State-of-art Separation Mass Spectrometry in Protein Characterization

by Siyuan Liu

B.S. in Chemistry, Wuhan University

A dissertation submitted to

The Faculty of
the College of Science of
Northeastern University
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

March 30, 2016

Dissertation directed by

Barry L. Karger
Professor of Chemistry and Chemical Biology
Acknowledgments

Completing my doctoral work has been a long journey, and I would not have been able to achieve this milestone without the support from many people. I would like to take this opportunity to express my gratitude to those who had helped me over the past five years.

I would like to give my deepest acknowledgment to my advisor, Prof. Barry L. Karger, for his intellectual contributions in my doctoral studies over these years. I would like to sincerely thank Dr. Karger for accepting me as a graduate student in his group, providing excellent research resources in the Barnett Institute, and offering me a valuable chance to work on these exciting projects that were extremely helpful to gain experience for my professional career. His knowledge, brilliant insight and tremendous enthusiasm always inspired me to seek for the highest standards on the doctoral work and also deeply inspired me beyond my research. Without his guidance, I would not grow from a student to a scientist.

Next, I am highly indebted to Dr. Shiaw-Lin (Billy) Wu and Dr. Alexander R. Ivanov, for their supervision and thoughtful guidance on my research projects. I gratefully appreciate Dr. Wu, who worked with me closely on the two projects. His expertise in biotechnology benefited me greatly both in scientific research and career development. My special thanks also goes to Dr. Ivanov. His solid knowledge, invaluable advice and detailed discussion on the CIEF-MS project helped me work effectively in the last year of my Ph.D. studies.

I would like give sincere acknowledgment to my committee members: Dr. William S. Hancock, Dr. Jeffrey N. Agar and Dr. Roger W. Giese. I appreciate their time and valuable suggestions on my dissertation and defense.
In addition, I also would like to acknowledge the colleagues from the Barnett Institute who had accompanied and helped me throughout these years. My thanks go to those former and current members both from Dr. Karger’s and Dr. Hancock’s group: Dr. Shujia (Daniel) Dai, Dr. David R. Bush, Dr. Zhenke Liu, Dr. Wenqin Ni, Dr. Chen Li, Dr. Siyang Li, Dr. Yue Zhang, Dr. Fan Zhang, Yuanwei Gao, Xianzhe Wang, Arseniy M. Belov, Simion Kreimer, Somak Ray, Victoria Berger, Yu Wang, Di Wu and Yanjun Liu.

I would also acknowledge the following faculty and staff from Northeastern University: Dr. Paul Vouros, Nancy Carbone, Richard Pumphrey, Andrew Bean, Emanuelle Hestermann, Cara Shockley and Jean Harris.

Thanks to my grandparents on my mother’s side, Mr. Changfu Tian and Ms. Xiumei Wang and other family members. Their unconditional love and continuous support made me who I am today. I am so fortunate to have them in my life and have them witness my growth.

Finally, I would like to dedicate this dissertation to the memory of my beloved grandparents on my father’s side, Mr. Jingzhong Liu and Ms. Daohua Zhang, the two persons whom I still miss every day.
Abstract of Dissertation

A therapeutic protein derived from recombinant DNA technology usually poses a high degree of complexity. A complete heterogeneity profile of the therapeutic protein is critical to ensure the quality, safety and efficacy of the drug. To achieve a complete mapping of all possible protein modifications including the various types of post-translational modifications and sequence variants, the incorporation of high resolution separation methods and high sensitivity detection techniques are required. The purpose of this dissertation is to develop and apply sensitive and robust mass spectrometry (MS)-based analytical platforms to address the challenges associated with protein identification and characterization.

In the first chapter, the current analytical techniques, including various gel-based, chromatographic-based and capillary electrophoresis-based separation methods frequently incorporated with mass spectrometry for protein identification and characterization are reviewed. The traditional sample preparation workflows for MS-based protein identification are introduced. In addition, the application of mass spectrometry for the determination of protein post-translational modifications and the associated analytical challenges are discussed.

The sample preparation steps involved in the LC-MS analysis of a therapeutic protein can potentially introduce disulfide scrambling leading to the incorrect characterization of the disulfide linkage structure. Chapter 2 focuses on the development of an LC-MS work flow for the determination of the disulfide mapping of proteins that are sensitive to disulfide scrambling. The recombinant human neublastin, a potential therapeutic agent with disulfide linkages partially elucidated via the traditional non-reduced multi-enzyme digestion strategy is studied. The modified workflow includes the optimization of the partial reduction and alkylation steps prior to
multi-enzyme digestion in order to minimize the breakage of a relatively unstable intermolecular disulfide bond and the rapid thiol-disulfide exchange between adjacent cysteines. Using the developed work flow, while scrambling remained, complete and reproducible disulfide mapping was achieved via in-depth LC-MS analysis with CID and ETD fragmentations.

In Chapter 3, a powerful methodology combining metabolic labeling, SDS-PAGE protein fractionation and nano LC-MS analysis is applied to isolate and identify the potential protein target to a novel small molecule drug from cell lysates. The non-covalent binding between the identified protein target and the drug is demonstrated with comprehensive characterization of the potential target.

In Chapter 4, a new design of a low-sheath-flow liquid junction interface is described to integrate commercial capillary electrophoretic instrumentation and electrospray ionization mass spectrometer for automatic peptide and protein analysis. Different separation modes including capillary zone electrophoresis (CZE) and capillary isoelectric focusing (CIEF) were applied to couple with mass spectrometry via the liquid junction interface. Confident identification and high separation efficiency of a BSA digest are demonstrated with the integrated CZE-MS system. In addition, separation of a mixture of model proteins with closely related isoelectric points was achieved with CIEF-MS analysis leading to the identification of different types of post-translational modifications. The system was also employed to study the potential degradation products of a therapeutic protein, recombinant human growth hormone. The low-sheath-flow liquid junction interface integrated CIEF-ESI-MS system potentially can be an alternative technique to the well-established LC-MS method for the online enrichment and identification of the charge heterogeneity in biopharmaceuticals.
Chapter 5 summarizes the work in the dissertation and provides a brief discussion on the future direction of each project.
# Table of Contents

Acknowledgements ................................................................................................................... ii

Abstract of Dissertation ........................................................................................................ iv

Table of Contents .................................................................................................................... vii

List of Figures ........................................................................................................................ x

List of Tables .......................................................................................................................... xii

Chapter 1: Overview of Protein Characterization by Mass Spectrometry-Based Methods and Related Analytical Techniques ................................................................. 1

1.1 Abstract ............................................................................................................................ 2

1.2 Introduction to Protein Characterization ........................................................................ 2

1.2.1 Significance of Protein Characterization in Biopharmaceuticals ................................ 3

1.2.2 Introduction of MS-based methods for Protein Characterization ............................ 6

1.3 Protein Structure .............................................................................................................. 8

1.3.1 Post-Translational Modifications ............................................................................. 10

1.3.2 Common Post-Translational Modifications ............................................................. 11

1.4 Protein and Peptide Separation Methods ................................................................... 18

1.4.1 Immunoprecipitation (IP) ...................................................................................... 19

1.4.2 Gel-Based Methods ............................................................................................... 20

1.4.3 Chromatography-Based Methods ....................................................................... 22

1.4.4 Capillary Electrophoresis-Based Methods ............................................................. 26

1.5 Mass Spectrometry ......................................................................................................... 34

1.5.1 Electrospray Ionization ......................................................................................... 34

1.5.2 Mass Analyzer ....................................................................................................... 36

1.5.3 Fragmentation ........................................................................................................ 41

1.5.4 Data Processing ...................................................................................................... 46

1.6 Application of Mass Spectrometry-Based Methods for Protein Characterization ...... 47

1.6.1 Sample Preparation for MS-based Characterization ............................................. 47

1.6.2 Applications of MS-Based Methods in PTM Analyses ......................................... 52

1.6.3 Challenges in PTM Determinations with MS-Based Methods ............................... 57

1.7 Summary ......................................................................................................................... 58
# Chapter 2: Complete Disulfide Mapping of Neublastin by Online LC-MS/MS Analysis following Partial Reduction and Alkylation

## 2.1 Abstract

## 2.2 Introduction

## 2.3 Experimental

### 2.3.1 Chemical and Reagents

### 2.3.2 Sample Preparation

### 2.3.3 LC-MS and LC-UV

## 2.4 Results and Discussion

### 2.4.1 Non-reduced Digestion Approach

### 2.4.2 Partial Reduction and Alkylation

## 2.5 Conclusion

## 2.6 Acknowledgments

## 2.7 References

---

# Chapter 3: Chemical Proteomic-Based Approach for Identification and Characterization of a Drug Target (Protein) Using Mass Spectrometry

## 3.1 Abstract

## 3.2 Introduction

## 3.3 Experimental

### 3.3.1 Chemicals and reagents

### 3.3.2 Samples

### 3.3.3 SDS-PAGE

### 3.3.4 Enzymatic digestion

### 3.3.5 LC-MS

### 3.3.6 Data analysis

## 3.4 Results and Discussion

### 3.4.1 Preparation of heavy and light proteins in cells (SILAC approach)

### 3.4.2 Analysis of Bound Proteins

### 3.4.3 Characterization of PAK4

## 3.5 Conclusion
Chapter 4: Integration of Capillary Isoelectric Focusing with Electrospray Ionization Mass Spectrometry using Liquid Junction Interface for Intact Protein Analysis

4.1 Abstract .............................................................................................................. 139
4.2 Introduction ........................................................................................................ 140
4.3 Experimental ..................................................................................................... 144
  4.3.1 Chemicals and Materials ............................................................................. 144
  4.3.2 Sample Preparation ................................................................................... 145
  4.3.3 Instrumentation ......................................................................................... 146
  4.3.4 Data Analysis .............................................................................................. 148
4.4 Results and Discussion ..................................................................................... 149
  4.4.1 Design rationale of the interface ................................................................. 149
  4.4.2 Configuration of the interface ................................................................... 152
  4.4.3 Operation and optimization of the interface ............................................. 154
  4.4.4 Evaluation of interface performance by CZE-MS analysis ..................... 159
  4.4.5 Online CIEF-MS with liquid junction interface for intact protein analysis.. 168
4.5 Conclusions ....................................................................................................... 197
4.6 References ......................................................................................................... 198
List of Figures

Figure 1-1. A general workflow of the MS-based the bottom-up approach for protein characterization.................................................................................................................. 10
Figure 1-2. Mechanism of thiol/disulfide exchange reaction. ......................................................... 12
Figure 1-3. Core structures of different types of glycans. .............................................................. 15
Figure 1-4. Mechanism of (A) asparagine deamidation and (B) aspartic acid isomerization .... 17
Figure 1-5. Scheme of the electric double layer formed inside a bare fused silica capillary.. .... 26
Figure 1-6. Schematic setup of three types of CE-MS interfaces .................................................. 32
Figure 1-7. Nomenclature of fragment ions for a peptide............................................................... 42
Figure 1-8. Scheme of CID fragmentation.......................................................................................... 44
Figure 2-1. Pictorial representation of the three-dimensional structure of the NBN disulfide linkages. ......................................................................................................................................... 75
Figure 2-2. Theoretical Asp-N + trypsin digestion of NBN. ............................................................. 75
Figure 2-3. Assignment of disulfide linkages in a scrambled peptide when a non-reduced digestion of NBN with Asp-N and trypsin was used.................................................................................. 85
Figure 2-4. Assignment of the disulfide linkage in a scrambled peptide with the two adjacent cysteines, C71 and C72........................................................................................................................................ 86
Figure 2-5. Optimization of conditions for partial reduction and alkylation................................. 92
Figure 2-6. Comparison of the peptide map of NBN digested with Asp-N and trypsin with (A) partial and (B) complete reduction with TCEP......................................................................................................... 96
Figure 2-7. Assignment of the NEM modification site on peptide P66 – R76 from the completely reduced digest of NBN......................................................................................................................... 97
Figure 2-8. Assignment of disulfide linkage C7-C72 with NEM alkylation on C71 ................. 99
Figure 2-9. CID-MS2 of the peptide with one disulfide linkage (Cys7 – Cys72) and one NEM modification (on Cys71). ............................................................................................................................................. 100
Figure 2-10. Assignment of disulfide linkages C34 - C100 and C38 - C102 for non-reduced NBN digested with Asp-N and trypsin.................................................................................................................. 102
Figure 2-11. Assignment of disulfide linkages C71-C71’ and C7-C72 ............................................. 104
Figure 3-1. Workflow of the MS-based SILAC approach for target identification ....................... 122
Figure 3-2. Histogram of proteins identified in forward and reverse order (reciprocal) experiments displayed by log2 SILAC ratios. ................................................................. 125

Figure 3-3. Scatter plot of log2 SILAC ratios in forward and reverse replicate experiments obtained from (A) In-gel digestion and (B) on-beads digestion approach. ......................... 126

Figure 3-4. Scheme of Michael addition of the side chains of Cys, His and Lys with 4-oxo-2-nonenal ........................................................................................................... 128

Figure 3-5. SDS-PAGE of PAK4 treated by solvent and SM drug. ........................................ 130

Figure 3-6. Summary of the identification of the peptide sequence of target protein PAK4. ... 130

Figure 3-7. Base peak chromatogram of (A) PAK4 and (B) drug KPT-7523 treated PAK4 .... 131

Figure 3-8. Identification of a Cys containing peptide in DMSO treated (top) and SM treated PAK4 (bottom). ........................................................................................................ 131

Figure 4-1. Direct infusion of myoglobin with different concentrations of ampholytes (AMF) at a flow rate of 50 nL/min ........................................................................................................ 150

Figure 4-2. Scheme of liquid junction interface. .................................................................... 154

Figure 4-3. Microphotos of three different shapes of capillary termini in the electrospray tips 156

Figure 4-4. CZE-MS of BSA digest. .......................................................................................... 161

Figure 4-5. Base peak electropherograms of BSA digest in triplicate CZE-MS runs. .......... 162

Figure 4-6. CZE-MS of four standard proteins with different concentrations of glycerol in the BGEs. .................................................................................................................. 167

Figure 4-7. Scheme of CIEF-MS process with the liquid junction interface......................... 169

Figure 4-8. CIEF-MS with 1 % (v/v) Pharmalyte 3-10 and Aeslytes SH3-10 using pressure mobilization. .............................................................................................................. 174

Figure 4-9. CIEF-MS with 0.5 % (v/v) Fluka 3-10 and Servalyt 3-10 using chemical mobilization. ................................................................................................................ 176

Figure 4-10. CIEF-MS with different mobilization methods. ................................................. 180

Figure 4-11. Summed ion electropherograms of the model proteins analyzed by CIEF-MS system in triplicate runs. .................................................................................... 184

Figure 4-12. CIEF-MS analysis of the model protein mixture using 30% glycerol in the separation medium. ................................................................................................. 187

Figure 4-13. Identification of post-translational modifications on Rnase .............................. 190

Figure 4-14. Identification of modifications on rhGH......................................................... 195
List of Tables

Table 2-1. Identity of major components in the four RPLC fractions. ........................................... 93
Table 2-2. Reproducibility peptides with expected disulfide linkages ........................................... 106
Table 2-3. Reproducibility peptides with scrambled disulfide linkages. ........................................... 106
Table 2-4. Reproducibility peptides with 2 NEM modifications. ..................................................... 107
Table 3-1. Differentially expressed proteins overlapped in both in-gel and on-beads digestion approach.......................................................................................................................... 127
Table 3-2. Summary of identified peptides in drug treated target protein PAK4. ......................... 132
Table 3-3. Summary of Cys containing peptides in DMSO (C) and drug treated PAK4 (S) ..... 134
Table 4-1. Repeatability in migration time measured in triplicate analysis of BSA digest. ...... 163
Table 4-2. Repeatability in migration time measured in triplicate CIEF-MS runs ...................... 185
Table 4-3. Summary of Identified Protein and Potential Protein Modifications. ....................... 191
Table 4-4. Summary of Identified Proteoforms in stressed rhGH and Potential Protein Modifications. ................................................................................................................................. 196
Chapter 1: Overview of Protein Characterization by Mass Spectrometry-Based Methods and Related Analytical Techniques
1.1  ABSTRACT

This chapter provides a general introduction to the background of MS-based methods for characterization of therapeutic proteins. First, the significance of protein characterization in biopharmaceutics is discussed. Secondly, major gel-based, HPLC-based and CE-based separation techniques, along with various MS detection methods are introduced. Finally, the application of MS-based methods for the determination of protein post-translational modifications is presented.

1.2  INTRODUCTION TO PROTEIN CHARACTERIZATION

Proteins are one of the most important class of biological molecules for all living systems, from unicellular bacteria to complex multicellular mammals. As the major component of cells, proteins occupy more than half the dry weight of a cell. Proteins play a critical role in forming structural components in a cell (e.g., membrane channels). Moreover, they participate in and regulate every aspect of biological activity, including DNA replication, oxygen transport and cell apoptosis\(^1\).

Many diseases originate from deficient, abnormal or structural alteration of specific proteins in the body. Understanding the structures of proteins, their interactions with other biomolecules, and their roles in biological pathways is essential in order to elucidate the mechanisms of diseases and, moreover, to develop new drugs to prevent, treat or cure diseases.
1.2.1 Significance of Protein Characterization in Biopharmaceuticals

In recent years, protein therapeutics, classified as monoclonal antibodies and therapeutic proteins, have shown significant commercial and clinical success and constitute an important class of modern pharmaceutics. Since the advent of recombinant DNA technologies and the approval of the first recombinant product by the Food and Drug Administration (FDA) in 1982, human insulin (trade name: Humulin, created by Genentech), the landscape of the pharmaceutical industry has been changed. Historically, small molecule drugs have dominated in the marketplace. However, protein therapeutics have represented fast growth and large marketing potential. From 2007 to 2013, the global biopharmaceutical market value increased from $94 billion to $140 billion, and it has been reported that the top-selling antibody drug in 2013 was the mAb-based drug, Humira (adalimumab), which has reached a global sale of $11 billion.

Compared to traditional small synthetic drugs, protein therapeutics offer some distinct advantages in terms of higher specificity, less interference with normal biological processes, larger immune tolerance, as well as improved long-term patent protection. Nevertheless, the structural complexity of protein therapeutics is much greater than that of small molecule drugs. Typically, a small molecule drug has a well-defined molecular weight (MW) between 100 to 1,000 Daltons (Da). It is a chemically synthesized, highly purified single molecule entity with a well-characterized chemical structure and very limited impurities. In contrast, protein therapeutics have larger molecular weights, varying from 18 to well over 150 kDa. A protein drug is never a single molecule entity; instead, it is made up of a population of protein molecules with similar but not identical structures. Nearly every step of the manufacturing process, including fermentation, purification, formulation and storage, could introduce variations into the
protein structure and result in heterogeneity of the final product. The major sources of protein heterogeneity are host cell protein (HCP) contaminants, post-translational modifications (PTMs) introduced during fermentation, and other modifications that may occur during the manufacturing process.

**Host Cell Proteins**

A limited number of protein therapeutics are extracted from their native sources. Most protein therapeutics instead are derived from living systems such as *Escherichia coli* (*E. coli*) bacteria, Chinese hamster ovary (CHO) cells, or mouse myeloma cells (NS0) with recombinant DNA technology. During the downstream purification process of recombinant protein products, small amounts of HCPs non-specifically bound with the protein of interest will inevitably be co-purified and become a source of heterogeneity. It is necessary to determine and quantitate the proteins remaining from the cell used to produce the biopharmaceutical.

**Post-Translational Modification**

PTMs introduce a large variety of structural heterogeneity to protein therapeutics. Most PTMs occur *in vivo*; however, some PTMs are very sensitive to manufacturing conditions and can occur *in vitro*. Common PTMs associated with protein therapeutics, such as disulfide linkages, deamidation, oxidation, and glycosylation will be discussed in section 1.3.

**Other Modifications**

There are other potential sources of protein heterogeneity such as sequence variance introduced by translational or transcriptional errors, and aggregation or degradation generated during manufacture or storage. Most of the production- and process-induced chemical
modifications are undesirable; however, some chemical modifications are introduced intentionally by protein engineering. A case in point is conjugating polyethylene glycol (PEG) for the benefit of reducing immunogenicity and elongating the serum half-life of the protein therapeutic. However, the number of PEG molecules incorporated into the protein molecule via chemical reaction may not be rigorously consistent across batches.

Protein heterogeneity is highly dependent on the production, manufacturing and storage conditions. A slight change in the manufacturing process can have a drastic effect on the final product. For example, the glycan structure of the innovator product Myozyme from Genzyme to treat Pompe disease was substantially altered when scaling up the size of the fermenter 12-fold. A small degree of protein heterogeneity may affect protein function and biological activity, ultimately resulting in adverse effects on drug efficacy, stability and even potentially inducing immunogenicity. Therefore, thorough and comprehensive characterization of the heterogeneity and impurity profile of protein therapeutics at high confidence levels, and rigorous monitoring during the manufacturing process, as well as the final product, are required by the FDA guidelines. In the following sections, the discussion will focus on how protein heterogeneity affects drug efficacy, stability and safety.

As an example, glycosylation heterogeneity not only affects drug efficacy and stability, but it also can give rise to safety risks by causing immunogenicity. Studies have shown that glycosylation is related to important physicochemical properties of a protein such as solubility, secretion, stability, and binding affinity. For example, the effect of different glycosylation structures on the in vivo biological activity of recombinant human erythropoietin (rhEPO) has
been demonstrated by site-direct mutagenesis experiments. Mutations on critical N-linked glycosylation sites not only decrease cellular secretion of EPO from COS1 and CHO cells, but also reduce the *in vivo* biological activity of the protein. A controlled deglycosylation study with different enzymes to selectively remove monosaccharide units indicated that the terminal sialic acids in the oligosaccharide units of EPO were essential to maintain its *in vivo* stability by protecting galactose residues from being recognized by hepatic receptors. Being aware of these and other issues, cell engineering has made great efforts to control the glycosylation heterogeneity and, more specifically, to increase the expression of sialic acid content (e.g. co-expression of Human α2, 3-sialyltransferase and β1, 4-galactosyltransferase), and to knock down enzymes involved in producing immunogenic glycans in the cell line in order to maximize the efficacy, stability and safety of protein therapeutics.

**1.2.2 Introduction of MS-based methods for Protein Characterization**

In-depth structural information of protein sequences and various PTMs not only ensure drug quality, safety and efficacy of protein therapeutics, but also provide insight into the complex cellular regulatory network associated with human disease. Currently, there are a large variety of analytical techniques for protein characterization, including Edman degradation for primary structure sequencing, Western blotting for the determination of PTMs, enzyme-linked immunoassay (ELISA) for host cell protein identification, and X-ray and NMR for protein 3D-structure analysis. Among these techniques, mass spectrometry (MS)-based methods have shown
important advantages in both qualitative identification and quantitative determination of proteins and proteomes because of the superb sensitivity, selectivity and specificity.\textsuperscript{16}

Initially, applications of MS-based methods were limited to small molecule studies because large biomolecules such as proteins had difficulty forming gaseous, charged ions necessary for MS transmission. However, advancements of matrix-assisted laser desorption ionization (MALDI) and electrospray ionization (ESI) in the 1980s, extended the application of MS systems to the analysis of large biomolecules.\textsuperscript{17,18} Moreover, the recent development of novel hybrid MS instrumentation with high resolution and mass accuracy (e.g. LTQ-Orbirap, Q Exactive) and the implementation of different ion activation methods (e.g. CID, ECD/ETD, HCD) as well as the combination of high resolving power from HPLC have enhanced the capability of MS-based methods to analyze proteins at ever-lower abundances and in ever more complex biological matrices.

In MS-based methods, the identification of a protein is based on the mass-to-charge ratio (m/z, units of Thomson, Th) of the precursor ions or product ions. In order to avoid signal suppression of low-abundance species by high-abundance components, sufficient separation prior to MS is necessary. Therefore, proteins or peptides are first separated by HPLC or capillary electrophoresis (CE), followed by desolvation, gas phase ionization and MS detection. In the mass spectrometer, an MS survey scan is first conducted to provide precursor m/z information, followed by isolation of a particular ion for fragmentation. A tandem MS/MS scan is performed to provide information about product ions. In an MS scan, the molecular weight of proteins or peptides is determined according to the m/z. In a tandem MS/MS experiment, by comparing experimental patterns of product ions with theoretical fragmentation patterns deduced from the
protein sequence, the identity of the protein sequence and the site-specific modifications can be determined.

Overall, with the integration of high resolving power from LC/CE to separate complex mixtures and the high sensitivity and high accuracy from MS to detect the mass of an unknown sample, LC-MS has been one of the most powerful tools for the structural analysis of proteins and target identification in large-scale proteomic samples.

1.3 PROTEIN STRUCTURE

Despite the fact that significant structural differences exist for proteins in living systems, all proteins are constructed from 20 amino acid residues. Two amino acid residues are linked together by the reaction between a free carboxyl group and a primary amino group to form a peptide bond. The linear order of amino acid residues from the N-terminus to the C-terminus of a polypeptide chain is referred to as the primary structure of a protein. The amino acid sequence defines the identity of the protein. The same class of proteins expressed by different species may have different primary structures. The specific order of amino acid residues in a polypeptide chain is derived from the gene sequence. However, the primary sequence of a protein cannot be exactly the same as that deduced from the DNA sequence, since transcriptional or translational errors may lead to changes in the amino acid residues on a polypeptide chain.

The local conformation of the polypeptide chain is termed the secondary structure of the protein. Three common secondary structures are the α-helix, β-strand and turns. The secondary structure helps to stabilize protein structures by forming backbone hydrogen bonds. Protein
tertiary structure is the overall spatial arrangement or the three-dimensional shape of a folded polypeptide chain. A polypeptide chain usually folds and twists in order to achieve the most stabilized structure with the lowest energy state. The formation of a stable tertiary structure is a result of disulfide linkages, other covalent linkages, charge-charge interactions, hydrophobic effects, hydrogen bonding and Van der Waals forces. If a protein contains multiple polypeptide chains, the interaction and spatial arrangement of all the polypeptide chains is referred to as the quaternary structure of the protein. Quaternary structures are stabilized by the same interactions as those applicable in tertiary structures.

Numerous analytical techniques are available for protein structural analysis, such as chemical derivatization, circular dichroism (CD) spectroscopy, X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy and MS-based methods. Compared with these methods, mass spectrometry presents advantages in its capabilities of analyzing protein structure in the solution phase with low sample consumption (at or below the femtomolar level). As a result, mass spectrometry has become the method of choice for protein characterization, especially for primary structure identification. Currently, the MS-based bottom-up approach for peptide sequencing is the most widely used method for identification. As shown in Figure 1-1, sequence-specific proteases are used to cut a protein of interest into smaller peptides, and peptide map experiments are performed with LC/MS to accurately measure the mass of each peptide and searched against online databases such as SwissProt for close matching. Tandem MS/MS experiments with different ion activation methods are applied on each digested peptide to generate fragment ions containing detailed sequence information for database searches. Each amino acid residue may be identified several times with different peptides: by overlapping and
reassembling all identified peptides, the primary structure of the protein will be covered, depending on the level of sequence coverage.

![Diagram](image)

**Figure 1-1.** A general workflow of the MS-based the bottom-up approach for protein characterization.

In addition to primary structure identification, mass spectrometry is able to assess the higher order structure of proteins, using the combination of chemical modifications and sequence-specific proteases. For example, hydrogen/deuterium exchange mass spectrometry (HDX MS) is widely used to study protein conformation in the solution phase. In summary, mass spectrometry is a versatile and powerful tool to study protein structure.

### 1.3.1 Post-Translational Modifications

It has been reported that the human genome is composed of up to 25,000 protein-encoding genes; the resulting number of proteins in the human proteome appears to exceed 1 million. The large populations in human proteome result from numerous modifications of proteins, such as alternative splice variants and post-translational modifications. More than 400 distinct types of PTMs have been identified, and each type contributes to a large number of protein species.
After initial translation of mRNA to produce a polypeptide chain, a series of reactions take place to modify the polypeptide chain before the final form is obtained; such modifications are termed post-translational modifications \(^1\). Some of the modifications occur immediately after translation to regulate protein folding or to direct protein transport, while other modifications occur after folding or transport to switch on or off a specific biological activity. PTMs give rise to structural heterogeneities by different mechanisms such as covalent attachment of functional groups to the side chains of amino acid residues (e.g., glycosylation), modification of existing amino acid residues on peptide backbones (e.g., deamidation), removal of specific sequences on polypeptide chains (e.g., proteolytic processing), or even degradation of the entire protein.

Protein PTMs display a large variety of functional diversities. Some PTMs are involved in signaling pathways to assemble complexes, or they may direct protein localization and modulate molecular interaction. Thus, some PTMs are crucial for biological regulation \(^2\), whereas others relate to the quality and stability of proteins and therefore are studied by the biopharmaceutical industry. In this section, some common PTMs such as disulfide linkages, deamidation, oxidation and glycosylation, which are characteristic for protein therapeutics, are discussed.

\[1.3.2 \text{ Common Post-Translational Modifications}\]

\[1.3.2.1 \text{ Disulfide Linkages}\]

Disulfide linkages are common modifications found in many membrane and secreted proteins but less often in the cytoplasmic environment because the reduced cytoplasmic condition does not support the stabilization of disulfide linkages \(^25\). Formation of a disulfide linkage between thiol groups (SH) on the side chain of two cysteine (Cys) residues leads to the loss of two
hydrogen atoms and a 2 Da mass shift. The chemistry in disulfide linkage formation has been extensively studied 26–28.

Disulfide linkage formation is the result of a series of redox reactions mediated by distinct classes of cellular enzymes. In vivo disulfide linkage formation occurs via nucleophilic attack, termed as the thiol/disulfide exchange reaction. The thiolate anion (S⁻) in a protein interacts with a substrate containing either a disulfide linkage or a compound containing a free thiol group (SH) under the catalysis of an oxidoreductase and isomerase, resulting in the formation of a mixed disulfide linkage, as shown in Figure 1-2 29.

![Figure 1-2. Mechanism of thiol/disulfide exchange reaction](image)

However, the thiol/disulfide reaction can also occur in vitro regardless of the catalysis of cellular enzymes. The presence of oxidants such as oxygen and iodine can trigger the redox reaction and lead to disulfide breakage, shuffling and rearrangement (scrambling). Additionally, the pKa value for a thiol group in a Cys residue is approximately 8.5; however, the pKa value can be lower due to the surrounding amino acid residues 1,28. Thus, in the manufacturing process or long-term storage for a biopharmaceutical, alkaline conditions may cause the deprotonation of thiol groups in unpaired Cys residues, initiating thiol/disulfide exchange reactions and consequently leading to disulfide scrambling 28.
The formation of correct disulfide linkages has great biological significance. First, disulfide linkage formation is essential for protein folding. Formation of disulfide linkages is thermodynamically favorable for protein folding since it can reduce the entropy of the unfolded state and stabilize the native conformation. Secondly, disulfide linkages are associated with the catalytic activity of many cellular enzymes, serving as molecular switches for a wealth of biological pathways in an allosteric manner. As an illustration, the immune-cell receptor CD4 protein, the major receptor for HIV-1 and regulator of cell-cell interaction, contains four extracellular domains. It has been found that cleavage of the disulfide linkage in the D2 domain catalyzed by thioredoxin occurring on the T-cell surface can induce the conformational change of CD4 to promote the entry and infection of HIV-1. Thirdly, the formation of disulfide bonds maintains the structural integrity of a protein in response to external stimuli. Disulfide linkages can stabilize the tertiary and quaternary structures of a protein and protect the protein against unfolding, aggregation or degradation. Furthermore, for the biopharmaceutical industry, correct disulfide linkage is associated with the safety and stability of a protein drug. As a result, analytical techniques to identify the status (paired or unpaired) of every cysteine residue and to map the disulfide linkages of a protein are critical. Chapter 2 presents a study on the disulfide linkage analysis of a homodimer protein whose two monomers are linked by a disulfide linkage.
1.3.2.2 Glycosylation

Glycosylation, is one of the most complex co/post-translational modifications in living species. It has been reported that over two thirds of protein entities within the SWISS-PROT database are potentially glycosylated \(^3^3\). Moreover, more than 40% of approved protein therapeutics on the market are glycoproteins \(^2^5\). The complexity of glycosylation stems from the structural heterogeneity of attached glycan moieties in terms of different glycan composition and linkage between monosaccharide units, linear/branched structures and glycosidic binding sites. More specifically, different numbers of monosaccharide units such as glucose (Glc), galactose (Gal), mannose (Man), fucose (Fuc), sialic acid (SA), N-acetylglucosamine (GlcNAc), and N-acetylgalactosamine (GalNAc) can join together with α- or β-linkages to form either linear or branched chains and further attach to either Asn or Ser/Thr residues on a polypeptide chain \(^3^4\). Usually, glycosylations are classified according to specific attachment sites of glycans on a protein, among which two major classes are N-linked glycosylation and O-linked glycosylation.

In N-linked protein glycosylation, a glycan is covalently bound to a nitrogen atom on the side chain of Asn in the consensus motif Asn-Xaa-Ser or Asn-Xaa-Thr, where Xaa stands for any amino acid residue except proline. In contrast, in O-linked protein glycosylation, a glycan is attached to an oxygen atom in Ser or Thr without any consensus sequence. The synthesis of N-linked glycoprotein in humans has been extensively studied and well-established. Briefly, the common core structure attached to Asn is Man\(^3\)GlcNAc\(^2\). Depending on the types and linkages of monosaccharides attached to this core structure, N-linked glycans can be further divided into three categories: high mannose, complex and hybrid type (as shown in Figure 1-3A).
Figure 1-3. Core structures of different types of glycans. (A) Schemes of three types of N-linked glycans. The core structure Man3GlcNAc2 is shown in a dashed square. (B) Some common core structures of O-linked glycans 35.

In the case of O-linked glycans, the first monosaccharide unit is usually GalNAc, which anchors on fully-folded and assembled proteins via covalent interaction with a hydroxyl group of Ser or Thr. Subsequently, various numbers of different monosaccharide units such as Fuc, SA, and Man further extend or terminate the core structure of O-linked glycans (see Figure 1-3B) 36. Besides Ser and Thr residues, O-linked glycosylation can also occur on hydroxylysine and hydroxyproline residues 37. Generally, the glycan structures in O-linked glycoproteins are simpler than those in N-linked glycoproteins in terms of having shorter glycan lengths and fewer branches 34.

A protein can have multiple N-linked or O-linked glycosylation sites. The attachment of large glycans may drastically change the molecular weight of a protein; moreover, the physicochemical properties such as hydrophobicity will be affected. Glycosylations play a key role in a wide array of biological processes, such as ligand recognition, intracellular trafficking,
cell adhesion, immune defense and programmed cell death. Since different cells produce
different monosaccharide units, the expression of a glycoprotein may result in the formation of
foreign glycoforms that may trigger immunogenicity. For example, it has been found that using
yeast cells to culture human glycoproteins inevitably results in the production of structural
heterogeneities with higher mannose contents \(^{38}\). In Chapter 2 of this dissertation, the therapeutic
protein being studied is a N-linked glycoprotein. The N-linked glycans play a critical role in the
stability of the therapeutic protein, thus optimization of sample preparation steps was conducted
to achieve unambiguous characterization of the protein.

1.3.2.3 Other Modifications

Deamidation and oxidation are other two post-translational modifications that are frequently
observed on proteins. Protein deamidation, is a spontaneous, irreversible reaction that occurs \(\textit{in vivo}\) without participation of an enzyme \(^{39}\). Deamidation primarily occurs at an asparagine
residue (Asn), converting Asn into aspartic acid (Asp) and isoaspartic acid (isoAsp). The
conversion results from an intramolecular rearrangement \(\textit{via}\) a \(\beta\)-aspartyl shift mechanism, as
shown in Figure 1-4. Deamidation is initiated with nucleophilic attack of the side chain carbonyl
group of an Asn residue from the \(\alpha\)-amino group of the adjacent amino acid, thereby forming a
five-membered ring succinimide/ cyclic imide (Asu) intermediate. The Asu intermediate rapidly
undergoes hydrolysis to produce a mixture of Asp and isoAsp with an approximate molar ratio of
1:3 \(^{39-41}\). Meanwhile, racemization reactions are often accompanied by the deamidation process,
resulting in a mixture of L- and D- Asp and isoAsp. Among these species, L-isoAsp is the major
product.
Figure 1-4. Mechanism of (A) asparagine deamidation and (B) aspartic acid isomerization $^{41,42}$.

Deamidation also occurs on glutamine residue (Gln) to yield glutamic acid (Glu) and isoglutamic acid (isoGlu) in a similar manner, but with lower reaction rates.

Formation of Asp or isoAsp from deamidation of Asn introduces additional negative charges and thereby gives rise to charge variants in protein products. This will alter the distribution of
charge variants, changing the pI of the protein. In Chapter 4, we developed an interface to
couple capillary isoelectric focusing with electrospray ionization mass spectrometer to separate
protein charge variants including deamidated variants. Importantly, introducing a methyl group
on a peptide backbone (isoAsp) leads to different physicochemical properties of polypeptide
chains, and induces conformational changes and potentially functional alterations of proteins,
thereby ultimately causing an immune response, or initiating protein degradation or inducing
other adverse effects. It is very important to elucidate whether deamidation occurs on the protein
biopharmaceutical.

Protein oxidation is another major post-translational modifications. Methionine (Met),
containing a sulfur atom, is most susceptible to oxidation\textsuperscript{43}. Oxidation on Met leads to the
formation of methionine sulfoxide and methionine sulfone. Other residues such as tryptophan
(Trp) may also undergo oxidation under more stringent conditions.

