4 Fc-UNLB, which is a mouse IgG1 antibody specific to the Fc region of human IgG4 antibody (Figure 4.2 A). Heavy and light chains were baseline separated within 10 min. Two major isoforms for both heavy chain and light chain were partially separated with peak width of about 30 seconds. Migration profile and retention times for duplicate runs were similar. The second run was performed two hours after the first run. Since both acetic acid and ACN are volatile, there was a decrease in sample volume after two hours. As a result, the peak intensity in the second run was about twice as high as in the first run, which
matched the increase in sample concentration estimated by observing the volume of sample solution.

Knowing that surfactants are commonly used to solubilize proteins, I also investigated the use of 0.01% octyl β-D-glucopyranoside (OG) and 50% ACN to elute the reduced antibody from the C4 ZipTip column. As shown in Figure 4.2 B, CZE-ESI-MS runs were performed 10 min, one hour, and 1.5 hour after desalting. There were two peaks (heavy chain and light chain) observed in the extracted ion electropherogram with reasonable intensities for the first run. After one-hour storage at 4 °C, the signal intensity reduced significantly, and no signal was observed after 1.5 hour. Other surfactants (CHAPS and digitonin) were tested, but no antibody signal was detected after two hours of storage. With the above results, it was concluded that the presence of high concentrations of acetic acid in the eluting solution is important for enhancing reduced antibody solubilization.

A standard protocol of reduced antibody sample preparation for CZE-MS analysis was developed, and sample preparation requires one hour, including reduction, alkylation and desalting. The desalted reduced antibody was eluted in 35% acetic acid and 50% ACN in water, which is compatible with CZE-MS analysis. It is worth mention that this reduced antibody sample can be kept for months with no observed sample loss.
4.3.3 Evaluating the performance of CZE-ESI-MS for separation and detection of single reduced antibodies

I next investigated the performance of the CZE-ESI-MS system with three monoclonal antibodies. Mouse anti-human IgG4 Fc-UNLB (antibody A), human/mouse/rat activin A beta A subunit antibody (antibody B), and anti-phosphotyrosine antibody (antibody C) were reduced, alkylated, desalted, and analyzed by our CZE-ESI-MS system. All antibodies were prepared at a starting concentration of 1 mg/mL. The presence of 50% ACN in the sample resulted in stacking, which allowed an increase in the sample volume that can be injected. Approximately 70 nL (~7% of total capillary volume) of each sample was injected into the capillary. Figure 4.3 presents the electropherograms for each antibody. The light chains and heavy chains for both antibody A and B were baseline separated. Moreover, two isoforms were partially separated and observed for each polypeptide chain. There was only ~1 Da difference between LC isoforms. This difference could be the result of deamidation or incomplete reduction of intrachain disulfide bond, which introduce only 1~2 Da mass shift but introduces a difference in charge. The average separation window was about three minutes and the average peak widths (full width at baseline) for each light chain and heavy chain peaks were about 30 seconds. Plate counts were about 15,000. The light chain and heavy chain peaks overlapped for antibody C, which was caused by an extremely wide heavy chain peak. One possible explanation for the wide migration time range is that the heavy chain for this antibody is highly modified.
The noise in the baseline was estimated as the standard deviation of the region between 2 and 4 minutes. The signal-to-noise ratio for the peaks ranged from 100 to 1,000, corresponding to extrapolated detection limits of ~3 to 30 µg/mL (~2 to 20 fmol) of antibody. These data were acquired using an LTQ-XL mass spectrometer; use of a higher sensitivity mass spectrometer will improve the detection limits.

4.3.4 Evaluating and improving the performance of CZE–ESI-MS for reduced antibody mixture separation

A two-antibody mixture was prepared by mixing 10 µg of antibody A and 2 µg of antibody B in 10 µL of a 35% acetic acid and 50% ACN solution. The mixture was analyzed by the system in quintuplicate runs, Figure 4.4 A. Heavy chains and light chains were partially resolved within 10 min. The light chains were assigned by calculating their masses, while the signal intensities were not strong enough to calculate the masses of the heavy chains. No flushing was performed between runs, but the protein signals (~4×10^5) were about 100-fold higher than the background signals (~5×10^3), which was measured at 0-2 min. Sample carryover was negligible in this experiment. The migration times for the polypeptide chains were highly reproducible with relative standard deviations of about 1%. The signal intensities were reasonable consistent with a relative standard deviation of about 13%.

