UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

DEVELOPMENTS AND APPLICATIONS OF HIGH-THROUGHPUT TOP-DOWN PROTEOMICS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

ZHE WANG
Norman, Oklahoma
2019

DEVELOPMENTS AND APPLICATIONS OF HIGH-THROUGHPUT TOP-DOWN PROTEOMICS

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF CHEMISTRY AND BIOCHEMISTRY

BY THE COMMITTEE CONSISTING OF

Dr. Si Wu, Chair

Dr. Mark Nanny

Dr. Robert White

Dr. Zhibo Yang

Dr. Rakhi Rajan

Acknowledgements

I would like to express my deepest appreciation to my mentor, Dr. Si Wu. She is not only my mentor on academia research but also a teacher of how to be a successful and decent person. If it were not for her advice, constant encouragement, support and nurturing, this doctoral dissertation would not have been possible.

In addition to my mentor, I am grateful for our collaborator Dr. Kenneth Smith, at Oklahoma Medical Research Foundation, for all his support on our collaboration projects.

I would like to extend my sincere thanks to my committee members, Dr. Mark Nanny, Dr. Robert White, Dr. Rakhi Rajan, and Dr. Zhibo Yang. Thank you so much for the help and guidance throughout the years.

I would like to thank and gratefully acknowledge the assistance that I received from all the lab members: Dahang Yu, Lushuang Huang, Graham Delafield, Mulin Fang, Yanting Guo, Toni Woodard, Jacob Klenke, and Kellye Cupp-Sutton. Their suggestions and help are much appreciated.

Finally, I would like to thank my family and friends for their love, support, and companion during the past years.

Table of Contents

Acknowledgements iv
Table of Contents v
List of Tables viii
List of Figures ix
Abstract xi
Chapter 1 Introduction 1
1.1 Background 1
1.2 High-throughput proteomics 3
1.2.1 Separation techniques for top-down proteomics 7
1.2.2 The development of mass spectrometry for top-down proteomics 10
1.3 Serum autoantibody proteomics 13
1.4 Dissertation synopsis 15
Chapter 2 Two-dimensional separation using high-pH and low-pH reversed phase liquid chromatography for top-down proteomics 17
2.1 Abstract 17
2.2 Introduction 18
2.3 Material and methods 21
2.3.1 Materials and reagents 21
2.3.2 Sample preparation 21
2.3.3 1st dimension RPLC separation and fractionation 22
2.3.4 2nd dimension top-down RPLC-MS/MS analysis 23
2.3.5 Bottom-up LC-MS/MS analysis of fractions from 1st dimension RPLC 24
2.3.6 Protein and proteoform identification 25
2.4 Results and discussion 25
2.4.1 Evaluation of orthogonality between high-pH RPLC and low-pH RPLC 25
2.4.2 2D high-pH and low-pH RPLC-MS/MS analysis of *E. coli* intact proteins 30
2.4.3 Identification of novel intact proteoforms of apo-acyl carrier protein 35
2.5 Conclusions 38
Chapter 3 Development of an online 2D ultra-high-pressure nano-LC system for high-pH and low-pH reversed phase separation in top-down proteomics 40
3.1 Abstract 40
3.2 Introduction 41

3.3 Materials and methods 42
3.3.1 Chemicals and materials 42
3.3.2 Sample preparation 43
3.3.3 Instrumental setup of 1D-LC system 43
3.3.4 Instrumental setup of online 2D-LC system 44
3.3.5 Top-down MS analysis 46
3.3.6 Data analysis 46
3.4 Results and discussion 47
3.5 Conclusion 57
Chapter 4 Top-down mass spectrometry analysis of human serum autoantibody antigen-binding fragments 58
4.1 Abstract 58
4.2 Introduction 59
4.3 Materials and methods 63
4.3.1 Materials and Reagents 63
4.3.2 Human subjects 63
4.3.3 Monoclonal antibodies 63
4.3.4 Sample preparation 64
4.3.5 UPLC-TD-HRMS 64
4.3.6 Data analysis 65
4.4 Results 67
4.4.1 Optimization of the UPLC-TD-HRMS platform 67
4.4.2 UPLC-TD-HRMS analysis of a 12-antibody mixture 69
4.4.3 UPLC-TD-HRMS analysis of monoclonal antibodies in human serum 73
4.5 Discussion 78
Chapter 5 Optimization of a subzero temperature LC separation in hydrogen/deuterium exchange mass spectrometry for conformational epitope mapping 80
5.1 Abstract 80
5.2 Introduction 81
5.3 Material and methods 83
5.3.1 Materials and reagents 83
5.3.2 Sample preparation 84
5.3.3 Low-temperature liquid chromatography 84

5.3.4 Hydrogen/Deuterium Exchange 86
5.3.5 Bottom-up MS analysis 86
5.3.6 Data analysis 87
5.4 Results and discussion 87
5.4.1 Evaluation of low-temperature RPLC separation 87
5.4.2 Back-exchange rate evaluation 90
5.4.3 Epitope mapping of Anthrax protective antigen 93
5.5 Conclusions 97
Chapter 6 Overall summary and future directions 99
6.1 Overall summary 99
6.2 Future directions 101
References 103

List of Tables

Table 4-1. Parameter settings of TopPIC in the analysis of the top-down MS/MS data of human serum samples 66

List of Figures

Figure 1-1. Central dogma of molecular biology..... 2
Figure 1-2. Scheme of bottom-up and top-down proteomics. 4
Figure 1-3. Bottom-up proteomics results in loss of proteoform information..... 5
Figure 1-4. Typical workflow of a high-throughput top-down proteomics study. 6
Figure 1-5. Orthogonality demonstration. 9
Figure 1-6. Peptide backbone fragmentation. 12
Figure 1-7. The VDJ recombination diversifying the antibody repertoire. 14
Figure 2-1. The separation of three standard proteins. 26
Figure 2-2. Comparison of low-pH RPLC and high-pH RPLC as the first-dimension separation. 27
Figure 2-3. Identification results of the 2DLC..... 31
Figure 2-4. 2D pH RP/RPLC/MS analysis on *E.coli* proteins..... 32
Figure 2-5. Five-minute segment (RT = 90 min to 95 min) from LC/MS runs of 1D and 2D methods. 34
Figure 2-6. Different proteoforms of apo-acyl carrier protein identified using 1D and 2D methods. 36
Figure 3-1. The schematic diagram of online comprehensive 2D-LC platform..... 48
Figure 3-2. Stepwise operation of online 2D nano-LC system and the gradient setup. ... 48
Figure 3-3. Validation of the dilution factor for 2DLC analysis. 49
Figure 3-4. The online 2D chromatograms of 2-fraction runs..... 50
Figure 3-5. Evaluation of ^{1}D gradient range for *E. coli* protein separation. 51
Figure 3-6. LC-MS base peak chromatograms (*m/z* 600-1100) of the online 2D-LC analysis with 10 online collected fractions. 52
Figure 3-7. Online 2D pH RP/RPLC MS analysis of *E.coli* proteins..... 53
Figure 3-8. Five-minute segment from 1DLC and online 2DLC (RT = 190 min – 195 min). 54
Figure 3-9. Proteoforms of the apo-acyl carrier protein identified using offline 2D and online 2D pH RP/RPLC-MS..... 55
Figure 3-10. The proteoform elution patterns in high pH fractions..... 56
Figure 4-1. Peak capacities of the LC-MS run of E.coli lysate proteins with different gradient length. 67
Figure 4-2. LC-MS of 12 standard antibody mixture. 69

Figure 4-3. The identifications of the Fab light chain and Fab heavy chain of an antibody. 70
Figure 4-4. Analysis of reduced and intact Fab fragments. 72
Figure 4-5. LC-MS analysis of SLE serum sample and healthy control samples. 73
Figure 4-6. Identification of light chains and heavy chains of the IgGs in SLE Patient Serum. 75
Figure 4-7. Run-to-run reproducibility of the developed platform. 77
Figure 5-1. UV traces of RPLC chromatograms of *E. coli* digest at different temperatures. 89
Figure 5-2. The mass error matching algorithm used for deuterated peptide identification. 91
Figure 5-3. Deuteration level vs. retention time of *E. coli* peptides under -9 ℃ and -20 ℃. 92
Figure 5-4. Mass spectra of deuterated peptides under different temperature conditions. 93
Figure 5-5. Heatmap of deuteration level of PA peptides obtained from differential HDX-MS of the free PA sample and the PA – anti-PA C01 antibody complex sample. 94
Figure 5-6. Mass spectra of PA peptides that were (A) involved and (B) not involved in the binding events. 95
Figure 5-7. Heatmap of the deuteration level of PA peptides obtained from differential HDX-MS of the free PA sample and the PA, anti-PA antibody C01, and F01 mixture... 96

Abstract

Emerging technologies for liquid chromatography (LC) separations and mass spectrometry (MS) have dramatically improved the throughput of top-down proteomics, which has allowed the identification of thousands of intact proteoforms in complex biological samples. Many proteins have multiple intact proteoforms (*i.e.* variants with different post-translational modifications, PTMs) and these proteoforms often exhibit differing functions in biological processes. Therefore, it is important to establish an accurate and high-throughput top-down approach for the analysis of intact proteoforms. The focus of this dissertation is the development and application of novel high-throughput quantitative top-down proteomics techniques to characterize intact proteins in complex biological samples (*e.g.,* serum autoantibody repertoire).

We have developed a two-dimensional separation platform for high-throughput top-down MS analysis, which employs high-pH and low-pH RPLC-MS for top-down proteomics. The high-pH and low-pH 2D RPLC has been widely applied in bottom-up proteomics; however, it has not been applied to intact protein separation. Our results demonstrated that the proposed high-pH and low-pH RPLC separation format has good orthogonality for intact protein separation. We have used our 2D Ph RP/RPLC platform for PTM characterization of complex biological samples such as *E. coli* and human cell lysates. This 'salt-free' RPLC process overcomes the disadvantage of other chromatographic methods, which require complicated sample processing (*i.e.,* buffer exchange, desalting) before MS analysis, and the two dimensions of separation could be easily coupled for online analysis.

One of the drawbacks of the platform is that a large amount of starting material is

required and the sample handling process between each dimension can introduce variations in the sample and can cause loss of sensitivity due to sample loss during the transfer, recovery, and desalting processes. Moreover, the off-line multidimensional separation techniques are often time consuming, labor intensive, and low-throughput. To improve the throughput and sensitivity of the offline approach, we developed an online 2D ultra-high-pressure nano-LC system for high-pH and low-pH RPLC separation in top-down proteomics. The online 2DLC platform enabled the characterization of 2000+ intact proteoforms from 5 micro-grams of intact *E. coli* cell lysate, presenting a potential for deep proteome characterization in mass-limited samples using top-down proteomics.

The human serum autoantibody repertoire is extremely complicated and autoantibody development is crucial to the function of the human immune system. We have applied top-down MS methods to characterize the serum autoantibody repertoire. We implemented a 1D UPLC-TDMS platform and achieved separation of a mixture of 12 antibody Fabs with highly homologous sequences. Furthermore, we characterized 86 Fab related mass features, which represents the first top-down analysis of the complexity of the human autoantibody repertoire. Using top-down proteomics techniques, the PTMs of monoclonal antibodies can also be characterized; this information could give valuable insight into autoantibody function and human immunity.

Antibody-antigen interaction is the fundamental reaction of immune system. The clear characterization of paratope/epitope of an immune complex can help the understanding of immune response and possibly the mechanisms of related diseases. To characterize the immune complexes, we coupled a low-temperature LC system with hydrogen/deuterium exchange mass spectrometry (HDX-MS), and applied it to

characterize the epitopes of the protective antigen.

Overall, our results demonstrate the utility of top-down proteomics for the analysis of the serum autoantibody repertoire and application of HDX-MS on immune complex characterization. Top-down proteomics techniques hold great promise for the discovery of novel serum autoantibody biomarkers, as well as the promotion of our understanding of pathogenic autoimmune processes.

Chapter 1 Introduction

1.1 Background

The proteome, the complex system of proteins expressed by a cell, is essential for the activity of complex biochemical processes. The study of the proteome provides complementary information to genomics and transcriptomics, which promotes the understanding of cellular functions and signaling pathways. During the transcription and translation processes that are used to synthesize proteins, any gene and protein processing event such as allelic variation (*e.g.* coding polymorphisms), alternative splicing, and post-translational modification may vary protein structure (**Figure 1-1**). Different protein products can be produced from a single gene due to these pre- and post-transcriptional/translational processes which results in the large diversity of the proteome compared to the diversity of the genome . These protein variants were designated with the term "proteoform" in 2013 [1]. Different proteoforms of the same protein may have different abundances and functions in the cellular environment and may be active in the regulation of gene expression. Thus, understanding the abundance and function of a proteoform in a biological system is extremely important to study and understand the mechanisms of different biological processes.

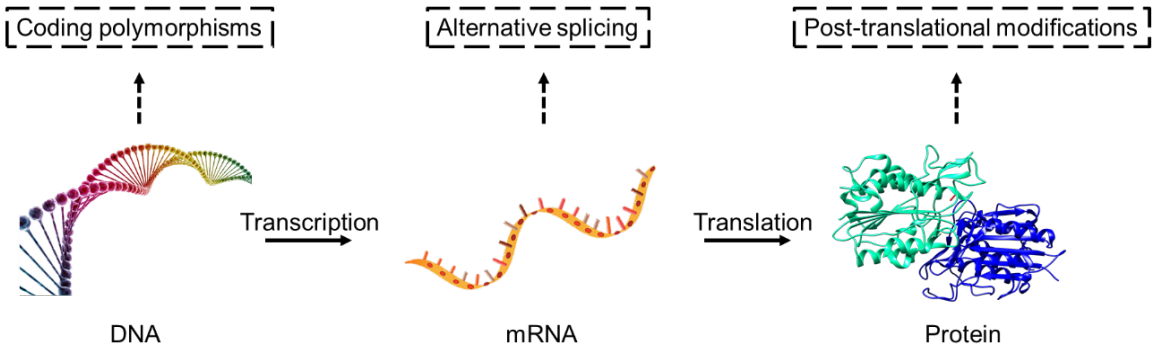


Figure 1-1. Central dogma of molecular biology.

Proteomics is the large-scale study of proteins[2, 3]. Unlike the relatively constant genome in a biological system, the proteome is dynamic. This means that, even from a distinct set of genes, proteoform identity and abundance can vary in different cells; PTMs and of abundance can also change over the lifetime of the cell[4]. To study proteins, ligand binding assays such as antibody-based immunoassays (*e.g.* enzyme-linked immunosorbent assay, ELISA) and mass spectrometry-based methods have been the most commonly applied methods in the past decades. Antibody-based immunoassays require known antibodies against proteins of interest for affinity purification. These antibodies can be time consuming and labor intensive to produce and antibody-based immunoassays cannot reveal protein structural information. Traditionally, mass spectrometry-based proteomics have utilized two-dimensional polyacrylamide gel electrophoresis (2D PAGE) for front end protein separation prior to mass spectrometry analysis. After the protein mixture is separated on a PAGE gel, the protein bands of interest are isolated, subjected to sample preparation (*e.g.* in-gel enzymatic digestion), and detected using a mass spectrometer. However, 2D PAGE is time consuming and low throughput.

The use of high-resolution front-end separation techniques such as reversed phase

liquid chromatography (RPLC) and capillary electrophoresis (CE) has emerged as an alternative to 2D PAGE for protein separation. Using these online separation techniques, thousands of proteins can be detected from a single sample using mass spectrometry. Additionally, the advancement of high-end mass spectrometry instruments (*e.g.* FTICR, Orbitrap, Time-Of-Flight mass analyzers, and etc.) has enabled the fast and accurate detection of large biological molecules (*e.g.* peptides and proteins). Through the application of online separation techniques, predominantly LC, coupled with mass spectrometric detection, large scale, routine, and high-throughput analysis of proteins is becoming feasible and reliable.

1.2 High-throughput proteomics

There are about 20,000 protein-coding genes in the human genome [5, 6], and an estimated 1,000,000 proteoforms from one cell type in the human proteome. Most of these different proteoforms are responsible for unique biological processes and control the function or misfunction of the human body[6]. The comparatively large diversity in the human proteome stems from sources such as alternative splicing, PTMs, etc. However, we have very limited knowledge on the function of proteoforms in health and disease. Understanding the identity, abundance, and function of important proteoforms will provide useful information for disease diagnosis and treatment. Considering the large size of the human proteome, high-throughput techniques are needed to map as many human proteoforms as possible.

The aim of high-throughput proteomics is deeper proteome coverage with decreased analysis time[7]. There are three main approaches to high-throughput mass

spectrometry-based proteomics: bottom-up, middle-down, and top-down proteomics (**Figure 1-2**).

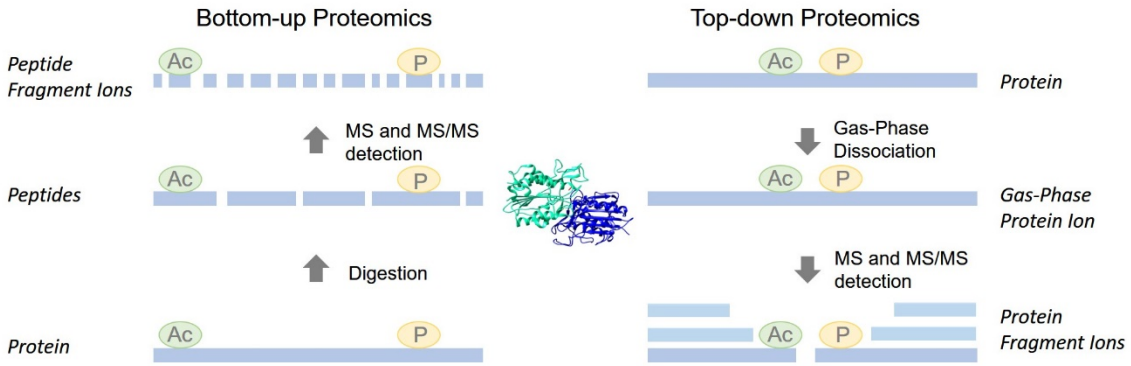


Figure 1-2. Scheme of bottom-up and top-down proteomics.

In bottom-up proteomics, proteins are normally enzymatically digested (*e.g.* tryptic digestion) and analyzed by LC-MS techniques. Full MS scans can be used for peptide quantification and tandem mass spectrometry is utilized for peptide sequencing and identification. In top-down proteomics, intact proteins are directly separated using LC and detected using high-resolution mass spectrometry. In middle-down proteomics, proteins undergo limited proteolysis to produce relatively large peptides (*i.e.* 3 – 9 kDa), the resulting peptides are then separated and detected by LC-MS/MS[8].

Bottom-up methods have been widely applied to proteomics studies and are capable of identifying thousands of peptides in a single analysis. Bottom-up proteomics is robust and sensitive, but it cannot provide accurate information on the combinatorial modification patterns of proteoforms (**Figure 1-3**). Top-down proteomics analyzes intact proteins and allows for the identification of PTM patterns, sequence variations, etc., providing a

complete description of the primary structure of the protein and its modifications. Middle-down proteomics is a popular method because it preserves combinatorial modification patterns on proteoforms and can approach the sensitivity of bottom-up proteomics. Middle-down proteomics is extensively applied for histone proteomics[9, 10].

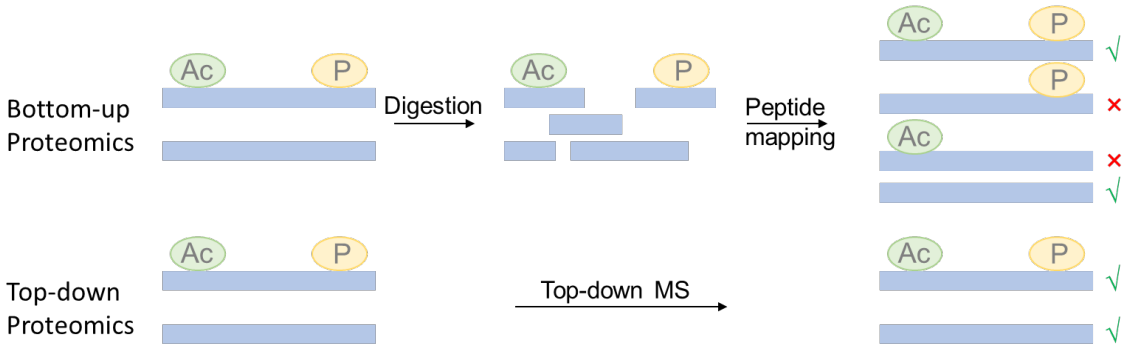


Figure 1-3. Bottom-up proteomics results in loss of proteoform information.

The biological functions of different proteoforms largely rely on the abundance and PTM of the proteoform (*e.g.* histone code). Bottom-up proteomics is more robust for protein identification; however, the proteoform information in terms of PTM location, pattern, and abundance will be lost. High-throughput top-down proteomics has become more accessible thanks to the advancement of online separation/offline fractionation workflows, high-resolution MS instrumentation, and fragmentation techniques for large gas phase molecules.

Top-down proteomics analyzes intact proteoforms which preserves the primary structure of different proteoforms and information of PTMs. Top-down proteomics provides a promising tool for the identification of novel proteoforms, deeper understanding

of protein sequences, and the quantification of functional PTMs[11, 12]. Top-down proteomics was previously limited to the analysis of pure intact proteins and targeted analysis of proteins of interest in its early development[13-15]. Over the past 5 to 10 years, top-down proteomics has been rapidly improved thanks to the fast development of separation techniques and high resolution mass spectrometers[12]. Now, global profiling of proteomes using top-down proteomics is feasible, enabling the identification and quantification of thousands of intact proteoforms in one single LC-MS/MS analysis[16-21].

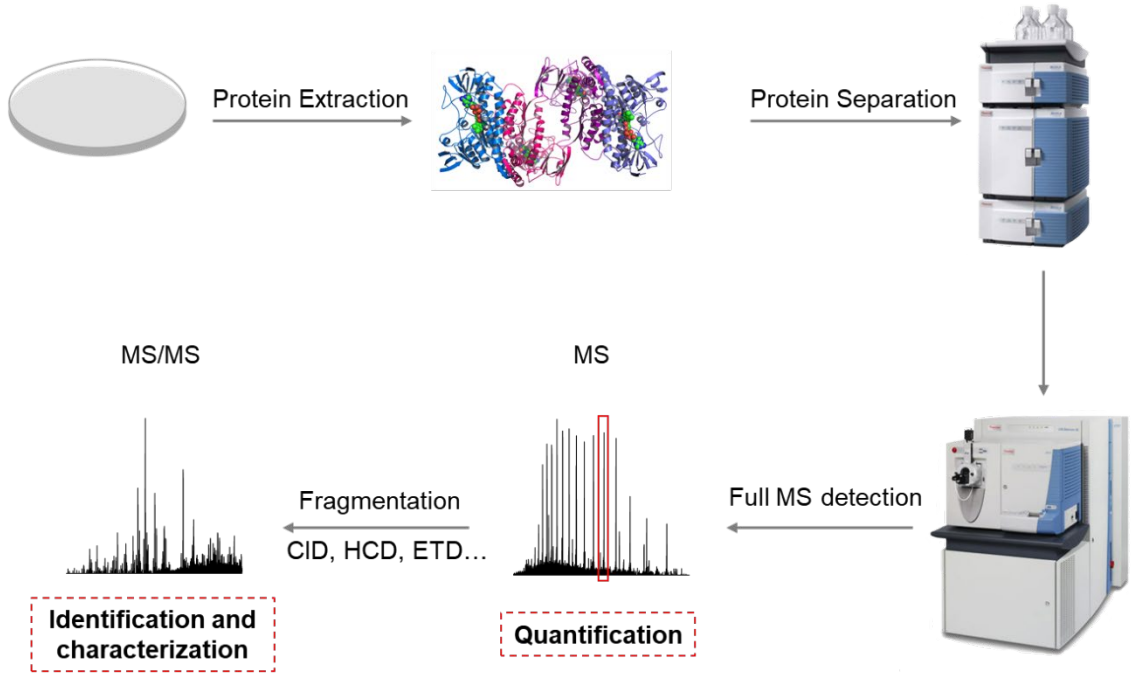


Figure 1-4. Typical workflow of a high-throughput top-down proteomics study.

Typically, a high-throughput top-down proteomics study starts from protein extraction (*i.e.,* sample preparation), followed by protein separation, and MS detection[22].

Top-down proteomics is developing quickly; however, there are still significant challenges posed by the complexity of protein samples, *e.g.* the wide dynamic range of proteoforms in a biological sample and the large size of intact proteins[23]. The high degree of complexity requires optimization or customization for each step of top-down analysis.

Different sample preparation strategies can be utilized based on the characteristics of protein samples to be analyzed. In the case of physiological samples that are not compatible with separation and MS detection, buffer exchange of the proteins can be performed. Traditionally, filter-aided sample preparation (FASP) has been applied where protein samples are centrifuged through a membrane of certain pore size. This method provides an easy process for protein sample buffer exchange and desalting. For samples that contain proteins of interest at low abundance, enrichment of selected proteins can be performed. For example, immunoprecipitation and immobilized metal affinity chromatography (IMAC) can be used to enrich phosphoproteins.

The throughput of the top-down proteomics workflow has also been improved with the advancement of high-resolution separation techniques, high-resolution MS instruments with fast scan rate, and accessible bioinformatics tools.

1.2.1 Separation techniques for top-down proteomics

One of the major goals of global profiling of proteomes using top-down proteomics is to identify and quantify as many proteoforms as possible from biological samples. The number of identified proteoforms directly correlates with the separation power of separation techniques, and MS and MS/MS scan quantity and quality. Separation techniques with higher resolution reduce the ionization competition from co-eluted

analytes because the proteoforms can be concentrated into a sharper elution peaks; this results in higher sensitivity of MS detection. A variety of separation methods that separate proteins based on different mechanisms (*e.g.* size, charge, polarity) have been evaluated for protein separation for top-down proteomics. Separation techniques that have been applied to the separation of intact proteins include reversed phase liquid chromatography (RPLC), hydrophilic interaction chromatography (HILIC), size exclusion chromatography (SEC), hydrophobic interaction chromatography (HIC), and ion exchange chromatography (IEC)[24, 25]. Electrophoresis-based methods such as gel-eluted liquid fraction entrapment electrophoresis (GELFrEE)[26] and capillary electrophoresis (CE) have also been applied to the separation of intact proteins[27]. These various separation techniques provide a pool of tools for the separation of intact proteins for MS analysis.

RPLC is the most prevalent approach for studying complex intact protein samples[7, 11, 28-30]. CE has also been applied as a complementary separation technique to RPLC for the separation of peptides and proteins in proteomics studies. While CE provides higher theoretical plate numbers and more efficient separation, the application of CE has been limited due to the characteristics of the method. For example, as CE is a nano-scale sample separation technique, the limited sample loading amount results in the inability to detect low abundant proteins. In addition, compared to LC based techniques, the reproducibility of CE is relatively poor, making it hard to apply to routine analysis[27, 31, 32]. Given the challenges of CE, top-down proteomics with LC separation techniques are still the most applicable and feasible approaches currently available. Various approaches have been developed to improve RPLC separation power [33-35] including the utilization of ultra-high pressure LC systems with longer columns [36] or smaller column

particle sizes. These advancements serve to improve the peak capacity of the 1D RPLC separation [37].

Since single dimensional separation techniques separate proteins based on a single physicochemical property, they can never resolve the whole proteome. Multi-dimensional separation platforms have emerged to separate proteins based on multiple physicochemical properties. Different separation techniques can be coupled together either online or offline (*i.e.,* pre-fractionation) to extend the separation power of front-end separation for top-down proteomics. Multidimensional separation can result in more identification of proteins and a deeper and more comprehensive understanding of proteoforms. In order to maximize the utilization of the separation window, the different dimensions of multidimensional separation techniques need to be orthogonal, meaning that the same protein will have different elution behaviors in the different dimensions (**Figure 1-5A**). If the separation mechanism of the first dimension is the same or similar with the second-dimension separation, the same protein will be eluted out at a similar retention time in the first- and second-dimension separations, resulting an inefficient usage of whole separation window (**Figure 1-5B**).

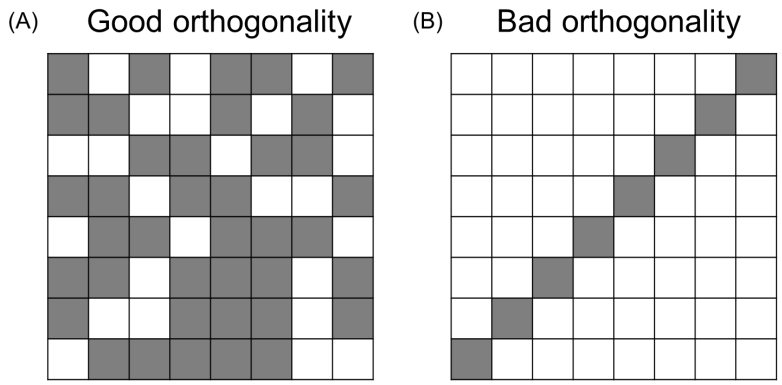


Figure 1-5. Orthogonality demonstration.

In the past decades, various multi-dimension techniques have been developed for top-down proteomics. In 2010, Tian and his coworkers developed weak cation exchange chromatography and hydrophilic interaction chromatography (WCX/HILIC) coupled to RPLC for histone studies[33]. In 2011, the Kelleher group reported a four-dimensional separation platform that coupled isoelectric focusing, multiplex GELFrEE, and nanocapillary LC-MS. This platform demonstrated greater than 20-fold increase in both separation power and proteome coverage[38]. In 2014 and 2015, the Ge group developed a method that coupled HIC to RPLC and three-dimensional separation using IEC, HIC, and RPLC for intact protein analysis [21, 39]. Traditionally, the multidimensional separation platforms involve sample processing between different dimensions (*e.g.,* desalting, concentrating), which can be time consuming and labor intensive. More importantly, offline fractionation requires a large amount of starting materials which limits the application of multi-dimensional separation for top-down proteomics in some situations. Although, the online 2DLC system has been reported for histone analysis [33], no automated online multi-dimensional separation platform has been reported for global profiling of proteomes. Multi-dimensional separation for top-down proteomics has been effective; however, the instrumentation of the multi-dimensional separation system will need to be further improved for more robust and high-throughput analysis of intact protein samples.

1.2.2 The development of mass spectrometry for top-down proteomics

The very early top-down MS characterization of intact protein was performed on quadrupole-based instruments. Compared to modern MS instrument, quadrupole-based

instruments are simpler but the mass resolving power is poor, which leads to inaccurate mass measurement. With the invention and development of high-resolution mass analyzers, the benefits of top-down proteomics started to shine. Fourier-transform ion cyclotron resonance (FTICR) mass analyzers and Orbitrap mass analyzers are currently more commonly used for top-down MS analysis. Other types of instruments, such as linear ion traps, time-of-flight (ToF) mass analyzers, can also be applied to analyze intact proteins. Compared to FTMS, ToF mass analyzer can detect intact proteins with very high *m/z*, making it preferable for large protein complex detection. However, the FTMS instruments provide excellent sensitivity and high mass accuracy, allowing better characterization of proteoforms for top-down proteomics. The improved mass accuracy and high resolving power of modern MS instruments enables the detection of monoisotopic masses of intact proteins. In addition, the high-resolution mass detection can also resolve overlapped fragment ions with similar masses, resulting in more confident assignment of fragment ions[40].