1.4 **PROTEIN AND PEPTIDE SEPARATION METHODS**

In recent years, enormous efforts have been made to improve the scan speed, dynamic range,
and sensitivity of MS instrumentation to provide fast, precise and reproducible results for protein
characterization. However, it is still difficult to characterize and quantify a target protein from a
complex sample using only an MS instrument. High resolution separation is required for
unambiguous identification of proteins or peptides, since in MS detection, the identification is
based on m/z values of precursor ions and isobaric species will have the same m/z precursor
ions. Ideally, proteins or peptide precursor ions with the same m/z ratios should not be co-eluted.
Additionally, the dynamic range of a complex biological sample can exceed $10^{10}$, which is far beyond the detection capability that a novel MS instrument can provide. Furthermore, the duty cycle of MS instruments may result in an undersampling of very complex samples. For these reasons, it is important to perform the separation prior to MS detection.

Currently, there are a broad range of separation techniques that are able to readily couple with MS instruments in on-line or off-line mode for protein identification and quantitation. Some biochemical fractionations, gel-based separations, chromatography-based separations and electrophoresis-based separations are discussed in this section.

### 1.4.1 Immunoprecipitation (IP)

IP is one of the most widely used biochemical fractionation methods to isolate and enrich low-abundance target proteins or protein complexes from a biological sample in order to provide insight into the components in a biological interaction network. Usually, a target-specific purified antibody is immobilized on a solid matrix such as agarose beads or magnetic beads via reactive chemical groups on the surface of the solid matrix. Subsequently, the immobilized antibody is used to pull out bait proteins or protein complexes from the biological sample.

Another approach for the immobilization of antibodies is using beads derivatized with streptavidin to allow strong binding by a biotinylated antibody. The protein complexes are further analyzed by MS for unbiased identification of the target protein and its interaction partners. MS-based IP experimentation is superior to traditional immunoblot methods in the simultaneous identification of multiple interaction partners, even without prior knowledge of the partners. However, one of the major limitations of IP is that it can pull down many non-specific interacting proteins at the same time. In Chapter 3, we describe LC-MS analysis of a pulled down sample in order to determine the target of a small molecule drug.
1.4.2 Gel-Based Methods

Gel-based methods are inexpensive, sensitive and well-established techniques primarily used for the separation and fractionation of biological samples. They offer the benefit of reducing the sample complexity and increasing the sample compatibility with subsequent MS analysis. The basis of gel electrophoresis is that charged protein molecules with distinct physicochemical properties migrate at different rates through a gel matrix under an electric field, thereby locating at different positions in the gel matrix at a given time. The most widespread gel-based separation methods are sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and isoelectric focusing (IEF).

In SDS-PAGE, the separation and fractionation of proteins rely on their different molecular weights. An anionic detergent, SDS, is applied to wrap around unfolded proteins with negative charges at a constant amount relative to the size, regardless of the intrinsic net charges of the proteins. The resulting overall shapes of the proteins are very similar; thus, the only factor that can be attributed to different migration rates under an electric field is their molecular weights. Small proteins move rapidly, while large proteins have lower mobilities. After SDS-PAGE separation, gels are visualized by staining. Each gel lane is then divided into several gel bands which are individually excised, enzymatically digested and analyzed by LC-MS/MS. Fang & Robinson have compared the proteome coverage at the protein and peptide levels with implementation of some commonly used fractionation methods including SDS-PAGE, IEF and SCX. They found that SDS-PAGE offered the highest identification of both proteins and peptides.

Another conventionally used gel separation technique is IEF. The principle for the separation of protein molecules by IEF is that each protein has its characteristic isoelectric point (pI).
Specifically, a pH gradient is established in the gel by carrier ampholytes when applying an electric field. A protein moving through the gel is either positively or negatively charged, depending on its pI and the surrounding pH. During the migration, the protein molecule will either acquire protons from the local environment or lose protons until the pH is equal to its pI. The protein will be focused in the specific position where its net charge is zero. The carrier ampholyte can either be used in solution phase mixing with buffer to establish a linear pH gradient or can be immobilized in the gel matrix by covalently grafting the ampholytes to form a fixed pH gradient. IEF offers the highest resolution among all electrophoresis techniques, and it can differentiate the charge variants induced by protein PTMs with a subtle pI change (as low as 0.1 pH unit or less).

Combining the orthogonal separation of proteins by IEF and by SDS-PAGE leads to the generation of 2-dimensional gel electrophoresis (2DE). One of the major features of this technique is the high resolving power of 2DE to separate thousands of proteins in a single gel. With advances in modern electrophoresis and immobilized pH gradient techniques, proteins with a difference in pI values as low as 0.001 pH unit can be separated by 2DE. Nevertheless, some disadvantages of 2DE should be taken into account, such as labor-intensive sample handling, relatively poor reproducibility, non-applicability to membrane proteins because of the solubility issue, and limited utility for some low abundant proteins due to limited dynamic ranges. Overall, gel-based methods are attractive analytical techniques to reduce the complexity of biological samples and are often employed together with chromatography-based methods for quantitative proteomics.
1.4.3 Chromatography-Based Methods

At present, HPLC has become the predominant chromatographic method for the purification, fractionation, identification and quantification of proteins. The fundamental principle of HPLC separation relies on the interaction of proteins or peptides with the stationary phase and the mobile phase. There are various types of chromatographic modes available for protein and peptide separation to fulfill different research objectives. In the following sections, different modes of HPLC that are typically used for protein characterization and quantitative proteomics are discussed.

1.4.3.1 Normal Phase Liquid Chromatography (NPLC)

NPLC separates compounds according to their polarities. The stationary phase is polar, while the mobile phase is less polar. An analyte with less polarity has less retention on the column and will be eluted earlier than a polar analyte. NPLC enables the separation of very hydrophilic compounds that are not retained on an RPLC column. However, the use of nonpolar eluents such as hexane in NPLC limits the application of NPLC in protein or peptide separation, since most proteins or peptides show low solubilities in nonpolar solvents. In addition, highly organic and nonpolar eluents with low volatility are not compatible with MS instrumentation.

1.4.3.2 Hydrophilic Interaction Chromatography (HILIC)

HILIC employs a polar stationary phase such as silica-, amino- and cyano-bounded phases; and an aqueous-organic mobile phase that can form a water layer on the surface of the stationary phase. The retention mechanism of analytes is based on the liquid-liquid partitioning between a water-rich layer and a hydrophobic eluent. Comparing with NPLC, HILIC does not suffer from the problem of low solubility of analytes due to the employment of aqueous-organic
mobile phases. Furthermore, HILIC is easy to interface with MS, especially for the online ESI-MS \(^60\). An important application of HILIC is separation of hydrophilic compounds that show poor retention and separation on RPLC. To date, HILIC has been successfully applied with online ESI-MS to study protein glycosylations \(^61\).

1.4.3.3 Reversed Phase Liquid Chromatography (RPLC)

Despite continuous developments in various separation modes, RPLC has remained the most powerful separation mode for proteins and peptides in complex biological samples due to its high efficiency, selectivity, resolution, and peak capacity, as well as its capability for directly interfacing with both UV and MS detection \(^62\). Proteins and peptides are separated by RPLC according to distinct hydrophobicities. The retention mechanism for proteins or peptides on a reversed phase column is based on analytes partitioned between a non-polar hydrophobic stationary phase and a polar hydrophilic mobile phase. Analytes that are more hydrophilic tend to exhibit less retention and pass through the column with the mobile phase. By increasing the hydrophobicity of the mobile phase \(\text{via}\) an increase in the organic content, analytes are subsequently separated and eluted in order of hydrophobicity \(^63\text{–}65\).

The hydrophobic stationary phase used in RPLC typically consists of porous silica particles bearing surface derivatization with aliphatic alkyl ligands. N-butyl (C4), octyl (C8) and octadecyl (C18) silica particles are most widely used for protein and peptide separation due to their suitable retention characteristics. The hydrophobicity of the aforementioned stationary phases increases in response to increasing alkyl length. Generally, proteins exhibit good retention on C4 and C8 columns, but are seldom separated by the C18 column because of the difficulty in recovering proteins from the too-hydrophobic stationary phase. In contrast, C18 and C8 columns
are the first choices for separating peptides. Protein or peptide separations on RPLC are usually carried out under acidic conditions (pH lower than 4) in order to minimize the undesirable interaction between positively-charged solutes and the residual silanol groups on the surface of stationary phases.

The mobile phase employed in RPLC consists of an aqueous buffer A (water) and an organic buffer B (acetonitrile, methanol, propanol or isopropanol), which are mixed to a specified ratio by an external pump system. Among the above four organic buffers, acetonitrile is the most popular, due to its appropriate viscosity, volatility, hydrophobicity and detector compatibility.

In addition, ion-pairing reagents such as acetic acid, formic acid (FA) and trifluoracetic acid (TFA) are added into mobile phases A and B to enhance separation efficiency by lowering the pH of the mobile phase, reducing protonation of the charged residues on solutes, and thereby increasing hydrophobicity and improved retention. TFA is a good ion-pairing reagent with which to perform peptide separation with good peak shape. However, it can introduce ion suppression effect on the analytes in ESI-MS. Consequently, FA is widely used as the acid of choice for reversed-phase separation. Although FA does not offer very strong ion pairing as TFA, it provides the best compromise in retention time and ionization.

Because of its high resolving power, selectivity and reproducibility, RPLC has been extensively used as a single-stage separation method for proteins and peptides. Additionally, in order to handle complex proteomic samples, especially the proteolytic peptide mixtures obtained via the bottom-up approach, RPLC can be integrated with other orthogonal separation methods for even higher resolving power. However, one limitation of RPLC is that the high levels of organic content in the mobile phase will denature proteins; as a result, higher-order conformation of proteins cannot be studied using the RPLC approach.
1.4.3.4 Ion Exchange Chromatography (IEX)

The fundamental principle of IEX separation primarily relies on electrostatic interactions between charged analytes and charged stationary phases. In IEX, the stationary phase typically consists of silica-based or polymer-based porous matrices grafted with a charged ligand. Elution of bound proteins or peptides is conducted by neutralization of ionic interactions with a salt gradient, the counterions from the mobile phase will competitively displace bound analytes.

An alternative mode for IEX, chromatofocusing (CF) utilizes the pH change of mobile phases for separation. In the CF mode, a pH gradient is developed in the column by migration of mobile phases containing ampholytes, which progressively titrates charge species and thereby neutralize ionic interactions between analytes and ligands.

In contrast to RPLC in which the employment of a high organic mobile phase tends to denature proteins and cause the loss of biological functions, IEX shows its major advantage in separating proteins under native conformation. Additionally, the unique selectivity and high resolving power of IEX contribute to its widespread application. For example, IEX can be applied to isolate and enrich PTMs or chemically modified proteins and peptides with charge heterogeneities.
1.4.4 Capillary Electrophoresis-Based Methods

1.4.4.1 Basic Principles of Capillary Electrophoresis

As slab-gel electrophoresis, capillary electrophoresis (CE) is an alternative form of electrophoresis in a capillary format. In CE, the electrophoretic mobility of an analyte and the electroosmotic mobility of the background electrolyte contribute to the net migration velocity of the analyte molecule. The electrophoretic mobility ($\mu_{ep}$) is characteristic for each charged analytes, which is defined as Equation 1-1:

$$\mu_{ep} = \frac{q}{6\pi \eta r}$$

where $q$ and $r$ are the charge and hydrodynamic radius of the ion, and $\eta$ is the viscosity of the solution. \(^{74,75}\) According to this equation, small and higher charged analytes has higher electrophoretic mobilities and therefore are separated from the large analytes carrying less charges.

![Scheme of the electric double layer formed inside a bare fused silica capillary.](image)

**Figure 1-5.** Scheme of the electric double layer formed inside a bare fused silica capillary. The direction of EOF is indicated as an arrow. And the neutral species are indicated as a letter N. \(^{76}\)
When CE is performed in a fused silica capillary, a bulk flow, electroosmotic flow (EOF) is generated by the background electrolyte under an electric field due to the influence of the charge density on the capillary inner wall. An electric double layer is formed between deprotonated silanol groups (when the pH > 2) on the inner wall of silica capillary and the cations hydrolyzed by the background electrolyte. When applying an electric field, the positively charged layer migrates towards the cathode end of the separation capillary and results in the formation of EOF, as illustrated in Figure 1-5. The electroosmotic mobility ($\mu_{eo}$) is defined by the dielectric constant ($\varepsilon$) and the viscosity ($\eta$) of the background electrolyte and the zeta potential ($\zeta$), and it can be described as Equation 1-1:

$$\mu_{eo} = \frac{\varepsilon \zeta}{4\pi\eta}$$

Thus, when conducting CE in a fused silica capillary, the overall mobility or apparent mobility ($\mu_{app}$) of a charged analyte is the result of electrophoretic mobility of the solute with the electroosmotic flow of the buffer. The apparent mobility ($\mu_{app}$) of a charged analyte can be calculated from Equation 1-3.

$$\mu_{app} = \mu_{ep} \pm \mu_{eo}$$

In most cases, cations move towards the same direction as EOF, resulting in less retention time during separation. On the contrary, anions migrate towards the opposite direction as EOF, resulting in longer retention in the capillary. In this way, different analytes are separated due to the distinct apparent mobility.
Nevertheless, CE separation can be performed in the absence of EOF. According to Equation 1-4, the zeta potential ($\zeta$) is proportional to the charge density of the capillary inner wall ($\sigma$)\textsuperscript{80}:

$$\zeta = \frac{\sigma}{\varepsilon \sqrt{\frac{2 c F^2}{\varepsilon R T}}}$$

where $c$ is the concentration of the BGE, $F$ is the Faraday constant, $R$ is the universal gas constant and $T$ is the temperature. Since the electroosmotic mobility is dependent on the charge density of capillary wall and the properties of BGE, changing buffer pH, ionic strength or adding buffer additives (e.g., polysaccharides) can effectively reduce the EOF and minimize protein adsorption to the capillary wall\textsuperscript{81}. Another strategy to control the EOF is applying static capilalry coating to shield the residual silanol group from exposing to the BGE. For example, neutral coated capillary with polyacrylamide, polyvinyl alcohol (PVA), and polyethylene glycol (PEG) are successfully employed for CE separation of proteins and peptides due to the suppressed EOF as well as the minimized sample adsorption\textsuperscript{82}.

Two parameters, the number of theoretical plates ($N$) and resolution ($R_s$) are typically used to evaluate the system performance of capillary electrophoresis. The separation efficiency of a system can be determined from $N$, which is defined as Equation 1-5, where $E$ is the field strenght, $l$ is the effective length of a capillary and $D$ is the diffusion coefficinecy of the analyte\textsuperscript{78,79}.

$$N = \frac{\mu_{ep} E l}{2D}$$

( 1-5 )
In practice, the theoretical plates can be calculated from the electropherogram according to Equation 1-6:

\[ N = 5.54 \times \left( \frac{t_r}{w_{1/2}} \right)^2 \]  

(1-6)

where \( t_r \) is the migration time of the analyte and \( w_{1/2} \) is the peak width at half height, respectively. Resolution (\( R_s \)) of two closely-eluted analytes is related to the separation efficiency (\( N \)), the difference in electrophoretic mobility of two analytes (\( \Delta \mu_{ep} \)) and the average electrophoretic mobility (\( \mu_{avg} \)) \(^{83} \):

\[ R_s = \frac{1}{4} \frac{\Delta \mu_{ep}}{\mu_{avg}} \sqrt{N} \]  

(1-7)

Resolution can be calculated from the migration time of the analyte and the peak width at half height, according to Equation 1-8:

\[ R_s = \frac{1.18 \times (t_{r2} - t_{r1})}{\sum w_{1/2}} \]  

(1-8)

Similar to HPLC, there are various separation modes applied in CE that can fulfill specific needs. Some major separation modes, such as capillary zone electrophoresis (CZE) and capillary isoelectric focusing (CIEF), will be discussed in the following section.
1.4.4.2 Different CE-Based Separation Modes

CZE is the simplest mode in all capillary electrophoresis-based separation methods. In CZE mode, charged analytes move in a capillary filled with background electrolyte and separate in discrete narrow zones according to distinct apparent velocities under the electric field. CZE has been widely applied, for example, to separate protein charge variants. Moreover, CZE can also be used for high-sensitivity peptide mapping experiments.

CIEF is another powerful capillary electrophoresis mode with high resolving power to separate complex protein and peptide mixtures and to differentiate microheterogeneities caused by PTMs, and other factors. Similar to slab gel IEF, CIEF separates proteins and peptides according to their characteristic isoelectric points (pIs). In a typical CIEF experiment, protein or peptide samples are mixed with carrier ampholytes supplemented in aqueous gel medium and filled into a whole capillary. An acidic solution and a basic solution are used as the anolyte and catholyte and placed at the two sides of the capillary. Upon applying a high voltage (e.g., 30 kV) across the capillary, a continuous pH gradient is established inside the capillary by the carrier ampholytes. The proteins or peptides are concentrated simultaneously while being focused into sharpe zones where their net charges are zero. Comparing to slab gel IEF, the strong electric field employed in CIEF not only accelerates separation speed, but also contributes to the large concentrating capability. Experiments have shown that CIEF enables the concentration of intact proteins at least 500-fold, while for proteolytic peptides, the enrichment is around 700-fold when using CIEF in the presence of carrier ampholytes within a neutral coated capillary. The large concentrating potential of CIEF is significant for the analysis of low-abundance species in proteomic samples. Due to its concentrating ability, CIEF is frequently employed as a preconcentration step in multidimensional separations. Additionally, the high resolving power...
of CIEF makes it suitable for reducing the complexity of proteomic samples to achieve a large number of protein identifications. Chapter 4 will describe efforts in coupling CIEF directly to MS.

1.4.4.3 CE-MS Interfaces

The major technical challenge involved in the implementation of CE-MS is the interface. With extensive research addressing interfacing techniques, both of the predominant ionization methods, ESI and MALDI, can be applied in CE-MS. On-line coupling of CE with ESI-MS favors highly-automated processes and analysis of a wide range of analytes, from small peptides to large intact proteins. In the following section, online CE-ESI-MS interfaces are discussed.

CE-ESI-MS Interface

The interfaces applied in the online coupling of CE with ESI-MS can be divided into two groups, sheath-flow interface and sheathless, depending on whether make-up flow is employed.

The most widely applied sheath-flow interface in CE-ESI-MS is coaxial designed by Smith and co-workers in 1988 for CZE-ESI-MS. As illustrated in Figure 1-6A, the basic design of the coaxial sheath-flow interface consists of three tubes in a coaxial arrangement, including a separation capillary to deliver samples in a BGE, a conductive metal tubing to deliver sheath liquid, and an outer tubing for nebulizer gas. Typically, the separation capillary slightly protrudes out of the metal tubing to minimize potential dead volume. The metal tubing acts as the electrode for both the separation capillary and ESI emitter. Employing the sheath liquid not only provides a contact closure between the metal tubing with the separation capillary outlet for delivering a stable CE flow, but also dilutes the BGE with MS-compatible solvents (e.g., methanol) by
adding volatile acid (e.g., acetic acid) to assist in analyte ionization. The nebulizer gas is responsible for the desolvation of analytes.

Figure 1-6. Schematic setup of three types of CE-MS interfaces: (A) a coaxial sheath flow interface, (B) a liquid junction interface and (C) a porous sheathless interface.

The advantages of sheath-flow interfaces are that they are robust and permit versatile choices of BGE. However, employing relatively high sheath flow (1~10 µL/min), in comparison with conventional CE flow (20~100 nL/min), inevitably leads to large sample dilution and thereby loss of sensitivity, since the signal in ESI-MS is over a wide flow range concentration-dependent. Additionally, parabolic flow introduced by the sheath liquid and nebulizer gas leads to lower separation efficiency.

Another type of sheath-flow interface is the liquid junction interface introduced. The liquid junction interface provides a low volume chamber between the exit of separation capillary and ESI emitter, which allows the low flow sheath liquid to modify the CE effluent with minimized sample dilution. In the original design, a short fused silica capillary with minimized dead volume was connected to the exit of a separation capillary through a microcross. The sheath liquid flow (<200 nL/min) was controlled by the restricted low pressure inside the ESI chamber. In the later modified design of liquid junction interface, as shown in Figure 1-6B, the separation
capillary terminus is coated with gold and placed into a tapered glass ESI emitter to further minimize the extra volume, while the ESI emitter is inserted into stainless steel tubing delivering nebulizing gas for desolvation. The sheath liquid filled in the space between the ESI emitter and the separation capillary creates a liquid junction to modify the BGE and to provide electric contact. The liquid junction interface design with the tapered emitter allows the coupling of CE with a nano-ESI source, which results in enhanced desolvation and increased sensitivity.

In Chapter 4, we present a novel design of liquid junction interface with miniaturized volume to couple CE-ESI-MS for protein and peptide analysis.

Sheathless interfaces do not include any make-up flow, the latter of which has the potential to cause sample dilution; thus, better sensitivity is typically achieved. On the contrary, the requirement for the BGE in the sheathless strategy is much more restricted than that in the sheath-flow interface since the BGE should provide both good CE separation and MS compatibility. The fundamental principle of the sheathless interface is the direct coupling of CE with ESI-MS via a ESI emitter, either by directly from the outlet of the CE separation capillary or by creating a separate emitter that connects to the outlet of the separation capillary. There are various designs for sheathless interfaces, such as metal coating of a tapered capillary outlet to generate a conductive ESI emitter or the insertion of an electrode in a tapered capillary terminus. However, in these designs, potential bubble formation at the electrode or contact area may perturb the electrospray and cause electric discharge. The porous sheathless interface developed by Moini in 2007 overcomes these issues. As illustrated in Figure 1-6C, the capillary terminus is chemically etched to form a porous tip and is directly inserted into an ESI metal needle filled with BGE. The porous section allows the transport of small ions into the capillary to provide electric connection, whereas proteins and peptides cannot pass through the pores in the low
nL/min flow rate. Additionally, no dead volume is added, and the bubbles formed outside of the separation capillary will not perturb the separation or ionization. However, peak broadening is observed with the employment of the porous sheathless interface, likely resulting from protein adsorption by the porous capillary wall \(^{103}\). This interface has been commercialized by Beckman Instruments \(^{104}\).

### 1.5 MASS SPECTROMETRY

Mass spectrometry (MS) is a comprehensive and versatile analytical technique that plays a central role in the characterization and quantitation of protein therapeutics and proteomic samples. It not only provides the measurement of the intrinsic properties of a protein, such as molecular weight, with high sensitivity and accuracy, but also offers the measurement of detailed structural features such as sites and compositions of various modifications. Fundamentally, any commercially available mass spectrometer consists of three major components, an ion source for the production of gas phase ions, a mass analyzer for the differentiation of ions by m/z ratios, and a detector for the detection and recording of relative abundances of each resolved ion \(^{24}\). In this section, discussion will be focused on two soft ionization techniques, some of the popular mass analyzers, and the predominant ion activation methods (MS2) applied in proteomic studies.

#### 1.5.1 Electrospray Ionization

The ion source is central to a mass spectrometer, where neutral or charged molecules are ionized to form charged species for introduction into the vacuum of the instrument. The first
challenge faced in analyzing proteins and peptides by MS is the introduction of polar and non-volatile molecules into the gas phase without extensive thermal degradation\textsuperscript{62}. The emergence of two soft ionization methods, MALDI and ESI, has revolutionized the application of MS for large biomolecules. In this dissertation, the major ionization method used is nano-ESI; therefore, the discussion below will solely deal with ESI ionization.

Since its introduction by Fenn in the late 1980s, ESI has become one of the predominant ionization methods applied for the online coupling of LC-MS or CE-MS systems for both qualitative and quantitative analysis of large biomolecules such as proteins\textsuperscript{105}. First, the intrinsic principle of ESI enables the ionization of continuous effluents from LC or CE systems, which allows online coupling of LC-MS or CE-MS. Secondly, there is no upper limit of the mass range for the analyte since ESI produces multiply-charged ions, which is especially beneficial when measuring large protein molecules. Thirdly, ESI-MS is highly sensitive to flow rate, which allows the reduction of flow rate to achieve higher sensitivity\textsuperscript{106}. Additionally, operation at atmospheric pressure, and solvents compatible with various modes in HPLC and CE all contribute to the universal usage of ESI\textsuperscript{62}. However, it should be noted that in comparison with MALDI, the tolerance of ESI to salts and detergents is relatively lower.

The transformation from aqueous phase proteins or peptide molecules to gaseous phase ions generally involves the creation of an electrically-charged spray, liberation of ions from droplets, and transport of gaseous ions from the atmospheric environment to high vacuum\textsuperscript{107}. Specifically, solution containing analyte molecules encounters a strong electric field at the spray needle which disperses thousands of charged micrometer-sized analyte-solvent droplets. The evaporation of solvent by elevated temperature or by applying a sheath gas flow leads to the reduction of droplet size, increasing the surface charge density until it reaches a critical point,
called the Rayleigh limit. At this point, where the electrostatic repulsion between surface charges competes with the surface tension, a Coulomb explosion occurs to generate smaller droplets. The repetition of these processes ultimately results in the ejection of charged ions from droplets into the gaseous phase. The free ions in the atmospheric pressure in the ion source region are sampled by a skimmer cone and introduced into the mass analyzer for subsequent analysis 18,107,108.

The development of nanospray techniques further increases the method’s sensitivity as well as decreased sample consumption compared with conventional ESI. In nano-ESI, the flow rate of sample solution decreases from the microliter-per-minute level to the nanoliter-per-minute level, which leads to the formation of a smaller-sized droplet containing more analytes and fewer impurities so that less ion suppression is observed. Meanwhile, in nano-ESI, solvents with high surface tension, such as water, can form stable sprays; and addition of organic solvent, such as acetonitrile and methonal improves the ionization efficiency and sensitivity. Lower voltages applied at the spray needle to maintain the spray stability have less tendency to cause electric discharge. Moreover, the spray needle that is placed much closer to the inlet of MS in the nano-ESI leads to less sample loss and a stronger signal 109.

1.5.2 Mass Analyzer

The mass analyzer is a compartment to store and differentiate ions based on their m/z ratios. The isolation of ions is performed either by an electrical field or a magnetic field 105. Currently, there are many different types of mass analyzers commercially available. Each is based on distinct physical principles and analytical performance characteristics. For example, some mass analyzers are based on the scanning of ion beams while others on the trapping of ions 62. Additionally, recent innovative instrument designs have permitted the emergence of hybrid MS instruments that can be applied to support a wide range of research strategies 110. The following
section will focus on the principles and features of different types of mass analyzers frequently used in proteomics and protein analysis studies. Furthermore, some advanced hybrid MS instruments related to the research in this thesis will be introduced as well.

1.5.2.1 Single-Stage Mass Analyzers

The major structure of a quadrupole (Q) mass analyzer is an assembly of four parallel metal rods in which one pair of diagonally-located rods has a direct current (DC) potential applied to them, while the remaining rods are superimposed with a radio-frequency (RF) potential. Under the influence of DC and RF potentials, a charged ion entering the mass analyzer moves towards the negatively-charged rods with oscillation. By changing the value and direction of the DC and RF potentials, the movement and oscillation of ions can be controlled so that only ions within the desirable m/z ranges will be on stable trajectories, travel through the rods, and reach the detector, whereas ions with m/z ratios outside the pre-defined range will have unstable trajectories and will eventually strike the rods \(^{105,108}\). In this way, the quadrupole functions as a mass filter to separate ions based on their m/z ratios. The attractive features of quadrupoles are their low cost, robustness and ease of maintenance. A quadrupole mass analyzer is typically coupled with continuous ionization methods and is not suitable for use with pulsed ionization from MALDI.

Time-of-flight (TOF) is an ion-scanning type of mass analyzer. Unlike the rejection of ions on unstable trajectories in quadrupole analyzers, all ions can reach the detector in a TOF analyzer, with each ion having a different flight time \(^{105}\). Ions bearing the same kinetic energy travel through a 1- to 2-meter-long tube from an ion source to a detector with different times that are proportional to the square root of the m/z ratio of each ion \(^{111}\). TOF analysis is characteristic for its resolving power (>12,000), low ppm level mass accuracy, and high mass range \(^{110}\). TOF is
suitable in conjunction with both ESI and MALDI ionization methods and is widely used in multi-stage MS instruments such as TOF-TOF and Q-TOF.

Ion trap (IT) enables trapping and accumulating ions over time with an oscillating electric field. There are two types of ion traps, the 3D ion trap (also called a quadrupole ion trap, QIT) and the 2D ion trap (also called a linear ion trap, LIT). The major components in a QIT are a pair of end-cap electrodes with a ring structure electrode located halfway between them. As in the quadrupole mass analyzer, a combination of RF and DC voltages is applied on the ring electrode to create a three-dimensional quadrupolar field for trapping and accumulating ions with m/z within the stable 3D trajectories. By changing the trapping potential, the ions with desirable m/z will be on unstable trajectories and will be expelled for detection via the exit end-cap electrode. Unlike the quadrupole mass analyzer, which spatially separates ions by selecting ions with specific m/z to pass through the rods, in QIT, ions are separated via time. All ions with different m/z ratios are accumulated in the trap and are ejected outside the trap sequentially. QIT features a fast scan rate, high sensitivity and remarkable tandem experimental capabilities. For tandem MS, only ions with pre-defined m/z accumulate in the trap and collide with an inert gas such as helium for fragmentation. The product ions can either be ejected for detection or trapped within the cavity for further fragmentation (tandem MS) to provide structural information. However, QIT has some limitations in terms of low resolution (~2,000), low mass accuracy (100 ppm), and limited ion trapping capacity. The advent of LIT allows a 10-fold increase in ion trapping capacity, which results in a wider dynamic range and higher sensitivity. Owing to its remarkable high-throughput feature, the LIT mass analyzer is used in hybrid MS instruments such as LTQ-FTICR and LTQ-Orbitrap to analyze bottom-up proteomic samples for large proteome coverage.
Orbitrap is a novel type of ion trap mass analyzer invented by Makarov that is widely used in high performance mass spectrometers for analyses of complex biological samples. The Orbitrap mass analyzer utilizes a static electric field created between an axial central spindle-shaped electrode and a coaxial outer barrel-like electrode to trap ions. Under the influence of the three-dimensional electric field, ions tangentially injected into the trap oscillate harmonically around the central electrode with an axial frequency that is dependent on m/z ratios but independent of the energy and spatial location of the ions. Fast Fourier transform algorithms (FFT) are then applied to transform the current image information to an MS spectrum containing m/z information. As a result, large trapping capacity is achieved with the Orbitrap mass analyzer, which leads to an increased space charge capacity. Remarkable features such as high resolution (~20,000), high mass accuracy (2-5 ppm), and wide dynamic range (> 10^3) enable the Orbitrap to be a powerful tool to manage the increasing complexity of biological samples.

1.5.2.2 Multi-Stage Mass Analyzers

In addition to the major types of single-stage mass analyzers described above, the trend towards advanced MS technology is to combine different types of mass analyzer in a hybrid MS instrument to improve its analytical performance and extend its applications. In this section, three hybrid MS instruments used in this dissertation, Q-TOF, Orbitrap Elite and Q Exactive, will be introduced.

The Q-TOF is a multi-stage hybrid MS instrument first introduced in 1996 that attaches two quadrupole mass analyzers with a TOF mass analyzer in an orthogonal configuration. The resulting Q-TOF mass spectrometer takes advantage of ion scanning and tandem MS capabilities from quadrupole technology and the high resolving power from TOF to perform MS and tandem
MS experiments with high resolution, high sensitivity and high mass accuracy \(^{105}\). In a single-stage MS experiment, the quadrupoles Q1 and Q2 are operated in the RF-only mode to isolate and scan ions within specific m/z ranges. In a tandem MS experiment, Q1 serves as a mass filter to allow precursor ions with desirable m/z ratios to pass through, while Q2 functions as a collision cell for fragmentation. The product ions are separated by a TOF mass analyzer based on different flight times. An additional quadrupole, Q0, is typically applied before Q1 for ion focusing to achieve better ion beams \(^{105}\). The Q-TOF has been extensively used in both top-down and bottom-up proteomics for protein identification and PTM analysis \(^{62}\). In Chapter 2 of this dissertation, a Q-TOF mass spectrometer from Agilent Technologies was used.

The LTQ-Orbitrap is a more recently developed hybrid MS instrument that integrates the remarkable features of fast scan rate, robustness, sensitivity and tandem MS capabilities from LIT mass analyzers, as well as high resolution and high mass accuracy from Orbitrap mass analyzers \(^{118}\). The LTQ-Orbitrap enables a parallel operation mode that provides protein or peptide mass measurements with outstanding resolution (~150 K) and mass accuracy (2-5 ppm) from the Orbitrap, and subsequent fragment ion measurement with exquisite sensitivity (fmol level) from the linear ion trap, resulting in a broad range of applications in proteomic research, such as intact protein characterizations, PTM analysis, and peptide quantifications \(^{115,119-121}\). The Orbitrap Elite is the novel generation of LTQ-Orbitrap from Thermo Scientific that was introduced on the market in 2011. The Orbitrap Elite employs a compact high field Orbitrap mass analyzer geometry and an advanced FT algorithm for signal deconvolution leading to ultrahigh resolution greater than 240,000, which is particularly attractive for study of intact proteins \(^{122,123}\). Additionally, the incorporation of multiple collision modes, including CID, ETD and HCD, allows multiple types of fragmentations to yield detailed structural information. This
extends the applications to fulfill a wide range of research purposes including characterization of protein PTMs and quantitation of peptides with tandem mass tags \(^{124-126}\). Overall, ultrahigh resolution, sub-ppm-level mass accuracy, large dynamic range exceeding 5,000 within a single scan, and multiple fragmentation capabilities are the key features of the Orbitrap Elite. In Chapter 3 of this dissertation, determination of complex disulfide linkages in a therapeutic protein were achieved by applying the highly accurate mass identification capability and multiple fragmentation (CID and ETD) modes from the Orbitrap Elite.

Another hybrid MS instrument, the Q Exactive, was launched in the same year as the Orbitrap Elite from Thermo Scientific. In the Q Exactive mass spectrometer, the ion scanning capability from the quadrupole mass analyzer and the high resolution, high mass accuracy from the Orbitrap mass analyzer are combined to achieve outstanding analytical performance. Due to its high resolution (>140,000), high mass accuracy (< 1 ppm), large dynamic range (>10⁴), femotogram-level sensitivity, and incorporation of fast HCD fragmentation technology, the Q Exactive is widely used in qualitative and quantitative proteomics \(^{127,128}\). In Chapter 3, the Q Exactive was applied for target identification in a proteomic sample.

### 1.5.3 Fragmentation

Ionization and isolation of proteins or peptides in a single-stage MS instrument can only provide molecular weight information. Tandem mass spectrometry is required to elucidate more detailed structural details of proteins or peptides for unambiguous identification. Tandem mass spectrometry (MS/MS) or MS\(^n\) experimentation involves the ion activation and detection of a precursor ion in the first MS stage, further excitation and dissociation of the precursor ion, and molecular weight measurements of product ions in the subsequent MS stages. Modern MS
instruments with multiple mass analyzers, such as triple quadrupole and LIT-Orbitrap, can offer multiple stages of mass measurements. There are various types of ion activation methods to fragment precursor ions for tandem mass spectrometry. In this section, three universally applied ion activation methods in protein characterization and proteomics studies, collision-induced (activated) dissociation (CID), electron-transfer dissociation (ETD), and higher-energy collision dissociation (HCD), will be introduced.

![Diagram of peptide structure](image)

**Figure 1-7.** Nomenclature of fragment ions for a peptide.

### 1.5.3.1 Collison-Induced Dissociation

CID remains the first choice of ion activation method applied in tandem MS experiments for peptide sequencing. The mobile proton model, introduced by Wysocki et al, has been widely used to explain the dissociation of a peptide upon ion activation. This model assumes that a heterogenous population of the protonated forms for a peptide can be generated during the CID process. A protonated peptide usually contains multiple protonation sites where the main protonation sites are the basic N-terminal amino groups, and the side chain of Arg, Lys and His. In the positive ESI mode, the proton is first attached to the most basic sites of the peptide precursor ion. Upon activation by collision with an inert gas such as helium or argon, the mobile proton then rapidly migrates between different thermodynamically favourable protonation sites.
on the peptide and results in the formation of different protonated structures, which dissociate via
different pathways and produces fragment ions containing sequence information. In low energy
CID, the protonation of a peptide backbone mainly triggers the charge-directed cleavages (i.e.,
cleavage occurs close to a protonated site) and leads to the formation of sequence informative
ions, accompanied by the formation of neutral species yielding minimal structural information,
such as water and ammonium $^{129}$. In low energy CID, peptide backbone dissociation involves a
rearrangement process and primarily generates b- and y-type ions containing the N- and C-
terminals of the precursor ion (the nomenclature is shown in Figure 1-7 $^{134,135}$). Briefly, the
protonation on N-terminal amide nitrogen decrease the amide bond strength, and allows the
nucleophilic attacks of the amide carbonyl group by the N-terminal neighbor carbonyl oxygen,
which leads to the formation of a protonated oxazolone ring complex $^{136-138}$. The dissociation of
the oxazolone ring complex results in the formation of b- and y-ions, as illustrated in Figure 1-8.
When a protein containing multiple protonated sites undergoes CID fragmentation, the presence
of heterogeneous protonation froms (as suggested by mobile proton model) usually resultes in
random cleavages of backbone amide bonds. The type of amino acid residue can also be deduced
via the mass difference of adjacent b ions ($b_n$ and $b_{n-1}$) and y ions ($y_n$ and $y_{n-1}$). In this way,
tandem mass spectrometry can be used for de novo sequencing $^{129,139,140}$. Ideally, large numbers
of b- and y- ions generated from random cleavages throughout all backbone amide bonds could
cover the entire protein or peptide sequence. However, in practice, only partial sequence
coverage is typically obtained due to the fact that CID fragmentation is affected by many factors,
such as amino acid composition and peptide length $^{141}$. In general, CID effectively dissociates
small peptides (<20 amino acid residues) with low charge states (< 3). Other limitations
associated with CID are not being suitable for analytes with multiple basic residues, due to the
side chain amides competing with backbone amides for protonation. Another limitation includes not being suitable for analytes containing labile PTMs, such as phosphorylation, as the labile modification groups are more preferentially dissociated 142.