I improved separation efficiency by use of a 5% acetic acid background electrolyte. It was previously demonstrated that 5% acetic acid effectively increases the width of separation window by minimizing electroosmotic flow and increasing viscosity. Figure 4.4 B shows triplicate runs of a two-antibody mixture. Using 5% acetic acid as
A background electrolyte generated a separation window of 4 min, which was twice the width produced by 0.1% FA. The analysis was finished within 14 min. Both heavy chains and light chains were separated with improved resolution compared with 0.1% FA.

4.4 Conclusion

This chapter describes the first study of the separation and analysis of reduced monoclonal antibodies on a CZE-ESI-MS system. The system provides fast separation, good reproducibility, and reasonable sensitivity, which shows the potential for CE-MS serving as a quick diagnostic tool for mAb quality control. Furthermore, I have demonstrated the separation and analysis of a two-antibody mixture, highlighting the possibilities for CZE-MS to characterizing mAb mixtures and even polyclonal antibodies. However, the resolution for antibody isoforms was limited in this experiment. A higher resolution mass spectrometer can yield more information on the molecular weight of separated antibody isoforms. Also, employing a pre-fractionation method before CZE-MS analysis would greatly improve the separation resolution for more complex antibody sample, which will be addressed in future work.

4.5 Acknowledgments

I would like to thank Mike Knierman and Eli Lilly for the collaboration work. They generously provided technique support, data analysis help and comments to the work described in this Chapter.
Figure 4.1: Base peak electropherograms of a four-standard protein mixture (cytochrome c - 0.1 mg/mL, myoglobin - 0.1 mg/mL, carbonic anhydrase - 0.2 mg/mL and β-casein - 0.8 mg/mL in 0.04% FA, 20% ACN in water). Blue trace: before capillary treatment; this trace was multiplied by 100 before plotting. Orange trace: after capillary treatment. Data were treated with a Lowess filter with five-point span before they were plotted. LTQ-XL mass spectrometer was used for detection.
Figure 4.2: Base and extracted ion electropherograms of reduced mAbs. (A) Duplicate base peak electropherogram of CZE-ESI-MS analysis of a reduced mAb (mouse anti-human IgG4 Fc-UNLB), including two isoforms of heavy chain and light chain. Mass spectra correspond to elution profiles of the designated polypeptide chains. (B) Extracted ion electropherograms of reduced mouse anti-human IgG4 Fc-UNLB dissolved in 0.01%OG and 50% ACN in water. Top, middle, bottom electropherograms correspond to 10 min, 1 hour and 1.5 hours after desalting of sample. LTQ-XL mass spectrometer was used for detection.
Figure 4.3: Base peak electropherograms of CZE-ESI-MS analysis of reduced antibody A, B and C. Data treated with a Lowess filter with five-point span before plotting. 50 cm long, 50 µm I.D. capillary was used for separation. About 70 nL sample was injected for each analysis. An LTQ-XL mass spectrometer was used for detection.
Figure 4.4: Base peak electropherograms for a two-antibody mixture (mouse anti-human IgG4 Fc-UNLB and human/mouse/rat activin A beta A subunit antibody), including migration times for each heavy chain and light chain. Major peaks were assigned in the 1st run. (A) Quintuplicate CZE-ESI-MS analysis of a two-antibody mixture. 50 cm long, 30 µm I.D. capillary was used for separation. About 18 nL liquid aliquot was injected for each analysis. Voltages: 21.3 kV/1.3 kV. LTQ-Orbitrap Velos mass spectrometer was used for detection with Orbitrap as detector. (B) Triplicate CZE-ESI-MS analysis of a two-antibody mixture. 50 cm long, 30 µm I.D. capillary was used for separation. About 30 nL liquid aliquot was injected for each analysis. Sheath liquid: 0.5% FA, 10% methanol. Voltages: 23.7 kV/1.2 kV. LTQ-XL mass spectrometer was used for detection. NL is the normalization level.
5.1 Introduction