One of the aims of top-down proteomics is the deep characterization of proteoforms with more accurate primary structure characterization, PTM localization, and relative quantification. Ion dissociation techniques have been development to increase the internal energy of analyte ions to the dissociation threshold to generate fragment ions. The fragment ions can be further analyzed by tandem mass spectrometry (*i.e.,* MS/MS) and used to characterize the proteins. Currently, collision-based dissociation methods are still commonly used for protein identification. However, to increase the sequence coverage and better characterize proteoforms (*e.g.* localization of PTMs), different dissociation techniques with complementary fragment features and patterns are sometimes required.

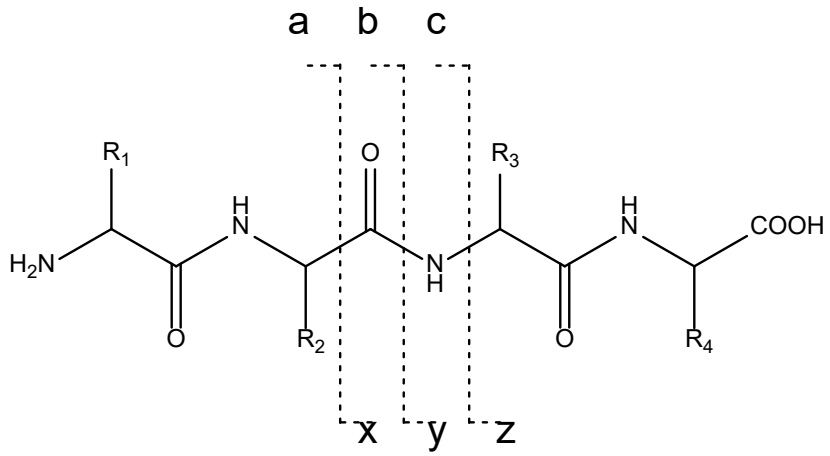


Figure 1-6. Peptide backbone fragmentation.

Different dissociation techniques fragment ions using different mechanisms, such as gas-phase collisions (CID, HCD), surface induced collisions (SID), photodissociation (UVPD), electron capture/transfer dissociation (ECD/ETD), etc. The peptide backbone in a protein can be cleaved at different sites depending on the type of dissociation technique used (**Figure 1-6**). b- and y- ions can be produced by "slow-heating" techniques, such as CID and HCD, c- and z- ions can be produced by ECD, ETD, UVPD, etc., and a- and x- ions are normally generated with electron-mediated techniques for anionic analytes, such as electron detachment dissociation (EDD), negative electron transfer dissociation (nETD), etc [41]. b- and y- ions, c- and z- ions are the most commonly used for characterization of proteins in top-down proteomics.

Different dissociation techniques can be utilized to address different experimental applications. For example, HCD is generally used with Orbitrap type instruments for high throughput protein identification. UVPD can be applied to sequence intact monoclonal antibodies[42]. ETD can be beneficial to PTM analysis, especially to glycoproteomics,

since the glycans can be easily dissociated with collision-based dissociation techniques[43].

1.3 Serum autoantibody proteomics

The human immune system protects human health against disease pathogens mainly through the adaptive immune system that is reliant on the diversification of the antibody repertoire in response to antigenic determinants (antigens)[44]. When the B cell antigen receptor (BCR) on the cell membrane recognizes a specific antigen, the B cell undergoes a series of somatic gene rearrangement events to produce a specific antibody that can bind to the antigen. This antibody can prevent future attack from the same disease pathogen[44, 45]. However, the immune system can be overactive and result in autoimmune diseases such as Sjögren's syndrome (SS) and systemic lupus erythematosus (SLE). Autoimmune diseases occur when the immune system recognizes a self-antigen and generates an autoantibody against the autoantigen which will attack healthy tissue, organs, etc[46, 47].

To understand the mechanisms of autoimmune diseases, current studies focus on autoantigens and their autoepitopes[46, 48-51]. However, such research provides no information on molecular characteristics, clonality, variable gene usage, or mutational status of the autoantibodies[45, 46]. Mass spectrometry-based proteomics techniques have been used for the detection and characterization of serum monoclonal antibodies. Several bottom-up and middle-down approaches have been developed to identify autoantibodies in serum[52-54]. However, there are inherent challenges with bottom-up approaches for the analysis of the highly diverse serum antibody repertoire. In order to produce an

autoantibody repertoire of high diversity, the VDJ recombination process produces an astronomical number of antibody sequences (**Figure 1-7**). A finite set of variable (V), diversity (D), and joining (J) germline gene segments will arrange together tandemly, with the random addition or deletion of N-nucleotides, to make a unique antibody sequence. The unique process of VDJ recombination diversifies the antibody repertoire, however, the majority of the antibody sequences share common primary sequences from common VDJ and C (constant) gene families. This recombination results in a pool of peptides with both shared and non-shared sequences. Even assuming 100% sequence coverage (which is nearly impossible to generate with bottom-up approaches), bottom-up MS is unable to identify the precise coordination of individual sequences for each IgG. Top-down proteomics uses high resolution MS for accurate mass detection distinguishing different antibodies. It also allows for post-translational modification mapping and preserves the light chain and heavy chain pairing information. The advantages of top-down proteomics make it a promising tool for comprehensive antibody proteomics.

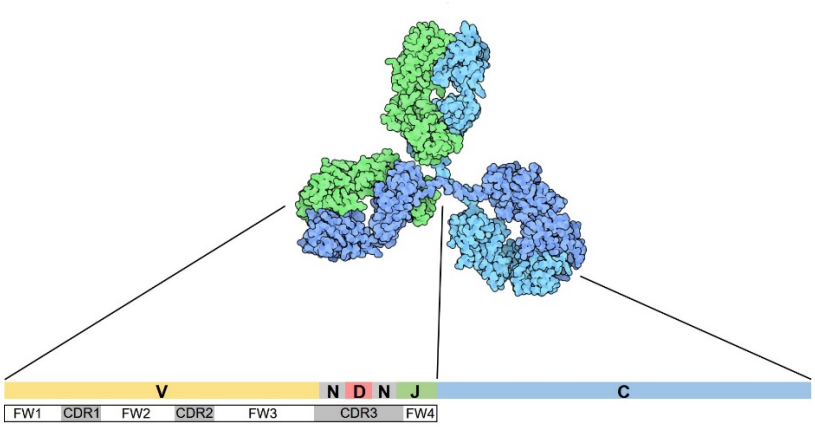


Figure 1-7. The VDJ recombination diversifying the antibody repertoire.

V: variable gene segment; D: diversity gene segment; J: joining gene segment; N: non-templated addition of N-nucleotides; C: constant gene segment; FW: framework; CDR: complementarity-determining region.

1.4 Dissertation synopsis

This dissertation presents the development and application of high-throughput top-down proteomics techniques to enable the deeper understanding of proteomes and related biochemical processes. A 1D high-performance liquid chromatography platform using a long capillary RPLC column was developed for complex intact protein sample analysis. The platform achieved a high peak capacity for intact proteome separation. To further improve the throughput, an online comprehensive 2DLC platform couples high-pH RPLC and low-pH RPLC for top-down proteomics (**Chapter 2**). The high-pH and low-pH 2D RPLC platform has previously been applied in bottom-up proteomics. This platform has good orthogonality for peptide separation, utilizes simple sample handling techniques, and 'salt'-free, MS friendly buffers. However, it has never been applied to intact protein separation. We evaluated the 2D pH RP/RPLC platform and observed orthogonality for intact protein separation. The 2DLC platform enables the deep identification of proteoform PTMs due to increased sensitivity. The 2DLC platform was further developed into an online automated system utilizing nano flow LC for both dimensions (**Chapter 3**). The automated online 2DLC system does not require sample handling between the two dimensions; as a result, less sample loss and artificial modifications occur during the experiment. More importantly, 100 times less starting material was needed (*i.e.*, 5 μg *vs.* 500 μg), to achieve similar performance to the offline 2DLC platform, which equates to a 100-fold increase in sensitivity.

Understanding the human autoantibody repertoire and immune complex is essential to explore the mechanism of autoimmune diseases for disease management and treatment. We applied high-throughput top-down proteomics to the analysis of antigen-binding

fragments from serum autoantibody repertoire (**Chapter 4**). To our knowledge, this work is the first, to date of publication, top-down MS analysis of the human serum autoantibody pool and has enabled the classification of light chains and heavy chains. Our top-down approach gives us a 'bird's-eye' view of the complexity of human serum autoantibodies. Further, we optimized and evaluated a low-temperature LC system coupled with HDX-MS to characterize the epitopes of the protective antigen (**Chapter 5**).

Chapter 2 Two-dimensional separation using high-pH and low-pH reversed phase liquid chromatography for top-down proteomics

2.1 Abstract

Advancements in chromatographic separation are critical to in-depth top-down proteomics of complex intact protein samples. Reversed-phase liquid chromatography is the most prevalent technique for top-down proteomics. However, in cases of high complexities and large dynamic ranges, 1D-RPLC may not provide sufficient coverage of the proteome. To address these challenges, orthogonal separation techniques are often combined to improve the coverage and the dynamic range of detection. In this study, a "salt-free" high-pH RPLC was evaluated as an orthogonal dimension of separation to conventional low-pH RPLC with top-down MS. The RPLC separations with low-pH conditions (pH=2) and high-pH conditions (pH=10) were compared to confirm the good orthogonality between high-pH and low-pH RPLC's. The offline 2D RPLC-RPLC-MS/MS analyses of intact *E. coli* samples were evaluated for the improvement of intact protein identifications as well as intact proteoform characterizations. Compared to the 163 proteins and 328 proteoforms identified using a 1D RPLC-MS approach, 365 proteins and 886 proteoforms were identified using the 2D RPLC-RPLC top-down MS approach. Our results demonstrate that the 2D RPLC-RPLC top-down approach holds great potential for in-depth top-down proteomics studies by utilizing the high resolving power of RPLC separations and by using mass spectrometry compatible buffers for easy sample handling for online MS analysis.

2.2 Introduction

Post-translational modifications (*e.g.,* phosphorylation, glycosylation, acetylation, etc.) and other cellular biochemical processes often result in various modified proteoforms of a single protein, which are recognized as important functional molecular signatures for disease diagnosis and potential drug targets[38]. Several mass spectrometry (MS) based proteomics approaches are currently available: bottom-up proteomics, middle-down proteomics, and top-down proteomics[7]. Bottom-up proteomics offers high sensitivity, high throughput, and good sequence coverage for studying complex protein samples[7, 55]. Middle-down proteomics refers to the partial cleavage of intact proteins into a few large fragments prior to LC-MS/MS analysis[56]. However, both techniques often result in the loss of information for different proteoforms. In top-down proteomics, intact protein samples are directly analyzed using MS without any pre-treatment. Compared to bottom-up proteomics and middle-down proteomics, top-down proteomics has the advantages of the preservation of intact proteoforms[57]. Therefore, top-down proteomics is being increasingly utilized for proteome studies. Additionally, in recent years, high-speed and high-resolution MS instrument development (*i.e.,* orbitrap and FTICR MS) has greatly advanced the field of top-down proteomics[58-60].

One of the main challenges of top-down proteomics is the lack of high-resolution separation techniques for complex protein samples[39, 61]. Reversed phase liquid chromatography (RPLC) coupled online with MS is the most prevalent approach for studying complex intact protein samples in top-down proteomics[7, 11, 28-30]. Various approaches have been developed to improve the separation power of the RPLC analysis[33-35]. One of the major efforts is to utilize ultra-high pressure LC systems with

longer columns[36] or smaller particle sizes, improving the peak capacity of the 1D RPLC separation[37]. However, due to extreme complexities, limited sample loading amounts, and large dynamic ranges of intact protein samples, 1D RPLC alone may not provide sufficient proteome coverage for top-down proteomics. Capillary electrophoresis (CE) has also been applied as a complementary separation technique to RPLC for the separation of peptides and proteins in proteomics studies. However, the application of CE has been limited for several reasons. Low concentration and ionic strength of the analytes is required to minimize peak broadening and peak distortion. Limited sample loading amounts and relatively poor reproducibility of CE have also limited its wide application[27, 31, 32]. In addition, confident intact protein and proteoform characterizations often require good quality fragmentation peaks by averaging several MS/MS scans, and this is a rate limiting step for both 1D RPLC analysis and CE based separations.

To address these challenges, various orthogonal separation techniques (*i.e.,* 2D LC) are often combined to improve proteome coverage and increase the dynamic range of detection[62, 63]. Different separation methods have been evaluated and optimized as complimentary separation techniques to RPLC, including hydrophilic interaction chromatography (HILIC), size exclusion chromatography (SEC), hydrophobic interaction chromatography (HIC), and ion exchange chromatography (IEC), etc. [24, 25]. Electrophoresis-based separation techniques, such as gel-eluted liquid fraction entrapment electrophoresis (GELFrEE)[26], and capillary zone electrophoresis (CZE) [27] have also been used. HILIC has been successfully applied to histone proteoform analysis[64, 65], but it is not widely applicable due to poor solubility of many proteins in organic loading buffers. SEC separates proteins based on the sizes of the proteins which provides a

complementary mechanism to RPLC and is useful for the identification of large proteins. However, SEC often has low peak capacity for protein separation[66, 67]. HIC is a technique that separates proteins based on the hydrophobicity with high resolution. It also provides complementary selectivity to RPLC[20, 68]. However, HIC provides inadequate protein retention and is often limited by the concentrations of salts in the mobile phase. These problems of HIC were overcome by several studies. Ge's group developed the online HIC-MS platform with the MS-compatible salt ammonium tartrate[39], and more recently, introduced a novel hydrophobic HIC packing material that can retain proteins with MS-compatible salts such as ammonium acetate[20]. IEC is another commonly applied pre-fractionation approach in top-down proteomics. It offers good separation power and is highly orthogonal to the RPLC separation. However, IEC buffers contain high concentrations of non-volatile salts that are incompatible with MS, and additional desalting steps are always required prior to MS analysis[69]. Therefore, there is a general need to develop complementary separation approaches to couple with RPLC-MS for top-down analysis.

Using high-pH RPLC as the first-dimension of separation offers the possibility of retaining the high resolution of RPLC and bringing orthogonality to the second dimension of the RPLC separation[70, 71]. It has been widely applied to bottom-up MS approaches due to the good or even better performance compared to current state-of-the-art strong cation exchange (SCX)-RPLC separation. In addition, the MS compatible mobile phases simplify the sample processing because no desalting step is required for the secondary dimension of the separation. However, this approach has not been applied to intact protein separation. In this study, we evaluate the high-pH and low-pH RPLC's for intact protein

separation using standard proteins and *E. coli* intact proteins. The optimized 2D separation platform was further applied to the separation and the identification of *E. coli* intact proteins and proteoforms. Our results demonstrate that the proposed platform provides high resolving power for both RPLC separations, good orthogonality between the two dimensions, and easy sample handling with mass spectrometry compatible buffers.

2.3 Material and methods

2.3.1 Materials and reagents

LC/MS CHROMASOLV® grade isopropanol (IPA), acetonitrile (ACN), and water were purchased from Sigma-Aldrich (St. Louis, MO). Analytical reagent (AR) grade ammonium formate (AF) and acetic acid (HAc) were also procured from Sigma-Aldrich. Pierce™ Trifluoroacetic Acid (TFA), formic acid (FA), and the Pierce™ BCA Protein Assay Kit were obtained from Thermo Scientific (Hanover Park, IL). Three standard proteins, α-Casein from bovine milk, carbonic anhydrase from bovine erythrocytes, and cytochrome c from bovine heart were obtained from Sigma-Aldrich. The packing materials for packing C5 (Jupiter particles, 5 μm diameter, 300 Å pore size) and C18 (Jupiter particles, 5 μm diameter, 300 Å pore size) columns were purchased from Phenomenex (Torrance, CA).

2.3.2 Sample preparation

Standard protein solutions (CAS, α-Casein from bovine milk; CAH, carbonic anhydrase from bovine erythrocytes; Cyt, cytochrome c from bovine heart) were prepared

by dissolving the lyophilized proteins into the HPLC mobile phase A (0.1% TFA in water for low-pH RPLC, 20 mM AF in water for high-pH RPLC) to a final concentration of 1.0 mg/mL. Intact soluble *E. coli* cell lysate proteins were obtained from the BL21 strain grown in house and by a bead-beating approach described in the literature[72]. Aliquots of protein solutions were stored at -80 ºC until further use.

2.3.3 1st dimension RPLC separation and fractionation

The first-dimension separation was performed on a Thermo Accela HPLC system (Thermo Scientific, Hanover Park, IL). An XBridge® Protein BEH C_4 column (300 Å, 3.5 μm, 2.1 mm × 250 mm) from Waters, Inc. (Milford, MA) was used. For high-pH (pH=10) 1st dimension fractionation, the mobile phase A (MPA)was 20 mM ammonium formate in water and the mobile phase B (MPB) was 20 mM ammonium formate in acetonitrile. The mobile phases were adjusted to pH 10. For low-pH (pH=2) 1st dimension fractionation, the mobile phase A was 0.1 % TFA in water and the mobile phase B was 0.1 % TFA in ACN. For both approaches, the LC flow rate was 150 μL/min, and the UV absorbance detection wavelength was set at 280 nm. Five hundred micrograms of *E. coli* intact proteins or 50 μg of the three standard proteins were loaded onto the column. For direct comparison between the high-pH 1st dimension RPLC separation and the low-pH 1st dimension RPLC separation, the same gradient was applied, and the same column was used. The LC method was set at 5 minutes for sample loading followed by a 60-minute separation gradient from 10% to 70% of MPB. The column was regenerated by running 90% of MPB over 10 minutes and equilibrated to 97% of MPA for the next run. Twenty-four fractions were collected by a fraction collector. Each fraction was vacuum dried and stored at -20 ºC.

Right before the analysis using second-dimension RPLC, each fraction was reconstituted by adding 100 µL of MPA for the second-dimension RPLC (0.01 % TFA, 0.585% HAc, 2.5% IPA, and 5% ACN in water). The protein content and concentration of each fraction were evaluated by SDS-PAGE gel and BCA Protein Assay Kit.

2.3.4 2nd dimension top-down RPLC-MS/MS analysis

An in-house packed nano-flow capillary C_5 column (5 µm, 75µm × 75 cm) was used on a modified Thermo Accela HPLC system. The RPLC conditions were similar to those previously reported[73, 74]. The mobile phase A was 0.01% TFA, 0.585% HAc, 2.5% IPA and 5% ACN in water, and the mobile phase B was 0.01% TFA, 0.585% HAc, 45% IPA and 45% ACN in water. 25 µL of each reconstituted fraction (*i.e.*, ¼ of each fraction) was loaded on an SPE column for sample trapping and cleaning. A 280-minute gradient from 10% A to 80% B at a flow rate of about 400 nL/min was applied for the separation and the column was regenerated by running 90% of MPB for 10 minutes and equilibrated to 97% of MPA. The second-dimension RPLC was coupled directly to an LTQ Orbitrap Velos Pro mass spectrometer for MS analysis. Eluents from the second-dimension RPLC were electro sprayed from a custom designed nano-ESI source into an LTQ Orbitrap Velos Pro mass spectrometer (ThermoFisher Scientific, Bremen, Germany). The electrospray voltage was set to 2.6 kV and the heated inlet capillary temperature was optimized to 300 °C. MS data were collected at the resolving power setting of 100 000 (at *m/z* 400) with two micro scans. Data-dependent MS/MS acquisition was performed by selecting the top five most abundant precursor ions in the MS scan with an isolation width of 3.0 and fragmenting them using collision induced dissociation (CID) with a normalized

energy of 35%. The MS/MS data were obtained at a resolving power setting of 60 000 (at *m/z* 400) with one micro scan. Ions with less than 4 charges were rejected for the selection of MS/MS scans. The maximum injection time for a full mass scan and MS/MS scan were set to 1000 ms. and 500 ms., respectively. The AGC target was set as 1×10^6 for full mass scans, and 5×10^5 for MS/MS scans. All the data were collected with Xcalibur 3.0 software (Thermo Fisher Scientific, Bremen, Germany).

The *E.coli* intact proteins were also analyzed by a single-dimensional RPLC-MS/MS (1D RPLC) to compare with the 2D RPLC-RPLC-MS/MS platform. The exact same conditions for LC and MS as the 2nd dimension top-down RPLC-MS/MS analysis were applied.

2.3.5 Bottom-up LC-MS/MS analysis of fractions from 1st dimension RPLC

The fractions from the first-dimension RPLC separation were tryptic digested and analyzed by the bottom-up approach for the evaluation of orthogonality between the low-pH RPLC and high-pH RPLC. Briefly, the vacuum dried fractions were reconstituted in 25 mM ammonium bicarbonate (ABC) and 6 M urea. Two hundred mM dithiothreitol (DTT) was used to reduce the disulfide bonds. Two hundred mM iodoacetamide (IAA) was used to protect the thiol groups from re-forming disulfide bonds. Trypsin was added to fractions with a protein to enzyme ratio of 50:1 and the digestion was performed overnight at 37 ºC. The digested fractions were desalted and loaded onto an in-house packed C_{18} column (5 µm, 75µm × 15 cm) for the bottom-up study. The mobile phases were 0.1% formic acid in water (MPA) and 0.1% formic acid in ACN (MPB). The gradient was from 3% MPB to 35% MPB over 40 minutes following a 15-minute sample loading step. The column was

regenerated with 90% MPB for 10 minutes and equilibrated to 3% MPB for 30 minutes.

2.3.6 Protein and proteoform identification

In bottom-up experiments, peptides were identified using MSGF+[75, 76] to search the mass spectra from the LC-MS/MS analysis against the annotated *E coli.* database and its decoy database. Peptide identifications were filtered using an MS-GF cut-off value of $1x10^{-10}$ (*i.e.*, the calculated FDR<1% at the unique peptide level). The intact protein MS/MS data were subjected to data analysis and protein identification using MS-Align+ (http://bix.ucsd.edu/projects/msalign/)[77] with the following search parameters: minimal precursor mass = 2500 Da; minimal fragment peaks per scan =10; maximum number of modifications = 2; fragment mass error tolerance = 15 ppm. MS-Align+ reported only the PrSM with the best E-value for each spectrum. LC-MS/MS data were searched against the annotated *E coli.* database. The false discovery rate (FDR) for protein/spectrum matches was estimated by searching top-down spectra against the human Uniprot database. A final E-value cutoff of $2x10^{-4}$[75] was used to achieve an FDR of 1%. All of the identified proteins and proteoforms were further manually evaluated.

2.4 Results and discussion

2.4.1 Evaluation of orthogonality between high-pH RPLC and low-pH RPLC

We first evaluated the orthogonality between high-pH RPLC and low pH-RPLC using three standard proteins, α-casein, carbonic anhydrase, and cytochrome c. The standard proteins were loaded on the same column with different mobile phases (low pH

and high pH) individually. The elution time was normalized to the percentage of mobile phase B for direct comparison (**Figure 2-1**).

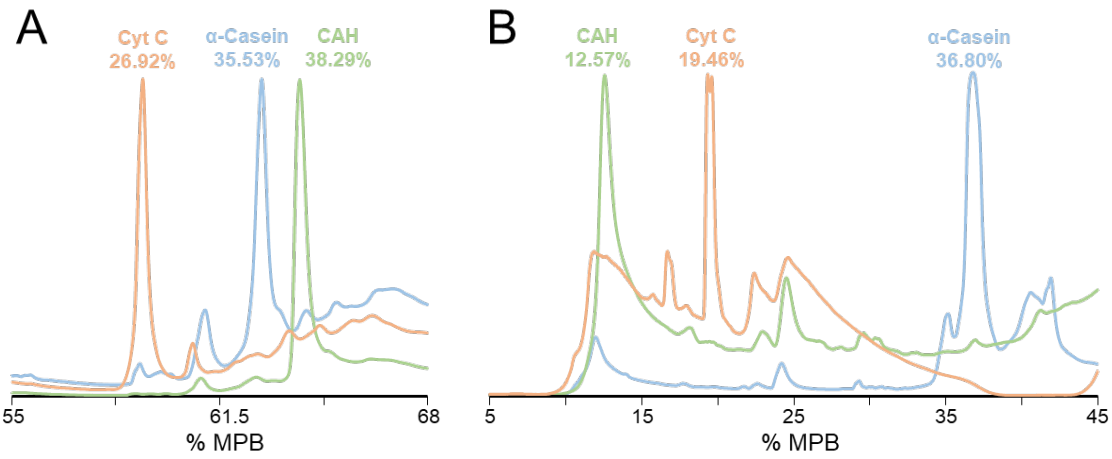


Figure 2-1. The separation of three standard proteins.

(A) low pH RPLC, and (B) high pH RPLC. The retention time was normalized to the percentage of mobile phase B.

Under low-pH conditions, the elution order of the three standard proteins was Cyt, Cas, and CAH, while under high-pH conditions, the elution order was CAH, Cyt, and Cas. The change in elution order of the standard proteins indicates that under different pH conditions, different proteins will have different retention behaviors which provides the possibility of achieving orthogonal separation using low-pH RPLC and high-pH RPLC. In addition, we observed that the separation window under high-pH conditions was wider than that under the low-pH conditions. The separation window under low-pH conditions was about 12% MPB, and the separation window under the high-pH conditions was about 24% MPB further indicating the differences in retention behaviors of different proteins between high-pH RPLC and low-pH RPLC.

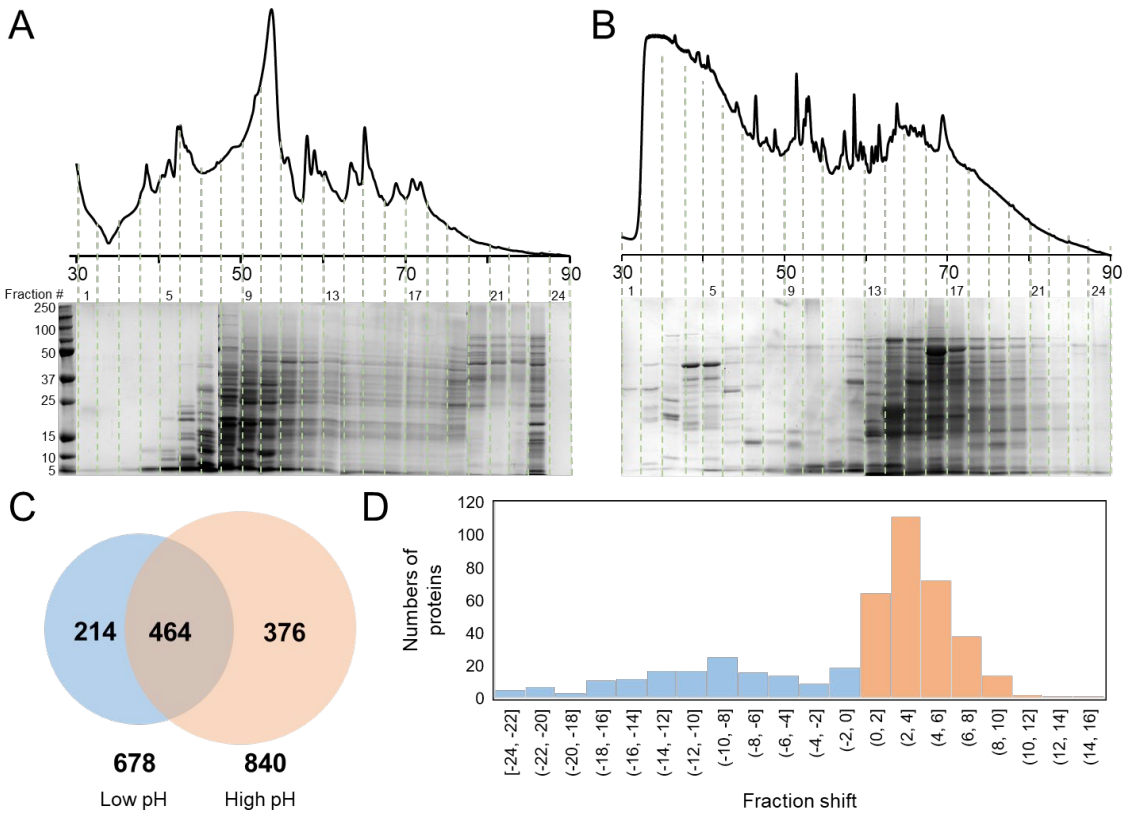


Figure 2-2. Comparison of low-pH RPLC and high-pH RPLC as the first-dimension separation.

(A) UV chromatograms SDS-PAGE using low-pH RPLC; (B) UV chromatogram and SDS-PAGE using high-pH RPLC. (C) Venn diagram of unique proteins identified using bottom-up proteomics. (D) Histogram of the fraction shift from low pH RPLC to high pH RPLC.

We further evaluated the performance of the high-pH RPLC and the low-pH RPLC for separating complex protein mixtures such as *E. coli* intact proteins. For both RPLC separations, the same elution gradient (10-70% of MPB) was employed for direct comparison. **Figure 2-2** shows the elution profiles (UV chromatograms) of *E.coli* soluble intact proteins under different pH conditions. The fractions from both low-pH RPLC and high-pH RPLC were analyzed by SDS-PAGE. A significant difference between the protein elution profiles was observed on both UV chromatograms and SDS-PAGE. All fractions

were analyzed by bottom-up MS for protein identification, and protein elution profiles were evaluated under different pH conditions (**Figure 2-2**). Overall, with bottom-up MS, a total of 678 unique proteins were identified using low-pH RPLC, and 840 proteins were identified using high pH RPLC with an overlap of 464 proteins between the two methods.

The fraction shift was calculated to evaluate the orthogonality between different pH conditions (The fraction shift = the fraction number where the protein is most abundant under the high-pH condition – the fraction number where the protein is most abundant under the low-pH condition). Most of the identified proteins have a fraction shift larger than 2 (about 5% of the organic component), indicating good orthogonality between the two methods. Previous studies suggest that protein retention behaviors depend heavily on the composition of amino acids and their three-dimensional conformations.[78] Five amino acids (Glu, Asp, Arg, His, and Lys) can be differentially ionized under different pH conditions. These five amino acids have either acidic side chains or basic side chains of which the polarities will be reversed when the conditions change from low-pH to high-pH. Thus, the retention behavior changes under different pH conditions and depends heavily on the ratio of these five amino acids in the sequence. Other than these five charged amino acids, all other types of amino acids have higher or similar retention coefficient under high-pH condition than under low-pH condition. Therefore, proteins are often eluted in later fractions under high-pH conditions, which is consistent with our observations (**Figure 2-2D**). Compared to the separation of peptides using a high-pH and low-pH RPLC platform reported by Gilar[70], the retention time shift of intact proteins was greater than the shift of tryptic-digested peptides, which may indicate that pH conditions have greater impacts on the hydrophobicity of proteins than that of peptides. We hypothesize that the large

properly folded proteins change properties more than small peptides because changing a single charge somewhere of a large protein can cause large conformational changes which will greatly affect retention behavior.