Figure 1-8. Scheme of CID fragmentation 129.

1.5.3.2 Higher-Energy Collision Dissociation

HCD is a novel fragmentation technique implemented on hybrid Orbitrap MS instruments, which can offer high-quality MS/MS spectra for protein and peptide precursor ions. Compared to conventional CID fragmentation, HCD employs higher collison energy (keV range) and shorter activation time (0.1 ms) to dissociate protein or peptide precursor ions 141. Olsen and co-workers developed a design in which an octopole collision cell was employed for HCD fragmentation. The product ions were ejected from the collision cell, sent back to the C-trap (a curved RF-only QIT), and detected in a high-resolution Orbitrap 143. This novel configuration allows the maintenance of a high trapping efficiency of C-trap and the avoidance of the 1/3 low mass cut-off restriction in the MS2 spectrum. Similar to CID, the primary cleavage in HCD fragmentation
for a protein or peptide precursor ion occurs at backbone amide bonds to form b- and y-type ions. However, the high collision energy enables the further fragmentation of b- ions to produce smaller species, so that the HCD MS2 spectrum is usually predominantly occupied by y- ions.\textsuperscript{144,145} Because of the high accuracy and lack of low mass cutoff MS2 spectra produced by HCD fragmentation, it is widely used in quantitative proteomics.\textsuperscript{146} In Chapter 3 of the dissertation, HCD fragmentation was implemented for label-free and labeling quantitative proteomics.

1.5.3.3 Electron-Transfer Dissociation

Electron capture dissociation (ECD), developed by McLafferty and co-workers in 1998, is another fragmentaiton technique for proteins and peptides.\textsuperscript{147} In the ECD process, a multiply-protonated protein or peptide precursor ion captures a thermal electron and results in the formation of an odd-electron hypervalent intermediate, which subsequently dissociates and induces cleavage at backbone N-C\textsubscript{α} bonds.\textsuperscript{148} However, ECD is technically exclusive to be used in FTICR instruments, which limits its widespread application. An alternative fragmentation method, electron transfer dissociation (ETD), developed by Hunt and co-workers, has recently enabled the implementation of this technique on various types of MS instruments such as linear ion trap, Q-TOF and FTICR; this promotes its universal applications.\textsuperscript{142,149}

The mechanism in ETD fragemtation is similar to that in ECD, in which an electron generated from a reagent ion (usually a fluoranthene radical anion) from a chemical ionization source is transferred to a multiply-protonated protein or peptide precursor ion. The resulting protein or peptide radical cation is usually not stable and dissociation occurs rapidly, driven by radical chemistry.\textsuperscript{150} ETD fragmentation primarily breaks the backbone N-C\textsubscript{α} bonds in a protein or peptide and forms c- and z- type ions. Unlike in CID, ETD is a nonergodic process (i.e. without a
distribution of the internal vibration energy) in which the cleavage occurs in a sequence-independent manner so that labile PTMs, such as phosphorylation and glycosylation, remain integrated after fragmentation. The most prominent feature of ETD is that it enables the localization and characterization of labile PTMs. In addition, ETD is well-suited to study the disulfide bonds which are not typically cleaved in low-energy CID. In some cases, the disulfide cleavages can also be obtained by low-energy CID in negative ESI-mode, and alternatively by high collision energy in HCD, however, with the low ionization efficiency in negative ESI mode and the competition between the backbone cleavage and disulfide bond cleavage usually ends up with MS/MS spectra containing limited fragment ions to elucidate the disulfide mapping. In contrast to CID, many studies have shown that ETD is particularly suitable for large peptides or intact proteins with higher charge states (>2). In practice, CID and ETD are typically applied together to provide complementary information on a protein or peptide analyte. In this thesis, the incorporation of CID and ETD fragmentation in a hybrid MS instrument, Orbitrap Elite, was employed.

1.5.4 Data Processing

Database algorithms such as SEQUEST and MASCOT are essential to employ automation to interpret large numbers of tandem MS/MS spectra generated in high-throughput bottom-up proteomic experiments to aid the identification of peptides and assignment of peptides to corresponding proteins. The fundamental principle of database searching is comparing experimental MS/MS spectra with theoretical fragmentation patterns and using a specific scoring model to sort the match between them. Briefly, the SEQUEST algorithm utilizes the precursor mass to select candidates with identical mass from the database and computes the cross-
correlation (Xcorr) score between the experimental spectra with theoretical spectra predicted from primary sequences of candidate peptides. Meanwhile, MASCOT employs a probabilistic algorithm to score the fit between experimental spectra with sequence database entries. In practice, SEQUEST and MASCOT are complementary in protein identification and are usually employed together to ensure identification with increased confidence and sensitivity. In this dissertation, SEQUEST and MASCOT are combined in the software (i.e., Proteome Discoverer) to aid the interpretation of large datasets in order to identify target proteins from proteomic samples.

1.6 APPLICATION OF MASS SPECTROMETRY-BASED METHODS FOR PROTEIN CHARACTERIZATION

Due to its superb specificity, selectivity and sensitivity, mass spectrometry has shown significant power and potential in the elucidation of protein structures. One interesting research area is the characterization of protein heterogeneities induced by PTMs. With the advanced LC- or CE-based separation techniques, modified proteins or peptides with different physicochemical properties can readily be separated. The subsequent MS analyses are able to precisely differentiate subtle levels of heterogeneities and provide structural details. The following section will focus on the introduction of key sample handling procedures associated with protein characterization and will discuss the determination of PTMs with MS-based methods.

1.6.1 Sample Preparation for MS-based Characterization

One of the major challenges in the determination of PTMs in a therapeutic protein or the discovery of a biomarker from complex proteomic samples is that the target analyte of interest
often is present at low concentration and is masked by bulk drug substances or high-abundance proteins. As a result, how to extract the target protein from complex matrices without causing any sample loss or inducing structural change remains an important consideration in these research areas. Sample preparation has a profound impact on the results of MS analysis. Appropriate sample preparation not only makes samples compatible and optimized for MS-based analysis, but also help to reduce sample complexity to achieve high-quality results. Some key processes in sample preparation for MS-based protein characterization will be introduced.

1.6.1.1 Enrichment/Fractionation

After the protein samples are extracted and isolated from the remaining cell constituents, enrichment or fractionation steps are often required prior to MS analyses. In some cases, when the concentration of the protein of interest is insufficient for subsequent analysis, enrichment processes are exploited to increase the relative abundance of target analytes in the sample and simultaneously to reduce interferences. In other scenarios, when the proteins or peptides in a sample exhibit a wide dynamic range, fractionations are typically used to reduce the sample complexity by dividing proteins or peptides into several groups according to their physicochemical properties, such as size, hydrophobicity and pI. Some commonly used enrichment methods include immunoprecipitation and affinity chromatography, which utilize specific and reversible biological reactions to isolate and enrich proteins of interest from background matrices. For fractionation purposes, 2DE is often applied in proteomic studies to fractionate proteins according to their pIs and molecular weights; each spot or gel section is treated as a specific fraction and undergoes subsequent derivatization, proteolysis and LC-MS analysis. On the other hand, RPLC and IEX are often employed to fractionate peptides.
after proteolysis in order to reduce the sample complexity and to increase overall protein identification in MS analysis \(^{163}\).

### 1.6.1.2 Derivatization

Derivatization strategies offer several benefits for MS analysis including improved protein identification, quantification, and tandem MS/MS spectra interpretation as well as database searching. As a result, these strategies are widely applied in protein characterization and proteomic research areas \(^{164}\). Derivatization methods can be divided into global derivatization, which utilizes stable isotopes such as \(^{15}\)N and \(^{13}\)C to label every single amino acid residue in a protein sample, or site-specific derivatization, which focuses on the chemical modification of specific amino acid residues. In the present section, the discussion will be focused on cysteine alkylation, the most effective site-specific derivatization strategy applied in the characterization of single-protein or proteomic studies \(^{164-166}\).

In the bottom-up approach, the tertiary structure of a protein is usually intentionally destroyed by using reducing reagents such as dithiothreitol (DTT), tris(2-carboxyethyl)phosphine (TCEP), or \(\beta\)-mercaptoethanol to partially or completely break disulfide linkages in the protein \(^{167-169}\). The reduction of protein disulfide linkages can improve the access of enzymes to cleavage sites and assist gas-phase fragmentation. Derivatization on Cys residues is performed after the reduction step to block any free thiol groups in unpaired Cys residues either generated during protein synthesis or produced by disulfide reduction. The strong nucleophilic Cys residues are readily derivatized by a wide array of reagents \(^{165}\). The reactions are based on the irreversible alkylation with iodoacetyl groups containing reagents or reversible exchange with disulfide bonds containing reagents \(^{164,170}\). Iodoacetamide (IAM), iodoacetic acid (IAA), and N-ethylmaleimide (NEM) are
routinely used as irreversible blocking reagents for Cys residues. The advantages of derivatization on Cys residues include, but are not limited to, stabilizing Cys-containing peptides by reducing the potential of disulfide scrambling that may lead to incorrect structural assignment and may complicate the subsequent MS analysis, increasing the hydrophobicity of Cys-containing peptides, thereby improving the chromatographic behavior, and adding a tag that is easily differentiated in tandem MS/MS spectra.

While the aforementioned derivatization strategies focus on the inhibition of unpaired Cys residues, another interesting derivatization method on Cys residues aims at introducing new proteolytic recognition sites for trypsin or Lys-C by converting Cys residues to S-aminoethylcysteine. This derivatization method is particularly valuable for Cys-rich proteins.

1.6.1.3 Proteolysis

Proteolytic digestion with protease is the most crucial sample preparation step because it has a direct impact on what species are subjected to LC-MS analysis to provide structural or quantitative details about the original sample. The selection of a protease or combination of proteases is highly dependent on the specific structural detail or the target analyte, as well as the research goal. Among various types of endoproteases, trypsin is the most prevalent choice for proteomic samples due to its superb proteolytic activity and exceptional cleavage specificity. Trypsin breaks peptide bonds at the C-terminus of basic Lys and Arg residues to generate peptides with optimal sizes and charges for subsequent chromatographic separation and gas phase fragmentation. Other enzymes are also employed to provide peptides with desired structural details. For example, unlike trypsin, Lys-C does not cut the C-terminal of Arg, and thus it can be used to generate peptides with larger size and higher charge states, which is
suitable for tandem MS/MS analysis with ETD fragmentation. Asp-N has high specificity at the N-terminal of Asp, which offers an alternative cleavage site to improve the sequence coverage. PNGase F is widely used to remove the N-linked glycan to simplify the characterization of deglycosylated peptides. Moreover, multi-enzymatic strategies are sometimes preferred to combine the proteolytic activity of different proteases to maximize the digestion efficiency and thereby increase the sequence coverage.

1.6.1.4 Relationship of Sample Preparation to PTM Determination

Sample preparation is the most variable and time-consuming process in protein characterization or proteomic studies. Appropriate sample preparation methods enable a reduction in sample complexity and lead to robust, reproducible and accurate PTM determinations. However, sample preparation methods may complicate the subsequent MS analysis, lead to incomplete determination, or increase the risk of method-induced artifacts. For example, in a study of in-gel digestion and in-solution digestion of kynurenine (KYN) and N-formylkyneurenine (NFK), heavy oxidative modifications on Trp residues were only observed in the former case, indicating that oxidation artifacts were introduced during SDS-PAGE separation. Additionally, significant deamidation artifacts (up to 80%) could be introduced on peptides with deamidation hotspots after the standard overnight trypsin digestion process (37°C and pH 8), compared with those not using such sample pretreatments. Hence, optimized sample preparations are essential to avoid misleading PTM determinations. In Chapter 2 of this dissertation, a stabilizing sample pretreatment strategy is designed for the determination of disulfide linkages in proteins, which are extremely sensitive to scrambling.
1.6.2 Applications of MS-Based Methods in PTM Analyses

In comparison with traditional methods such as Western blot and ELISA, MS-based methods have been tremendously successful in the research area of protein PTMs. Physicochemical property differences between modified proteins or peptides with their unmodified counterparts are readily separated by various LC-based or CE-based methods, whereas the mass increment or decrement induced by PTMs can easily be differentiated with high-resolution mass spectrometry. One of the most attractive features of MS-based methods is that it allows comprehensive measurements of protein PTMs, including localization of modified residues, characterization of the structural details of modifications, and quantification of the abundance changes of each modification. Moreover, MS-based methods are not only powerful to determine PTMs on a purified protein, but also enable mapping of the global PTMs in large protein populations. In the following section, MS determinations of common protein PTMs such as disulfide linkages, deamidation, oxidation and glycosylation will be introduced.

1.6.2.1 Characterization of Disulfide Linkages

Disulfide linkage is a key structural feature to maintain the activity and stability of protein molecules. The characterization of disulfide linkages can yield an understanding of the higher-order structures of proteins. Two important aspects in the characterization of disulfide linkages includes the determination of the status of Cys residues (unpaired or paired) and the determination of the connection patterns between each half-cystinyl constituent involved in a disulfide bridge. Determination of the status of Cys residues is the first necessary step, since unpaired Cys residues are very likely to induce disulfide scrambling, which often complicates and misleads the assignment of correct disulfide linkages. One of the most frequently used
non-MS-based methods for unpaired Cys residue analysis is the chemical derivatization of free thiol groups with Ellman’s reagent (dithionitrobenzoic acid, DTNB) to form a thiol-nitrothiobenzoate complex that has intensive UV absorbance at 416 nm 1,178. Even though Ellman’s reagent allows the rapid measurement of the total content of unpaired Cys residues in a protein, the location of each unpaired Cys residue is still undeciphered. Instead, mass spectrometry provides an alternative approach to accurately identify and localize each unpaired Cys residue.

With respect to the analysis of disulfide linkages, although the top-down approach avoids any sample preparation that may potentially cause disulfide scrambling, tandem MS/MS analysis of intact protein with disulfide bonds preserved often results in limited numbers of fragment ions, most of MS-based disulfide linkage characterizations rely on the bottom-up approach 179. A conventional strategy for disulfide linkage determination is separating Cys residues on different peptide backbones by proteolytic digestion, comparing proteolytic peptide profiles before and after reduction to identify the disulfide-linked peptides and their half-cystinyl constituents, and isolating disulfide-linked peptides for tandem MS experimentation to assign linkage information 180. However, the proteolytic digestion of a non-reduced protein is very difficult, since it is sometimes not possible to locate the proper enzyme to obtain disulfide peptides with a desirable size for chromatographic separation and MS detection. Moreover, highly bridged disulfide linkages prevent the access of enzymes into cleavage sites. Multi-enzyme digestion and partial reduction strategies have been developed to circumvent these problems 181. Selective reduction of some relatively weak disulfide linkages in the protein by low concentration of tris(2-carboxyethyl)phosphine (TCEP) or 2-mercaptoethanol (2-ME) followed by alklyation has been
successfully applied for disulfide linkage assignments in IgG proteins to assist subsequent digestion and MS detection \(^{182,183}\).

Tandem mass spectrometry is an efficient technique to assign the disulfide linkages. Fragment ions generated in a tandem MS experiment not only provide the sequence information of half-cystinyl peptides, but also identify the exact linkage site and connection pattern. CID is routinely used as an ion activation method in tandem MS experiments for peptide sequencing. However, the ergodic nature of the CID fragmentation process leads to the preferential dissociation of the weakest bond in a peptide. Peptide backbone amide bonds are more likely to be cleaved than disulfide bonds in a CID fragmentation process, because higher activation energy is required to dissociate a disulfide bond (40-70 kcal/mol) compared with that for an amide bond (25-40 kcal/mol) \(^{152,184}\). In contrast, the nonergodic ECD/ETD fragmentation process allows cleavages of disulfide linkages in a peptide to yield disulfide-dissociated peptides containing either a thyl radical (-S•) or a free thiol (-SH) group \(^{185}\). Recently, Wu et al. established a strategy to combine the CID and ETD fragmentation modes for online LC-MS analysis of disulfide linkages in therapeutic proteins and antibodies \(^{155}\). Specifically, the precursor ion from a full MS scan is isolated and subjected to CID MS2 and ETD MS2 tandem MS experiments. In CID MS2, disulfide linkages remain intact and peptide backbones are fragmented to generate sequence information, whereas in ETD MS2, disulfide linkages are cleaved to yield partially or completely disulfide-dissociated peptide ions, which are characterized in a subsequent CID MS3 scan employing higher energy to break the peptide backbone for sequencing. This strategy enables the determination of complex and intertwined disulfide linkages, which are difficult to elucidate with a single fragmentation stage.
1.6.2.2 Deamidation Characterization

The MS-based method to identify protein deamidation is straightforward and well-established. The conversion of Asn to Asp or isoAsp in a deamidation process leads to changes in charge, hydrophobicity and pI of a protein. Using these properties, deamidated protein and its unmodified counterparts can be separated with chromatographic-based techniques such as IEX or RPLC, as well as electrophoresis-based methods such as IEF. Furthermore, the conversion of an amine group to a hydroxyl group in deamidation modification introduces a mass increment of 0.984 Da, which is readily detected by mass spectrometry. Additionally, gas-phase fragmentation in CID mode preserves deamidation modifications on protein or peptide backbones, thus enabling the accurate identification of deamidation sites.

1.6.2.3 Glycosylation Characterization

Determination of glycosylation site, elucidation of glycan structures including monosaccharide components and linkages, and quantification of glycosylation occupancy are involved in a comprehensive analysis of protein glycosylation. To date, the most widely applied MS-based glycosylation characterizations have been performed via the bottom-up approach. In general, there are two MS-based bottom-up strategies for glycosylation characterization, glycan release analysis and glycopeptide analysis.

In glycan release analysis, glycans are removed from the backbone of their carrier proteins by enzymatic (e.g., PNGase digestion) or chemical reaction (e.g., β-elimination). Released glycans are further analyzed by MS to obtain composition and linkage information, whereas the carrier protein is digested by endoproteases (e.g., trypsin) and undergo peptide mapping by MS.
to yield the protein sequence. Unfortunately, the site-specific information of each glycan is always lost when glycan release analysis is used \textsuperscript{187}.

In glycopeptide analysis, glycans remain intact whereas peptide backbones are directly cleaved by endoproteases to generate glycopeptides, which are then characterized by mass spectrometry to obtain site-specific glycan structural information \textsuperscript{189}. The mass of a glycopeptide can be determined from the mass of the attached glycan with its carrier peptide. Tandem mass spectrometry incorporated with multiple fragmentation methods is very powerful to fully characterize glycopeptides. CID is the most commonly used ion activation method in a tandem MS/MS experiment. For a glycopeptide, the activation energy required to break a glycosidic bond is much lower than that required for cleavage of a peptide backbone in the CID fragmentation \textsuperscript{190}. In CID positive ion mode, predominant fragment ions resulting from breakages of glycosidic bonds provide information on glycan structures. In contrast, both glycosidic bond cleavages and cross-ring cleavages are observed in CID negative ion mode, which provides information on glycan structures as well as peptide sequences \textsuperscript{191}. An alternative fragmentation method for glycopeptide analysis is ECD or ETD. In ECD/ETD fragmentation mode, cleavages occur primarily on the peptide backbone to yield c- and z- ions to elucidate the peptide sequence, while glycan moieties remain attached to peptide backbones. Moreover, the cross-ring cleavages on glycans further enhance the characterization of various linkage types and branching patterns \textsuperscript{192}. CID and ECD/ETD fragmentation techniques are often combined and used in a multistage activation strategy for the comprehensive characterization of glycopeptide, including the assignment of glycosylation site, elucidation of the composition of each glycan, and quantification of the relative abundance of each glycoform \textsuperscript{191,193}. Additionally, it has been reported that peptide backbones and cross-rings are both cleaved in an HCD fragmentation;
accordingly, the HCD fragmentation technique can also be applied to assist glycopeptide characterization\textsuperscript{194}.

\textbf{1.6.3 Challenges in PTM Determinations with MS-Based Methods}

The determination of protein PTMs by mass spectrometry still faces great challenges. Compared with large-scale protein identification, a comprehensive analysis of protein PTMs is even more difficult. This is because, for the unambiguous identification of a protein in a large-scale proteome, the identifications of several unique peptides are sufficient. However, for large-scale protein PTM determinations, each modified peptide must be detected, and sufficient tandem MS\textsuperscript{n} spectra containing signature ions should be obtained to localize each modification to a single amino acid residue\textsuperscript{195}. The difficulty in the determination of protein PTMs with MS-based methods results from several factors. First, not all types of PTMs lead to mass shift (e.g., Asp isomerization), and characterizations of some PTMs with small mass shifts require high-resolution mass spectrometers. Secondly, most modified forms are present in low stoichiometry and are usually masked by highly abundant proteins in biological matrices. Thirdly, the lability of some PTMs gives rise to challenges in MS and tandem MS/MS analysis. Fourthly, some PTMs, such as glycosylation, may have influences on MS sensitivity by inducing an ion suppression effect\textsuperscript{196}. Fifthly, sample preparations are susceptible to introduce method-induced artifacts. Finally, data interpretation of some PTMs is not fully automatic. For example, the assignment of disulfide linkages in a protein primarily relies on manual interpretation. Despite these difficulties, MS-based methods are sensitive and robust for the rigorous determination of large-scale protein PTMs.
1.7 SUMMARY

Comprehensive characterization of protein is critical to ensure the biological activity and safety of protein therapeutics. Great efforts have been made to explore robust, sensitive and reproducible protocols for determination of protein structure and its PTM forms. This chapter gives an overview of MS-based analytical techniques for protein characterization. In Chapter 2 of the dissertation, an effective approach using online LC-MS analysis with CID and ETD fragmentation has been developed for unambiguous identification of complex cystine knot disulfides in a therapeutic protein. In Chapter 3, we present a protocol incorporated GeLC separation (SDS-PAGE plus nano LC), metabolic labeling method with quantitative mass spectrometry to identify the drug target and study the nature of drug-protein binding. In Chapter 4, a low sheath flow liquid junction interface was designed to integrate commercial CE and ESI-MS instrument for charge heterogeneity profile of intact proteins by CIEF-MS analysis.

1.8 REFERENCES


(50) Chrambach, A.; Baumann, G. In *Isoelectric Focusing*; Catsimpoolas, N., Ed.; Elsevier,
2012; p 278.


(79) Engelhardt, H.; Beck, W.; Schmitt, T. *Capillary Electrophoresis*; Vieweg Press:


(175) Krokhin, O. V; Antonovici, M.; Ens, W.; Wilkins, J. A.; Standing, K. G. *2006*, *78* (18), 6645–6650.


Chapter 2: Complete Disulfide Mapping of Neublastin by Online LC-MS/MS Analysis following Partial Reduction and Alkylation
2.1 ABSTRACT

Neublastin is a naturally secreted homo-dimeric glycoprotein with a complex disulfide structure, including a cystine knot on each chain formed from three intramolecular disulfide bonds and the monomers held together by an intermolecular disulfide bond. Applying the conventional non-reduced multi-enzyme digestion coupled with online LC-MS/MS method resulted in the identification of two intramolecular disulfide bonds, formed from two pairs of closely spaced cysteines (Cys34 and Cys38, Cys100 and Cys102). However, significant disulfide scrambling was observed presumably due to the breakage of the intermolecular disulfide bond (Cys71 – Cys71’) and formation of new bond between two adjacent cysteines (Cys71 and Cys72) during the sample preparation process. To overcome the difficulties, we explored conditions to selectively reduce and alkylate the intermolecular disulfide linkage in Neublastin prior to enzymatic digestion. While scrambling remained, the complete and reproducible disulfide mapping of Neublastin was achieved with the approach involving partial reduction/alkylation, multi-enzyme digestion, followed by LC-MS/MS analysis with multi-fragmentation methods (CID, ETD and CRCID). The protocol described herein offers an approach for determination of the complex disulfide patterns in proteins that are susceptible to disulfide scrambling during conventional sample preparation steps.
2.2 INTRODUCTION

Neublastin, known as artemin or enovin, is a member of the glial cell-derived neurotrophic factor ligand (GFL) family\(^1,2\), involved in improving the survival, proliferation and regeneration of several distinct types of neurons\(^1,3-6\). In animal models, studies have demonstrated the neuroprotective role of Neublastin after peripheral nerve injuries, thus Neublastin was investigated as a potential therapeutic agent for neurodegeneration disease\(^7-10\). While the Phase II clinical trial was very recently found to be unsuccessful\(^11\), Neublastin represents a good model for advanced protein structure analysis.

Recombinant human Neublastin is a homo-dimer glycoprotein with a disulfide cystine knot on each monomer, as illustrated in Figure 2-1. The primary structure of the NBN monomer, contains seven cysteines, including two that are adjacent to each other (Cys71 and Cys72) and two pairs of closely spaced cysteines (Cys34 and Cys38, Cys100 and Cys102). The crystal structure of NBN homo-dimer revealed two cystine knots were formed from the three intra-molecular disulfide bonds on each monomer with an additional inter-molecular disulfide bond formed between Cys71 and Cys71’ linking the two monomers together as shown in Figure 2-1\(^12,13\).
Figure 2-1. Pictorial representation of the three-dimensional structure of the NBN disulfide linkages.

Figure 2-2. Theoretical Asp-N + trypsin digestion of NBN. Theoretical digestion of the NBN dimer. The sites of cysteine residues are colored red, the sites of Arg residues green, and the site of N-glycosylation underlined. The purple arrows are the cleavage positions for trypsin, and the blue arrows are cleavage sites for Asp-N.
Current advances in peptide mapping with high-resolution LC-MS $^{14-19}$ can achieve disulfide linkage mapping. Such analysis is important in higher order characterization, as LC-MS provides a means for monitoring biopharmaceutical product. Determination of the disulfide linkages in cystine knots, however, is still challenging because the highly cross-linked disulfide structure causes limited enzymatic digestion. Recently, a strategy using multi-enzyme digestion in combination with online LC-MS analysis incorporating CID and ETD fragmentation was used by this lab to successfully solve the cystine knot and nested disulfide linkages in a therapeutic protein, recombinant arylsulfatase $^{20}$.

In the present study, the characterization of Neublastin cystine knot was initially attempted using a similar strategy to that published $^{20}$. Although two of the three intra-chain linked disulfides of Neublastin could be mapped, significant scrambling was observed, with loss of the intermolecular disulfide linkage. We then modified the sample preparation using partial reduction and alkylation. With carefully optimized conditions, the labile inter-chain disulfide in Neublastin was reduced with the other cysteine knot disulfides remaining largely intact. Using this approach, the inter-chain disulfide was unambiguously identified by the on-line LC-MS with CID and ETD fragmentation through the determination of the alkylation site on C71 and C71’. The study provides an LC/MS strategy for determining disulfide linkages where scrambling can be significant.
2.3 EXPERIMENTAL

2.3.1 Chemical and Reagents

**Neublastin Samples:** Neublastin was manufactured at Biogen (Cambridge, MA). In Stage I of the project, Neublastin was provided with 4.7 μg/μL in formulation buffer consisting of 10 mM sodium succinate, 75 mM sodium chloride, 100 mM L-arginine HCl, pH 5.5) (Lot 1). In Stage II, Neublastin was provided at a higher concentration 79.7 μg/μL, again in formulation buffer (Lot 2). Samples were aliquoted as 20 μL per vial for Lot 1 and 2 μL per vial for Lot 2, and stored at -80 °C before analysis.

**Reagents:** PNGase F, Asp-N, guanidine hydrochloride, sodium acetate, N-ethylmaleimide (NEM), Tris, HCl, and DMSO were purchased from Sigma-Aldrich (St. Louis, MO). Tris(2-carboxyethyl)phosphine hydrochloride (TCEP), formic acid (FA), trifluoroacetic acid (TFA), HPLC grade water and acetonitrile were obtained from Thermo Fisher Scientific (Fairlawn, NJ). Sequencing grade modified trypsin was from Promega (Madison, WI).

2.3.2 Sample Preparation

**Non-reduced Digestion:** 20 μL Lot 1 Neublastin (4.7 μg/μL) was denatured by 400 μL of 6 M GnHCl in 100 mM Tris-HCl buffer, pH 6.8, at room temperature for 30 min and then buffer exchanged to 100 mM Tris-HCl (pH 6.8) using a 10 KDa molecular weight cutoff filter (EMD Millipore, Bedford, MA) followed by concentration to 1 μg/μL (100 μL). Asp-N (100:1, w/w) together with trypsin (40:1, w/w) was added to the protein solution and incubated at 37 °C for 4 hours. The two enzymes were again added to the sample, and the solution was incubated at 25 °C for an additional 12 hours. For some initial experiments, PNGase F (20:1, w/U) was further
added to the sample, and the mixture of three enzymes used under the above conditions. In all cases, digestion was terminated with addition of 5% formic acid.

For identification of disulfide dissociated peptides, an aliquot of the peptide digest was completely reduced with 10 mM TCEP at room temperature for 60 min and analyzed by nanoLC-MS.

**Partial Reduction:** 25 μL of 6 M GnHCl in 0.75 M sodium acetate buffer, pH 4.6, was added to 1.25 uL Lot 2 Neublastin (79.7 μg/μL). A preparation of 0.1M TCEP in 6M GnHCl/0.75 M sodium acetate buffer was prepared. 1.38 uL of a 20 mM TCEP in 6M GnHCl/0.75 M sodium acetate buffer, pH 4.6, was added to the Neublastin sample solution (final concentration of TCEP at 1 mM), and the solution was incubated at 37 °C for 20 min. 0.71 µL of 2M NEM in DMSO was added to the sample for a final concentration of 50 mM NEM, and the solution was incubated at 37 °C for 60 min in the dark.

**Desalting:** The solution was desalted on an Agilent Technologies (Santa Clara, CA) 1200 series HPLC system equipped with a UV detector. The sample was then manually injected onto a reversed phase Polaris C18-A column (50 × 2.0 mm, 5 μm, 180Å) (Agilent Technologies). Mobile phase A was 0.1% FA in water, and mobile phase B 0.1% FA in acetonitrile. A 60 min isocratic gradient of 2% B was used for desalting at room temperature at a flow rate of 0.8 mL/min. The flow rate was then decreased to 0.2 mL/min, and the fraction collected with a 12 min gradient of 10-80% B for 7 min, followed by an isocratic hold at 80% B for 3 min, and finally, a column reconditioning step from 80-2% B for 2 min. The UV detector was set at 280 nm to monitor the separation.
The collected fraction was lyophilized with a vacuum concentrator from Labconco (Kansas City, MO) and reconstituted to 100 mM Tris-HCl (pH 7.2) to a concentration at 1 μg/μL. The sample was digested with two enzymes as described in the previous non-reduced digestion section in the absence of denaturation and buffer exchange processes.

2.3.3 LC-MS and LC-UV

**LC-MS System for Disulfide Characterization (System 1):** LC/MS analysis was performed on a Thermo Fisher Ultimate 3000 nano-LC pump (Mountain View, CA) and a self-packed C18 column (Magic C18, 200 Å pore and 5 μm particle size, 75 μm i.d. × 15 cm) (Bruker Daltonics Inc., Billerica, MA). The self-packed column was connected to a 4 cm-long uncoated fused silica nanoelectrospray emitter (360 μm o.d., 20 μm i.d., and 10 μm tip i.d.) (New Objective, Woburn, MA) by a 250 μm i.d. Teflon connector (Thermo Scientific, Rockwood, TN). The column was coupled online to a Thermo Fisher LTQ-Orbitrap-Elite-ETD mass spectrometer (San Jose, CA) through a New Objective nanospray ion source. 2 μg of the digested sample (the enzyme-digest) was used per LC-MS analysis. Mobile phase A (0.1% formic acid in water) and mobile phase B (0.1% formic acid in acetonitrile) were used for the gradient consisting of (i) 20 min at 2% B for sample loading and 5 min for desalting at 0.3 μL / min; (ii) linear from 2 to 5% B for 2 min; (iii) linear from 5 to 40% B for 60 min; (iv) linear from 40 to 90% B for 10 min; and finally (v) isocratic at 90% B for 5 min. The flow rate of the column was maintained at 0.2 μL/ min. The LTQ-Orbitrap-Elite-ETD mass spectrometer was operated initially in the data-dependent mode as follows: survey full-scan MS spectra (m/z 300-2000) were acquired in the Orbitrap with a mass resolution of 60,000 at m/z 200, followed by operation in the data-
dependent mode to switch automatically between CID-MS2 (scan 2 at the LTQ), and ETD-MS2 (scan 3 at the LTQ). Briefly, after surveying the MS spectrum from m/z 300 to m/z 2000, subsequent CID-MS2 and ETD-MS2 steps were performed on the same precursor ion with a ±2.5 m/z isolation width. Any inadequate information (assignment) in the CID-MS2 and ETD-MS2 spectra was repeated by targeting the desired ions, e.g. the same precursor but with a different charge state, in order to gain additional linkage information. These targeted approaches were repeated (e.g. targeting multiple charges of a precursor ion or the same disulfide-linked peptide but with different enzymatic cleavages or miscleavages) until the linkage information was determined. In addition, as explained in the Results and Discussion Section, a targeted CID-MS3 after ETD for the ions of interest was further performed as necessary.

**Disulfide Assignment**: The disulfide linkage assignment was achieved with manual inspection of the LC-MS chromatogram, along with tandem MS2 and MS3 spectra. The theoretical masses of an anticipated disulfide peptide at different charge states were calculated from each disulfide dissociated peptide, with NEM alkylation (+125.0476 Da) as the dynamic modification on cysteines. The theoretical mass of the precursor ion was then matched to the observed mass in the LC-MS chromatogram with mass accuracy < 5 ppm for a high abundant ion and < 20 ppm for a low abundant ion. Further confirmation utilized CID-MS2 and ETD-MS2 spectra. As an illustration, for a disulfide peptide P1-P2 (P1 and P2 being the two peptide chains), the molecular weight of disulfide peptide P1-P2 corresponded to the sum of the masses of peptides P1 and P2 with loss of two protons. The assignment of the disulfide linkage was performed by assuming that the Cys residue on P1 was modified by the peptide chain P2. The observed fragment ions in CID-MS2 spectra (b and y type ions) were compared with the theoretical fragmentation pattern with P2 attached. The analysis was then repeated by assuming
that the Cys residue on P2 was modified by the peptide chain P1. In the ETD activation step, the disulfide linkages were dissociated, and the identity of disulfide dissociated peptides P1 and P2 was determined in ETD-MS2 (c and z type ions) without any modification on the Cys residues. If required, the charged reduced species (i.e., the disulfide bond was reduced, whereas peptide P1 and P2 were still linked by noncovalent interaction) in the ETD-MS2 spectrum were isolated and subjected to CID-MS3 fragmentation, and the fragment ions (both b/y and c/z type ions) were searched against the theoretical fragmentations. Additionally, for the disulfide peptide, product ions with internal cleavages (i.e. simultaneous cleavages on both peptide chains) were frequently observed.

**LC-UV-MS System for Routine Monitoring (System 2):** Enzymatic peptides were monitored by a 1200 HPLC pump and a 300SB-C18 HPLC column (3.5 μm particle size, 2.1 mm i.d. × 15 cm) coupled online with a UV detector (G1315D photodiode array) and a 6520 quadrupole time-of-flight (QTOF) mass spectrometer (Agilent Technologies). 20 μg of the enzyme-digested samples was used per LC-UV-MS analysis. Mobile phase A was 0.1% formic acid in water, and mobile phase B 0.1% formic acid in acetonitrile. The separation was carried out by a linear gradient starting from 2% B to 5% in 4 min; increased to 40% B in 80 min; increased to 80% B in 3 min; and then isocratic at 80% B for 5 min. The flow rate was maintained at 200 μL/min. The QTOF mass spectrometer was run in the positive-ion mode with a capillary voltage of 4 KV, gas temperature of 325 °C, drying gas at 11 L/min, nebulizer at 45 psi, and m/z scan range of 350-1800 with an acquisition rate of 7 spectra/sec.
2.4 RESULTS AND DISCUSSION

As noted in the Introduction, our group previously developed a strategy employing non-reduced multi-enzyme digestion followed by LC-MS analysis with CID and ETD fragmentation to determine cystine knot disulfide linkages\textsuperscript{20}. As described below, to map the disulfide linkages in NBN, we first tried to apply the same strategy and were able to successfully solve the linkage of two out of three intra-molecular cysteine knot disulfide bonds. However, the remaining intramolecular and the intermolecular disulfide bond were not successfully determined due to disulfide scrambling. We therefore modified the sample preparation steps, adding partial reduction and alkylation before digestion to protect the inter-molecular disulfide bond. We describe a sample optimization procedure for disulfide linkage analysis of NBN.

2.4.1 Non-reduced Digestion Approach

The disulfide structure of the NBN homo-dimer was described in the Introduction and is illustrated in Figure 2-1 and Figure 2-2. Our initial study explored the possibility of developing an LC/MS method for the analysis of non-reduced proteolytic digested peptides. Generally, native proteins with disulfide bonds intact are resistant to enzymatic digestion\textsuperscript{21}. Thus, in this study, Neublastin was denatured by 6 M GnHCl to enhance the accessibility of enzymes to the designated cleavage sites. We then focused on finding an optimum digestion strategy to generate peptides with the desired sizes to facilitate subsequent LC-MS and tandem MS/MS analysis. Peptides with molecular weights of 1 – 4 kDa are preferred\textsuperscript{22,23}. After examination of the overall structure of NBN, a multi-enzyme digestion platform of trypsin plus Asp-N was selected. As illustrated in Figure 2-2A, the theoretical digest of the NBN homo-dimer with the combination of these two enzymes would yield two peptides, one with two intramolecular disulfide bonds and
the other with two intramolecular plus the single intermolecular disulfide bond, both peptides in the desired molecular weight range.