The performance of CZE can be degraded due to both the small sample injection volume that is typically employed\textsuperscript{191} and sample adsorption to inner capillary wall.\textsuperscript{190} Sample injection volume can be increased by the use of dynamic pH junction based CZE,\textsuperscript{192,193} which has been recently applied to bottom-up proteomics.\textsuperscript{194} Sample adsorption has been partially solved by the use of linear polyacrylamide (LPA)-coated capillaries. However, commercially available LPA-coated capillaries often provide poor performance. A thermally-initiated coating protocol developed by Zhu et al. improved the uniformity and stability of this coating for peptides and standard proteins.\textsuperscript{195} These two techniques were combined in this study to improve the overall performance of CZE for top-down proteomics.

\textsuperscript{4} A portion of this chapter was published as: Zhao Y, Sun L, Zhu G, Dovichi NJ. Coupling Capillary Zone Electrophoresis to a Q Exactive HF Mass Spectrometer for Top-down Proteomics: 580 Proteoform Identifications from Yeast. J Proteome Res. 2016; 15: 3679-3685. DOI: 10.1021/acs.jproteome.6b00493
In this work, I prefractionated the yeast proteome by RPLC and explored the potential of CZE-ESI-MS/MS for intact protein characterization. The parameters of the Q-Exactive HF mass spectrometer were first optimized for top-down proteomics using a mixture of seven model proteins; I observed that intact protein mode with trapping pressure of 0.2 and normalized collision energy of 20% produced the highest intact protein signals and most protein identifications. Then, I applied the optimized parameters for analysis of the fractionated yeast proteome. 580 proteoforms and 180 protein groups were identified via database searching of the MS/MS spectra. This number of proteoform identifications is two times larger than previous CZE-MS/MS studies. An additional 3,243 protein species were detected based on the parent ion spectra. Post-translational modifications including N-terminal acetylation, signal peptide removal, and oxidation were identified.

5.2 Experimental section

5.2.1 Materials and reagents

All reagents were purchased from Sigma-Aldrich (St. Louis, MO), unless stated otherwise. Formic acid (FA) and glacial acetic acid (HOAc) were purchased from Fisher Scientific (Pittsburgh, PA). Methanol was purchased from Honeywell Burdick & Jackson (Wicklow, Ireland). Water was deionized by a NanoPure system from Thermo Scientific (Marietta, OH). Fused silica capillary tubing was purchased from Polymicro Technologies (Phoenix, AZ). The RPLC column (Jupiter 5 μm C5 300 Å, LC Column 250 x 4.6 mm) was purchased from Phenomenex (Torrance, CA).
5.2.2 Sample preparation

Cytochrome c, myoglobin, carbonic anhydrase, β-casein, insulin, superoxide dismutase and ubiquitin were dissolved in 0.07% formic acid and 30% acetonitrile in water with concentrations of 0.1 mg/mL, 0.2 mg/mL, 0.2 mg/mL, 0.8 mg/mL, 0.1 mg/mL, 0.1 mg/mL and 0.1 mg/mL, respectively. For the yeast proteome, a small portion (~ 0.2 g) of commercial baker’s yeast (S. cerevisiae) (Red Star Active Dry Yeast) was added to 150 mL of yeast extract peptone dextrose (YPD) medium (50 g DifcoTM YPD in 150 mL distilled water, autoclaved at 120 °C for 15 min) and grown overnight at 37 °C on a shaker. Yeast lysis was performed as described.119 Briefly, a yeast suspension was centrifuged at 4000 g for 5 min and washed with PBS for 3 times. After adding lysate buffer (100 mM DTT, 5% SDS), the cell suspension was heated at 95 °C for 5 min, followed by sonication for 15 min at the maximum power. Finally, the lysate was centrifuged at 16000 g for 5 min and the supernatant was collected. Next, 100 μL of 1 M iodoacetamide were added to 200 μL of yeast protein extract and reacted for 20 min at room temperature. Cold acetone precipitation was performed by adding 1.2 mL of cold acetone (-20 °C) to 300 uL of sample, incubating at -20 °C overnight, centrifuging at 18000 g for 15 min, washing with cold acetone, and centrifuging again. An ~800 μg of protein pellet was resuspended in 400 μL of a solution containing 8 M urea and 100 mM NH4HCO3. The sample solution was passed through a 30 kDa centrifugal filter (Millipore, MA). The protein was extracted from the filter membrane with 8 M urea and 100 mM NH4HCO3 to form a final protein solution of 500 μL (~1.6 mg/mL based on BSA analysis of the sample before acetone precipitation).
5.2.3 RPLC fractionation