We also compared the fraction width of each protein identified under high-pH and low-pH conditions (Fraction width refers to the number of fractions where a certain protein is identified). Fraction width under high pH is less than the fraction width under low pH, which is consistent with the greater identification number under high-pH conditions. Both the identification number and fraction width proved that RPLC with high-pH conditions provides similar or even better separation power than conventional low-pH RPLC, which shows high potential for high-pH RPLC as an orthogonal separation technique to yield high resolution 2D separations.

One interesting observation was that under high-pH conditions, the retention behaviors of proteins seem to be less molecular weight dependent. According to previous reports [79, 80], larger proteins tend to have greater retention times in RPLC due to greater hydrophobicity under low-pH conditions. One possible explanation of our observation is that protein charge distributions are different under different pH conditions. Basic amino acids such as Lys, Arg, and His are positively charged under low-pH conditions, and acidic amino acids such as Glu and Asp are negatively charged under high-pH conditions. The location of these charged groups can have effects on protein elution. It has been reported that the proteins are more positively charged at their N-terminal and C-terminal, and more negatively charged at their core regions[81]. Therefore, there are less charges in the core regions under low-pH conditions and there are possibly more chain-length-dependence effects[82]. On the other hand, there are possibly more nearest-neighbor effects[83] on

protein elution due to neighboring negatively charged amino acids at their core region under high-pH conditions. Still, the mechanisms of these proposed effects are unknown, and additional research is needed to further explore these observations.

Overall, our results demonstrate the good orthogonality between low-pH RPLC and high-pH RPLC, which may be combined as a 2D separation approach for top-down MS analysis. The low-pH RPLC was selected as the second-dimension separation to directly couple with MS because the protein ionization efficiency is higher under acidic conditions and proteins tend to have higher charge state distributions that can be efficiently detected under the normal Orbitrap MS scan range (*i.e.*, 400-2000).

2.4.2 2D high-pH and low-pH RPLC-MS/MS analysis of *E. coli* intact proteins

The two-dimensional separation platform using high-pH RPLC as the first dimension and low-pH RPLC as the second dimension was applied to the identification of complex *E. coli* proteins. To further evaluate the orthogonality between the 1st dimension separation and 2nd dimension separation, a total of 24 sequential fractions (*i.e.*, 1-min per fraction, **Figure 2-2A**) were analyzed using the 2nd dimension RPLC-MS. For future applications, the fractionation scheme can be optimized to justify the time investment and its benefits. In addition, a targeted fraction analysis can be applied to increase the proteoform coverage of a specific protein or several proteins in a complicated background setting without significant increases in analysis time (*i.e.,* with pre-fractionation, only several fractions need to be analyzed with 2nd RPLC-MS/MS reducing the total analysis time significantly).

To evaluate the improvement from the 1D low pH RPLC to the 2D high- and low-

pH RPLC/MS platform, we analyzed *E. coli* intact proteins with 1D RPLC as well. Our results show that 2D pH-PRLC-RPLC separation allows better separation of protein mixtures with more proteins and proteoforms identified in the *E. coli* lysate. A total of 365 proteins and 886 proteoforms were identified with the 2D pH-RPLC-RPLC-MS/MS analysis, which is a significant improvement over 163 proteins and 328 proteoforms identified using the 1D RPLC method. An overlap of 121 proteins and 139 proteoforms between the 2D and 1D method was observed (**Figure 2-3A and 2-3B**).

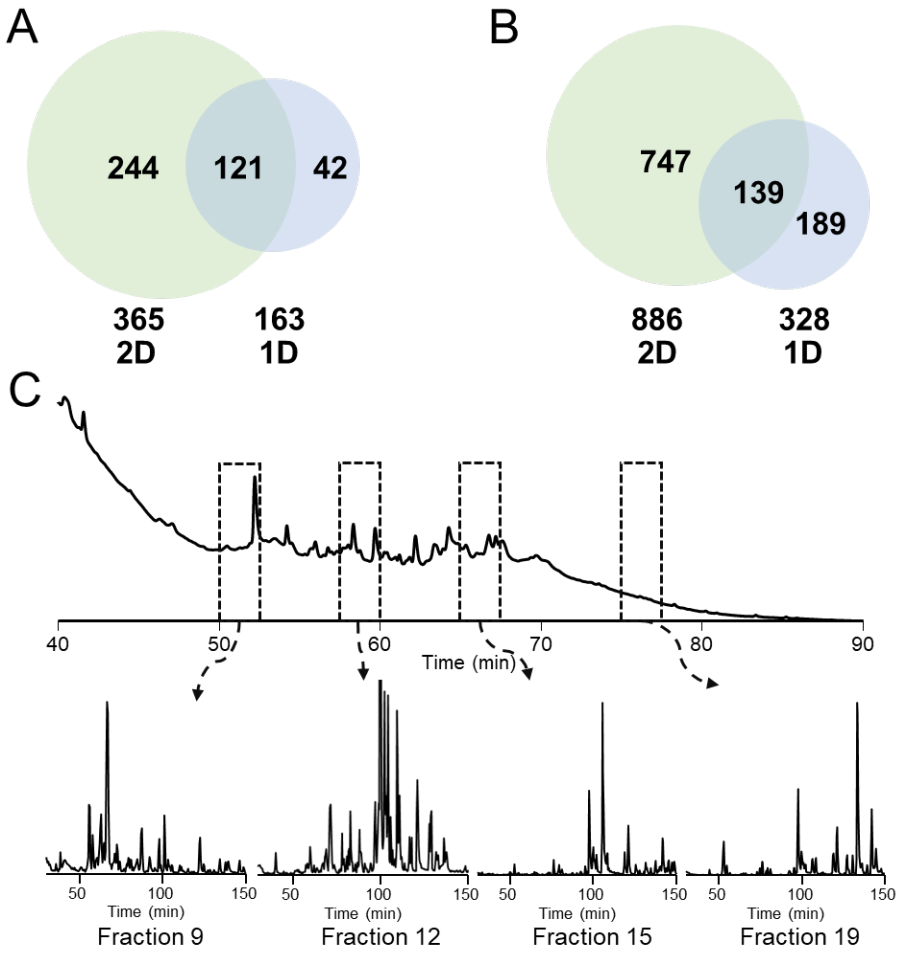


Figure 2-3. Identification results of the 2DLC.

Venn diagram of numbers of (A) proteins and (B) proteoforms identified using 1D and 2D methods. (C) Base peak chromatograms of 4 representative fractions.

However, there were some proteins that were only identified in the 1D RPLC-MS/MS analysis. One possible explanation is that some proteins were unable to bind to the packing material under high pH conditions resulting in loss of identification of those proteins. A flow-through fraction will be incorporated in future studies to ensure the identification of unbound proteins.

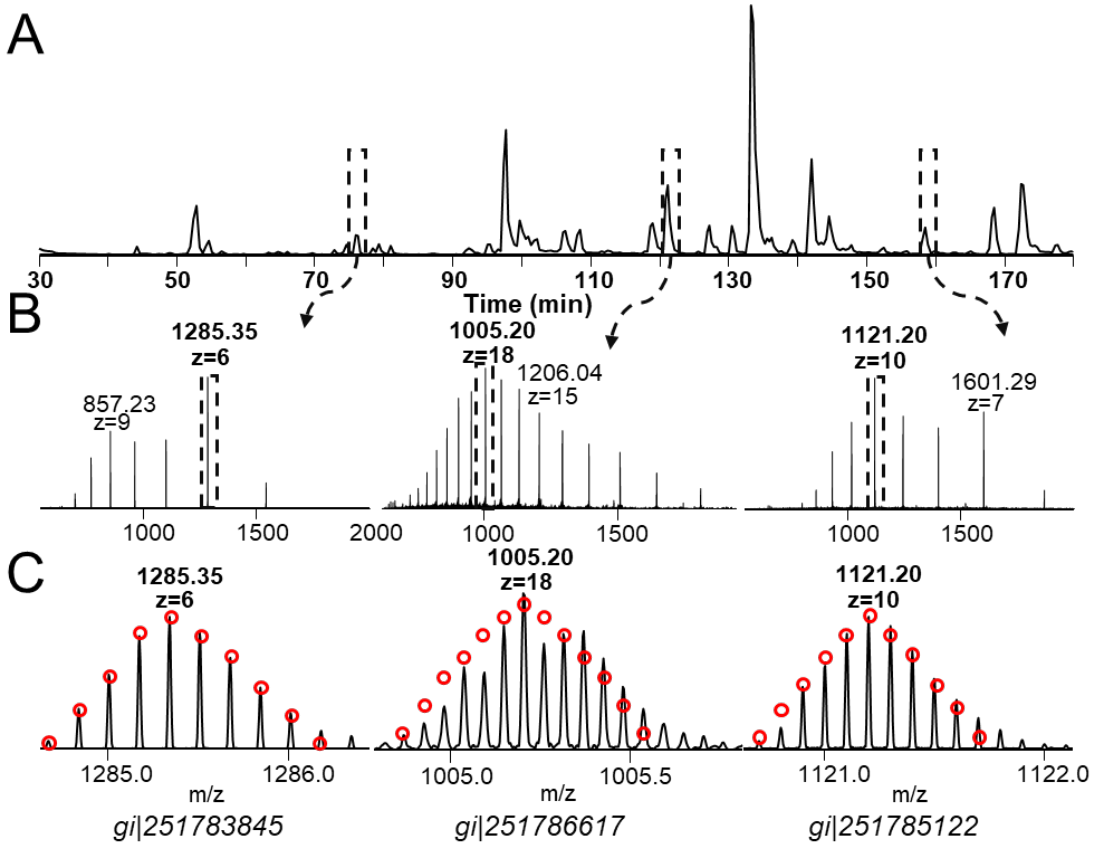


Figure 2-4. 2D pH RP/RPLC/MS analysis on *E.coli* proteins.

(A) Base peak chromatogram of second dimension top-down analysis of fraction 19. (B) Representative mass spectra of three proteins identified in fraction 19. The three proteins are periplasmic protein, peptidyl-prolyl cis-trans isomerase A, and mono-oxygenase, subunit of predicted monooxygenase. (C) Overlay of observed isotopic distribution and theoretical isotopic distribution (red cycles).

The base peak chromatograms (BPCs) of individual fractions in the second dimension were evaluated (*i.e.,* the BPCs of fraction 9, 12, 15, and 19 were demonstrated in **Figure 2-3C**). **Figure 2-4** shows the examples of three identified proteins with isotopic distributions and molecular weight from 7.7 kDa to 18 kDa. For all 4 fractions, the BPC profiles indicated good orthogonality between different pH RPLCs because most of the proteins in these fractions were eluted out over the entire separation gradient (**Figure 2-3C**). In addition, the BPCs of these 4 selected fractions have significantly different patterns as indicated in **Figure 2-3C**. These two observations further supported the good orthogonality between high pH RPLC and low pH RPLC.

To further evaluate the improvement of top-down MS performance using 2D pH-RPLC, the 5-minute segments (retention time from 90 minutes to 95 minutes) in the gradient of the 1D separation and the 2D separations of each fraction were compared (**Figure 2-5**). In the 1D RPLC-MS analysis, a total of 9 mass features were found in the selected 5-minute segment where6 unique proteins were identified. On the other hand, a total of 65 unique mass features were found in all the fractions in the 2D method, and 28 unique proteins were identified. Some of the identified proteins were color coded in **Figure 2-5**. A protein with the *m/z* of 1211.46 and charge state of 13 was identified in the 1D RPLC-MS analysis as a superoxide dismutase precursor (Cu-Zn) protein. This protein was also identified in the fraction 17 of the 2D analysis. Interestingly, with the 2D separation, another precursor ion with similar *m/z* (i.e., m/z=1211.57) but different charge state (i.e., z=15) was observed in the same elution window in fraction 6 of the 2D analysis, which was confirmed as gi|251785751. With 1D separation, only high-abundance proteins can be observed when the m/z of the detected ions are overlapping. Due to less sample

complexity in the 2D method, more proteins can be identified, especially some low-abundance proteins or from some overlapped peaks. The identification of intact proteins benefited from the orthogonality between the different pH RPLC separations. In addition, due to the improved separation power with the 2D pH-RPLC method, more proteoforms can be identified for the same proteins. For example, only one proteoform of the stress response protein *(gi|251787301)* was identified with the 1D RPLC method, however, 4 different proteoforms (unmodified protein, lysine acetylation, methionine oxidation, and C-terminal degradation) were identified with the 2D method.

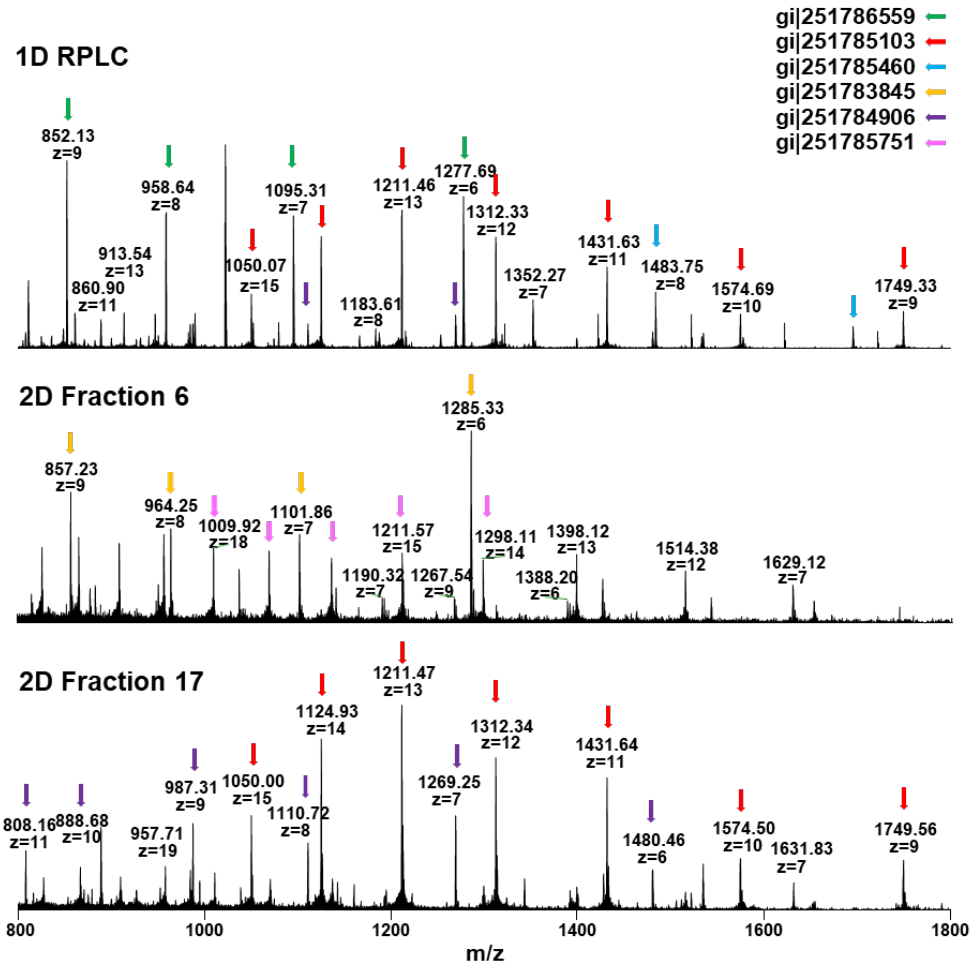


Figure 2-5. Five-minute segment (RT = 90 min to 95 min) from LC/MS runs of 1D and 2D methods.

It is known that the methionine residue can be oxidized through the interaction with molecular oxygen during sample processing steps. Therefore, we compared the identified methionine oxidized proteoforms between the 1D RPLC-MS (280 min) analysis and the 2D RPLC-MS analysis. Overall, we detected 2 methionine oxidized proteoforms in the 1D analysis, and 22 methionine oxidized proteoforms in 2D analysis. We further compared the intensities between the oxidized proteoforms and the non-oxidized proteoforms, and the average ratio was about 1:10 for both 1D and 2D. Our results do not suggest extensive oxidation arising from the additional processing steps and time in the 2D analysis. Still, the detected proteoforms with oxidized methionine are likely from sample preparation steps such as cell lysis.

Our 2D separation approach is relatively simple when compared with some other multidimensional separation approaches such as the GELFrEE [26] where an SDS removal step is necessary or and ion exchange chromatography where desalting step is often required before sample concentration. The only step after fraction collection and before sample injection is that fractions are concentrated using vacuum drying where the frozen solution sublimes under vacuum (*i.e.,* low oxygen level).

2.4.3 Identification of novel intact proteoforms of apo-acyl carrier protein

The apo-acyl carrier protein *(gi|251784624)* was identified in both the 1D and 2D methods. Apo-acryl carrier protein (ACP) is a unique protein working as a coenzyme in fatty acid and polyketide biosynthesis[84]. The protein is expressed in an inactive form. The phosphopantetheinyl transferase activates the protein after the expression by adding the phosphopantetheine moiety to serine 37 on ACP[85]. During the biosynthesis process,

the growing fatty acid chain is tethered to the thiol group of the phosphopantetheine on serine 37.

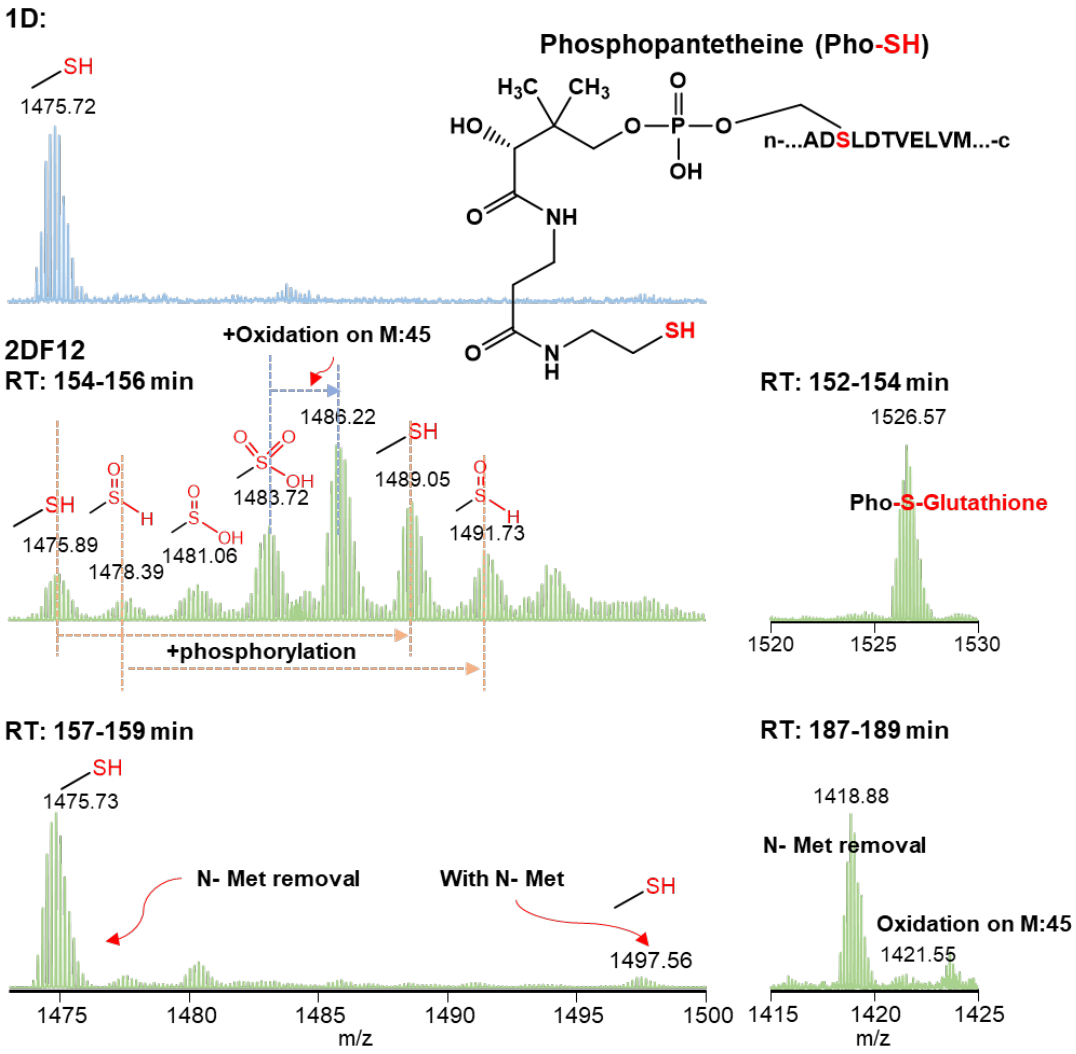


Figure 2-6. Different proteoforms of apo-acyl carrier protein identified using 1D and 2D methods.

From our 2D results, we identified both inactive and active forms of ACP, namely the protein with and without phosphopantetheine modifications (short as Pho-SH in **Figure 2-6**) on serine 37 in the ACP sequence. Due to the improved separation, we were able to identify other low-abundance proteoforms, such as the proteoforms with and without N-

terminal methionine, the proteoform with an oxidation on methionine 45, etc. Interestingly, we also observed several novel proteoforms with different types of oxidized thiol groups of the phosphopantetheine on serine 37. We confirmed the proteoforms with Pho-SH, Pho-SOH, Pho-SO_2H, Pho-SO_3H by the MS/MS results as well as by comparing the isotopic distribution of the detected peaks with the theoretical distribution (*data not shown*). Another oxidative proteoform was also confirmed as the glutathionylation on the thiol group of the phosphopantetheine on serine 37, which is a common non-enzymatic modification of cysteine thiol groups resulting from oxidative stress and preventing irreversible oxidation of thiol groups[86-89]. Interestingly, all of the identified oxidative forms on the thiol group of the phosphopantetheine mimic the PTMs of cysteine thiol groups, which indicates the similar reactivity of these two types of thiol groups. One thing we would like to point out is that most of the different proteoforms of ACP were observed in one fraction (i.e., fraction 12), which indicates the modifcations on the thiol group of phosphopantetheine do not greatly affect the retention times of the protein. Using the 1D method, however, we only observed one proteoform as the active apo form of ACP with the removal of the N-terminal methionine. This enhanced identification of PTMs of ACP has convincingly proven that the 2D pH-RPLC-RPLC method improved proteoform identification by simplifying the complexity of the samples. All of the identifications of PTMs were characterized by MS^2 spectra generated by collision induced dissociation. [17]

Among all the 11 proteoforms of apo-acyl carrier protein in our study, we identified 4 oxidized forms of the thiol group. We confirmed all of the Pho-S oxidative forms by MS/MS spectra. Because the reactivity of the Pho-SH thiol group has not been studied before, it is difficult for us to conclude if the observed oxidative proteoforms are from the

sample process. Interestingly, a proteoform with a free thiol group on phosphopantetheine was observed. Thiol groups normally have high reactivity in biological samples, especially with other thiol groups, forming disulfide bonds. It would be interesting to study on the function of the thiol group on the phosphopantetheine in the future.

2.5 Conclusions

We report the development and evaluation of a two-dimensional separation technique using a high- and low- pH RPLC/MS/MS platform for top-down proteomics. It achieves orthogonal separation by altering the pH conditions of the mobile phases while taking advantage of the high resolving power of RPLC. This allows the use of RPLC in both dimensions and offers higher resolution and better sensitivity than 1D techniques. From both the retention times of standard proteins as well as the bottom-up results of fractions from *E. coli* proteins using different pH RPLC's, we have proven that altering the pH conditions of the mobile phases used in reversed phase chromatography changes the retention times of proteins in a useful manner. The platform was further applied to the identification of intact proteins in an *E. coli* lysate and it allowed the identification of greater numbers of proteins and showed higher proteoform coverage compared to 1D RPLC/MS top-down proteomics. This 'salt-free' RPLC process overcomes the disadvantage of other chromatographic methods which require complicated sample processing (*i.e.* buffer exchange, desalting) before MS analysis and provides the potential of easily coupling two dimensions of separation for online analysis. This is an important advance in separating complex intact protein samples for high throughput top-down proteomics.

*The materials in Chapter 2 are adapted from

Wang, Z., Ma, H., Smith, K. and Wu, S., 2018. Two-dimensional separation using high-pH and low-pH reversed phase liquid chromatography for top-down proteomics. *International journal of mass spectrometry*, 427, pp.43-51.

Chapter 3 Development of an online 2D ultra-high-pressure nano-LC system for high-pH and low-pH reversed phase separation in top-down proteomics

3.1 Abstract

The development of novel high-resolution separation techniques is crucial for advancing the complex sample analysis necessary for high-throughput top-down proteomics. Ultra-high-pressure long capillary column RPLC separation has been applied to top-down proteomics for the improvement of separation resolution because it allows the identification of hundreds of intact proteoforms from complex biological samples at low microgram sample amounts. Recently, our group developed an offline 2D high-pH RPLC/low-pH RPLC separation method and demonstrated good orthogonality between these two RPLC formats. To further improve the throughput and sensitivity of the offline approach, we developed an online 2D ultra-high-pressure nano-LC system for high-pH and low-pH RPLC separations in top-down proteomics. A micro-trap column with an online dilution setup was used to collect eluted proteins from the high-pH separation and to inject fractions for ultra-high-pressure long capillary column low-pH RPLC separation in the second dimension. This platform enables the characterization of 1000+ intact proteoforms from 5 micrograms of intact *E. coli* cell lysate in 10 online-collected fractions. Here, we have demonstrated that our online 2D pH RP/RPLC system coupled with top-down proteomics holds the potential for deep proteome characterization of mass-limited samples.

3.2 Introduction

Top-down proteomics analyzes intact proteoforms resulting from genetic variations, alternative RNA splicing, and post translational modifications [1]. Whereas bottom-up proteomics is robust and sensitive but provides limited information on proteoforms, top-down proteomics can provide a complete description of the primary structure of a protein and its modifications [90]. However, there are significant challenges posed to top-down proteomics by the complexity of protein samples [23]. Proteomes normally present a wide dynamic range of thousands of proteoforms, which makes the separation, accurate mass spectrometric detection, and data analysis extremely complicated. One of the major efforts to address these challenges is the improvement of separation resolution prior to MS analysis.

Various separation techniques have been developed to be directly coupled to MS for high-throughput top-down proteomics, including RPLC [16, 18, 19] and capillary electrophoresis (CE) [91-94]. Multidimensional separation platforms have also emerged to improve the separation resolution of intact proteins from complex biological samples based on their orthogonal separation mechanisms [17, 21, 38, 39, 64, 91, 95-99]. Recently, our group reported the development of a 2D-LC platform that coupled high-pH RPLC to low-pH RPLC for top-down proteomics. Fractions from the 1st-dimension high-pH RPLC were further separated using an ultra-high-pressure long capillary column for high-resolution low-pH RPLC-MS analysis. The offline platform was applied to characterize intact proteoforms from *E. coli* lysate [17] and from human *HeLa* lysate[100]. Our results suggested that orthogonal separation by altering the pH conditions of the mobile phases of RPLC was be achieved and the use of RPLC in both dimensions offered higher resolution

and better sensitivity than 1D-LC techniques. However, additional sample handling steps in offline 2D-LC often make the techniques lower-throughput, introduce degradation products, and result in poor sample recovery.

Online 2D-LC has been developed to overcome some limitations of offline 2D-LC [19, 95, 96, 101, 102]. Recently, Baghdady *et al.* reported a proof-of-principle online 2D system that couples high-pH RPLC and low-pH RPLC for intact protein separation [101]. In Baghdady's work, microflow RPLC columns were used in both dimensions and two sample storage loops were used for online coupling. In the 2nd dimension low pH RPLC separation, a very steep gradient (<1 min) was utilized which has been reported to enable the injection of sample volumes up to 70% of the column dead volume without band broadening for peptides and proteins [101, 103]. However, the separation resolution of the 2nd dimension RPLC was relatively low, with a peak capacity of 19. This low resolution in the second dimension makes this approach less practical for improving proteoform identification without significantly improved 1st-separation resolution since online MS detection largely relies on decreased protein complexity eluted from the 2nd dimension RPLC.

3.3 Materials and methods

3.3.1 Chemicals and materials

LC/MS grade acetonitrile (ACN), isopropanol (IPA), water, analytical reagent (AR) grade ammonium formate (AF), and acetic acid (HAc) were purchased from Sigma-Aldrich (St. Louis, MO). Pierce™ Trifluoroacetic Acid (TFA), formic acid (FA), and

Pierce™ BCA Protein Assay Kit were obtained from Thermo Scientific (Hanover Park, IL). The packing material for packing C5 (Jupiter particles, 5 µm diameter, 300 Å pore size) column was purchased from Phenomenex (Torrance, CA), and the packing material for high-pH C4 columns was obtained from an ACQUITY UPLC Protein BEH C_4 300 Å column (3.5 µm particles, 300 Å pore size, 4.6 mm × 100 mm) from Waters (Milford, MA). The 15,000 psi nano-volume 6-port switching valve and UPLC fittings were purchased from VICI Valco Instruments Co. Inc. (Houston, TX). Capillary tubing of different sizes was purchased from Molex (Lisle, IL).

3.3.2 Sample preparation

BL21 strain *Escherichia coli* (*E. coli*) cells were incubated in LB medium at 37 ℃ for 12 hours. The harvested *E. coli* cells were centrifuged at 10,000 × g at 4 ℃ for 10 minutes to separate the cell pellets from culture medium. Cells were lysed using a high-pressure homogenizer unit for 3 minutes (Avestin C3 EmulsiFlex homogenizer, pressure maintained between 1,500 psi and 2,000 psi). The cell lysate was then centrifuged at 10,000 × g at 4 ℃ for 30 minutes to get rid of the cell debris. The supernatant was desalted using a 3K MWCO filter tube (Amicon®, MilliporeSigma). The concentrations of the *E. coli* lysate were determined using BCA Protein Assay from Thermo Scientific. Aliquots of protein solutions were stored at -80 ℃ until further use.

3.3.3 Instrumental setup of 1D-LC system

All the 1D-LC experiments were performed on a modified Thermo Accela HPLC system as described in our previous work.[17] An in-house packed nano-flow capillary

column (75 cm, 75 μm I.D. × 360 μm O.D., C5, 5 μm particle size) was used. The separation utilized 0.01% TFA, 0.585% HAc, 2.5% IPA and 5% ACN in water as mobile phase A, and 0.01% TFA, 0.585% HAc, 45% IPA and 45% ACN in water as mobile phase B. Twenty microliters of 0.25 μg/μL *E. coli* proteins were loaded on an SPE column (C5, 5 μm, 150 μm I.D. × 5 cm) using an autosampler installed with a 25-μL sample loop. The sample was loaded onto the SPE with a flowrate of 6 μL/min for 15 minutes using buffer A, then the valve was switched to connect the SPE to the analytical capillary column for sample separation, where a 200-minute gradient from 10% A to 70% B at a flow rate of 400 nL/min was applied. After the elution gradient, the gradient was ramped to 90% B and the column was flushed with 90% B for column washing. The gradient was then shifted back to 100% A to re-equilibrate the column for 20 minutes.