NBN has a partially occupied glycosylation site at Asn86 which could potentially block proteolytic cleavage and complicate MS analysis. We explored deglycosylation with PNGase F in combination with the above two proteolytic enzymes; however, protein precipitation was observed, likely due to the hydrophobicity of the large peptides. Hence, deglycosylation was not utilized further in this work.

Reduction of disulfide bonds through β-elimination can occur in alkaline pH, potentially leading to disulfide scrambling (in the β-elimination reaction, the attack by hydroxide anion on the α-proton of a cysteine in a disulfide bond leads to the elimination of β-proton and the yield of persulfide and dehydroalanine)\(^{24-27}\). As noted in the Introduction, given the structure of NBN, scrambling is of major concern. As a consequence, it was decided to conduct trypsin plus Asp-N digestion of NBN homo-dimer at pH 6.8, rather than the optimal pH of greater than 8.0.

Simultaneous digestion of the two enzymes at pH 6.8 of the non-reduced homo-dimer and LC-MS/MS analysis with CID and ETD fragmentation led to correct assignment of two intramolecular disulfide bonds (Cys34 – Cys100 and Cys38 – Cys102). However, only trace amounts (1000-fold lower intensity compared to the peptide with two intramolecular disulfide bonds) of the remaining three expected disulfide bonds (two intramolecular Cys7 – Cys72 and an intermolecular Cys71 – Cys71’) were observed. Instead of observing the Cys71 – Cys71’ intermolecular disulfide bond, a variety of highly abundant scrambled forms containing the intermolecular cysteines were found. Figure 2-3 shows an example of a disulfide scrambled peptide (major scrambled peptides were summarized in Table 2-2). The disulfide dissociated ion P1 and P2 in the ETD-MS2 spectrum revealed the two peptides (F\(^{33}\) – R\(^{39}\) and P\(^{66}\) – R\(^{76}\)) were
connected by two disulfide bonds. In addition, the fragment ion with a residual disulfide bond (e.g., c₃ and c₄ on P1, and c₆ on P2) demonstrated that the Cys71 was disulfide linked with Cys34. Among all the peptides with scrambled disulfide bonds, the peptide P₆₆–R₇₆ with an intra-chain disulfide bond between Cys71 and Cys72 was the most abundant form identified from the protein digest (see Figure 2-4 for disulfide assignment). Thus, in contrast to earlier work ²⁰, the use of non-reduced peptide digests was not successful in determining disulfide linkages for NBN, even conducting the enzymatic digestion at pH 6.8.
Figure 2-3. Assignment of disulfide linkages in a scrambled peptide when a non-reduced digestion of NBN with Asp-N and trypsin was used. (A) Precursor ion spectrum of the peptide; (B) ETD-MS2 spectrum of the most abundant 4+ precursor ion (m/z 500.47); (C) CID-MS2 spectrum of the same precursor ion as in (B). Spectra (B) and (C) are measured in the linear ion trap. The peptide sequences with the observed fragment ions are shown in the figure. If the product ion contains a disulfide bond, a red label is used.
Figure 2-4. Assignment of the disulfide linkage in a scrambled peptide with the two adjacent cysteines, C71 and C72. (A) Precursor ion spectrum of the peptide; (B) CID-MS2 spectrum of the peptide of precursor ion of m/z 621.30 (2+). The CID-MS2 spectrum is measured under high resolution in the Orbitrap.

The high level of scrambled disulfide peptides was likely in part a consequence of the facile breakage of one or more disulfide linkages by the sample preparation steps. Individual disulfide bonds in a protein bear different stabilities (e.g. intra and inter-molecular disulfide bonds in IgG\textsuperscript{28}). Due to the conformational energy and level of solvent exposure, the intermolecular disulfide linkage in homo-dimeric NBN is the most susceptible linkage that can be reduced in neutral to alkaline pH. Breaking the relatively unstable intermolecular disulfide bond during enzymatic digestion would generate a reactive thiol group on Cys71 and Cys71'. This would be followed by nucleophilic attacks on the intramolecular disulfide bonds, thereby triggering a
series of thiol-disulfide exchange reactions, ending up with peptides containing scrambled disulfide bonds. Compared to other cysteines, the close proximity position of Cys72 to Cys71 (and C72’ to C71’) would be expected to be the most kinetically favored positions for thiol-disulfide exchange and new bond formation. The adjacent position of the two cysteines may explain the presence of the high abundant scrambled disulfide peptide P$^{66}$ – R$^{76}$ in which Cys71 and Cys72 were held together (Figure 2-4).

Attempts to suppress disulfide scrambling by carrying out non-reduced two enzyme digestion of NBN homo-dimer at pH 6.8 in presence of 5 mM NEM, an alkylating agent that could block free cysteines from thiol-disulfide exchange failed. Peptide P$^{66}$ – R$^{76}$ with the scrambled intra-chain disulfide bond Cys71 – Cy72 was still found, whereas peptides with NEM protection on the cysteines were not detected. This result indicates that the kinetics of the alkylation reaction on free cysteines cannot compete with the rapid thiol-disulfide exchange of the adjacent cysteines.

Since the non-reduced multi-enzyme digestion strategy was not successful, in generating the complete disulfide linkage pattern of NBN from the peptides formed, we modified the sample preparation steps. Clearly, we needed to avoid the cleavage of the labile intermolecular disulfide bond in order to minimize a series of thiol-disulfide exchanges catalyzed by this breakage. The approach explored, involved partial reduction of the NBN dimer under mild conditions, and rapid protection of the resultant free cysteines by alkylation, as described below.
2.4.2 Partial Reduction and Alkylation

The relative stability of individual disulfide bonds in a protein is highly dependent on its conformational energy and solvent exposure level. As discussed above, in the Neublastin homo-dimer, the inter-molecular disulfide is more labile than the intra-molecular linkages involved in a cysteine knot structure. Thus, under a well-controlled reduction condition, it might be possible to initially break the intra-molecular disulfide bond in NBN while preserving the others for cystine knot assignment. We then could seek conditions where the free cysteines might be alkylated prior to thiol-disulfide exchange. We explore this strategy.

To prevent or at least minimize pH-induced disulfide scrambling, the partial reduction and alkylation steps were performed at pH 4.6 (instead of pH 6.8). TCEP and NEM, both with high reactivity at acid pH, were selected as the reduction and alkylation reagent, respectively. The partial reduction was performed under denaturing conditions with 6 M GnHCl to unfold the protein and improve the accessibility of TCEP to each disulfide bond and enhance the subsequent alkylation and digestion efficiency. As described in the Experimental Section, the inter-molecular disulfide bond in homo-dimeric Neublastin was selectively broken with low concentration of TCEP. The reduced sample was then alkylated with an excess of NEM to terminate the reduction and block the unpaired cysteines from further reaction. This partial reduction and alkylation step would allow the proteolytic enzymes to achieve a high digestion efficiency and diminish the labile-disulfide induced scrambling.

Initial attempts of the partial reduction and alkylation reactions were performed under the conditions reported in the literature. Approximately 94 µg denatured Neublastin (12 µM) was incubated with 20 mM TCEP at 37 °C for 60 min, followed by alkylation with 100 mM NEM at 37 °C for 60 min (i.e. NEM: Neublastin = 8300, conc. / conc.). As the Neublastin
homo-dimer contains both glycosylated and non-glycosylated forms, products of partially reduced and alkylated Neublastin are expected to be a mixture of components with different residual disulfide linkages, with and without glycosylation modification.

The partially reduced/alkylated mixture was separated by HPLC (reversed phase) with UV detection at 280 nm. As shown in Figure 2-5A, four peaks were observed, each corresponding to the species either glycosylated or non-glycosylated and reduced fully or partially with NEM modification. The assignment was determined by accurate mass measurement to assess the degree of NEM addition (see Table 2-1). Briefly, peak 1 was observed as the masses of 13,554.50 Da and 13,625.49 Da for the highest intensity ion. The mass difference of 71.01 Da resulted from the known N-terminal Ala truncation during secretion. Since Neublastin is glycosylated, after PNGase F digestion of the collected fraction, the observed masses decreased to 11,202.62 Da and 11,273.64 Da, which agreed well with the predicted molecular masses of 11,202.77 Da and 11,273.85 Da for monomers with 1 NEM and three residual disulfide bonds. Presumably, the single disulfide breakage would occur for the intermolecular bond, the weakest disulfide linkage. Based on the deglycosylated result, peak 2 was assigned to the same non-glycosylated monomer. From the accurate mass and PNGase F treatment, peak 3 consisted of the glycosylated and peak 4 the non-glycosylated forms. Both peaks consisted of all disulfides broken and alkylated with NEM.

With the degree of reaction monitored by the HPLC elution profile, we were able to explore the partial reduction and alkylation conditions to maximize the desirable products peak 1 and 2 and reduce peak 3 and 4 where a high potential existed for disulfide scrambling. In this way, we believed we could preserve the linkage information of the three disulfides involved in the cystine
knot structure by protecting the reduced cysteines of the inter-molecular disulfide bond with NEM alkylation to minimize disulfide scrambling.

The reduction conditions of TCEP concentration (1 – 20 mM) and reaction time (1 – 60 min) were investigated in a search for optimum conditions. As shown in Figure 2-5B, the relative abundance of peak 2 was significantly increased by decreasing the TCEP concentration from 20 mM to 1 mM, for a fixed reaction time of 60 min. Additional improvement was achieved by reducing the reaction time from 60 min to 20 min (Figure 2-5C). A further decrease in the reduction time (below 20 min) did not lead to significant enrichment of desired products. Additionally, the alkylation condition was explored by performing NEM concentration from 20 to 200 mM at 37 °C for 60 min, while maintaining the reduction condition at 1 mM TCEP and 20 min. Alkylation with 50 mM NEM (approximately 20 molar excess to Cys) improved the desired peak 2 yield (~90%) and was therefore selected as the optimum alkylation condition Figure 2-5D).
Figure 2-5. Optimization of conditions for partial reduction and alkylation. (A) LC-UV (280 nm) analysis of intact NBN homo-dimer (1) and NBN (12 µM) partially reduced by 20 mM TCEP at 37 °C for 60 min, followed by alkylation by 100 mM NEM at 37 °C in the dark for 60 min (2); (B) LC-UV (280 nm) analysis of NBN (12 µM) reduced by different concentrations of TCEP for a fixed time of 60 min; (C) LC-UV (280 nm) analysis of NBN (12 µM) reduced by the same concentration of TCEP (1 mM) for different time periods; (D) LC-UV (280 nm) analysis of NBN (12 µM) reduced by 1 mM TCEP for 20 min, followed by alkylated by different concentration of NEM for a fixed time of 60 min. The intact NBN is a mixture of non-glycosylated and glycosylated forms. The identities of major components in the four peaks are as follows: 1. glycosylated monomer with 1 NEM; 2. non-glycosylated monomer with 1 NEM modification and a minor amount of glycosylated dimer; 3. a mixture of glycosylated monomers with different numbers of NEM modifications (7 NEMs is the most abundant form); 4. non-glycosylated monomer with 7 NEM modifications. The concentration of TCEP and the reduction time are indicated in the figure. The percentage represents the relative abundance of peak 2, calculated by the area of peak 2 divided by the sum of the area of the four peaks in the chromatogram. HPLC conditions are described in the Experimental Section.
Table 2-1. Identity of major components in the four RPLC fractions.

<table>
<thead>
<tr>
<th>Peak</th>
<th>Mass observed(^b)</th>
<th>Numbers of NEM modifications</th>
<th>Numbers of disulfide linkages</th>
<th>Identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1356.4557(10+) &amp; 1363.5568 (10+) &amp; 1019.4280(11+) &amp; 1025.8845 (11+) (^b)</td>
<td>1</td>
<td>3</td>
<td>Glycosylated monomers</td>
</tr>
<tr>
<td>2</td>
<td>1245.6372(9+) &amp; 1253.5291(9+)</td>
<td>1</td>
<td>3</td>
<td>Non-glycosylated monomers</td>
</tr>
<tr>
<td>3</td>
<td>842.7594(17+) &amp; 846.9964 (17+) &amp; 1197.0015(10+) &amp; 1204.1046 (10+) (^b)</td>
<td>7</td>
<td>0</td>
<td>Glycosylated monomers</td>
</tr>
<tr>
<td>4</td>
<td>789.2715(15+) &amp; 803.0063 (15+)</td>
<td>7</td>
<td>0</td>
<td>Non-glycosylated monomers</td>
</tr>
</tbody>
</table>

\(a\) Most abundant isotopic peak  
\(b\) Following PNGase F digestion

It is interesting to note in Figure 2-5B and C that, as the TCEP concentration and reduction time decreased, the desired product of non-glycosylated monomer with 1 NEM addition (peak 2) apparently increased, whereas the glycosylated monomer with 1 NEM modification did not increase (peak 1). This result may have been due in part to the glycosylated Neublastin homodimer being more chemically resistant to reduction compared to the non-glycosylated forms. In addition, peak 2 was a mixture of the glycosylated dimer and non-glycosylated monomer. This was evident by the observed MW at 24,580.19 Da and 24,651.16 Da for the glycosylated homodimer and at 11,201.66 Da and 11,272.69 Da for the non-glycosylated monomer. The increased relative abundance of peak 2 also resulted from less glycosylated dimer being reduced with a lower concentration of TCEP and shorter reduction time. It should be noted that the remaining non-reduced homodimer could still potentially lead to disulfide scrambling, but at a much lower
level under the optimized partial reduction and alkylation condition than the previous harsher conditions. The MS result was applied to monitor the extent of the remaining non-reduced dimer for optimizing the partial reduction condition.

Using the optimized conditions of partial reduction and alkylation, the next step was to collect peak 2 and conduct accurate LC/MS analysis with ETD and CID to determine the disulfide linkages. The collected fraction should have the preserved cystine knot along with the molecular tag of NEM on the reduced intermolecular cysteine.

With the protection of inter-molecular disulfide bond, we then applied the established cystine knot characterization strategy to successfully obtain the complete disulfide mapping of Neublastin. Multi-enzyme digestion of partially reduced and alkylated Neublastin leads to the formation of a peptide containing two intra-molecular disulfide bonds and another peptide containing one disulfide bond and one NEM modified cysteine.

Figure 2-6 shows a comparative base peak chromatogram of the digest before and after full reduction. In Figure 2-6B, the cysteine containing peptide with one NEM modification can be identified. A doublet peak (at 20.4 and 20.8 min) with identical mass (1,367.64 Da) was assigned to the peptide with one NEM modification. Alkylation with NEM introduced a new chiral center at the thiol-ether, leading to formation of stereoisomers for the cysteine containing peptides, which are separated by RPLC, resulting in the doublet peak 39. The exact site of NEM modification (at one of the two adjacent cysteines) was pinpointed by MS/MS analysis. As seen in Figure 2-7, the precursor ion of 684.82 (2+) and 456.88 (3+) were fragmented by CID-MS2. The critical fragment ions generated from backbone cleavage between the two adjacent cysteine
residues allowed determination of the site of NEM modification. In Figure 2-7B and 2-7C, the \( y_4 \) (m/z 529.5) and \( y_5 \) (m/z 632.4) fragment ions indicated Cys72 is an unmodified free cysteine, while \( b_4 \) (m/z 412.2) and \( b_6 \) (m/z 737.3) indicated one NEM modified cysteine at Cys71. Additionally, the internal cleavage ion (m/z 325.1, 1+) generated from cleavage of both N- and C-terminal side further confirmed the NEM modification at Cys71.

A doublet peak was also observed at 13.9 and 14.4 min in the peptide map with partial reduction (Figure 2-6A), corresponding to the disulfide peptide with one intra-molecular disulfide linkage and with one NEM adduct. The assignment of this peptide is shown in spectra of Figure 2-8. Multiple charge states of this disulfide peptide were observed, with the triple charge of highest intensity. The accurate mass of the observed precursor ion agrees closely with the predicted mass calculated from peptide G\(^6\)CR\(^8\) and peptide P\(^{66}\) – R\(^{73}\) with one NEM modification (mass accuracy of 0.2 ppm). The precursor ions of this peptide at different charge states (2+ to 4+) were subjected to MS/MS analysis. CID and ETD fragmentations were observed for the peptide in charge state 4+, allowing assignments to be made. In Figure 2-8B, the disulfide linkage was broken by the ETD fragmentation and the resultant two disulfide dissociated ions P1 and P2, along with the c and z fragment ions verified the sequence of the two associated peptides. In the target CID-MS2 analysis (Figure 2-8C), the backbone cleavage between Cys71 and Cys72 was observed. The fragment ion \( y_2 \) of m/z 305.7 (2+) and 610.3 (1+) confirm the linkage of Cys7 and Cys72. The mass differences of the diagnostic fragment ions \( y_2 \) and \( y_3 \) at m/z 305.7 (2+) and 419.7 (2+), respectively, are exactly the same as an NEM modified cysteine, suggesting the NEM modification site was on Cys71. Further, no trace level of fragment ions related to NEM modified Cys72 was detected. Additionally, the CID fragmentation of this peptide was performed in the Orbitrap MS for two different charge states at
m/z 337.40 (4+) and m/z 673.79 (2+). The critical fragment ions $y_2$, $y_3$ (shown in Figure 2-8E) and $b_6$, $b_7$ (shown in Figure 2-9) observed under high resolution CID-MS2 provide more convincing evidence for the assignment of the intramolecular disulfide linkage. This result is in agreement with that obtained from the NEM alkylated peptide P$^{66-76}$ observed in post-digestion reduction peptide map, confirming that the intermolecular disulfide linkage in NBN homo-dimer was Cys71 – Cys71’ and that the intramolecular disulfide linkage was located between Cys7 and Cys72.

**Figure 2-6.** Comparison of the peptide map of NBN digested with Asp-N and trypsin with (A) partial and (B) complete reduction with TCEP. (A) The NBN dimer was partially reduced/alkylated and digested with Asp-N and trypsin; (B) Half of the digest from (A) was then reduced with 10 mM TCEP at room temperature for 1 hour for complete reduction. The peaks of peptides with expected disulfide linkages are labeled in red arrows, and the scrambled disulfide peptides are marked with blue arrows. Only disulfide peptides and cysteine containing peptides are labeled in (A) and (B). The unlabeled peaks are non-cystine containing peptides.
Figure 2-7. Assignment of the NEM modification site on peptide P^{66–76} from the completely reduced digest of NBN. (A) Precursor ion spectrum of the peptide P^{66–76} with one NEM modification; (B) CID-MS2 spectrum of the precursor ion from (A); (C) CID-MS2 spectrum of the same peptide with a different charge state (3+). Spectra (B) and (C) are measured in the ion trap.
Figure 2-8. Assignment of disulfide linkage C7-C72 with NEM alkylation on C71. (A) Precursor ion spectrum of the peptide linked by C7-C72; (B) ETD-MS2 spectrum of the precursor ion of m/z 337.40 (4+); (C) CID-MS2 spectrum of the same precursor ion. (D) and (E) are measured in the linear ion trap. (D) and (E) are, respectively, the ETD-MS2 and CID-MS2 spectra of the same precursor ion measured in the Orbitrap. The sequences with the observed fragment ions are shown in the figure. The critical ions $y_2$ and $y_3$ on P2 indicate the site of alkylation was on Cys 71.
Figure 2-9. CID-MS2 of the peptide with one disulfide linkage (Cys7 – Cys72) and one NEM modification (on Cys71). The spectrum is measured in the Orbitrap.

The remaining two intra-chain disulfide linkages were also identified and the assignments are present in Figure 2-10, with the expected linkages (Cys34 – Cys100 and Cys38 – Cys102) shown in the accurate precursor mass measurement (Figure 2-10A), and corresponding ETD and CID spectra, (Figure 2-10B and C) respectively. In Figure 2-10A, the observed monoisotopic mass of the precursor is in agreement with the predicted mass of the peptide F<sup>34</sup> – R<sup>39</sup> and L<sup>95</sup> – G<sup>104</sup> linked by two disulfide linkages (mass accuracy of 0.5 ppm). In the ETD-MS2 spectrum (Figure 2-10B), the high abundant disulfide dissociated ions P1 and P2, as well as the c and z fragment ions formed by peptide backbone cleavages indicate the sequence of the two peptides (F<sup>34</sup> – R<sup>39</sup> and L<sup>95</sup> – G<sup>104</sup>). The backbone cleavages within the ring structure on the partially disulfide dissociated species define the exact linkage pattern. Fragment ions c<sub>2</sub>, c<sub>4</sub>, z<sub>3</sub> and z<sub>5</sub> on P1 indicate that the disulfide linkage is cleaved at the C-terminal of the peptide, yielding a free cysteine.
residue on Cys38, whereas the Cys34 is still linked to P2 via a disulfide bond. However, no fragment ion on P2 could define the exact position of the P1 attachment (either on Cys100 or Cys102), presumably due to the low backbone fragmentation efficiency in ETD inside a ring structure. The precursor ion at a lower charge state (2+) was selected for CID-MS2 experiment since CID more effectively dissociates low charge state species (z<3+) [41]. In CID-MS2 spectrum (Figure 2-10C), the two disulfide linkages remain intact whereas the fragment ions from backbone cleavages lead to the proposed structure. The signature fragment ion m/z 855.5 (1+), obtained from the internal cleavages within the ring structure (cleavage of Cys34-Ser35 on P1 and Cys100-Gly101 on P2) define the disulfide linkages C34-C100 and C38-C102. The accurate mass and charge state of the signature fragment ion was further confirmed by a high resolution MS2 spectrum measured in the Orbitrap (data not shown). In summary, the peptide sequence analysis using both ETD and CID fragmentation confirmed the two expected intramolecular disulfide linkages Cys34-Cys100 and Cys38-Cys102 in Neublastin.
Figure 2-10. Assignment of disulfide linkages C34 - C100 and C38 - C102 for non-reduced NBN digested with Asp-N and trypsin. (A) Mass and charge of peptides linked by C34 - C100 and C38 - C102 disulfides; (B) ETD-MS2 spectrum of the most abundant 3+ precursor isotopic ion (m/z 550.56); (C) CID-MS2 spectrum of the same peptide with a different charge state (2+). Spectra (B) and (C) are measured in the ion trap of the Thermo Scientific Elite mass spectrometer.

As we have noted, some non-reduced dimer remain under the partial reduction conditions, as evident by the observation of a disulfide-linked peptide with Cys71 – Cys71’ and two Cys7 – Cys72 disulfides, as shown in Figure 2-11. The fragment ions P1+P2, and P1+2P2 in the ETD and CID-MS2 spectra indicate the presence of the expected intra-molecular and intermolecular disulfide bonds in this peptide. Although these disulfides have linkages in agreement with the
crystal structure, the non-reduced molecule could still be susceptible to form the kinetically favorable scrambled disulfides between two adjacent cysteines after enzymatic digestion (i.e. after breaking the cystine knot)⁴⁰. As described earlier, the reason for not pushing to fully reduce all the dimer molecule was that once the reaction was slightly stronger than the chosen conditions, NBN was observed as having >2 NEM on the monomer, as well as NEM-modified to other reduced cystine knot cysteines after digestion, as illustrated in Table 2-4. On the other hand, if even milder reaction conditions were used, the amount of non-reduced dimer could increase, thus triggering the potential formation of higher amount of kinetically favorable Cys71 – Cys72 scrambled disulfide formation.

Applying partial reduction and alkylation, followed by multi-enzyme digestion, and orthogonal LC-MS analysis with both CID and ETD fragmentation methods, we successfully determined the highly bridged structure formed by seven disulfide linkages in Neublastin. Digestion of the partially reduced, NEM alkylated monomer with trypsin and Asp-N led to the cleavage of the cystine knot structure and separation of three intramolecular disulfide linkages. As a result, a disulfide peptide containing two intertwined intramolecular disulfide linkages formed by closely spaced cysteines (Cys34 and Cys38, Cys100 and Cys102) were identified. More importantly, the missing information on the linkage of the third intramolecular disulfide linkage and the intermolecular disulfide linkage were unequivocally determined through our partial reduction/ alkylation approach, leading to a well characterized disulfide peptide containing a single intramolecular disulfide linkage (Cys7 – Cys72) and one NEM modified cysteine (Cys71).
Figure 2-11. Assignment of disulfide linkages C71-C71’ and C7-C72. (A) Precursor ion spectrum peptides linked by C71-C71’ and C7-C72; (B) ETD-MS2 spectrum of the precursor ion with the most abundant m/z 525.74 (6+); (C) CID-MS2 spectrum of the same precursor ion. (B) and (C) are measured in the Orbitrap. The sequences with the observed fragment ions are shown in the figure.
We have seen that, with the optimized partial reduction and alkylation conditions, we were able to selectively break the intermolecular disulfide bond and alkylate the reduced cysteines with NEM. Compared to the conventional non-reduced digestion approach, dramatic enhancement (~1000-fold) were found for the peptides with the expected intermolecular disulfide bond which allowed us to unequivocally observe the complete disulfide mapping of Neublastin. However, disulfide scrambling was still observed after digestion of the NEM-modified Neublastin monomer (Table 2-3 and 2-4). The presence of peptide P^{66} – R^{73} and P^{66} – R^{76} with a scrambled disulfide bond (Cys71 – Cys72) in the digest indicated that the thiol-disulfide exchange between the adjacent cysteines was too rapid, making it not possible to completely eliminate scrambling.

Nevertheless, our studies indicate that we were able to reproducibly measure the disulfide linkages in Neublastin with the established partial reduction/alkylation plus multi-enzyme digestion protocol. The reproducibility test was assessed by two independent sample preparations and four consecutive LC-MS runs. A non-cysteine containing peptide, D^{46} – R^{66} is taken as an internal standard. The peak area of each peptide with expected disulfide linkages at all charge states were summed and corrected by the peptide D^{46} – R^{66}, and the results are presented in Table 2-2 for the peptides with expected disulfide bonds, and in Table 2-3 and Table 2-4 for peptides with scrambled disulfide bonds. In all cases, the small %CVs indicate that the method can be an effective LC-MS based tool to determine the expected disulfide linkages and control the quality of Neublastin product.
Table 2-2. Reproducibility peptides with expected disulfide linkages

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Average ratio</th>
<th>Standard Deviation</th>
<th>% CV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4.34E-03</td>
<td>6.66E-05</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>3.96E-03</td>
<td>2.03E-04</td>
<td>5.1</td>
</tr>
<tr>
<td></td>
<td>1.34E-04</td>
<td>7.76E-06</td>
<td>5.8</td>
</tr>
<tr>
<td></td>
<td>2.75E-03</td>
<td>1.21E-04</td>
<td>4.4</td>
</tr>
</tbody>
</table>

The ratio are obtained from the peak area of the corresponding peptide detected at all charge states divided by the peak area of peptide D^{46} – R^{65}.

Table 2-3. Reproducibility peptides with scrambled disulfide linkages.

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Average ratio</th>
<th>Standard Deviation</th>
<th>% CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>PVSQPCRPTR</td>
<td>8.63E-03</td>
<td>2.12E-04</td>
<td>2.5</td>
</tr>
<tr>
<td>FCSGSCR</td>
<td>3.74E-03</td>
<td>1.51E-04</td>
<td>4.0</td>
</tr>
<tr>
<td>PVSQPCRPTR</td>
<td>6.22E-04</td>
<td>5.74E-06</td>
<td>0.9</td>
</tr>
<tr>
<td>LSATAACGLG</td>
<td>4.73E-04</td>
<td>1.37E-05</td>
<td>2.9</td>
</tr>
<tr>
<td>LSATAACGLG</td>
<td>1.22E-03</td>
<td>3.30E-05</td>
<td>2.7</td>
</tr>
</tbody>
</table>

For peptide with multiple scrambled disulfide linkages (n>1), only one connectivity pattern is shown above. The peak areas of all scrambled disulfide linkage patterns are summed. The ratio are obtained from the peak area of the corresponding peptide detected at all charge states divided by the peak area of peptide D^{46} – R^{65}.
Table 2-4. Reproducibility peptides with 2 NEM modifications.

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Average ratio</th>
<th>Standard Deviation</th>
<th>% CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEHM</td>
<td>1.21E-03</td>
<td>1.03E-05</td>
<td>0.9</td>
</tr>
<tr>
<td>NEHM</td>
<td>2.98E-04</td>
<td>6.00E-06</td>
<td>2.0</td>
</tr>
<tr>
<td>NEHM</td>
<td>2.02E-03</td>
<td>1.39E-04</td>
<td>6.9</td>
</tr>
<tr>
<td>NEHM</td>
<td>1.36E-02</td>
<td>1.59E-04</td>
<td>1.2</td>
</tr>
<tr>
<td>NEHM</td>
<td>7.69E-03</td>
<td>2.82E-04</td>
<td>3.7</td>
</tr>
<tr>
<td>NEHM</td>
<td>1.08E-02</td>
<td>6.02E-04</td>
<td>5.6</td>
</tr>
</tbody>
</table>

2.5 CONCLUSION

In the present study, for the first time, the complete cystine knot structures of a therapeutic protein, Neublastin has been elucidated. The challenging for the disulfide linkage mapping of this protein arises from several factors. The highly bridged structure formed by two cystine knots and an intermolecular disulfide linkage across them has significantly reduced the efficiency of enzymatic digestion and the gas-phase dissociation to provide any structural details. The seven disulfide linkages exhibit different stability and the labile bond is susceptible to break during sample preparation steps and induce disulfide scrambling. The kinetically favored thiol-disulfide exchange between two adjacent cysteines can result in various artifacts and mistaken the assignment of correct disulfide linkages.
To overcome these problems, we have developed a robust and reproducible protocol to selectively break the most labile intermolecular disulfide linkages, rapidly protect the reduced cysteines and accurately assign the complete disulfide mapping by in-depth LC-MS analysis. The established strategy incorporating optimized partial reduction/alkylation techniques, multi-enzyme digestion and LC-MS analysis is powerful to solve the disulfide mapping of protein with complex and intertwined disulfide linkages formed by closely spaced or adjacent cysteines. More importantly, our method is particularly valuable for characterization of proteins containing labile disulfide linkages that are sensitive to method-induced disulfide scrambling.

2.6 ACKNOWLEDGMENTS

I would like to thank to Dr. Li Zang and Dr. Andrew Weiskopf at Biogen Idec for providing samples and sharing the preliminary results on partial reduction.

2.7 REFERENCES


Chapter 3: Chemical Proteomic-Based Approach for Identification and Characterization of a Drug Target (Protein) Using Mass Spectrometry

Co-authors’ work in this chapter: The SILAC experiments were performed by Karyopharm Therapeutics Inc. and the samples were provided. The SDS-PAGE, in-gel digestion and LC-MS/MS analysis were done by Siyuan Liu in Prof. Karger’s group.
3.1 ABSTRACT

The potential protein target, p21 protein kinase 4 (PAK4), to the candidate small molecule drug PAK4 allosteric modulators (PAMs; KPT-7523) was identified by using KPT-7523 as bait to isolate the target protein from a sea of background proteins. The identification strategy was demonstrated by combined SILAC labeling, using heavy and light labeling in both forward and reverse order, with the power of mass spectrometry (MS) for quantitative comparisons. The binding nature between the drug and its target was further characterized by MS and shown to be a result of non-covalent association.

3.2 INTRODUCTION

Drug target identification involves the determination of the protein(s) in a cell specifically associated with a drug candidate (often a chemically synthesized small molecule), an important process in the analysis of drug function and mechanism of action. A wide range of techniques are available to identify protein targets, such as expression cloning, biochemical suppression, and protein microarrays. Recently, affinity pull-down has been used to isolate target protein(s) using the drug or an antibody to a drug’s tag as a bait from large mixtures of background proteins. Mass spectrometry then can be used to identify the isolated protein(s). In order to distinguish true affinity from non-specific interactions, isotopically labeled approaches are now being used.

Two types of affinity pull-down strategies are typically employed in chemical proteomics - activity-based profiling, and drug-centric. Activity-based protein profiling focuses on the design of a small molecule drug with an active functional group that allows the isolation of a
particular protein or class of proteins from cells. On the other hand, the drug-centric chemical proteomic approach focuses on the target discovery and characterization of the drug mode of action specific for the drug. Both approaches involve (1) immobilization of the small molecule drug on an affinity matrix (e.g., Sepharose beads), (2) incubation of the immobilized small molecule ligand with protein extracts, (3) wash to remove unbound proteins, (4) recovery bound proteins from the affinity matrix and (5) finally followed by proteomic-wide MS approach for quantitative analysis (i.e. quantitative proteomics).

Quantitative proteomics has been used as a powerful tool to study the drug target and to understand the related cellular functional networks. Generally, MS-based quantitative proteomics can be performed by labeling and label-free methods. To eliminate the potential bias from sample preparation, the labeling method conducted at the beginning of sample preparation is often used, such as metabolic labeling technique, i.e. stable-isotope labeling by amino acid in cell culture (SILAC). In SILAC experiments, the entire proteome is metabolically labeled with either natural (light) or isotopically substituted (heavy) essential amino acids (e.g., lysine or arginine). In this regard, the drug treated sample group could contain cells from heavy amino acids, and the control (untreated) group could contain cells from light amino acids. The two groups are mixed together in a pre-defined ratio (e.g., 1:1 ratio) followed by affinity extraction and then quantitative LC-MS analysis. Proteins with a large ratio difference as measured by quantitative LC-MS analysis (i.e., far from 1:1 ratio) are likely the potential drug target proteins. Thus, the combination of SILAC, affinity purification and quantitative MS can provide a robust approach to identify the protein target to the small molecule drug.

In this study, a novel benzyodihydrofuranyl derivative drug developed by Karyopharm was rationally designed containing an α, β-unsaturated amide structure moiety, which could either
covalently or non-covalently interact with free cysteines of the target protein $^{14,15}$. As described in this chapter, the SILAC metabolic labeling method (performed by Karyopharm) was utilized to identify PAK4 as the potential target protein to the small molecule drug with high confidence. In addition, using mass spectrometry to obtain high sequence coverage of PAK4, we concluded the small molecule was not covalently bound to any of the cysteine residues of the protein (either free Cys or disulfide bonded forms) or any particular amino acids of PAK4, which indirectly indicated that the drug was non-covalently bound to the potential target protein PAK4.

3.3 EXPERIMENTAL

3.3.1 Chemicals and reagents

Dithiothreitol (DTT), iodoacetamide (IAA), calcium chloride, ammonium bicarbonate were purchased from Sigma-Aldrich (St. Louis, MO). Sodium dodecyl sulfate (SDS) was obtained from GE Healthcare (Pittsburgh, PA), and Odyssey® One-Color Protein Molecular Weight Marker from LI-COR (Lincoln, NE). NuPAGE® Novex® 4-12% Bis-Tris gels, NuPAGE® LDS sample buffer (4X), NuPAGE® MES SDS running buffer (20X), and SimpleBlue™ SafeStain were purchased from Life Technologies (Carlsbad, CA). LC-MS grade water, acetonitrile and formic acid were from Thermo Fisher Scientific (Fairlawn, NJ). Sequence grade modified trypsin was from Promega (Madison, WI).

3.3.2 Samples

All samples used in this chapter were provided by collaborators at Karyopharm Therapeutics Inc. (Newton, MA).
The SILAC experiments were performed at Karyopharm under the guidance of Monica Schenone, the Broad Institute. For 2-plex SILAC labeling, the normal arginine and lysine in MS-751 cell cultures were replaced by L-arginine: HCl ($^{13}$C$_6$, $^{15}$N$_4$) and L-lysine: 2HCl ($^{13}$C$_6$, $^{15}$N$_2$), yielding a +10 Da and +8 Da mass difference, for the two amino acid residues respectively. The labeled MS-751 cells were cultured for at least five generations to ensure the full incorporation of labeled arginine and lysine. The lysates from the heavy or light labeled cells were collected and incubated with either the free small molecule drug KPT-7523 or DMSO, respectively, followed by incubation with drug bound beads overnight. Samples from heavy and light labeled cells were mixed in 1:1 ratios, washed and either digested directly on-beads or released from the beads, and fractionated by SDS-PAGE for subsequent in-gel digestion.

In the target protein characterization experiment, recombinant human PAK4 protein was either treated with DMSO or KPT-7523 drug, respectively.

3.3.3 SDS-PAGE

In all cases, prior to loading, 30 µL of the concentrated sample was mixed with NuPAGE® LDS sample buffer (4X), 1M DTT and deionized water, and incubated at 90 °C for 10 min for denaturation. The mixture was then loaded onto a gel lane of a NuPAGE® Novex® 4-12% Bis-Tris gel, along with the molecular marker on a separated gel lane. SDS-PAGE was run at 200 V for 35 min with NuPAGE® MES SDS running buffer (1X). After triple wash with deionized water, the gel bands were visualized by SimplyBlue™ SafeStain containing Coomassie® G-250, de-stained in deionized water overnight to remove excessive color to obtain a clear background for photography.
3.3.4 Enzymatic digestion

For in-gel digestion of the proteins recovered from the beads, the entire gel lane was cut into 5-6 sections, and each section was further chopped into 1 mm × 1 mm cubes and transferred to a clean Eppendorf tube. For in-gel digestion of the single protein sample, the gel bands corresponding to the PAK4 protein were excised and chopped. The residual Coomassie® G-250 dye was removed by dehydration of the gel cubes in acetonitrile, followed by rehydration in 100 mM ammonium bicarbonate buffer. The dehydration and rehydration were repeated for 3 cycles. The de-stained gel cubes were incubated with 10 mM DTT in 100 mM ammonium bicarbonate buffer at 56 °C for 30 min for protein reduction, followed by incubation with 55 mM IAA in 100 mM ammonium bicarbonate buffer at 25 °C for 60 min in the dark for alkylation. The gel cubes were washed with 100 mM ammonium bicarbonate and dried with a vacuum concentrator. The dried gel cubes were saturated with digestion buffer containing 12.5 ng/µL trypsin, 5 mM calcium chloride and 25 mM ammonium bicarbonate at 4 °C for 30 min. The supernatant was removed and digestion buffer without trypsin was added to cover the gel cubes. The tubes were placed in an incubator at 37 °C overnight with gentle shaking. Acetonitrile was added to dehydrate the gel cubes and recover tryptic peptides. The supernatant was collected and dried with a vacuum concentrator. The digest was stored at -80°C and was reconstituted in 8 µL 0.1% formic acid prior to LC-MS analysis.