Intact protein fractionation was performed on a Waters e2695 HPLC system with a C5 stationary phase Jupiter column, 5 μm particle diameter, 300 Å pore size, 250 x 4.6 mm column size. Mobile phase A was composed of water with 0.1% FA. Mobile phase B was composed of acetonitrile with 0.1% FA. The operating flow rate was 0.8 mL/min. The RPLC system was activated with 80% mobile phase B for 10 min, and then equilibrated with 5% mobile phase B for 10 min. A ~320 μg yeast protein sample was injected, followed by 10 min of washing with 5% of mobile phase B. A 70 min linear gradient was set from 5% mobile phase B to 80% mobile phase B. A total of 46 fractions were collected from 15 min to 61 min (one fraction per minute). The fractions collected were then lyophilized and suspended in 5 mM NH4HCO3. The 46 RPLC fractions were further combined to 23 fractions with a final volume of 5 μL and were subjected to CZE-ESI-MS/MS analysis.

5.2.4 CZE-ESI-MS/MS analysis

The preparation of a LPA-coated capillary was described elsewhere. Briefly, the silica capillary was pretreated with gamma-methacryloxypropyl-trimethoxysilane, and then the monomer mixture and ammonium persulfate initiator were introduced into the capillary without TEMED initiator. The filled capillary was heated in a water bath to initiate polymerization to form LPA coatings. CZE was coupled to a Q Exactive HF mass spectrometer (Thermo Fisher Scientific, San Jose, CA). Electrospray was generated using an electrokinetically pumped sheath flow through a nanospray emitter. The borosilicate glass emitter (1.0 mm o.d. × 0.75 mm i.d., 10 cm length) was pulled with a
Sutter P-1000 flaming/brown micropipet puller. The emitter inner diameter was 15-20 µm. Separation was performed in 60 cm and 80 cm long, 50 µm ID, 150 µm OD LPA-coated fused silica capillaries for standard proteins and yeast fractions. The separation buffer was 5% (v/v) HOAc. The electrospray sheath liquid was 10% (v/v) methanol and 0.5% (v/v) FA. About 20 nL standard protein solution or 120 nL (240 nL for fractions 2-8) yeast fraction solution was injected for each CE-MS experiment. The separation voltages were 18 kV and 24 kV for standard proteins and yeast fractions, respectively. The electrospray voltages were 1.5 kV and 1.8 kV for standard proteins and yeast fractions, respectively.

The inlet ion transfer tube was held at 300°C and the S-lens rf level was set at 60%. For standard proteins, full MS scans were acquired in the Orbitrap over the m/z 400-1800 range with a resolving power of 60,000 at m/z 200. The three most intense peaks with charge state ≥5 were selected in data-dependent fashion for fragmentation. Detection for all tandem mass spectra was performed in the Orbitrap with a resolving power of 60,000 at m/z 200. The MS1 AGC target value was 1,000,000 with a maximum injection time of 100 ms, while MS/MS scans have an AGC target value of 500,000 and a maximum injection time of 200 ms. Seven and three microscans were used in MS1 and MS2 scans, respectively. An exclusion window of ±10 ppm was constructed around the monoisotopic peak of each selected precursor for 5 seconds. For yeast fractions, most of the mass spectrometer parameters were the same and the exceptions were listed below. Full MS scans were acquired over the m/z 600-2000 range. MS/MS scans have an AGC
target value of 100,000 and a maximum injection time of 300 ms. Dynamic exclusion
time was set at 20 seconds.