3.3.4 Instrumental setup of online 2D-LC system

The online 2D-LC platform was built using a Thermo Accela LC system (one autosampler, two Accela pumps), with two 6-port switching valves from Valco Instruments Co. Inc. (**Figure 3-1** and **3-2**). The 1st dimension RPLC column (5 cm, 150 μm I.D. × 360 μm O.D., Waters BEH C4 3.5 μm particles, 300 Å pore size), the 2nd RPLC column (75 cm, 75 μm 8.D. × 360 μm O.D., C5, 5 μm particles), and the SPE micro-trap column (5 cm, 150 μm I.D. × 360 μm O.D., Waters BEH C4, 3.5 μm particles, 300 Å pore size) were packed in-house. For high-pH RPLC separation, 20 mM ammonium formate in water (pH=10) was used as mobile phase A and 20 mM ammonium formate in acetonitrile (pH=10) was used as mobile phase B. The pH values for buffers were adjusted with ammonium hydroxide. (1) Sample loading: 5 μg of *E. coli* proteins was loaded onto the

1st dimension high-pH C4 column using an autosampler when the switching valve #1 was set at position A to allow the sample loop to be connected to the 1st dimension column. Meanwhile, the switch valve #2 was set at position A so the SPE micro-trap column was connected to the 1st dimension column to collect the flow-through fraction (*e.g.,* proteins did not bind to the 1st dimension column); (2) Sample elution from 1st dimension column: The switching valve #1 was set to position B to bypass the sample loop after sample loading to ^{1}D column. The first-dimension separation utilized step-gradient to elute the proteins from ^{1}D column to the SPE. To enable the binding of eluted proteins from the first-dimension separation to the ^{2}D column, a flow from the ^{2}D pump was used to dilute the ^{1}D elution with a 1:10 dilution ratio. The flowrate was regulated using a splitting column to achieve a flowrate of 3 μL/min through the 1D column and a flowrate of 25 μL/min through the SPE. For each step, the proteins on the ^{1}D column were eluted with a certain mobile phase B percentage range to the SPE at position A of the switching valve #2. (3) Sample separation and detection from 2nd dimension column: At 70 minutes, the switching valve #2 was switched to position B where the SPE was connected to the ^{2}D column for second-dimension separation. After that, the gradient of the second-dimension separation started at 10% B and ramped to 70% B over 200 minutes. Meanwhile, the first-dimension gradient was set back to 100% A. After the elution gradient, 270 minutes, the second-dimension gradient was set to 90% B to wash the ^{2}D column for 10 minutes, and then back to 100% A to re-equilibrate the column. Repeating methods with different mobile phase B ranges for the first-dimension separation were applied to perform the online comprehensive 2DLC. The flow-through from the first-dimension separation was also analyzed after ^{2}D separation.

3.3.5 Top-down MS analysis

The analytes from LC separation were injected into an LTQ Orbitrap Elite mass spectrometer (ThermoFisher Scientific, Bremen, Germany) via a custom designed Nano-ESI interface. The electrospray voltage was set to 2.6 kV and the inlet capillary temperature was optimized to 300 °C. The MS was set to a resolving setting of 120,000 (at *m/z* 400) with three micro scans with max ion time of 1,000 ms. The precursor ions were selected using an isolation window width of 6.0 under the data-dependent MS/MS acquisition mode where the top six most abundant precursor ions in the MS scan was selected. The selected precursor ions were fragmented using collision induced dissociation (CID) with a normalized energy of 35%. The MS/MS data were collected at a resolving power setting of 60,000 (at *m/z* 400) with three micro scans with max ion time of 500 ms. Ions with less than 4 charges were rejected for the selection of MS/MS scans. The dynamic exclusion was enabled where repeat count was set to 1, repeat and exclusion duration was set to 90 s with an exclusion list size of 500 count. The AGC target was set as 1×10^6 for full mass scans, and 3×10^5 for MS/MS scans. All the data were collected with Xcalibur 3.0 software (Thermo Fisher Scientific, Bremen, Germany).

3.3.6 Data analysis

The RAW files were converted into mzXML files with msconvert.[104] Top-down data were deconvoluted and searched against the annotated *E coli.* Database from UniProt using top-down mass spectrometry based Proteoform Identification and Characterization (TopPIC) suite[105] with the following search parameters: error tolerance for precursor and fragment masses was set to 15 ppm, the unexpected mass shifts were set to -500 to

500, maximum number of unexpected modifications was set to 2, the spectrum level and proteoform level cut off were both set to EVALUE less than 0.01. TopPIC only reported one PrSM with the best E-value for each spectrum. To remove the redundant proteoforms, the proteoforms with mass differences less than 3.7 Da from the sample protein were manually evaluated and removed if they shared a sequence with a proteoform with a mass shift less than 3.7 Da.

3.4 Results and discussion

We have recently demonstrated a peak capacity of 200 for a single low-pH RPLC separation using an ultra-high-pressure long capillary column in top-down proteomics[16]. To further improve the separation resolution, here we report a microtrap column-based online 2D-LC platform using high-pH RPLC and low-pH RPLC separation with the ultra-high-pressure long capillary column being used in the low-pH RPLC separation. Nanoscale UPLC columns were also used in the 1st-dimension of high-pH RPLC to improve the sensitivity of the system. To operate these columns, ultra-high-pressure nanoflow pumps are often required. However, commercially available nanoflow UPLC systems are often expensive and hard to maintain. To provide an alternative affordable solution, we customized a high-pressure normal flow LC system with two Thermo Accela LC pumps (operational pressure 10,000+ psi, operation flow rate 10-1000 μL/min), two high-pressure 6-port switching valves and splitting columns [16], and one Thermo Accela autosampler. The schematic diagram of the online 2D system is shown in **Figure 3-1**. The experimental details are shown in **Figure 3-2**.

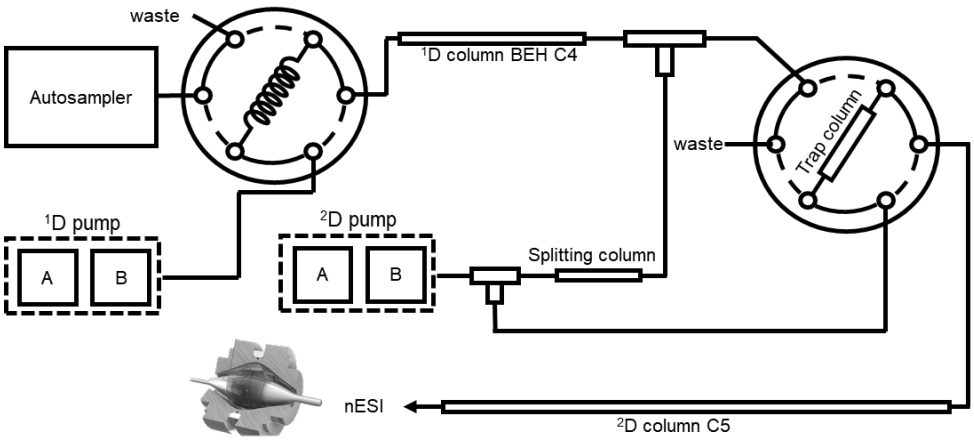
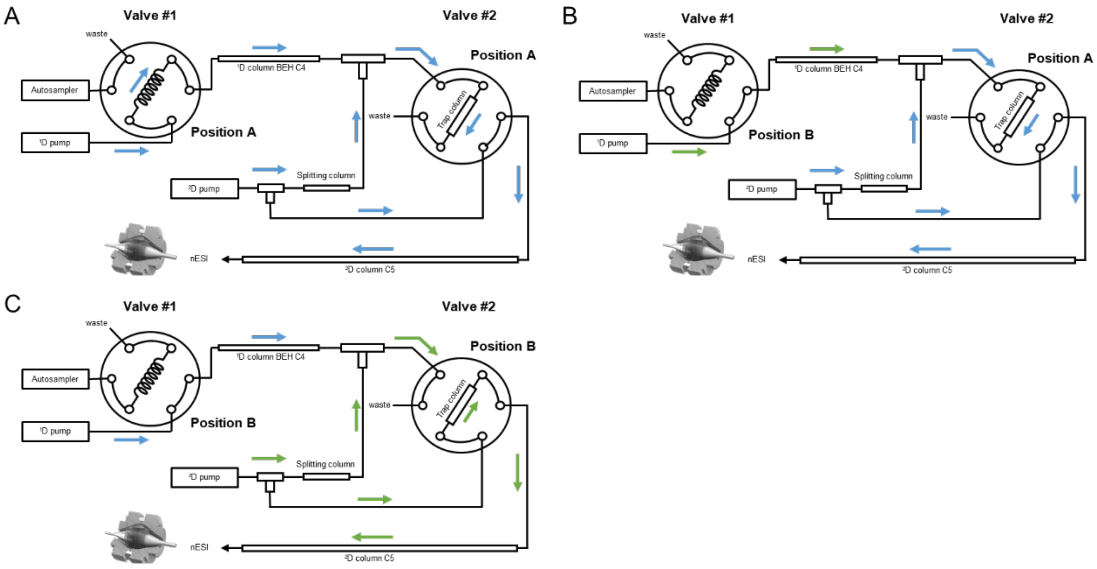


Figure 3-1. The schematic diagram of online comprehensive 2D-LC platform.

The setup of the online comprehensive 2D-LC. A 25-µL sample loop was installed on a valve in the autosampler. The 1D pump (high-pH mobile phase) was used to pumping the mobile phases for sample loading and first-dimension separation. The 2D pump (low-pH mobile phase) was used to provide dilution mobile phase to dilute the elution from the first-dimension separation, and to provide the gradient for second-dimension separation. A splitting column was used to regulate the pressure and flowrate of the SPE sample loading and first-dimension separation.



Gradient setup for ^{1}D:

Online Fraction	1	2	3	4	5	6	7	8	9	10
^{2}D MPB%	Flow through (0%)	0 - 40%	40 - 45%	45 - 50%	50 - 55%	55 - 60%	60 - 65%	65 - 70%	70 - 75%	75-100%

Figure 3-2. Stepwise operation of online 2D nano-LC system and the gradient setup.

In the 1st-dimension, we utilized a custom-packed 10 cm C4 capillary column (Waters BEH300, 3.5 µm, 300 Å, 75 µm I.D.) for high-pH RPLC separation (pH = 10). The eluate from the 1st-dimension was diluted online (1:10) and collected using a micro-trap SPE column (Waters BEH300, 3.5 µm, 300 Å, 20 mm × 150 µm). In the 2nd dimension, we utilized a 100 cm C5 column (Jupiter particles, 300 Å, 5 µm, 75 µm I.D.) for low-pH RPLC separation (pH = 2) that was directly coupled to an LTQ Orbitrap Velos Pro mass spectrometer for top-down MS analysis.

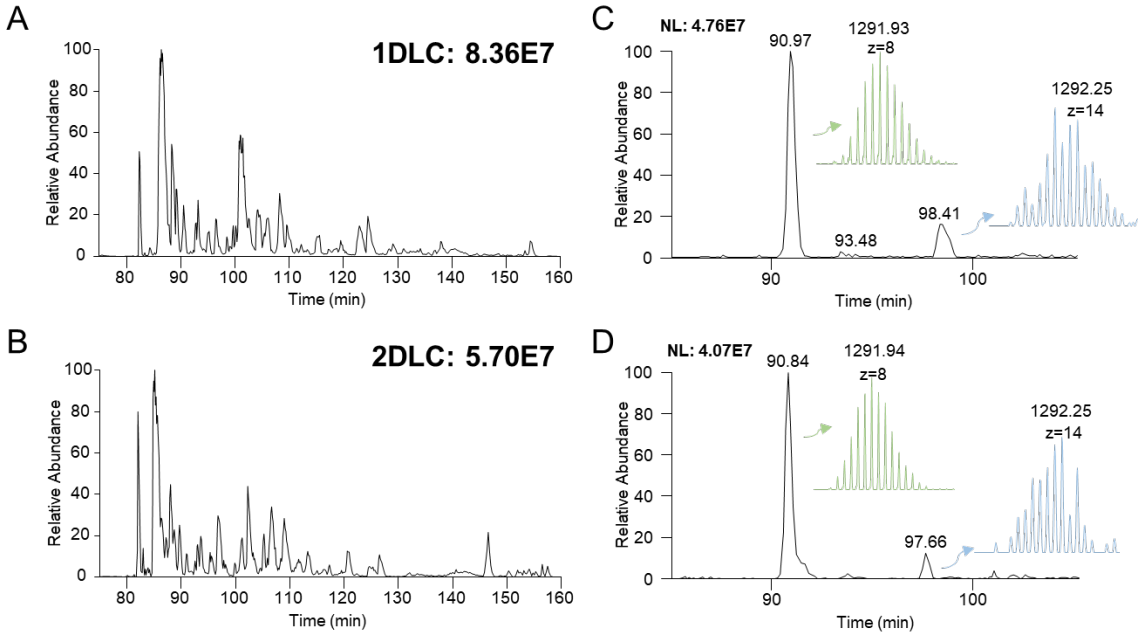


Figure 3-3. Validation of the dilution factor for 2DLC analysis.

The comparison of the base peak chromatograms of (A) 1DLC and (B) 2DLC analysis and the extracted ion chromatograms of two detected mass features from (C) 1DLC and (D) 2DLC analysis indicate that the eluate of the 1D separation can be efficiently diluted in a Tee with the dilution ratio of 1:10 (v/v).

To evaluate the trapping efficiency of the micro-trap column (*e.g.,* if the organic content was sufficiently diluted and pH was low enough for 2nd dimension RPLC separation), 5 µg of *E. coli* protein sample was injected onto the 1st-dimension column and

then directly eluted using 100% of high-pH mobile phase B to the trap column. Both flow-through and eluate were collected on the trap column during the trapping process. After trapping, the sample was separated using the 2nd dimension low-pH RPLC column for MS detection. The protein elution profile, protein intensity, and protein identification from this run were compared to a 1D low-pH RPLC separation of the same amount of *E. coli* protein sample (**Figure 3-3**). These results show that proteins can be efficiently eluted from the high-pH RPLC separation and trapped using the online micro-trap for the 2nd dimension low-pH RPLC separation. We further evaluated the reproducibility of the online 2D system with duplicated 2-fraction runs (1st-fraction is flow-through and 2nd fraction is eluate from 0% to 100% buffer B) (**Figure 3-4**). The two 2D-LC runs had similar elution profiles in both flow-through and elution experiments.

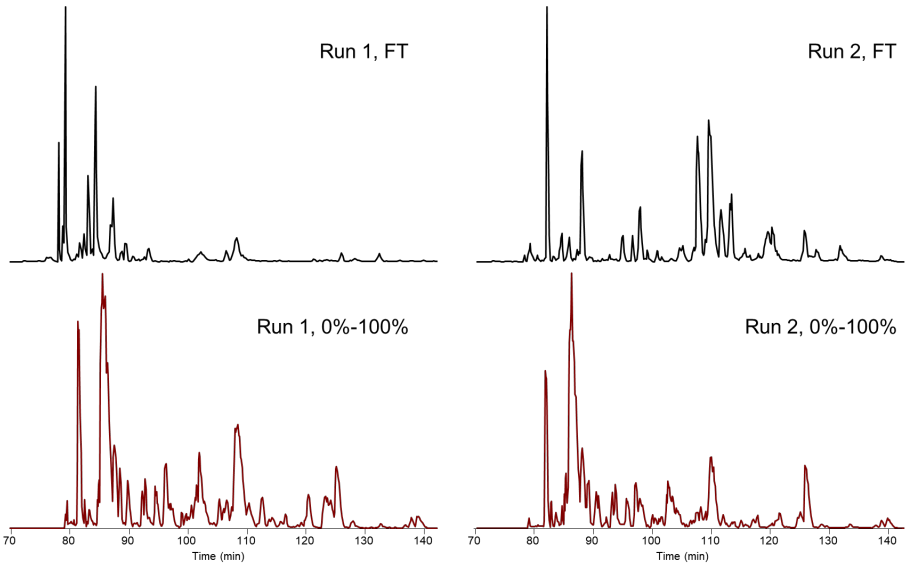


Figure 3-4. The online 2D chromatograms of 2-fraction runs.

We then optimized the 1st-dimension fraction collection steps. An initial test of 11 fractions (**Figure 3-5**) was evaluated with the following elution steps: 0% for 1st-fraction

and 0% to 10% for 2nd fraction followed by a 10% increase in the concentration of the elution buffer for each subsequent fraction. Analysis of intact *E. coli* cell lysate suggested that the majority of proteins were eluted in the range of 20% to 70% of mobile phase B of the 1st-dimension high-pH RPLC separation with the most mass features were detected in the fractions from 40% to 70% of mobile phase B.

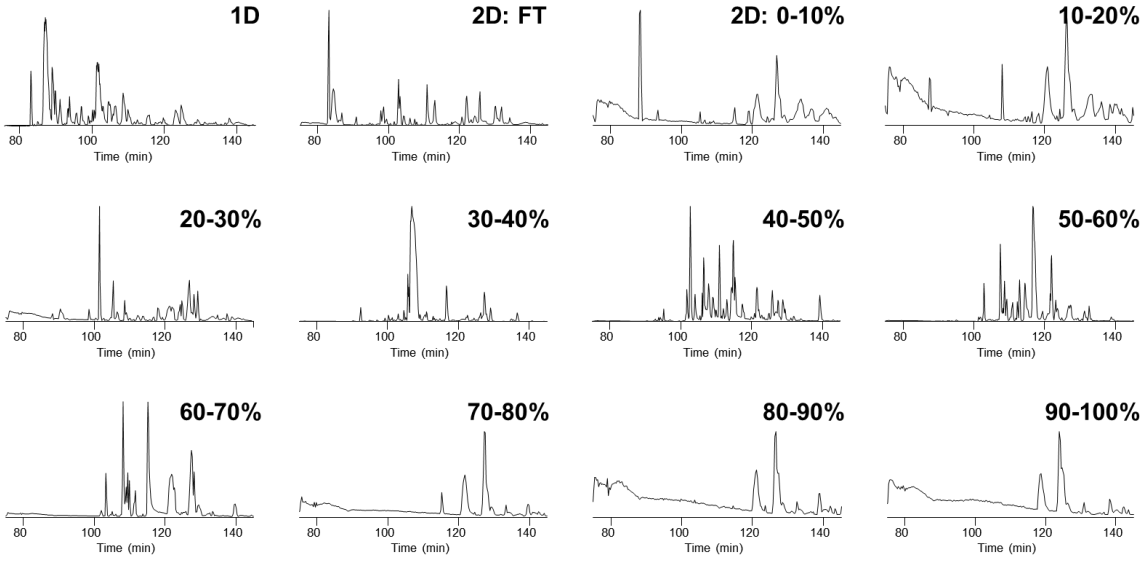


Figure 3-5. Evaluation of 1D gradient range for *E. coli* protein separation.

Different percentage ranges of mobile phase were applied to elute the samples from the 1D column. The eluates were separated and analyzed on the 2D column.

Based on these observations, we combined 0% – 40% into one online fraction, and 70% - 100% into another online fraction. The optimized a 10-fraction gradient setup for online 2D nano-LC was included in **Figure 3-2** which was then applied to the characterization of intact *E. coli* proteins. The base peak chromatograms (BPCs) of the 2nd-dimension low-pH RPLC-MS separation for each fraction are shown in **Figure 3-6**, demonstrating different elution profiles in different fractions. The BPSs of fractions after 60% MPB of 1st-dimension present similar profiles. In last three fraction, there were only

87 unique proteoforms identified. More precise optimization of gradient setup for different types of samples may be needed to better utilize the instrument time.

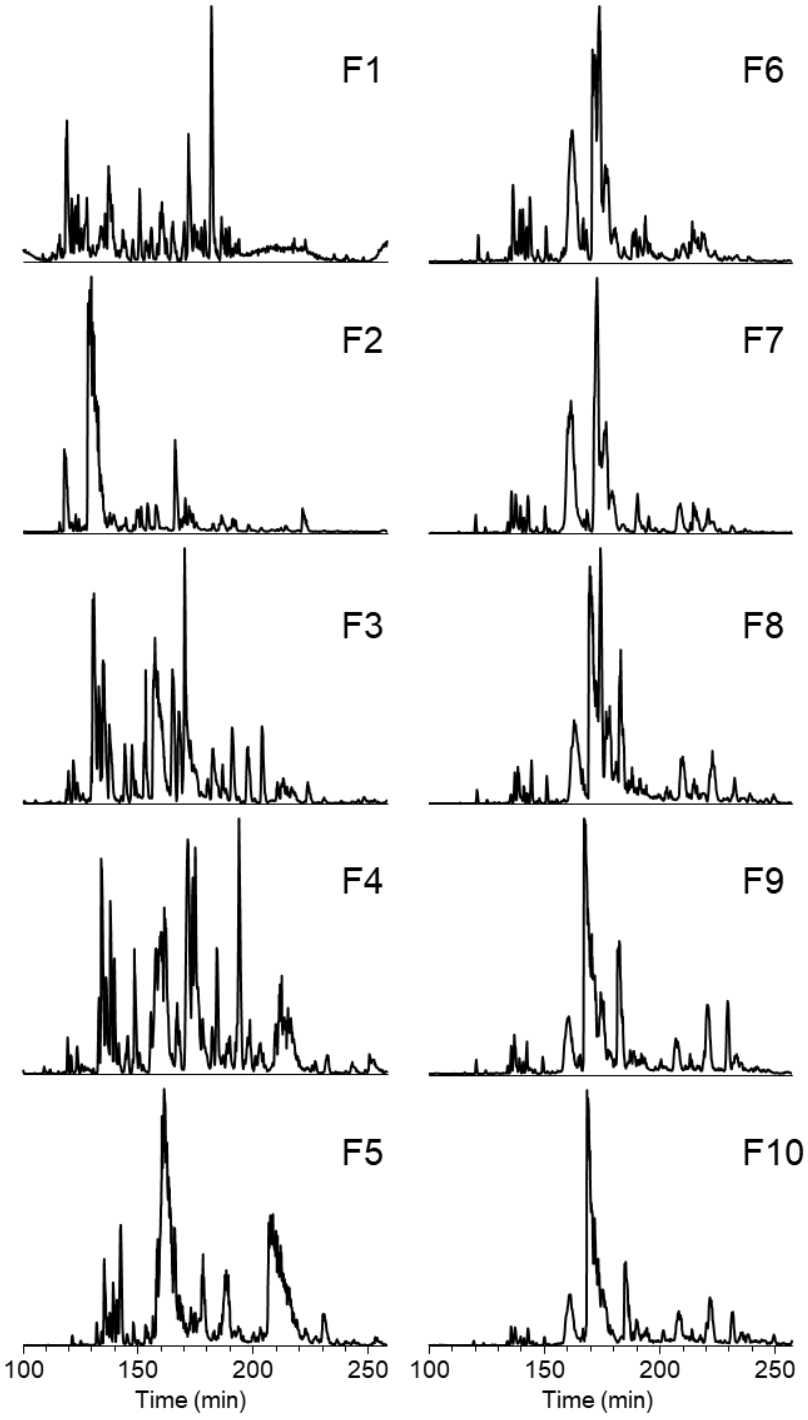


Figure 3-6. LC-MS base peak chromatograms (*m/z* 600-1100) of the online 2D-LC analysis with 10 online collected fractions.

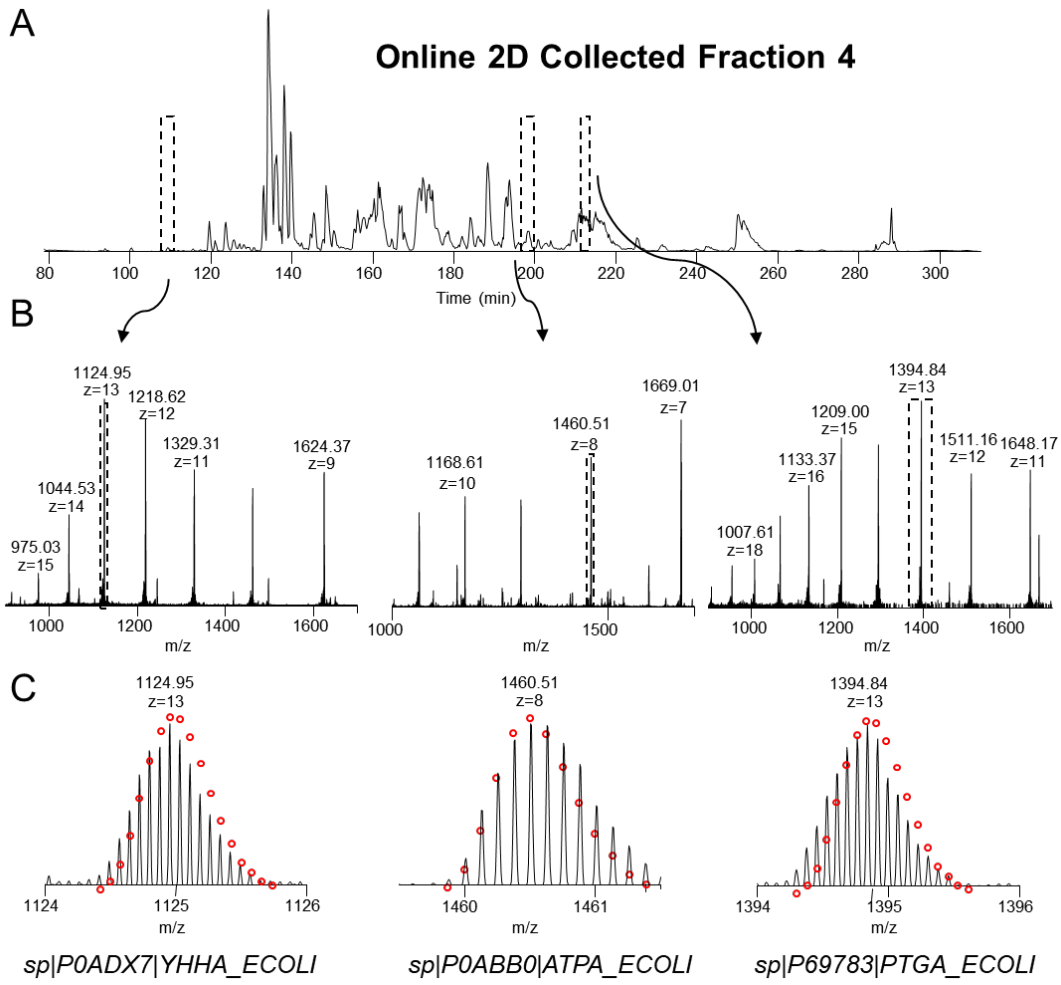


Figure 3-7. Online 2D pH RP/RPLC MS analysis of *E.coli* proteins.

(A) Base peak chromatogram of second dimension top-down analysis of online collected fraction 4. (B) Representative mass spectra of three proteins identified in fraction 4. (C) Overlay of observed isotopic distribution and theoretical isotopic distribution (red circles)**.**

Examples of three identified proteins with isotopic distributions are shown in **Figure 3-7**. To further evaluate the improvement of the online 2D-LC platform, we evaluated the 5-min 1D-LC gradient segment and the same retention time segment in each online 2D-LC collected fraction (**Figure 3-8**). Using 1D-LC, 19 mass features were detected in the 5-min separation window and 9 non-redundant proteoforms were identified.

Using online 2D-LC, we identified 24 unique proteoforms from 57 detected mass features in the 5-min segment in all the 10 online collected fractions. In total, 1,507 non-redundant proteoforms in 308 unique proteins were identified in 5 µg intact *E. coli* cell lysate with our online 2D separation, which is comparable with our previous reported results in 500 µg intact *E. coli* cell lysate using the offline 2D separation [17].

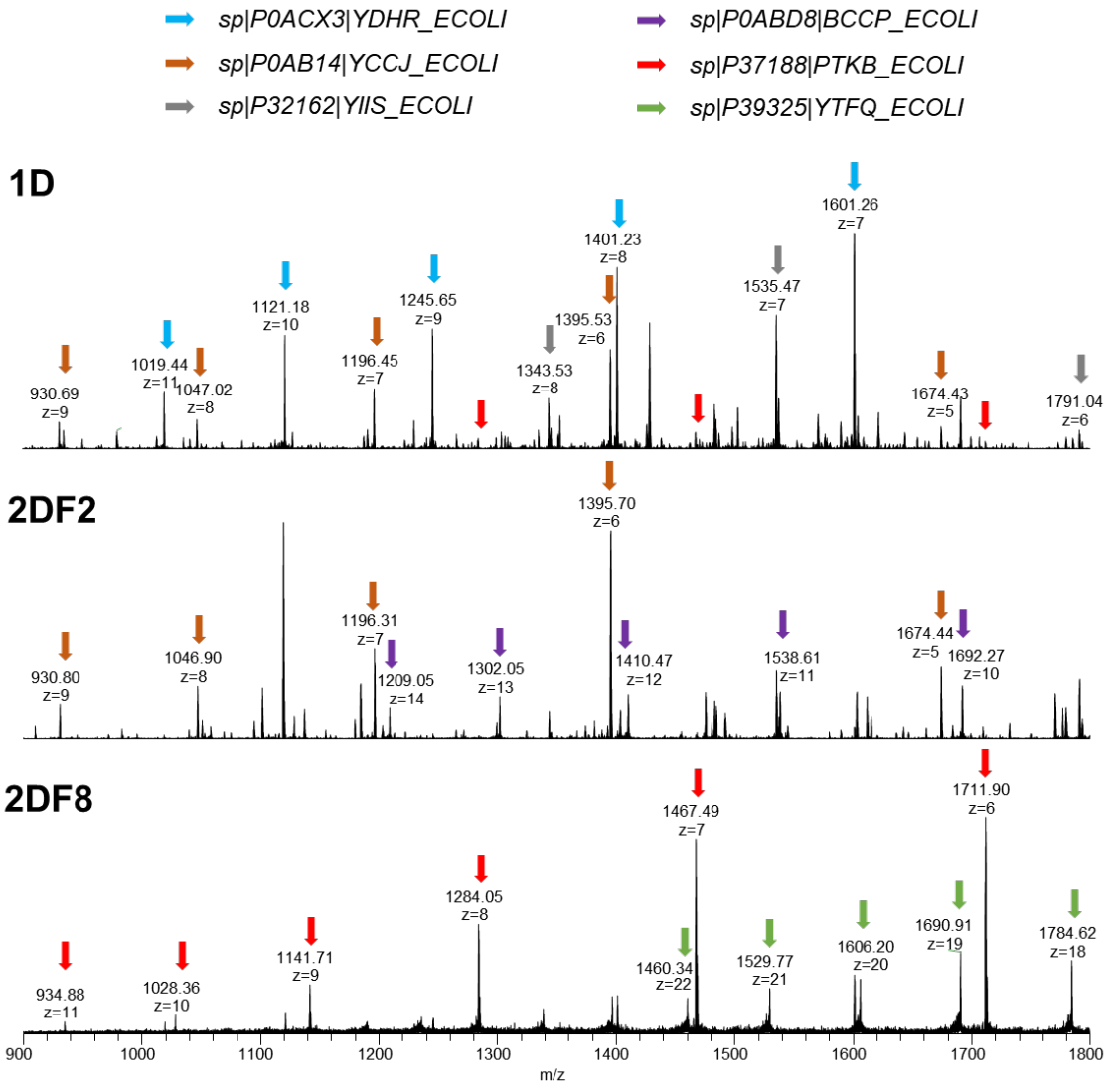


Figure 3-8. Five-minute segment from 1DLC and online 2DLC (RT = 190 min – 195 min).