3.3.5 LC-MS

The LC-MS experiments were performed on an Ultimate 3000 UPLC system (Thermo Fisher Scientific, Mountain View, CA) coupled to a Q-Exactive mass spectrometer (Thermo Fisher Scientific, San Jose, CA) through a Nanospray Flex™ ion source (Thermo Fisher Scientific, San Jose, CA). A self-packed C18 column (Magic C18, 3 µm, 200 Å, 75 µm i.d. × 20 cm) (Bruker
Daltonics Inc., Billerica, MA) was used for separation. The column was connected to a 4 cm-long uncoated fused silica nanoelectrospray emitter (360 μm o.d., 20 μm i.d., and 10 μm tip i.d.) (New Objective, Woburn, MA) by a 250 μm i.d. Teflon connector (Thermo Scientific, Rockwood, TN). Mobile phase A was 0.1% FA in water, and mobile phase B was 0.1% FA in acetonitrile. 5 μL of the digest was desalted with an isocratic gradient of 2% B for 60 min at a flow rate of 350 nL/min. Linear gradient elution (200 nL/min) was performed as follows: (i) 2 to 32% B for 60 min; (ii) 32 to 95% B in 10 min; (iii) isocratic at 95% B for 7 min; (iv) 95% to 2% B in 1 min and isocratic at 2% B for 2 min.

The heated capillary temperature for the ion transfer tube was set at 285 °C, the S-Lens RF level was fixed at 55, and the nanospray voltage was generated with an electrospray potential of 2.1 kV. The mass spectrometer was operated in the data-dependent scan mode. The full MS survey scan was from m/z 380 to 1650 with a mass resolution of 70,000 at m/z 400 and an automatic gain control (AGC) target value of 1E6. The most abundant 20 precursor ions were isolated with a width of ± 2.5 m/z and dynamic exclusion of 30s, and analyzed by HCD-MS2 (with 28% normalized collision energy). The MS2 scan was performed in the Orbitrap with a resolution of 17,500 at m/z 400 and an AGC target at 1E5.

3.3.6 Data analysis

For the protein identification and quantitation from the SILAC data, the raw data was combined and search against the Uniprot human database with Sequest HT engines (Thermo Fisher Scientific) and validated by Percolator to achieve 1% FDR. Search parameters were used as follows: the limit of the mass range of a singly charged precursor for tryptic peptide was set from 350 Da to 6000 Da. A precursor ion mass tolerance of 10 ppm and fragment ion mass tolerance of 0.05 Da were used. Enzyme specificity was trypsin, and up to 2 missed cleavages.
were allowed. Cysteine carbamidomethylation was utilized as the fixed modification, and asparagine and glutamine deamidation, methionine oxidation and N-terminal acetylation were considered as dynamic modifications. Additionally, isotopically labeled arginine (\(^{13}C_6, ^{15}N_4\); + 10.008 Da) and lysine (\(^{13}C_6, ^{15}N_2\); + 8.014 Da), were added to the dynamic modifications. The quantification of precursor ions of each SILAC pair was processed by the Precursor ion quantifier algorithm in the Proteome Discoverer software under the user guide from Thermo Fisher Scientific. The extracted ion chromatograms of the precursor ions for a SILAC peptide pair were generated with a mass precision of 2 ppm and a retention time tolerance of 0.2 min. For identified peptides with both isotopically labeled and unlabeled forms, the relative peptide intensity for the SILAC pair was determined by the ratio of the isotopically labeled precursor ion intensity relative to the unlabeled precursor ion intensity. If either the heavy or light form of the peptide was missing, the minimal base peak intensity was applied to substitute the ratio of relative intensity of the peptide from the sample over control for quantitation. The protein ratio was determined from the ratio of unique peptides for the given protein. The protein ratio was normalized by the median protein ratio. A quality control sample of the isotopically labeled cell lysates provided by Karyopharm was in-gel digested and analyzed by LC-MS to evaluate the labeling efficiency of arginine and lysine. A quality control of the mixture of heavy labeled and unlabeled cell lysates (mixed with 1:1 ratio) was used to determine the mixing ratio of the two samples.

For the characterization of target protein PAK4, the MS/MS data was searched against the single protein database of human PAK4 found from Uniprot with the SEQUEST algorithm and validated with Percolator with a fixed FDR of 1%. The database search was performed twice, with the addition of the drug KPT-7523 (+ 517.188 Da) as a dynamic modification on the entire
amino acid sequence first, and then only on the cysteine residues. Manual inspection on the precursor ions and MS/MS spectra of each cysteine containing peptides was also performed.

3.4 RESULTS AND DISCUSSION

The overall workflow is illustrated in Figure 3-1 as: preparation of heavy and light proteins in cells, incubation of protein extract with the immobilized small molecule, combination of the protein immobilized matrix from two groups in a fixed 1:1 ratio, wash to remove unbound proteins, recovery bound proteins from the affinity matrix, follow by SDS-PAGE separation of bound proteins, and in-gel LC-MS for quantitative analysis. The work in this thesis involved analysis of the precipitated proteins at the stage of SDS-PAGE separation. The results of the work flow are shown and discussed herein.

3.4.1 Preparation of heavy and light proteins in cells (SILAC approach)

Cells were split into two populations, and each population was cultured in the medium with standard normal amino acids (light labeled) or isotopically incorporated amino acids (heavy labeled) by Karyopharm. Proteins from heavy labeled cells had mass shift of + 10 Da on each Arg and + 8 Da on each Lys, respectively. To eliminate potential bias of the proteins due to preferential growth on either heavy or light amino acid media, the forward and reverse order of heavy and light media for control and drug treated cells were used. In the forward experiment (labeled as replicate 1), the heavy labeled cells were treated with the drug by Karyopharm,
whereas the light labeled cells were treated with the solvent DMSO as a control. In the reverse order experiment (labeled as replicate 2), the heavy and light labeling were switched, as the heavy labeled cells were treated with the solvent DMSO and the light labeled cells were treated with the drug.

Proteins from the two cell lysates (either in the forward or reverse order experiment) were pulled down by drug tagged beads. Next, the two groups were mixed with a defined ratio (1:1) and the stringent washes were applied to remove non-specific bindings. We received the beads for analysis. The pulled down proteins were either released from beads for further SDS-PAGE separation followed by in-gel digestion and LC-MS analysis, or the pulled down proteins were subjected to direct digestion on the beads for subsequent LC-MS analysis (as a control check to determine if using SDS-PAGE and in-gel digestion could induce bias). The in-gel digestion approach enabled the fractionation of the pulled down proteins to reduce the complexity of samples prior to LC-MS analysis; however, the protein covalently bound to the small molecule drug could not be mobilized from and thus could not be identified. The on-bead digestion approach allowed direct digestion of all pulled down proteins from the beads and could provide a complementary result for target identification.
The potential drug target proteins should exhibit a large difference in the ratio from non-target proteins, the latter being 1:1. In the forward experiment (Rep 1), proteins specifically bound to the drug should exhibit a high heavy/light (H/L) ratio, whereas a low H/L ratio should be expected in the reverse order experiment (Rep 2). For the non-specifically bound proteins, the H/L ratio in both the forward and reverse order experiments should close to 1.

Before the quantitative comparison, we first evaluated the efficiency of SILAC labeling from the quality control sample received from Karyopharm. The heavy labeled whole cell lysates were run on SDS-PAGE, along with Rep1 and Rep 2. A gel band from MW range of 40 – 90 kDa was excised and subjected to in-gel digestion followed by LC-MS/MS analysis. The incorporation of the heavy isotope, estimated from the most abundant 30 peptides, was found to be 99.4% for Arg10 and 99.0% for Lys8, indicating no bias. The precision of sample mixing was also assessed by a quality control sample which consisted of equally combined proportions of heavy and light
labeled whole cell lysates. The quality control sample was separated on SDS-PAGE, and the gel band from MW range of 40 – 90 kDa was selected, processed and analyzed by LC-MS. The ratio of heavy/light labeled Arg and Lys was found to be 0.90 and 0.89, respectively. The result showed that the overall experimental mixing ratios of the two groups was differed by 10%. The labeling efficiency (~99%) and the mixing ratio (0.90) were thus adopted for normalization of the SILAC ratio for each protein prior to the quantitative comparison.

3.4.2 Analysis of Bound Proteins

After affinity extraction, proteins identified with at least 2 peptides in both replicates were selected for quantitatively compared (i.e. H/L ratio). The distributions of the protein groups are displayed in Figure 3-2 (using log2 ratio). As seen, the majority of the log2 ratios for identified proteins were close to zero (after normalization) on either in-gel (Figure 3-2 A) or direct on-bead digestion analysis (Figure 3-2 B). These proteins were enriched in both heavy and light states and were assumed to be non-specifically interacting with the magnetic beads. Proteins with ratios above 1.96 SD of the distribution was considered to be highly enriched proteins by the drug 16. For the in-gel digestion experiment, 51 proteins overlapped in Rep1 and Rep2 were found to be significantly up-regulated. In this regard, the non-overlapped proteins were eliminated since the identified proteins with forward and reverse order experiments should have resulted in a consistent difference in the ratio (the same rule applied for on-bead digestion analysis as well).

To further visualize the result of specific and non-specific binding of the drug, a scatter plot was generated for the quantitative comparison in biological replicates from in-gel digestion
and on-beads digestion (see Figure 3-3). The non-specific binders of the drug are tightly clustered around zero. Proteins that are specifically bound to the drug are located in the top right quadrant of the plot. Overall, only 4 proteins were identified with the log2 SILAC ratio higher than 1.96 SD across two replicates both from in-gel digestion and on-beads digestion approach. As shown in Table 3-1, the 4 proteins were serine/threonine protein kinase PAK4, keratin type II cytoskeletal 7, keratin type II cytoskeletal 8 and Histone H2AZ. After considering their biological functions (i.e. signaling pathway for the intended drug function), the serine/threonine protein kinase PAK4 was selected as the potential drug target. Nevertheless, applying a more stringent cutoff (2-fold change cutoff), only PAK4 had a fold change over 2. Consequently, PAK4, was determined as the potential target protein of the drug KPT-7523, with over 10.6-fold enrichment and 50% sequence coverage.

The unbiased identification of PAK4 as the potential target protein to the novel benzyodihydrofuranly derivative drug is critical to the drug development. We have shown that with the combination of SILAC method, affinity enrichment and quantitative LC-MS analysis were able to characterize protein changes between two cell populations, determine the differentially expressed proteins and identify the potential target proteins to the small molecule drug.
Figure 3-2. Histogram of proteins identified in forward and reverse order (reciprocal) experiments displayed by log2 SILAC ratios for (A) in-gel digestion and (B) on-beads digestion approach. The H/L and L/H ratios obtained from in-gel digestion approach are normalized by the isotope labeling efficiency of Arg (99.4%) and Lys (99.0%) and mixing ratios (0.90). The mean and standard deviation are shown as inset. Proteins above the threshold value are considered as significant (outliers).
**Figure 3-3.** Scatter plot of log2 SILAC ratios in forward and reverse replicate experiments obtained from (A) In-gel digestion and (B) on-beads digestion approach. Each dot represents a single protein. Proteins above the threshold considered to be significant (p<0.05) are shown in red.
Table 3-1. Differentially expressed proteins overlapped in both in-gel and on-beads digestion approach.

<table>
<thead>
<tr>
<th>Protein Accession</th>
<th>Gene</th>
<th>On-beads Coverage</th>
<th>In-gel Coverage</th>
<th>On-beads Rep1 Log 2 H/L</th>
<th>On-beads Rep2 Log2 L/H</th>
<th>In-gel Rep1 Log 2 H/L</th>
<th>In-gel Rep2 Log2 L/H</th>
<th>Protein Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>O96013</td>
<td>PAK4</td>
<td>13.54</td>
<td>50.08</td>
<td>2.71</td>
<td>3.02</td>
<td>50.08</td>
<td>2.71</td>
<td>Serine/threonine-protein kinase PAK 4</td>
</tr>
<tr>
<td>P0C0S5</td>
<td>H2AFZ</td>
<td>56.25</td>
<td>31.25</td>
<td>0.43</td>
<td>0.84</td>
<td>31.25</td>
<td>0.43</td>
<td>Histone H2A.Z</td>
</tr>
<tr>
<td>P08729</td>
<td>KRT7</td>
<td>11.94</td>
<td>62.26</td>
<td>0.65</td>
<td>1.03</td>
<td>62.26</td>
<td>0.65</td>
<td>Keratin, type II cytoskeletal 7</td>
</tr>
<tr>
<td>P05787</td>
<td>KRT8</td>
<td>39.54</td>
<td>73.08</td>
<td>0.46</td>
<td>0.67</td>
<td>73.08</td>
<td>0.46</td>
<td>Keratin, type II cytoskeletal 8</td>
</tr>
</tbody>
</table>

3.4.3 Characterization of PAK4

Once PAK4 was identified as the potential target protein, the binding of the small molecule drug KPT-7523 to PAK4 was further evaluated. The binding domain of the drug KPT-7523 was located by Karyopharm to the kinase domain of PAK4 through several parallel experiments, e.g. isothermal calorimetry analysis, thermal stability shift, and surface plasmon resonance methods. To study the exact binding location of the drug KPT-7523 to PAK4, the full sequence of PAK4 was examined by targeted MS analysis. In addition, literature has suggested that a related protein could likely form a covalent bound with a model compounds, 4-oxo-2-nonenal, via Michael addition, with Cys ranked on the top list, followed by His, Lys and Arg as the attached locations (Figure 3-4).
In this study, recombinant PAK4 (untreated) and the drug KPT-7523 treated PAK4 were run on an SDS-PAGE gel. A gel band in the MW range 75 – 100 kDa represented the untreated and drug treated PAK4 (see details in Figure 3-5). In this experiment, the recombinant human PAK4 contained a GST tag at the N-terminus (increased the MW to ~ 90 kDa). The gel bands corresponding to the untreated and drug treated PAK4 were excised for reduction, alkylation and in-gel trypsin digestion. The protein digest was analyzed by LC-MS.

LC-MS peptide mapping was performed on untreated and drug treated PAK4. Similar base peak chromatography was observed for the two PAK4 digests (see Figure 3-7), with all identified peptide sequences summarized in Table 3-2. A total of 91% sequence coverage was obtained for the KPT-7523 treated PAK4 and 85% for the untreated PAK4. The unidentified peptides were short tryptic peptides (likely too short to be retained on a RPLC column), and none of the unidentified peptides contained any cysteine residues.
As suggested by the information from Uniprot and DiANNA (i.e., an online tool for prediction of cysteine status in a protein), the five cysteine residues on recombinant human PAK4 are all unpaired and are likely to form a covalent bond with the KPT-7523 molecule. To further examine the five cysteines, the targeted analysis was performed on the cysteine containing peptides in untreated PAK4 and KPT-7523 treated PAK4. All the precursor ions of the peptides with addition of KPT-7523 molecule (+ 517.1880 Da) on the Cys were examined; however, none of the cysteines contained KPT modification. As illustrated in Figure 3-8, peptide T10 with Cys58 on the sequence was identified from untreated and KPT-7523 treated PAK4 with similar intensity. The fragment ions b9, b10 and b11 from the MS/MS spectra confirmed the carboxymethylation modification on Cys58. All the cysteine containing peptides, including their typical modifications and miscleavage forms, were summarized in Table 3-3. Further comparison of the peptide intensity between untreated and KPT-7523 treated PAK4 indicated that all the five cysteines in PAK4 were fully alkylated with iodoacetamide, which strongly indicated the PAK was not covalently bound to the KPT-7523 molecule. In addition, the mass (precursor ion) of the expected covalent bound between KPT-7523 and cysteine residues on PAK4 was not found. Further, other potential binding sites to the KPT-7523 molecule (i.e., His, Lys, and Arg) were also examined, and none of the peptide sequences could be found with the KPT-7523 molecule (i.e. + 517.1880 Da).

In summary, the MS results from this study demonstrated the binding nature of the drug KPT-7523 with PAK4 was not through covalent interaction. The strategy of using peptide mapping for full sequence analysis for potentially modified sites on a protein can serve as a robust means of characterization of small-molecule drug – protein complex and determination of the binding sites.
Figure 3-5. SDS-PAGE of PAK4 treated by solvent and SM drug. Lane 1: MW marker; lane 2: PAK4 treated by DMSO; lane 3: PAK4 treated by drug. The bands shown in red squares are excised for in-gel tryptic digestion.

Figure 3-6. Summary of the identification of the peptide sequence of target protein PAK4. The sequences in red were not identified by LC-MS analysis. The sequence coverage of 91% was obtained.
Figure 3-7. Base peak chromatogram of (A) PAK4 and (B) SM drug KPT-7523 treated PAK4. The major peaks are labeled with peptide numbers. Peaks marked with asterisks come from the trypsin autolysis.

Figure 3-8. Identification of a Cys containing peptide in DMSO treated (top) and SM treated PAK4 (bottom). (A) Extracted ion chromatogram and accurate mass of precursor ion 534.0517 (4+); (B) CID-MS2 spectrum of the same precursor ion.
Table 3-2. Summary of identified peptides in drug treated target protein PAK4.

<table>
<thead>
<tr>
<th>#</th>
<th>AA position</th>
<th>Sequence</th>
<th>Charge</th>
<th>RT /min (observed)</th>
<th>m/z (Observed)</th>
<th>Intensity</th>
<th>Modification</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>1-4</td>
<td>MFGK</td>
<td></td>
<td></td>
<td></td>
<td>N.D.</td>
<td></td>
</tr>
<tr>
<td>T2</td>
<td>5-5</td>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td>N.D.</td>
<td></td>
</tr>
<tr>
<td>T3</td>
<td>6-6</td>
<td>K</td>
<td></td>
<td></td>
<td></td>
<td>N.D.</td>
<td></td>
</tr>
<tr>
<td>T4-T6</td>
<td>7-20</td>
<td>KRVEISAPSNP EHR</td>
<td>3+</td>
<td>26.98</td>
<td>557.2972</td>
<td>9.63E+06</td>
<td>Miscleavage</td>
</tr>
<tr>
<td>T7</td>
<td>21-31</td>
<td>VHTGFDQHEQ K</td>
<td>3+</td>
<td>15.29</td>
<td>442.5458</td>
<td>1.05E+06</td>
<td></td>
</tr>
<tr>
<td>T8</td>
<td>32-37</td>
<td>FTGLPR</td>
<td>2+</td>
<td>27.07</td>
<td>345.6998</td>
<td>2.02E+09</td>
<td></td>
</tr>
<tr>
<td>T9</td>
<td>38-48</td>
<td>QWQSLIEESAR</td>
<td>2+</td>
<td>42.45</td>
<td>673.8395</td>
<td>2.48E+09</td>
<td></td>
</tr>
<tr>
<td>T10</td>
<td>49-68</td>
<td>RPKPLVPDACI TSIQPGAPK</td>
<td>4+</td>
<td>38.80</td>
<td>537.0541</td>
<td>2.50E+08</td>
<td>C Carbamidomethylation</td>
</tr>
<tr>
<td>T11</td>
<td>69-72</td>
<td>TIVR</td>
<td>1+</td>
<td>29.47</td>
<td>488.3212</td>
<td>5.51E+05</td>
<td></td>
</tr>
<tr>
<td>T12</td>
<td>73-75</td>
<td>GSK</td>
<td></td>
<td></td>
<td></td>
<td>N.D.</td>
<td></td>
</tr>
<tr>
<td>T13-T14</td>
<td>76-96</td>
<td>GAKGDGALTLL LDEFENMSVTR</td>
<td>3+</td>
<td>63.95</td>
<td>766.0547</td>
<td>1.00E+06</td>
<td>M modification; Miscleavage</td>
</tr>
<tr>
<td>T15-T16</td>
<td>97-102</td>
<td>SNSLRR</td>
<td>2+</td>
<td>15.46</td>
<td>388.2093</td>
<td>4.37E+05</td>
<td>Miscleavage; Miscleavage</td>
</tr>
<tr>
<td>T17</td>
<td>103-110</td>
<td>DSSPPPAR</td>
<td>2+</td>
<td>13.30</td>
<td>418.8167</td>
<td>3.06E+05</td>
<td></td>
</tr>
<tr>
<td>T18-T19</td>
<td>111-126</td>
<td>ARQENGMPEE PATTAR</td>
<td>3+</td>
<td>22.95</td>
<td>586.9413</td>
<td>2.84E+08</td>
<td>N deamidation</td>
</tr>
<tr>
<td>T20-T22</td>
<td>127-137</td>
<td>GGPKGKAGSRG R</td>
<td>2+</td>
<td>40.95</td>
<td>500.2765</td>
<td>1.34E+06</td>
<td>Miscleavage</td>
</tr>
<tr>
<td>T23-T24</td>
<td>138-152</td>
<td>FAGHSEAGGG SGDRR</td>
<td>4+</td>
<td>10.65</td>
<td>365.921</td>
<td>7.15E+04</td>
<td>Miscleavage</td>
</tr>
<tr>
<td>T25</td>
<td>153-153</td>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td>N.D.</td>
<td></td>
</tr>
<tr>
<td>T26</td>
<td>154-158</td>
<td>AGPEK</td>
<td></td>
<td></td>
<td></td>
<td>N.D.</td>
<td></td>
</tr>
<tr>
<td>T27</td>
<td>159-161</td>
<td>RPK</td>
<td></td>
<td></td>
<td></td>
<td>N.D.</td>
<td></td>
</tr>
<tr>
<td>T28</td>
<td>162-164</td>
<td>SSR</td>
<td></td>
<td></td>
<td></td>
<td>N.D.</td>
<td></td>
</tr>
<tr>
<td>T29</td>
<td>165-175</td>
<td>EGSGGPQESSR</td>
<td>2+</td>
<td>9.22</td>
<td>545.7417</td>
<td>8.72E+04</td>
<td></td>
</tr>
<tr>
<td>T30-T31</td>
<td>176-198</td>
<td>DKRPLSGPDVG TPQAPLAGSA K</td>
<td>4+</td>
<td>31.86</td>
<td>555.5254</td>
<td>1.26E+09</td>
<td>Miscleavage</td>
</tr>
<tr>
<td>T32</td>
<td>199-210</td>
<td>LAAGRPFNTYP R</td>
<td>3+</td>
<td>30.43</td>
<td>454.9141</td>
<td>6.71E+08</td>
<td></td>
</tr>
<tr>
<td>T33</td>
<td>211-218</td>
<td>ADTDHPSR</td>
<td></td>
<td></td>
<td></td>
<td>N.D.</td>
<td></td>
</tr>
<tr>
<td>T34</td>
<td>219-252</td>
<td>GAQGEPHHDVA PNGPSAGGLAI PQSSRSSRPT R</td>
<td>4+</td>
<td>33.12</td>
<td>818.4019</td>
<td>1.23E+09</td>
<td>N deamidation</td>
</tr>
<tr>
<td>T35</td>
<td>253-254</td>
<td>AR</td>
<td></td>
<td></td>
<td></td>
<td>N.D.</td>
<td></td>
</tr>
<tr>
<td>T36</td>
<td>255-290</td>
<td>GAPSPGVLGPH ASEPQLAPPAC TPAAPAVPGPP GPR</td>
<td>3+</td>
<td>43.16</td>
<td>1125.9139</td>
<td>3.40E+08</td>
<td>C Carbamidomethylation</td>
</tr>
<tr>
<td>T37</td>
<td>291-294</td>
<td>SPQR</td>
<td></td>
<td></td>
<td></td>
<td>N.D.</td>
<td></td>
</tr>
<tr>
<td>T38</td>
<td>295-298</td>
<td>EPQR</td>
<td></td>
<td></td>
<td></td>
<td>N.D.</td>
<td></td>
</tr>
<tr>
<td>T39</td>
<td>299-305</td>
<td>VSHEQFR</td>
<td>3+</td>
<td>14.26</td>
<td>301.4875</td>
<td>5.30E+05</td>
<td></td>
</tr>
<tr>
<td>T40</td>
<td>306-318</td>
<td>AALQLVVDPG DPR</td>
<td>2+</td>
<td>41.29</td>
<td>675.8734</td>
<td>2.43E+09</td>
<td></td>
</tr>
<tr>
<td>T41</td>
<td>319-326</td>
<td>SYLDNFIK</td>
<td>2+</td>
<td>45.24</td>
<td>500.261</td>
<td>2.91E+09</td>
<td></td>
</tr>
<tr>
<td>Peptide</td>
<td>Start - End</td>
<td>Sequence</td>
<td>Charge</td>
<td>M/Z</td>
<td>Intensity</td>
<td>Remarks</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>-------------</td>
<td>---------------------------</td>
<td>--------</td>
<td>-----------</td>
<td>--------------</td>
<td>----------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>T42</td>
<td>327-341</td>
<td>IEGSTGIVCIA TVR</td>
<td>2+</td>
<td>41.28</td>
<td>766.9088</td>
<td>C Carboxamidomethylation</td>
<td></td>
</tr>
<tr>
<td>T43</td>
<td>342-345</td>
<td>SSGK</td>
<td></td>
<td></td>
<td>N.D.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T44</td>
<td>346-350</td>
<td>LVAVK</td>
<td>2+</td>
<td>15.36</td>
<td>265.189</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T45-T46</td>
<td>351-355</td>
<td>KMDLR</td>
<td>2+</td>
<td>18.96</td>
<td>352.6917</td>
<td>N-terminal Acetylation; Miscleavage</td>
<td></td>
</tr>
<tr>
<td>T47</td>
<td>356-356</td>
<td>K</td>
<td></td>
<td></td>
<td>N.D.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T48</td>
<td>357-359</td>
<td>QKR</td>
<td></td>
<td></td>
<td>N.D.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T49-T50</td>
<td>360-371</td>
<td>RELIFNEVVM R</td>
<td>3+</td>
<td>49.41</td>
<td>506.9538</td>
<td>Miscleavage</td>
<td></td>
</tr>
<tr>
<td>T51</td>
<td>372-411</td>
<td>DYQHENVVEM YNSLYVGDEL WVVMEFLEGG ALTDIVTHTR</td>
<td>4+</td>
<td>71.1</td>
<td>1168.8063</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T52</td>
<td>412-439</td>
<td>MNEEOIAAVC LAVLQASLVL HAQGVIHR</td>
<td>4+</td>
<td>69.76</td>
<td>768.4125</td>
<td>C Carboxamidomethylation</td>
<td></td>
</tr>
<tr>
<td>T53-T54</td>
<td>440-453</td>
<td>DIKSDSILLTHGDGR</td>
<td>3+</td>
<td>34.93</td>
<td>523.9463</td>
<td>Miscleavage</td>
<td></td>
</tr>
<tr>
<td>T54-T55</td>
<td>440-455</td>
<td>SDSILLTHGDGR VK</td>
<td>3+</td>
<td>27.58</td>
<td>480.9312</td>
<td>Miscleavage</td>
<td></td>
</tr>
<tr>
<td>T56-T57</td>
<td>456-471</td>
<td>LSDGFGCAQVS KEVPR</td>
<td>3+</td>
<td>43.63</td>
<td>613.9736</td>
<td>C Carboxamidomethylation</td>
<td></td>
</tr>
<tr>
<td>T58</td>
<td>472-472</td>
<td>R</td>
<td></td>
<td></td>
<td>N.D.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T59</td>
<td>473-473</td>
<td>K</td>
<td></td>
<td></td>
<td>N.D.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T60</td>
<td>474-489</td>
<td>SLVGTPYWM APHELISR</td>
<td>2+</td>
<td>35.77</td>
<td>910.4816</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T61</td>
<td>490-522</td>
<td>LYPGPEVDIWS LGIMVIEVMVG EPPYFNEPLK</td>
<td>3+</td>
<td>71.76</td>
<td>1248.9558</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T62-T63</td>
<td>523-528</td>
<td>AMKMIR</td>
<td>2+</td>
<td>25.86</td>
<td>396.2169</td>
<td>N-terminal Acetylation; Miscleavage</td>
<td></td>
</tr>
<tr>
<td>T64</td>
<td>529-534</td>
<td>DNLPPR</td>
<td>2+</td>
<td>16.81</td>
<td>377.1978</td>
<td>N-terminal Acetylation</td>
<td></td>
</tr>
<tr>
<td>T65</td>
<td>535-536</td>
<td>LK</td>
<td></td>
<td></td>
<td>N.D.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T66-T67</td>
<td>537-546</td>
<td>NLHKVPSLKL</td>
<td>3+</td>
<td>27.02</td>
<td>388.8962</td>
<td>N-terminal Acetylation; Miscleavage</td>
<td></td>
</tr>
<tr>
<td>T68</td>
<td>547-551</td>
<td>GFLDR</td>
<td>2+</td>
<td>21.09</td>
<td>304.164</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T69-T70</td>
<td>552-560</td>
<td>LLVRDPAQR</td>
<td>3+</td>
<td>24.54</td>
<td>356.549</td>
<td>Miscleavage</td>
<td></td>
</tr>
<tr>
<td>T71</td>
<td>561-566</td>
<td>ATAAELIK</td>
<td>2+</td>
<td>27.37</td>
<td>408.7455</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T72</td>
<td>569-574</td>
<td>HPFLAK</td>
<td>2+</td>
<td>18.51</td>
<td>356.7103</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T73</td>
<td>575-586</td>
<td>AGPPASIVPLM R</td>
<td>2+</td>
<td>44.54</td>
<td>604.8455</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T74</td>
<td>587-589</td>
<td>QNR</td>
<td></td>
<td></td>
<td>N.D.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T75</td>
<td>590-591</td>
<td>TR</td>
<td></td>
<td></td>
<td>N.D.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Peptides marked as “N.D.” were not detected.
Table 3-3. Summary of Cys containing peptides in DMSO (C) and drug treated PAK4 (S).

<table>
<thead>
<tr>
<th>#</th>
<th>Cys position</th>
<th>Sequence</th>
<th>Modification</th>
<th>RT/min (C)</th>
<th>RT/min (S)</th>
<th>m/z (C)</th>
<th>m/z (S)</th>
<th>Ratio (S/C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T10</td>
<td>58</td>
<td>RPKPLVDPACITSIQPGAPK</td>
<td>C Carbamidomethylation</td>
<td>39.53</td>
<td>38.80</td>
<td>537.0527 (4+)</td>
<td>537.0541 (4+)</td>
<td>2.63</td>
</tr>
<tr>
<td>T36</td>
<td>276</td>
<td>GAPSPGVLGPHASEPOPLAPACTPAAPAVPGP</td>
<td>C Carbamidomethylation</td>
<td>43.71</td>
<td>43.16</td>
<td>1125.9137 (3+)</td>
<td>1125.9150 (3+)</td>
<td>1.50</td>
</tr>
<tr>
<td>T36</td>
<td>276</td>
<td>GAPSPGVLGPHASEPOPLAPACTPAAPAVPGP</td>
<td>C Carbamidomethylation; N-terminal Acetylation</td>
<td>44.38</td>
<td>43.89</td>
<td>1139.9185 (3+)</td>
<td>1139.9181 (3+)</td>
<td>1.44</td>
</tr>
<tr>
<td>T42</td>
<td>336</td>
<td>IGEGSTGIVCIATVR</td>
<td>C Carbamidomethylation</td>
<td>41.89</td>
<td>41.28</td>
<td>766.9089 (2+)</td>
<td>766.9088 (2+)</td>
<td>1.44</td>
</tr>
<tr>
<td>T42</td>
<td>336</td>
<td>IGEGSTGIVCIATVR</td>
<td>C Carbamidomethylation; N-terminal Acetylation</td>
<td>45.23&amp;</td>
<td>44.74&amp;</td>
<td>787.9144 (2+)</td>
<td>787.9141 (2+)</td>
<td>1.84</td>
</tr>
<tr>
<td>T52</td>
<td>421</td>
<td>MNEEQIAAVCLAVLQALSVLHAQGVHIHR</td>
<td>C Carbamidomethylation</td>
<td>69.81</td>
<td>69.76</td>
<td>768.4121 (4+)</td>
<td>768.4125 (4+)</td>
<td>2.27</td>
</tr>
<tr>
<td>T52</td>
<td>421</td>
<td>MNEEQIAAVCLAVLQALSVLHAQGVHIHR</td>
<td>C Carbamidomethylation; M oxidation</td>
<td>69.65</td>
<td>69.57</td>
<td>772.4111 (4+)</td>
<td>772.4116 (4+)</td>
<td>2.72</td>
</tr>
<tr>
<td>T52</td>
<td>421</td>
<td>MNEEQIAAVCLAVLQALSVLHAQGVHIHR</td>
<td>C Carbamidomethylation; N-terminal Acetylation</td>
<td>70.31</td>
<td>70.29</td>
<td>778.9172 (4+)</td>
<td>778.9158 (4+)</td>
<td>2.47</td>
</tr>
<tr>
<td>T56</td>
<td>462</td>
<td>LSDFGFCAQVSK</td>
<td>C Carbamidomethylation</td>
<td>41.54</td>
<td>40.97</td>
<td>679.8239 (2+)</td>
<td>679.8245 (2+)</td>
<td>2.27</td>
</tr>
<tr>
<td>T56</td>
<td>462</td>
<td>LSDFGFCAQVSK</td>
<td>C Carbamidomethylation; N-terminal Acetylation</td>
<td>46.62</td>
<td>46.04</td>
<td>700.8292 (2+)</td>
<td>700.8295 (2+)</td>
<td>2.47</td>
</tr>
<tr>
<td>T56-T57</td>
<td>462</td>
<td>LSDFGFCAQVSKR</td>
<td>C Carbamidomethylation; Miscleavage</td>
<td>44.24</td>
<td>43.63</td>
<td>613.9731 (3+)</td>
<td>613.9736 (3+)</td>
<td>2.72</td>
</tr>
<tr>
<td>T56-T57</td>
<td>462</td>
<td>LSDFGFCAQVSKR</td>
<td>C Carbamidomethylation; Miscleavage; N-terminal Acetylation</td>
<td>46.62</td>
<td>46.04</td>
<td>700.8292 (2+)</td>
<td>700.8295 (2+)</td>
<td>2.47</td>
</tr>
<tr>
<td>T55-T56</td>
<td>462</td>
<td>VKLSDFGFCAQVSK</td>
<td>C Carbamidomethylation</td>
<td>44.52</td>
<td>43.90</td>
<td>529.2728 (3+)</td>
<td>529.2730 (3+)</td>
<td>2.47</td>
</tr>
</tbody>
</table>

Note: The ratio (S/C) was obtained by the intensity of cysteine containing peptide detected in the drug treated PAK4 sample divided by the intensity of the same peptide detected in the DMSO treated PAK4 sample.
3.5 CONCLUSION

The chemical proteomic-based approach has successfully identified the potential PAK4 protein target to a novel benzyodihydrofuranly derivative drug (KPT-7523) as a bait. The use of the SILAC approach is effective for quantitative comparison between the control and drug treated cells, leading to the unbiased determination of drug target from complex samples with high abundant background of non-specific binding proteins. In this case, PAK4 was unambiguously determined as the potential target protein of the small molecule drug KPT-7523 from 1,158 proteins identified from two cell populations. Both sample processing, using either the on-bead digestion or in-gel digestion approaches, obtained consistent results for drug target. Additionally, the nature of the drug bound to PAK4 target was comprehensively evaluated by a targeted LC-MS approach, leading to the conclusion that the binding nature was through non-covalent interaction. The approaches presented in this chapter provides valuable strategies for potential drug target identification and characterization. Further studies on target validation by biochemical or genetic methods (e.g., western blotting, transgenic animal models) are required for the potential drug target. The native protein mass spectrometry and hydrogen deuterium exchange (HDX)-MS analysis can be applied to characterize the noncovalent bound complex \(^{21-24}\).

3.6 ACKNOWLEDGMENTS

I would like to thank to Dr. William Senapedis at Karyopharm Therapeutics Inc. for providing samples and helpful discussions and Dr. Monique Schenone from the Broad Institute for helpful guidance on SILAC experiment and data analysis.
3.7 REFERENCES


(23) A Loo, J.; A Benchaar, S.; Zhang, J. *Mass Spectrom. (Tokyo, Japan)* **2013**, 2 (Spec Iss), S0013.

Chapter 4: Integration of Capillary Isoelectric Focusing with Electrospray Ionization Mass Spectrometry using Liquid Junction Interface for Intact Protein Analysis
4.1 ABSTRACT

A low-sheath-flow liquid junction interface was developed to integrate commercial CE and ESI-MS instruments for online CIEF-MS analysis. To evaluate the interface performance, CZE-MS analysis of a BSA digest was performed on a bare fused silica capillary. Under a 30 min separation window, confident identification of BSA with 59.5% sequence coverage and an average of 157,000 theoretical plates were obtained. Online CIEF-MS analysis of intact proteins with the integrated system was demonstrated with a mixture of model proteins. CIEF parameters, including concentration of glycerol, injection ratio of catholyte/sample-ampholytes/anolyte, ampholytes composition, sheath liquid composition, and mobilization method were investigated and optimized to achieve high resolution and ionization efficiency. Reasonably high resolution for model proteins with pI ranges from 9.5 to 4.5 was demonstrated. Protein modifications including deamidation, oxidation, and the addition of heme and lactose groups were identified by the change in intact mass. In addition, CIEF-MS analysis was carried out on basic and heat stressed therapeutic protein recombinant human growth hormone, resulting in the separation and identification of the potential degradation products including deamidation and chain-cleavage. The results demonstrate the capability of the liquid junction interface CIEF-MS system in achieving high resolution CIEF separation along with modifying the effluent to overcome the incompatibility problems with ESI-MS. The integrated CIEF-MS system can be a powerful tool for top-down analysis of protein charge variants.
4.2 INTRODUCTION

In recent years, protein therapeutics have gained increasing attention in the biotechnology industry due to the significant commercial and clinical success. As discussed in Chapter 1, protein charge variants such as deamidation, C-terminal Lys loss, and aggregation introduced by post-translational modifications and degradations during manufacturing and storage processes may have great impact on the bioactivity and stability of a drug. Thus, the analysis of charge heterogeneity of protein therapeutics is critical to ensure the drug quality, safety and efficacy.