5.2.5 Data analysis

The tandem spectra were decharged and deisotoped by MS-Deconv (version
0.8.0.7370),\textsuperscript{143} followed by database searching with TopPIC software (version
0.9.1).\textsuperscript{144} Raw files from Q Exactive HF were first converted to mzXML files with ReAdW (version
4.3.1). Then, MS-Deconv (v 0.8.0.7370) was used to generate msalign files with mzXML
files as the input. Finally, TopPIC software
(http://proteomics.informatics.iupui.edu/software/toppic/) was used for database
searching with msalign files as the input. Uniprot protein database for yeast (reviewed,
23,525 entries) was used for database searching. The parameters for database searching
included N-terminal variable PTM as methionine exclusion and acetylation, number of
the unexpected PTMs as 2, mass error tolerance as 10 ppm, cysteine protecting group as
carbamidomethylation, cutoff type as EVALUE, and cutoff value as 0.001. The results of
23 fractions were combined manually and all the proteoforms with mass shift of 1 Da
were counted as the same proteoforms.

5.3 Results and discussion

5.3.1 Q Exactive HF mass spectrometer intact protein mode optimization

The Q Exactive HF mass spectrometer combines a state-of-art segmented
quadrupole with a high-resolution ultra-field Orbitrap mass analyzer. An intact protein
MS mode is included in the instrument setting; this setting is designed for analysis of
intact proteins with optimized pressure and ion optic parameters. Besides turning on/off the intact protein mode, the trapping gas pressure can be manipulated; the optimized value is 1.0 for non-intact and 0.2 for intact protein modes. In this experiment, I first investigated the effect of this intact protein mode and the trapping gas pressure for intact protein detection, fragmentation and identification. Figure 5.1 A shows the extracted ion electropherogram for a seven-protein mixture detected under intact protein mode with trapping pressure of 0.2. All seven proteins were separated with reasonable peak shape in 20 min. Compared to other settings, intact protein mode with trapping pressure of 0.2 detected proteins with significantly higher signal intensities for all seven proteins except myoglobin which showed comparable signal intensity to the result of intact mode with 0.3 trapping pressure (Figure 5.1 B).

The inserted spectra in Figure 5.1A are butterfly plots of the averaged MS1 spectra for each protein, with the intact mode spectra on top and the non-intact mode spectra on the bottom. With the intact protein mode, spectra with higher charge ions were generated. This effect is increasingly significant as the size of the protein increases. For example, the signal of higher charged ions was only slightly enhanced for ubiquitin and insulin, and was moderately enhanced for cytochrome c, myoglobin and superoxide dismutase. When the protein sizes were greater than 20 kDa (i.e. carbonic anhydrase and β-casein), the signal intensities for higher charged ions as well as the overall signal intensity were significantly enhanced. The fragmentation efficiency for higher charged ions is generally better than lower charged ions due to their higher charge densities, and thus resulted in higher quality tandem spectra and improved identification rate.
To further evaluate the intact protein mode, the protein identification rate for both intact protein and non-intact mode were evaluated. First, the normalized collision energy (NCE) was optimized for the best identification results. NCE of 15, 20 and 25% were applied under intact protein mode and as listed in Table 5.1, where 20% NCE produced the most identifications. Six out of seven proteins were identified with intact protein mode and 20% NCE. Table 5.1 also lists the identification results for non-intact mode with 20% NCE, where only three proteins were identified. Although ubiquitin was not identified by tandem spectra either in intact mode or non-intact mode, it can be easily identified based on its parent ion mass. Similarly, all proteins were identified by their masses even though some of them were not identified by tandem spectra. Therefore, intact protein mode with NCE of 20% was employed in the following characterization of yeast fractions.

5.3.2 RPLC-CZE-ESI-MS/MS

By employing dynamic pH junction and LPA-coated capillary, I not only obtained a high peak capacity for CZE but also generated high protein signals to ensure quality tandem spectra for identifications. The average peak capacity for CZE separation was about 50 (excluding 6 fractions in which no useful electrophoretic peaks were generated), and the average separation window was ~20 min. The average base peak signal intensity for 23 fractions was about $3 \times 10^7$. The best identification result was generated from fraction 14, which identified 180 proteoforms within a 40 min CZE-ESI-MS/MS run and increased the number of total identifications by 137 (Table 5.2),
demonstrating good performances for both the CZE-ESI-MS/MS system and RPLC fractionation. The peak capacity was about 75, and the total separation window was about 20 min, generating over four identifications per minute for this run.