We previously reported the identification of the apo-acyl carrier protein which is a unique protein working as a coenzyme in fatty acid and polyketide biosynthesis. Due to improved separation using the offline 2D pH RP/RPLC platform, low abundant proteoforms of the apo-acyl protein were characterized with the phosphopantetheine modification and its unique different oxidative forms of the thiol group on this modification. In this study, we were able to characterize most of the low abundant proteoforms of the apo-acyl protein using the online comprehensive 2D-LC platform (**Figure 3-9**).

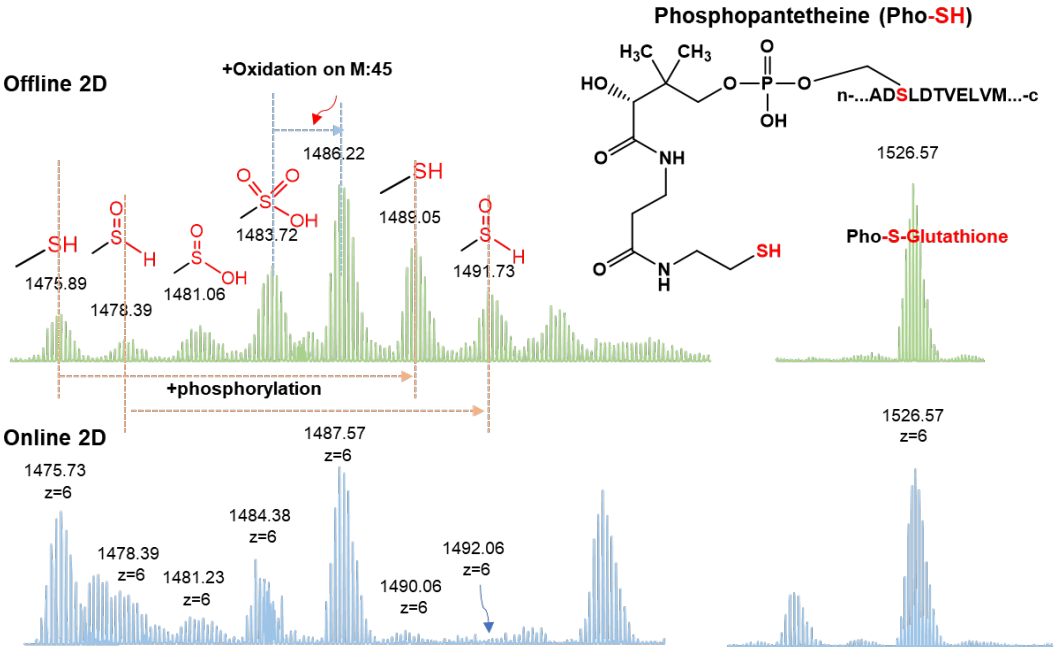


Figure 3-9. Proteoforms of the apo-acyl carrier protein identified using offline 2D and online 2D pH RP/RPLC-MS.

Despite the success of offline 2D high-pH/low-pH RPLC separations in bottom-up proteomics, there has been limited success for online 2D separation for bottom-up analyses. One major reason is that high-pH RPLC and low-pH RPLC are only partially orthogonal for peptide separations. Therefore, fraction concatenation is applied to more efficiently

utilize the separation space[106]. In top-down proteomics, as we discussed in previous offline studies, we observed good orthogonality between high-pH RPLC separation and low-pH RPLC separation. The improved orthogonality of the 2D platform allows efficient utilization of separation space without fraction concentration, which makes the online coupling of high-pH RPLC and low-pH RPLC feasible. We here evaluated the orthogonality between high-pH RPLC separation and low-pH RPLC separation with the developed online platform. For each "fraction", a heatmap was generated using the relative number of uniquely identified proteoforms in each bin (10-minute windows) (**Figure 3-10**). Our results suggest good orthogonality of the online 2D platform, which is comparable with the results from our previously reported offline 2D platform.

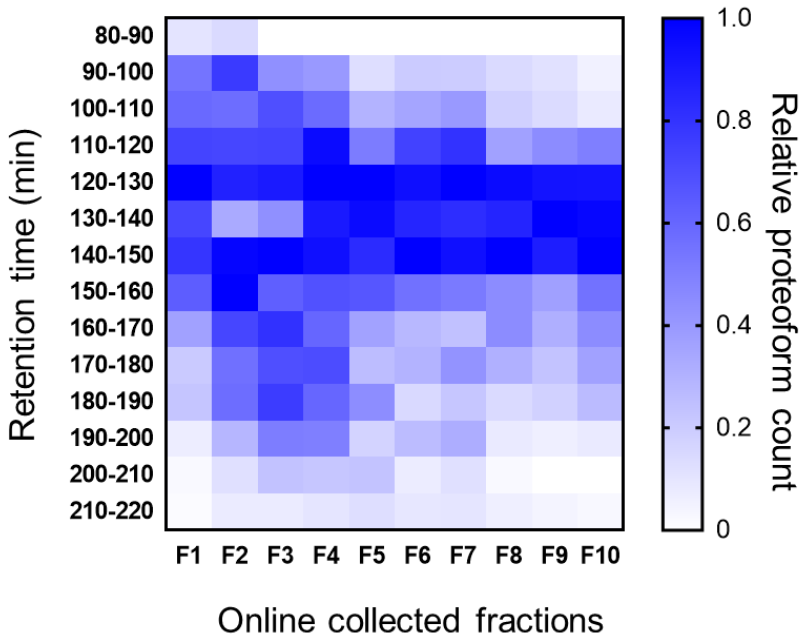


Figure 3-10. The proteoform elution patterns in high pH fractions.

Each column bin in the heatmap represents the relative count number of uniquely identified in the 10-min elution window in the low pH RPLC separation.

3.5 Conclusion

While offline fractionation is still the most commonly applied method, online 2D-LC platforms offer the advantage of less sample loss and higher sensitivity, which enables the analysis of small quantity samples. In addition, since the online 2D-LC system is automated and no sample processing between the two dimensions is involved, artificial modifications such as degradation can potentially be reduced. We here report the development of an online 2D ultra-high-pressure nano-LC system for high-pH and low-pH reversed phase separation for the top-down analysis of complex biological samples. With the online 2D system, small sample quantities could be analyzed to achieve comparable proteoform identification results with the offline 2D-LC using much higher starting sample amounts. Good orthogonality between 1st dimension high-pH RPLC and 2nd dimension low-pH RPLC were demonstrated using the online 2D system, which is consistent with our previously reports using the offline 2D system. The performance of the proposed on-line 2D system can be further optimized. For example, the incorporation of duel trap columns could reduce the MS dead time in between runs through the parallel trapping and elution process[107]. Additionally, high-performance and high frequency MS instrumentation can be adapted for improved proteoform identifications in relatively shorter 2nd dimension gradient with columns using smaller particle size. Also, 1st dimension RPLC separation conditions such as buffer composition and column temperature can be optimized to improve the elution of hydrophobic proteins, which can increase column life and avoid back pressure problems.

*Authors: Zhe Wang, Dahang Yu, Xiaowen Liu, Kenneth Smith, and Si Wu

Chapter 4 Top-down mass spectrometry analysis of human serum autoantibody antigen-binding fragments

4.1 Abstract

Detecting autoimmune disease at an early stage is crucial for effective treatment and disease management to slow disease progression and prevent irreversible organ damage. In many autoimmune diseases, disease-specific autoantibodies are produced by B cells in response to soluble autoantigens due to defects in B cell tolerance mechanisms. Autoantibodies accrue early in disease development, and several are so disease-specific they serve as classification criteria. In this study, we established a high-throughput, sensitive, intact serum autoantibody analysis platform based on the optimization of a one dimensional ultra-high-pressure liquid chromatography high-resolution top-down mass spectrometry platform (1D UPLC-TD-HRMS). Combined with our customized sequencing software, this approach has been successfully applied to a 12 standard monoclonal antibody Fab mixture, demonstrating the feasibility to separate and sequence intact antibodies with high sequence coverage and high sensitivity. We then applied the optimized platform to characterize serum autoantibody Fabs in a systemic lupus erythematosus (SLE) patient sample and compared it to healthy control samples. From this analysis, we show that the SLE sample has many dominant antibody Fab-related mass features unlike the healthy controls. To our knowledge, this is the first top-down demonstration of serum autoantibody pool analysis. Overall, our proposed approach holds great promise for discovering novel serum antibody biomarkers that are of interest for

diagnosis, prognosis, and tolerance induction, as well as improving our understanding of pathogenic autoimmune processes.

4.2 Introduction

Autoimmune diseases are a leading cause of death and disability in young minority women and collectively affecting more than 23.5 million Americans [108]. Many of the autoimmune diseases share similar symptoms, especially during the early stage of disease, which makes the diagnosis of autoimmune diseases extremely difficult[109]. Most autoimmune diseases are chronic conditions which can be controlled to varying extents by medication, but there is no permanent cure and these medications often leave the patient at increased risk of infection[110, 111]. Current researches increase the chances of more accurate diagnosis[46, 53, 54], however, many of the autoimmune diseases share similar serum biomarkers further contribute to an unreliable diagnosis[109]. Therefore, detecting systematic autoimmune diseases at an early stage is crucial for effective treatment and disease management to slow disease progression and prevent irreversible organ damage. However, this remains a significant clinical challenge due to the lack of unique biomarkers with both specificity and sensitivity.

Autoantibodies are a hallmark of many autoimmune diseases and can be present in serum years before clinical symptoms arise[112] and are occasionally present even in healthy individuals[113]. Current analysis approaches only measure the total concentrations of the autoantigen specific autoantibodies that are often polyclonal and may contain highly homologous clonal sequences[45, 46]. However, the presence of specific autoantibodies in patients with autoimmune diseases is of interest for diagnosis, prognosis,

drug targets, and for our understanding of various disease processes. DNA deep sequencing of the B cell antibody repertoire can be used to analyze humoral immune responses[114], but few of the detected sequences are represented in the circulating pool of serum immunoglobulins, and it is essentially impossible to determine which sequences are specific to an antigen of interest. To elucidate functionally relevant autoantibodies that mediate autoimmune responses, protein-level characterization of autoantibodies in the patient serum (*i.e.,* proteomics) is needed to precisely determine which of these autoantibody clones are predictive of autoimmune disease progression.

Mass spectrometry-based proteomics techniques have been used for the detection and characterization of serum monoclonal antibodies. Several bottom-up approaches have been developed to identify antigen-specific autoantibodies in serum [52, 53, 115]. These approaches often start with affinity purification of polyclonal autoantibodies from human serum with an autoantigen of interest. The purified antibodies are then digested with proteases such as trypsin to produce peptide fragments that are analyzed by LC-MS/MS. Identification of the peptide sequences corresponding to antibody fragments can be performed either with reference databases or through *de novo* sequencing. In a very recent report, Gordon's group demonstrated a possible "clonotypic sharing" by several shared peptide sequences in unrelated patients with SLE through bottom-up MS and *de novo* sequencing [46]. However, there are inherent challenges with bottom-up approaches for serum antibody analysis. Serum autoantibodies are likely to be highly homologous with very similar sequences from common V gene families. Bottom-up proteomics on serum autoantibodies, starting with digested peptides, will result in a pool of peptides with both shared and non-shared sequences. Even assuming 100% sequence coverage (which is

nearly impossible to generate with bottom-up approaches), without additional information, bottom-up MS is unable to identify the precise coordination of individual sequences for each IgG. Moreover, the affinity purification of autoantibodies often requires extensive optimization processes, large initial sample volumes, selects for the highest affinity antibodies, and cannot provide a "bird-eye's" view of the total autoantibody composition to all autoantigens in patient serum samples.

Top-down proteomics has unique advantages in analyzing proteoforms with sequence variations and post-translational modifications (PTMs) because it analyzes intact proteoforms rather than short peptides. Recent developments in MS instrumentation and protein separation have paved the way for proteome-wide analysis of complex[61, 116]. A top-down proteomics approach (*i.e.*, miRAMM) has been demonstrated for monitoring the light chain of a single monoclonal therapeutic IgG in spiked-in serum. Recently, the miRAMM was applied with the ultrahigh resolution MS (*i.e.*, 21T FTICR-MS) and nano-scale RPLC separation to analyze several spiked-in monoclonal antibodies in human serum offering the high mass accuracy and high sequence coverage[117]. However, because multiple autoantigens co-exist in autoimmune diseases, sera of autoimmune disease patients are very complex, likely containing at least hundreds of highly homologous monoclonal autoantibodies. Thus, miRAMM or similar approaches cannot be directly applied to analyze serum autoantibodies without significantly advancing the analytical capability to separate many highly homologous autoantibodies from the serum antibody background. In addition, advanced bioinformatics tools need to be developed to confidently sequence these autoantibodies.

With top-down proteomics, reversed phase liquid chromatography (RPLC) is the

most commonly applied high throughput separation approach that can be coupled directly online with MS. Similar to bottom-up MS, longer column and higher pressure pumps are used to improve the peak capacity of RPLC separation[18]. We have previously reported a single-dimension top-down proteomics platform analysis with home-built nanoflow columns and Waters nanoAcquity pumps (maximum pressure 10,000 psi, operation pressure 7,000 psi)[29, 74, 118]. This platform identified 563 intact proteins in *Salmonella* including 1,665 proteoforms generated by posttranslational modification (PTMs) at 5% false discovery rate (FDR), representing one of the largest single-dimension top-down datasets reported to date[29]. In this study, we optimized an automatic single-dimension RPLC platform through a custom-modified ultra-high pressure nano LC system (UPLC, maximum pressure 14,000 psi, operation pressure 10,000 psi) to improve the separation of highly homologous autoantibodies (*i.e.,* intact Fab's, light chains, and heavy chains of the Fab portion) in serum samples. This approach has been successfully applied to a 12 standard monoclonal antibody Fab mixture, demonstrating the feasibility to separate and sequence intact antibodies with high sequence coverage and high sensitivity. We then applied the optimized platform to characterize serum autoantibody Fabs in a systemic lupus erythematosus (SLE) patient sample compared to healthy control samples, showing that 80+ dominant antibody Fab related mass features are only observed in the SLE sample, which is the first top-down demonstration of a serum autoantibody pool analysis.

4.3 Materials and methods

4.3.1 Materials and Reagents

LC/MS CHROMASOLV® grade isopropanol (IPA), acetonitrile (ACN), and water were purchased from Sigma-Aldrich (St. Louis, MO). Pierce™ Trifluoroacetic Acid (TFA) and Bond-Breaker™ TCEP Solution were obtained from Thermo Scientific (Hanover Park, IL). The packing materials for packing C5 (Jupiter particles, 5 μm diameter, 300 Å pore size) was purchased from Phenomenex (Torrance, CA).

4.3.2 Human subjects

SLE and healthy control plasma samples were obtained in accordance with the Helsinki Declaration and were approved by the Institutional Review Board at the Oklahoma Medical Research Foundation. Blood was collected via venipuncture into ACD vacutainers (BD Biosciences, San Jose, CA), spun and the plasma was removed and stored at -20 ^{o}C until used. The plasma from this particular SLE patient contains ~0.5 mg/mL of anti-Sm, as well as smaller quantities of anti-nRNP, anti-Ro and anti-La (data not shown).

4.3.3 Monoclonal antibodies

Fully human, full-length monoclonal antibodies were produced by the Human antibody core facility at the Oklahoma Medical Research Foundation as previously reported[119]. These antibodies were obtained from single cell-sorted antibody secreting cells or naïve B cells and are expressed with human IgG1 heavy chains and kappa or lambda light chains.

4.3.4 Sample preparation

Escherichia coli cell lysate proteins were obtained from the BL21 strain grown in house. Cell lysate was obtained by bead-beating with zirconia silica beads[72]. Aliquots of protein solutions were stored at -80 ^{o}C until further use. Protein A/agarose beads were used for the antibody purification from serum samples. In detail, protein A beads were incubated with diluted plasma (3-5 ml total plasma, diluted 1:5 in 1 × PBS) at 4 ^{o}C overnight. The antibodies were eluted with 0.1 M glycine-HCl (pH 2.7) and concentrated using Amicon concentrators (30 kDa cutoff). The purified IgG fractions, as well as standard monoclonal antibodies were then digested using insoluble papain suspension with incubation at 37 ^{o}C for 4 hours. After the incubation, protein A/agarose beads were used for the removal of Fc fragments from the antibody digests with the same conditions as previous described. Fab fragments from either plasma or monoclonal antibodies were then concentrated to > 1 mg/mL total protein concentration with Amicon protein concentrators (10 kDa cutoff). The Fab fragments were reduced by reacting with 1 μL of 0.5 M TCEP prior to the UPLC-TD-HRMS analysis.

4.3.5 UPLC-TD-HRMS

An in-house packed nano-flow capillary RPLC-C_5 column (5 μm, 75μm × 100 cm) was used on a custom modified UPLC (maximum pressure 14,000 psi) system. The mobile phase A was 0.01% TFA, 0.585% HAc, 2.5% IPA and 5% ACN in water, and the mobile phase B was 0.01% TFA, 0.585% HAc, 45% IPA and 45% ACN in water. Ten micrograms of *E. coli* lysate proteins, 8 micrograms of 12-antibody-Fab mixture, and 8 micrograms of purified human serum antibody Fab samples were loaded on the column for the top-down

MS analysis individually. A gradient from 10% to 70% of mobile phase B over 70 minutes or 280 minutes at a flow rate of 400 nL/min was applied for the separation and the column was regenerated by flushing with 90% of mobile phase B for 10 minutes and equilibrated to 100% of mobile phase A. The nano-LC column was directly coupled to an LTQ Orbitrap Velos Pro mass spectrometer (ThermoFisher Scientific, Bremen, Germany) for online MS/MS analysis with a custom designed nano-ESI interface under positive mode. The electrospray voltage was set to 2.6 kV and the heated inlet capillary temperature was optimized to 250 °C. MS data were collected at the resolving power setting of 100,000 (at *m/z* 400) with two micro scans. MS/MS acquisition was performed by selecting the top five most abundant precursor ions in the full MS scan using collision induced dissociation (CID) with the normalized energy of 35 %, higher-energy collisional dissociation (HCD), or electron-transfer dissociation (ETD). The MS/MS data were obtained at a resolving power setting of 60,000 (at *m/z* 400) with one micro scan. Ions with less than 4 charges were rejected for the selection of MS/MS scans. The maximum injection time for a full mass scan and a MS/MS scan were set to 1000 ms. and 500 ms., respectively. The AGC target was set at 1×10^6 for full mass scans, and 5×10^5 for MS/MS scans. All of the data were collected with the Xcalibur 3.0 software (Thermo Fisher Scientific, Bremen, Germany).

4.3.6 Data analysis

The MS raw data were converted to centroid mzXML files with msconvert (a tool in ProteoWizard[104] and deconvoluted with MS-Deconv[120]. A constant region sequence database (55 heavy, 5 kappa, and 13 lambda sequences) were downloaded from

the IMGT database[121]. The deconvoluted data were searched against the sequence database separately using TopPIC [105, 122], in which the error tolerances for precursor and fragment masses were 15 ppm and at most 2 unknown mass shifts were allowed in a proteoform spectrum-match. Other parameter settings in TopPIC can be found in **Table 1**.

Table 4-1. Parameter settings of TopPIC in the analysis of the top-down MS/MS data of human serum samples

Parameter	**Value**
Fragmentation method	FILE
Fixed modifications	None
N-terminal forms of proteins	NONE, NME, NME+ACETYLATION
Using a decoy database	No
Error tolerance	15 ppm
Maximum number of unexpected mass shifts in a proteoform spectrum-match	2
Spectrum level cutoff type	E-value
Spectrum level cutoff value	0.01
Number of combined spectra	1
E-value computation method	Lookup table

The filtered PrSMs were evaluated and filtered based on fragmentation patterns and matched signature sequence tags. The identifications are manually evaluated using ProSight Lite[123]. The detected mass features were deconvoluted using Informed Proteomics[124] and evaluated manually.

4.4 Results

4.4.1 Optimization of the UPLC-TD-HRMS platform

(A)

Peak	a	b	c	d	e
Peak width (base)	0.78 min	1.34 min	1.82 min	1.91 min	1.22 min
Average width	1.414 min				
Gradient time	280 min				
Peak capacity	200				

(B)

a	b	c	d	e
0.80 min	1.3 min	0.8 min	1.0 min	1.5 min
1.08 min				
70 min				
66				

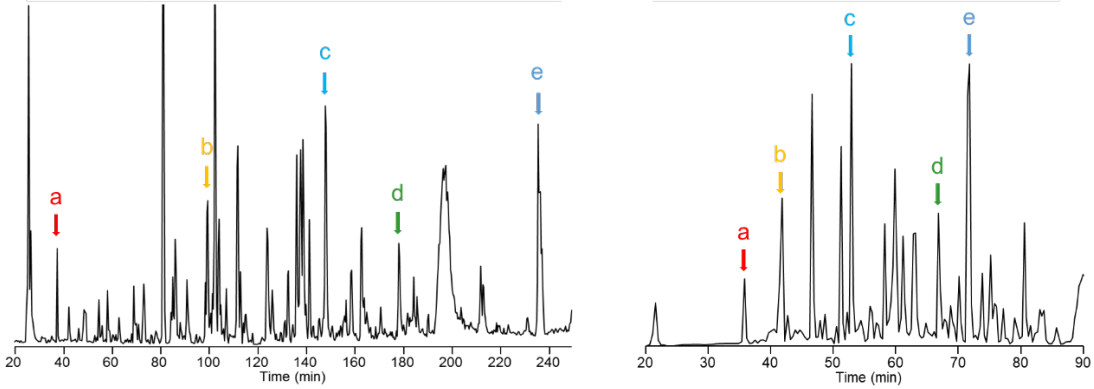


Figure 4-1. Peak capacities of the LC-MS run of E.coli lysate proteins with different gradient length.

(A) 280 minutes, and (B) 70 minutes. a, b, c, d, and e represent the randomly chosen peaks used for the calculation of the peak capacity.

The 1D UPLC-TD-HRMS was developed and optimized using a Waters NanoAcquity HPLC system with the maximum pressure of 10,000 psi in our previous work[29, 73, 74, 118, 125, 126]. However, the routine operational pressure of the system was limited to 7,000 psi due to commercial system limitations and automation. The operational pressure often changes with the organic components in the elution buffer, and the run would be interrupted when the pressure reached the maximum pressure. Recently, Shen *et al.* demonstrated high peak capacities in long-column RPLC separation at ultra-high pressure operation pressure (maximum operation pressure at 14,000 psi) using a manually operated constant pressure syringe pumps with customized gradient mixers[18].

To improve the throughput for long-column RPLC separation at higher operation pressure limits, we modified a commercially available normal-flow system (*e.g.,* Thermo Accela pumps, maximum pressure 14,000 psi) through the customized splitting system to establish an automatic UPLC system that can be routinely operated at a pressure higher than 10,000 psi for long-column nano-LC separation (*e.g.,* operation flow rate between 200 nL/min to 400 nL/min).

The elution gradient of the customized UPLC system was optimized using a custom packed C5 long column (*e.g.,* C5, 100 cm length, 360 µm o.d., 75 µm i.d.). Ten micrograms of *E. coli* lysate proteins were loaded on the column and an elution gradient from 10 % to 70% of mobile phase B was applied over 70 minutes and 280 minutes separately. The peak capacities with different gradient times were calculated by comparing the base peak widths of five randomly selected proteins from the LC/MS runs (**Figure 4-1**). The average base peak widths were 1.08 minutes for a 70 minute-gradient and 1.41 minutes for a 280 minute-gradient, respectively. The peak capacity of a 70-minute run was calculated as 66, and the peak capacity of a 280-min run was 200. Our results suggested that the peak widths did not increase significantly with the gradient length was increased. With the ultra-high pressure applied, the improved resolution of the separation from longer columns can partially overcome the resolution loss from the diffusion with the longer separation time. Based on the results, we here chose 280 minutes as the gradient time (10% - 70% of mobile phase B) for the separation of 12-Fab mixture and Fab fragments enriched from human serum samples.

4.4.2 UPLC-TD-HRMS analysis of a 12-antibody mixture

To evaluate the capacity of the separation and identification of intact antibody Fabs using the optimized UPLC-TD-HRMS platform, we papain-digested 12 fully human monoclonal antibodies and enriched the Fab fragments using protein A agarose beads as described above. These 12 Fabs were mixed in equal quantities and eight micrograms of the 12-Fab mixture was reduced by TCEP and loaded on the column for top-down MS analysis.

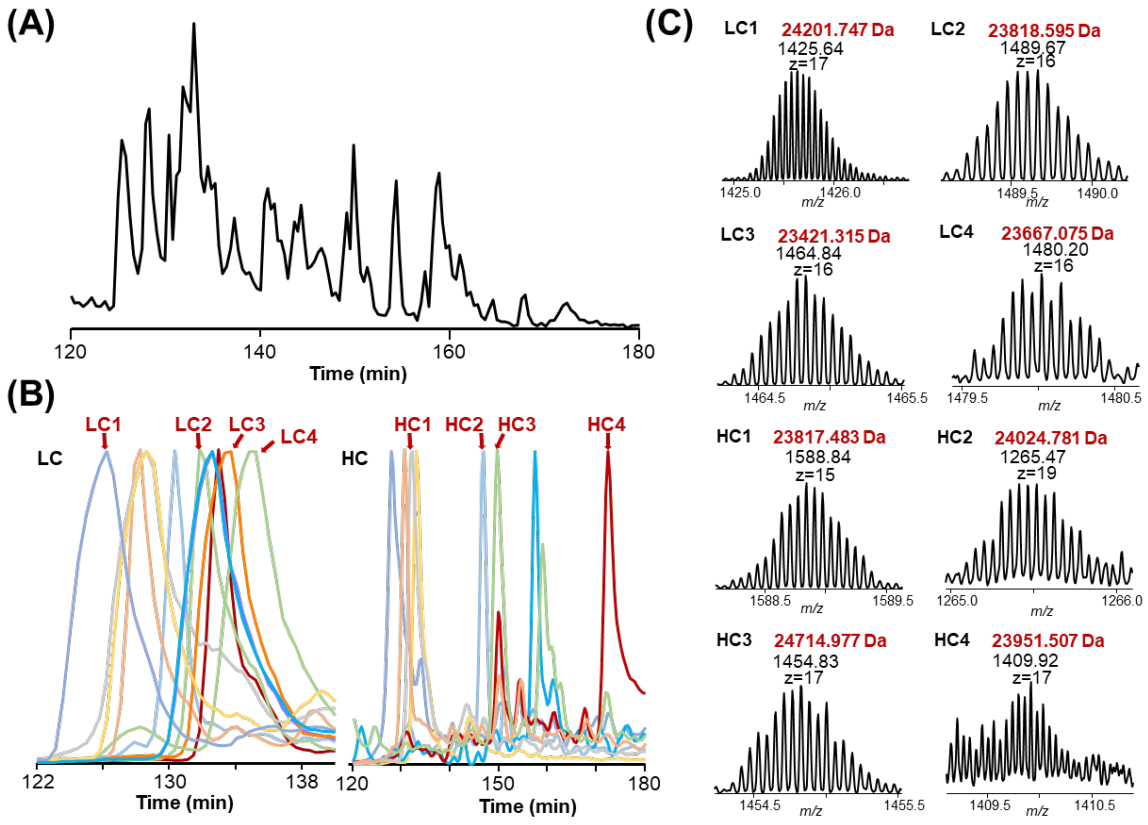


Figure 4-2. LC-MS of 12 standard antibody mixture.

(A) Base peak chromatogram of the LC-MS. (B) Extracted ion chromatograms of light chains and heavy chains from 4 standard antibodies. (C) Mass spectra of the 4 light chains and 4 heavy chains.

Figure 4-2A shows the base peak chromatogram of the separation of the 12-Fab mixture as well as the extracted ion chromatograms (**Figure 4-1B**) of the light chains and heavy chains for each Fab fragment. Overall, the Fab light chains eluted earlier (between 27% and 30% B) compared with the Fab heavy chains (between 27% and 38% B). Combined with high-resolution MS spectra and online MS/MS spectra, most of the standard antibody Fabs can be confidently assigned. Representative high-resolution MS spectra were demonstrated for 4 light chains and 4 heavy chains (**Figure 4-1C**). Our analysis demonstrated the feasibility of separating reduced intact Fab fragments in complex samples such as serum autoantibodies.

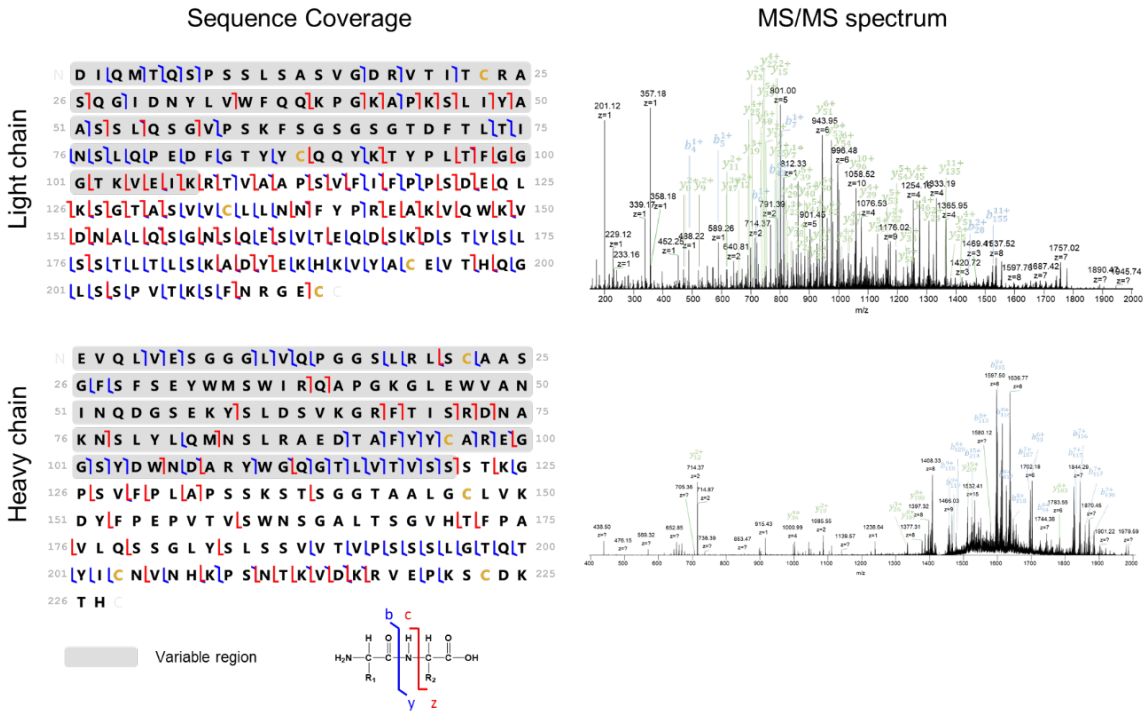


Figure 4-3. The identifications of the Fab light chain and Fab heavy chain of an antibody. Different fragmentation methods (e.g. HCD and ETD) were used to improve the sequence coverage. Examples of the MS/MS spectra of Fab light chain and Fab heavy chain were shown in the figure.