A wide array of analytical techniques have been applied to characterize protein charge variants, including ion exchange chromatography, chromatofocusing, size exclusive chromatography, slab-gel isoelectric focusing, capillary isoelectric focusing and imaged capillary isoelectric focusing, and capillary zone electrophoresis. In this chapter, we focus on the capillary isoelectric focusing (CIEF) method coupled to MS.

As described in Chapter 1, CIEF is a high resolution and high sensitivity method to separate amphoteric molecules according to their pI. The high resolving power from CIEF enables the separation of protein charge variants derived from modification on a single amino acid (with small pI difference as low as 0.02 units). In addition, the focusing effect of CIEF offers a concentration factor of 50 – 100 fold, which enhances the detection sensitivity and permits the analysis of low abundant charge variants.

Currently, the single-point UV detection and the whole-column imaging techniques have been employed in conjunction with CIEF to study protein charge variants. However, limited information on the structural details of the detected proteins can be directly obtained by the online CIEF-UV analysis. MS is an alternative detection technique which permits the direct
identification of the proteins from CIEF effluent and provides an additional dimension to
differentiate the co-migrated proteins by the molecular weight with high sensitivity and high
accuracy. The first offline combination of CIEF with MALDI-TOF-MS detection for protein
analysis was demonstrated by Foret et al. 24.

Despite the advantages of online CIEF-ESI-MS in identification and characterization of
protein charge variants, some major challenges have been encountered in direct hyphenation of
the two systems. First, the use of non-volatile electrolytes (e.g., phosphoric acid and sodium
hydroxide), anti-convective media (e.g., ethylene glycol and poly(ethylene oxide)), and carrier
ampholytes in a traditional CIEF experiment lead to electrospray instability, ion suppression and
source contamination 25,26. To improve the compatibility of CIEF with ESI-MS, the non-volatile
inorganic acids and bases have been replaced by ESI-MS friendly compounds, such as acetic
acid and ammonium hydroxide 27. Glycerol-water solution has been applied as a substitute to the
traditional anti-convective gel 28-30.

Considerable effort has been made to minimize the interference from the carrier ampholytes.
One strategy is to modify the CIEF effluent by an ESI-MS compatible sheath liquid. Tang et al.
utilized a commercial coaxial sheath flow interface to combine CIEF separation with an FTICR
mass spectrometer for the analysis of a standard protein mixture and hemoglobin variants 31,32.
The use of a high sheath liquid flow (i.e., 5 µL/min) effectively reduced the ion suppression from
the carrier ampholytes; however, it also contributed to the sample dilution. In addition, the use of
a nebulizing gas could cause suction at the capillary terminus and introduce laminar flow to
disturb the separation. The second strategy for the interface design of CIEF-ESI-MS was to
remove the interferences before entering into the ESI source. Interfaces incorporated a micro-
dialysis unit 21,25,33 or free flow electrophoresis chip 34 between the CIEF separation capillary and
ESI source were developed by different groups to filter out the low molecular weight (i.e., MW < 3000 Da) interferences from the system. RPLC was also investigated as a tool for online removal of the carrier ampholytes and other ESI incompatible reagents 35–38.

The first CIEF-nano RPLC-ESI MS system was developed by Chen et al. using an injection loop to sequentially load and transfer the focused peptides from the CIEF separation capillary to C<sub>18</sub> trap columns followed by separation by nano-RPLC 27. Online separation and characterization of intact yeast proteins with the integrated system was demonstrated 36. However, such interface designs require relative sophisticated instrument design and the dead volume can lead to the loss of separation efficiency. Other approaches focus on exploring new strategies to generate a stable pH gradient inside the separation capillary for isoelectric focusing. Zhang and coworkers developed a pH gradient immobilized capillary column in which the carrier ampholytes were bound to the monolith via the reaction of the amine and carboxyl groups with the epoxide group 39–41. Chingin et al. introduced an ampholyte-free multiple-junction CIEF system, where the separation capillary was divided into seven segments, connected by a Nafion membrane and each segment was surrounded by an external electrolyte with a well-defined pH 42. A non-linear pH gradient was established inside the separation capillary by the local buffer exchange of the internal and external electrolytes 42.

Another major technical challenge presented in the direct hyphenation of CIEF with ESI-MS is to maintain the closed electric circuit at the capillary terminus during the entire CIEF processes in the absence of a terminal electrolyte reservoir. As already discussed in Chapter 1, the configuration of the sheath flow interface allows using the coaxial sheath liquid flow to provide the electric contact at the capillary terminus during the focusing and mobilization stages. The coupling has been demonstrated by different groups with the coaxial sheath flow interface.
Nevertheless, early CIEF-ESI-MS experiments with the coaxial sheath flow interface usually involved steps to interrupt the CE potential, reposition the capillary terminus and manually pump the acid buffer to the sheath liquid capillary to replace the catholyte between the focusing and mobilization stages. These processes can lead to diffusion of the focused protein zones. Mokaddem et al. demonstrated using a plug of catholyte left at the capillary terminus, instead of the traditional catholyte reservoir, to establish the pH gradient and to ensure stable electric contact during the focusing step. No interruption of the electric potential or switching of the sheath liquid was required during the mobilization step.

As discussed earlier, some of the major constraints associated with the sheath flow interface are the potential of significant sample dilution and loss of sensitivity. To overcome the difficulties, interface designs accommodating a lower sheath liquid flow have been investigated. Dovichi’s group developed an electro-kinetically pumped sheath-flow interface to perform CIEF-MS for peptide analysis using a mixture of six amino acids instead of carrier ampholytes to establish the pH gradient. Chen’s group introduced a junction-at-tip interface with a separation capillary terminus inserted into a beveled stainless stain electrospray emitter, and the separation of model protein mixture by CIEF-MS with both uncoated and coated capillary was demonstrated.

In the present study, based on the liquid junction interface developed previously for rapid CZE-MS analysis, we report a novel version of interface design for online CIEF-MS analysis of intact proteins. The special tip geometry provided a narrow region (i.e., nanoliter volume) enclosed by the interior of the electrospray tip and the exterior of separation capillary terminus, allowing the delivery of controlled low-sheath flow to provide stable electric contact and to modify the CIEF effluent without introducing significant sample dilution. In addition, the
interface design is amenable to apply hydrodynamically driven or electrokinetically driven sheath liquid flow, or the combination of both to facilitate separation and ionization. Moreover, the stainless steel needle is replaced by silica emitter to reduce the risk of bubble formation and corona discharge, and to enhance the lifetime of the emitter. The new interface is easy to fabricate and integrates commercial CE and MS instruments for automatic CIEF-MS sequence runs. The effective coupling of the high resolving power of CIEF separation with the high resolution, high mass accuracy of Orbitrap Elite mass spectrometer using the new liquid junction interface facilitate identification and characterization of protein charge variants. In this chapter, we demonstrated the feasibility of the interface for CIEF-MS, as well as CZE-MS, analyses.

4.3 EXPERIMENTAL

4.3.1 Chemicals and Materials

HCl, sodium hydroxide, ammonium acetate, ammonium bicarbonate, guanidine hydrochloride, dithiothreitol (DTT), iodoacetamide (IAA), Tris, acetaminophen, LC-MS grade acetic acid, ammonium hydroxide and glycerol (≥99.5%) were purchased from Sigma-Aldrich (St. Louis, MO). Tris(2-carboxyethyl)phosphine hydrochloride (TCEP), LC-MS grade formic acid, HPLC grade water and methanol were purchased from Thermo Fisher Scientific (Fairlawn, NJ). Sequence grade modified trypsin was from Promega (Madison, WI). BSA (bovine serum albumin), Cyt c (cytochrome C, bovine heart), CA (carbonic anhydrase, bovine erythrocytes), α-CN (α-casein, bovine milk), Rnase (ribonuclease A, bovine pancreas), Myo (myoglobin, horse skeletal muscle), β-LG (β-lactoglobulin, bovine milk) and α-LA (α-lactalbumin, bovine milk) were purchased from Sigma-Aldrich (St. Louis, MO). The CIEF peptide marker kit was obtained.
from Beckman Coulter Inc. (Brea, CA). Servalyt 3-10, Servalyt 4-6 and Servalyt 7-9 were purchased from Serva Electrophoresis (Heidelberg, Germany), Fluka 3-10 was from Sigma-Aldrich (St. Louis, MO) and Pharmalyte 3-10 from GE Healthcare (Piscataway, NJ) and Aeslytes SH3-10 was from a kind gift from Advanced Electrophoresis Solution Ltd (Cambridge, ON, Canada). Centrifugal concentrators (5 KDa cutoff) were obtained from Sartorius (Goettingen, Germany). Bare fused silica capillaries and linear polyacrylamide coated capillaries were purchased from Polymicro Technologies (Phoenix, AZ). Standard PEEK cross, nuts, ferrules and sleeves were purchased from Idex Health & Science (Oak Harbor, WA).

4.3.2 Sample Preparation

A BSA digest was prepared for sheath flow interface optimization. A 5 µg/µL solution of BSA in 25 mM ammonium bicarbonate (pH 8.0) was reduced with 5 mM TCEP in 25 mM ammonium bicarbonate buffer for 30 min at 37 °C, and then alkylated with 10 mM IAA in 25 mM ammonium bicarbonate in the dark for 45 min at 37 °C. The reaction was quenched with addition of TCEP to a final concentration of 5 mM, and the protein solution was incubated at room temperature for 1 hour. Prior to digestion, the sample was diluted with 25 mM ammonium bicarbonate solution to 3 µg/µL, and the buffer pH was adjusted 8.0. Trypsin was added to the protein solution with a protein/ enzyme ratio of 1: 50 (w/w). Digestion was performed for 16 hours at 40 °C. The digest was lyophilized with a vacuum concentrator (Labcono, Kansas City, MO), stored at -80°C and thawed promptly before use. For CZE-MS experiments, the digest was reconstituted in 1% acetic acid to a final concentration of 1 µg/µL.

A mixture of four model proteins containing 0.8 µg/µL β-LG, 0.2 µg/µL Cyt c, 0.4 µg/µL Myo and 0.8 µg/µL α-CN was prepared in 1% acetic acid with various concentrations of glycerol (0, 10%, and 20%) for CZE-MS analysis.
A mixture of five model proteins containing 0.3 µg/µL Rnase, 0.1 µg/µL Myo, 0.2 µg/µL CA, 0.2 µg/µL β-LG and 0.15 µg/µL α-LA was prepared for CIEF-MS analysis. The stock solution (5 µg/µL) of each protein was prepared in 20% (v/v) glycerol/water medium, aliquoted and stored at -80 °C. The sample solution was prepared daily by mixing peptide pI markers (2% v/v each), protein stock solutions and the carrier ampholyte mixture pH 3-10 (0.5-2.5% v/v) supplemented in 20% (v/v) glycerol/water. Anolyte and catholyte was composed of 50 mM formic acid and 100 mM ammonium hydroxide solution in 20% (v/v) glycerol/water, respectively. The sheath liquid was 10:89:1 (v/v/v) methanol/water/acetic acid (pH 3) or 10 mM ammonium acetate (pH 5).

Base and heat stressed human recombinant growth hormone (rhGH) was also prepared for CIEF-MS analysis. The rhGH was buffer exchanged with 50 mM sodium tetraborate buffer (pH 9.5) to a concentration of 2 mg/ml and incubated at 37°C for up to 2 weeks with gentle shaking. The aliquots were taken and stored at -80 °C. Prior to CIEF-MS analysis, each aliquot was buffer exchanged to 20 mM ammonium acetate buffer (pH 3.5). The sample solution was prepared by adding peptide markers (2% v/v each), cleaned protein solution and carrier ampholytes supplemented in 30% (v/v) glycerol/water.

4.3.3 Instrumentation

CE system. Capillary zone electrophoresis and capillary isoelectric focusing experiments were performed on a PA800 plus Pharmaceutical Analysis System (Beckman Coulter, Brea CA) using a standard capillary cartridge. Unless stated otherwise, an 80 cm-long bare fused silica capillary (50 µm i.d. and 150 µm o.d.) was used for separation, and a 100-cm long bare fused silica capillary (50 µm i.d. and 360 µm o.d.) was used for delivery of sheath liquid. The inlet sides of the two capillaries were loaded in the capillary cartridge. The CE instrument was
operated in the normal polarity mode with the anode located at the inlet side of the separation capillary. The capillary temperature was maintained at 25 °C during the experiments. Electrophoresis and isoelectric focusing were controlled by 32 Karat software (Beckman Coulter, Brea CA).

The new separation capillary was preconditioned by flushing with 1M NaOH, 0.1M NaOH, water, methanol and water for 5 min. Prior to each CZE (or CIEF) run, the separation capillary was rinsed with 0.1M NaOH, 0.1M HCl, water and the BGE (or catholyte) at 70 psi for 5 min, and the sheath liquid line was filled with the designated sheath liquid.

**MS system.** An LTQ ion trap mass spectrometer (Thermo Scientific, San Jose, CA) was coupled to CE for the initial interface optimization experiments. In all cases, the mass spectrometer was operated in the positive ion mode. For the CZE-MS experiments, the instrument was operated with an electrospray voltage +1.1 kV, a tube length voltage 120 V and a capillary temperature 285 °C. The AGC (automatic gain control) target value was set at $3.0 \times 10^4$ for full MS. The maximum ion injection time for MS scan was 50 ms, and 5 microscans were applied. Full MS scans were acquired over an m/z range of 380 – 1800 Th for BSA peptides and 600 – 1800 Th for intact proteins. For the intact proteins, only MS1 data was acquired. For the tandem analysis of the BSA digest, the data-dependent acquisition mode was used. The top 9 most abundant precursor ions with charge state ≥2 were isolated with a width of ± 3.0 m/z, and fragmented in the ion trap by CID with 35% normalized collision energy. The AGC target for MS2 was $1.0 \times 10^4$, and one microscan was applied.

The CIEF-MS experiments were first carried out on the LTQ ion trap mass spectrometer and then on an LTQ-Orbitrap Elite mass spectrometer (Thermo Scientific) to generate high resolution data for MS1 deconvolution. The Orbitrap Elite mass spectrometer was operated in the Protein
Mode, a mode specifically designed to provide best trapping efficiency for intact protein ions (i.e., ions were trapped in the HCD cell with high pressure). The electrospray potential was between +4.0 – 4.5 kV. The ion transfer tube was held at 325 °C, and the S-lens RF level was set at 65%. A source fragmentation voltage between 20-35 V was used for the stressed protein samples, and no source fragmentation was utilized for standard proteins. The AGC target value for MS data acquisition was 1×10⁶. A maximum of 200 ms injection time and 5 averaged microscans were applied for the full MS scan. Full MS scans were acquired in the Orbitrap mass analyzer over the range of m/z 200 – 1000 with resolution 30,000 (at m/z 200) for the peptide pI markers, and over the range of m/z 1000 – 2000 with resolution 120,000 (at m/z 200) for intact standard proteins and 240,000 (at m/z 400) for rhGH.

4.3.4 Data Analysis

For peptide identification of the BSA digest peptides, the raw data with acquired MS2 spectra was searched against the BSA sequence using the Sequest algorithm (Thermo FisherScientific) integrated in the software Proteome Discoverer (PD) 1.4 (Thermo Fisher Scientific, San Jose, CA). The database search was performed with the following search criteria: The mass range of a singly charged precursor was set from 300 Da to 6000 Da. The precursor ion mass tolerance of 1.5 Da and fragment ion mass tolerance of 0.8 Da were used. Enzyme specificity was trypsin (C-terminal to R and K) and up to 2 missed cleavages were allowed. Cysteine carbamidomethylation was utilized as a fixed modification, and asparagine and glutamine deamidation and methionine oxidation were considered as dynamic modifications. The resulting peptides from the search were filtered by Percolator (Thermo Scientific) with a maximum of 1% FDR.
The isotopically resolved MS spectra for the intact proteins from the Orbitrap Elite were deconvoluted using the Xtract algorithm incorporated in the Xcalibur software version 2.2 (Thermo Fisher Scientific). The averaging spectra in a specific migration time window were manually selected and converted to a zero charged mass peak list with the following parameters: m/z = 1000 – 2000, S/N threshold = 5, fit factor and remainder = 30%, and maximum charge state = 20.

4.4 RESULTS AND DISCUSSION

4.4.1 Design rationale of the interface

As noted in the Introduction, one of the major constraints in interfacing CIEF with ESI-MS is the presence of carrier ampholytes. A high concentration of carrier ampholytes is critical to ensure high resolution. In general, ampholyte concentrations of 1 – 5% (v/v) are typically used in a CIEF-UV experiment. It has been reported that reducing the ampholyte concentration from 5% to 1% resulted in 20% loss in resolution and 25% loss in peak height for two model proteins due to broader peaks. On the other hand, a high concentration of ampholytes can lead to a decrease in MS detection sensitivity. Previous studies demonstrated that infusion of proteins with 1% ampholytes at a flow rate of 5 µL/min led to a 90% loss in protein signal intensity. Compared to conventional electrospray ionization, applying the nanospray ionization technique leads to considerable improvement in MS sensitivity due to the higher ionization efficiency, lower ion suppression and better ion transmission. Thus, we first explored the feasibility of using an ultralow flow rate to reduce the ion suppression from ampholytes.
Figure 4-1. Direct infusion of myoglobin with different concentrations of ampholytes (AMF) at a flow rate of 50 nL/min. (A) Mass spectra of myoglobin obtained at different concentrations of AMF, and (B) plot of protein intensity as a function of AMF concentration. Experimental conditions: 0.5 mg/ml myoglobin mixed with 0 – 5% (v/v) Pharmalyte 3-10; BGE = 1% formic acid. Sample was infused at a flow rate 50 nL/min. ESI voltage was 1.4 kV. The distance between the electrospray emitter and the MS inlet was adjusted to achieve stable spray.

In preliminary experiments, direct infusion was carried out with a sheathless capillary electrophoresis-mass spectrometer consisting of a PA 800 plus CE instrument equipped with an autosampler and pressure delivery system for sample infusion, a sheathless CESI interface (Beckman Coulter, Brea, CA), and a LTQ mass spectrometer. A mixture of myoglobin at a concentration of 0.5 mg/mL (30 μM) and Pharmalyte 3-10 at different concentrations (1 – 5%) in 1% formic acid was continuously infused into a 100-cm long capillary (50 μm i.d.) at a low flow rate (50 nL/min) with pressure (5.5 psi). The volumetric flow rate \( Q \) was estimated from the
capillary i.d. (dc) and length (L), applied pressure (∆P) and the viscosity of the fluid (η) according to the Hagen-Poiseuille Equation 4-1.

\[ Q = \frac{\pi d_c^4 \Delta P}{128 \eta L} \quad (4-1) \]

The mass spectra of myoglobin with various concentrations of ampholytes are shown in Figure 4-1. 60% loss in protein intensity was observed with the addition of 1% (v/v) ampholytes in the sample, and 76% loss was found for addition of 5% (v/v) ampholytes. The presence of ampholytes not only caused the strong suppression on myoglobin signal intensity, but also led to the shift of the protein net charges to lower charge states. The shift in protein net average charge states was likely due to the charge neutralization effect where the anions in ampholytes paired with the basic groups on a protein and the ion pairs dissociated in the gas phase to produce neutral acid and protein with lower charge states \(^{31,52}\). Additionally, it was found infusing sample containing non-volatile and high concentrated ampholytes increased the risk of clogging in the electrospray emitter. A further decrease of the flow rate from 50 nL/min to 10 nL/min did not lead to improvement (i.e., a suppression of protein intensity by 52% was observed in the presence of 1% ampholyte). These results suggest that reducing the flow rate cannot overcome the incompatibility problem of ampholytes to ESI-MS.

To alleviate the adverse effects caused by ampholytes, another strategy is using a makeup flow to modify the characteristics of the capillary effluent. MS-compatible organic solvent (e.g., methanol or acetonitrile) and volatile acid (e.g., formic acid or acetic acid) are typically added in the makeup flow to mix with the aqueous capillary effluent, change the properties of electrospray (e.g. volatility and surface tension) and improve the ionization and desolvation of the analytes \(^{53}\).
In addition, since the response of ESI-MS is considered as concentration-sensitive within the operating flow rate from subnanoliter per minute to submicroliter per minute regime, diluting the capillary effluent with a makeup flow results in a lower concentration of interferences (e.g., ampholytes) and less pronounced ion suppression\textsuperscript{54,55}. The traditional sheath-flow interface is not ideal for coupling CIEF to ESI-MS since the large sheath liquid flow (e.g. 2-6 µL/min) can lead to significant sample dilution. In addition, using nebulizing gas may introduce laminar flow by a suction effect which can reduce the separation\textsuperscript{56}. On the other hand, a low sheath flow (i.e., estimated at ~ 50 nL/min) driven by electroosmotic flow may not be efficient to reduce the ion suppression from the high concentration of ampholytes\textsuperscript{57}.

In this chapter, we developed a sheath flow interface to allow delivery of a controlled low flow of sheath liquid to modify the CIEF effluent with minimum degradation of the separation. The interface facilitated direct coupling of commercial CE and MS instruments for automatic CIEF-MS analysis.

4.4.2 Configuration of the interface

The design of the interface is illustrated in Figure 4-2. The extra column volume has a great impact on the interface performance (peak tailing and post-column broadening)\textsuperscript{56}. In the present design, the dimensions of the components were carefully chosen to minimize the extra column volume. A bare fused silica separation capillary (100-cm long, 50 µm i.d., 150 µm o.d.) was threaded through a standard PEEK cross (Upchurch, Oak Harbor, WA) with a 150 µm i.d. central hole, so that almost zero dead volume was created inside the cross. The distal end of the separation capillary was then threaded into an electrospray emitter (4-cm long, 10 µm tip i.d.) pulled from a 180 µm i.d. × 360 µm o.d. silica capillary with a P-2000 puller (Sutter Instrument,
Novato, CA). The distance between the capillary terminus and the electrospray tip was optimized (details to be discussed later in the tip geometry section). The other two arms of the cross were connected to a capillary (100-cm long, 50 µm i.d., 360 µm o.d.) to deliver the sheath liquid, and a short capillary (15 cm-long, 20 µm i.d., 360 µm o.d.) inserted into the sheath-liquid outlet reservoir where a platinum electrode was housed for applying the ESI voltage. The inlet of the separation capillary and sheath liquid line were loaded in a standard capillary cartridge and set in the CE instrument. All liquid reservoirs, along with the electrospray emitter were placed at the same height to avoid hydrodynamic flow. The electrophoretic separation potential was controlled by the power system in the CE instrument. The built-in pressure delivery system allowed control of the flow rate either in the separation capillary or sheath liquid inlet capillary by applying of supplementary pressure (details on supplementary pressure will be discussed later). With the 32Karat and Xcalibur software, the parameters for electrophoretic separation and MS data acquisition could be optimized independently. In addition, the sequence runs could be programmed and analyzed automatically.

The small space enclosed by the interior of the electrospray tip and the exterior of separation capillary tip introduced only low nL extra dead volume into the system and did not contribute to apparent sample loss or peak broadening (see later). The conductive liquid of the sheath liquid helped maintain the electrical connection at the capillary terminus. In a CIEF-MS experiment, the capillary tip inner volume functioned as a catholyte reservoir during the focusing stage, to establish a pH gradient inside the separation capillary. During the mobilization stage, a low-flow sheath liquid was pumped from the inlet of the sheath liquid line at 0.1 to 2 psi pressure, to mix with the capillary effluent. ESI voltage was applied on the sheath liquid in the inner tip to establish electrospray. Instead of using a metal tip or coated tip, a non-coated silica electrospray
tip was chosen for several reasons, including cost, easy of fabrication, less potential for a redox reaction to disturb the electrospray, and longer lifespan.

Figure 4-2. Scheme of liquid junction interface.

4.4.3 Operation and optimization of the interface

One of the major concerns associated with the liquid junction interface configuration is the potential efficiency loss that could occur at the junction. Some operational parameters (e.g., the geometry of the electrospray tip and the supplementary pressure) are known to impact the analyte transport process inside the junction and affect separation efficiency, as already revealed by computational simulation studies carried out on different types of liquid junction designs $^{58-60}$. To better understand and control our integrated system, we investigated the effects of the working parameters on the interface performance under CZE-MS analysis.
4.4.3.1 Geometry of electrospray tip

The geometry of the electrospray tip, including the width of the gap between the electrospray tip orifice and the separation capillary terminus, and the tip size, significantly affects the interface performance. In our liquid junction interface, the ESI potential is directly applied on the conductive sheath liquid inside the gap. The tip size affects the electrospray quality. It has been demonstrated an electrospray tip with smaller orifice could accommodate lower flow rates in nano-flow regime, yielding higher sensitivity \(^{59,61,62}\), but at the risk of increased tip clogging. To obtain a stable electrospray within the flow rate region for CE and CIEF modes (typically less than 500 nL/min), the size of the electrospray tip was determined to be 10 µm.

Apart from the tip size, another important feature in tip geometry is the width of the gap between the end of the CE capillary and the end of the electrospray tip. The mixing of the capillary effluent with a steady flow of MS compatible sheath liquid occurs in the gap. A narrow gap is preferred to minimize the post-column broadening; however, a broader gap results in a higher mixing ratio of the capillary effluent with the sheath liquid to yield higher ionization efficiency. To find the best compromise between separation and ionization efficiency, the capillary-to-tip distance was investigated by CZE-MS.

In our interface configuration, since the separation capillary with 150 µm o.d. was fully inserted into an electrospray tip with limited volume (pulled from a 180 µm. i.d. capillary), the end of the separation capillary stopped at where the local i.d. of the electrospray tip was the same as the o.d. of the separation capillary. As shown in Figure 4-3, the o.d. of the capillary distal end could be considerably reduced to approximately 45 µm via mechanical polishing or chemical etching with hydrofluoric acid, which allowed the capillary terminus to approach closer to the electrospray tip orifice (0.5 mm capillary-to-tip distance). However, it was difficult to get a
symmetric shape and a reproducible o.d. at the capillary terminus, and the end of the capillary after treatment became fragile. To simplify the interface assembly process and to improve the robustness of the interface setup, we chose a capillary with an unmodified rectangular terminus and performed optimization studies on the capillary-to-tip distance.

Figure 4-3. Microphotos of three different shapes of capillary termini in the electrospray tips. (A) Flat capillary terminus; (B) polished capillary terminus; (C) etched capillary terminus. The capillary-to-tip distance is (A) 0.7 mm, (B) 0.6 mm and (C) 0.5 mm.

To ensure that a low flow of sheath liquid could be continuously delivered, the distal end of an 80-cm long bare fused silica capillary was inserted as far as possible in an electrospray tip and then pulled back for 50, 100 and 150 µm, yielding an overall capillary-to-tip distance of 0.85, 0.90 and 0.95 mm, respectively. CZE-MS analyses of 1 µg/µL (15 µM) BSA digest were performed with the integrated CE-ESI-MS system to investigate the optimum capillary-to-tip distance. Compared to 0.85 mm distance, the 0.95 mm capillary-to-tip distance led to a 1.8 min increase expand in the separation window and an average of 4.0 min delay in migration time (t) for the BSA peptides. When the capillary terminus was almost touching the conical part of the
electrospray tip, the sheath liquid flow rate was likely limited to tens nanoliter per minute. While pulling the distal separation capillary back a certain distance, a higher sheath liquid flow resulted, applying a low pressure (0.5 psi) at the inlet side of the sheath liquid line. The increase in sheath liquid flow created a higher back pressure on the separation capillary and led to a longer separation window and migration time for the peptides. The 0.95 mm capillary-to-tip distance produced an average of 21.5% broader peak with 6-fold higher signal intensity for the BSA peptides than the 0.85 mm distance. A larger gap introduced more post-column broadening; on the other hand, it provided more mixing space for the capillary effluent and the sheath liquid, resulting in less ion suppression and higher ionization efficiency. Among the three separations, the highest resolution for the selected BSA peptides ($R_{s1,2}=2.0$ and $R_{s2,3}=3.1$) and the highest theoretical plates of 31, 215 were found at the 0.90 mm capillary-to-tip distance, along with a reasonable ionization efficiency. Hence, the overall capillary-to-tip distance for the liquid junction interface with a flat cutting separation capillary was determined to be 0.9 mm. The separation efficiency for CZE-MS of BSA peptides with the integrated system could be further improved upon optimization of other parameters.

### 4.4.3.2 Effect of supplementary pressure

Another important parameter, the applied pressure at the separation capillary inlet and sheath liquid was optimized to achieve enhanced interface performance. Our interface configuration allowed the control of flow rate inside the separation capillary or/and the sheath liquid line via applying a supplementary pressure on the inlet side of one or both capillaries. Applying supplementary pressure could effectively reduce the total analysis time, and this was especially beneficial when the highly viscous separation medium (e.g., glycerol) was employed (see later).
It should be noted that due to the physical setting of the commercial CE instrument, it was impossible to independently control the level of pressure applied on the separation capillary and sheath liquid line. When the pressure on the separation capillary was increased, the sheath liquid flow rate also increased. Due to the tip geometry, increasing the applied pressure led to different increments in the CE bulk flow rate and the sheath liquid flow rate. The effect of supplementary pressure on the interface performance was studied. It was found increasing the pressure from 0.1 to 0.5 psi resulted in a 9 min decrease in the separation window and a 7.3 min decrease in the average migration time for BSA peptides. With 0.5 psi pressure, 40 % loss in resolution, however, was observed for the selected BSA peptides, and 73% decrease in theoretical plates and 45% loss in maximum peak capacity were found than with only 0.1 psi pressure. Clearly, higher supplementary pressure introduced higher laminar flow inside the separation capillary, which degraded the separation. Nevertheless, with 0.5 psi supplementary pressure, a 4-fold enhancement in peptide intensity was observed due to the lower ion suppression and better ionization efficiency induced by the higher sheath liquid flow. Taking into account analysis time, separation efficiency and ionization efficiency, the 0.1 psi supplementary pressure was considered best and was employed.
4.4.4 Evaluation of interface performance by CZE-MS analysis

4.4.4.1 Evaluation of the interface performance for peptide separation

CZE-MS experiments were performed on a LTQ ion trap mass spectrometer to evaluate the performance of the in-house constructed liquid junction interface. A BSA tryptic digest was used as the sample for system evaluation with a 100 cm-long bare fused silica capillary for separation. The background electrolyte (BGE) contained 5% acetic acid (pH 2.4) and the sheath liquid was composed of 1% acetic acid in 10% methanol (pH 3.0). The BSA digest was prepared in 1% acetic acid to a concentration of 1 µg/µL (15 µM). Approximately 530 nL (corresponding to 27% of the total capillary volume) of the peptide sample was hydrodynamically injected into the separation capillary. Due to the low conductivity of the sample buffer (1% acetic acid) compared to the BGE (5% acetic acid), sample stacking occurred resulting in concentrated and narrow sample zones. An electric field strength of 160 V/cm was applied across the separation capillary (calculated from +17 kV electrophoretic voltage and +1.1 kV ESI voltage), and a minimum supplementary pressure (0.1 psi) was applied on the inlet side of the capillary to assist separation and to stabilize the electrospray. The sheath liquid flow was delivered with 0.1 psi. Figure 4-4 shows the density map and base peak electropherogram for the BSA peptides. A 30 min separation window was produced for peptide separation. A total of eight peptides which migrated out at different positions within the separation window were extracted with a mass tolerance of 1.0 Da to evaluate the separation efficiency out of roughly 90 peaks that were resolved in the separation window. An average of 157,000 theoretical plates was calculated using the migration time (t) and the electrophoretic peak width at the half height ($w_{1/2}$) for each peptide, according to Equation 1-6 (see Chapter 1). Figure 4-4C shows the summed ion electropherogram of the three selected peptides with baseline separation ($R_s 1, 2=6.4$ and $R_s 2$).
Narrow and symmetrical peaks were obtained for the three peptides with \( \frac{1}{2} \) ranged from 11 to 24 s. The performance of the interface indicates that high efficiency separations achieved in the CE column were not significantly degrade by the sheath flow electrospray interface.

To identify the tryptic peptides, CZE-MS/MS analysis was performed in the data dependent acquisition mode. An overall sequence coverage of 59.5\% was obtained from the 52 unambiguously identified peptides (with 1\% FDR).

Overall, a reasonable efficiency, comparable to the efficiency from a published low sheath flow interface, was obtained with the liquid junction integrated CZE-MS system \(^6^5\). The results suggest the integrated system is capable to study the protein digest, or even complex proteomic samples. With the employment of a controlled low sheath flow in our liquid junction interface, a variety types of background electrolyte and chemical additives can be used in the CZE-MS experiments to enhance the separation.
Figure 4-4. CZE-MS of BSA digest. (A) Ion density map of BSA digest, (B) base peak electropherogram and (C) summed extracted ion electropherogram of three tryptic peptides. The start and end of the separation window are labeled in the ion density map. Capillary: 100 cm-long bare fused silica capillary; sample: 1µg/µL BSA digest in 1% acetic acid solution; BGE: 5% acetic acid; sheath liquid: 1% acetic acid in 10% methanol; injection: 2psi for 25s; separation: 17 kV and 0.1 psi applied on the inlet of the separation capillary; sheath liquid delivered with 0.1 psi; ESI voltage: 1.1 kV; scan range: m/z 380 – 1800. Three peptides: HLVDEPQNLIK at m/z 653.36 (2+), YLYEIAR at m/z 464.25 (2+) and DAFLGSFLYER at m/z 785.38 (2+). Linear ion trap data was filtered by 7-point Gaussian filter.
4.4.4.2 Repeatability of the peptide separation

The robustness of interface performance was evaluated with programmed sequence runs using the integrated CZE-MS system. Triplicate analysis of BSA digest was performed on a bare fused silica capillary. Figure 4-5 presents a comparison of the base peak electropherogram of the technical replicates. Similar separation profiles were obtained across the three runs, with a % CV of 9.5 for the base peak intensities. The % CVs of the migration time run-to-run replicates remained below 3.0 % BSA peptides from the extracted ion electropherograms, as presented in Table 4-1 for eight high intensity peptides. Reasonable reproducibility in peptide intensity for extracted ion electrophoretic peaks (i.e., %CV < 20.0) was obtained. The results demonstrate the technical repeatability of the liquid junction interface in coupling commercial CE instrument and mass spectrometer for CZE-MS analysis.

Figure 4-5. Base peak electropherograms of BSA digest in triplicate CZE-MS runs.
Table 4-1. Repeatability in migration time measured in triplicate analysis of BSA digest.

<table>
<thead>
<tr>
<th>peptide #</th>
<th>m/z</th>
<th>Average migration time (min)</th>
<th>Standard deviation</th>
<th>%CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>653.98</td>
<td>23.68</td>
<td>0.54</td>
<td>2.3</td>
</tr>
<tr>
<td>2</td>
<td>488.51</td>
<td>23.78</td>
<td>0.59</td>
<td>2.5</td>
</tr>
<tr>
<td>3</td>
<td>465.05</td>
<td>25.89</td>
<td>0.68</td>
<td>2.6</td>
</tr>
<tr>
<td>4</td>
<td>462.99</td>
<td>26.10</td>
<td>0.60</td>
<td>2.3</td>
</tr>
<tr>
<td>5</td>
<td>583.10</td>
<td>26.88</td>
<td>0.73</td>
<td>2.7</td>
</tr>
<tr>
<td>6</td>
<td>450.91</td>
<td>22.25</td>
<td>0.49</td>
<td>2.2</td>
</tr>
<tr>
<td>7</td>
<td>537.63</td>
<td>22.76</td>
<td>0.52</td>
<td>2.3</td>
</tr>
<tr>
<td>8</td>
<td>867.14</td>
<td>33.07</td>
<td>1.00</td>
<td>3.0</td>
</tr>
</tbody>
</table>

4.4.4.3 Intact protein separation with CZE-MS

In the preliminary experiment, the separation of four standard proteins, including cytochrome c (Cyt c, MW 12.3 kDa, pI 9.6), myoglobin (Myo, 16.9 kDa, pI 7.2 & 6.8), β-lactoglobulin (β-LG, MW 18.4 kDa, pI 5.2 & 5.3) and α-casein (α-CN, MW 23.6 kDa, pI 4.6) was carried out in a bare fused silica capillary using aqueous acidic buffer (5% acetic acid, pH 2.4). Poor resolution and significant peak tailing were observed for protein separation (Figure 4-6A, i). The failure in separation of the proteins can also be attributed to the strong nonspecific protein-wall interaction and the high EOF. Proteins contain multiple charged amino acids which under the electrophoretic conditions were more likely to interact with the uncoated capillary wall. To investigate the potential of protein separation with our liquid junction integrated system, we first focused on reducing the protein adsorption to the capillary wall and suppressing the strong EOF in a bare fused silica capillary.

Due to the low dielectric constant and high viscosity, glycerol has been used in the separation medium to suppress the EOF $^{28}$. It has also been reported that the addition of glycerol reduced
the adsorption of basic proteins to the capillary wall and enhanced the solubility of hydrophobic proteins \(^{66-68}\). However, the low volatility and high surface tension of glycerol contributed to the decrease in ionization efficiency \(^{69}\).