In total, 580 proteoforms and 180 protein groups were identified from 23 CZE-ESI-MS/MS runs of the fractionated proteome. The combined identified protein groups and the masses of their proteoforms are listed in the supporting materials. C5 RPLC column provides efficient fractionation, so the overlaps were reasonably small between fractions as shown in Figure 5.2 and Table 5.2. All fractions (except fraction 3 with only one identification) contributed to unique identifications when combined with the previous fractions. Later fractions generated more identifications compared to the earlier ones. One possible explanation is that the earlier fractions may contain more impurities that resulted in the large peaks in the electropherograms (data not shown), which suppressed the protein signal and possibly reduced the separation efficiency of CZE. The increases in total identifications were less for the last few fractions compared to the middle fractions, indicating that the number of identifications would not be improved if more fractions were generated in the same dimension. A second dimension of fractionation (i.e. size-based separation) before CZE separation would help further improve the identification result. Despite the limited number of proteoforms being identified from tandem spectra, the number of detected species from MS1 spectra was much larger. A total of 3243 species with masses greater than 5 kDa were detected from the 23 fractions. Therefore, there is potential for CZE-ESI-MS/MS to identify more proteoforms given a higher resolution and sensitivity mass spectrometer.
When examining the molecular weight (MW) distribution of the identifications, I found a MW bias towards the low MW region, Figure 5.3 A. Most of the proteoforms identified were small, truncated proteins (~5 kDa), which were preferentially identified by the Orbitrap detector and more effectively fragmented. Only 1/10 of the proteoforms have MW close to or larger than 10 kDa. This result was not surprising because the Orbitrap generally shows decreased sensitivity for larger proteins that have more charge states and isotopic peaks, which resulted in poorly resolved isotopic envelope and low quality tandem spectra. For example, higher concentrations were used for larger proteins such as β-casein in the standard protein experiment and signal intensities of $10^7$ were reached for efficient fragmentations. In contrast, I found that larger MW species in the yeast sample typically have signals as low as $10^5$, which decreases successful deconvolution and tandem spectra identification. There were more large mass species detected in MS1 but not identified with tandem spectra, Figure 5.3 A.

As a result, incorporating a size-based fractionation dimension prior to RPLC fractionation would greatly improve the number of identifications by concentrating large proteins and optimizing fragmentation conditions for narrower mass ranges.

The integrated protein abundances listed in PaxDb (http://pax-db.org/) were used to evaluate the protein abundance distribution of identified protein groups, Figure 5.3 B. All the protein groups identified fell in the middle- to high-region of the yeast total protein abundance distribution. Most of them have abundance at around $10^3$ ppm, and the total identifications spanned a range from 10 ppm to $10^5$ ppm, indicating that
our CZE-ESI-MS/MS system is able to identify a reasonably wide dynamic range in a biological sample.

The identification of proteoforms and PTMs are the most valuable advantages of top-down proteomics. In this study, an average of three proteoforms was identified for each protein group. PTMs including N-terminal acetylation, signal peptide removal, and oxidation were identified for a number of proteins. Since the protein sample were reduced and alkylated, the fixed modification, carbamidomethylation, was also successfully identified on cysteine residues.

5.4 Conclusion

I used reversed-phase liquid chromatography to separate the yeast proteome into 23 fractions. These fractions were then analyzed using capillary zone electrophoresis (CZE) coupled to a Q Exactive HF mass spectrometer using an electrokinetically pumped sheath flow interface. The parameters of the mass spectrometer were first optimized for top-down proteomics using a mixture of seven model proteins; I observed that intact protein mode with trapping pressure of 0.2 and normalized collision energy of 20% produced the highest intact protein signals and most protein identifications. Then, I applied the optimized parameters for analysis of the fractionated yeast proteome. 580 proteoforms and 180 protein groups were identified via database searching of the MS/MS spectra. This number of proteoform identifications is two times larger than previous CZE-MS/MS studies. An additional 3,243 protein species were detected based on the parent ion spectra. Post-translational modifications
including N-terminal acetylation, signal peptide removal, and oxidation were identified.
Figure 5.1: A. Extracted ion electropherogram for the seven-protein mixture. The signal for each protein was amplified to the same height as shown in the graph. The inserted graphs represent the averaged mass spectra for each protein. Top spectrum: intact mode with pressure at 0.2. Bottom spectrum: non-intact mode. B. The signal intensities for each protein with different instrument settings. Non-intact mode with trapping gas pressure at 1.0 and intact protein mode with trapping gas pressure at 0.1, 0.2, 0.3, and 0.5 were listed here. Duplicate runs were performed for all conditions, except non-intact mode where triplicate runs were performed.
Figure 5.2: Summary of the identification results from CZE-ESI-MS/MS top-down analysis of 23 fractions isolated from the yeast proteome. The number of mass species detected only includes the mass species greater than 5 kDa. The number of total identifications is the cumulative number of produced by combining the new fraction to all the previous ones.
Figure 5.3: Summary of identifications. (A) Molecular weight distributions of yeast proteoforms identifications and mass species detected (> 5 kDa) in MS1 combined from 23 CZE-ESI-MS/MS experiments. (B) Protein abundance distributions of identified protein groups. X-axis is plotted on a log_{10} scale.
TABLE 5.1