Different dissociation methods available on commercial orbitrap instruments were applied to the online fragmentation of proteins ions, including collisional induced dissociation (CID), higher-energy collisional dissociation (HCD), and electron transfer dissociation (ETD). The combination of different dissociation methods can significantly improve the sequence coverage of reduced Fab fragments. The representative Fab fragment was well characterized with 51% residue cleavages (67% for light chain and 40% for Fab heavy chain) (**Figure 4-3**). All of the MS/MS spectra used for Fab identification were manually checked and confirmed.

Analysis of Fabs with proper heavy and light chain pairing is crucial for antigen-binding studies of antibodies and this cannot be achieved with reduced samples. Therefore, the intact Fab fragment mixture (no reduction) was also analyzed using the UPLC-TD-HRMS platform. We have optimized our current Velos Orbitrap Pro's performance and were able to partially resolve intact non-reduced Fabs (e.g., M.W. ~48 kDa). The results (**Figure 4-4**) indicated that intact non-reduced Fabs can be efficiently separated and analyzed using our developed UPLC-TD-HRMS platform. The detected masses from the non-reduced samples were used to pair the light chain and heavy chain masses detected from the reduced samples (**Figure 4-4B**). Our results demonstrated that the developed UPLC-TD-HRMS platform is capable of separating and characterizing antibody Fab mixtures (both reduced and non-reduced) with high similarities, which can be applied to analyze enriched Fabs from human serum samples.

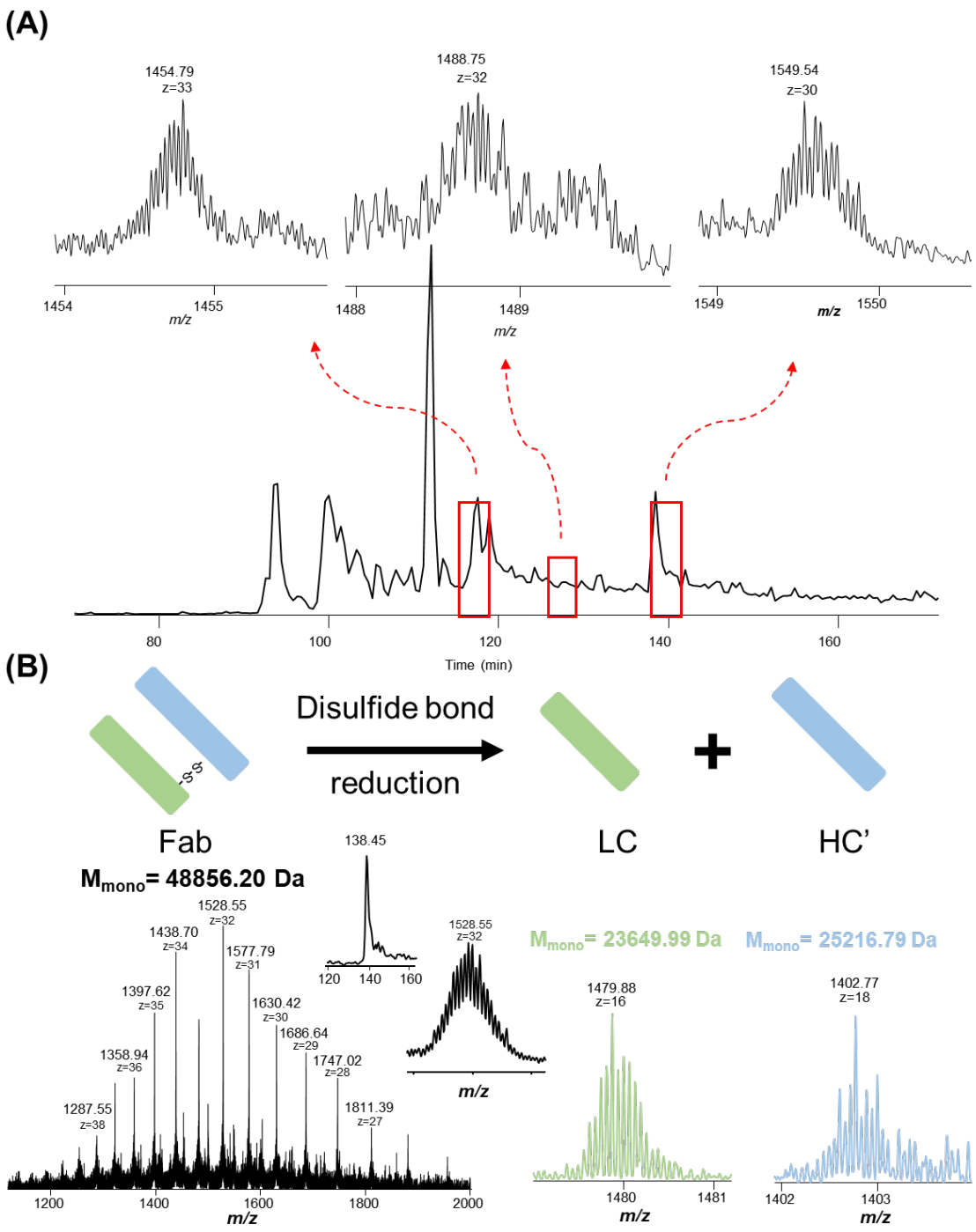


Figure 4-4. Analysis of reduced and intact Fab fragments.

(A) Base peak chromatogram of one LC-MS/MS run of the intact Fab mixture. Two examples of the detected mass spectra. (B) One example of light chain and heavy chain pairing. The mass spectrum of the intact Fab was from the LC-MS/MS run of the 12 intact Fab mixture. The detected mass was compared with all the detected masses from the LC-MS/MS run of the 12 reduced Fab mixture to obtain the LC/HC pairing.

4.4.3 UPLC-TD-HRMS analysis of monoclonal antibodies in human serum

With the optimized platform, we analyzed several human serum samples and characterized higher abundance antibodies that we suspect are autoantibodies in the serum with the ability to determine the light chains and heavy chains. Three serum samples were obtained from a SLE patient at different time points and the control samples were obtained from two healthy control individuals (autoantibody negative). All of the serum samples were purified using Protein A beads to enrich the antibodies from serum. After the enrichment, the samples were papain-digested and Fc portions were removed by Protein A beads. The purified Fab mixtures were reduced with TCEP and analyzed using the 1D UPLC-TD-HRMS platform. We first evaluated the summed MS spectra among different samples (**Figure 4-5A**).

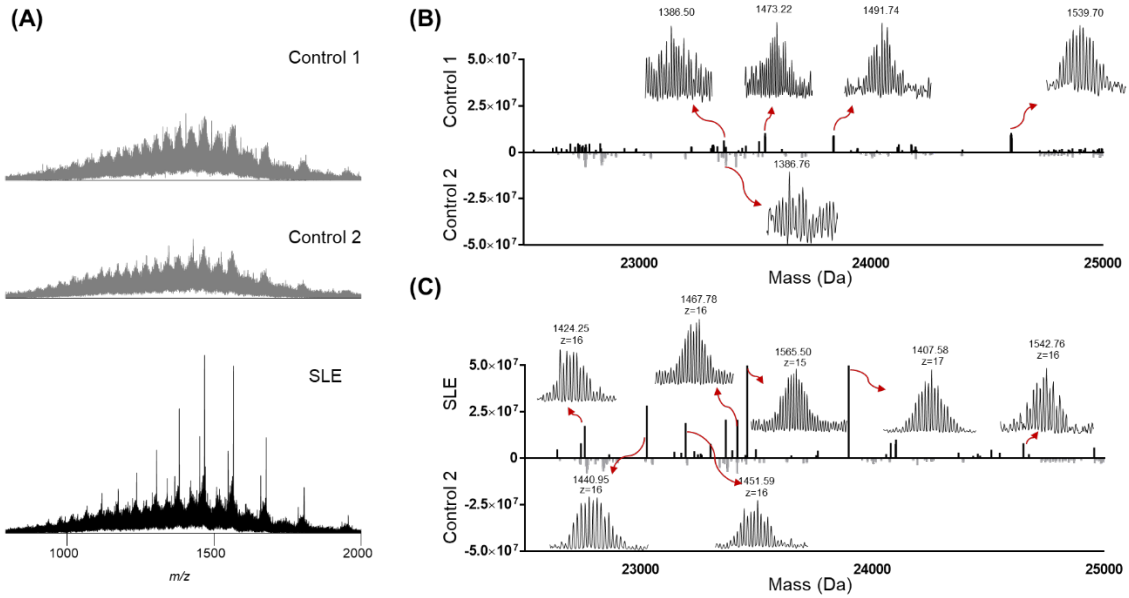


Figure 4-5. LC-MS analysis of SLE serum sample and healthy control samples.

(A) Summed mass spectra of control samples and the SLE serum sample. Deconvoluted mass features from 22,500 Da to 25,000 Da from different samples. Comparison between (B) two control samples, (C) SLE serum and one of the control samples. Some of the mass spectra of the deconvoluted mass features were also shown above.

For control samples, the summed spectra showed a wide range of unresolved normal distributed peaks similar to previous reports. For the SLE sample, several mass features were observed, indicating possible presence of autoantibodies. The UPLC-MS datasets were then deconvoluted to generate a "bird's-eye" view on the total autoantibody composition in patient serum samples. **Figure 4-5B and 4-5C** showed the deconvoluted mass features between 22,500 Da to 25,000 Da, representing the molecular weight range of both Fab heavy chains and Fab light chains. In the control samples, most of the deconvoluted mass features have low intensities and are not well-resolved (*i.e.,* less than 1E5 with the S/N less than 2). In control sample 1 (**Figure 4-5B**), we observed 4 mass features with relatively good S/N ratios. We manually averaged the related scans for these mass features, and all of them have good isotopic distributions that can be putative Fab chains. These putative Fab chains are less likely from autoantibodies because these control serum samples tested negative against all known autoantigens, but they are likely from ongoing immune responses. Moreover, the measured intensities of these putative Fab chains are relatively low (*i.e.,* less than 2E5). We further compared one of the SLE serum samples with one of the control samples (**Figure 4-5C**). In the SLE serum sample, 40 mass features were confidently detected with the total intensity larger than 1E6. All of these mass features are manually evaluated to ensure that they stood out from the serum antibody background (*i.e.,* S/N larger than 5).

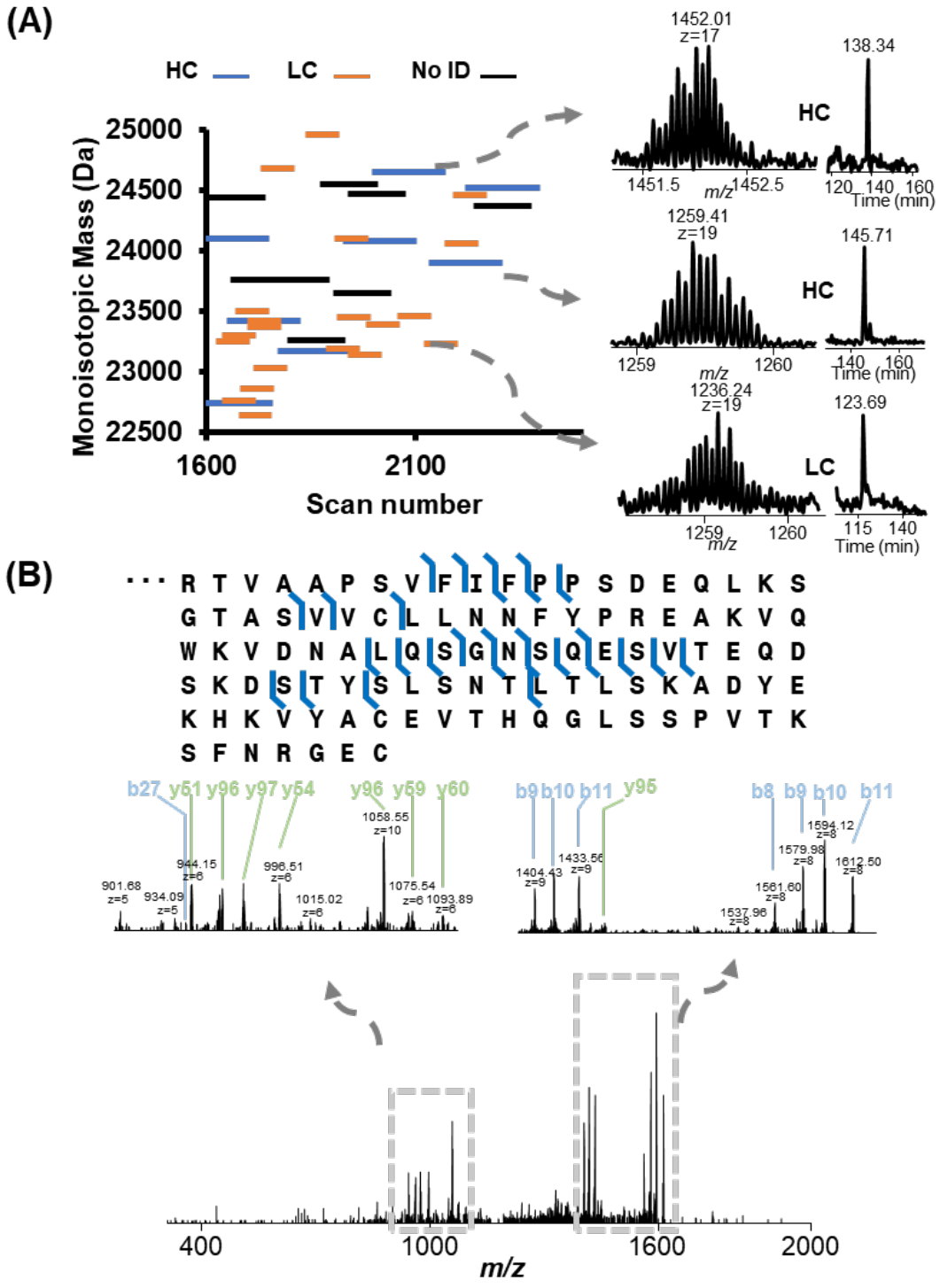


Figure 4-6. Identification of light chains and heavy chains of the IgGs in SLE Patient Serum.

(A) Plotting of the distribution of the detected mass features in one of the SLE patient serum samples. (B) MS/MS identification of one kappa 1 immunoglobulin light chain.

The detected mass features were deconvoluted and plotted against the LC elution time (**Figure 4-6A)**. Mass spectra and EICs of two representative detected mass features are also shown in the figure. Baseline resolved monoisotopic distribution of the detected mass feature was achieved and the EIC showed the high resolution of the UPLC-TD-HRMS platform even with extreme complicated and similar background serum antibodies. In order to identify the detected mass features, we first used our customized sequencing software to perform a database search, and 20 PrSMs were confidently identified as either light chains or heavy chains of intact Fabs. After manually checking the MS/MS spectra, we noticed a common pattern in all detected light chain MS/MS spectra. The sequence tag, VFIFPP, was confidently identified in all MS/MS spectra of the light chains, which was then applied as the criteria of the mass features being light chains. One of the identified light chain from the SLE patient serum sample was shown in **Figure 4-6B**. The constant region of the light chain was well characterized with fully cleavage site coverage of the desired region (VFIFPP). Thus, in the filtered PrSMs, a mass feature was assigned as a light chain if more than two product ions of VFIFPP were observed. In addition, we manually analyzed the identified CID MS/MS spectra and confirmed 9 of them as intact Fab light chains in the SLE serum sample. We further performed the UPLC-TD-HRMS analysis on two additional SLE serum samples from the same patient. Overall, we confidently detected 80+ unique mass features in the putative Fab range in samples from 3 different years. The collected MS/MS raw data were processed with TopPIC and IMGT database[121] was used to search the data. Manual evaluation was also performed to ensure the quality of the identifications. In total, 47 unique light chains and 16 unique heavy chains were identified in the samples from one SLE patient collected over three successive

years. Both biological and technical duplicates were also performed to evaluate the run-to-run reproducibility. The biological duplicate samples of SLE serum samples collected in 3 different years were obtained by performing the same Fab produce process described above and analyzed by the developed platform. All of the detected mass features were evaluated manually. The detected mass features from each run were then compared showing in **Figure 4-7A**. Technical duplicates were performed as well. **Figure 4-7B** shows the overlap of the detected mass features from two runs of year 3 sample. From the result, the platform was proven to be highly reproducible between technical duplicates where run 1 and run 2 of year 3 sample shared 39 detected mass features out of 47 and 43 detected mass features, respectively. Biological duplicates of the 3 samples also showed good reproducibility. However, the overlap between the biological duplicates of year 1 sample showed small overlap. One of the possible reasons is the difference brought in during the sample process. Overall, the overlap between biological duplicates was over 50%.

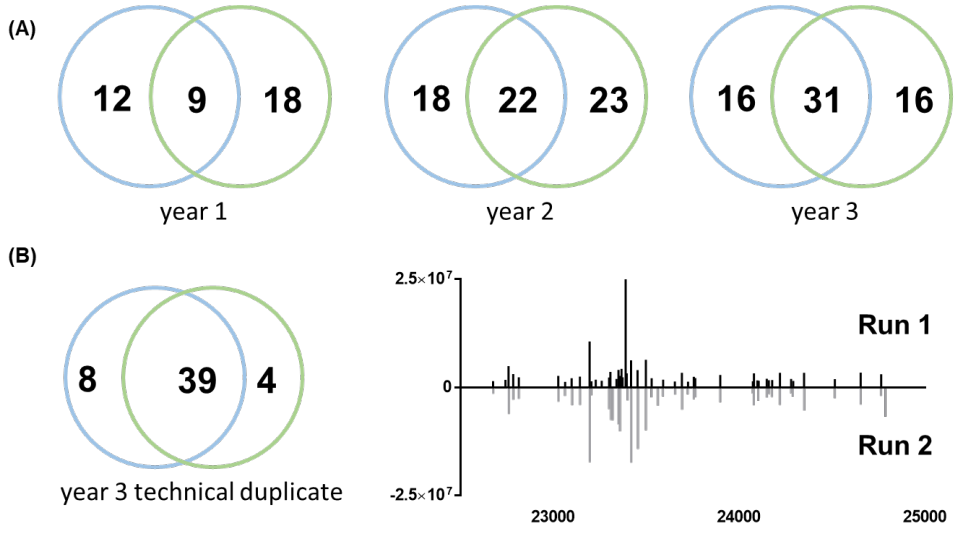


Figure 4-7. Run-to-run reproducibility of the developed platform.

(A) The overlaps between biological duplicates of samples collected in different years are shown above. (B) The technical duplicate of the sample from year 3 is also compared showing overlapping of 39 detected mass features with small difference.

4.5 Discussion

We developed and optimized a single dimensional UPLC-Top-down-High resolution mass spectrometry platform (UPLC-TD-HRMS) for the separation and characterization of autoantibodies in human serum. The UPLC-TD-HRMS platform provided sufficient separation of complicated antibody Fab mixtures and comprehensive characterization of Fab for light chain/heavy chain classification with high mass accuracy, high run-to-run reproducibility and in a high-throughput manner. To achieve better LC and mass resolution on our accessible instruments, the Fab fragments were reduced. However, due to the reduction of disulfide bonds, the light chain and heavy chain pairing information cannot be preserved with one LC-MS/MS analysis, but it is possible to regain this information by LC-MS/MS analysis of intact Fab fragments.

Using our platform, we were able to identify a total of47 light chains and 16 heavy chains from SLE patient serum. To our knowledge, this work is the first, to date, top-down MS demonstration of the human serum autoantibody pool analysis enabling the classification of light chains and heavy chains, which gives us a 'bird's-eye' view of the complexity of human serum autoantibodies. Our results showed the potential of using high resolution separation methods coupled to high resolution and highly sensitive mass spectrometry detection to help the understanding of human serum autoantibodies. Future studies can be done to improve the LC separation resolution of larger proteins, such as multidimensional separation techniques[127]. We note that currently selected patient samples are from a patient with advanced disease. More sensitive separation approaches such as capillary electrophoresis[128] and targeted approaches can be incorporated for early stage autoantibody detections. High-end mass spectrometers (*i.e.*, 21T FTICR-MS)

with different fragmentation techniques, such as front-end electron transfer dissociation[129] and Ultraviolet Photodissociation (UVPD)[130], can also be applied to increase the mass detection range and mass accuracy.

One of the challenges of serum autoantibody sequencing and characterization is the complexity of human antibody repertoire. The DNA sequencing of B-cells from particular patient can be performed to enlarge the existing database for more identifications of antibodies in human serum samples. In addition, we also observed some of the MS/MS spectra with good quality fragmentation were not identified which might be another result of the incomplete database. Another reason may relate to the sequencing software. In future works, the optimization of the sequencing software can also be done to expand the identifications. Our results, overall, demonstrated the ability of our platform performing top-down MS analysis on complicated human serum samples and detecting antibodies developed in patient serum over years, which gives it the possibility to monitor the development of antibodies during autoimmune disease progression.

*The materials in Chapter 4 are adapted from

Wang, Z., Liu, X., Muther, J., James, J.A., Smith, K. and Wu, S., 2019. Top-down Mass Spectrometry Analysis of Human Serum Autoantibody Antigen-Binding Fragments. *Scientific reports*, 9(1), p.2345.

Chapter 5 Optimization of a subzero temperature LC separation in hydrogen/deuterium exchange mass spectrometry for conformational epitope mapping

5.1 Abstract

Hydrogen deuterium exchange coupled with mass spectrometry (HDX-MS) provides a powerful and desirable protein foot-printing method to monitor protein dynamics and protein-protein interaction. In HDX-MS, sub-zero temperature LC separation has been developed to minimize the back-exchange and maintain a high-sequence coverage, which are crucial aspects for HDX-MS analysis. We here optimized the LC separation for HDX-MS using different mobile phases under different subzero temperatures with the tryptically digested *E. coli* cell lysate. The back-exchange rates were evaluated by comparing the theoretical maximum deuterium uptake to the detected mass of each identified peptide using customized software. The optimized sub-zero temperature RPLC platform (i.e., -9 ℃, water/acetonitrile-based buffers, 10,000 psi) was then used to characterize the conformational changes in the anthrax protective antigen (PA) and its known antibody complex samples. We identified a potential binding site of the PA and one anti-PA antibody. Furthermore, we performed the same technique to characterize the binding sites of PA and two different anti-PA antibodies in a one-pot sample, showing the potential of the sub-zero temperature RPLC HDX-MS platform to identify site-specific information in complex samples.

5.2 Introduction

Proteins are of great importance in every living organism. They perform a variety of essential functions in different biological processes such as immune response, energy conversion, transport, catalysis, etc. [131]. The functionality of a protein depends on the structures and dynamics of that protein; thus, understanding the structures and dynamics of proteins is essential to the studies of protein functions. Hydrogen deuterium exchange mass spectrometry (HDX-MS) has become a promising technique for characterizing protein structure and dynamics, as well as protein/protein interactions [132]. HDX-MS monitors the exchange of protein backbone amide hydrogens to deuterium to reveal conformational information regarding the structures of proteins or protein complexes. HDX-MS technique is complementary to traditional structural biology techniques such as X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and cryogenic electron microscopy (cryo-EM) [133, 134]. However, protein crystals are required for X-ray crystallography, purified protein is required for NMR techniques, and special sample treatment is needed for cryo-EM techniques. The crystallization of proteins is normally time consuming and X-ray crystallography and NMR techniques are limited to small size proteins[135]. Compared to these traditional structural biology methods, HDX-MS offers the advantage of simple sample preparation, which only requires D_2O to produce label proteins. Additionally, since the deuterated proteins are digested into small peptides for MS or LC-MS analysis, HDX-MS can be applied to analyze large proteins, which is difficult to achieve using other methods [136]. Furthermore, HDX-MS also has minimal background matrix effect, which means that it can be applied to analyze proteins of interest in complex samples [133].

HDX-MS, however, does offer its own drawback including the sensitivity of deuteration to experimental conditions such as pH and temperature. Therefore, careful optimization and control of HDX experimental conditions is essential to deliver a reliable and unbiased result [133, 137]. Another issue in the implementation of HDX-MS is the back exchange of deuterium with hydrogens from the solvent after quenching, during protein digestion, reduction, and LC separation [138]. Hydrogen back-exchange affects the measurement of amide hydrogen exchange rates which results in loss of site-specific information and can bias the identification of sites of interest.

Extensive efforts have been made to optimize the HDX experimental conditions, especially the conditions of post HDX quenching steps, to minimize the back-exchange of deuterium [138-141]. These efforts include low-pH and low-temperature protein digestion [142] and fast chromatographic separation at low temperature (*e.g.* 0 ℃) [143, 144]. However, fast chromatographic separation limits the identification of peptides, reduces sequence coverage, and cannot sufficiently separate complex samples. The loss of site-specific information may impede the identification of binding sites if the corresponding peptides cannot be identified. To minimize back-exchange during longer chromatographic separation sub-zero temperature LC for HDX-MS has been developed and evaluated [138, 139]. Bai and Englander have demonstrated that decreasing the temperature of separation from 0 to -30 ℃ can effectively reduce the deuterium exchange rate by about 40-fold[145]. With appropriate mobile phase modifier, to prevent the mobile phase from freezing, reversed phase liquid chromatography (RPLC) at sub-zero temperature demonstrated a negligible loss of deuterium.

We here evaluated LC separation at subzero temperatures using different mobile

phases to minimize the back exchange and maintain a high sequence coverage. The optimized sub-zero temperature (*i.e.,* -9 ℃) RPLC separation system were used for the HDX-MS of the antibody-antigen interactions of the adaptive immune system (*i.e.,* anthrax protective antigen (PA) and the interactions of various anti-PA antibody complex mixtures). We have identified a potential binding site of PA with one anti-PA antibody and characterized the binding sites of PA with two different anti-PA antibodies in a one-pot sample. Clear characterization of paratope/epitope interaction of an immune complex can further the understanding of immune response and possibly elucidate the mechanisms of immune diseases [146, 147]. Moreover, our results demonstrate the potential of a sub-zero temperature RPLC HDX-MS platform to identify site-specific interactions in complex samples.

5.3 Material and methods

5.3.1 Materials and reagents

All chemicals, including Protease Type XIII from Aspergillus saitoi (≥0.6 unit/mg) and Deuterium oxide (≥99.6 atom % D), were purchased from Sigma-Aldrich (Milwaukee, WI) unless noted otherwise. Phenylmethylsulphonyl fluoride (PMSF) was purchased from VWR (Radnor, PA) and trypsin (TPCK treated) was obtained from ThermoFisher (Rockford, IL). The desalting column, Strata C18-U (55 μm, 70 Å, 100 mg/mL) was purchased from Phenomenex (Torrance, CA). An ACE® Excel® SuperC18™ column (100 mm × 2.1 mm, 1.7 μm, 90 Å) was purchased from Advanced Chromatography Technologies Ltd (Aberdeen, Scotland).

5.3.2 Sample preparation

Escherichia coli (*E. coli*) K12 cells were grown in 2% LB media and cultured at 37 °C with gentle shaking at 250 rpm overnight. The cultured *E. coli* was collected and centrifuged at 13,000 rpm for 60 minutes. The supernatant was discarded, and the resulting *E. coli* pellets were resuspended in 25 mM ammonium bicarbonate at a ratio of 1 gram of cell pellets to 5 mL of buffer. 0.1% (v/v) PMSF was added to inhibit the protease from degrading proteins. Then, the *E. coli* pellets were lysed using EmulsiFlex-C3 homogenizer. The cell lysate was then centrifuged at 4 °C at 13,000 rpm for 60 minutes to remove cell membrane. The proteins in the supernatant were denatured using 6 M urea and reduced using 200 mM dithiothreitol and 200 mM iodoacetamide. The proteins were then digested with trypsin at 37 °C overnight with a protein to enzyme ratio of 50:1 (m/m). The peptides were then desalted using solid phase extraction columns and vacuum dried. Non-deuterated *E. coli* peptides were prepared by reconstituting the vacuum dried peptides into LC-MS grade water to a concentration of 1 μg/μL. Fully deuterated *E. coli* peptides were prepared by reconstituting the vacuum dried peptides into D_2O to a concentration of 1 μg/ μL and incubated at room temperature for three days.

5.3.3 Low-temperature liquid chromatography

Low-temperature liquid chromatography was achieved by placing a six-port valve and a C18 RPLC column (100 mm × 2.1 mm, 1.7 μm, 90 Å) in a portable freezer. The temperature of the portable freezer can be controlled at a range from 20 °C to - 20°C. Two different mobile phase systems (mobile phase system 1 and 2) were utilized to avoid freezing of solvent in the LC system. Mobile phase system 1 was made up of 89.9% HPLC

water, 10% acetonitrile, and 0.1% formic acid for mobile phase A and 99.9% acetonitrile and 0.1% formic acid for mobile phase B. Mobile phase system 2 was made up of 64.9% water, 35% methanol, and 0.1% formic acid for mobile phase A and 99.9% acetonitrile and 0.1% formic acid for mobile phase B.

LC separation performance of the two different mobile phase systems were evaluated using *E. coli* digest and detected with a photodiode array (PDA) detector. For mobile phase system 1, 25 µg of non-deuterated *E. coli* lysate peptides were manually injected into the LC system at room temperature, 4 °C and -9 °C. The flow rate was 150 µL/min. The gradient started with 0% mobile B for sample loading over 10 minutes followed by an increase from 0% to 35% mobile phase B over 30 minutes for peptide separation. Then the column was washed by 90% mobile phase B for 5 minutes, followed by a decrease to 0% mobile phase B in 3 minutes. The column was re-equilibrated by flushing with 100% mobile phase A for 10 minutes. For mobile phase set 2, the same sample and gradient procedures were used, except mobile phase B was ramped to 40% instead of 35% over 30 minutes for peptide separation. Mobile phase system 2 was also evaluated at -20 °C.

The back-exchange rates of the two different mobile phase systems at different temperatures were evaluated using fully deuterated *E. coli* digest and detected using MS. 25 µg fully deuterated *E. coli* lysate peptides were manually injected into LC system at room temperature, 4 °C, -9 °C, and -20 °C. Mobile phase system 1 and 2 were both used for room temperature, 4 °C, and -9 °C. Mobile phase system 2 was used for -20 °C.. A longer gradient, 10% mobile phase B to 40 % mobile phase B over 90 minutes with a flow rate of 150 µL/min, was also applied to evaluate the back-exchange of fully deuterated *E.*

coli peptides.

5.3.4 Hydrogen/Deuterium Exchange

Differential HDX experiments were performed for PA epitope mapping. 6 μM PA was prepared by diluting 12 mM PA into HPLC water at a ratio of 1:1 (v/v). The PA and anti-PA antibody immunocomplex sample was prepared by premixing 12 μM PA with 15 μM anti-PA antibody at ratio of 1:1 (v/v) at room temperature. For deuterium labeling, 4 μL of free PA sample and 4 μL of complex sample were diluted with 20 μL of D_2O and incubated for 3.5 minutes at room temperature, individually. Experiments were conducted in triplicate. The deuteration was quenched by adding 24 μL of chilled 1% formic acid to a final pH of 2.4 and incubated at 0 °C for 2 minutes. The quenched solution then was incubated with 48 μL of 1.2 mg/mL protease XIII at 0 °C for 4 minutes for protein digestion. The digested peptides were quickly injected into the low-temperature LC system with a gradient from 0% to 35% mobile phase B over 30 minutes at -9 °C followed by mass spectrometer analysis. 4 μL of 6 mM PA sample was incubated with 20 μL of H_2O, followed by same quenching, digestion, LC separation, and MS analysis steps for peptide identification.