To study the concentration of glycerol that can be applied in our system, CZE-MS of intact proteins was carried out with glycerol in the BGE. The mixture of four standard proteins was prepared in 1% acetic acid with different concentrations of glycerol, and the BGE was 5% acetic acid with the same glycerol content as the sample buffer. Approximately 295 nL (corresponding to 15% of the total capillary volume) of the protein sample was used, and the injection parameters (i.e., pressure and time) were recalculated for each sample according to the Hagen-Poiseuille equation due to the change in viscosity. Separation was carried out with high voltage (~ +20 kV) and with a minimal supplementary pressure (0.1 psi) applied on the inlet of the separation capillary. The sheath liquid was delivered with 0.1 psi. The ESI potential was optimized for each CZE-MS run to obtain a stable electrospray. Compared to the separation with aqueous BGE (Figure 4-6A, i), it was found the addition of 10% (v/v) glycerol led to an improvement in the resolution for the four proteins, and baseline separation was observed with 20% (v/v) glycerol in the BGE (Figure 4-6A, ii & iii). Upon increasing glycerol concentration to 20% (v/v), longer migration times and broader peak widths were observed for the proteins.

The increase in glycerol concentration in the separation medium significantly increased the viscosity and led to a dramatic decrease in the apparent mobility of the proteins, since both of the electrophoretic mobility and electroosmotic flow were inversely proportional to the viscosity of the BGE (Chapter 1, Equation 1-1 and 1-2). In addition, due to the use of a supplementary pressure, at higher glycerol concentration, more glycerol was introduced into the liquid junction as evidenced by the higher ESI potential required at the higher glycerol concentration (1.2 kV at
0% glycerol, 2.9 kV at 10% glycerol and 4.0 kV at 20% glycerol). The increase in the viscosity of the sheath liquid present at the junction led to a decrease in sheath liquid flow rate. Thus, the overall flow rate decreased and resulted in the longer migration times and broader peak widths for the proteins. In addition, the change in migration order for Cyt c, Myo and β-LG as the glycerol concentration increased also suggested the level of protein-wall interaction was significantly reduced. When capillary electrophoresis was performed in the absence of glycerol, due to the presence of strong EOF in the bare fused silica capillary, the apparent mobility of each protein was primarily determined by the high electroosmotic mobility. Thus, the apparent mobility of four proteins was not differentiated. Proteins with smaller size and higher positive charges under the electrophoretic condition (5% acetic acid, pH 2.4) have the highest electrophoretic mobilities and are expected to migrate faster. In this regard, the migration order should follow Cyt c (r=1.7 nm, pI =9.6), Myo (r=1.9 nm, pI= 7.2 & 6.8), β-LG (r=2.6 nm, pI=5.2 & 5.3) and α-CN (r=2.9 nm, pI=4.6) 70–75. However, compared to the acid protein β-LG, the adsorption-desorption process of two basic proteins, Cyt c and Myo slowed down the migration inside the separation capillary. When 20% glycerol was added to the BGE, the interaction of Cyt c and Myo with the capillary wall was significantly reduced, though not completely eliminated. The apparent mobility of each protein was primarily determined by the electrophoretic mobility, thus the Cyt c migrated faster than the β-LG.

As the concentration of glycerol increased to 20%, with a high ESI voltage (i.e., 4 kV) and high heated capillary temperature (i.e., 325 °C), a stable electrospray was obtained in the absence of nebulizing gas and high sheath flow. It was found that a higher concentration of formic acid in the sheath liquid was required to assist ionization of proteins when higher glycerol was applied in the BGE. Therefore, 20% glycerol was considered to be sufficient for the protein
analysis in a bare-fused silica capillary. It was also observed that the charge envelope of proteins shifted to a higher m/z range with increasing concentration of glycerol in BGE (Figure 4-6B). Due to the high surface tension of glycerol, electrospray droplets accumulated more charges at the surface prior Rayleigh fission, thereby the higher charge states of the protein ions were observed using 20% glycerol 69.

Employment of 20% glycerol in the electrolyte is also beneficial for the CIEF-MS analysis of intact proteins in a bare fused silica capillary 19,28,30. The minimization of protein-wall interaction and suppression of EOF stabilize the focused bands during entire CIEF process and ensure the high-resolution of CIEF separation. In addition, the enhancement in protein solubility by glycerol reduces the potential of protein precipitation and eliminates the risks of tip clogging during the analysis. Although using a neutral coated capillary can minimize completely avoid the protein-wall interaction and the presence of EOF, the short lifetime and the non-compatibility of the coating to the highly basic pH is an issue. In the following sections, the intact protein separation in the glycerol-water medium is demonstrated using the liquid junction interface in an integrated CIEF-ESI-MS system.
Figure 4-6. CZE-MS of four standard proteins with different concentrations of glycerol in the BGEs. (A) Base peak electropherogram of four proteins separated in 5% acetic acid supplemented in (i) 0, (ii) 10% (v/v) and (iii) 20% (v/v) glycerol-water solution. (B) Representative ESI mass spectra of myoglobin at different concentrations of glycerol. Capillary: 100 cm-long bare fused silica capillary; sample: 0.2 µg/µL cytochrome C, 0.4 µg/µL myoglobin, 0.8 µg/µL β-lactoglobulin and 0.8 µg/µL α-CN in 1% acetic acid supplemented with different concentrations of glycerol; sheath liquid: (i) 0.5% formic acid, (ii) 1% formic acid and (iii) 2% formic acid in 10% methanol; injection: 15% of the total capillary volume; separation: (i) and (ii) 20 kV and 0.1 psi applied on the inlet of the separation capillary, and (iii) 17 kV and 0.1 psi applied on the inlet of the separation capillary; sheath liquid delivered with 0.1 psi; ESI voltage: (i) 1.2 kV, (ii) 2.9 kV and (iii) 4.0 kV; scan range: m/z 600 – 1800. Linear ion trap data was filtered by 7-point Gaussian filter.
4.4.5 Online CIEF-MS with liquid junction interface for intact protein analysis

The system used in the online CIEF-MS experiments integrates the PA800 plus CE instrument with the LTQ ion trap mass spectrometer or the Orbitrap Elite mass spectrometer through the liquid junction interface. Figure 4-7 illustrates the scheme of operation of a CIEF-MS experiment with the integrated system. Electrolyte containing 100 mM ammonium hydroxide was applied as the catholyte and 50 mM formic acid as the anolyte. The sheath liquid was composed of 1% (v/v) acetic acid in 10% (v/v) methanol. The catholyte, anolyte, and the mixture of sample and ampholytes were prepared in 20% glycerol-water to maintain the same viscosity, reduce EOF, suppress protein-wall interaction and minimize protein precipitation during the focusing stage. The catholyte, the mixture of sample and ampholytes, and anolyte were sequentially injected into an 80-cm long bare fused silica capillary. At the focusing stage, upon applying a high voltage with positive polarity (+22 kV) across the separation capillary, the ampholytes were self-focused and a linear pH gradient was formed inside the capillary. The proteins and peptides were focused into sharp zones at their pI values where the net charges were zero. The time for the focusing step was monitored by the electric current profile. Once the current dropped by more than 90% of the maximum and a plateau was reached (the residual current inside the capillary was typically 0.1 – 0.3 µA), the focusing was considered complete.

While maintaining the high electric field, the sheath liquid was then continuously delivered to the low volume junction using the built-in pressure delivery system in the PA 800 plus instrument. This changed the pH at the cathodic boundary from basic to acidic, resulting in the mobilization of the focused sample zones (along with the ampholytes) towards the exit of the separation capillary (i.e., chemical mobilization). A low pressure was often applied at the inlet of the separation capillary to introduce hydrodynamic flow in the capillary to assist the mobilization
(i.e., pressure mobilization). A +4 kV ESI potential was applied at the outlet electrode equipped in the sheath liquid outlet reservoir to establish the electrospray. The dilution of the CIEF effluent containing ampholytes with the low sheath liquid flow at the junction resulted in reduced ion suppression and enhanced ionization efficiency for the proteins.

Figure 4-7. Scheme of CIEF-MS process with the liquid junction interface. (A) Focusing stage and (B) mobilization stage.
4.4.5.1 Relative injection volume

As noted in the Introduction, Varenne’s group has previously demonstrated replacing the traditional catholyte reservoir by a plug of catholyte left at the capillary terminus to provide a basic pH and complete electric circuit during the focusing stage. Chen’s group further modified the partial injection technique by adding a plug of anolyte after the sequential injection of catholyte and sample-ampholytes zones. In the present study, we adopted the “sandwich injection” method developed by Chen’s group and studied the effect of relative injection volume of catholyte, sample-ampholytes and anolyte on the separation and detection with the liquid junction integrated CIEF-MS system.

A mixture of five model proteins, ribonuclease A (Rnase, pI 9.6), myoglobin (Myo, pI 7.2 and 6.8), carbonic anhydrase (CA, pI 5.9), β-lactoglobulin (β-LG, pI 5.3 – 5.2) and α-lactalbumin (α-LA, pI 4.5) were mixed with 0.5% Servalyt 3-10 to study the CIEF-MS performance. Five peptide markers (pI 10.0, 9.5, 7.0, 5.5 and 4.1) were added to the sample to evaluate the linearity of the pH gradient. To precisely control the actual injection volume, the time and pressure required for pushing a plug of protein sample through the capillary filled with 20% glycerol-water solution were recorded. For a given injection volume, the injection parameters (pressure and time) were then calculated according the percentage of the separation capillary to be occupied by the sample plug. For example, to obtain an injection ratio of 47/50/3 for the catholyte/ sample–ampholytes/ anolyte, the capillary was first filled with catholyte, and then the sample-ampholytes was injected at 15 psi for 59 s, followed by injection of anolyte at 3 psi for 29 s.

While maintaining the injection volume for sample-ampholytes zone at 50% of the total capillary volume (i.e., injected at 15 psi for 59 s), increasing the anolyte zone from 3% to 25%
moved the position of the sample-ampholytes zone towards the exit of the separation capillary and led to a 6.33 min decrease in the migration times for the proteins. Decreasing the volume of sample-ampholytes zone from 50% to 25% of the total capillary volume significantly reduced the width and the linearity of the pH gradient. A broad pH gradient with good linearity (R²=0.986) were obtained for CIEF-MS run with the sample-ampholytes zone occupied at 50% and the anolyte zone occupied at 25% of the total capillary volume, which was favored for protein separation and also in consistent with Chen’s result. Therefore, the relative injection volume of catholyte/ sample-ampholytes/ anolyte of 1:2:1 was selected in the subsequent experiments.

4.4.5.2 Carrier ampholyte composition

Ampholyte is a distribution of amphoteric compounds with small ΔpH intervals to create a stable and linear pH gradient in the presence of a high electric field. The composition of carrier ampholytes, including the pH range, concentration and type (commercial brand) of ampholytes, is the most critical factor to determine the CIEF-MS performance. A broad range ampholyte (pH 3-10) was selected to separate the mixture of acidic and basic model proteins with different pIs (4.5 – 9.6). Although a high resolution protein separation can be achieved with a high concentration of ampholytes (e.g., 5%), the strong ion suppression is a problem for ESI-MS detection. The effect of ampholyte concentration on CIEF-MS performance has already been studied with the high-sheath flow interface. To investigate on the concentration of ampholytes favored for ESI-MS detection with our low sheath-flow interface, CIEF-MS was carried out on an LTQ ion trap mass spectrometer with 0.2 mg/mL Myo and different concentrations (i.e., 0.5 – 2.5% v/v) of Fluka 3-10. In all cases, the sheath liquid was delivered with 0.5 psi to reduce the concentration of ampholytes entering into the electrospray. The protein
peak was almost non-detectable with 2.5% Fluka, due to the significant ion suppression. A low intensity peak (1.3E+04) for Myo (pI 7.2) was observed with 1.25% Fluka 3-10, and a further decrease the ampholyte concentration to 0.5% resulted in 3-fold increase in signal intensity for Myo and the identification of the low abundant isoform (pI 6.8) (data not shown). Considering the resolution and sensitivity, 0.5% ampholyte concentration was sufficient to provide a smooth and continuous pH gradient and to differentiate proteins with a minor difference in pI without introducing significant ion suppression.

Ampholytes from different commercial sources have different chemical structures in the amphoteric species, and thus the type of the ampholytes affects the stability of a pH gradient. We evaluated the CIEF-MS performance with four different types of ampholytes (Fluka 3-10, Pharmalyte 3-10, Servalyt 3-10 and Aeslytes SH3-10) on the liquid junction interface integrated system. As shown in the ion density maps in Figure 4-8, a relatively clean background over the MS scan range m/z 1000 – 2000 were observed for Pharmalyte 3-10 and Aeslytes SH3-10, probably due to less components with high molecular weights. Compared to the other two types of ampholytes, relative higher concentrations (i.e., 1% v/v) of the two ampholytes were compatible with the low sheath flow liquid junction interface for intact protein analysis. For Pharmalyte 3-10 and Aeslytes SH3-10, mobilization of the focused protein bands solely by electroosmotic flow via chemical mobilization method resulted in unreasonably long migration times. This has been reported by others, and was likely a result of the change of EOF induced by the dynamic modification of ampholytes on the capillary wall. Thus, cathodic pressure mobilization was used for protein mobilization. Good pH linearity (R^2= 0.990 and 0.937) was established for both ampholytes. The two isoforms of Myo with a pI difference of 0.4 pH unit were separated by Pharmalyte 3-10 and almost baseline resolved by Aeslytes SH3-10 from the
base peak electropherograms. Minor peaks for Rnase (pI 9.6) and α-LA (pI 4.5) were also detected in the CIEF-MS run with 1% Aeslytes SH3-10. The clean backgrounds provided by Pharmalyte and Aeslytes SH were attractive for protein analysis since it reduced the interference of ampholytes species at the MS scan range for the protein charge envelopes. However, due to the limited choice on the mobilization method, the two ampholytes were not further investigated.
Figure 4-8. CIEF-MS with 1 % (v/v) Pharmalyte 3-10 and Aeslytes SH3-10 using pressure mobilization. (A) Plot of peptide pI as a function of migration time, (B) ion density map and (C) base peak electropherogram of proteins separated with (i) Pharmalyte 3-10 and (ii) Aeslytes SH3-10. Experimental conditions: injection: catholyte/sample-ampholytes/anolyte=1/2/1 (v/v/v); focusing: (i) 22 kV, 35 min and (ii) 23 kV, 40 min; mobilization: (i) 22 kV, 0.7 psi for and (ii) 23 kV, 1.5 psi; sheath liquid delivered with (i) 0.7 psi and (ii) 1.5 psi.
With Fluka 3-10 and Servalyt 3-10, both pressure mobilization and chemical mobilization were applicable to observe protein peaks in a reasonable total analysis time. Figure 4-9 summarizes the CIEF-MS results obtained with the two ampholytes using chemical mobilization. As shown in the ion density maps and base peak electropherograms, compared to Pharmalyte and Aeslytes SH, a highly complicated background in the m/z range of 1000 – 2000 was generated with 0.5% Fluka 3-10 due to a different mass distribution of amphoteric species. The background signal for 0.5% Servalyt 3-10 was 50% lower than Fluka, and less interferences from ampholytes was observed in the m/z range 1500 – 2000, which was well-suited for the analysis of large proteins (e.g., MW>30 kDa). A 34 min-broad pH gradient zone was produced by both Fluka and Servalyt, with good pH linearity ($R^2=0.985$ and 0.997). However, based on the extracted ion electropherogram of peptide markers, separation with Servalyt resulted in sharp peaks for both basic and acidic peptide markers (data not shown). The broad peaks caused by the defocusing effect from moving ionic boundary was observed in the acidic region for Fluka (details on the defocusing effect will be discussed later). The results revealed that the ability of Servalyt to stabilize the pH gradient during the focusing and mobilization stages was better. In addition, separation with 0.5% Servalyt 3-10 led to improved resolution and higher intensities for the two isoforms of Myo, and increased sensitivities for the Rnase (pI 9.6) and α-LA (pI 4.5). Finally, Servalyt 3-10 was considered optimum for the liquid junction integrated CIEF-MS system, since it provides good focusing, moderate background noise, and satisfactory resolution in a reasonable total analysis time.
Figure 4-9. CIEF-MS with 0.5 % (v/v) Fluka 3-10 and Servalyt 3-10 using chemical mobilization. (A) Plot of peptide pI as a function of migration time, (B) ion density map, (C) base peak electropherogram and (D) summed extracted ion electropherogram of proteins separated with (i) Fluka 3-10 and (ii) Servalyt 3-10. Experimental conditions: focusing: 22 kV, 35 min; mobilization: 22 kV; sheath liquid delivered with 0.5 psi.
To improve the resolution for proteins (and protein isoforms) with small differences in pIs (CA, pI 5.9; β-LG B, pI 5.3; β-LG A, pI 5.2; α-LA, pI 4.5), a narrow range ampholyte, Servalyt 4-6, was employed alone or in combination with the broad range ampholyte Servalyt 3-10. Servalyt 4-6 provided a 2-fold wider pH gradient with a 53% smaller slope (0.338 vs. 0.180) over the pH range 4-6 compared to Servalyt 3-10. However, due to the poor focusing ability of 0.5% Servalyt 4-6, broad peaks were observed for proteins (average $w_{1/2}=32$ s). Increasing the concentration of Servalyt 4-6 to 1% improved the focusing for β-LG and α-LA, and the average $w_{1/2}$ was reduced for 2.4-fold; however, CA with pI (5.9), close to the basic boundary of the pH gradient zone, failed to get focused. CIEF-MS with a combination of 0.5% Servalyt 3-10 and 0.5% Servalyt 4-6 led to small improvements in the resolution of the two isoforms of β-LG (increased from 0.6 to 0.8) and the resolution between β-LG A and α-LA (increased from 1.2 to 1.6). However, the signal intensity of proteins decreased 8-fold due to the ion suppression from the higher concentration of ampholytes. Eventually, 0.5% Servalyt 3-10 was used for further optimization of CIEF-MS conditions.

4.4.5.3 Mobilization parameters

**Effect of pressure on CIEF-MS performance**

Mobilization of the focused protein zones has been demonstrated using electroosmotic flow, electrophoretic flow, hydrodynamic flow or a combination. The versatility of our interface configuration allows the employment of different mobilization methods. Figure 4-10 presents the separation of proteins and peptide markers using pressure (top) and chemical mobilization (bottom). Mobilization with 0.5 psi applied at the capillary inlet produced a more compressed pH
gradient zone, which was 7 min-narrower and with 2-fold larger slope than chemical mobilization. Sharper peaks with an average of 6 s smaller $w_{1/2}$ were detected with significantly reduced migration times using pressure mobilization. We noticed that the acidic proteins and peptides experienced more peak broadening than the neutral or basic proteins and peptides, likely caused by the moving ionic boundary. At the mobilization stage, when the acetate ions from the sheath liquid penetrated into the basic pH boundary at the capillary terminus to replace the hydroxide ions, a moving ionic boundary was developed inside the capillary, traveling against the focused sample zones. The sharp changes in pH and conductivity at the ionic boundary can potentially degrade the separation of the focused protein or peptide bands. The acidic proteins and peptides travelled longer distances than the basic and neutral species inside the capillary, thus experiencing more defocusing effect from the moving ionic boundary. However, compared to chemical mobilization, less peak broadening was found for the acidic proteins and peptides when pressure mobilization was used. This likely resulted from the hydrodynamic flow introduced by the small pressure difference between the two sides of the capillary which minimized the propagation of the moving ionic boundary. Nevertheless, employment of chemical mobilization resulted in a much cleaner background in the ion density map and a 5-fold increase in protein signal intensity. Without applying the pressure on the separation capillary, the amount of ampholytes and glycerol introduced into the electrospray was minimized; thus reduced ion suppression and enhanced ionization efficiency were obtained with chemical mobilization. In addition, an improvement in peak resolution was found for the model proteins. CA (pI 5.9) and β-LG (pI 5.2 – 5.3), with small differences in pI, co-migrated in the extracted ion electropherogram from pressure mobilization; however, they were resolved when chemical
mobilization was employed. Further, the two deamidated forms of Rnase were baseline separated from the non-deamidated form (see later with the high resolution MS data).

From comparison of the results for two mobilization methods, it is clear that pressure mobilization offers rapid separation of proteins and protein isoforms with high efficiency. On the contrary, chemical mobilization provides high resolution for proteins with minor pI differences, enhanced protein signal intensity and improved electrospray stability at the price of longer analysis times.
Figure 4-10. CIEF-MS with different mobilization methods. (A) Plot of peptide pI as a function of migration time, (B) summed extracted ion electropherogram of peptide markers, (C) ion density map and (D) summed extracted ion electropherogram of proteins with (i) pressure and (ii) without pressure applied at the inlet of the separation capillary during the mobilization stage. The peak labeled as "D" and “DD” are the mono-deamidated and di-deamidated Rnase, respectively. Experimental conditions: proteins were mixed with 0.5% Servalyt 3-10; injection: catholyte/sample-ampholytes/anolyte=1/2/1 (v/v/v); focusing: 22 kV, 30 min; mobilization: 22 kV, (i) 0.5 psi and (ii) no pressure applied at the capillary inlet; sheath liquid delivered with 0.5 psi.
The hydrodynamic flow introduced by the pressure difference on the separation capillary can overcome the adverse effects caused by the moving ionic boundary, but it also reduces the resolution. To understand the impact of the pressure on CIEF-MS performance with the liquid junction integrated system, mobilization at three different pressures (0.1, 0.5 and 1.0 psi) was performed. Mobilization with 1.0 psi resulted in the fastest separation but the shortest time span for the pH gradient zone. An average of 9.8 s w_{1/2} was obtained for the peptides and proteins, which was 60% narrower compared to mobilization with 0.1 psi. The rapid mobilization and less diffusion caused by moving ionic boundary led to narrower peaks with 10-fold higher intensities for the α-LA (pI 4.5) and Rnase (pI 9.6), two proteins located at the end of the pH boundary (pH 3.0 – 10.0). Although applying relatively high pressure (1.0 psi compared to 0.1 psi) potentially introduced more ampholytes and glycerol into the electrospray, higher pressure could also applied at the inlet of the sheath liquid line. As a consequence, the signal intensities for proteins were not reduced significantly. However, a broad separation window and less hydrodynamic flow caused by 0.1 psi contributed to the enhanced resolution for the proteins and protein isoforms. The two isoforms of Myo were baseline separated, and the acidic proteins (CA, β-LG and α-LA) were better resolved with 0.1 psi. In addition, the maximum peak capacity also increased from 80 to 130 upon decreasing the inlet pressure from 1.0 to 0.1 psi. Taken together, we have demonstrated how to optimize the inlet pressure on pressure mobilization to achieve a balance in separation efficiency and MS response for protein separation.
Effect of sheath liquid flow on CIEF-MS performance

In the liquid junction integrated CIEF-MS system, the composition and flow rate of sheath liquid not only affect the mobilization process, but also impact the electrospray ionization. We first evaluated the effect of sheath liquid composition on CIEF-MS performance, using a hydro-organic sheath liquid, 1% acetic acid and 10% methanol (pH 3.0); and a volatile aqueous sheath liquid, 10 mM ammonium acetate (pH 5.0). Pressure mobilization with the two sheath liquids both generated a 26-min wide pH gradient zone with good linearity \((R^2=0.975\) and 0.994). Due to the low pH of the sheath liquid composed of 1% acetic acid and 10% methanol, the mobilization process of the focused proteins was faster than with 10 mM ammonium acetate at pH 5.0. The relatively high contents of protons and organic modifier (methanol) were favored for the electrospray process; hence, 20% enhanced signal intensities for the model proteins and a 4-fold improvement in signal intensities were observed when a higher sheath flow rate was applied (data not shown). On the other hand, pressure mobilization with 10 mM ammonium acetate resulted in narrower peaks with a 37% decrease in \(w_{1/2}\) for the model proteins. In addition, the resolution for Myo I and Myo II was improved from 2.0 to 3.4. It was likely that the use of 10 mM ammonium acetate as the sheath liquid suppressed the moving ionic boundary effect during the mobilization process. Thus the focused protein bands were stabilized, and improved resolution was achieved. Furthermore, mixing the capillary effluent with either the hydro-organic sheath liquid or the aqueous sheath liquid led to different protein conformations in the gas phase. As shown for the Myo ions, the most abundant charge state shifted from +13 to +12, indicating somewhat less denaturation with 10 mM ammonium acetate (pH 5.0). Shifting the protein charge envelopes to higher m/z scan range can potentially minimize interference of ampholyte ions and simplify the interpretation of the MS spectrum. Additionally, by varying the sheath liquid composition, the analysis of a protein either in the denatured or native state can be accomplished.
by online CIEF-MS, as demonstrated by Przybylski et al. using a commercial high sheath-flow interface 81.

The influence of sheath liquid flow rate on the separation and detection was next investigated. The sheath liquid composed of 1% acetic acid and 10% methanol was delivered to the junction by 0.5 and 1.0 psi applied at the inlet of the sheath liquid capillary to initiate the mobilization of the focused protein bands. The increase in the sheath liquid flow rate led to a moderate decrease in the linear correlation between pI and migration time. However, a 3.7-fold increase in the protein signal intensity was obtained due to the decrease in electrospray ion suppression and improved the ionization efficiency with higher sheath flow. A maximum peak capacity of 130 was achieved with chemical mobilization and 1.0 psi for the sheath liquid flow. Further increase the sheath liquid flow rate may contribute to a strong moving ionic boundary effect and high dilution to reduce the separation efficiency; thus only pressures up to 1.0 psi was employed in the integrated CIEF-MS system for chemical mobilization.

4.4.5.4 Repeatability of the protein separation

The repeatability of the protein separation with the integrated CIEF-MS system was evaluated with the triplicate CIEF-MS runs. Figure 4-11 presents a comparison of the summed extracted ion electropherogram of the proteins in the technical replicates. Similar separation profiles were obtained across the three runs, with a % CV of 10.8 for the base peak intensities. As shown in Table 4-2, the % CVs of the migration time for the model proteins in the run-to-run replicates remained below 2.1 %, using the peptide marker (pI 10.0) as an internal standard for calibration. The results demonstrate good technical repeatability of the liquid junction interface in coupling commercial CE instrument and mass spectrometer for CIEF-MS analysis.
Figure 4-11. Summed ion electropherograms of the model proteins analyzed by CIEF-MS system in triplicate runs.
Table 4-2. Repeatability in migration time measured in triplicate CIEF-MS runs

<table>
<thead>
<tr>
<th>Protein</th>
<th>Average MT ratio *</th>
<th>Standard deviation</th>
<th>%CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rnase</td>
<td>1.02</td>
<td>0.01</td>
<td>0.8</td>
</tr>
<tr>
<td>Myo I</td>
<td>1.37</td>
<td>0.03</td>
<td>2.1</td>
</tr>
<tr>
<td>Myo II</td>
<td>1.52</td>
<td>0.03</td>
<td>1.8</td>
</tr>
<tr>
<td>CA</td>
<td>1.86</td>
<td>0.02</td>
<td>1.2</td>
</tr>
<tr>
<td>β-LG B</td>
<td>1.93</td>
<td>0.03</td>
<td>1.5</td>
</tr>
<tr>
<td>β-LG A</td>
<td>1.97</td>
<td>0.04</td>
<td>2.0</td>
</tr>
<tr>
<td>α-LA</td>
<td>2.14</td>
<td>0.03</td>
<td>1.3</td>
</tr>
</tbody>
</table>

* The migration time (MT) ratio was calculated using the migration time for the protein divided by the migration time for peptide marker (pI 10.0).

4.4.5.5 Characterization of model protein mixture

With the liquid junction integrated CIEF-MS system, high resolution separation of a mixture of model proteins was achieved. A maximum peak capacity of 75 was obtained with the employment of pressure mobilization and 190 with chemical mobilization when CIEF-MS was performed in a bare fused silica capillary with 20% glycerol in the separation medium (data not shown). To further enhance the protein solubility and reduce the EOF, CIEF-MS was performed in the separation medium containing 30% glycerol. A mixture of 0.5% (v/v) Servalyt 3-10 and 0.5% (v/v) Servalyt 4-6 was employed as the carrier ampholytes to achieve improved separation of acidic proteins. Focusing was carried out with 30 kV voltage applied on the 80-cm long separation capillary. After 20 min of focusing, the sheath liquid composed of 1% acetic acid and 10% methanol was delivered to the liquid junction with 1.0 psi to initiate the chemical mobilization of focused protein bands.

To fully identify and characterize the proteins and protein isoforms, an Orbitrap Elite mass spectrometer was used to provide high resolution and high precision intact mass measurement. In
order to differentiate the minor mass shift of protein modifications such as deamidation, the mass spectrometer was operated at 120,000 (m/z 400) resolution with a scan range of 1200 – 2000 Th for the model proteins. The cutoff range (m/z > 1200) allowed effective removal of strong interferences from ampholyte ions and enabled the accumulation of large numbers of protein ions to improve the detection sensitivity. Figure 4-12 presents the CIEF-MS separation of the mixture of model proteins. Separation in the 30% glycerol-water medium yielded a 22.3 min-wide pH gradient zone. All proteins were separated in a reasonable total analysis time with good resolution. A linear correlation (R²=0.969) of protein pI and migration time was obtained with the chemical mobilization method. The model proteins were detected as narrow and sharp peaks (averaged w₁/₂ =13.5 s) in the extracted ion electropherograms. A maximum peak capacity of 210 demonstrated high resolution. In addition, the relatively high intensity (8E6 – 3E7) for the model proteins revealed the improved detection sensitivity upon using 30% glycerol. In low-sheath flow rate system, separation with 1% ampholytes in 20% glycerol-water resulted in ion suppression of the proteins. However, the significant decrease in apparent mobilities of the proteins due to the high viscosity of 30% glycerol led to lower ion suppression and higher ionization efficiency.
Figure 4-12. CIEF-MS analysis of the model protein mixture using 30% glycerol in the separation medium. (A) Plot of protein pI as a function of migration time, (B) summed extracted ion electropherogram (top) and extracted ion electropherograms, and (C) ESI-MS spectra of model proteins. The PA800 plus CE system was coupled to the Orbitrap Elite mass spectrometer with the liquid junction interface. Experimental conditions: proteins were mixed with 0.5% Servalyt 3-10 and 0.5% Servalyt 4-6; injection: catholyte/sample-ampholytes/anolyte=1/2/1 (v/v/v); focusing: 30 kV, 20 min; mobilization: 30 kV; sheath liquid delivered with 1.0 psi.
The relatively clean background and the isotopically resolved MS spectrum facilitated the unambiguous identification of protein modifications with small mass shifts on the co-migrated proteins and protein isoforms. Figure 4-13 illustrates the identification of modifications on the intact Rnase. The extracted ion electropherogram of intact Rnase (m/z 1521) with a 0.5 Da mass tolerance showed three peaks almost baseline resolved (Rs=1.5 and 2.0). The monoisotopic MW (13673.30) of the major peak determined from the deconvolution spectrum matched the mass of intact Rnase. The two minor peaks had the identical charge distribution as the major peak and the isotopic envelopes of the most abundant charge state (9+) showed a small shift of m/z 0.1105 and 0.2177, corresponding to a mass shift of +0.99 and +1.96 Da. The deconvolution results revealed that the two peaks corresponded to the Rnase with mono- and di-deamidation. Deamidation on a single site introduced a negative charge and led to a shift of protein pI to a more acidic pH region. Compared to the non-deamidated Rnase, the deamidated forms focused at positions closer to the cathodic side and resulted in longer migration times. The experimentally determined ΔpI for non-deamidated, mono-deamidated (-0.10 unit) and di-deamidated Rnase (-0.29 unit) were in good agreement with the theoretical values. In addition, protein oxidation, which did not contribute to any change in protein pI, was also identified in the deconvoluted spectra with a characteristic mass shift of +15.98 Da. The highly accurate delta mass also allowed us to assign other modifications on Rnase to the combinations of different numbers of deamidation and oxidation.

The monoisotopic mass of the model proteins measured from the electrospray mass spectra are summarized in Table 4-3, and the potential modifications are assigned to each identified protein and protein isoform according to the delta mass. Besides deamidation and oxidation, other protein modifications were also detected with the integrated system. The noncovalent
bindings of one and two heme prosthetic groups on Myo were identified with a delta mass of +615.22 and + 1230.33 Da \(^{83,84}\). The addition of 42.04 Da mass to Myo could be due to acetylation, which shifted the protein pI to acidic region and to a 1.7 min increase in the migration time. The two natural variants in β-LG with two different amino acids present on the sequence were detected. The change of Ala to Val did not contribute any difference in protein net charge; however, the substitution of the Gly to Asp in β-LG A introduced a negative charge and shift of the protein pI by - 0.1 unit. The expected mass difference of the four amino acids was 86.03 Da, which agreed with the experimentally determined delta mass (85.99 Da) between the two variants of β-LG. Furthermore, glycation was observed on both β-LG B and β-LG A with a mass discrepancy of + 324.13 Da for each lactose unit \(^{85}\). Such modification did not lead to any change in protein pI; thus the glycated modified β-LG isoforms co-migrated with the unmodified forms. The online coupling of CIEF-ESI ME allowed the simultaneous measurement of pI and accurate mass on the intact protein, providing valuable information on the identities of protein modifications.
Figure 4-13. Identification of post-translational modifications on Rnase. (A) Extracted ion electropherogram of Rnase, (B) intact MS spectra of the three peaks shown in (A), (C) comparison of the isotope distributions of the most abundant charge state (9+) and (D) deconvolution spectra of the three peaks. Experimental conditions were the same as Figure 4-12. Modifications: N, non-deamidation; D, mono-deamidation; DD, di-deamidation; O, mono-oxidation.
Table 4-3. Summary of Identified Protein and Potential Protein Modifications.

<table>
<thead>
<tr>
<th>Protein</th>
<th>MT (min)</th>
<th>MW Obs. (Da)</th>
<th>Delta mass (Da)</th>
<th>Potential Modification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rnase</td>
<td>57.62</td>
<td>13673.30</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>Rnase</td>
<td>57.62</td>
<td>13689.28</td>
<td>15.98</td>
<td>Oxidation (1)</td>
</tr>
<tr>
<td>Rnase</td>
<td>58.01</td>
<td>13674.32</td>
<td>1.01</td>
<td>Deamidation (1)</td>
</tr>
<tr>
<td>Rnase</td>
<td>58.01</td>
<td>13690.29</td>
<td>16.99</td>
<td>Deamidation (1) + Oxidation (1)</td>
</tr>
<tr>
<td>Rnase</td>
<td>58.70</td>
<td>13675.24</td>
<td>1.93</td>
<td>Deamidation (2)</td>
</tr>
<tr>
<td>Rnase</td>
<td>58.70</td>
<td>13691.27</td>
<td>17.97</td>
<td>Deamidation (2) + Oxidation (1)</td>
</tr>
<tr>
<td>Myo</td>
<td>61.63&amp;63.29</td>
<td>16941.01</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>Myo</td>
<td>61.63</td>
<td>17556.22</td>
<td>615.22</td>
<td>Heme-binding (1)</td>
</tr>
<tr>
<td>Myo</td>
<td>61.63</td>
<td>18171.34</td>
<td>1230.33</td>
<td>Heme-binding (2)</td>
</tr>
<tr>
<td>Myo</td>
<td>63.29</td>
<td>16983.05</td>
<td>42.04</td>
<td>Acetylation (1)</td>
</tr>
<tr>
<td>Myo</td>
<td>63.29</td>
<td>16999.03</td>
<td>58.02</td>
<td>Acetylation (1) + Oxidation (1)</td>
</tr>
<tr>
<td>CA</td>
<td>67.79</td>
<td>29006.72</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>CA</td>
<td>68.27</td>
<td>29007.75</td>
<td>1.03</td>
<td>Deamidation (1)</td>
</tr>
<tr>
<td>β-LG B</td>
<td>68.27</td>
<td>18265.49</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>β-LG B</td>
<td>68.27</td>
<td>18589.50</td>
<td>324.01</td>
<td>Lactose (1)</td>
</tr>
<tr>
<td>β-LG A</td>
<td>68.86</td>
<td>18351.48</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>β-LG A</td>
<td>68.86</td>
<td>18675.61</td>
<td>324.13</td>
<td>Lactose (1)</td>
</tr>
<tr>
<td>β-LG A</td>
<td>68.86</td>
<td>18999.67</td>
<td>648.19</td>
<td>Lactose (2)</td>
</tr>
<tr>
<td>β-LG A</td>
<td>68.86</td>
<td>18367.43</td>
<td>15.95</td>
<td>Oxidation (1)</td>
</tr>
<tr>
<td>α-LA</td>
<td>70.62</td>
<td>14168.82</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>α-LA</td>
<td>70.62</td>
<td>14184.75</td>
<td>15.93</td>
<td>Oxidation (1)</td>
</tr>
</tbody>
</table>
4.4.5.6 Analysis of recombinant human growth hormone

The feasibility of the integrated CIEF-MS system for separation and identification of degradation products in a therapeutic protein was investigated. Recombinant human growth hormone (rhGH, pI 5.1), a 191-amino acids single chain protein with two intra-chain disulfide bonds, was used as the model protein. To determine the degradation products, online CIEF-MS analysis was carried out on non-stressed and heat and alkaline stressed rhGH. The forced degradation of rhGH was performed by incubating the protein at 37 °C for 2 weeks in a strong alkaline buffer (pH 9.5). Significant precipitation of rhGH was observed when the protein was concentrated into a narrow zone where the local pH was the same as the intrinsic pI, and the same phenomena was also reported earlier by Frenz et al. Increasing the glycerol content (to 30%) in the sample and background electrolyte, decreasing the sample concentration of rhGH (to 0.1 mg/mL) and reducing the focusing time were investigated to overcome protein precipitation during the CIEF process. Employment of the non-ionic detergent, octyl β-D-glucopyranoside (OG) was also studied. It was found using 0.7% (w/v) OG in the sample buffer could considerably improve the protein solubility and reduce the protein precipitation in a CIEF-MS experiment. Due to the minimized EOF in a bare fused silica capillary when using 30% glycerol, the low level of OG introduced into the ESI ion source did not contribute to apparent ion suppression. In addition, the charge envelope of rhGH was shifted to a higher m/z range upon adding 0.7% OG, further reducing the interference of ampholyte ions on the mass spectrum of the protein. To improve separation, 1% Servalyt 4-6 was used as the carrier ampholyte to create a wide pH gradient in the pH region 4-6 for improved resolution of the protein isoforms with very similar pI values (close to 5.1). The sample was mixed with the ampholyte and 0.7% OG supplemented in 30% glycerol and injected into the capillary with a
catholyte/sample/anolyte ratio of 1:2:1. The relatively high concentration of salts in the rhGH formulation can disturb the stable pH gradient. To remove the interference of the salts, a 5 min-desalting ramp voltage (0 – 10 kV) was applied prior to the focusing step. The focusing step was carried out at 30 kV constant voltage for 20 min, and the focused bands were mobilized using 0.5 psi applied at the capillary inlet.