THE STANDARD PROTEIN IDENTIFICATIONS WITH DIFFERENT MODES AND NCE VALUES SHOWN IN THE BRACKET. NON-INTACT MODE WAS PERFORMED WITH NCE OF 20%. *

<table>
<thead>
<tr>
<th></th>
<th>Intact Mode (20%)</th>
<th>Intact Mode (15%)</th>
<th>Intact Mode (25%)</th>
<th>Non-intact Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cytochrome C</td>
<td>○</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Ubiquitin</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Myoglobin</td>
<td>○</td>
<td>○</td>
<td>×</td>
<td>○</td>
</tr>
<tr>
<td>Carbonic Anhydrase</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Superoxide Dismutase</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>×</td>
</tr>
<tr>
<td>Insulin</td>
<td>○</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>β-casein</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>

*“○” means that the protein was identified and “×” means that the protein was not identified.
TABLE 5.2
MASS SPECIES DETECTED FROM MS1 SPECTRA (> 5 KDA), AND FRACTION AND TOTAL PROTEOFORMS IDENTIFICATIONS FOR EACH FRACTION CORRESPONDING TO THE VALUES SHOWN IN FIGURE 5.2

<table>
<thead>
<tr>
<th>RPLC Fraction Number</th>
<th>CZE Fraction Number</th>
<th>Number of Mass Species</th>
<th>Fraction Identifications</th>
<th>Total Identifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>1, 24</td>
<td>1</td>
<td>334</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>2, 25</td>
<td>2</td>
<td>769</td>
<td>8</td>
<td>22</td>
</tr>
<tr>
<td>3, 26</td>
<td>3</td>
<td>100</td>
<td>1</td>
<td>22</td>
</tr>
<tr>
<td>4, 27</td>
<td>4</td>
<td>202</td>
<td>8</td>
<td>26</td>
</tr>
<tr>
<td>5, 28</td>
<td>5</td>
<td>247</td>
<td>32</td>
<td>49</td>
</tr>
<tr>
<td>6, 29</td>
<td>6</td>
<td>106</td>
<td>2</td>
<td>50</td>
</tr>
<tr>
<td>7, 30</td>
<td>7</td>
<td>100</td>
<td>2</td>
<td>51</td>
</tr>
<tr>
<td>8, 31</td>
<td>8</td>
<td>47</td>
<td>1</td>
<td>52</td>
</tr>
<tr>
<td>9, 32</td>
<td>9</td>
<td>138</td>
<td>19</td>
<td>70</td>
</tr>
<tr>
<td>10, 33</td>
<td>10</td>
<td>77</td>
<td>13</td>
<td>76</td>
</tr>
<tr>
<td>11, 34</td>
<td>11</td>
<td>89</td>
<td>65</td>
<td>133</td>
</tr>
<tr>
<td>12, 35</td>
<td>12</td>
<td>106</td>
<td>89</td>
<td>194</td>
</tr>
<tr>
<td>13, 36</td>
<td>13</td>
<td>83</td>
<td>33</td>
<td>210</td>
</tr>
<tr>
<td>14, 37</td>
<td>14</td>
<td>307</td>
<td>181</td>
<td>347</td>
</tr>
<tr>
<td>15, 38</td>
<td>15</td>
<td>144</td>
<td>42</td>
<td>368</td>
</tr>
<tr>
<td>16, 39</td>
<td>16</td>
<td>213</td>
<td>151</td>
<td>471</td>
</tr>
<tr>
<td>17, 40</td>
<td>17</td>
<td>154</td>
<td>23</td>
<td>478</td>
</tr>
<tr>
<td>18, 41</td>
<td>18</td>
<td>196</td>
<td>24</td>
<td>492</td>
</tr>
<tr>
<td>19, 42</td>
<td>19</td>
<td>410</td>
<td>48</td>
<td>521</td>
</tr>
<tr>
<td>20, 43</td>
<td>20</td>
<td>397</td>
<td>64</td>
<td>550</td>
</tr>
<tr>
<td>21, 44</td>
<td>21</td>
<td>227</td>
<td>22</td>
<td>558</td>
</tr>
<tr>
<td>22, 45</td>
<td>22</td>
<td>265</td>
<td>35</td>
<td>570</td>
</tr>
<tr>
<td>23, 46</td>
<td>23</td>
<td>273</td>
<td>18</td>
<td>580</td>
</tr>
</tbody>
</table>