5.3.5 Bottom-up MS analysis

An LTQ Orbitrap Elite mass spectrometer (Thermo Fisher Scientific, Hanover Park, IL, USA) with a custom nano-ESI interface was used for LC-MS/MS. The heated capillary temperature was set to 275°C with a spray voltage of 3.5 kV. MS scans were obtained using the Orbitrap MS with a resolution setting of 100,000 and *m/z* range from

350 to 1350. The AGC was set to 5E6 and the max ion time was set to 1000 ms with 2 micro scans. MS/MS scans were acquired using the LTQ MS with collisional induced dissociation (CID) at a normalized collision energy setting of 35%. The ten most abundant precursor ions were selected for MS/MS. The AGC for MS/MS was set to 1E6 and the max ion time was set to 1000 ms with 2 micro scans.

5.3.6 Data analysis

5.3.6.1 Peptide identification

Non-deuterated E. coli lysate peptides were identified using MSGF+ to search the mass spectra from the LC-MS/MS analysis against the annotated E. coli database and its decoy database. The peptide identifications were filtered using a SpecE cut-off value of 1E−10 (i.e., the calculated FDR <1% at the unique peptide level).

5.3.6.2 Deuterium uptake calculation

Deuterium incorporation of peptides was calculated using the in-house developed software. Deuteration level of each peptide was calculated as shifted weighed mass of each peptide divided by the total number of exchangeable amide hydrogens [148].

5.4 Results and discussion

5.4.1 Evaluation of low-temperature RPLC separation

RPLC is often performed at room temperature or higher for more efficient mass transfer of analytes to achieve better separation. The RPLC mobile phase A utilizes water

which has a freezing point of 0 ℃. To operate LC under sub-zero temperatures, mobile phase modifiers can be used to decrease the freezing points of mobile phases to avoid freezing the LC system. The modifiers should have good electrospray efficiency to minimize the effects on the sensitivity of MS detection. Ethylene glycol, dimethyl formamide, formamide, and methanol have been previously used as mobile phase modifiers [139]. We chose methanol as the modifier to perform LC at -20 ℃ because methanol has good electrospray efficiency and low hydrophobicity which should have a limited impact on the separation resolution. Acetonitrile is commonly used as mobile phase B for RPLC separation of peptides. The freezing point of 10% acetonitrile is slightly lower than -9 ℃. Thus, we evaluated four different temperatures: room temperature, 4 ℃, -9 ℃, and -20 ℃. In order to run LC under sub-zero temperature, we used two different mobile phase systems: (1) *Mobile phase system 1*: 0.1 % formic acid in water as mobile phase A, and 0.1 % formic acid in acetonitrile as mobile phase B; (2) *Mobile phase system 2*: 0.1 % formic acid, 30 % methanol in water for mobile phase A, and 0.1 % formic acid in acetonitrile as mobile phase B.

For mobile phase system 1, RPLC separation of E. coli digest with a gradient from 10% mobile phase B to 35% mobile phase B was conducted at room temperature, 4 ℃, and -9 ℃. For mobile phase system 2, we added 30% methanol into mobile phase A (0.1 % formic acid, 30 % methanol in water) to lower the freezing points to below -20 ℃. With this mobile phase composition, we performed RPLC separation of *E. coli* digest with a gradient from 0% B to 40% B at room temperature, 4 ℃, -9 ℃, and -20 ℃. To compare the separation efficiency using different mobile phases at different temperatures, a UV/Vis PDA detector was used. As shown in **Figure 5-1**, the separation performance under sub-

zero temperatures was similar to the performance at 4 ℃ with both mobile phase systems 1 and 2. However, as the temperature decreased, the peaks became broader. Meanwhile, some peptides did not bind to the column at lower temperatures, resulting in peaks eluted in flow through time. For mobile phase system 1 at 4 ℃ and -9 ℃, the elution profiles were very similar which indicates that the decrease in temperature did not affect the separation performance. In general, LC separation resolution was relatively low using mobile phase system 2 (*e.g.,* broader peaks). While -20 ℃ is favorable for HDX to minimize the back-exchange of deuterium, the RPLC separation under -20 ℃ using mobile phase system 2 may not provide sufficient separation resolution.

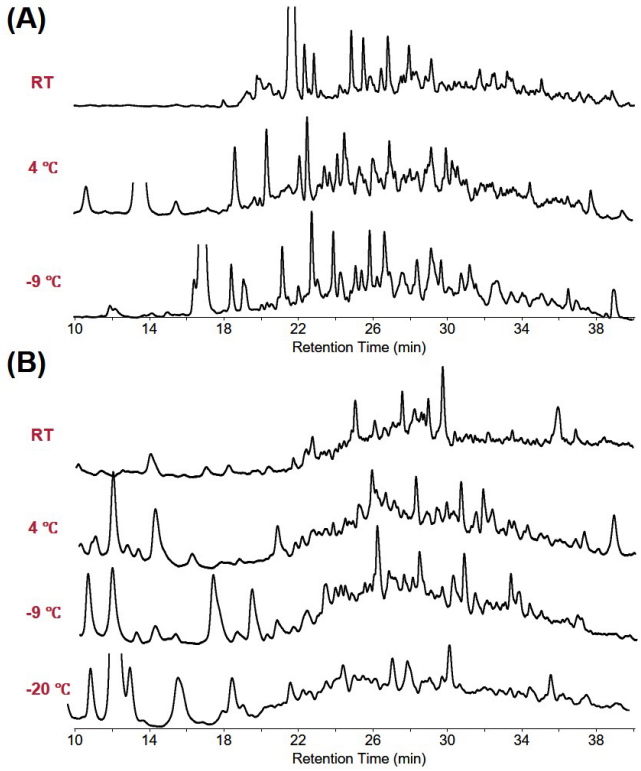


Figure 5-1. UV traces of RPLC chromatograms of *E. coli* digest at different temperatures. (A) Mobile phase system 1: 89.9% HPLC water, 10% acetonitrile, and 0.1% formic acid as A and 99.9% acetonitrile and 0.1% formic acid as B. (B) Mobile phase system 2: 64.9% water, 35% methanol, and 0.1% formic acid as A and 99.9% acetonitrile and 0.1% formic acid as B.

5.4.2 Back-exchange rate evaluation

Lower temperatures are favorable to HDX separation and it has been reported that by decreasing the temperature from 0 ℃ to -30 ℃, deuterium exchange rate can be reduced the by about 40 fold [145]. However, low temperatures are not favorable for RPLC separation of peptides due to the lower mass transfer rate under low temperatures. Our data indicated that lower resolution was obtained for RPLC separation using a C18 column at -20 ℃ than at -9 ℃. The low separation resolution resulted in insufficient separation of peptides and fewer peptide identifications. In complex samples, fewer peptide identifications decreases the overall peptide coverage and can result in the loss of information about binding sites.

In order to examine the back-exchange status of deuterated peptides, fully deuterated *E. coli* digest was prepared by incubating *E. coli* digest with D_2O for 3 days. The fully deuterated *E. coli* digest was then analyzed by LC-MS using mobile phase system 2 at -9 ℃ and -20 ℃. The mass of each peptide eluted at different time from the fully deuterated sample were measured. The deuterium level of each peptide was calculated as follow: the average masses of the peptide detected in the fully-deuterated and non-deuterated samples were calculated. The difference between the average masses of the same peptide in the fully-deuterated and non-deuterated samples was calculated to indicate the mass increase due to the incorporation of deuterium. Then, the mass difference was divided by the maximum deuterium uptake of the peptide to calculate the deuterium level of the peptide. The non-deuterated *E. coli* digest was analyzed by LC-MS/MS for peptide identification. An in-house developed software was used to match the deuterium labeled peptides to the non-deuterated peptides to identify the deuterated peptides. Briefly

speaking, a peptide identification list from the non-deuterated peptide sample with their corresponding m/z ratios was imported into the software. For each identified peptide, a list of m/z ratios with different deuterium uptake (*i.e.* +1 deuterium, +2 deuterium, +3 deuterium, ……) was generated. The software then calculated the mass error between the m/z on the list and each m/z detected in the spectrum from the deuterated peptide samples with the similar retention time window (time tolerance can be changed). If a cluster of m/z values (monoisotopic distribution) of the deuterated peptide samples with mass error less than 15 ppm can be matched with the m/z values on the list, the deuterated peptide will be assigned with the sequence of the peptides identified from the non-deuterated sample. The mass error matching algorithm was shown in **Figure 5-2**.

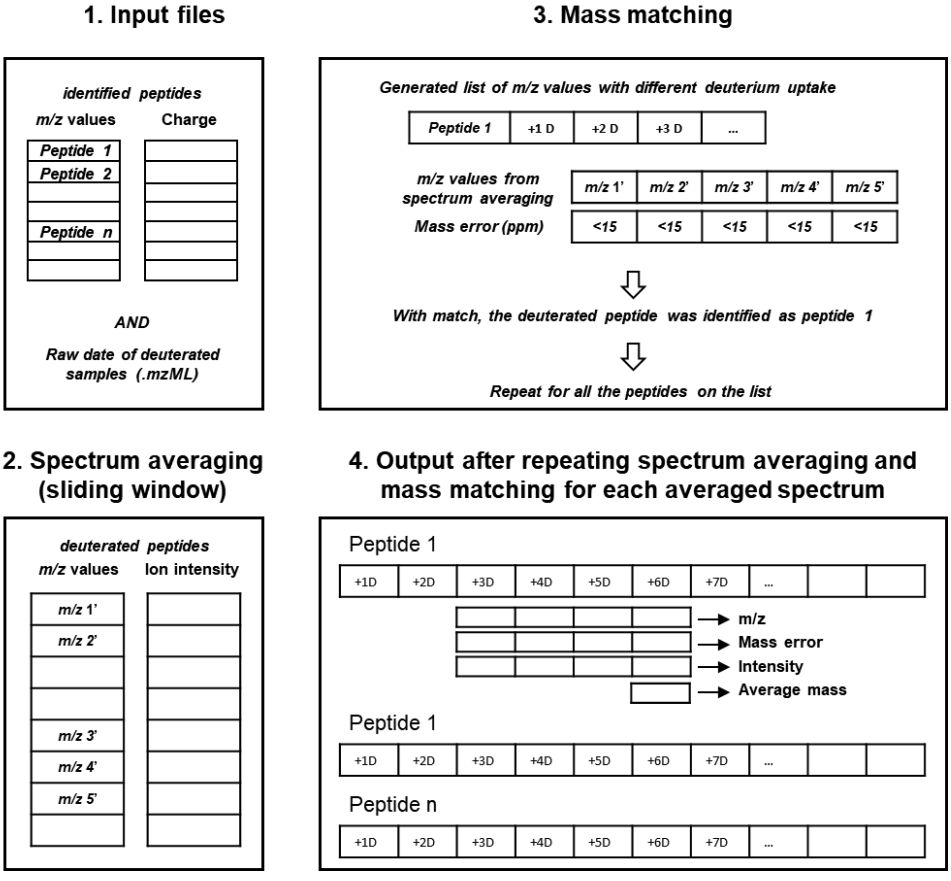


Figure 5-2. The mass error matching algorithm used for deuterated peptide identification.

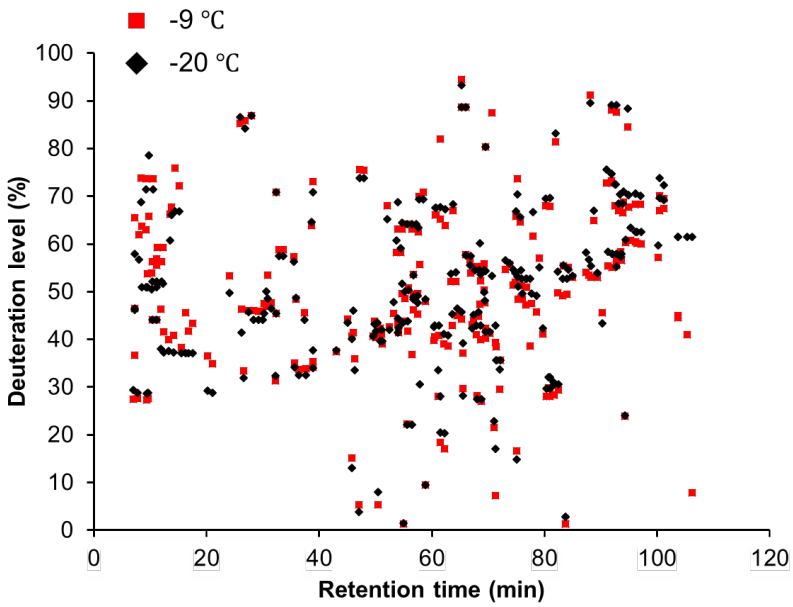


Figure 5-3. Deuteration level vs. retention time of *E. coli* peptides under -9 ℃ and -20 ℃.

The deuteration level of each identified peptide from the *E. coli* digest under different temperature conditions was calculated to examine the deuterium back-exchange rate, as shown in **Figure 5-3**. Using mobile phase system 2 at -9 ℃ and -20 ℃, similar deuteration levels of identified peptides from the *E. coli* digest were observed, which indicated that the rate of deuterium back-exchange at -9 ℃ and -20 ℃ were similar. The back-exchange rate at higher temperature was also examined using mobile phase system 1. Increasing deuterium uptake was observed with the decrease in temperature from room temperature to -9 ℃ with the use of mobile phase system 1, **Figure 5-4**. When mobile phase system 2 was used, similar deuterium uptake was observed for the two peptides at -9 ℃ and -20 ℃. Considering RPLC separation at -9 ℃ offered better separation resolution, -9 ℃ and mobile phase system 1 were used for our low-temperature LC HDX-MS experiments.

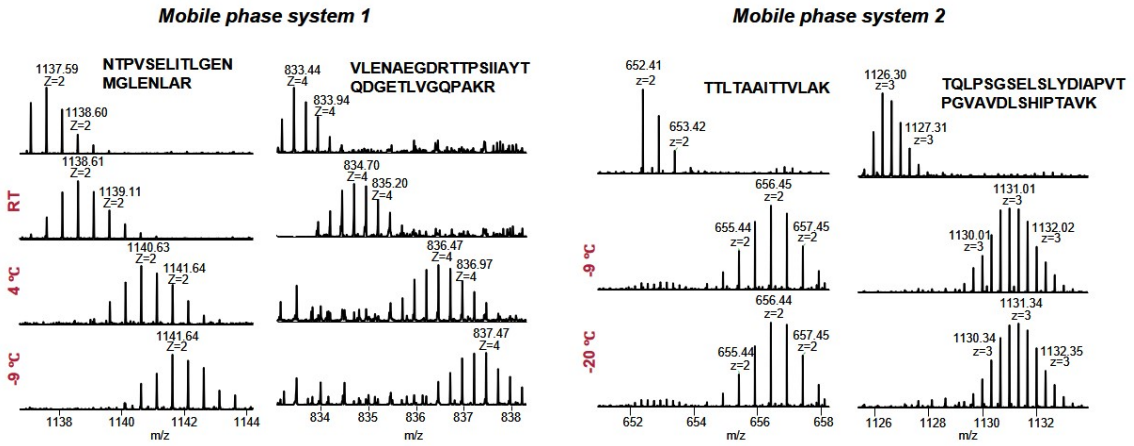


Figure 5-4. Mass spectra of deuterated peptides under different temperature conditions.

5.4.3 Epitope mapping of Anthrax protective antigen

After the optimization of the subzero temperature LC separation, we applied the platform to characterize the epitope of PA to an anti-PA antibody (C01). Differential HDX-MS was performed, where the free PA antigen sample and the PA – anti-PA complex sample were analyzed using the low-temperature RPLC platform. The non-deuterated PA sample was analyzed using low-temperature LC for peptide identification. Then, the free PA and complex samples were diluted into D_2O for HDX reaction. After the reaction was quenched, the samples were quickly digested using protease XIII and separated using low-temperature LC for direct mass measurement. Triplicate runs were performed for each sample.

The in-house developed software was used to calculate the average mass of each identified PA peptide after deuteration in the free PA and complex samples. The average mass of each peptide from the triplicate runs were then averaged and the deuteration level (%) was calculated by comparing the mass difference of the peptide in the free PA and

complex samples with its theoretical maximum deuterium uptake. Deuteration levels less than ± 5% were considered to not have a significant mass shift and it was concluded that these peptides were not involved in the binding event or the environment of these peptides did not change due to binding. For all the peptides with deuteration levels greater than ± 5%, manual evaluation of the raw data was performed to ensure the validity of the identification. A heatmap of each identified PA peptide was generated to visualize the deuteration level (**Figure 5-5**).

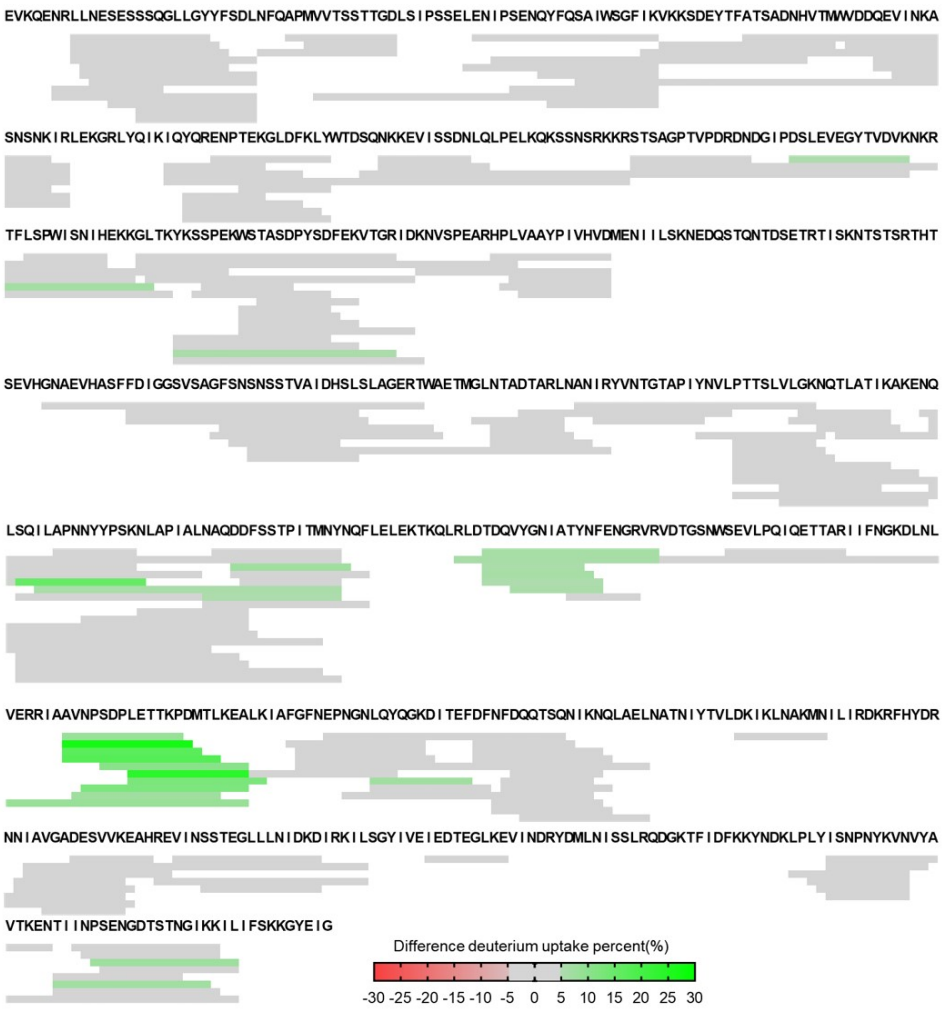


Figure 5-5. Heatmap of deuteration level of PA peptides obtained from differential HDX-MS of the free PA sample and the PA – anti-PA C01 antibody complex sample.

From the heatmap, we observed several peptides with consistent deuteration level increase comparing the peptides in the free PA sample with the peptides in the complex sample. Thus, we concluded that the epitope of PA binding to anti-PA C01 antibody was AAVNPSDPLETTKPDMTLKEA. The spectra of peptides that were in this region is shown in **Figure 5-6A**. For comparison, peptides that were not affected by the binding event are shown in **Figure 5-6B**.

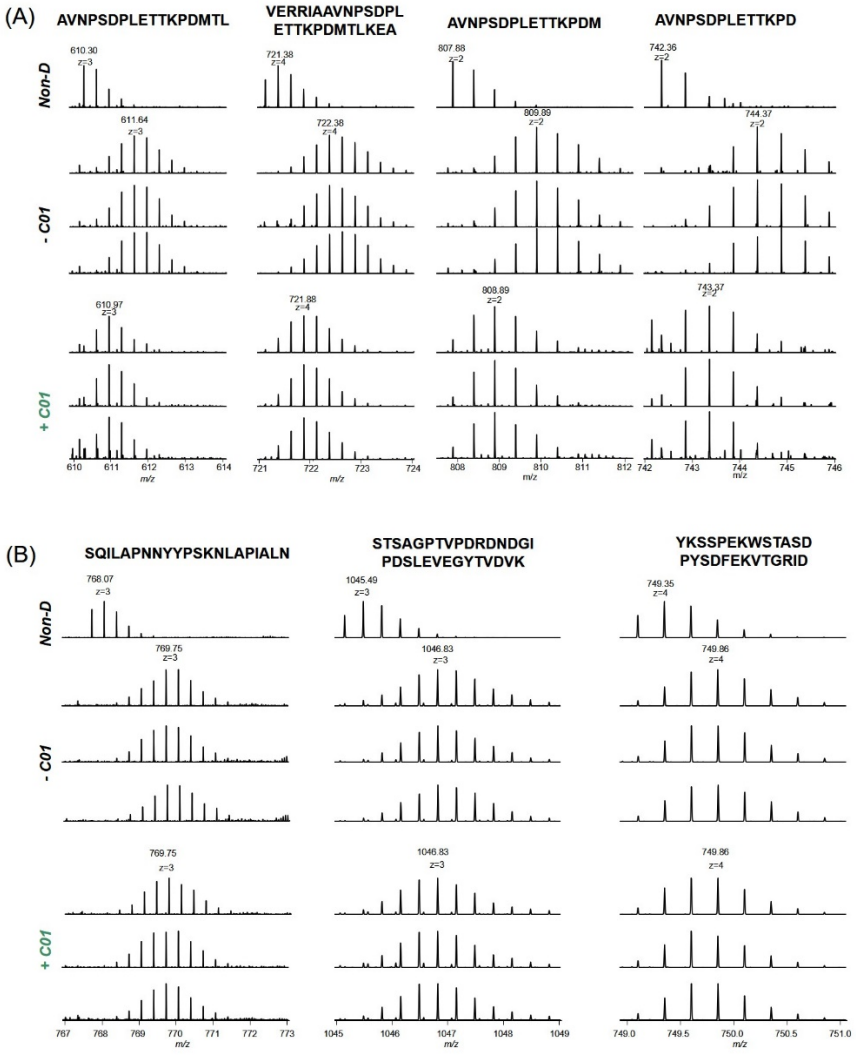


Figure 5-6. Mass spectra of PA peptides that were (A) involved and (B) not involved in the binding events.

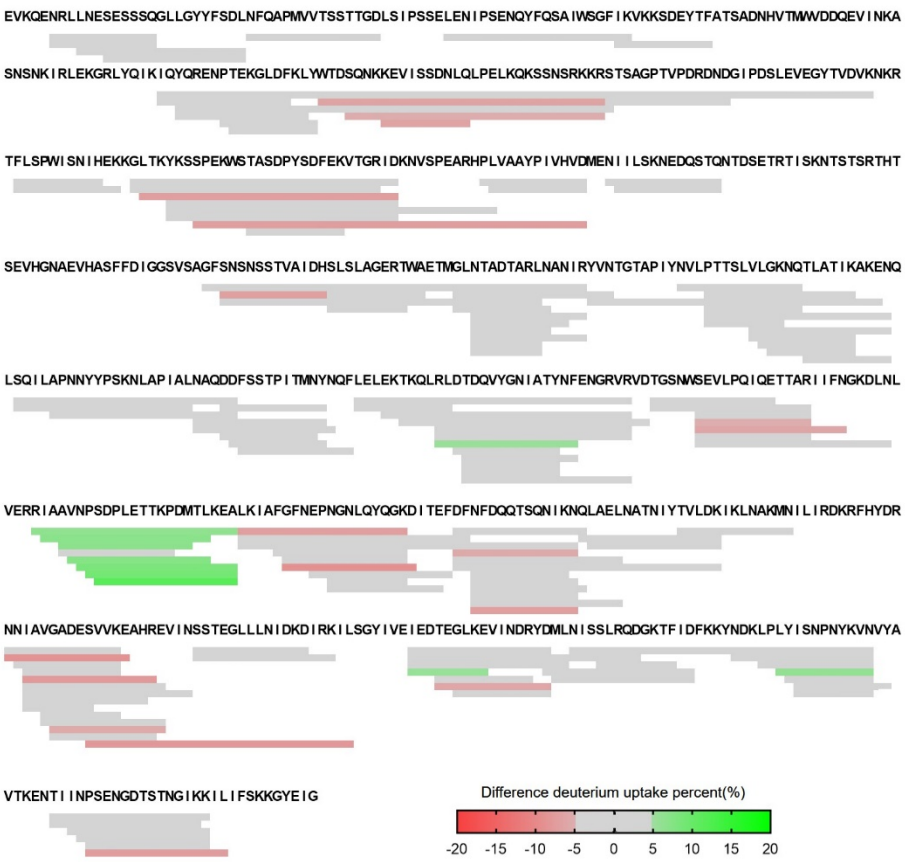


Figure 5-7. Heatmap of the deuteration level of PA peptides obtained from differential HDX-MS of the free PA sample and the PA, anti-PA antibody C01, and F01 mixture.

Furthermore, the same experiments were performed for the free PA sample and a mixture of PA, anti-PA antibody, C01 and F01, respectively. The heatmap (**Figure 5-7**) indicates the same binding site between PA and anti-PA antibody C01 as was identified using when pure anti-PA antibody C01 was bound with PA. The results indicate that the low-temperature LC HDX-MS can be used to analyze the protein-protein interaction in a relatively complex sample. However, we were not able to identify the binding site between PA and anti-PA antibody F01 and there are two possible reasons. The binding of PA to C01 may be more favorable than the binding of PA to F01, so the PA may have been bound primarily to C01. Alternatively, PA epitope to F01 may be contain only a few amino acid

residues. As the mass shift for shorter peptides is smaller, back-exchange is more of an issue for these peptides and the differential deuterium uptake may not have been measurable even under low-temperature separation conditions. To address these issues, the low-temperature LC HDX-MS platform could be further optimized with lower temperature to further minimize back-exchange during LC separation and MS detection.

5.5 Conclusions

We have developed a low-temperature LC system to separate deuterated peptide samples for differential HDX-MS analysis. Using mobile phase system 1, the separation resolution and deuterium back-exchange effect were optimized at -9 ℃. The platform was then successfully applied to analyze the epitope of PA to anti-PA antibodies in both pure antibody sample and a mixture of two anti-PA antibodies sample. However, we were not able to identify the binding site of anti-PA F01 when anti-PA C01 was present. Future work can be done to further optimize the platform and experimental conditions to identify multiple binding sites. For example, capillary packed RPLC columns may be used to increase the separation resolution and sensitivity of the platform. Lower temperature with suitable mobile phase modifiers could also be applied to further minimize the back-exchange. Modifications to the ionization process and implementation of different MS technology may also be considered to improve HDX identification. The back-exchange of peptides during the ESI process may be reduced by the incorporation of a flow of chilled nitrogen gas to the ESI spray. Additionally, the resolution of HDX-MS can be improved by using electron transfer dissociation (ETD), rather than collision induced dissociation (CID), which will not cause the rearrangement of deuterium on the peptides during the

dissociation process. Single amino acid resolution could even be achieved with the careful optimization of experimental conditions [149]. Overall, HDX-MS is a promising technique that is complementary to traditional structural biology techniques. With the incorporation of low-temperature LC, the level of deuterium labeling retained by peptide amides after a prolonged peptide separation has been improved, which will broaden the utility of HDX-MS [138].

*Authors: Mulin Fang, Zhe Wang, Thomas Welborn, Kellye A. Cupp-Sutton, Jiwon Kang, Kenneth Smith, and Si Wu

MF and ZW contributed equally.

Chapter 6 Overall summary and future directions

6.1 Overall summary

The work described here within this dissertation was about the developments of high-throughput top-down proteomics and the applications on analyzing human serum autoantibody repertoire. The need for developing high-throughput top-down proteomics arises from the complexity of biological samples. Proteomics techniques provide the tools to the analysis of proteins which are essential to life. Top-down proteomics makes it possible to characterize and quantify intact proteoforms and provides valuable information to the understanding of protein functions. As current proteomics techniques can only reveal limited information about proteomes, the development of techniques is needed to explore more of proteomes.

A two-dimensional separation platform using high-pH RPLC and low-pH RPLC was shown in Chapter 2 and 3, to improve the separation power of intact proteins for top-down proteomics. The 2DLC system utilized RPLC which is the most commonly applied technique for intact protein separation on both of the two dimensions. By varying the pH conditions of the mobile phases used in the two dimensions, orthogonal separation was achieved for intact proteins in complex samples. The protein conformation changes under different pH condition due to electrostatic force changes, which alters the hydrophobicity of proteins. The orthogonality of RPLC separation of intact proteins under different pH conditions was evaluated using protein standards and the whole *E. coli* proteome. The 2DLC was proved to be able to separate intact proteins efficiently and orthogonally.

We further developed a comprehensive online 2DLC system based on the work in Chapter 2 to increase the sensitivity of the top-down proteomics platform for the analysis of low-quantity biological samples. The two dimensions utilized Nano-flow capillary RPLC to minimize the usage of sample amounts. The online 2DLC system achieved comparable results to the offline 2DLC system with 100 times less sample injected. The online 2DLC system could potential provide a promising and sensitive tool for the analysis of low-volume samples.

Autoantibodies are a hallmark of many autoimmune diseases and can be present in serum years before clinical symptoms arise, thus the detection and characterization of autoantibody repertoire have been drawing attentions.

Based on previous work in our lab, we optimized an automatic single-dimension RPLC platform through a custom-modified ultra-high pressure nano LC system (UPLC, maximum pressure 14,000 psi, operation pressure 10,000 psi) to improve the separation of highly homologous autoantibodies in serum. This approach has been successfully applied to a 12 standard monoclonal antibody Fab mixture, demonstrating the feasibility to separate and sequence intact antibodies with high sequence coverage and high sensitivity. We then applied the optimized platform to characterize serum autoantibody Fabs in a systemic lupus erythematosus (SLE) patient sample compared to healthy control samples, showing that 80+ dominant antibody Fab related mass features are only observed in the SLE sample, which is the first top-down demonstration of a serum autoantibody pool analysis.

The characterization of immune complexes promotes the understanding of human immunity and provides the information about the mechanisms of related diseases. We applied low-temperature LC coupled to HDX-MS to characterize the binding between

Anthrax protective antigen and different anti-PA antibodies.