Figure 4-14 depicts the separation of rhGH before and after the forced degradation treatment. A single peak was observed in the extracted ion electropherogram of the non-stressed rhGH, corresponding to the unmodified intact protein (observed monoisotopic MW= 22111.1296, + 4.0 ppm mass error). At a similar migration time, a peak with a shoulder was found in the extracted ion electropherogram of the stressed rhGH. Deconvolution of the mass spectra revealed the components in the stressed rhGH significantly changed compared to the non-stressed sample. The reconstructed ion electropherograms of the major components (after mass deconvolution) in the stressed rhGH are shown in Figure 4-14B. The non-deamidated form was present at less than 3% of the whole sample. High abundant components (observed monoisotopic MW=22112.1128, 22113.1732 and 22114.1790) with delta masses of + 0.94, +2.00 and +3.00 Da were found in the stressed sample. These forms were likely singly, doubly and triply deamidated rhGH, as further supported by the increased migration time due to the more negative charges introduced by the higher level of deamidation. Oxidation on single to triple sites were also identified in the mass spectra. In addition, a minor constituent with an observed mass of 22129.1137 Da was detected co-migrating with the intact rhGH, and a mass discrepancy of + 17.94 Da (+ H2O) indicated the occurrence of chain-cleavage on the intact rhGH. Furthermore, a mass difference of – 244.13 Da was found on the non-deamidated and deamidated forms of rhGH. The shift in mass potentially resulted from the elimination of the N-terminal two amino acids (Phe and Pro) under
the strong alkaline conditions, as also reported by Taichrib et al. A trace amount of the mutation variant (His18 to Gln18) was found to co-migrate with the triply-deamidated rhGH with a characteristic delta mass of -9.06 Da. The change of His to Gln shifted the protein pI to a more acidic region and resulted in a longer migration time. Table 4-4 summarizes the identified isoforms of rhGH and the potential modifications. Although the different modified forms of rhGH were not fully separated in the electropherogram, the accurate mass determined from high resolution Orbitrap mass spectrometer still provided a powerful means to identify the types of the potential modifications.
Figure 4-14. Identification of modifications on rhGH. (A) Extracted ion electropherogram of rhGH (i) before and (ii) after forced degradation and (B) reconstructed ion electropherogram of the non-, mono-, di- and tri-deamidated rhGH after deconvolution. Modifications: N, non-deamidation; D, mono-deamidation; DD, di-deamidation; DDD, tri-deamidation. The data was filtered by a 7-point Gaussian filter.
Table 4-4. Summary of Identified Proteoforms in stressed rhGH and Potential Protein Modifications.

<table>
<thead>
<tr>
<th>MT (min)</th>
<th>MW Obs. (Da)</th>
<th>Delta mass (Da)</th>
<th>Potential Modification</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.72</td>
<td>22111.1711</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>11.72</td>
<td>22129.1137</td>
<td>+ 17.94</td>
<td>Chain cleavage (1)</td>
</tr>
<tr>
<td>12.25</td>
<td>22112.1128</td>
<td>+ 0.94</td>
<td>Deamidation (1)</td>
</tr>
<tr>
<td>12.25</td>
<td>21867.0334</td>
<td>- 244.13</td>
<td>Elimination (Phe1Pro2)</td>
</tr>
<tr>
<td>12.25</td>
<td>22130.1330</td>
<td>+ 18.96</td>
<td>Chain cleavage (1) + deamidation (1)</td>
</tr>
<tr>
<td>12.25</td>
<td>22143.9448</td>
<td>+ 31.94</td>
<td>Oxidation (2)</td>
</tr>
<tr>
<td>12.25</td>
<td>22144.1378</td>
<td>+ 32.97</td>
<td>Oxidation (2) + deamidation (1)</td>
</tr>
<tr>
<td>12.25</td>
<td>22160.1338</td>
<td>+ 48.97</td>
<td>Oxidation (3) + deamidation (1)</td>
</tr>
<tr>
<td>12.65</td>
<td>22113.1732</td>
<td>+ 2.00</td>
<td>Deamidation (2)</td>
</tr>
<tr>
<td>12.65</td>
<td>21868.0206</td>
<td>- 243.15</td>
<td>Elimination (Phe1Pro2) + deamidation (1)</td>
</tr>
<tr>
<td>12.65</td>
<td>22145.1040</td>
<td>+ 33.93</td>
<td>Oxidation (2) + deamidation (2)</td>
</tr>
<tr>
<td>13.04</td>
<td>22102.1123</td>
<td>- 9.06</td>
<td>His18 to Gln18</td>
</tr>
<tr>
<td>13.04</td>
<td>22114.1790</td>
<td>+ 3.00</td>
<td>Deamidation (3)</td>
</tr>
<tr>
<td>13.04</td>
<td>22146.1149</td>
<td>+ 34.92</td>
<td>Oxidation (2) + deamidation (3)</td>
</tr>
<tr>
<td>13.04</td>
<td>22162.0949</td>
<td>+ 50.92</td>
<td>Oxidation (3) + deamidation (3)</td>
</tr>
</tbody>
</table>
4.5 CONCLUSIONS

The studies presented here demonstrate a simple and robust interfacing strategy to hyphenate commercial CE and MS instrument for automatic peptide and protein analysis. The assembly of the interface only requires the alignment of the separation capillary column and the electrospray tip using the standard fittings from the commercially available sources. The interface design decouples the electrophoresis and electrospray ionization process and enables the independent optimization of CE and ESI-MS operational parameters. The interface utilizes (1) the built-in auto-sampler in the commercial CE instrument to perform the precise sample injection, (2) the power control system to conduct the voltage ramp for electrophoresis and (3) the pressure delivery system to control the sheath liquid flow. In addition, the interface design allows the maintenance of a stable electric contact at the outlet of the separation column during the entire electrophoretic and ESI processes. The special tip geometry guarantees the efficient mixing of the capillary effluent with the MS compatible sheath flow with minimized post-column broadening.

Several electrophoretic modes have been applied to the liquid junction interface CE-ESI-MS system. Reasonably high separation efficiency was obtained for the CZE-MS analysis of a BSA digest using an uncoated capillary. In addition, online CIEF-MS analysis of intact proteins was demonstrated with the integrated system. Separation of a mixture of model proteins resulted in near baseline separation of the deamidated variants of the ribonuclease and reasonable high resolution of β-lactoglobulin isoforms with closely related pIs. Finally, the technique has been applied to study the degradation products of the therapeutic protein rhGH. Incorporating the high resolution-high mass accuracy mass spectrometer to the CIEF-ESI-MS system allowed the identification of protein modifications with only minor change (e.g. + 0.998 Da) in mass even if
the modified and unmodified forms were not fully resolved from the first separation dimension (i.e., the CIEF process). The low sheath flow liquid junction integrated system should be a useful approach to study the charge variants of biopharmaceuticals.

4.6 REFERENCES


Chapter 5: Summary and Perspectives
This dissertation attempts to incorporate various high resolution separation techniques with MS detection to enhance the identification and characterization of proteins and protein variants.

Despite the advances in disulfide structure analysis by LC-MS, the sample preparation workflow can potentially introduce disulfide scrambling and mislead the assignment of the correct disulfide linkages in therapeutic proteins. The work in Chapter 2 has demonstrated a robust and straightforward sample preparation strategy designed to address this analytical challenge. The strategy described in this part incorporates optimized chemical derivatization conditions to minimize the potential artifacts introduced by improper sample handling. With the established methodology, both peptides containing the correct linkages and the scrambled disulfide linkages can be identified and quantified in a highly reproducible manner.

Therapeutic protein origin from different lots may present microheterogeneity (e.g., different glycosylation) and thus the partial reduction and alkylation conditions should be optimized. Using a combination of LC-UV detection and LC-MS analysis to separate and determine and quantitate the partial reduction and alkylation products facilitates the exploration of the optimized conditions for proteins from different lots and also provides insight into the monitoring of the lot-to-lot difference. Another promising approach to effectively avoid method-induced disulfide scrambling is to utilize top-down mass spectrometry for protein characterization. Although top-down analysis of the intact neublastin dimer failed to elucidate the disulfide linkages in preliminary experiments due to the limited fragmentation generated from the compact structure formed by the two cystine knots, top-down analysis of the isolated, partially reduced and alkylated monomer containing only one cystine knot may produce some information-rich fragment ions and assist the mapping of the disulfide linkages in neublastin.
In Chapter 3, the chemical proteomic-based approach combining SILAC labeling, SDS-PAGE fractionation and LC-MS analysis has been applied for the unbiased identification of the potential protein target PAK4 to a small molecule drug. Although PAK4 is found differentially expressed in the cells treated with and without the small molecule drug, it may be a member of a protein complex enriched by the small molecule drug. To guarantee that PAK4 is the correct protein target, target validation via target LC-MS analysis or generic approaches (e.g., overexpression, knock-down or mutation of PAK4) should be performed.

In Chapter 4, a new interface to integrate CE and MS for online protein analysis has been developed. In the liquid junction interface CIEF-ESI-MS system, proteins and protein variants with minor differences in pI can be separated with high resolution CIEF and then the co-migrated variants further differentiated by high resolution MS. The capability of the liquid junction interface CIEF-ESI-MS system in the enrichment and separation of protein charge variants has been demonstrated with a mixture of model proteins and the degraded therapeutic protein.

Some aspects can be studied to improve the performance and extend the application of the liquid junction interface integrated system. First, the reproducibility and separation efficiency can be enhanced by using a neutral coated capillary to replace the bare fused silica capillary. A coating such as polyvinyl alcohol and polyacrylamide can effectively minimize the interaction between the capillary wall and proteins/ampholytes and further suppress the EOF inside the capillary. However, the interface operational parameters, such as the sample injection method and capillary reconditioning steps should be optimized since the coating should not be exposed to strong acidic, basic and organic environments for a long time. Secondly, we have demonstrated that the tip geometry has a great impact on the separation and electrospray
ionization processes. To further improve the interface performance, the tip-to-capillary distance and opening of electrospray tip could be further optimized using different combinations of capillary and electrospray tip dimensions. For example, an electrospray tip with a larger i.d. would allow the separation capillary to be positioned closer to the tip orifice. Nevertheless, the dilution ratio of the capillary effluent to the sheath liquid flow will be affected by the limited mixing volume. Additionally, another approach to facilitate the online CIEF-MS analysis of protein would be to explore new synthesized ampholytes with less components and with lower molecular weights to yield lower ion suppression and produce less interference of the mass spectrum of the protein. Novel ampholytes should be able to provide a stable pH gradient in the designated pI range of the protein analyte in order to separate the various types of protein charge variants.

The potential application of the low-sheath-flow liquid junction interface developed in this dissertation can be integrated to commercial CE and MS instruments for automatic online analysis for charge heterogeneity profiling (e.g., glycol-profile) of biopharmaceuticals. Meanwhile, another attractive feature of the system is it allows the protein to maintain a folded state during separation. Further, the sheath flow enables addition of chemical adducts in order to modify the protein at the capillary exit to fulfill different experimental purposes. For example, charge-enhancing compounds can be added to improve the ionization of the protein. Possibly a reducing agent can be infused to break the disulfide bonds of the intact protein after separation and improve the fragmentation in tandem MS/MS analysis.
This is a License Agreement between Siyuan Liu ("You") and Nature Publishing Group ("Nature Publishing Group") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Nature Publishing Group, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

<table>
<thead>
<tr>
<th>License Number</th>
<th>3822760811861</th>
</tr>
</thead>
<tbody>
<tr>
<td>License date</td>
<td>Mar 05, 2016</td>
</tr>
<tr>
<td>Licensed content publisher</td>
<td>Nature Publishing Group</td>
</tr>
<tr>
<td>Licensed content publication</td>
<td>Nature Reviews Molecular Cell Biology</td>
</tr>
<tr>
<td>Licensed content title</td>
<td>Formation and transfer of disulphide bonds in living cells</td>
</tr>
<tr>
<td>Licensed content author</td>
<td>Carolyn S. Sevier and Chris A. Kaiser</td>
</tr>
<tr>
<td>Licensed content date</td>
<td>Nov 1, 2002</td>
</tr>
<tr>
<td>Volume number</td>
<td>3</td>
</tr>
<tr>
<td>Issue number</td>
<td>11</td>
</tr>
<tr>
<td>Type of Use</td>
<td>reuse in a dissertation / thesis</td>
</tr>
<tr>
<td>Requestor type</td>
<td>academic/educational</td>
</tr>
<tr>
<td>Format</td>
<td>print and electronic</td>
</tr>
<tr>
<td>Portion</td>
<td>figures/Tables/Illustrations</td>
</tr>
<tr>
<td>Number of figures/tables/illustrations</td>
<td>1</td>
</tr>
<tr>
<td>High-res required</td>
<td>no</td>
</tr>
<tr>
<td>Figures</td>
<td>BOX1 Thiol-disulphide exchange between proteins</td>
</tr>
<tr>
<td>Author of this NPG article</td>
<td>no</td>
</tr>
<tr>
<td>Your reference number</td>
<td>None</td>
</tr>
<tr>
<td>Title of your thesis / dissertation</td>
<td>STATE-OF-ART SEPARATION – MASS SPECTROMETRY IN PROTEIN CHARACTERIZATION</td>
</tr>
<tr>
<td>Expected completion date</td>
<td>Apr 2016</td>
</tr>
<tr>
<td>Estimated size (number of pages)</td>
<td>219</td>
</tr>
<tr>
<td>Total</td>
<td>0.00 USD</td>
</tr>
</tbody>
</table>

Terms and Conditions

Nature Publishing Group hereby grants you a non-exclusive license to reproduce this material for this purpose, and for no other use, subject to the conditions below:

https://10.1038/s41560-018-0128-9
1. NPG warrants that it has, to the best of its knowledge, the rights to license reuse of this material. However, you should ensure that the material you are requesting is original to Nature Publishing Group and does not carry the copyright of another entity (as credited in the published version). If the credit line on any part of the material you have requested indicates that it was reprinted or adapted by NPG with permission from another source, then you should also seek permission from that source to reuse the material.

2. Permission granted free of charge for material in print is also usually granted for any electronic version of that work, provided that the material is incidental to the work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version. Where print permission has been granted for a fee, separate permission must be obtained for any additional, electronic re-use (unless, as in the case of a full paper, this has already been accounted for during your initial request in the calculation of a print run). NB: In all cases, web-based use of full-text articles must be authorized separately through the 'Use on a Web Site' option when requesting permission.

3. Permission granted for a first edition does not apply to second and subsequent editions and for editions in other languages (except for signatories to the STM Permissions Guidelines, or where the first edition permission was granted for free).

4. Nature Publishing Group's permission must be acknowledged next to the figure, table or abstract in print. In electronic form, this acknowledgement must be visible at the same time as the figure/table/abstract, and must be hyperlinked to the journal's homepage.

5. The credit line should read:
   Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)
   For AOP papers, the credit line should read:
   Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME], advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].XXXX)

   Note: For republication from the British Journal of Cancer, the following credit lines apply.
   Reprinted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication) For AOP papers, the credit line should read:
   Reprinted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME], advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].XXXX)

6. Adaptations of single figures do not require NPG approval. However, the adaptation should be credited as follows:
   Adapted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)

   Note: For adaptation from the British Journal of Cancer, the following credit line applies.
   Adapted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)

7. Translations of 401 words up to a whole article require NPG approval. Please visit http://www.macmillanmedicalcommunications.com for more information. Translations of up to 400 words do not require NPG approval. The translation should be credited as follows:
   Translated by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication).

   Note: For translation from the British Journal of Cancer, the following credit line
We are certain that all parties will benefit from this agreement and wish you the best in the use of this material. Thank you.

Special Terms:
v1.1

Questions? customerservice@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.
Identification and Quantification of Protein Glycosylation

Ziv Roth, Galt Yehezkel, and Isam Khalaila

Received 25 December 2011; Accepted 23 January 2012

Abstract

This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to Cite this Article


Download citation as EndNote
This is a License Agreement between Siyuan Liu ("You") and Springer ("Springer") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Springer, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

<table>
<thead>
<tr>
<th>License Number</th>
<th>3822761111473</th>
</tr>
</thead>
<tbody>
<tr>
<td>License date</td>
<td>Mar 05, 2016</td>
</tr>
<tr>
<td>Licensed content publisher</td>
<td>Springer</td>
</tr>
<tr>
<td>Licensed content publication</td>
<td>Cellular and Molecular Life Sciences</td>
</tr>
<tr>
<td>Licensed content title</td>
<td>Deamidation and isoaspartate formation in proteins: unwanted alterations or surreptitious signals?</td>
</tr>
<tr>
<td>Licensed content author</td>
<td>K. J. Reissner</td>
</tr>
<tr>
<td>Licensed content date</td>
<td>Jan 1, 2003</td>
</tr>
<tr>
<td>Volume number</td>
<td>60</td>
</tr>
<tr>
<td>Issue number</td>
<td>7</td>
</tr>
<tr>
<td>Type of Use</td>
<td>Thesis/Dissertation</td>
</tr>
<tr>
<td>Portion</td>
<td>Figures/tables/illustrations</td>
</tr>
<tr>
<td>Number of figures/tables/illustrations</td>
<td>1</td>
</tr>
<tr>
<td>Author of this Springer article</td>
<td>No</td>
</tr>
<tr>
<td>Order reference number</td>
<td>None</td>
</tr>
<tr>
<td>Original figure numbers</td>
<td>Figure 1</td>
</tr>
<tr>
<td>Title of your thesis / dissertation</td>
<td>STATE-OF-ART SEPARATION – MASS SPECTROMETRY IN PROTEIN CHARACTERIZATION</td>
</tr>
<tr>
<td>Expected completion date</td>
<td>Apr 2016</td>
</tr>
<tr>
<td>Estimated size(pages)</td>
<td>219</td>
</tr>
<tr>
<td>Total</td>
<td>0.00 USD</td>
</tr>
</tbody>
</table>

Terms and Conditions

Introduction
The publisher for this copyrighted material is Springer. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

Limited License
With reference to your request to reuse material on which Springer controls the copyright,
SPRINGER LICENSE TERMS AND CONDITIONS

Mar 05, 2016

This is a License Agreement between Siyuan Liu ("You") and Springer ("Springer") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Springer, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

<table>
<thead>
<tr>
<th>License Number</th>
<th>3822770018012</th>
</tr>
</thead>
<tbody>
<tr>
<td>License date</td>
<td>Mar 05, 2016</td>
</tr>
<tr>
<td>Licensed content publisher</td>
<td>Springer</td>
</tr>
<tr>
<td>Licensed content publication</td>
<td>Journal of The American Society for Mass Spectrometry</td>
</tr>
<tr>
<td>Licensed content title</td>
<td>Quantitating the relative abundance of isoaaspartyl residues in deamidated proteins by electron capture dissociation</td>
</tr>
<tr>
<td>Licensed content author</td>
<td>Jason J. Cournoyer</td>
</tr>
<tr>
<td>Licensed content date</td>
<td>Jan 1, 2007</td>
</tr>
<tr>
<td>Volume number</td>
<td>18</td>
</tr>
<tr>
<td>Issue number</td>
<td>1</td>
</tr>
<tr>
<td>Type of Use</td>
<td>Thesis/Dissertation</td>
</tr>
<tr>
<td>Portion</td>
<td>Figures/tables/illustrations</td>
</tr>
<tr>
<td>Number of figures/tables/illustrations</td>
<td>1</td>
</tr>
<tr>
<td>Author of this Springer article</td>
<td>No</td>
</tr>
<tr>
<td>Order reference number</td>
<td>None</td>
</tr>
<tr>
<td>Original figure numbers</td>
<td>Figure 1</td>
</tr>
<tr>
<td>Title of your thesis / dissertation</td>
<td>STATE-OF-ART SEPARATION – MASS SPECTROMETRY IN PROTEIN CHARACTERIZATION</td>
</tr>
<tr>
<td>Expected completion date</td>
<td>Apr 2016</td>
</tr>
<tr>
<td>Estimated size(pages)</td>
<td>219</td>
</tr>
<tr>
<td>Total</td>
<td>0.00 USD</td>
</tr>
</tbody>
</table>

Terms and Conditions

Introduction
The publisher for this copyrighted material is Springer. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

Limited License
With reference to your request to reuse material on which Springer controls the copyright,
permission is granted for the use indicated in your enquiry under the following conditions:
- Licenses are for one-time use only with a maximum distribution equal to the number stated in your request.
- Springer material represents original material which does not carry references to other sources. If the material in question appears with a credit to another source, this permission is not valid and authorization has to be obtained from the original copyright holder.
- This permission
  • is non-exclusive
  • is only valid if no personal rights, trademarks, or competitive products are infringed.
  • explicitly excludes the right for derivatives.
- Springer does not supply original artwork or content.
- According to the format which you have selected, the following conditions apply accordingly:
  • **Print and Electronic:** This License include use in electronic form provided it is password protected, on intranet, or CD-Rom/DVD or E-book/E-journal. It may not be republished in electronic open access.
  • **Print:** This License excludes use in electronic form.
  • **Electronic:** This License only pertains to use in electronic form provided it is password protected, on intranet, or CD-Rom/DVD or E-book/E-journal. It may not be republished in electronic open access.
For any electronic use not mentioned, please contact Springer at permissions.springer@spi-global.com.
- Although Springer controls the copyright to the material and is entitled to negotiate on rights, this license is only valid subject to courtesy information to the author (address is given in the article/chapter).
- If you are an STM Signatory or your work will be published by an STM Signatory and you are requesting to reuse figures/tables/illustrations or single text extracts, permission is granted according to STM Permissions Guidelines: [http://www.stm-assoc.org/permissions-guidelines/](http://www.stm-assoc.org/permissions-guidelines/)
For any electronic use not mentioned in the Guidelines, please contact Springer at permissions.springer@spi-global.com. If you request to reuse more content than stipulated in the STM Permissions Guidelines, you will be charged a permission fee for the excess content.

Permission is valid upon payment of the fee as indicated in the licensing process. If permission is granted free of charge on this occasion, that does not prejudice any rights we might have to charge for reproduction of our copyrighted material in the future.

- If your request is for reuse in a Thesis, permission is granted free of charge under the following conditions:
  This license is valid for one-time use only for the purpose of defending your thesis and with a maximum of 100 extra copies in paper. If the thesis is going to be published, permission needs to be reobtained.
  - includes use in an electronic form, provided it is an author-created version of the thesis on his/her own website and his/her university’s repository, including UMI (according to the definition on the Sherpa website: http://www.sherpa.ac.uk/romeo/);
  - is subject to courtesy information to the co-author or corresponding author.

**Geographic Rights: Scope**
Licenses may be exercised anywhere in the world.

**Altering/Modifying Material: Not Permitted**
Figures, tables, and illustrations may be altered minimally to serve your work. You may not alter or modify text in any manner. Abbreviations, additions, deletions and/or any other
alterations shall be made only with prior written authorization of the author(s).

Reservation of Rights

Springer reserves all rights not specifically granted in the combination of (i) the license
details provided by you and accepted in the course of this licensing transaction and (ii) these
terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

License Contingent on Payment

While you may exercise the rights licensed immediately upon issuance of the license at the
end of the licensing process for the transaction, provided that you have disclosed complete
and accurate details of your proposed use, no license is finally effective unless and until full
payment is received from you (either by Springer or by CCC) as provided in CCC's Billing
and Payment terms and conditions. If full payment is not received by the date due, then any
license preliminarily granted shall be deemed automatically revoked and shall be void as if
never granted. Further, in the event that you breach any of these terms and conditions or any
of CCC's Billing and Payment terms and conditions, the license is automatically revoked and
shall be void as if never granted. Use of materials as described in a revoked license, as well
as any use of the materials beyond the scope of an unrevoked license, may constitute
copyright infringement and Springer reserves the right to take any and all action to protect
its copyright in the materials.

Copyright Notice: Disclaimer

You must include the following copyright and permission notice in connection with any
reproduction of the licensed material:

"Springer book/journal title, chapter/article title, volume, year of publication, page, name(s)
of author(s), (original copyright notice as given in the publication in which the material was
originally published) "With permission of Springer"

In case of use of a graph or illustration, the caption of the graph or illustration must be
included, as it is indicated in the original publication.

Warrants: None

Springer makes no representations or warranties with respect to the licensed material and
adopts on its own behalf the limitations and disclaimers established by CCC on its behalf in
its Billing and Payment terms and conditions for this licensing transaction.

Indemnity

You hereby indemnify and agree to hold harmless Springer and CCC, and their respective
officers, directors, employees and agents, from and against any and all claims arising out of
your use of the licensed material other than as specifically authorized pursuant to this
license.

No Transfer of License

This license is personal to you and may not be sublicensed, assigned, or transferred by you
without Springer's written permission.

No Amendment Except in Writing

This license may not be amended except in a writing signed by both parties (or, in the case
of Springer, by CCC on Springer's behalf).

Objection to Contrary Terms

Springer hereby objects to any terms contained in any purchase order, acknowledgment,
check endorsement or other writing prepared by you, which terms are inconsistent with these
terms and conditions or CCC's Billing and Payment terms and conditions. These terms and
conditions, together with CCC's Billing and Payment terms and conditions (which are
incorporated herein), comprise the entire agreement between you and Springer (and CCC)
concerning this licensing transaction. In the event of any conflict between your obligations
established by these terms and conditions and those established by CCC's Billing and
Payment terms and conditions, these terms and conditions shall control.
Jurisdiction
All disputes that may arise in connection with this present License, or the breach thereof, shall be settled exclusively by arbitration, to be held in the Federal Republic of Germany, in accordance with German law.

Other conditions:
V 12AUG2015

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.
This Agreement between Siyuan Liu ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number: 3822770534552
License date: Mar 05, 2016
Licensed Content Publisher: John Wiley and Sons
Licensed Content Publication: Mass Spectrometry Reviews
Licensed Content Title: CE-ESI-MS for bottom-up proteomics: Advances in separation, interfacing and applications
Licensed Content Author: Anthonius A.M. Heemskerk, André M. Deelder, Oleg A. Mayboroda
Licensed Content Date: May 23, 2014
Pages: 13
Type of use: Dissertation/Thesis
Requestor type: University/Academic
Format: Print and electronic
Portion: Figure/table
Number of figures/tables: 3
Original Wiley figure/table number(s): Figure 1, Figure 2 and Figure 3
Will you be translating? No
Title of your thesis / dissertation: STATE-OF-ART SEPARATION – MASS SPECTROMETRY IN PROTEIN CHARACTERIZATION
Expected completion date: Apr 2016
Expected size (number of pages): 219
Requestor Location: Siyuan Liu
360 Huntington avenue
BOSTON, MA 02115
United States
Attn: Siyuan Liu
Billing Type: Invoice
Billing Address: Siyuan Liu
360 Huntington avenue
BOSTON, MA 02115
United States
Attn: Siyuan Liu
**JOHN WILEY AND SONS LICENSE TERMS AND CONDITIONS**

Apr 13, 2016

This Agreement between Siyuan Liu ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

<table>
<thead>
<tr>
<th>License Number</th>
<th>3845860717766</th>
</tr>
</thead>
<tbody>
<tr>
<td>License date</td>
<td>Apr 13, 2016</td>
</tr>
<tr>
<td>Licensed Content Publisher</td>
<td>John Wiley and Sons</td>
</tr>
<tr>
<td>Licensed Content Publication</td>
<td>Mass Spectrometry Reviews</td>
</tr>
<tr>
<td>Licensed Content Title</td>
<td>Fragmentation pathways of protonated peptides</td>
</tr>
<tr>
<td>Licensed Content Author</td>
<td>Béla Paizs, Sándor Suhai</td>
</tr>
<tr>
<td>Licensed Content Date</td>
<td>Jul 12, 2004</td>
</tr>
<tr>
<td>Pages</td>
<td>41</td>
</tr>
<tr>
<td>Type of use</td>
<td>Dissertation/Thesis</td>
</tr>
<tr>
<td>Requestor type</td>
<td>University/Academic</td>
</tr>
<tr>
<td>Format</td>
<td>Print and electronic</td>
</tr>
<tr>
<td>Portion</td>
<td>Figure/table</td>
</tr>
<tr>
<td>Number of figures/tables</td>
<td>2</td>
</tr>
<tr>
<td>Original Wiley figure/table number(s)</td>
<td>Scheme 1 and Scheme 3</td>
</tr>
<tr>
<td>Will you be translating?</td>
<td>No</td>
</tr>
<tr>
<td>Title of your thesis / dissertation</td>
<td>STATE-OF-ART SEPARATION - MASS SPECTROMETRY IN PROTEIN CHARACTERIZATION</td>
</tr>
<tr>
<td>Expected completion date</td>
<td>Apr 2016</td>
</tr>
<tr>
<td>Expected size (number of pages)</td>
<td>219</td>
</tr>
<tr>
<td>Requestor Location</td>
<td>Siyuan Liu 360 Huntington avenue</td>
</tr>
</tbody>
</table>

BOSTON, MA 02115 United States Attn: Siyuan Liu

Billing Type | Invoice |
Billing Address | Siyuan Liu 360 Huntington avenue |

BOSTON, MA 02115 United States Attn: Siyuan Liu
TERMS AND CONDITIONS

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a "Wiley Company") or handled on behalf of a society with which a Wiley Company has exclusive publishing rights in relation to a particular work (collectively "WILEY"). By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the billing and payment terms and conditions established by the Copyright Clearance Center Inc., ("CCC's Billing and Payment terms and conditions"), at the time that you opened your RightsLink account (these are available at any time at http://myaccount.copyright.com).

Terms and Conditions

• The materials you have requested permission to reproduce or reuse (the "Wiley Materials") are protected by copyright.

• You are hereby granted a personal, non-exclusive, non-sub licensable (on a stand-alone basis), non-transferrable, worldwide, limited license to reproduce the Wiley Materials for the purpose specified in the licensing process. This license, and any CONTENT (PDF or image file) purchased as part of your order, is for a one-time use only and limited to any maximum distribution number specified in the license. The first instance of republication or reuse granted by this license must be completed within two years of the date of the grant of this license (although copies prepared before the end date may be distributed thereafter). The Wiley Materials shall not be used in any other manner or for any other purpose, beyond what is granted in the license. Permission is granted subject to an appropriate acknowledgement given to the author, title of the material/book/journal and the publisher. You shall also duplicate the copyright notice that appears in the Wiley publication in your use of the Wiley Material. Permission is also granted on the understanding that nowhere in the text is a previously published source acknowledged for all or part of this Wiley Material. Any third party content is expressly excluded from this permission.

• With respect to the Wiley Materials, all rights are reserved. Except as expressly granted by the terms of the license, no part of the Wiley Materials may be copied, modified, adapted (except for minor reformatting required by the new Publication), translated, reproduced, transferred or distributed, in any form or by any means, and no derivative works may be made based on the Wiley Materials without the prior permission of the respective copyright owner. For STM Signatory Publishers clearing permission under the terms of the STM Permissions Guidelines only, the terms of the license are extended to include subsequent editions and for editions in other languages, provided such editions are for the work as a whole in situ and does not involve the separate exploitation of the permitted figures or extracts, you may not alter, remove or suppress in any manner any copyright, trademark or other notices displayed by the Wiley Materials. You may not license, rent, sell, loan, lease, pledge, offer as security, transfer or assign the Wiley Materials on a stand-alone basis, or any of the rights granted to you hereunder to any other person.
The Wiley Materials and all of the intellectual property rights therein shall at all times remain the exclusive property of John Wiley & Sons Inc, the Wiley Companies, or their respective licensors, and your interest therein is only that of having possession of and the right to reproduce the Wiley Materials pursuant to Section 2 herein during the continuance of this Agreement. You agree that you own no right, title or interest in or to the Wiley Materials or any of the intellectual property rights therein. You shall have no rights hereunder other than the license as provided for above in Section 2. No right, license or interest to any trademark, trade name, service mark or other branding ("Marks") of WILEY or its licensors is granted hereunder, and you agree that you shall not assert any such right, license or interest with respect thereto.

NEITHER WILEY NOR ITS LICENSORS MAKES ANY WARRANTY OR REPRESENTATION OF ANY KIND TO YOU OR ANY THIRD PARTY, EXPRESS, IMPLIED OR STATUTORY, WITH RESPECT TO THE MATERIALS OR THE ACCURACY OF ANY INFORMATION CONTAINED IN THE MATERIALS, INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTY OF MERCHANTABILITY, ACCURACY, SATISFACTORY QUALITY, FITNESS FOR A PARTICULAR PURPOSE, USABILITY, INTEGRATION OR NON-INFRINGEMENT AND ALL SUCH WARRANTIES ARE HEREBY EXCLUDED BY WILEY AND ITS LICENSORS AND WAIVED BY YOU.

WILEY shall have the right to terminate this Agreement immediately upon breach of this Agreement by you.

You shall indemnify, defend and hold harmless WILEY, its Licensors and their respective directors, officers, agents and employees, from and against any actual or threatened claims, demands, causes of action or proceedings arising from any breach of this Agreement by you.

IN NO EVENT SHALL WILEY OR ITS LICENSORS BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR ENTITY FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, EXEMPLARY OR PUNITIVE DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, PROVISIONING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

Should any provision of this Agreement be held by a court of competent jurisdiction to be illegal, invalid, or unenforceable, that provision shall be deemed amended to achieve as nearly as possible the same economic effect as the original provision, and the legality, validity and enforceability of the remaining provisions of this Agreement shall not be affected or impaired thereby.
• The failure of either party to enforce any term or condition of this Agreement shall not constitute a waiver of either party’s right to enforce each and every term and condition of this Agreement. No breach under this agreement shall be deemed waived or excused by either party unless such waiver or consent is in writing signed by the party granting such waiver or consent. The waiver by or consent of a party to a breach of any provision of this Agreement shall not operate or be construed as a waiver of or consent to any other or subsequent breach by such other party.

• This Agreement may not be assigned (including by operation of law or otherwise) by you without WILEY’s prior written consent.

• Any fee required for this permission shall be non-refundable after thirty (30) days from receipt by the CCC.

• These terms and conditions together with CCC’s Billing and Payment terms and conditions (which are incorporated herein) form the entire agreement between you and WILEY concerning this licensing transaction and (in the absence of fraud) supersedes all prior agreements and representations of the parties, oral or written. This Agreement may not be amended except in writing signed by both parties. This Agreement shall be binding upon and inure to the benefit of the parties’ successors, legal representatives, and authorized assigns.

• In the event of any conflict between your obligations established by these terms and conditions and those established by CCC’s Billing and Payment terms and conditions, these terms and conditions shall prevail.

• WILEY expressly reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC’s Billing and Payment terms and conditions.

• This Agreement will be void if the Type of Use, Format, Circulation, or Requestor Type was misrepresented during the licensing process.

• This Agreement shall be governed by and construed in accordance with the laws of the State of New York, USA, without regards to such state’s conflict of law rules. Any legal action, suit or proceeding arising out of or relating to these Terms and Conditions or the breach thereof shall be instituted in a court of competent jurisdiction in New York County in the State of New York in the United States of America and each party hereby consents and submits to the personal jurisdiction of such court, waives any objection to venue in such court and consents to service of process by registered or certified mail, return receipt requested, at the last known address of such party.

WILEY OPEN ACCESS TERMS AND CONDITIONS
Wiley Publishes Open Access Articles in fully Open Access Journals and in Subscription journals offering Online Open. Although most of the fully Open Access journals publish open access articles under the terms of the Creative Commons Attribution (CC BY) License only, the subscription journals and a few of the Open Access Journals offer a choice of
Creative Commons Licenses. The license type is clearly identified on the article.

The Creative Commons Attribution License
The Creative Commons Attribution License (CC-BY) allows users to copy, distribute and transmit an article, adapt the article and make commercial use of the article. The CC-BY license permits commercial and non-

Creative Commons Attribution Non-Commercial License
The Creative Commons Attribution Non-Commercial (CC-BY-NC) License permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes. (see below)

Creative Commons Attribution-Non-Commercial-NoDerivs License
The Creative Commons Attribution Non-Commercial-NoDerivs License (CC-BY-NC-ND) permits use, distribution and reproduction in any medium, provided the original work is properly cited, is not used for commercial purposes and no modifications or adaptations are made. (see below)

Use by commercial "for-profit" organizations
Use of Wiley Open Access articles for commercial, promotional, or marketing purposes requires further explicit permission from Wiley and will be subject to a fee.
Further details can be found on Wiley Online Library
http://olabout.wiley.com/WileyCDA/Section/id-410895.html

Other Terms and Conditions:

v1.10 Last updated September 2015

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.
PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

If credit is given to another source for the material you requested, permission must be obtained from that source.