Total Number of Mass Species Detected 3243
Total Number of Protein Groups 180
CHAPTER 6:
CONCLUSIONS AND FUTURE WORK

With the inherent advantage of CZE separating proteins and their isoforms, CZE has become increasingly popular in top-down proteomics. In my study, I successfully employed the sheath-flow CZE-ESI-MS system to characterize proteins including \textit{M. marinum} secretome, monoclonal antibodies and yeast proteome. I demonstrated that CE-MS can be not only applied to single protein analysis, but also used for large scale proteomics studies.

Although CE-MS has identified proteins in a proteomic scale, the number of protein identifications, compared to RPLC-MS, is still limited. This performance can be attributed to the combination of both CE technique itself and mass spectrometry. First, the sample amount that can be injected in the CE system is low compared to LC, even with online preconcentrating methods, such as tIPT and pH junction. Microliters of sample is typically injected in LC while only nanoliters can be injected in CE. Given the lower sensitivity for proteins, it is challenging for a mass spectrometer to identify low abundance species and modifications. Second, long gradient can be applied in LC to separate very complex sample. However, employing longer capillary in CE increases the separation window, but also broadens peaks, which results in little or no improvement in resolution. With these limitations, CE is more effective for separation of relatively
simple mixtures, and is especially useful for therapeutic protein characterizations. When analyzing complex samples, extensive fractionation should be employed. For example, in Chapter 5, I described using RPLC as the fractionation method for yeast, which yielded 580 proteome identifications. Finally, the limitation in sample injection volume and resolution for complex sample leads to higher requirement for mass spectrometers. Theoretically, if mass spectrometry has sufficient sensitivity, scan speed and resolution, CE would have performed much better because there is no loss of proteins during the separation.

In order to further expand the application of CE-MS for protein characterization, multi-stage prefractionation needs to be involved for complex proteome analysis. For example, SEC can be performed followed by RPLC fractionation. SEC separates proteins by size, so the MS parameters can be optimized for specific size range. RPLC separates proteins by hydrophobicity, and also serves as a desalting method prior to CE separation. On the other hand, the analysis of a single protein is important for the therapeutic protein development. CE-MS is a promising platform for determining their heterogeneity. Similar to the reduced mAb analysis described in Chapter 4, the characterization can be performed on intact or even native state mAbs to determine the modifications and sequence variants.
REFERENCES


(45) Giddings, J. C. Analytical Chemistry 1984, 56, 1258.


(54) Jorgenson, J. W.; Lukacs, K. D. Science 1983, 222, 266.


(147) Qiao, Y.; Di, Z. G.; Ma, Y. G.; Ma, P. S.; Xia, S. Q. Chinese Journal of Chemical Engineering 2010, 18, 446.


(160) Gregorich, Z. R.; Ge, Y. Proteomics 2014, 14, 1195.


Horn, D. M.; Ge, Y.; McLafferty, F. W. Anal Chem 2000, 72, 4778.


Xia, Y.; Han, H.; McLuckey, S. A. Analytical Chemistry 2008, 80, 1111.