6.2 Future directions

Proteins play important roles in human body, including transporting molecules, building cell/organism structures, catalyzing metabolic reaction, and etc. Proteomes consist of numerous proteins with wide dynamic ranges in term of abundance and different modifications. On the long way of developing top-down proteomics techniques, every step will need to be considered and promoted by the collaboration of researchers in different fields, including the separation techniques, mass spectrometry techniques and instrumentation, bioinformatic tools for top-down proteomics. More efficient multidimensional separation with higher separation power will be beneficial to the separation of proteomes. The development of more affordable mass spectrometers will help the promotion of top-down proteomics in more laboratories and boost the pace of top-down proteomics development. Higher resolution mass spectrometers help to study larger proteins with better mass accuracy. Top-down proteomics data is normally tedious to analyze, advanced bioinformatic tools for top-down proteomics is in urgent need to provide robust and accurate tools for top-down proteomics data interpretation.

Analyzing autoantibodies in serum samples is extremely challenging due to their similarity and complexity. With current application of UPLC-HRMS platform described in chapter 4, only very limited information can be provided. The online 2DLC coupled to high resolution MS could be applied to analyze autoantibody repertoire, potentially providing deeper characterization of human serum autoantibodies. In addition, to fully understand the function of autoantibodies of interest, techniques that can provide structural

information, such as native proteomics, hydrogen deuterium exchange mass spectrometry, and etc., can be developed and applied to elucidate the structures of proteins and protein complexes, such as epitope/paratope mapping for immune complexes.

References

1. Smith, L.M., N.L. Kelleher, and P. Consortium for Top Down, *Proteoform: a single term describing protein complexity.* Nat Methods, 2013. **10**(3): p. 186-7.
2. Anderson, N.L. and N.G. Anderson, *Proteome and proteomics: new technologies, new concepts, and new words.* Electrophoresis, 1998. **19**(11): p. 1853-61.
3. Blackstock, W.P. and M.P. Weir, *Proteomics: quantitative and physical mapping of cellular proteins.* Trends Biotechnol, 1999. **17**(3): p. 121-7.
4. Rogers, S., et al., *Investigating the correspondence between transcriptomic and proteomic expression profiles using coupled cluster models.* Bioinformatics, 2008. **24**(24): p. 2894-900.
5. Gaudet, P., et al., *The neXtProt knowledgebase on human proteins: 2017 update.* Nucleic Acids Res, 2017. **45**(D1): p. D177-D182.
6. Aebersold, R., et al., *How many human proteoforms are there?* Nat Chem Biol, 2018. **14**(3): p. 206-214.
7. Zhang, Z., et al., *High-throughput proteomics.* Annu Rev Anal Chem (Palo Alto Calif), 2014. **7**: p. 427-54.
8. Moradian, A., et al., *The top-down, middle-down, and bottom-up mass spectrometry approaches for characterization of histone variants and their post-translational modifications.* Proteomics, 2014. **14**(4-5): p. 489-97.
9. Arnaudo, A.M., R.C. Molden, and B.A. Garcia, *Revealing histone variant induced changes via quantitative proteomics.* Crit Rev Biochem Mol Biol, 2011. **46**(4): p. 284-94.
10. Meyer, B., D.G. Papasotiriou, and M. Karas, *100% protein sequence coverage: a modern form of surrealism in proteomics.* Amino Acids, 2011. **41**(2): p. 291-310.
11. Gregorich, Z.R. and Y. Ge, *Top-down proteomics in health and disease: challenges and opportunities.* Proteomics, 2014. **14**(10): p. 1195-210.
12. Toby, T.K., L. Fornelli, and N.L. Kelleher, *Progress in Top-Down Proteomics and the Analysis of Proteoforms.* Annu Rev Anal Chem (Palo Alto Calif), 2016. **9**(1): p. 499-519.
13. Fenn, J.B., et al., *Electrospray ionization for mass spectrometry of large biomolecules.* Science, 1989. **246**(4926): p. 64-71.
14. McLafferty, F.W., *A century of progress in molecular mass spectrometry.* Annu Rev Anal Chem (Palo Alto Calif), 2011. **4**: p. 1-22.
15. Senko, M.W., S.C. Beu, and F.W. McLafferty, *High-resolution tandem mass spectrometry of carbonic anhydrase.* Anal Chem, 1994. **66**(3): p. 415-8.
16. Wang, Z., et al., *Top-down Mass Spectrometry Analysis of Human Serum Autoantibody Antigen-Binding Fragments.* Sci Rep, 2019. **9**(1): p. 2345.
17. Wang, Z., et al., *Two-Dimensional Separation Using High-pH and Low-pH Reversed Phase Liquid Chromatography for Top-down Proteomics.* Int J Mass Spectrom, 2018. **427**: p. 43-51.
18. Shen, Y., et al., *High-resolution ultrahigh-pressure long column reversed-phase liquid chromatography for top-down proteomics.* J Chromatogr A, 2017. **1498**: p. 99-110.

19. Gargano, A.F.G., et al., *Increasing the Separation Capacity of Intact Histone Proteoforms Chromatography Coupling Online Weak Cation Exchange-HILIC to Reversed Phase LC UVPD-HRMS.* J Proteome Res, 2018. **17**(11): p. 3791-3800.
20. Chen, B., et al., *Online Hydrophobic Interaction Chromatography-Mass Spectrometry for Top-Down Proteomics.* Anal Chem, 2016. **88**(3): p. 1885-91.
21. Valeja, S.G., et al., *Three dimensional liquid chromatography coupling ion exchange chromatography/hydrophobic interaction chromatography/reverse phase chromatography for effective protein separation in top-down proteomics.* Anal Chem, 2015. **87**(10): p. 5363-5371.
22. Lindon, J.C., G.E. Tranter, and D. Koppenaal, *Encyclopedia of spectroscopy and spectrometry*. 2016: Academic Press.
23. Chen, B., et al., *Top-Down Proteomics: Ready for Prime Time?* Anal Chem, 2018. **90**(1): p. 110-127.
24. Bloh, A.M., et al., *Determination of N-formimidoylthienamycin concentration in sera from pediatric patients by high-performance liquid chromatography.* J Chromatogr, 1986. **375**(2): p. 444-50.
25. Yang, Y., et al., *Studies on the effect of column angle in figure-8 centrifugal counter-current chromatography.* J Chromatogr A, 2011. **1218**(36): p. 6128-34.
26. Melani, R.D., et al., *CN-GELFrEE - Clear Native Gel-eluted Liquid Fraction Entrapment Electrophoresis.* J Vis Exp, 2016(108): p. 53597.
27. Wojcik, R., et al., *Capillary electrophoresis with Orbitrap-Velos mass spectrometry detection.* Talanta, 2012. **88**: p. 324-9.
28. Camerini, S. and P. Mauri, *The role of protein and peptide separation before mass spectrometry analysis in clinical proteomics.* J Chromatogr A, 2015. **1381**: p. 1-12.
29. Ansong, C., et al., *Top-down proteomics reveals a unique protein S-thiolation switch in Salmonella Typhimurium in response to infection-like conditions.* Proc Natl Acad Sci U S A, 2013. **110**(25): p. 10153-8.
30. Anderson, L.C., et al., *Identification and Characterization of Human Proteoforms by Top-Down LC-21 Tesla FT-ICR Mass Spectrometry.* J Proteome Res, 2017. **16**(2): p. 1087-1096.
31. Zhang, W., T. Hankemeier, and R. Ramautar, *Next-generation capillary electrophoresis-mass spectrometry approaches in metabolomics.* Curr Opin Biotechnol, 2017. **43**: p. 1-7.
32. Fonslow, B.R. and J.R. Yates, 3rd, *Capillary electrophoresis applied to proteomic analysis.* J Sep Sci, 2009. **32**(8): p. 1175-88.
33. Tian, Z., et al., *Two-dimensional liquid chromatography system for online top-down mass spectrometry.* Proteomics, 2010. **10**(20): p. 3610-20.
34. Vellaichamy, A., et al., *Size-sorting combined with improved nanocapillary liquid chromatography-mass spectrometry for identification of intact proteins up to 80 kDa.* Anal Chem, 2010. **82**(4): p. 1234-44.
35. Catherman, A.D., et al., *Top down proteomics of human membrane proteins from enriched mitochondrial fractions.* Anal Chem, 2013. **85**(3): p. 1880-8.
36. Shen, Y., et al., *High-resolution ultrahigh-pressure long column reversed-phase liquid chromatography for top-down proteomics.* J Chromatogr A, 2017.
37. Jorgenson, J.W., *Capillary liquid chromatography at ultrahigh pressures.* Annu Rev Anal Chem (Palo Alto Calif), 2010. **3**: p. 129-50.

38. Tran, J.C., et al., *Mapping intact protein isoforms in discovery mode using top-down proteomics.* Nature, 2011. **480**(7376): p. 254-8.
39. Xiu, L., et al., *Effective protein separation by coupling hydrophobic interaction and reverse phase chromatography for top-down proteomics.* Anal Chem, 2014. **86**(15): p. 7899-906.
40. Loo, J.A., et al., *High-resolution tandem mass spectrometry of large biomolecules.* Proc Natl Acad Sci U S A, 1992. **89**(1): p. 286-9.
41. Zhurov, K.O., et al., *Principles of electron capture and transfer dissociation mass spectrometry applied to peptide and protein structure analysis.* Chem Soc Rev, 2013. **42**(12): p. 5014-30.
42. Cotham, V.C. and J.S. Brodbelt, *Characterization of Therapeutic Monoclonal Antibodies at the Subunit-Level using Middle-Down 193 nm Ultraviolet Photodissociation.* Anal Chem, 2016. **88**(7): p. 4004-13.
43. Cao, L., et al., *Intact glycopeptide characterization using mass spectrometry.* Expert Rev Proteomics, 2016. **13**(5): p. 513-22.
44. Georgiou, G., et al., *The promise and challenge of high-throughput sequencing of the antibody repertoire.* Nat Biotechnol, 2014. **32**(2): p. 158-68.
45. Wine, Y., et al., *Serology in the 21st century: the molecular-level analysis of the serum antibody repertoire.* Curr Opin Immunol, 2015. **35**: p. 89-97.
46. Al Kindi, M.A., et al., *Secreted autoantibody repertoires in Sjogren's syndrome and systemic lupus erythematosus: A proteomic approach.* Autoimmun Rev, 2016. **15**(4): p. 405-10.
47. Mu, Q., H. Zhang, and X.M. Luo, *SLE: Another Autoimmune Disorder Influenced by Microbes and Diet?* Front Immunol, 2015. **6**: p. 608.
48. McCluskey, J., et al., *Determinant spreading: lessons from animal models and human disease.* Immunol Rev, 1998. **164**: p. 209-29.
49. Farris, A.D., et al., *Epitope mimics and determinant spreading: pathways to autoimmunity.* Cell Mol Life Sci, 2000. **57**(4): p. 569-78.
50. Mahler, M. and M.J. Fritzler, *Epitope specificity and significance in systemic autoimmune diseases.* Ann N Y Acad Sci, 2010. **1183**: p. 267-87.
51. Wolin, S.L. and K.M. Reinisch, *The Ro 60 kDa autoantigen comes into focus: interpreting epitope mapping experiments on the basis of structure.* Autoimmun Rev, 2006. **5**(6): p. 367-72.
52. Safonova, Y., et al., *IgRepertoireConstructor: a novel algorithm for antibody repertoire construction and immunoproteogenomics analysis.* Bioinformatics, 2015. **31**(12): p. i53-61.
53. Cheung, W.C., et al., *A proteomics approach for the identification and cloning of monoclonal antibodies from serum.* Nat Biotechnol, 2012. **30**(5): p. 447-52.
54. Al Kindi, M.A., et al., *Serum SmD autoantibody proteomes are clonally restricted and share variable-region peptides.* J Autoimmun, 2015. **57**: p. 77-81.
55. Zhang, Y., et al., *Protein analysis by shotgun/bottom-up proteomics.* Chem Rev, 2013. **113**(4): p. 2343-94.
56. Zhang, Z., H. Pan, and X. Chen, *Mass spectrometry for structural characterization of therapeutic antibodies.* Mass Spectrom Rev, 2009. **28**(1): p. 147-76.

57. Soler, L., et al., *Data on endogenous chicken sperm peptides and small proteins obtained through Top-Down High Resolution Mass Spectrometry.* Data Brief, 2016. **8**: p. 1421-5.
58. Catherman, A.D., O.S. Skinner, and N.L. Kelleher, *Top Down proteomics: facts and perspectives.* Biochem Biophys Res Commun, 2014. **445**(4): p. 683-93.
59. Gregorich, Z.R., Y.H. Chang, and Y. Ge, *Proteomics in heart failure: top-down or bottom-up?* Pflugers Arch, 2014. **466**(6): p. 1199-209.
60. Bogdanov, B. and R.D. Smith, *Proteomics by FTICR mass spectrometry: top down and bottom up.* Mass Spectrom Rev, 2005. **24**(2): p. 168-200.
61. Shishkova, E., A.S. Hebert, and J.J. Coon, *Now, More Than Ever, Proteomics Needs Better Chromatography.* Cell Syst, 2016. **3**(4): p. 321-324.
62. Roth, M.J., et al., *Sensitive and reproducible intact mass analysis of complex protein mixtures with superficially porous capillary reversed-phase liquid chromatography mass spectrometry.* Anal Chem, 2011. **83**(24): p. 9586-92.
63. Petersson, P., K. Haselmann, and S. Buckenmaier, *Multiple heart-cutting two dimensional liquid chromatography mass spectrometry: Towards real time determination of related impurities of bio-pharmaceuticals in salt based separation methods.* J Chromatogr A, 2016.
64. Tian, Z., et al., *Enhanced top-down characterization of histone post-translational modifications.* Genome Biol, 2012. **13**(10): p. R86.
65. Zhou, M., et al., *Profiling Changes in Histone Post-translational Modifications by Top-Down Mass Spectrometry.* Methods Mol Biol, 2017. **1507**: p. 153-168.
66. Wen, J., T. Arakawa, and J.S. Philo, *Size-exclusion chromatography with on-line light-scattering, absorbance, and refractive index detectors for studying proteins and their interactions.* Anal Biochem, 1996. **240**(2): p. 155-66.
67. Simpson, D.C., et al., *Using size exclusion chromatography-RPLC and RPLC-CIEF as two-dimensional separation strategies for protein profiling.* Electrophoresis, 2006. **27**(13): p. 2722-33.
68. Queiroz, J.A., C.T. Tomaz, and J.M. Cabral, *Hydrophobic interaction chromatography of proteins.* J Biotechnol, 2001. **87**(2): p. 143-59.
69. Mohammed, S. and A. Heck, Jr., *Strong cation exchange (SCX) based analytical methods for the targeted analysis of protein post-translational modifications.* Curr Opin Biotechnol, 2011. **22**(1): p. 9-16.
70. Gilar, M., et al., *Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions.* J Sep Sci, 2005. **28**(14): p. 1694-703.
71. Yang, F., et al., *High-pH reversed-phase chromatography with fraction concatenation for 2D proteomic analysis.* Expert Rev Proteomics, 2012. **9**(2): p. 129-34.
72. Chowdhury, S.M., et al., *Identification of cross-linked peptides after click-based enrichment using sequential collision-induced dissociation and electron transfer dissociation tandem mass spectrometry.* Anal Chem, 2009. **81**(13): p. 5524-32.
73. Wu, S., et al., *An integrated top-down and bottom-up strategy for broadly characterizing protein isoforms and modifications.* J Proteome Res, 2009. **8**(3): p. 1347-57.

74. Wu, S., et al., *Integrated workflow for characterizing intact phosphoproteins from complex mixtures.* Anal Chem, 2009. **81**(11): p. 4210-9.
75. Kim, S., N. Gupta, and P.A. Pevzner, *Spectral probabilities and generating functions of tandem mass spectra: a strike against decoy databases.* J Proteome Res, 2008. **7**(8): p. 3354-63.
76. Kim, S. and P.A. Pevzner, *MS-GF+ makes progress towards a universal database search tool for proteomics.* Nat Commun, 2014. **5**: p. 5277.
77. Liu, X., et al., *Protein identification using top-down.* Mol Cell Proteomics, 2012. **11**(6): p. M111 008524.
78. Guo, D.C., et al., *Prediction of Peptide Retention Times in Reversed-Phase High-Performance Liquid-Chromatography .1. Determination of Retention Coefficients of Amino-Acid-Residues of Model Synthetic Peptides.* Journal of Chromatography, 1986. **359**: p. 499-517.
79. Geng, X. and F.E. Regnier, *Retention model for proteins in reversed-phase liquid chromatography.* J Chromatogr, 1984. **296**: p. 15-30.
80. Mant, C.T., N.E. Zhou, and R.S. Hodges, *Correlation of protein retention times in reversed-phase chromatography with polypeptide chain length and hydrophobicity.* J Chromatogr, 1989. **476**: p. 363-75.
81. Requiao, R.D., et al., *Protein charge distribution in proteomes and its impact on translation.* PLoS Comput Biol, 2017. **13**(5): p. e1005549.
82. Chow, C.C., et al., *Chain length dependence of apomyoglobin folding: structural evolution from misfolded sheets to native helices.* Biochemistry, 2003. **42**(23): p. 7090-9.
83. Jha, A.K., et al., *Helix, sheet, and polyproline II frequencies and strong nearest neighbor effects in a restricted coil library.* Biochemistry, 2005. **44**(28): p. 9691-702.
84. Vanaman, T.C., S.J. Wakil, and R.L. Hill, *The complete amino acid sequence of the acyl carrier protein of Escherichia coli.* J Biol Chem, 1968. **243**(24): p. 6420-31.
85. Beld, J., et al., *The phosphopantetheinyl transferases: catalysis of a post-translational modification crucial for life.* Nat Prod Rep, 2014. **31**(1): p. 61-108.
86. Chung, H.S., et al., *Cysteine oxidative posttranslational modifications: emerging regulation in the cardiovascular system.* Circ Res, 2013. **112**(2): p. 382-92.
87. Garcia-Santamarina, S., S. Boronat, and E. Hidalgo, *Reversible cysteine oxidation in hydrogen peroxide sensing and signal transduction.* Biochemistry, 2014. **53**(16): p. 2560-80.
88. Kim, H.J., et al., *ROSics: chemistry and proteomics of cysteine modifications in redox biology.* Mass Spectrom Rev, 2015. **34**(2): p. 184-208.
89. Dalle-Donne, I., et al., *Protein S-glutathionylation: a regulatory device from bacteria to humans.* Trends Biochem Sci, 2009. **34**(2): p. 85-96.
90. Chait, B.T., *Chemistry. Mass spectrometry: bottom-up or top-down?* Science, 2006. **314**(5796): p. 65-6.
91. McCool, E.N., et al., *Deep Top-Down Proteomics Using Capillary Zone Electrophoresis-Tandem Mass Spectrometry: Identification of 5700 Proteoforms from the Escherichia coli Proteome.* Anal Chem, 2018. **90**(9): p. 5529-5533.

92. Zhang, Z.B., et al., *Preparation of linear polyacrylamide coating and strong cationic exchange hybrid monolith in a single capillary, and its application as an automated platform for bottom-up proteomics by capillary electrophoresis-mass spectrometry.* Microchimica Acta, 2017. **184**(3): p. 921-925.
93. Choi, S.B., et al., *Tapered-Tip Capillary Electrophoresis Nano-Electrospray Ionization Mass Spectrometry for Ultrasensitive Proteomics: the Mouse Cortex.* J Am Soc Mass Spectrom, 2017. **28**(4): p. 597-607.
94. Wang, Y., et al., *Improving the comprehensiveness and sensitivity of sheathless capillary electrophoresis-tandem mass spectrometry for proteomic analysis.* Anal Chem, 2012. **84**(20): p. 8505-13.
95. Kwiatkowski, M., et al., *Application of Displacement Chromatography to Online Two-Dimensional Liquid Chromatography Coupled to Tandem Mass Spectrometry Improves Peptide Separation Efficiency and Detectability for the Analysis of Complex Proteomes.* Anal Chem, 2018. **90**(16): p. 9951-9958.
96. Stoll, D.R., et al., *Development of Comprehensive Online Two-Dimensional Liquid Chromatography/Mass Spectrometry Using Hydrophilic Interaction and Reversed-Phase Separations for Rapid and Deep Profiling of Therapeutic Antibodies.* Anal Chem, 2018. **90**(9): p. 5923-5929.
97. Berkecz, R., et al., *Comprehensive phospholipid and sphingomyelin profiling of different brain regions in mouse model of anxiety disorder using online two-dimensional (HILIC/RP)-LC/MS method.* J Pharm Biomed Anal, 2018. **149**: p. 308-317.
98. Zhu, Z., et al., *Two-dimensional chromatographic analysis using three second-dimension columns for continuous comprehensive analysis of intact proteins.* Talanta, 2018. **179**: p. 588-593.
99. Chen, D., X. Shen, and L. Sun, *Capillary zone electrophoresis-mass spectrometry with microliter-scale loading capacity, 140 min separation window and high peak capacity for bottom-up proteomics.* Analyst, 2017. **142**(12): p. 2118-2127.
100. Yu, D., et al., *Deep Intact Proteoform Characterization in Human Cell Lysate Using High-pH and Low-pH Reversed-Phase Liquid Chromatography.* J. Am. Soc. Mass Spectrom., 2019.
101. Baghdady, Y.Z. and K.A. Schug, *Online Comprehensive High pH Reversed Phase x Low pH Reversed Phase Approach for Two-Dimensional Separations of Intact Proteins in Top-Down Proteomics.* Anal Chem, 2019. **91**(17): p. 11085-11091.
102. Cao, J.L., et al., *Online comprehensive two-dimensional hydrophilic interaction chromatographyxreversed-phase liquid chromatography coupled with hybrid linear ion trap Orbitrap mass spectrometry for the analysis of phenolic acids in Salvia miltiorrhiza.* J Chromatogr A, 2018. **1536**: p. 216-227.
103. Sarrut, M., A. D'Attoma, and S. Heinisch, *Optimization of conditions in on-line comprehensive two-dimensional reversed phase liquid chromatography. Experimental comparison with one-dimensional reversed phase liquid chromatography for the separation of peptides.* J Chromatogr A, 2015. **1421**: p. 48-59.
104. Chambers, M.C., et al., *A cross-platform toolkit for mass spectrometry and proteomics.* Nat Biotechnol, 2012. **30**(10): p. 918-20.

105. Kou, Q., L. Xun, and X. Liu, *TopPIC: a software tool for top-down mass spectrometry-based proteoform identification and characterization.* Bioinformatics, 2016. **32**(22): p. 3495-3497.
106. Wang, Y., et al., *Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells.* Proteomics, 2011. **11**(10): p. 2019-26.
107. Cao, J.L., et al., *Comprehensively qualitative and quantitative analysis of ginsenosides in Panax notoginseng leaves by online two-dimensional liquid chromatography coupled to hybrid linear ion trap Orbitrap mass spectrometry with deeply optimized dilution and modulation system.* Anal Chim Acta, 2019. **1079**: p. 237-251.
108. AARDA. *Autoimmune Disease Statistics*. 2018; Available from: https://www.aarda.org/news-information/statistics/.
109. Barturen, G., et al., *Moving towards a molecular taxonomy of autoimmune rheumatic diseases.* Nat Rev Rheumatol, 2018. **14**(2): p. 75-93.
110. Pozsgay, J., Z. Szekanecz, and G. Sarmay, *Antigen-specific immunotherapies in rheumatic diseases.* Nat Rev Rheumatol, 2017. **13**(9): p. 525-537.
111. Cohen, I.R., *Activation of benign autoimmunity as both tumor and autoimmune disease immunotherapy: a comprehensive review.* J Autoimmun, 2014. **54**: p. 112-7.
112. Arbuckle, M.R., et al., *Development of autoantibodies before the clinical onset of systemic lupus erythematosus.* N Engl J Med, 2003. **349**(16): p. 1526-33.
113. Slight-Webb, S., et al., *Autoantibody-Positive Healthy Individuals Display Unique Immune Profiles That May Regulate Autoimmunity.* Arthritis Rheumatol, 2016. **68**(10): p. 2492-502.
114. Boyd, S.D. and J.E. Crowe, Jr., *Deep sequencing and human antibody repertoire analysis.* Curr Opin Immunol, 2016. **40**: p. 103-9.
115. Lavinder, J.J., et al., *Next-generation sequencing and protein mass spectrometry for the comprehensive analysis of human cellular and serum antibody repertoires.* Curr Opin Chem Biol, 2015. **24**: p. 112-20.
116. Chen, B., et al., *Online Hydrophobic Interaction Chromatography-Mass Spectrometry for the Analysis of Intact Monoclonal Antibodies.* Anal Chem, 2018. **90**(12): p. 7135-7138.
117. He, L., et al., *Analysis of Monoclonal Antibodies in Human Serum as a Model for Clinical Monoclonal Gammopathy by Use of 21 Tesla FT-ICR Top-Down and Middle-Down MS/MS.* J Am Soc Mass Spectrom, 2017. **28**(5): p. 827-838.
118. Wu, S., et al., *An integrated top-down and bottom-up strategy for characterization of protein isoforms and modifications.* Methods Mol Biol, 2011. **694**: p. 291-304.
119. Smith, K., et al., *Rapid generation of fully human monoclonal antibodies specific to a vaccinating antigen.* Nat Protoc, 2009. **4**(3): p. 372-84.
120. Liu, X., et al., *Deconvolution and database search of complex tandem mass spectra of intact proteins: a combinatorial approach.* Mol Cell Proteomics, 2010. **9**(12): p. 2772-82.
121. Dekker, L., et al., *An integrated top-down and bottom-up proteomic approach to characterize the antigen-binding fragment of antibodies.* Proteomics, 2014. **14**(10): p. 1239-48.

122. Kou, Q., et al., *Characterization of Proteoforms with Unknown Post-translational Modifications Using the MIScore.* J Proteome Res, 2016. **15**(8): p. 2422-32.
123. Fellers, R.T., et al., *ProSight Lite: graphical software to analyze top-down mass spectrometry data.* Proteomics, 2015. **15**(7): p. 1235-8.
124. Park, J., et al., *Informed-Proteomics: open-source software package for top-down proteomics.* Nat Methods, 2017. **14**(9): p. 909-914.
125. Wu, S., et al., *Top-Down Characterization of the Post-Translationally Modified Intact Periplasmic Proteome from the Bacterium Novosphingobium aromaticivorans.* Int J Proteomics, 2013. **2013**: p. 279590.
126. Wu, S., et al., *Quantitative analysis of human salivary gland-derived intact proteome using top-down mass spectrometry.* Proteomics, 2014. **14**(10): p. 1211-22.
127. Wang, Z., et al., *Two-Dimensional Separation Using High pH and Low pH Reversed Phase Liquid Chromatography for Top-down Proteomics.* International Journal of Mass Spectrometry, 2017.
128. Han, X., et al., *In-line separation by capillary electrophoresis prior to analysis by top-down mass spectrometry enables sensitive characterization of protein complexes.* J Proteome Res, 2014. **13**(12): p. 6078-86.
129. Weisbrod, C.R., et al., *Front-End Electron Transfer Dissociation Coupled to a 21 Tesla FT-ICR Mass Spectrometer for Intact Protein Sequence Analysis.* J Am Soc Mass Spectrom, 2017. **28**(9): p. 1787-1795.
130. Cotham, V.C., et al., *Middle-Down 193-nm Ultraviolet Photodissociation for Unambiguous Antibody Identification and its Implications for Immunoproteomic Analysis.* Anal Chem, 2017. **89**(12): p. 6498-6504.
131. Kaiser, C.A., et al., *Molecular cell biology*. 2007: WH Freeman.
132. Konermann, L., J. Pan, and Y.H. Liu, *Hydrogen exchange mass spectrometry for studying protein structure and dynamics.* Chem Soc Rev, 2011. **40**(3): p. 1224-34.
133. Masson, G.R., et al., *Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (HDX-MS) experiments.* Nat Methods, 2019. **16**(7): p. 595-602.
134. Earl, L.A., et al., *Cryo-EM: beyond the microscope.* Curr Opin Struct Biol, 2017. **46**: p. 71-78.
135. Zhang, Z. and D.L. Smith, *Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation.* Protein Sci, 1993. **2**(4): p. 522-31.
136. Harrison, R.A. and J.R. Engen, *Conformational insight into multi-protein signaling assemblies by hydrogen-deuterium exchange mass spectrometry.* Curr Opin Struct Biol, 2016. **41**: p. 187-193.
137. Moroco, J.A. and J.R. Engen, *Replication in bioanalytical studies with HDX MS: aim as high as possible.* Bioanalysis, 2015. **7**(9): p. 1065-7.
138. Venable, J.D., et al., *Subzero temperature chromatography for reduced back-exchange and improved dynamic range in amide hydrogen/deuterium exchange mass spectrometry.* Anal Chem, 2012. **84**(21): p. 9601-8.
139. Wales, T.E., et al., *Subzero Celsius separations in three-zone temperature controlled hydrogen deuterium exchange mass spectrometry.* J Chromatogr A, 2017. **1523**: p. 275-282.

140. Chalmers, M.J., et al., *Differential hydrogen/deuterium exchange mass spectrometry analysis of protein-ligand interactions.* Expert Rev Proteomics, 2011. **8**(1): p. 43-59.
141. Ahn, J., et al., *Pepsin immobilized on high-strength hybrid particles for continuous flow online digestion at 10,000 psi.* Anal Chem, 2012. **84**(16): p. 7256-62.
142. Zhang, H.M., et al., *Enhanced digestion efficiency, peptide ionization efficiency, and sequence resolution for protein hydrogen/deuterium exchange monitored by Fourier transform ion cyclotron resonance mass spectrometry.* Anal Chem, 2008. **80**(23): p. 9034-41.
143. Tao, Y., et al., *Mapping the contact surfaces in the Lamin A:AIMP3 complex by hydrogen/deuterium exchange FT-ICR mass spectrometry.* PLoS One, 2017. **12**(8): p. e0181869.
144. Zhang, H.M., et al., *Fast reversed-phase liquid chromatography to reduce back exchange and increase throughput in H/D exchange monitored by FT-ICR mass spectrometry.* J Am Soc Mass Spectrom, 2009. **20**(3): p. 520-4.
145. Bai, Y., et al., *Primary structure effects on peptide group hydrogen exchange.* Proteins: Structure, Function, and Bioinformatics, 1993. **17**(1): p. 75-86.
146. Lu, L.L., et al., *Beyond binding: antibody effector functions in infectious diseases.* Nat Rev Immunol, 2018. **18**(1): p. 46-61.
147. Abbott, W.M., M.M. Damschroder, and D.C. Lowe, *Current approaches to fine mapping of antigen-antibody interactions.* Immunology, 2014. **142**(4): p. 526-35.
148. Walters, B.T., et al., *Minimizing Back Exchange in the Hydrogen Exchange-Mass Spectrometry Experiment.* Journal of the American Society for Mass Spectrometry, 2012. **23**(12): p. 2132-2139.
149. Landgraf, R.R., M.J. Chalmers, and P.R. Griffin, *Automated hydrogen/deuterium exchange electron transfer dissociation high resolution mass spectrometry measured at single-amide resolution.* J Am Soc Mass Spectrom, 2012. **23**(2): p. 301-9.